Advancing Scientific Computational Imaging through Data-driven and Physics-based Priors

Thesis by
Berthy Tianyu Feng

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended August 8, 2025

Berthy Tianyu Feng
ORCID: 0000-0002-1843-2165

ACKNOWLEDGEMENTS

First, I would like to thank Katie Bouman. I lucked out when she chose me to be one of her PhD students, even though I had no prior experience in computational imaging. In turn, choosing to join her lab at Caltech was the best decision I could have made for my PhD. She has taught me countless lessons about what it means to be a great researcher, a great teacher, and a great mentor. And I welcome these lessons without hesitation because in my six years of working with Katie, she has been right 99% of the time—it is incredible how many times I initially resisted a project idea, experiment suggestion, or writing edit, only to realize that she was right all along. I admire the calm dedication with which she approaches research. She has taught me to be patient and find joy in the act of research rather than in the outcome. Above all, she has taught me that empathy goes a long way as a researcher, teacher, and mentor.

I would like to thank my thesis committee members: Katie Bouman, Chiara Daraio, Yisong Yue, Georgia Gkioxari, and Bill Freeman. Chiara welcomed me as an honorary member in her lab and brings infectious enthusiasm to research meetings. Yisong has been an indispensable part of my Caltech experience, providing guidance and mentorship since my first day here. I am grateful to Georgia for making Annenberg a lively work environment and for entrusting me with opportunities to organize workshops and panels. I would like to extend a special thanks to Bill Freeman, who welcomed me into his group at Google Research for the summer internship that turned out to be pivotal in my PhD. When I get asked about the type of research I want to do, I often say that I want to do what Bill does: come up with creative ideas and have fun making them come to life.

I would like to thank my wonderful collaborators, including but not limited to Jamie Smith, Ricardo Baptista, Ray Wu, Yu Sun, and Aviad Levis.

Alex Ogren is an especially trusted collaborator and friend. He is incredibly smart, thoughtful, and kind, and he reminds me that research is better with friends. Many of my core Caltech memories involve us working together in the lab, writing papers together, and chatting at the brewery.

The folks in the Bouman Group and Perona Lab inspire me in our group meetings with their wit, curiosity, and friendliness.

Thank you to my first advisors at Princeton, Adam Finkelstein and Olga Russakovsky.

Adam took a chance on an undergrad junior who cold emailed him. Olga encouraged me to get into computer vision, recognizing potential that I didn't know I had.

Thank you to the friends who made this experience so joyful.

Thank you to my family, especially my parents: my mom, Lei Tan, and my dad, Chris Feng. My parents sacrificed a lot so that I can realize my dreams. No matter how far I may go, they will always be my home.

ABSTRACT

The core idea of computational imaging is to supplement limited observable data with human-imposed assumptions, or priors. One could formulate a prior as a statistical model or physics model of the object being imaged. However, incorporating such assumptions in the imaging process poses computational challenges, including efficiently expressing sophisticated priors, appropriately balancing priors with observations, and gently enforcing physics constraints. This thesis addresses such challenges with principled methods for bringing informative assumptions into computational imaging. We emphasize applications in scientific imaging, and we focus on two categories of priors as well as the intersection between them: data-driven statistics and physics knowledge.

On the data-driven side, this thesis presents work on score-based priors, including a posterior-estimation method and results of re-imagining the famous M87* black hole from real data with score-based priors. On the physics-based side, we have been able to tackle extremely under-determined imaging problems by enforcing physics constraints, including performing single-viewpoint dynamic tomography of emission near a black hole and characterizing interior material properties from video. As a means towards integrating data-driven and physics-based assumptions, we have developed a method to enforce physics constraints on generative models. In this thesis, we explore each aforementioned project with an emphasis on technical novelty and experimental validation on simulated and real data. By opening new routes for bringing in both data-driven and physics-based assumptions, the methods presented in this thesis enable visualizing scientific phenomena beyond the reach of conventional sensors.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Berthy T. Feng. "EHT Black-Hole Imaging". Tutorial. Tutorial. ICCP Summer School. Toronto, ON, CA, July 19, 2025. URL: https://github.com/berthyf96/eht_imaging_tutorial.
B.T.F. developed all the course slides and code notebooks and taught the course.

[2] Berthy T. Feng, Ricardo Baptista, and Katherine L. Bouman. "Neural Approximate Mirror Maps for Constrained Diffusion Models". In: The Thirteenth International Conference on Learning Representations. Singapore, 2025. URL: https://openreview.net/forum?id=vgZDcUetWS.
B.T.F. formulated the problem statement, created the method, performed all experiments, and wrote the paper.

[3] Berthy T. Feng and Katherine L. Bouman. "Seeing Beyond the Blur with Generative AI". In: *XRDS* 31.2 (Jan. 8, 2025), pp. 28–31. ISSN: 1528-4972. DOI: 10.1145/3703400. URL: https://dl.acm.org/doi/10.1145/3703400.
B.T.F. conceptualized and wrote the article.

[4] Alexander C. Ogren et al. "Visual Surface Wave Elastography: Revealing Subsurface Physical Properties via Visible Surface Waves". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2025.
B.T.F. helped with the development of the method and experiments and led the writing of the paper.

[5] Bingliang Zhang et al. *STeP: A Framework for Solving Scientific Video Inverse Problems with Spatiotemporal Diffusion Priors*. June 10, 2025. DOI: 10.48550/arXiv.2504.07549. arXiv: 2504.07549[cs]. URL: http://arxiv.org/abs/2504.07549.
B.T.F. provided feedback on the method, experiments, and manuscript.

[6] Sreemanti Dey et al. "Score-Based Diffusion Models for Photoacoustic Tomography Image Reconstruction". In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 2470–2474.
B.T.F. served a mentorship role on this project, helping with the method development, experiments, and writing of the manuscript.

[7] Berthy T. Feng and Katherine L. Bouman. "Variational Bayesian Imaging with an Efficient Surrogate Score-based Prior". In: *Transactions on Machine Learning Research* (Mar. 2, 2024). ISSN: 2835-8856. URL: https://openreview.net/forum?id=db2pFKVcm1.
B.T.F. formulated the problem statement, created the method, performed all experiments, and wrote the manuscript.

[8] Berthy T. Feng, Katherine L. Bouman, and William T. Freeman. "Event-horizon-scale Imaging of M87* under Different Assumptions via Deep Generative Image Priors". In: *The Astrophysical Journal* 975.2 (Nov. 2024). Publisher: The American Astronomical Society, p. 201. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ad737f. URL: https://dx.doi.org/10.3847/1538-4357/ad737f.
B.T.F. developed and performed all experiments and wrote the manuscript.

[9] Alexander C. Ogren et al. "Gaussian process regression as a surrogate model for the computation of dispersion relations". In: *Computer Methods in Applied Mechanics and Engineering* 420 (Feb. 15, 2024), p. 116661. ISSN: 0045-7825. DOI: 10.1016/j.cma.2023.116661. URL: https://www.sciencedirect.com/science/article/pii/S0045782523007843.
B.T.F. participated in research discussions and helped with the method development, experiments, and manuscript writing.

[10] Yu Sun et al. "Provable Probabilistic Imaging Using Score-Based Generative Priors". In: *IEEE Transactions on Computational Imaging* 10 (2024), pp. 1290–1305. ISSN: 2333-9403. DOI: 10.1109/TCI.2024.3449114. URL: https://ieeexplore.ieee.org/document/10645293.
B.T.F. provided code and baseline comparisons for the black-hole imaging experiments, and she provided feedback on the manuscript.

[11] Hongkai Zheng et al. "InverseBench: Benchmarking Plug-and-Play Diffusion Priors for Inverse Problems in Physical Sciences". In: The Thirteenth International Conference on Learning Representations. Oct. 4, 2024. URL: https://openreview.net/forum?id=U3PBITXNG6.
B.T.F. assisted with the conceptualization of the project and performed baseline experiments for the black-hole imaging task.

[12] Berthy T. Feng et al. "Score-Based Diffusion Models as Principled Priors for Inverse Imaging". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (ICCV)*. Paris, France: IEEE, 2023. DOI: 10.1109/ICCV51070.2023.00965.
B.T.F. led the project, including method development and experiments, and wrote the paper.

[13] Berthy T. Feng et al. "Visual Vibration Tomography: Estimating Interior Material Properties from Monocular Video". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022*. ISSN: 2575-7075. IEEE, June 2022, pp. 16210–16219. DOI: 10.1109/CVPR52688.2022.01575. URL: https://ieeexplore.ieee.org/document/9880380.
B.T.F. led the project, including method development and experiments, and wrote the paper.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

In the quest for scientific knowledge, seeing is believing. Whether it is viruses, weather systems, or exoplanets, making the unseen visible helps inform, motivate, and educate. Pushing the boundaries of imaging therefore goes hand-in-hand with pushing the boundaries of scientific knowledge.

However, as the boundaries of science continue to expand, the demands on imaging outpace the capabilities of direct imaging with sensors. Even though physical sensors can capture some amount of data, that information is often incomplete, highly corrupted, or both. For example, in astronomy, phenomena such as black holes and exoplanets are so distant that telescopes on Earth are nowhere near large enough to fully resolve them, leading to incomplete data. Furthermore, atmospheric turbulence and sensor aberrations corrupt the little amount of data that is available. To overcome the limitations of physical sensors, we need to fill in the missing information. Computational imaging aims to visualize scientific phenomena beyond the reach of conventional cameras by incorporating additional assumptions about the object being imaged.

As humans and scientists we often carry priors on images. We refer to imaging assumptions as priors because they describe image characteristics that are expected *a priori*, before any data is observed. These priors can range from the statistics of natural images that most people observe to domain-specific scientific knowledge. When imaging a black hole, for instance, we may wish to enforce spatial correlations that would make the image appear similar to a natural photograph, or we may rather make the image fit a physics model of a black hole. As such, two powerful categories of priors emerge: data-driven priors and physics-based priors. Data-driven priors integrate the statistics of large-scale image datasets, while physics-based priors impose constraints rooted in physical equations. Both categories require thoughtfully-designed imaging algorithms that integrate prior knowledge with observed data in a principled way. This thesis explores the two paradigms of data-driven priors and physics-based priors, focusing on principled methods for applying them to scientific imaging problems.

Our work has established new avenues for leveraging data-driven and physics-based priors. On the data-driven side, we developed a method for using diffusion models, a state-of-the-art type of generative model, as sophisticated data-driven priors for general imaging problems [118, 116]. We applied this to image the M87* black hole to target a wider range of imaging assumptions than previously possible [117]. On the physics-based side, we developed a method for recovering the dynamic 3D gas

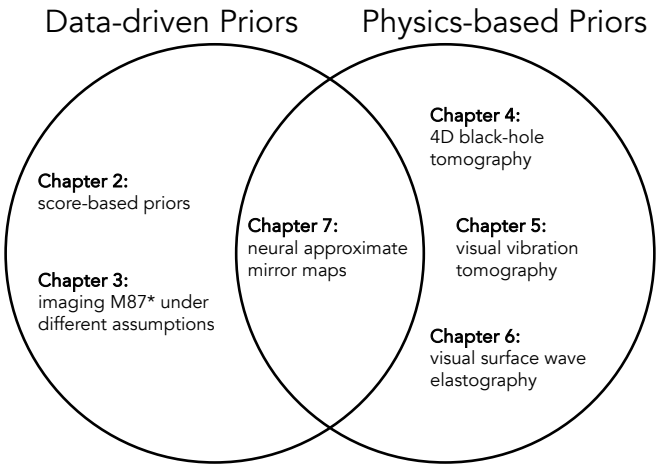


Figure 1.1: Venn diagram organizing the thesis chapters according to two paradigms of priors for scientific computational imaging. Both data-driven priors and physics-based priors help overcome extreme limitations in observable data when it comes to scientific imaging.

around a black hole given only a single 2D viewpoint by making use of soft and hard physics constraints. We also addressed problems in materials science with physics-based optimization approaches for material characterization from videos [119, 242]. We even created an approach for merging the two paradigms of data-driven priors and physics-based priors by enforcing physics constraints in deep generative models [114]. Figure 1.1 organizes the chapters in this thesis into a Venn diagram of data-driven and physics-based priors. Before covering these topics throughout the thesis, this section covers background on how to formalize imaging as an inverse problem in Section 1.0.1 and the different types of priors used in related work in Section 1.0.2.

1.0.1 Imaging as an inverse problem

Imaging can be framed as an inverse problem in which a hidden image $\boldsymbol{x}^* \in \mathbb{R}^D$ must be recovered from measurements $\boldsymbol{y} \in \mathbb{R}^M$, where

$$\boldsymbol{y} = f(\boldsymbol{x}^*) + \boldsymbol{n}. \tag{1.1}$$

Here $f : \mathbb{R}^D \to \mathbb{R}^M$ is the forward model, and $\boldsymbol{n} \in \mathbb{R}^M$ is the measurement noise. The forward model is typically derived from knowledge of the physical sensing system and describes how the underlying image gets transformed into a set of measurements. Although the additive noise is not exactly known, we assume that its statistics are known (often we assume that $\boldsymbol{n}$ is a Gaussian random variable).

1.0.1.1 Bayesian inference

Bayesian imaging accounts for the uncertainty added by the measurement process in Equation 1.1 by formulating a posterior distribution $p(\boldsymbol{x} \mid \boldsymbol{y})$. By Bayes' rule, the log-density of the prior comprises a likelihood term and a prior term:

$$\log p(\boldsymbol{x} \mid \boldsymbol{y}) = \log p(\boldsymbol{y} \mid \boldsymbol{x}) + \log p(\boldsymbol{x}) + \text{const.} \tag{1.2}$$

The log-likelihood function $\log p(\boldsymbol{y} \mid \boldsymbol{x})$ depends on the forward model and measurement noise statistics, whereas the prior log-probability function $\log p(\boldsymbol{x})$ is independent of the imaging task at hand. Given $\log p(\boldsymbol{y} \mid \boldsymbol{x})$ and $\log p(\boldsymbol{x})$, we can use optimization or sampling techniques to find images from the posterior.

MAP estimation Many imaging approaches are based on estimating the *maximum a posteriori* (MAP), which is the image that maximizes the log-posterior density:

$$\boldsymbol{x}_{\text{MAP}} = \arg\max_{\boldsymbol{x}} \left\{ \log p(\boldsymbol{y} \mid \boldsymbol{x}) + \log p(\boldsymbol{x}) \right\}. \tag{1.3}$$

Often the prior takes the form of regularization, such as total variation (TV) [27], maximum entropy [143], or L1 sparsity [59]. Equation 1.3 can be solved via gradient-based optimization [250, 28].

Posterior sampling Often it would be better to sample from the image posterior rather than solve for a single image. Posterior sampling offers a greater diversity of possible image reconstructions and supports downstream tasks like uncertainty quantification. We can use established techniques for sampling from the posterior, such as Markov chain Monte Carlo (MCMC) [55] or variational inference (VI) [40]. MCMC algorithms generate a Markov chain whose stationary distribution is the posterior, but they are generally slow to converge for high-dimensional data like images. VI instead approximates the posterior with a tractable distribution. The variational distribution is usually parameterized to efficiently represent high-dimensional data distributions. Our work on data-driven priors leverages VI [118, 116, 117].

1.0.2 Types of priors in previous work

Assuming a known likelihood, the challenge is defining a prior on images that reflects their complicated statistics. Traditional imaging uses handcrafted regularizers that have simple analytical forms. With advances in deep learning, many methods have been proposed that leverage either the implicit prior of a neural network or a deep generative model as the prior.

1.0.2.1 Regularizers

Regularization-based approaches enforce simple assumptions about natural images. Examples include total variation (TV) and total squared variation (TSV) for spatial smoothness [27, 45] and L1 norm for sparsity [59]. The weightings of these regularizers are typically set by hand.

1.0.2.2 Data-driven priors

Classical examples of data-driven models are Gaussian mixture models [361, 362] and independent components analysis (ICA) [31, 170]. These are simplified, data-driven generative models of images, and since they have analytical probability functions $p(\boldsymbol{x})$, they can be easily incorporated as a prior in Equation 1.2. On the other hand, deep-learned priors enable much more sophisticated image reconstruction.

Implicit deep priors End-to-end learning approaches involve training a neural network on a paired dataset of images and measurements [248, 171, 343, 351, 350, 342, 263, 95, 94]. They are effective when there is a sufficient amount of paired data and when measurements at inference time are not too different from the training data. However, they require re-training for new tasks, and uncertainty cannot be analyzed under the prior implicitly learned by the network. Bayesian networks [127] can account for uncertainty, but they have not been demonstrated on complicated posteriors. Deep Image Prior [311] showed that the inductive bias of a convolutional neural network (CNN) can act as an implicit prior. Other methods like Plug-and-Play (PnP) [316] and Regularization by Denoising (RED) [258] use an image denoiser as an implicit regularizer to provide a point solution.

Data-driven generative models Deep generative models, including generative adversarial networks (GANs) [138] and variational autoencoders (VAEs) [187, 255] have proven effective at generating visually impressive images, leading to some approaches that solve inverse problems by finding a solution in the range of the generative model that agrees with the measurements [43, 147, 148]. Flow-based models like discrete normalizing flows [186, 101, 156, 140, 30, 100, 233] and autoregressive flows [199, 136, 314, 244] support exact log-probability computation [22] so can be easily plugged into Equation 1.2, but they are restricted to certain network architectures and tend to not generalize well outside of training data [189, 233]. Diffusion models [155, 188, 279, 239, 283] have recently emerged as the

state-of-the-art in generative modeling [99] and have already been demonstrated as effective priors for solving inverse problems [355].

1.0.2.3 Physics-based priors

Any imaging method that makes use of physics knowledge, whether as a hard constraint or soft constraint, can be viewed as using a physics-based prior. Unlike with data-driven priors, physics knowledge often gets embedded into the imaging process in ways besides plugging it in as a prior in the Bayesian formula (Equation 1.2). Here we list some broad approaches to incorporating physics.

Model-fitting Model-fitting imposes hard physics constraints by fitting parameters of a physical model to observed measurements. This has been done for black-hole imaging [8, 322, 132, 181, 344, 238] and is one of the most constraining ways to approach imaging, as it requires defining a model that can fully describe the object being imaged and forces the reconstructed image to agree with that model even if the observed measurements are imperfect.

Hard physics constraints Constraints can be enforced in optimization-based [48, 93] and sampling-based approaches [160] via feasible sets. For example, one can propose candidate solutions that satisfy the desired physics equation [202]. When it comes to deep-learning-based approaches, one approach is to build inductive biases into the neural network to satisfy desired constraints [179]. For example, equivariant neural networks [266, 133, 301] enforce symmetry constraints [15, 240, 190, 285, 49, 256, 265, 192, 227, 97, 338, 161, 341]. One can also modify the architecture of diffusion models to satisfy certain types of constraints [44, 165, 121, 215, 211, 122], such as those that can be characterized as a Riemannian manifold [44, 165] or a convex constraint set [121, 211]. Such neural networks can be used to solve inverse problems in an end-to-end manner or as deep generative priors.

Soft physics constraints For optimization-based approaches, the constraint can be relaxed as an additional penalty term [34]. Physics-informed neural networks (PINNs) [254] are trained to solve inverse problems with an additional loss that enforces physics as a soft constraint. Other approaches include adding constraint-violating examples during training [137] and using reinforcement learning to align pretrained models to reward functions that encourage satisfying constraints [290, 358].

Part I

Data-driven Priors

Chapter 2

SCORE-BASED DIFFUSION MODELS AS PRINCIPLED PRIORS

The inverse problem of imaging become more ill-posed as imaging sensors exceed their limits, making priors even more crucical in the imaging process. When many different images agree with observed measurements, a prior helps constrain solutions according to desired image statistics. How to incorporate a sophisticated prior, however, is not straightforward. In this chapter, we address the problem of incorporating a rich prior into principled approaches to inverse problems by leveraging a diffusion model as a fully data-driven and powerful prior.

Previous methods pose a tradeoff: principled methods require simple priors, while deep-learned priors preclude precise analysis. On the principled side, recall that Bayesian-inference methods model the posterior distribution of images $\boldsymbol{x}$ conditioned on measurements $\boldsymbol{y}$:

$$p(\boldsymbol{x} \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{x})\, p(\boldsymbol{x}).$$

This Bayesian framework supports a modular approach to inverse problems where the likelihood $p(\boldsymbol{y} \mid \boldsymbol{x})$ is defined by an expert based on knowledge of how measurements are obtained, and the prior $p(\boldsymbol{x})$ is defined independently. Furthermore, it allows for principled solutions. MAP estimation can be done by optimizing the posterior probability. Posterior sampling can be done with MCMC or VI. But since such methods require the value or gradient of $p(\boldsymbol{x})$, they have been limited to simple priors (e.g., Gaussian) and weighted regularizers (e.g., total variation). In practice, the relative weights of the prior and likelihood terms are usually tuned by hand, introducing a human bias that is unsatisfactory for scientific applications.

On the deep-learning side, solutions leveraging an implicit, deep-learned prior may look convincing but do not lend themselves to principled analysis. For example, a CNN can be trained in a supervised way to output images given measurements, but its prior cannot be probed and does not generalize to new tasks. Previous work showed how to condition a diffusion model—a type of generative model whose prior is captured in a learned image denoiser—on arbitrary measurements [79, 80, 77, 78, 139, 280, 284], but the methods depend on hand-tuned hyperparameters and do not sample from a true posterior except in auspicious cases. To get the best of both worlds (traditional Bayesian inference and modern deep learning), we need a way to incorporate the expressive prior of a deep-learned model into a traditional, principled Bayesian approach.

We propose employing a diffusion model as the prior in Bayesian inference for imaging. A *score-based prior* is the distribution under a score-based diffusion

model [283], which is the most sophisticated type of generative model to-date and has been proven to allow for exact probabilities (a feature under-explored in practice). The main contribution of our work is establishing score-based priors as an interface between modern deep-learning and traditional inverse problem-solving, giving proven, principled approaches direct access to learned, rich priors. Under our framework, we train a score-based prior once on a dataset of images. Paired with any likelihood $p(\boldsymbol{y} \mid \boldsymbol{x})$, this prior can be plugged into any inference algorithm that uses the value or gradient of the posterior.

We first present background on diffusion models in Section 2.1. We then discuss a different class of methods for solving inverse problems with diffusion models and their limitations in Section 2.2. In Section 2.3, we define and validate the log-probability function of a score-based diffusion model. In Section 2.4, we propose an approach to use this log-probability function in a principled VI framework for posterior estimation. In Section 2.5, we present results of posterior sampling for various imaging problems. We propose a more efficient surrogate for the score-based prior in Section 2.6, and we present results on improved efficiency and image reconstruction with the surrogate in Section 2.7.

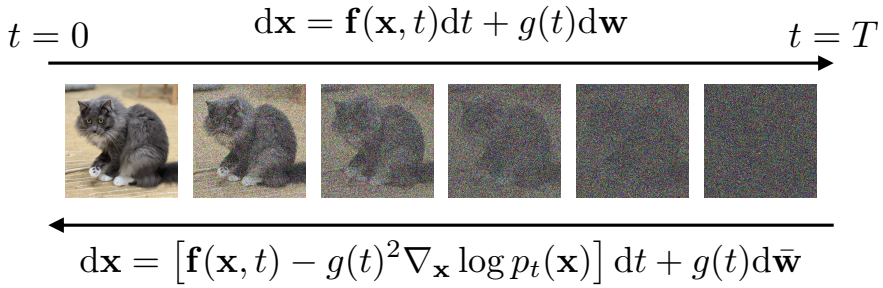


Figure 2.1: An illustration of the diffusion process. The forward-time SDE dictates how to transform an image distribution p_{data} at $t = 0$ into a noise distribution at $t = T$. The reverse-time SDE dictates how to transform the noise distribution $\mathcal{N}(\mathbf{0}_D, \boldsymbol{I}_D)$ into a clean image distribution p_0 that is indistinguishable from p_{data}.

2.1 Diffusion Models

A diffusion model transforms a simple distribution into a complex one, with image generation usually done by gradually denoising a sample from a Gaussian distribution until it becomes a clean image. Denoising diffusion probabilistic models (DDPMs) [277, 155] treat this transformation as a discrete-time process with a fixed number of denoising steps. Score-based diffusion models [281] generalize to continuous time. We present diffusion models within the score-based framework.

2.1.1 Score-based diffusion models

In the continuous-time setting, a stochastic differential equation (SDE) describes how the data distribution should look at each step of denoising. Each noise level is associated with a diffusion time t, where a larger t corresponds to more noise present in the image. A *forward-time SDE* defines the diffusion process of an image $\boldsymbol{x}(0)$ from $t = 0$ to $t = T$:

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}, \tag{2.1}$$

where $\boldsymbol{w} \in \mathbb{R}^D$ denotes Brownian motion; $g(t) \in \mathbb{R}$ is the diffusion coefficient, which controls the rate of noise increase; and $\boldsymbol{f}(\cdot, t) : \mathbb{R}^D \to \mathbb{R}^D$ is the drift coefficient, which controls the deterministic evolution of $\boldsymbol{x}(t)$. By defining a stochastic trajectory $\{\boldsymbol{x}(t)\}_{t\in[0,T]}$, this SDE gives rise to a time-dependent probability distribution p_t, which is the marginal distribution of $\boldsymbol{x}(t)$. We construct $\boldsymbol{f}(\cdot, t)$ and $g(t)$ so that if $p_0 = p_{\text{data}}$, then $p_T \approx \pi$.

Sampling an image from p_0 requires reversing the diffusion process. The *reverse-time SDE* [18] is given by

$$\mathrm{d}\boldsymbol{x} = \left[\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2 \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}}. \tag{2.2}$$

The problematic term here is $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$, which is the (Stein) score of $\boldsymbol{x}$ under p_t. In words, undoing diffusion is difficult because it requires knowing how to nudge each perturbed distribution p_t closer to the clean data distribution. Being the only data-dependent component of this SDE, the score function is learned by a neural network $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$ with parameters $\boldsymbol{\theta}$ [283, 282]. Because each p_t is p_0 perturbed by Gaussian noise, the time-dependent score model $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$ can be thought-of as an image denoiser: it takes a noisy image as input and estimates the noise, and t indicates the level of noise applied. To sample a new image, a noise image $\boldsymbol{x}(T) \sim \mathcal{N}(\mathbf{0}_D, \boldsymbol{I}_D)$ is drawn whose final state $\boldsymbol{x}(0)$ is given by solving the reverse-time SDE with the assumption that $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) \approx \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$. If the score model is well trained, then we can assume that $p_0 \approx p_{\text{data}}$.

2.2 Related Work on Diffusion Models for Inverse Problems

Many methods have been proposed to leverage pretrained diffusion models as priors for inverse problems, but they operate within a completely different framework than ours. These methods generally apply the denoising network within the diffusion model as an implicit prior, iteratively alternating between denoising steps and measurement-consistency steps. This concept is similar to the idea of "plug-and-play" (PnP) denoising priors [316] and regularization by denoising (RED) [258], whereby a sequence of steps consisting of denoising and enforcement of measurement consistency gradually transforms an initial guess of an image into a clean image. Diffusion models are preferable because they offer an alternative theoretical framework for employing the denoiser while typically resulting in richer images than a classical PnP/RED approach.

We use the term *plug-and-play diffusion priors* (PnPDPs) to refer to the class of methods that use diffusion models to solve inverse problems specifically by incorporating both the measurement likelihood and the denoiser of the diffusion model in an iterative process (often according to the diffusion SDE). In contrast, our approach directly optimizes the posterior probability that arises from the measurement likelihood and the prior probability under the diffusion model. A major limitation of PnPDP methods is that they entail many hyperparameters that need to be tuned, whereas our approach is hyperparameter-free.

In this section, we cover the three main categories of PnPDP methods: guidance-based, variable splitting, and sequential Monte Carlo. To evaluate PnPDP methods in a systematic and easily-extensible manner, we introduced a comprehensive benchmarking framework known as InverseBench [355]. We also note that while this chapter focuses on imaging, PnPDP methods have been applied to inverse problems in physical sciences [355] and video inverse problems [349].

Guidance-based methods Arguably the most popular approach to solving inverse problems with a pretrained diffusion model is guidance-based methods [280, 326, 180, 260, 79, 98], which modify Equation 2.2 by adding a likelihood score term, $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{y}|\boldsymbol{x})$, along the diffusion trajectory. However, guidance-based methods are fundamentally limited due to the intractibility of $\log p_t(\boldsymbol{y}|\boldsymbol{x}(t))$ [144]. Various approximations have been proposed [284, 79, 280, 50], but as we show in this chapter, they may lead to inaccurate estimations of the posterior.

Variable splitting Variable splitting is a widely-used strategy for solving regularized optimization problems and conducting Bayesian inference [320, 72, 200]. The core idea is to split the inference into two alternating steps [334, 360, 206, 278, 348, 339]. The first step uses the forward model to update or sample in the neighborhood of the most recent $\boldsymbol{x}(t)$. The second step runs unconditional inference on $p(\boldsymbol{x}(t))$, which amounts to running Equation 2.2 for a small amount of time.

Sequential Monte Carlo Sequential Monte Carlo (SMC) methods draw samples iteratively from a sequence of probability distributions. These methods represent probability distributions by a set of particles with associated weights, which asymptotically converge to a target distribution following a sequence of proposal and reweighting steps. Recent work extended SMC methods to the sequential diffusion sampling process [332, 306, 61, 103], enabling zero-shot posterior sampling with diffusion priors. However, these methods are typically applicable only to inverse problems with linear forward models.

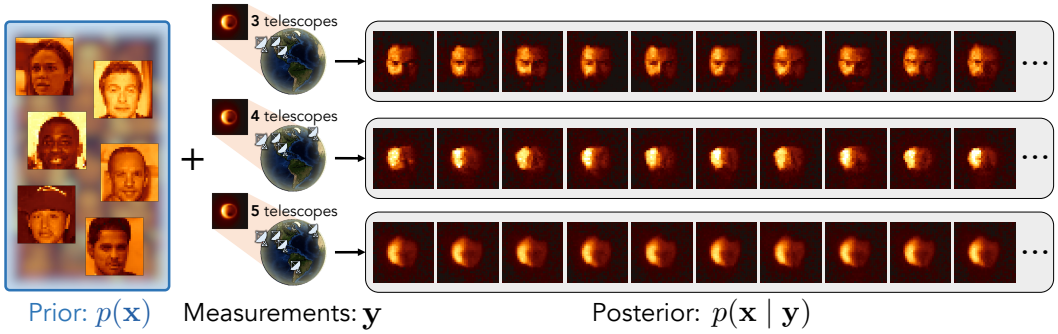


Figure 2.2: A score-based prior is a **hyperparameter-free, probabilistic prior** that is also **expressive and data-driven**. Paired with a set of measurements, the prior can be used for principled inference of a full posterior. In this example, a score-based prior was trained on face images ("Prior" shows samples from the learned prior). The inverse problem is interferometic imaging of a synthetic black hole. We simulated interferometric measurements from the actual telescope array used to capture the first black-hole image [7] and sampled images from the posterior via VI. From the top to bottom row, the posterior stably moves away from the prior given more constraining measurements. With measurements from only three telescopes, the posterior shows strong influence from the prior and contains images resembling faces that are brighter on the left half. As more telescopes (measurements) are added, the posterior reveals the ring-like structure of the underlying image. Our framework finds the proper relative strengths of the prior and measurements automatically.

2.3 Score-based Priors

We propose score-based priors as differentiable image priors that can be trained on any dataset and employed for principled inverse imaging. For example, one can learn a prior on face images by training a score model $\boldsymbol{s}_{\theta}(\boldsymbol{x}, t)$ on the CelebA [213] dataset of faces. Then one can model any posterior with a face prior by appealing to the function $\log p_{\theta}(\boldsymbol{x})$. See Figure 2.2 for an example of our method applied to black-hole imaging with such a face prior.

2.3.1 Log-probability computation

Our method leverages previous work that shows how to compute image probabilities under the score-based SDE framework [283]. Prior to our work, this feature was mostly discussed in theory but not demonstrated in practice. In this section, we discuss log-probability computation under a score-based diffusion model and empirically validate it.

Probability flow ODE Computing probabilities requires inverting the sampling process of a diffusion model: the probability of an image $\boldsymbol{x}(0)$ depends on the

probability of the $\boldsymbol{x}(T)$ that would have resulted in that image through reverse diffusion. There is an ordinary differential equation (ODE) known as the *probability flow ODE* [283] that makes both the forward and reverse SDEs (Equations 2.1 and 2.2) invertible, inducing the same time-dependent probability distribution p_t but without Brownian motion. The ODE is the same forward and backward in time, defining a bijective mapping between p_t and $p_{t'}$ for any two times $t, t' \in [0, T]$.[1] For a score model $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) \approx \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$, the learned probability flow ODE is given by

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{x}, t) - \frac{1}{2} g(t)^2 \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) =: \tilde{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}, t). \tag{2.3}$$

Log-probability formula By the continuous-time change-of-variables formula [71], the log-probability of an image $\boldsymbol{x} = \boldsymbol{x}(0)$ under the p_0 distribution is given by the log-probability of $\boldsymbol{x}(T)$ under the $p_T = \pi$ Gaussian, plus a normalization factor accounting for the change in probability density from $\boldsymbol{x}(0)$ to $\boldsymbol{x}(T)$. We compute the log-probability under the ODE (Equation 2.3) by solving an initial-value problem:

$$\log p_0(\boldsymbol{x}(0)) = \log \pi(\boldsymbol{x}(T)) + \int_0^T \nabla \cdot \tilde{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}(t), t) \mathrm{d}t, \tag{2.4}$$

where $\boldsymbol{x}(0) = \boldsymbol{x}$. The divergence $\nabla \cdot \tilde{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}(t), t)$ quantifies the instantaneous change in log-probability of $\boldsymbol{x}(t)$ caused by applying $\tilde{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}(t), t)$ in either time direction. It can be estimated with Hutchinson-Skilling estimation of the trace of $\frac{\partial}{\partial \boldsymbol{x}(t)} \tilde{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}(t), t)$ [283]. We denote the log-probability function of a score-based prior as $\log p_{\boldsymbol{\theta}} := \log p_0$.

2.3.2 Log-probability validation

In several experiments (Figures 2.3, 2.4, and 2.5), we compare a learned score-based prior to a known ground-truth distribution. The ground-truth is a Gaussian distribution of 16×16 grayscale images, whose mean and preconditioned covariance were fit to CelebA face images. The score model was trained on samples from this Gaussian. Figure 2.3 shows our empirical analysis of log-probabilities under the score-based prior versus the true prior. We verify accuracy on both in-distribution and out-of-distribution images.

Gradients As many Bayesian-inference approaches require a *differentiable* prior, we validate that the gradient function $\nabla_{\boldsymbol{x}} \log p_{\boldsymbol{\theta}}(\boldsymbol{x})$ provides more accurate gradients

[1] The implicit assumption is that the score model is well-trained such that $p_0 \approx p_{\text{data}}$, where p_{data} is the true distribution of the training data.

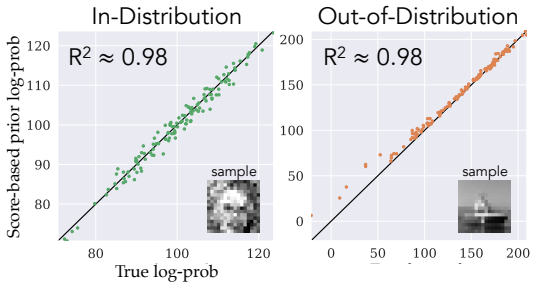


Figure 2.3: Log-probabilities of the score-based prior vs. ground-truth. Black line indicates perfect agreement. **In-Distribution.** The log-probabilities of 128 samples from the Gaussian ground-truth distribution were evaluated (shown as scatter points). Score-based log-probabilities are strongly correlated with ground-truth log-probabilities ($R^2 \approx 0.98$). **Out-of-Distribution.** The log-probabilities of test images from CIFAR-10 (scaled to 16×16) are shown. The score-based prior generalizes well out of distribution.

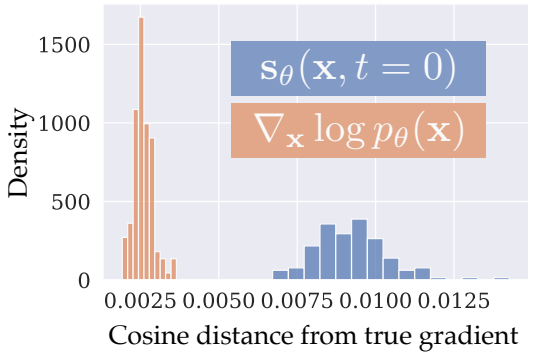


Figure 2.4: ODE gradients vs. score-model outputs. Histogram shows density of cosine distance between the estimated gradient and true gradient for 128 samples from the ground-truth Gaussian. Here $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t = 0)$ is the score-model-approximated gradient ($t = 0.001$ is actually used for numerical stability), and $\nabla_{\boldsymbol{x}} \log p_{\boldsymbol{\theta}}(\boldsymbol{x})$ is computed numerically according to the probability flow ODE.

than the score model $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t = 0) \approx \nabla_{\boldsymbol{x}} \log p_0(\boldsymbol{x})$. It is tempting to apply the score model as a cheap gradient approximator since it is designed in theory as such, but in practice, it does not generalize to out-of-distribution data. It is more reliable to compute the gradient of the entire ODE solve of $\log p_{\boldsymbol{\theta}}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$. Figure 2.4 shows that given samples from the ground-truth Gaussian, gradients computed according to the ODE are closer to the true gradients (in terms of cosine similarity) than score-model outputs. Figure 2.5 (appearing in Section 2.4) shows that using score-model outputs as gradients leads to an incorrect posterior.

2.4 Posterior Sampling with Score-based Priors

We accomplish posterior sampling by plugging the score-based prior into a VI method. This both highlights the applicability of score-based priors to established optimization methods and provides a solution to the open problem of posterior sampling with diffusion-model priors. In this section, we describe our chosen approach, *Deep Probabilistic Imaging* (DPI), which converges faster than MCMC and provides efficient sampling [288].

2.4.1 Main idea for posterior sampling

Our main idea for unlocking the rich prior of a diffusion model is to remove the concept of diffusion and consider the prior as a fixed distribution just like the likelihood. This lets us directly model the posterior $\log p(\boldsymbol{x}|\boldsymbol{y})$.

As discussed in Section 2.2, guidance-based PnPDP methods face an intractable time-dependent score function. Posterior reverse diffusion requires the *posterior* score function

$$\nabla_{\boldsymbol{x}(t)} \log p_t(\boldsymbol{x}(t) \mid \boldsymbol{y}) = \nabla_{\boldsymbol{x}(t)} \log p_t(\boldsymbol{y} \mid \boldsymbol{x}(t)) + \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}(t), t)$$

for all $t \in [0, T]$. The score model gives the prior score, but the likelihood score is only defined for $t = 0$. For every diffusion time t, the true likelihood is defined by an intractable integral over all $\boldsymbol{x}(0)$:

$$p_t(\boldsymbol{y} \mid \boldsymbol{x}(t)) = \int_{\boldsymbol{x}(0)} p(\boldsymbol{y} \mid \boldsymbol{x}(0)) p(\boldsymbol{x}(0) \mid \boldsymbol{x}(t)) \mathrm{d}\boldsymbol{x}(0).$$

To circumvent this challenge, previous methods either abandon the true measurement uncertainty [74, 80, 77, 78, 139, 284, 3] or strongly approximate $p_t(\boldsymbol{y} \mid \boldsymbol{x}(t))$ [79, 173, 180, 280]. This necessitates hyperparameter(s) for determining the importance of measurements versus the prior, making the estimated posterior more of a *conditional* distribution rather than a principled posterior. Different hyperparameter settings have drastic effects on the estimated posterior: if measurements are under-weighted, then the posterior is overly-biased toward the prior and may contain misleading data; if measurements are over-weighted, then the samples may collapse onto a subspace that does not make sense under the prior. Figure 2.16 in Section 2.7 demonstrates that even with ideally-tuned hyperparameters (which require knowledge of ground-truth), previous methods cannot capture the true posterior.

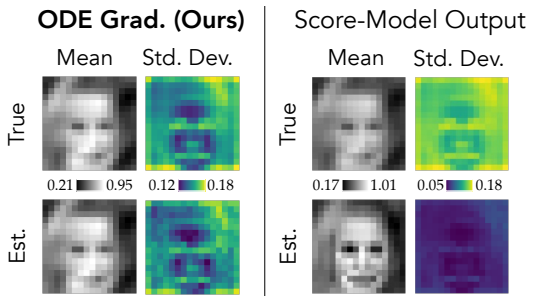


Figure 2.5: True vs. estimated posterior. Measurements are the 6.25% lowest DFT spatial-frequencies of an image from the prior and have i.i.d. noise with $\sigma = |1|$. The true mean and variance were derived analytically since the prior is the Gaussian ground-truth distribution, and the likelihood distribution is also Gaussian. The mean and variance were estimated from 10240 samples.

2.4.2 Selected approach: variational inference with a normalizing flow

Our approach for directly modeling the posterior is rooted in VI. Following the method proposed in DPI [288], we define a family of distributions q_ϕ via a RealNVP [101] normalizing flow with parameters ϕ, which we optimize to approximate the desired posterior. For a score-based prior with parameters $\boldsymbol{\theta}$, the objective is

$$\begin{aligned} \phi^* &= \arg\min_\phi D_{\mathrm{KL}}\left(q_\phi \| p_{\boldsymbol{\theta}}(\cdot \mid \boldsymbol{y})\right) \\ &= \arg\min_\phi \mathbb{E}_{\boldsymbol{x}\sim q_\phi}\left[-\log p(\boldsymbol{y} \mid \boldsymbol{x}) - \log p_{\boldsymbol{\theta}}(\boldsymbol{x}) + \log q_\phi(\boldsymbol{x})\right]. \end{aligned} \tag{2.5}$$

This objective includes the log-posterior and an entropy term, $\mathbb{E}_{\boldsymbol{x}\sim q_\phi}[\log q_\phi(\boldsymbol{x})]$, which can be tractably computed under the RealNVP. During fitting, the expectation is Monte-Carlo approximated with a batch of samples from the RealNVP. The **absence of hyperparameters** in Equation 2.5 is a feature of the score-based prior. Because our prior is truly probabilistic, there is no need for a hand-tuned weight.

As evidence that our method correctly samples the posterior, Figure 2.5 compares an estimated posterior to ground-truth. The score-based prior was trained on the ground-truth Gaussian used in Section 2.3, and the task was to deblur an image from the true prior. The estimated mean and covariance closely agree with the analytical mean and covariance. Additionally, we find that directly using the score-model output leads to an incorrect posterior.

We note that accurate posterior sampling is dependent on accurate log-probabilities in the first place. We verify that log-probabilities under a score-based prior well-approximate those under a ground-truth Gaussian prior in Figure 2.3, but log-probabilities are difficult to validate for complex priors (which motivates our work).

2.5 Results of Posterior Sampling with Score-based Priors

We validate our approach on several imaging tasks: denoising, a version of deblurring, and interferometry. Interferometry is used for black-hole imaging (Figure 2.2) and highlights the benefits of score-based priors for scientific applications, which call for exact posterior sampling given standalone priors to accurately quantify uncertainty.

2.5.1 Baseline methods

We include comparisons to three PnPDP methods for posterior sampling with an unconditional diffusion model, described in the following paragraphs.

SDE+Proj [284] The image $\boldsymbol{x}(t)$ is projected onto a measurement subspace at each t throughout reverse diffusion. A hyperparameter λ determines measurement weight. This approach is purported to give plausible solutions rather than exact posterior samples. It is restricted to compressed-sensing linear inverse problems in which the forward matrix has fewer rows than columns.

Score-based annealed Langevin dynamics (Score-ALD) [173] The authors propose ALD with approximate posterior scores:

$$\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x} \mid \boldsymbol{y}) \approx \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) + \frac{\boldsymbol{A}^H(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})}{\sigma^2 + \gamma_t^2}, \tag{2.6}$$

where $\frac{\boldsymbol{A}^H(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})}{\sigma^2+\gamma_t^2}$ is the log-likelihood gradient assuming linear measurements ($\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{n}$) and Gaussian noise ($\boldsymbol{n} \sim \mathcal{N}(0_M, \sigma^2 \boldsymbol{I}_M)$). Here γ_t is a hyperparameter for the weight of the log-likelihood term at step t, thus calling for a hand-tuned approximation of the likelihood score at each t. One way to automatically adjust the magnitude of the likelihood score is to renormalize it to have the same magnitude as the prior score, which is what the authors do in their experiments. Without this trick, we find image posterior sampling highly sensitive to the γ_t annealing schedule.

Diffusion posterior sampling (DPS) [79] The authors propose an approximation of $\log p_0(\boldsymbol{y} \mid \boldsymbol{x}(t))$ throughout reverse diffusion. A hyperparamter ζ_t determines the magnitude of the log-likelihood gradient at each t. Similar to ALD, the authors use a trick to determine the magnitude of the likelihood score, setting $\zeta_t := \zeta / \|\boldsymbol{y} - f(\hat{\boldsymbol{x}}(\boldsymbol{x}(t)))\|$, where f is the forward model, and $\hat{\boldsymbol{x}}(\boldsymbol{x}(t)) := \mathbb{E}[\boldsymbol{x}(0) \mid \boldsymbol{x}(t)]$.

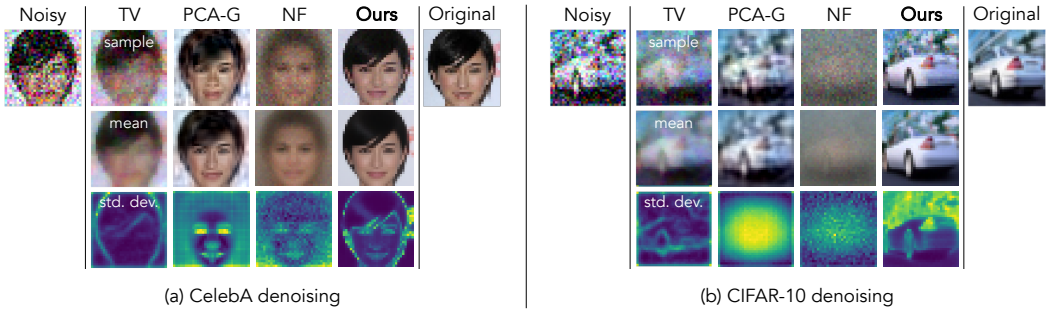


Figure 2.6: Denoising with different priors. The data-driven priors (PCA-G, NF, Ours) were trained on the CelebA training set in (a) and the CIFAR-10 training set in (b). **TV** regularization weight is 0.00025. **PCA-G** is a Gaussian based on the top 512 principal components of the training data. **NF** is a RealNVP with 64 affine-coupling layers. TV, PCA-G, and NF are relatively simple priors that do not give rich posteriors. Score-based priors (**Ours**) capture complex spatial correlations (see sample quality and data-driven std. dev. maps).

2.5.2 Imaging inverse problems

There are two comparisons to perform: (1) score-based priors vs. other *priors* and (2) our posterior-sampling approach vs. *PnPDP methods*. (1) is done with a denoising experiment, and (2) is done with a deblurring experiment, although the findings are not task-dependent. We then focus on black-hole imaging as an endeavor that could benefit from score-based priors as tools in the scientific process.

2.5.2.1 Validation of prior (denoising)

We denoise images corrupted by i.i.d. Gaussian noise with a standard deviation of $\sigma = 0.2$ (20% of the dynamic range). Figure 2.6 shows denoised posteriors given different priors: (1) **TV** regularization, (2) PCA-Gaussian (**PCA-G**), (3) RealNVP normalizing flow (**NF**), and (4) our score-based prior. DPI was used to approximate the posterior for each prior except PCA-G, which has an analytical Gaussian posterior.

Score-based priors provide more informative posteriors than traditional priors do. This is reflected in sample quality: as Table 2.1 shows, the average SSIM and PSNR of posterior samples are highest when using a score-based prior. Score-based priors also provide richer posteriors: as shown by the empirical standard deviations in Figure 2.6, our priors result in full posteriors with a data-driven uncertainty (e.g., uncertainty is higher on facial features like the eyes, ears, and hair of the CelebA image, whereas other priors are more unaware of facial structure). The RealNVP NF performed poorly as an image prior, perhaps because it struggles to generalize

Image \ Prior	CelebA (Ours)	CIFAR (Ours)	Same dataset (PCA-G)	Same dataset (NF)	TV
CelebA	SSIM: **0.88**	0.71	0.76	0.31	0.53
	PSNR: **25.1**	21.6	20.5	11.3	17.4
CIFAR	SSIM: 0.79	**0.85**	0.74	0.22	0.57
	PSNR: 19.0	**21.0**	19.4	13.4	17.6

Table 2.1: Denoising metrics. Row refers to test image in Figure 2.6. Column refers to prior used for denoising. Average SSIM and PSNR were computed across 128 posterior samples for each (test image, prior) pair. The "correct" score-based prior (Ours) performs best for each test image, while even the "incorrect" score-based prior performs well compared to the other priors.

to non-training images and thus leads to unstable optimization of the variational posterior.

2.5.2.2 Validation of posterior sampling (deblurring)

We consider the task of reconstructing an image from measurements of the lowest DFT spatial frequencies, which we call "deblurring" to simplify terminology. In our experiments, we observed the lowest 6.25% DFT spatial frequencies with complex-valued measurement noise with a standard deviation of $|\sigma| = 1.0$ (~0.2% of the magnitude of the zero-frequency component). Figures 2.7 and 2.8 show results for a CelebA and CIFAR-10 source image, respectively, comparing our method, Score-ALD, DPS, and SDE+Proj. We used the same two score models (one trained on CelebA and one trained on CIFAR-10) for all methods. Our method outperforms others in terms of MSE, PSNR, and SSIM (e.g., in Figure 2.7(a), our posterior samples have an average PSNR of 24.75, while DPS, the best-performing method in terms of PSNR, achieved an average of 20.37).

Our posterior sampling method is much more robust to mismatched priors than baselines are. Consider the CelebA prior applied to measurements of a CIFAR-10 image in Figure 2.8(b). When the measurement weight in baselines is lower, they hallucinate faces, resulting in a posterior that lies within the prior. When it is higher, they introduce unnatural artifacts to fit the measurements. Hyperparameter-dependent methods make it easy to mistakenly over-bias the prior or the measurements. Without access to a ground-truth, one cannot reliably interpret a posterior if it is not the result of principled Bayesian inference.

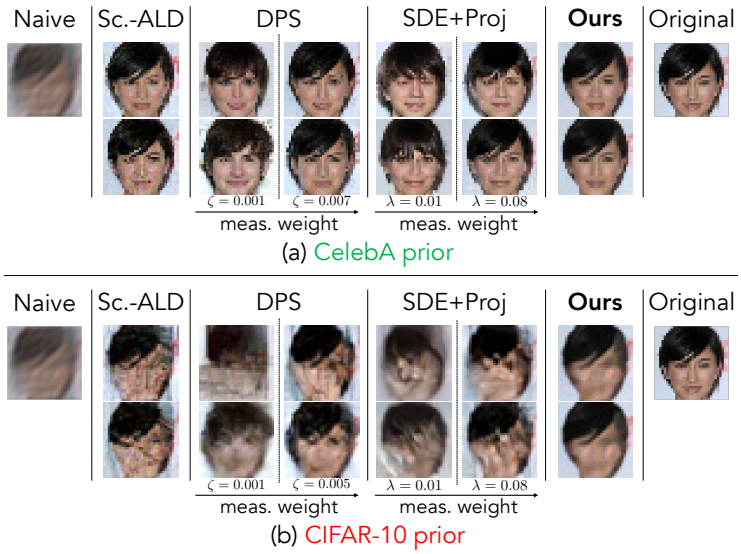


Figure 2.7: Deblurring (CelebA). Two samples from each method are shown. The source is a CelebA test image (Original). Score-ALD uses the likelihood-gradient renormalization trick. **(a) Using the "correct" prior.** All methods recover plausible posterior samples when the prior includes the true source image. With baselines, posterior variance depends on the meas. weight, which is difficult to set without knowing the ground-truth posterior. **(b) Using an "incorrect" prior.** All methods (expectedly) are not able to recover a face. But baseline methods introduce heavy artifacts, while samples from Ours still look natural.

2.5.2.3 Interferometric imaging

The scientific venture of black-hole imaging calls for principled inference with priors that are minimally hand-tuned. Radio interferometry is a technique for imaging astronomical targets with high angular resolution by using a distributed array of radio telescopes. An interferometer collects sparse spatial-frequency measurements of the sky's image. The Event Horizon Telescope (EHT) notably used this technique to take the first image of a black hole [7]. In this work, we simulated interferometric measurements from the EHT telescope array using the `ehtim` package [65]. These simulated measurements contain thermal noise but exclude realistic atmospheric noise that results in additional phase corruption. We discuss the EHT measurements and forward model in-depth in Chapter 3 and Appendix A.

The first image of a black hole, while the result of carefully-obtained measurements from the EHT array, was only possible with image priors (technically formulated as regularizers). EHT scientists handcrafted many priors, each bringing different biases to the image reconstruction. Only the common structure found between these priors could be reliably interpreted, such as the diameter of the photon ring

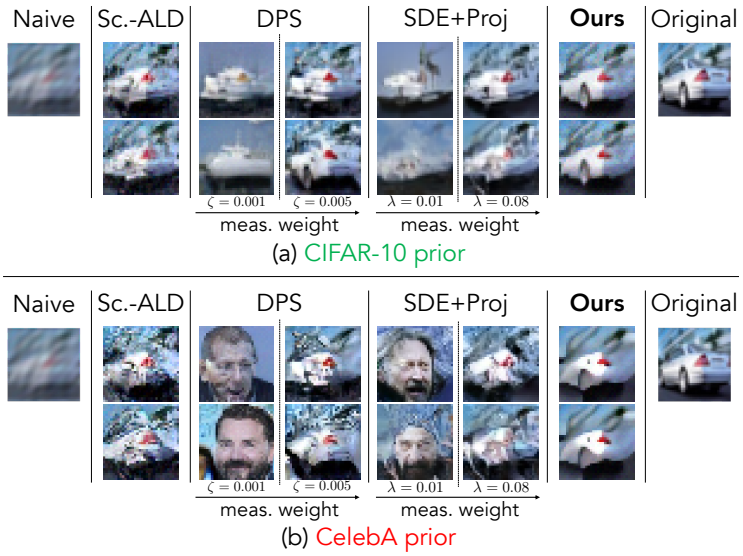


Figure 2.8: Deblurring (CIFAR-10). **(a) Using the "correct" prior.** For DPS and SDE+Proj, both weights achieve plausible samples, but variance differs drastically. **(b) Using an "incorrect" prior.** ALD, DPS, and SDE+Proj with high meas. weight struggle to produce natural images when faced with out-of-distribution measurements. With lower meas. weight, DPS and SDE+Proj hallucinate faces. This is also troubling since the posterior should not lie inside the prior in this case.

surrounding the black hole [7]. Score-based priors could streamline this process as principled, hyperparameter-free priors that can be easily trained on different image distributions. Plugged into the imaging algorithm, they provide a collection of posteriors that incorporate different image statistics while maintaining measurement consistency.

It is important to remember that a posterior exists for *any* combination of a prior and measurements, no matter how far the prior is from the source image. Faithfully modeling the posterior that arises from a given prior is especially crucial in a task like black-hole imaging. Since it is impossible to train a prior on real black holes, any data-driven prior likely does not perfectly agree with the source image. Similar to our deblurring experiments (Figures 2.7 and 2.8), Figure 2.9 highlights that score-based priors are robust to mismatches between the prior and true underlying distribution. Figure 2.9 shows results using score-based priors trained on CelebA and CIFAR-10. The simulated measurements, from all five telescopes in the array, contain enough information for both priors to recover the underlying image structure. We compare our results to those of Score-ALD and DPS (SDE+Proj does not support this type of forward model). [2] Unlike these methods, which introduce either prior-

[2]The EHT forward matrix has more rows than columns, but it is ill-conditioned because mea-

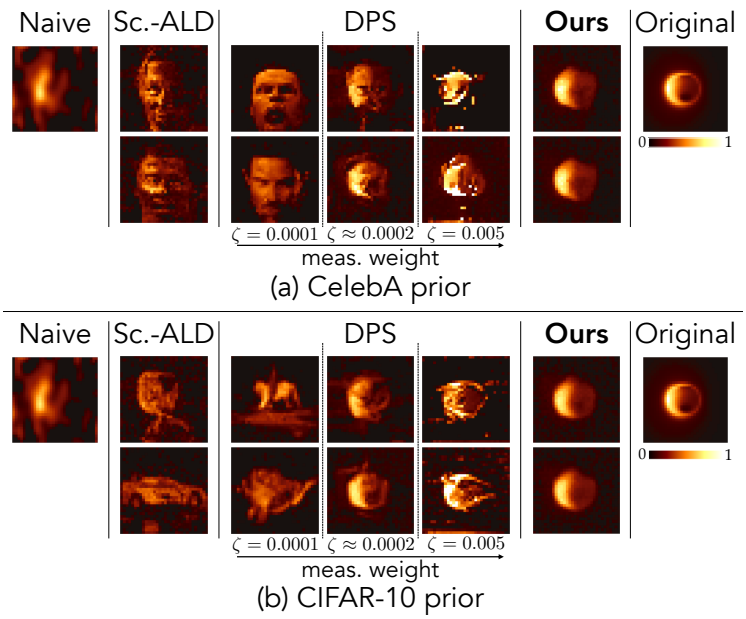


Figure 2.9: Interferometric imaging of a synthetic black-hole image. Two random samples from each method are shown (images are all shown with the same color scale). Both baselines struggle to balance measurements and prior, hallucinating faces in (a) and a car and horse in (b). Posterior samples become unstable as meas. weight increases, as evidenced by the unnatural structures resulting from the highest meas. weight in DPS. Regardless of meas. weight, variance of baseline samples far exceeds ours (ALD max. std. dev. is 0.202; DPS max. std. devs. are 0.237, 0.242, 0.444, resp., as meas. weight increases; our max. std. dev. is 0.036). **Ours** produces samples that agree with the true structure, automatically balancing likelihood and prior.

related features (e.g., face) or unstructured artifacts (e.g., random pixels to agree with measurements), our hyperparameter-free method automatically balances the measurements and prior.

As shown in Figure 2.2, a score-based prior visibly affects the posterior wherever measurements are not sufficient to constrain image structure. For instance, when applying a score-based prior trained on CelebA, as measurements are removed, more facial features appear in the posterior images. Given enough measurements, the recovered structure in the posterior samples can be reliably analyzed for scientific interpretation. With our framework, it is also possible to train a collection of score-based priors and look for common features between the posteriors that arise from the different priors. These common features are more likely to reflect the true underlying image. This idea is explored in Feng and Bouman [115], and we investigate it in-

surements are highly correlated.

depth in Chapter 3 .

2.5.3 Implementation details

Gaussian ground-truth distribution The Gaussian distribution used for Figures 2.3, 2.4, and 2.5 is defined for 16×16 grayscale images. The mean and covariance were fit by expectation-maximization to images from the CelebA training set (each image was first center-cropped to 140×140 and then rescaled to 16×16). The covariance was preconditined by adding 0.01 along the diagonal. Training data were samples randomly drawn from the resulting Gaussian distribution.

Score model All score models that were trained on 32×32 images had an NCSN++ architecture [283] with 64 filters in the initial layer. The score model trained on the Gaussian ground-truth distribution in Section 2.3.2 had 128 filters in the initial layer.

DPI implementation We adapted the PyTorch implementation of DPI [287] for JAX/Flax. For all presented results on image posterior sampling, we used a RealNVP architecture with 64 affine-coupling layers. The RealNVP was optimized with stochastic gradient descent (SGD) with a batch size of 64. We used the Adam optimizer with a learning rate of 0.0002 and clipped gradients to have norm 1.

DPI sampling Once optimized, the RealNVP can be sampled to obtain samples from the approximate posterior. Occasionally the RealNVP produces a clearly out-of-distribution sample, so we remove such outliers by discarding any sample with a pixel value whose magnitude is greater than 2. Although not needed in most cases, we applied this postprocessing step before computing statistics of DPI-estimated posteriors.

DPI optimization time The main computational bottleneck is computing log-probabilities for each batch. Since we use adaptive step-size controllers, the time required for each SGD step is variable. In our experiments, we found it ranged from 30 seconds/step to 200 seconds/step. The time required for each ODE solve could also depend on the complexity of the distribution underlying the score-based prior. For example, we found CelebA priors to be faster (about 50 seconds/step for interferometric imaging experiments) and CIFAR-10 priors to be slower (about 200 seconds/step for deblurring a CIFAR-10 image, which was the slowest case). The RealNVP generally converges within 5000–10000 SGD steps, although we ran the

Solver	$D_{\mathrm{KL}}(q\|p)$ (↓)	NFE low. bound (↓)
Euler* (1st order)	0.848	4092
Heun (2nd order)	0.478	312
Bosh3 (3rd order)	0.453	81
Tsit5 (5th order)	0.521	255
Dopri5 (5th order)	0.284	65
Dopri8 (8th order)	0.422	1440

Table 2.2: KL divergence depending on the solver used for log-probability computation. "Euler" used a fixed step-size of $1/4092$. All other solvers used adaptive step-sizing, with the number of function evaluations ("NFE") calculated as the number of solver steps times the order of the solver. The KL divergence was estimated from 512 samples from the ODE sampler.

optimization for 20000–50000 steps to be sure of convergence. We used v4-8 TPUs to perform the optimization.

2.5.3.1 Computation of log-probability value and gradient

Log-probability estimation The two implementation decisions that most affect log-probability accuracy are: (1) which ODE solver to use and (2) how to estimate the divergence in Equation 2.4. To deal with (1), our code[3] uses Diffrax [183], a JAX library for differential equations, to easily swap out solvers and adaptively select time steps. As for (2), we use Hutchinson-Skilling estimation with multiple trace estimators to reduce the variance of log-probability and gradient calculations. For our experiments, we used 5th-order Runge-Kutta solvers [102, 307] and Hutchinson-Skilling trace estimation [169, 276]. We approximated gradients with the continuous adjoint method [183].

ODE solver Table 2.2 shows how different solvers affect time-efficiency and KL divergence to a ground-truth distribution. The ground-truth is the Gaussian distribution used in Section 2.3.2. This suggests that Bogacki–Shampine's 3/2 method and Dormand–Prince's 5/4 method offer a good balance between efficiency and accuracy. Note, however, that score-based priors trained on different datasets may show different trends. It is always a good idea to evaluate the runtime of different solvers for a given score-based prior to find the most efficient solver.

[3]Website: `http://imaging.cms.caltech.edu/score_prior`

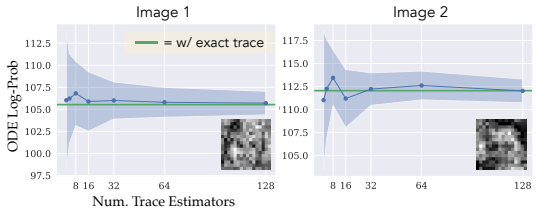


Figure 2.10: Mean and variance of log-probability values vs. number of trace estimators. The score-based prior was fit to the ground-truth Gaussian distribution used in Main Sec. 3.2. For each number of trace estimators (1, 2, 4, 8, 16, 32, 64, or 128), 50 trials of log-probability estimation were done with different random seeds (using the Dopri5 solver with adaptive step-sizing). The solid blue line indicates the mean of these trials, and the shaded region indicates one std. dev. above and below the mean. The solid green line shows the value resulting from exact trace calculation. The evaluated image is inset. As more trace estimators are used, the variance of the log-probability decreases.

Trace estimation For high-dimensional data, trace estimation is necessary to estimate the divergence in Equation 2.4. This causes variance in the estimated log-probabilities and gradients. Song et al. [283] used Hutchinson-Skilling with one trace estimator, but we use multiple trace estimators to reduce variance. In our implementation, the same trace estimators are applied to each image in a batch. Figures 2.10 and 2.11 show the variance of densities and gradients, respectively, depending on the number of trace estimators used.

Adjoint ODE To compute the exact gradient $\nabla_{\boldsymbol{x}} \log p_{\boldsymbol{\theta}}(\boldsymbol{x})$, we would need to backpropagate through the ODE solve. This is too memory-intensive, so we opt for the continuous adjoint method [71, 183], which solves a secondary ODE that gives the gradient of the idealized continuous-time primary ODE. This adjoint method best balances our memory, speed, and accuracy requirements. Direct back-propagation through the probability flow ODE could be possible with improved gradient-checkpointing.

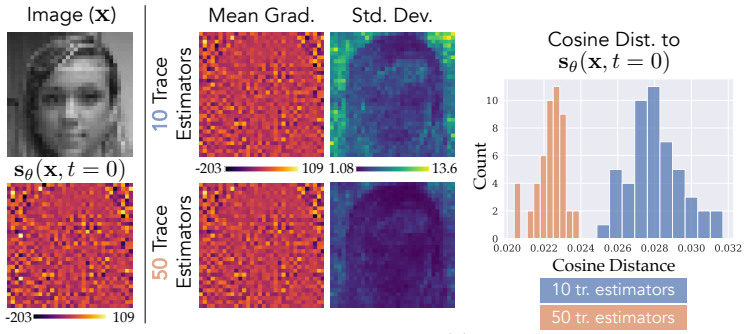


Figure 2.11: Mean and variance of $\nabla_{\boldsymbol{x}} \log p_{\boldsymbol{\theta}}(\boldsymbol{x})$ with 10 vs. 50 trace estimators. The score-based prior was trained on 32×32 grayscale CelebA images. **(a)** Test image $\boldsymbol{x}$ and gradient according to the learned score model, $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$, evaluated at $t = 0$. (In reality, we set $t = 10^{-3}$ for numerical stability and perturbed $\boldsymbol{x}$ with noise accordingly.) Since the test image was drawn as $\boldsymbol{x} \sim p_0$, the score-model output should equal the true $\nabla_{\boldsymbol{x}} \log p_{\boldsymbol{\theta}}(\boldsymbol{x})$. **(b)** Results of estimating the gradient $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x})$ with the probability flow ODE including trace estimation. For both "10 trace estimators" and "50 trace estimators," 50 trials of gradient estimation were done with the continuous adjoint method. "Mean Grad." and "Std. Dev." are the average gradient and std. dev. of the gradient of all these runs. "Cosine Dist. to $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t = 0)$" shows the histogram of the cosine distance between each gradient estimate and the score-model output, which we consider to be ground-truth. The results in (b) are evidence that trace estimation gives a good approximation of the gradient in expectation, but using fewer trace estimates causes higher variance. With 10 trace estimates, the median relative std. dev. of the gradient is 16%. With 50 trace estimates, it is 8.6%. (Relative std. dev. is computed as $|\sigma|/|\mu|$.) Also note that regions of highest variance are in the image background.

2.6 Efficient Surrogate Score-based Priors

A drawback of score-based priors is that computing the exact probability of an image is computationally expensive. Even so, the VI approach is appealing because it offers the same theoretical guarantees as traditional VI. This can sometimes be worth the computational cost for imaging applications in which measurements are expensive or difficult to obtain (e.g., black-hole imaging [7], cryo-electron microscopy [106, 204], and X-ray crystallography [331, 226]). Still, it would be beneficial to have an efficient "surrogate" for this approach that can be used in the development and testing stages or simply to reduce computational costs. In this section, we propose a surrogate score-based prior for efficient VI.

It turns out that computing the evidence lower bound (ELBO) of a score-based prior [282, 155] is computationally efficient and feasible for high-dimensional images. We propose to use this lower bound as a surrogate for the exact score-based prior. In particular, we use the ELBO function as a substitute for the exact log-probability function, and it can be plugged into any inference algorithm that requires the value or gradient of the posterior log-density. For VI of an image posterior, we find at least two orders of magnitude in speedup of optimizing the variational distribution. Furthermore, the surrogate approach reduces GPU memory requirements, as there is no need to evaluate and backpropagate through an ODE.

2.6.1 Evidence lower bound of score-based priors

In Section 2.3 we evaluated image probabilities under the p_0 distribution induced by the probability flow ODE. We now denote that p_0 distribution by $p_{\boldsymbol{\theta}}^{\mathrm{ODE}}$. To derive the ELBO, we turn to the following reverse-time SDE using the approximate score:

$$\mathrm{d}\boldsymbol{x} = \left[\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2 \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}}, \tag{2.7}$$

which is the same as Equation 2.2, except that $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$ replaces $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$. Sampling from the diffusion model involves solving this approximate reverse-time SDE, and it induces another distribution at $t = 0$, which we denote as $p_{\boldsymbol{\theta}}^{\mathrm{SDE}}$. Song et al. [282] proved that if $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) \equiv \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}, t)$ for all $t \in [0, T]$ and $p_T = \pi$, then $p_{\boldsymbol{\theta}}^{\mathrm{ODE}} = p_{\boldsymbol{\theta}}^{\mathrm{SDE}} = p_{\mathrm{data}}$. We assume that $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) \approx \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}, t)$ for almost all $\boldsymbol{x} \in \mathbb{R}^D$ and $t \in [0, T]$ and that $p_T \approx \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, so that $p_{\boldsymbol{\theta}}^{\mathrm{ODE}} \approx p_{\boldsymbol{\theta}}^{\mathrm{SDE}} \approx p_{\mathrm{data}}$.

Song et al. [282] derived an ELBO, denoted as $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}$, for $p_{\boldsymbol{\theta}}^{\mathrm{SDE}}$ such that $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x}) \leq \log p_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ for any proposed image $\boldsymbol{x}$. Essentially, this lower bound corresponds to how well the diffusion model is able to denoise a given image: an image with high

probability under the diffusion model is easy to denoise, whereas a low-probability image is difficult.

The lower bound, or negative "denoising score-matching loss" [282], is defined as

$$b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x}) := \mathbb{E}_{p_{0T}(\boldsymbol{x}'|\boldsymbol{x})}\left[\log \pi(\boldsymbol{x}')\right] - \frac{1}{2}\int_0^T g(t)^2 h(t)\mathrm{d}t, \tag{2.8}$$

where

$$h(t) := \mathbb{E}_{p_{0t}(\boldsymbol{x}'|\boldsymbol{x})}\Bigg[\|\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}', t) - \nabla_{\boldsymbol{x}'} \log p_{0t}(\boldsymbol{x}' \mid \boldsymbol{x})\|_2^2 \tag{2.9}$$
$$- \|\nabla_{\boldsymbol{x}'} \log p_{0t}(\boldsymbol{x}' \mid \boldsymbol{x})\|_2^2 - \frac{2}{g(t)^2}\nabla_{\boldsymbol{x}'} \cdot \boldsymbol{f}(\boldsymbol{x}', t)\Bigg].$$

Here $p_{0t}(\boldsymbol{x}' \mid \boldsymbol{x})$ denotes the transition distribution from $\boldsymbol{x}(0) = \boldsymbol{x}$ to $\boldsymbol{x}(t) = \boldsymbol{x}'$. For a drift coefficient that is linear in $\boldsymbol{x}$, this transition distribution is Gaussian: $p_{0t}(\boldsymbol{x}' \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}'; \alpha(t)\boldsymbol{x}, \beta(t)^2\boldsymbol{I})$. This means that the gradient $\nabla_{\boldsymbol{x}'} \log p_{0t}(\boldsymbol{x}' \mid \boldsymbol{x})$ is directly proportional to the Gaussian noise that is subtracted from $\boldsymbol{x}'$ to get $\boldsymbol{x}$. Equation 2.8 is efficient to compute since we can evaluate it by adding Gaussian noise to $\boldsymbol{x}$ without having to solve an initial-value problem as with the ODE. In fact, Equation 2.8 is closely related to the denoising score-matching objective used to train diffusion models [283].

Intuitively, we can interpret Equation 2.8 as associating an image's probability with how well the score model $\boldsymbol{s}_{\boldsymbol{\theta}}$ could denoise that image if it underwent diffusion. This is represented by the first term in $h(t)$ (Equation 2.9). To assess the probability of an image $\boldsymbol{x}$, we perturb it with Gaussian noise to get $\boldsymbol{x}'$ and then ask the score model to estimate the noise that was added. If $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$ accurately estimates the noise, then $\|\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}', t) - \nabla_{\boldsymbol{x}'} \log p_{0t}(\boldsymbol{x}' \mid \boldsymbol{x})\|_2^2$ is small, and the value of $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ becomes larger. The remaining terms in $h(t)$ are normalizing factors independent of $\boldsymbol{\theta}$, and $\mathbb{E}_{p_{0T}(\boldsymbol{x}'|\boldsymbol{x})}\left[\log \pi(\boldsymbol{x}')\right]$ accounts for the probabilities of the noise images $\boldsymbol{x}(T)$ that could result from $\boldsymbol{x}$ being entirely diffused.

2.6.2 Approximating the posterior with VI

Recall from Section 2.4 that we would like to find the variational distribution $q_{\boldsymbol{\phi}}$ that minimizes the KL divergence to the target posterior:

$$\boldsymbol{\phi}^* = \arg\min_{\boldsymbol{\phi}} D_{\mathrm{KL}}(q_{\boldsymbol{\phi}} \| p_{\boldsymbol{\theta}}(\cdot \mid \boldsymbol{y})) \tag{2.10}$$
$$= \arg\min_{\boldsymbol{\phi}} \mathbb{E}_{\boldsymbol{x}\sim q_{\boldsymbol{\phi}}}\left[-\log p(\boldsymbol{y} \mid \boldsymbol{x}) - \log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x}) + \log q_{\boldsymbol{\phi}}(\boldsymbol{x})\right]. \tag{2.11}$$

Now, to circumvent the computational challenges of evaluating the prior term $\log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$, we replace it with the surrogate $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ to get the following objective:

$$\boldsymbol{\phi}^* = \arg\min_{\boldsymbol{\phi}} \mathbb{E}_{\boldsymbol{x}\sim q_{\boldsymbol{\phi}}} \left[-\log p(\boldsymbol{y} \mid \boldsymbol{x}) - b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x}) + \log q_{\boldsymbol{\phi}}(\boldsymbol{x}) \right]. \tag{2.12}$$

We can also think of $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}$ as replacing the intractable $\log p_{\boldsymbol{\theta}}^{\mathrm{SDE}}$ in Equation 2.10. Since $-\log p_{\boldsymbol{\theta}}^{\mathrm{SDE}} \leq -b_{\boldsymbol{\theta}}^{\mathrm{SDE}}$, *our surrogate objective minimizes the upper bound of the KL divergence involving* $p_{\boldsymbol{\theta}}^{SDE}$.

2.6.3 Implementation details

Evaluating $b_{\boldsymbol{\theta}}^{\mathbf{SDE}}(\boldsymbol{x})$ The formula for $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ (Equation 2.8) contains a time integral and expectation over $p_{0t}(\boldsymbol{x}' \mid \boldsymbol{x})$ that can be estimated with numerical methods. Following Song et al. [282], we use importance sampling with time samples $t \sim p(t)$ for the time integral and Monte-Carlo approximation with noisy images $\boldsymbol{x}' \sim \mathcal{N}(\alpha(t)\boldsymbol{x}, \beta(t)^2\boldsymbol{I})$ for the expectation. The proposal distribution $p(t) := \frac{g(t)^2}{\beta(t)^2 Z}$ was empirically verified to result in lower variance in the estimation of $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ [282]. We take N_t time samples and $N_{\boldsymbol{z}}$ noise samples to approximate the time integral and expectation over $p_{0t}(\boldsymbol{x}' \mid \boldsymbol{x})$, respectively. In our experiments, $N_t = N_{\boldsymbol{z}} = 1$. We find that increasing the number of time and noise samples does not efficiently decrease variance in the estimated value of $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$.

Optimization We use SGD to optimize $\boldsymbol{\phi}$, Monte-Carlo-approximating the expectation in Equation 2.12 with a batch of $\boldsymbol{x} \sim q_{\boldsymbol{\phi}}$. Estimating $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ has higher variance than estimating $\log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$. For instance, in Figure 2.15, $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ with $N_t = 2048, N_{\boldsymbol{z}} = 1$ shows higher variance than $\log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$ with 16 trace estimators. A lower learning rate can help mitigate training instabilities caused by variance. For example, in Figure 2.14b the learning rate with the exact prior was 0.0002, while the learning rate with the surrogate prior was 0.00001.

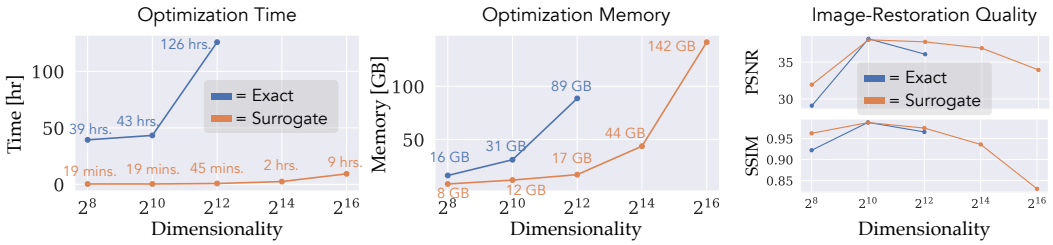


Figure 2.12: Computational efficiency of surrogate prior vs. exact prior. For each image size, we estimated a posterior of images conditioned on 4×-accelerated MRI measurements of a knee image, using a Gaussian distribution with diagonal covariance as the variational distribution. The hardware was 4x NVIDIA RTX A6000. The surrogate prior allows for VI of image sizes that are prohibitively large for the exact prior. For image sizes supported by the exact prior, the surrogate improved total optimization time by over 120× while using less memory and scaling better with image size. "Image-Restoration Quality" verifies that optimization with the surrogate was done fairly, as the PSNR and SSIM of the converged posterior (averaged over 128 samples) are at least as high as with the exact prior.

Image size	Surrogate	Exact
16×16	0.029	19.5
32×32	0.038	41.9
64×64	0.090	123
128×128	0.294	N/A
256×256	1.115	N/A

Table 2.3: Iteration time [sec/step]. A step of gradient-based optimization of the variational distribution is two to three orders of magnitude faster with the surrogate prior.

2.7 Results with Surrogate Score-based Priors

We validate our approach with surrogate score-based priors on the tasks of accelerated MRI, denoising, and reconstruction from low spatial frequencies (deblurring). See Appendix B for details about the forward models.

2.7.1 Efficiency improvements

In Table 2.3 and Figure 2.12, we quantify the efficiency improvements of the surrogate prior for an accelerated MRI task at different image resolutions. We drew a test image from the fastMRI knee dataset [345] and resized it to 16×16, 32×32, 64×64, 128×128, and 256×256. For each image size, we trained a score model on images of the corresponding size from the fastMRI training set of single-coil knee scans. We then optimized a Gaussian distribution with diagonal covariance to approximate the posterior. The batch size was 64 for the surrogate and 32 for

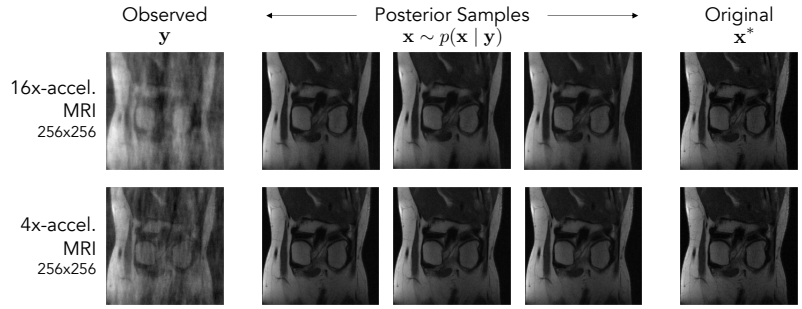
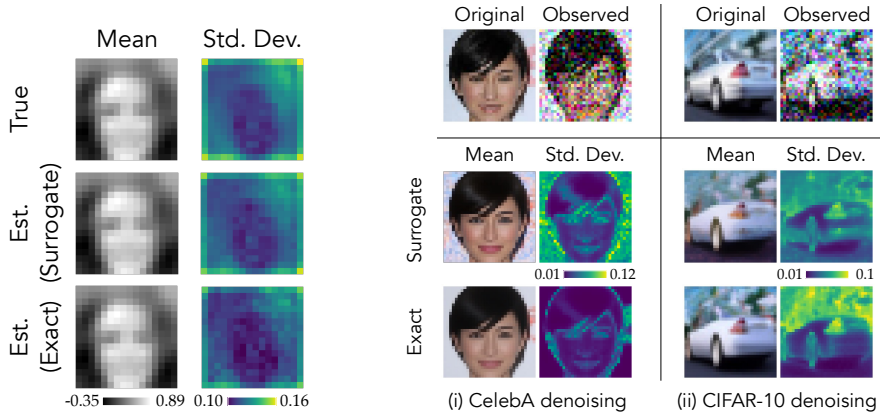


Figure 2.13: High-dimensional Bayesian inference with a surrogate score-based prior. Here we show posterior samples for accelerated MRI of 256×256 knee images, approximated via VI with a surrogate score-based prior. The first row shows reconstruction from 16×-reduced MRI measurements. The second row shows reconstruction given more κ-space measurements, i.e., 4×-reduced MRI. Bayesian imaging at this image resolution is computationally infeasible with the ODE-based approach. Our proposed surrogate enables efficient yet principled inference with diffusion-model priors, resulting in inferred posteriors where the true image is within three standard deviations of the posterior mean for 96% and 99% of the pixels for 16×- and 4×-acceleration, respectively.

the exact prior (a smaller batch size was needed to fit 64×64 optimization into GPU memory). Convergence was defined via a minimum acceptable change in the estimated posterior mean between optimization steps.

We find at least two orders of magnitude in time improvement with the surrogate prior. Table 2.3 compares the iteration time between the two priors. Figure 2.12 compares the total time it takes to optimize the variational distribution. The surrogate also improves memory consumption, which in turn enables optimizing higher-dimensional posteriors. Following standard practice, we just-in-time (JIT) compile the optimization step to reduce time/step at the cost of GPU memory. Figure 2.12 shows how the surrogate prior significantly reduces memory requirements and scales better with image size. The exact prior could only handle up to 64×64 before exceeding GPU memory (we tested on 4x 48GB GPUs). While memory could be reduced with a smaller batch size, this would make optimization more time-consuming. On the other hand, our surrogate prior supports much larger images, as we demonstrate in Figure 2.13 for 256×256[4] MRI with a Gaussian-approximated posterior. This type of principled inference of high-dimensional image posteriors is not practical with the exact score-based prior.

[4]Larger images may be feasible, but their larger memory footprint might restrict the possible batch size and complexity of the variational distribution.



(a) Ground-truth (Gaussian) posterior. (b) Complex posteriors.

Figure 2.14: Estimated posteriors under surrogate vs. exact prior. For each task, the variational distribution is a RealNVP, and the score model is the same between both prior functions. **(a)** Both prior functions help recover the correct (Gaussian) posterior. The score-based prior was trained on samples from a known Gaussian distribution (originally fit to 16×16 face images), and the measurements are the lowest 6.25% spatial frequencies of a test image from the prior. Since the prior and likelihood are both Gaussian, we know the ground-truth Gaussian posterior. **(b)** Estimated posteriors for (i) denoising a CelebA image and (ii) denoising a CIFAR-10 image. Std. dev. is averaged across the three color channels. The score-based prior was trained on CelebA in (i) and CIFAR-10 in (ii). Both prior functions result in comparable image quality; visual differences appear mostly in the image background.

2.7.2 Posterior estimation under surrogate versus exact

The surrogate prior b_{θ}^{SDE} may not be an identical substitute for the exact prior $\log p_{\theta}^{\text{ODE}}$. Importantly, though, we verify in Figure 2.14a that both the surrogate and exact prior recover a ground-truth Gaussian posterior derived from a Gaussian likelihood and prior, using a RealNVP as the variational distribution. The score model (used by both the surrogate and exact prior) was trained on samples from the known prior.

Nonetheless, the surrogate could result in a different locally-optimal variational posterior, particularly if the posterior leads to many local minima in the variational objective. Figure 2.14b compares posteriors (with unknown true distributions) approximated by a RealNVP under the surrogate versus exact prior. For each task (CelebA denoising and CIFAR-10 denoising), both prior functions used the same trained score model. We observe in these comparisons that most of the differences appear in the image background and that both priors result in a plausible mean reconstruction and uncertainty.

Visualizing the bound gap throughout optimization helps shed light on why the two priors converge to different solutions even with the same underlying score model. Figure 2.15 shows probabilities of samples generated by q_ϕ (in this case a RealNVP) as optimization progresses. At each checkpoint of q_ϕ, we plot $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ versus $\log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$ for samples $\boldsymbol{x} \sim q_\phi$ coming from both the exact and surrogate optimization of q_ϕ. We find that the surrogate is a valid bound for the ODE log-density: $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x}) \leq \log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$ for all $\boldsymbol{x} \sim q_\phi$, except for some outliers due to variance of $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$. However, optimization follows a different trajectory depending on the prior. With the surrogate, samples $\boldsymbol{x} \sim q_\phi$ tend toward a region where the bound gap is small. Meanwhile, the exact prior follows a loss landscape whose structure appears to be independent of the lower bound. Note that samples from q_ϕ optimized under the exact prior obtain higher values of $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ than samples obtained under the surrogate. Figure 2.15 suggests that gradients under the surrogate tend to push the q_ϕ distribution along the boundary of equality between $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}$ and $\log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}$. This constrains the path taken through gradient descent and subsequently the converged solution.

2.7.3 Quality of posterior estimation with Bayesian vs. PnPDP approaches

As discussed in Section 2.2, many guidance-based PnPDP methods attempt to sample from the posterior by incorporating measurements throughout the reverse diffusion process of the diffusion model that has been trained on images from the prior. Such approaches provide fast conditional sampling, but they may severely mischaracterize the posterior. Although our approach also approximates the posterior (due to using both a surrogate prior and a variational distribution), we find that being grounded in VI helps us obtain a more accurate posterior *and* images that more accurately reflect the true image than guidance-based PnPDP methods that make ad-hoc approximations of the posterior. In the following experiments, we compare to the same three baselines as in Section 2.5: **SDE+Proj** [284], **Score-ALD** [173], and **DPS** [79]. Our approach is DPI with the surrogate prior.

2.7.3.1 Accuracy of posterior

A simple 2D example illustrates the accuracy gap between the PnPDP baselines and our VI approach. We consider a bimodal posterior: the prior is a bimodal mixture-of-Gaussians and the forward model a linear projection with Gaussian noise, making the posterior a bimodal mixture-of-Gaussians. This setup lets us evaluate with a true posterior and over a reasonable space of hyperparameters for the PnPDP baselines.

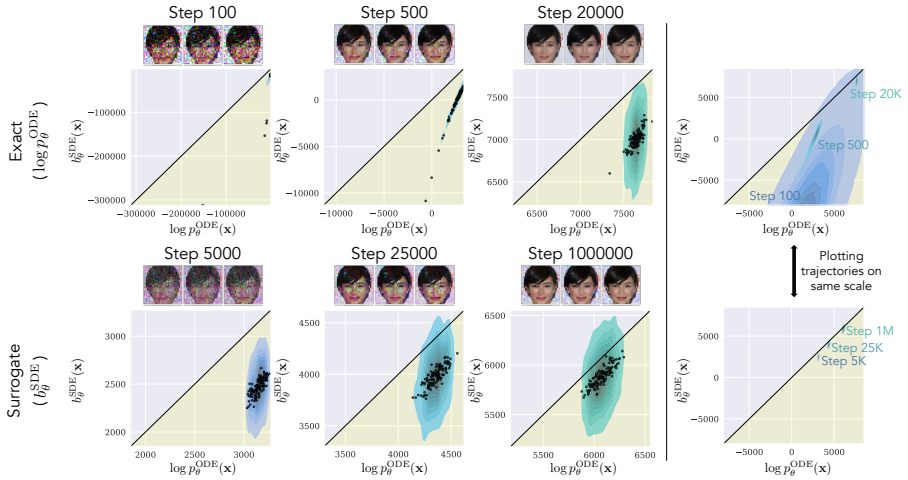


Figure 2.15: $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ vs. $\log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$ for samples $\boldsymbol{x} \sim q_{\boldsymbol{\phi}}$ as optimization of $\boldsymbol{\phi}$ progresses. The task is from Figure 2.14b(i). For each plot, we took 128 samples $\boldsymbol{x} \sim q_{\boldsymbol{\phi}}$ and performed 20 estimates each of $\log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$ and $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ (approximated with $N_t = 2048$ for reduced variance). The density map is a KDE plot of all $128 \cdot 20 = 2560$ values; the 128 scatter points represent the mean estimate for each $\boldsymbol{x}$. The black line indicates perfect agreement between $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ and $\log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$. We expect all points to lie below this black line for $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}$ to be a lower bound. We find that $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x}) \leq \log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$ (up to variance error), but the optimization progresses differently depending on the prior. Gradients under the surrogate push $q_{\boldsymbol{\phi}}(\boldsymbol{x})$ along the black line to increase $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ without exceeding $\log p_{\boldsymbol{\theta}}^{\mathrm{ODE}}(\boldsymbol{x})$. Optimization under the exact prior proceeds more freely, although eventually achieves higher $b_{\boldsymbol{\theta}}^{\mathrm{SDE}}(\boldsymbol{x})$ at convergence. This visualization may help explain differences in the posterior estimated with the surrogate vs. exact prior.

	KL (↓)	time/optimization step (↓)
DPI + exact	**0.030**	130 ms
DPI + surr.	0.037	**22 ms**
DPS (oracle)	0.064	
Score-ALD (oracle)	0.10	
SDE+Proj (oracle)	0.12	

Table 2.4: Quantitative results for Figure 2.16. A two-component Gaussian mixture model (GMM) was fit to estimated samples to obtain a PDF. "Ours" achieves a much lower KL div. (i.e., reverse KL from estimated posterior to true posterior) than guidance-based PnPDP baselines at their best. Time/step for DPI optimization is lower with our surrogate than with the exact score-based prior without sacrificing much accuracy.

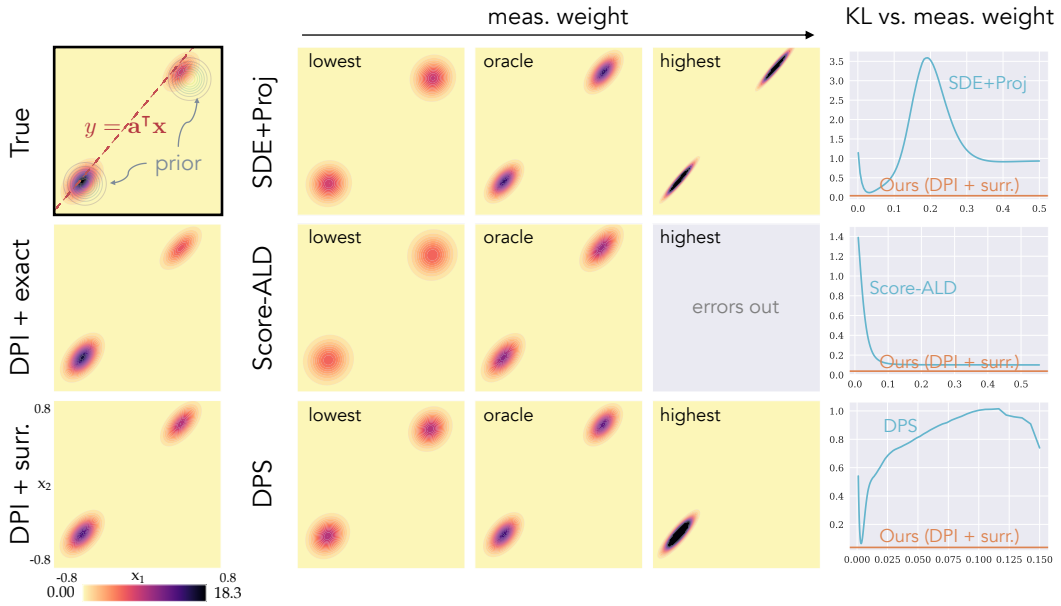
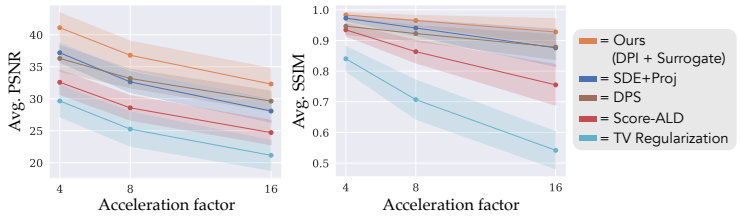
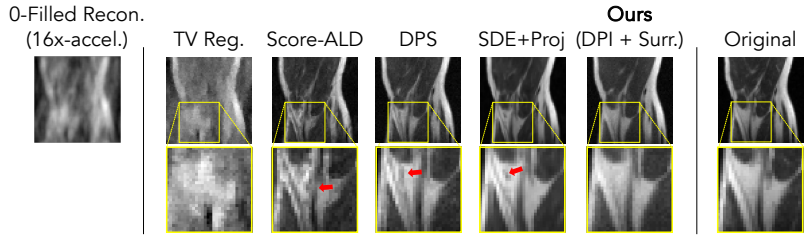


Figure 2.16: Comparing our VI approach with a surrogate score-based prior to baselines on a bimodal posterior. In this example, the prior is a bimodal mixture-of-Gaussians, and the likelihood is Gaussian, making the posterior a bimodal mixture-of-Gaussians (shown in "True"). Assuming access to the true prior score function, we tested how well each method recovers the true posterior. Guidance-based PnPDP methods depend on hand-tuned meas. weights. Even the meas. weight giving the best KL divergence ("oracle") does not rival using our hyperparameter-free VI approach ("DPI + surr."). Note that this "oracle" weight would not be accessible in practice, as it is determined by comparing to the ground-truth posterior. The baselines either (1) incorrectly place equal weight on both posterior modes or (2) miss one of the modes. DPI with either the surrogate or the exact score-based prior recovers the relative weights of both modes. **(KL vs. meas. weight)** Regardless of hyperparameters, the PnPDP baselines do not reach our KL divergence.

We tested how well each of the methods could recover the ground-truth posterior when given the true score function of the bimodal prior (thus avoiding potential error caused by learning the score function). Figure 2.16 shows estimated probability density functions (PDFs) for the evaluated methods. None of the PnPDP baselines correctly recover the bimodal posterior for any hyperparameter value. In particular, they struggle to find the correct balance between the two posterior modes—in the best case, they incorrectly place equal weight on each mode; in the worst case, they only recover one mode. Our VI approach with the exact or surrogate score-based prior recovers both modes in correct proportion. As shown in Figure 2.16 and Table 2.4, even the best KL divergence obtained by the PnPDP baselines does not rival that of VI with a score-based prior. We emphasize that the hyperparameter values



(a) Image-restoration metrics.



(b) Example image reconstructions for 16× acceleration.

Figure 2.17: Accelerated MRI. **(a)** For each accel. factor (4×, 8×, 16×), we estimated posteriors for ten knee images measured at that accel. rate. For each method, we computed the average PSNR and SSIM of 128 estimated posterior samples. The line plot shows the average result across the ten tasks; the shaded region shows one std. dev. above and below the average. **(b)** An example of 16×-accel. MRI. The cropped region exemplifies how PnPDP baselines hallucinate more features than necessary. (a) and (b) are evidence that a principled Bayesian approach can get closer to the true image than previous unsupervised methods.

resulting in the "best" KL for PnPDP methods can only be found with knowledge of the ground-truth, which is inaccessible in real-world scenarios. In contrast, our method automatically finds a better KL divergence by following the Bayesian formula.

2.7.3.2 Image-reconstruction quality

It would be reasonable to assume that the guidance-based PnPDP baselines we consider, though less principled, may lead to better visual quality than a Bayesian approach. However, we find that in addition to providing more reliable uncertainty, our approach achieves higher-fidelity reconstructions. We note that similarity to a ground-truth image does not indicate a correct posterior. Still, for a good prior, it might be desirable for posterior samples to accurately recover the original image.

We performed multiple MRI tasks at different acceleration rates and compared our approach to the PnPDP baselines, as well as a total variation (TV) baseline. Implementations and hyperparameter settings for SDE+Proj and Score-ALD were

provided by Song et al. [284]. For DPS, we followed the implementation of Chung et al. [79] and performed a hyperparameter search on an 8×-acceleration test image to find the optimal PSNR. For the TV baseline, we performed DPI with TV regularization with a regularization weight of 10^5 instead of the surrogate score-based prior.

We simulated MRI at three acceleration factors for ten test images, resulting in thirty posteriors to be estimated. As baseline implementations do not assume measurement noise, we gave the baselines noiseless measurements and set a near-zero measurement noise for our method. The test images were randomly drawn from the fastMRI dataset and resized to 64×64. The score model s_θ was trained on 64×64 fastMRI knee images and stayed fixed across all methods.

Our method achieves a marked improvement in PSNR and SSIM (Figure 2.17). Across all acceleration factors and PnPDP baselines, our method improves PSNR by between 2.7 and 8.5 dB. Even though our method and the PnPDP baselines all use the same score model, restoration quality depends on how the prior is used for inference; whereas baselines loosely approximate the posterior and involve hyperparameters, our approach treats the diffusion model as a standalone prior in Bayesian inference. Furthermore, Figure 2.17 confirms that a diffusion-model prior far outperforms a traditional regularizer like TV.

2.7.4 Implementation details

For the sake of reproducibility, we detail the experimental setup behind each figure. Some common implementation details are that the exact prior ($\log p_\theta^{\mathrm{ODE}}$) was always estimated with 16 trace estimators. The RealNVP had 32 affine-coupling layers unless stated otherwise.

2.7.4.1 Variational distributions

We first describe the two types of variational distributions considered in our experiments: a RealNVP and a multivariate Gaussian with diagonal covariance.

RealNVP For images up to 64×64, we used 32 affine-coupling layers and set the number of hidden neurons in the first layer to $1/8$ of the image dimensionality (e.g., $32 \cdot 32 \cdot 3/8$ for 32×32 RGB images). We used batch normalization in the network.

Gaussian The parameters of the diagonal-covariance Gaussian distribution were the mean image and the pixel-wise standard deviation. We initialized the mean at 0.5 and the standard deviation at 0.1 for all pixels. We took the absolute value of the standard deviation to construct the diagonal covariance matrix.

2.7.4.2 MRI efficiency experiment (Figure 2.12, Table 2.3)

Score model For each image size, the score model was an NCSN++ architecture with 64 filters in the first layer. It was trained with the VP SDE with $\beta_{\text{min}} = 0.1$, $\beta_{\text{max}} = 10$.

Variational optimization For each task (i.e., each image size and prior), the variational distribution was a multivariate Gaussian with diagonal covariance. The batch size was 64, learning rate 0.0002, and gradient clip 1. A convergence criterion based on the loss value is difficult to define due to high variance of the loss (we used 1 time sample to estimate $b_{\boldsymbol{\theta}}(\boldsymbol{x})$). We defined a convergence criterion based on the change in the mean of the variational distribution. Specifically, every 10000 steps, we evaluated a snapshot of the variational Gaussian and computed $\delta = \left\|\mu_{\text{curr}} - \mu_{\text{prev}}\right\| / \|\mu_{\text{prev}}\|$, where μ_{curr} and μ_{prev} are the current and previous snapshot means, respectively. If $\delta < \varepsilon$ for some threshold ε two snapshots in a row, then the optimization was considered converged. Since convergence rate depends on the image size and the prior used, we set a different ε for each task:

- 16×16 (surrogate): $\varepsilon = 0.002$
- 32×32 (surrogate): $\varepsilon = 0.003$
- 64×64 (surrogate): $\varepsilon = 0.005$
- 128×128 (surrogate): $\varepsilon = 0.007$
- 256×256 (surrogate): $\varepsilon = 0.009$
- 16×16 (exact): $\varepsilon = 0.0025$
- 32×32 (exact): $\varepsilon = 0.0027$
- 64×64 (exact): $\varepsilon = 0.005$

We were conservative in defining the convergence and checked that optimization under the surrogate actually achieved better sample quality than optimization under the exact prior (see Figure 2.12).

Data The test image is from the fastMRI [345] single-coil knee test dataset and was resized to 64×64 with antialiasing.

2.7.4.3 256x256 MRI examples (Figure 2.13)

The 4×-acceleration result is from the efficiency experiment (Figure 2.12 and Table 2.3) on the 256×256 test image. The 16×-acceleration result came from a similar setup, where the variational distribution was Gaussian with diagonal covariance. Optimization was done with a batch size of 64, learning rate of 0.00001, and gradient clip of 0.0002. We ran optimization for 270K steps (optimization for 4×-acceleration was done in 100K steps with the convergence criterion).

In the figure caption, we report that the true image is within three standard deviations of the inferred posterior mean for 96% and 99% of the pixels for 16×- and 4×-acceleration, respectively. This was computed based on the mean and standard deviation of 128 samples from the inferred posterior. We find the same result when using the exact mean and standard deviation of the inferred posterior: with respect to the inferred posterior, the true image is within three standard deviations of the mean for 96.7% and 99.0% of the pixels for 16×- and 4×-acceleration, respectively.

2.7.4.4 Ground-truth posterior (Figure 2.14a)

Data The mean and covariance of the ground-truth Gaussian prior were fit with PCA (with 256 principal components) to training data from the CelebA dataset [213]. The CelebA images were resized to 16×16 with antialiasing.

Score model The score model was based on the DDPM++ deep continuous architecture of Song et al. [283] with 128 filters in the first layer. It was trained with the VP SDE with $\beta_{\text{min}} = 0.1$ and $\beta_{\text{max}} = 20$ for 100K steps.

Variational optimization The variational distribution was a RealNVP. Under the surrogate prior, optimization was done with a learning rate of 0.00005 and gradient clip of 1. Under the exact prior, the learning rate was 0.0002 and gradient clip 1. Both priors used a batch size of 64.

2.7.5 32x32 image denoising (Figure 2.14b)

Variational optimization For both CelebA denoising (i) and CIFAR-10 denoising (ii), the variational distribution was a RealNVP. Optimization under the exact prior

was done with a learning rate of 0.0002 and gradient clip of 1 for 20K steps. Optimization under the surrogate prior was done with a learning rate of 0.00001 and gradient clip of 1. For CelebA, the batch size was 64, and training was done for 1.72M steps (convergence was probably achieved earlier, but we continued training to be conservative). For CIFAR-10, the batch size was 128 and training was done for 550K steps.

Score model For both (i) and (ii), the score model had an NCSN++ architecture with 64 filters in the first layer. For the CelebA prior, it was trained with the VP SDE with $\beta_{\text{min}} = 0.1$ and $\beta_{\text{max}} = 20$ and with images that were resized without antialiasing. For the CIFAR-10 prior, it was trained with the VP SDE with $\beta_{\text{min}} = 0.1$ and $\beta_{\text{max}} = 10$.

Data The CelebA and CIFAR-10 images are both 32×32. The CelebA image was resized without antialiasing.

2.7.5.1 Bound gap (Figure 2.15)

Visualization of the bound gap is shown for optimization of the RealNVP from Fig.2.14b(i) (i.e., 32×32 CelebA denoising). For the plots comparing the lower-bound to the ODE log-probability, we used 2048 time samples to estimate $b_{\boldsymbol{\theta}}(\boldsymbol{x})$.

2.7.5.2 Accuracy of posterior (Figure 2.16, Table 2.4)

Variational optimization For both the exact score-based prior and surrogate score-based prior, the variational distribution was a RealNVP with 16 affine-coupling layers, and it was optimized for 12K iterations with a batch size of 2560 and learning rate of 10^{-5}. For the surrogate, $b_{\boldsymbol{\theta}}(\boldsymbol{x})$ was approximated with $N_t = N_z = 1$.

Baselines For this 2D experiment, we implemented the diffusion-based baselines exactly according to their proposed algorithms. For SDE+Proj, we tested the following values for the measurement weight λ: `linspace(0.001, 0.5, num=100)`. For Score-ALD, we distilled all hyperparameters into one global hyperparameter $1/\gamma_T$ and tested the following values for γ_T: `linspace(100, 0.8, num=100)`. For DPS, we tested the following values for the scale parameter ζ: `exp(linspace(log(0.001), log(0.15), num=100))`.

Evaluation Since the diffusion-based approaches only provide samples (not probability densities), we approximated the PDF from the estimated posterior samples. For each method, we fit a two-component Gaussian mixture model (GMM) to 10000 samples. The reverse KL divergence was approximated with the log-density function of the fitted GMM and the log-density function of the true posterior, evaluated on these 10000 samples.

2.7.5.3 Image-restoration metrics (Figure 2.17)

Score model The score model was the same as the one used for the 64×64 image in the MRI efficiency experiment (Figure 2.12).

Variational optimization The variational distribution was a RealNVP. Optimization was done with a learning rate of 0.00001 and gradient clip of 0.0002. We used the same convergence criterion as the one used in the MRI efficiency experiment with $\varepsilon = 0.005$. The same convergence criterion was also used for the TV results but with a maximum number of steps of 50000. The TV regularization weight was 10^5.

Baseline hyperparameters For SDE+Proj, we used the `projection` CS solver provided by Song et al. [284] with the hyperparameters `snr=0.517, coeff=1`. For Score-ALD, we used the `langevin` CS solver with the hyperparameters `n_steps_each=3, snr=0.212, projection_sigma_rate=0.713`. For DPS, we used `scale=0.5`. This was the best scale out of $[10, 1, 0.9, 0.5, 0.3, 0.1, 0.001]$ for a test image in terms of PSNR with respect to the true image.

Chapter 3

BLACK-HOLE IMAGING UNDER DIFFERENT ASSUMPTIONS WITH SCORE-BASED PRIORS

In 2019, the Event Horizon Telescope (EHT) Collaboration obtained the first picture of M87* through computational imaging methods [4, 7, 9] The published images gave humans a glimpse of the shadow cast by the supermassive black hole [218, 109, 216] in the galaxy M87 based on data that EHT telescopes had collected in 2017 April [5, 6], and more recent images showed the persistence of the shadow one year later [13]. However, because measurements obtained from VLBI [81, 302] with EHT telescopes are corrupted and limited in number, image reconstruction requires assumptions about plausible image statistics in order to constrain the space of possible images.

Designing a prior is not a straightforward task, especially considering the absence of true images of black holes. We address this problem with a principled strategy: we collect a range of priors that each impose different visual biases and plug these priors into the Bayesian imaging algorithm introduced in Section 2.6 along with EHT VLBI data. Whereas the EHT Collaboration explored different imaging assumptions via the use of different imaging pipelines (e.g., CLEAN [159, 270, 82, 268, 85, 273, 219] and regularized maximum-likelihood (RML) methods [46, 64, 12]), we explore different priors within the same imaging pipeline. Our imaging approach allows us to assess how visual characteristics and uncertainty, as quantified through a Bayesian posterior, vary with the choice of prior.

Our goal is to move along the spectrum between strong constraints and weak constraints on the image. On one side of the spectrum lie model-fitting strategies, which find the parameters of an underlying geometric [9, 317, 232, 214, 289], physical [8, 322, 132, 181, 344, 238], or statistical [224] model that best match the observations. On the other side lie traditional imaging approaches using weak regularizers like TV [11, 196] and maximum entropy [234]. However, each side has its own limitations: model-fitting prevents discovering new features that cannot be explained by the assumed model, whereas traditional regularizers struggle to produce visually rich images. This motivates using score-based priors to obtain a diverse array of priors, ranging from those akin to model-fitting (e.g., by assuming black-hole structure) to those similar to weak regularizers (e.g., by assuming basic properties of natural images).

With score-based priors, we achieve a collection of M87* images that all fit the observed data but differ in certain visual characteristics. Specifically, we trained a score-based prior on each of the following datasets: CIFAR-10 [195] (generic natural images), general relativistic magneto-hydrodynamic (GRMHD) simulations [329],

radially inefficient accretion flow (RIAF) simulations [54], and CelebA [213] (celebrity faces). We use our Bayesian imaging technique to apply each prior to the M87* observations, resulting in a collection of image posteriors. Each posterior is a probability distribution of images conditioned on the M87* data but incorporating a different prior. The visual biases of images from different posteriors are strikingly different, yet the images share structure that is prior-invariant, such as the ring shape and brightness asymmetry. We thus present two contributions based on our results: (1) a collection of possible M87* images that humans can selectively interpret based on their trust of the assumed biases and (2) analysis of which extracted black-hole features are robust to the prior and can be reliably used in scientific analysis.

In this chapter, we first present background on EHT VLBI measurements in Section 3.1 and our method in Section 3.2. In Section 3.3, we validate the imaging approach on simulated data using a collection of score-based priors ranging from weak biases (e.g., a prior trained on generic natural images) to strong biases (e.g., a prior trained on RIAF images). In Section 3.4, we present image posteriors of M87* based on the same collection of priors. Next, in Section 3.5, we analyze the influence of the prior on characteristic ring features, including diameter, width, and orientation, by performing tests on both the simulated-data and M87* images.

3.1 EHT Measurements

The EHT performs VLBI with a global array of radio telescopes. Each pair of telescopes i, j, known as a *baseline*, provides a Fourier measurement called a *visibility* v_{ij} [81, 347, 302]. However, the baselines only sparsely sample the complex 2D Fourier plane, or (u, v) space. Moreover, the visibilities are affected by thermal noise, station-dependent gain errors, and station-dependent phase errors [6].

To overcome station-dependent errors, we use robust data products known as closure quantities. A *closure phase* [175] is given by a triplet of telescopes i, j, k and computed as $\angle\left(v_{ij}v_{jk}v_{ki}\right)$. A *log closure amplitude* [310] is given by a combination of four telescopes i, j, k, ℓ and computed as $\log\left(\frac{|v_{ij}||v_{k\ell}|}{|v_{ik}||v_{j\ell}|}\right)$. We denote the set of all linearly independent observed closure phases as $\boldsymbol{y}_{\text{cp}} \in \mathbb{R}^{N_{\text{cp}}}$ and that of log closure amplitudes as $\boldsymbol{y}_{\text{logca}} \in \mathbb{R}^{N_{\text{logca}}}$. In the case of visibilities with a high signal-to-noise ratio (i.e., SNR > 1), closure phases and log closure amplitudes approximately experience mean-zero Gaussian thermal noise [257, 39, 53] with standard deviations $\boldsymbol{\sigma}_{\text{cp}} \in \mathbb{R}^{N_{\text{cp}}}$ and $\boldsymbol{\sigma}_{\text{logca}} \in \mathbb{R}^{N_{\text{logca}}}$, respectively. We assume the high-SNR setting. Conditioned on an image $\boldsymbol{x}$, the measurement distribution can be modeled as Gaussian with log likelihoods

$$\log p(\boldsymbol{y}_{\text{cp}} \mid \boldsymbol{x}) = -\frac{1}{2\boldsymbol{\sigma}_{\text{cp}}^2}\left\|f_{\text{cp}}(\boldsymbol{x}) - \boldsymbol{y}_{\text{cp}}\right\|_2^2 \quad \text{and} \tag{3.1}$$

$$\log p(\boldsymbol{y}_{\text{logca}} \mid \boldsymbol{x}) = -\frac{1}{2\boldsymbol{\sigma}_{\text{logca}}^2}\left\|f_{\text{logca}}(\boldsymbol{x}) - \boldsymbol{y}_{\text{logca}}\right\|_2^2, \tag{3.2}$$

where f_{cp} and f_{logca} are nonlinear forward models. We note that the Gaussian noise is not purely (statistically) independent, but it is approximately independent under high-SNR visibilities or can be made independent under a linear transformation [39, 53, 21]. Please refer to [113] for a tutorial on EHT imaging. Appendix A.1 provides details about VLBI data products and their forward models.

It is possible to use the same imaging algorithm with visibility amplitudes instead of log closure amplitudes. Visibility amplitudes, which have been used for other imaging results [7, 224], are more constraining than closure amplitudes, but they require calibration according to assumptions such as station-dependent systematic noise. In this work, we focus on using log closure amplitudes in order to avoid tuning the calibration assumptions. The original M87* work includes reconstructions from both types of data products for reference [7].

3.2 Method

To sample from an image posterior given EHT measurements and a score-based prior, we use the method discussed in Section 2.6 with a measurement likelihood based on closure quantities (Equation 3.1). We optimize a slightly modified version of the objective in Equation 2.12:

$$\phi^* = \arg\min_{\phi} \mathbb{E}_{\boldsymbol{x}\sim q_\phi}\Big[-\log p(\boldsymbol{y}_{\text{cp}} \mid \boldsymbol{x}) - \log p(\boldsymbol{y}_{\text{logca}} \mid \boldsymbol{x}) - b_{\boldsymbol{\theta}}^{\text{SDE}}(\boldsymbol{x}) + \big(V(\boldsymbol{x}) - \bar{V}\big)^2 + \log q_\phi(\boldsymbol{x})\Big]. \tag{3.3}$$

We include a flux-constraint objective $-\big(V(\boldsymbol{x}) - \bar{V}\big)^2$, where $V(\boldsymbol{x})$ is the total flux (i.e., the sum of the pixel values) of the image $\boldsymbol{x}$ and $\bar{V}$ is the target total flux, because closure quantities do not constrain the total flux. We set $\bar{V}$ as the median total flux of images sampled from the score-based prior $p_{\boldsymbol{\theta}}$ and then scale posterior images to the original total flux as measured in the zero-baseline visibility. Note that since our priors were trained on compact centered images, we do not need an explicit center-of-light constraint. We approximately solve Equation 3.3 with SGD, iteratively Monte-Carlo approximating the KL upper bound with a batch of samples from q_ϕ and computing the gradient with respect to ϕ.

Annealing the weight of the data-fit terms gradually from 0 to 1 helps prevent bad local minima. Furthermore, optimization can be sensitive to the chosen target flux $\bar{V}$ and data annealing schedule. We can mitigate this by making sure the diffusion model has a median total flux that is close to the median total flux of the training images. The data annealing may need to be tuned to achieve a local minimum at which the posterior images exhibit characteristic features of the prior (e.g., posterior images should be centered if all the training images for the prior are centered). Once ϕ is optimized, samples $\boldsymbol{x} \sim q_\phi$ can be efficiently obtained as approximate posterior samples. The RealNVP occasionally outputs slightly negative pixel values, so we clip samples to a minimum value of 0 to impose a positivity constraint.

3.2.1 Score-based priors used in this work

We focus on the following score-based priors, each trained on a dataset of images assuming a 128 μas field of view (FOV). Figure 3.1 shows samples from each prior.

- The **CIFAR-10** prior was trained on the CIFAR-10 [195] dataset of 32×32 images from 10 object classes (e.g., airplane, automobile, dog). We used a training set of 45K grayscale images. The images were tapered on the edges to incorporate assumptions of a black background and centered

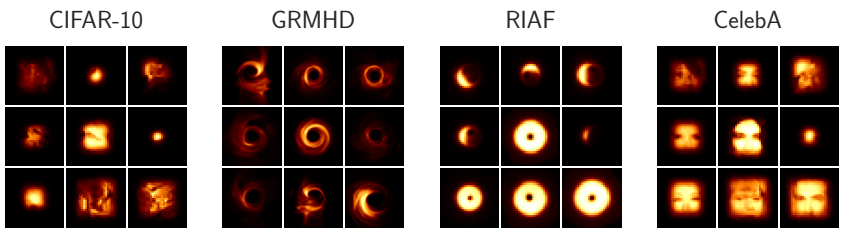


Figure 3.1: Samples from score-based priors used in this work.

object. A tapering effect was created by defining a binary mask with a center square region of pixels set to 1 and everywhere else set to 0, then applying a Gaussian blur kernel with a standard deviation of 8 µas, and then element-wise multiplying the blurred mask with the image. The size of the taper was randomly varied during training by randomly varying the size of the center square region of the mask, resulting in a centered compact region of between 12.8 µas × 12.8 µas and 83.2 µas × 83.2 µas.

- The **GRMHD** prior was trained on 100K images from GRMHD simulations [329] of Sgr A* resized to 64×64 pixels. During training, the GRMHD images were randomly flipped horizontally and randomly zoomed between -16.7% (zoomed-in) and $+14.5\%$ (zoomed-out), thus varying the diameter of the thin ring to be between 35 µas and 48 µas.
- The **RIAF** prior was trained on 9070 images of RIAF [54] simulations. The RIAF images were downloaded[1] with all available spin and inclination parameters and resized to 32×32. During training, they were also randomly zoomed between -16.7% and $+14.5\%$, randomly flipped horizontally, and randomly rotated between -2π and $+2\pi$.
- The **CelebA** prior was trained on the CelebA [213] dataset of celebrity faces. We used a training set of 160K images that were resized to 32×32. The same tapering effect that was used for CIFAR-10 was applied. Although far from a reasonable prior for astronomical images, a prior trained on faces helps us see what happens when strong but probably incorrect assumptions are made.

The RIAF and GRMHD priors incorporate strong assumptions about a ring structure. The CIFAR-10 and CelebA priors, on the other hand, do not assume any ring structure or even the presence of an astronomical object. One might make the following order of priors from weak to strong assumptions: CIFAR-10, GRMHD, RIAF. The CelebA prior makes specific assumptions against our expectations.

[1] `http://vlbiimaging.csail.mit.edu/myData`

3.3 Results with Simulated Data

We validate our imaging approach on simulated observations of synthetic source images. Whereas previous work [7] had to validate on simulated data to calibrate hyperparameters, the experiments presented in this section simply verify the efficacy of our approach, which does not require hyperparameter-tuning.

Figure 3.2 shows results for a dataset of eight source images. Among the images are validation images used in the original M87* imaging work [7] and two images of an elliptical object used in the Sgr A* imaging work [10]. All observations were simulated based on the April 6 observing array using code provided by Akiyama et al. [7]. We used closure phases and log closure amplitudes of the combined low-band and high-band data and followed the same preprocessing steps as the `eht-imaging` algorithm (assuming non-closing fractional systematic noise of 0.03), except we did not add station-dependent systematic noise since we do not need to calibrate the visibility amplitudes. Although imaging was done with a prior-dependent total flux and either 32×32 or 64×64 pixels depending on the prior, we re-scale images to have a total flux of 0.6 Jy and resize them to 128×128 for visualization.

The quality of image reconstruction heavily depends on the prior. For example, the GRMHD reconstruction of GRMHD 1 appears more convincing than the GRMHD reconstruction of the Double image in Figure 3.2. On the other hand, the RML methods used in previous EHT imaging efforts [7, 10] achieve overall cleaner reconstructions of synthetic data. One reason for the better performance of those RML methods is that they use regularization parameters chosen based on a calibration dataset that is very similar to their test images. In contrast, we consider priors that may be profoundly different from the true source image (e.g., CelebA prior applied to the Double data). Another reason is that RML methods produce a mean image that tends to be cleaner than individual posterior samples, which are shown in Figure 3.2. We emphasize that the goal of our work is not to achieve the cleanest or most accurate reconstructions; rather, we aim to showcase the effects of different priors, even when those priors might not lead to the best reconstruction due to mismatch with the data.

Overall, the reconstructed images make sense according to the biases of the prior. The CelebA prior introduces face-like features, especially when the ground-truth source object is fairly "flat" (e.g., the Disk and Elliptical images), and it is the only prior that leads to multimodal estimated posteriors. Images under the RIAF prior are always centered and ring- or disk-like. The GRMHD prior always prefers the

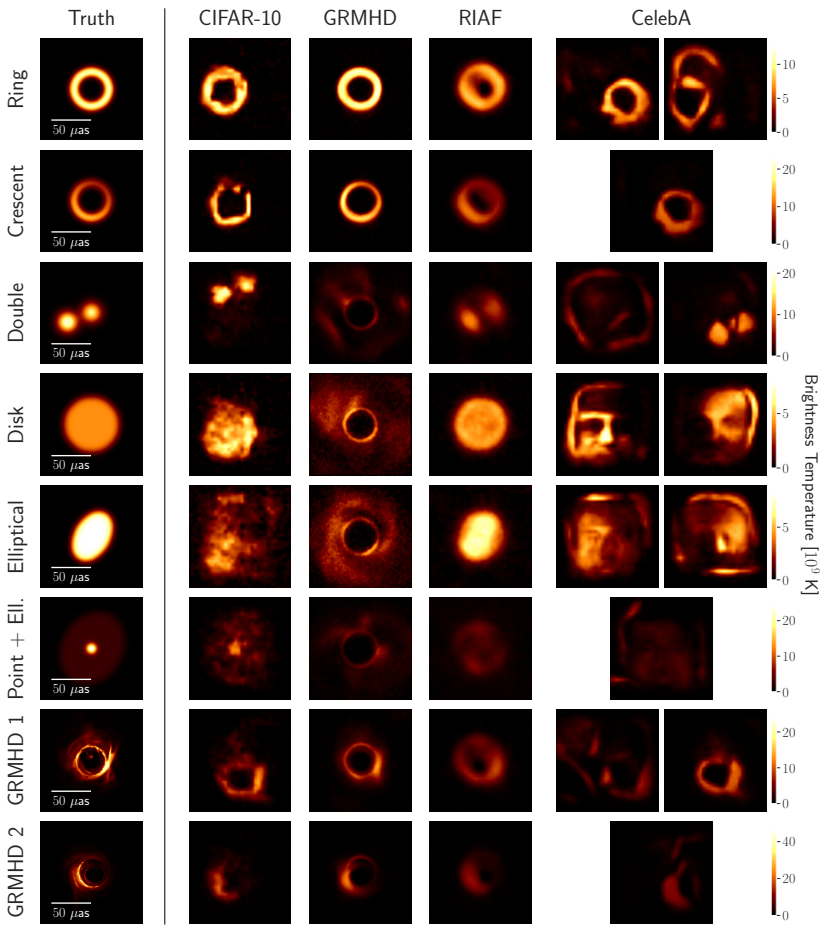


Figure 3.2: Image reconstructions from simulated data. A sample (one sample from each mode if the posterior is bimodal) is shown from each estimated posterior. Qualitatively, the CIFAR-10 prior adds the least amount of bias, producing reasonable reconstructions of each image in this dataset. The GRMHD prior strongly prefers a centered ring in the image. The RIAF prior prefers a centered ring- or disk-like structure in the image. The CelebA prior struggles to recover these source images, in some cases adding face features, and it leads to the most multimodal posteriors. However, it performs decently well on certain images like the Crescent and GRMHD images. When the source image is known to be well-approximated by a GRMHD or RIAF model, the more constrained GRMHD or RIAF prior may be the best choice.

presence of a thin ring at the center of the image. The CIFAR-10 prior imposes weak biases and appears to assemble images from small, locally-smooth patches.

Table 3.1 quantifies the performance of the various priors on each source image. As in previous work [7, 10], we evaluate the normalized cross-correlation (NCC). Since our approach does not explicitly constrain the center-of-light, we use a shift-invariant NCC metric, which is computed as the maximum NCC between all shifted versions of the reconstructed image and the ground-truth image. As expected, the

Image	CIFAR-10	GRMHD	RIAF	CelebA
Ring	0.95 ± 0.009	**0.98** ± 0.024	0.90 ± 0.003	0.88 ± 0.090
Crescent	0.85 ± 0.019	**0.96** ± 0.007	0.92 ± 0.005	0.95 ± 0.007
Double	**0.94** ± 0.009	0.49 ± 0.028	0.93 ± 0.004	0.39 ± 0.130
Disk	0.95 ± 0.007	0.59 ± 0.027	**0.99** ± 0.002	0.80 ± 0.020
Elliptical	0.79 ± 0.037	0.50 ± 0.029	**0.97** ± 0.002	0.75 ± 0.026
Point + Ell.	**0.87** ± 0.022	0.58 ± 0.031	0.68 ± 0.010	0.43 ± 0.026
GRMHD 1	0.84 ± 0.015	**0.86** ± 0.009	0.83 ± 0.004	0.46 ± 0.089
GRMHD 2	0.85 ± 0.012	**0.90** ± 0.003	0.85 ± 0.006	0.77 ± 0.012

Table 3.1: Normalized cross-correlation (NCC) metrics for simulated data. The avg. ± std. dev. of the NCC metric for 128 samples from the posterior is reported (highest NCC in each row is shown in bold). In general, the closer the prior is to the ground-truth image, the closer its posterior samples are to the ground truth. For example, the Ring, Crescent, and GRMHD images are best reconstructed with the GRMHD prior, whereas non-ring-like images are poorly reconstructed with the GRMHD prior. The CIFAR-10 prior may be the best "general-purpose" prior, giving NCC values between about 0.80 and 0.95 across these images.

closer the prior is to the ground-truth image, the more accurately it recovers the ground-truth image. For example, the GRMHD prior excels at recovering the Ring, Crescent, and GRMHD images but struggles with the non-ring-like images. The RIAF prior performs well on ring-like images and disk-like images. The CelebA prior, unsurprisingly, performs poorly on this dataset of images. The CIFAR-10 prior does best compared to the other priors at recovering the non-ring-like images (e.g., Double and Point+Elliptical). It performs generally well across all the source images, suggesting that it serves as an effective "general-purpose" prior.

Table 3.2 quantifies agreement with the simulated data using the reduced χ^2 metric—we note that it is not a true reduced χ^2 since we only incorporate image pixels as degrees of freedom, but it is useful as a proxy metric of data consistency. There does not appear to be a correlation between the χ^2 statistics in Table 3.2 and the NCC statistics in Table 3.1. The results in Table 3.2 simply confirm data consistency of the reconstructed images, as χ^2 values are consistently less than 2 and often close to 1 (lower χ^2 corresponds to more data consistency, and $\chi^2 \approx 1$ is considered a sign of a good balance between data and prior). The RIAF prior results in the highest χ^2 values, perhaps because it is the most constraining prior. Overall, our tests on simulated data confirm that the score-based priors impose the expected biases on the image reconstruction while allowing for reasonable data consistency.

Image	Data Product	CIFAR-10	GRMHD	RIAF	CelebA
Ring	χ^2_{cp}	0.87 ± 0.02	0.90 ± 0.03	0.96 ± 0.02	0.89 ± 0.04
	χ^2_{logca}	0.72 ± 0.03	0.75 ± 0.05	1.04 ± 0.06	0.72 ± 0.03
Crescent	χ^2_{cp}	0.73 ± 0.02	0.76 ± 0.02	0.84 ± 0.03	0.73 ± 0.02
	χ^2_{logca}	0.67 ± 0.02	0.78 ± 0.03	0.95 ± 0.05	0.68 ± 0.02
Double	χ^2_{cp}	0.96 ± 0.02	1.00 ± 0.03	1.01 ± 0.03	0.99 ± 0.03
	χ^2_{logca}	0.77 ± 0.02	0.81 ± 0.04	1.31 ± 0.08	0.84 ± 0.03
Disk	χ^2_{cp}	1.82 ± 0.04	1.83 ± 0.05	1.80 ± 0.03	1.81 ± 0.05
	χ^2_{logca}	1.30 ± 0.04	1.56 ± 0.14	1.41 ± 0.04	1.42 ± 0.06
Elliptical	χ^2_{cp}	1.76 ± 0.04	1.76 ± 0.04	1.91 ± 0.04	1.79 ± 0.04
	χ^2_{logca}	1.44 ± 0.03	1.64 ± 0.06	1.64 ± 0.05	1.43 ± 0.03
Point + Ell.	χ^2_{cp}	1.22 ± 0.02	1.21 ± 0.02	1.42 ± 0.04	1.24 ± 0.02
	χ^2_{logca}	0.83 ± 0.02	0.83 ± 0.03	0.96 ± 0.03	0.84 ± 0.02
GRMHD 1	χ^2_{cp}	0.90 ± 0.02	0.89 ± 0.07	1.02 ± 0.04	0.90 ± 0.05
	χ^2_{logca}	0.72 ± 0.03	0.70 ± 0.03	1.08 ± 0.07	0.83 ± 0.03
GRMHD 2	χ^2_{cp}	0.59 ± 0.02	0.59 ± 0.01	0.62 ± 0.04	0.60 ± 0.02
	χ^2_{logca}	0.51 ± 0.02	0.56 ± 0.03	1.01 ± 0.08	0.53 ± 0.02

Table 3.2: Data-consistency metrics (χ^2) for closure quantities of simulated data. χ^2_{cp} and χ^2_{logca} are the χ^2 metrics for closure phases and log closure amplitudes, respectively. The avg. ± std. dev. of 128 samples from the estimated posterior is reported. Lower χ^2 is a sign of higher data consistency. $\chi^2 \approx 1$ is considered an indication of a good balance between the observed data and the prior. The Disk, Elliptical, and Point+Elliptical images are the most challenging cases for these particular priors, as evidenced by the high χ^2 values that indicate data-fitting difficulty.

3.3.1 Characterizing the simulated-data posteriors

In addition to evaluating single samples from the posterior (Figure 3.2), we can assess aspects of the posterior distribution such as uncertainty and multimodality. Figure 3.3 shows the mean and pixel-wise standard deviation of posteriors under the CIFAR-10, GRMHD, and RIAF priors. We find that uncertainty decreases as the prior becomes stronger. For a weak prior like CIFAR-10, which leads to high posterior uncertainty, it can be helpful to consider the mean reconstruction instead of noisier individual samples.

The CelebA prior leads to bimodal posteriors, which are characterized in Figure 3.4. We note that the number of modes in the estimated posterior may be due to the variational family being used; in the future, a more expressive family of

distributions parameterized through a better network architecture may identify more modes. It is perhaps reassuring that a prior with such erroneous assumptions (i.e., that face statistics well describe these particular source images) is able to account for mismatches with the data through a wider posterior. For example, for the Ring, Double, and GRMHD 1 data, the posterior covers two modes, one of which accurately recovers the ground-truth image.

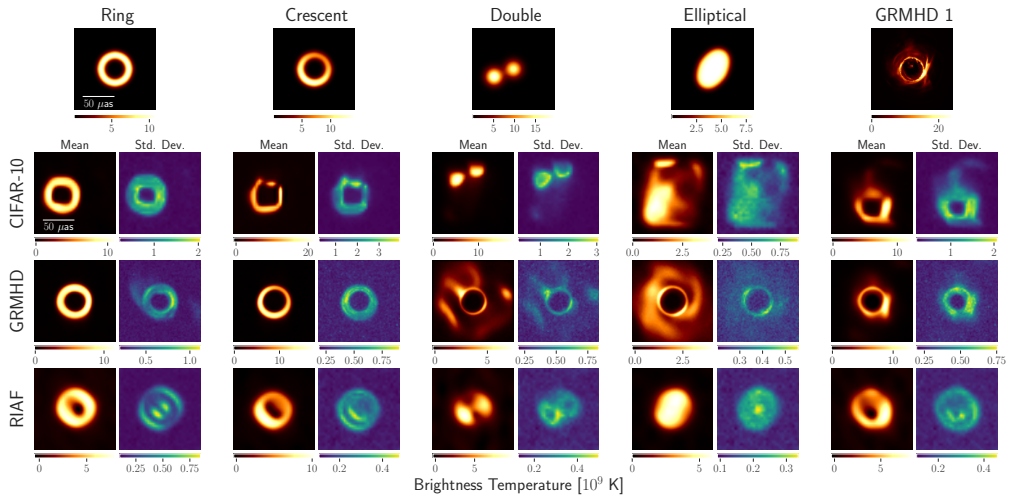


Figure 3.3: Mean and std. dev. of simulated-data posteriors. We find that uncertainty decreases with a stronger prior (i.e., maximum std. dev. decreases from CIFAR-10 to GRMHD to RIAF). CIFAR-10 exhibits the most uncertainty given that it is the most flexible of the priors. Compared to the individual samples in Figure 3.2, the mean images appear much cleaner.

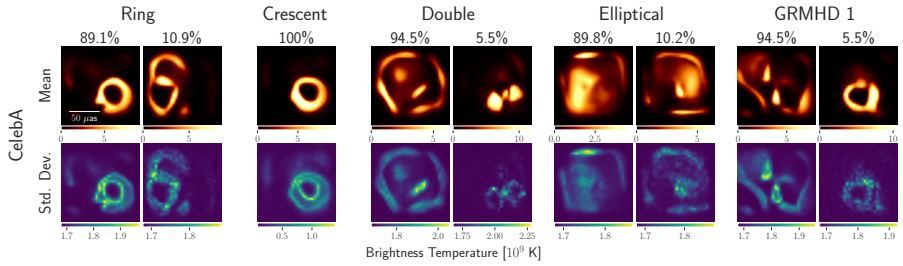
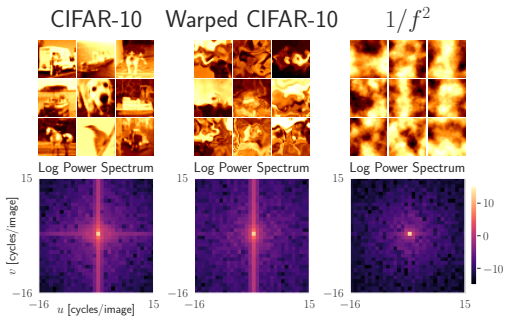


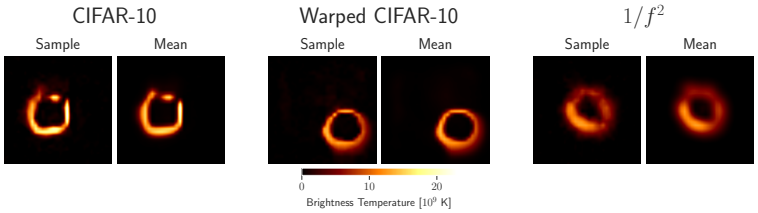
Figure 3.4: Mean and std. dev. of simulated-data posteriors under the CelebA prior. The CelebA prior is the only one that gives rise to bimodal posteriors. For each posterior, a one- or two-component Gaussian mixture model (one for a single-mode posterior and two for a bimodal posterior) was fit to 128 samples. The mean, pixel-wise std. dev., and weight of each Gaussian component are shown. For the Ring, Double, and GRMHD 1 data, one of the two modes is quite similar to the ground-truth image, while the other is not.

3.3.2 Biases of the CIFAR-10 prior

While CIFAR-10 represents a "generic" natural-image prior, the dataset itself still contains biases. The CIFAR-10 dataset comprises upright images of animals and man-made objects, which tend to exhibit horizontal or vertical lines. As Figure 3.5a shows, the average log power spectrum of CIFAR-10 images has the most power in the purely horizontal or vertical spatial frequencies. This preference for vertically- or horizontally-oriented edges results in images of objects that look somewhat rectangular instead of circular, even given measurements of a ring structure. See, for example, the CIFAR-10 reconstructions from the Crescent and GRMHD 1 data in Figures 3.2 and 3.3 or the April 10 and 11 CIFAR-10 reconstructions of M87* in the following section. In Figure 3.5, we demonstrate on the Crescent data how boxy artifacts can be mitigated with a prior trained on warped CIFAR-10 images or a prior trained on images with a $1/f^2$ spectral distribution. By distorting CIFAR-10 images with warped random affine transforms, we expect to reduce the presence of straight lines by perturbing them to have more curvature. Alternatively, by randomly sampling from a $1/f^2$ spectral distribution, we create a dataset of images that follow a simplified statistical model without any preference for straight lines.



(a) Image statistics of CIFAR-10 and alternative natural-image datasets.



(b) Crescent reconstructions under alternative natural-image priors.

Figure 3.5: Reducing the CIFAR-10 bias for horizontal/vertical lines. For example, given simulated data of the Crescent image from Figure 3.2, this bias results in posterior images containing a boxy object, even though the actual Crescent is a ring without any straight edges. This is likely due to the statistics of the CIFAR-10 dataset, which contains mostly man-made objects and animals. Such images include many sharp corners and lines arising from boxy objects like cars, the legs of standing animals, or horizon lines. Although CIFAR-10 is our choice of a generic natural-image prior, it is possible to define an alternative natural-image prior with less preference for horizontal/vertical lines. We tested a prior trained on CIFAR-10 images that underwent warped random affine transforms and a prior trained on 32×32 images with power spectral density proportional to $1/f^2$ (since the power spectra of natural images have been shown to follow a $1/f^\alpha$ trend, where f is the spatial frequency in cycles per image [267], and α is typically between 1 and 3). The CIFAR-10, Warped CIFAR-10, and $1/f^2$ priors were all trained with the same number of training images (45K), with the same tapering effect described in Section 3.2.1, and for the same number of iterations (100K). **(a)** Nine training samples and the average log power spectrum of 10K training samples (without the taper) for each prior. The average spectral power of CIFAR-10 images is relatively high in the horizontal and vertical frequencies. Warping the CIFAR-10 images reduces the preference for vertical frequencies, but a large presence of horizontal frequencies remains (perhaps because horizons are still distinguishable after warping, as can be seen in some of the training samples). The $1/f^2$ noise images have isotropically-distributed spectral power. **(b)** Results of imaging the Crescent simulated data under the different priors. A posterior sample and the average of 128 posterior samples are shown for each prior. With the regular CIFAR-10 prior, there is a sharp corner at the top-right of the ring-like object that makes it look squarish. This artifact is not present in the Warped CIFAR-10 or $1/f^2$ reconstructions.

3.4 Results with Real M87* Data

We estimated image posteriors of M87* given the EHT observations from 2017 April. For each of the four observation days, we gathered the closure phases and log closure amplitudes from the combined low-band and high-band public data.[2] We used the same data preprocessing and visualization steps as for the synthetic data.

Figure 3.6 shows a posterior sample for each of the observation days, under each prior. The CIFAR-10 and CelebA priors, which incorporate no assumptions about black holes, lead to images with a ring-like structure. The CelebA reconstructions include some face-like details, especially for April 10, the day with the fewest observations and thus the day whose image is least constrained by measurements. Nonetheless, the fact that both the black-hole-agnostic priors recover ring shapes is strong evidence of a ring-like structure in the measurements. It also reveals that basic natural-image statistics shared by CIFAR-10 and CelebA may be sufficient to retrieve the ring structure under a constrained FOV.

The GRMHD and RIAF priors lead to images with a well-defined ring structure. The GRMHD prior is visually richer, encouraging a thin ring with wispy features in the bright region of the ring. The RIAF prior constrains the image according to a simplified geometric model without adding any lower-level details.

Table 3.3, which reports reduced χ^2 values, confirms that all the reconstructed images are consistent with the EHT measurements. Across priors and observation days, the χ^2 values are < 1.5, which indicates fitting the measured data. As with the simulated data, the RIAF prior struggles most to fit the data (most of its χ^2 values are greater than 1) because it imposes the most constraining black-hole model. The rest of the priors tend to result in $\chi^2 < 1$, which means that they are flexible enough to somewhat overfit the data and that any differences between their posteriors are likely due to the visual biases of the priors that are not constrained by the data.

3.4.1 Characterizing the M87* posteriors

Our imaging approach leads to single-mode M87* image posteriors, except for some CelebA posteriors that are bimodal. Figure 3.7 shows Gaussian fits of the posteriors estimated under the CIFAR-10, GRMHD, and RIAF priors. Under all these priors, the mean image shows a clear progression of the bright spot of the ring moving counter-clockwise over the four observation days. The pixel-wise standard

[2] `https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/EHTC_FirstM87Results_Apr2019`

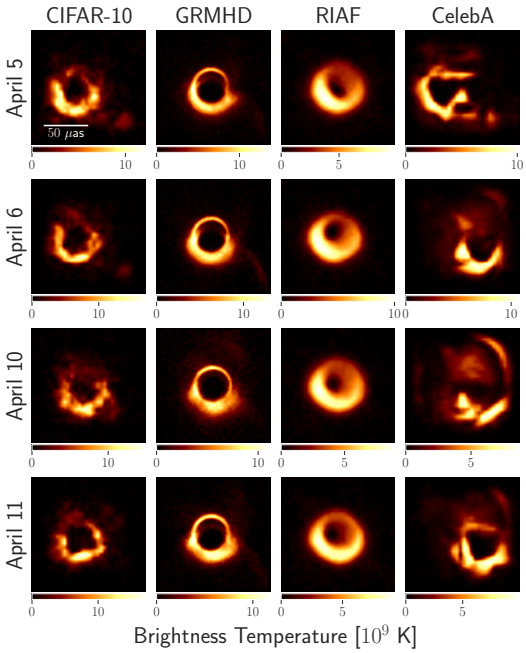


Figure 3.6: M87* posterior samples. A random posterior image is shown for each observation day and prior. The CIFAR-10 prior was trained on images of everyday objects and makes no assumptions of black-hole structure, yet it recovers a ring-like structure for all four observation days. The GRMHD prior assumes a fluid-flow model of black holes, which helps it recover visually-striking images of a thin ring with wisps. The RIAF prior assumes a simplified crescent model of black holes, which results in simplified crescent images of M87*. The CelebA prior was trained on images of human faces, so its preferred images are presumably far away from the true source image. Even with its incorrect and strong biases, the CelebA prior recovers a ring-like structure, here favoring an eye from the face prior to explain the ring. These images under various priors all fit the EHT observations but incorporate different visual biases.

deviation shows areas of uncertainty around the mean. Similarly to our results with simulated data, the results in Figure 3.7 show that as the prior becomes stronger (i.e., from CIFAR-10 to GRMHD to RIAF), the uncertainty goes down. Under the CIFAR-10 and GRMHD priors there is uncertainty throughout the ring, whereas under the RIAF prior uncertainty lies mainly along the edges of the ring.

Figure 3.8 visualizes the bimodal posteriors under the CelebA prior. Like with CIFAR-10, the magnitude of the uncertainty is higher than that of the GRMHD and RIAF priors. The presence of multiple modes further reflects higher uncertainty under the CelebA prior, which makes sense for a prior that is so mismatched with the observed data.

Image	Data Product	CIFAR-10	GRMHD	RIAF	CelebA
April 5	$\chi^2_{\rm cp}$	0.80 ± 0.03	0.83 ± 0.15	1.27 ± 0.06	0.80 ± 0.03
	$\chi^2_{\rm logca}$	0.75 ± 0.04	0.76 ± 0.09	1.50 ± 0.10	0.71 ± 0.04
April 6	$\chi^2_{\rm cp}$	0.92 ± 0.02	0.93 ± 0.05	1.08 ± 0.03	0.93 ± 0.03
	$\chi^2_{\rm logca}$	0.76 ± 0.02	0.74 ± 0.04	1.00 ± 0.05	0.80 ± 0.06
April 10	$\chi^2_{\rm cp}$	0.92 ± 0.05	0.89 ± 0.04	1.25 ± 0.07	0.90 ± 0.05
	$\chi^2_{\rm logca}$	0.73 ± 0.06	0.71 ± 0.04	1.35 ± 0.12	0.72 ± 0.06
April 11	$\chi^2_{\rm cp}$	1.09 ± 0.02	1.06 ± 0.02	1.26 ± 0.04	1.08 ± 0.03
	$\chi^2_{\rm logca}$	0.69 ± 0.04	0.65 ± 0.04	1.01 ± 0.06	0.65 ± 0.03

Table 3.3: Data-consistency metrics (χ^2) for closure quantities of M87* data. $\chi^2_{\rm cp}$ and $\chi^2_{\rm logca}$ are the χ^2 metrics for closure phases and log closure amplitudes, respectively. The avg. ± std. dev. of 128 samples from the estimated posterior is reported. $\chi^2 \approx 1$ indicates a good balance between fitting the observed data and fitting the prior. Lower χ^2 means more data consistency. The RIAF prior leads to the highest χ^2 values, meaning the strength of this prior relative to the data is strongest. This is probably because the RIAF model is the most constraining of the priors.

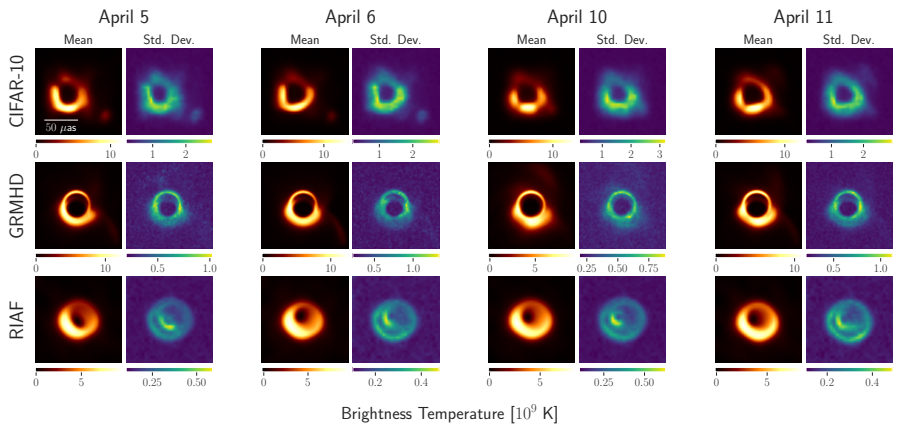


Figure 3.7: Mean and std. dev. of posterior samples. Like with the simulated data in Figure 3.3, uncertainty decreases as the prior becomes stronger. Note that under the CIFAR-10 prior the day with the most uncertainty is April 10, the day with the least data. The stronger priors (i.e., GRMHD and RIAF) do not exhibit more uncertainty for this day compared to other days.

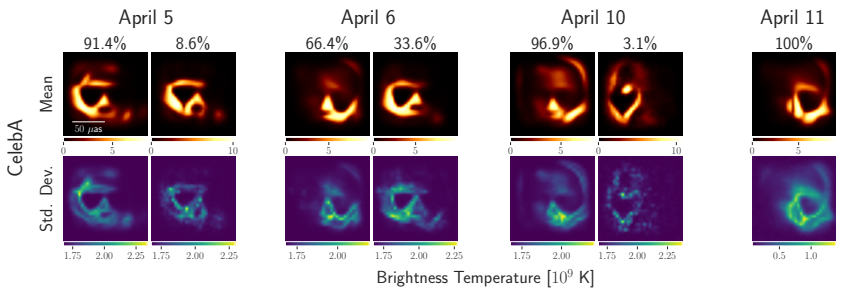


Figure 3.8: Bimodal M87* image posteriors under the CelebA prior. A two-component Gaussian mixture model was fit to 128 posterior samples for April 5, 6, and 10. The mean, std. dev., and weight of each Gaussian mixture component are shown. The only single-mode posterior is for April 11, which is the day with the most (u, v) coverage.

A feature that stands out in the CIFAR-10, GRMHD, and CelebA reconstructions is a southwest region of extended flux outside the ring. It is especially noticeable in the April 5 and 6 images. See, for example, the faint spot of brightness in the CIFAR-10 reconstructions and the faint wisp to the southwest in the GRMHD reconstructions in Figures 3.6 and 3.7. A disconnected southwest region of brightness also appears in both modes of the CelebA posterior on April 5 and in the second mode of the April 6 posterior in Figure 3.8. Such a feature is not visible in the RIAF reconstructions. With previous imaging results that only incorporated one prior, it would have been difficult to conclude whether this feature was an artifact of imaging or a clue from the data. Our results in Figures 3.7 and 3.8 suggest that it is a prior-dependent feature, with different priors placing different amounts of brightness in that southwest region.

To summarize our findings from the estimated M87* posteriors, the most notable result is that all priors recover ring-like structure. The priors that do not assume a black hole (i.e., CIFAR-10 and CelebA) exhibit most uncertainty in the posterior and are most flexible with adding flux outside of the ring. The priors based on a black-hole model (i.e., GRMHD and RIAF) reconstruct the clearest rings, with the GRMHD prior providing the most visual detail. In general, the more constraining the prior, the less uncertainty there is in the posterior, but the more potential there is to overfit to prior assumptions.

3.5 Extracted Ring Features

An important stage of analysis is to extract ring features from reconstructed images, distilling any ring structure into a few parameters dictating its geometry and brightness profile. In Akiyama et al. [7, 9], the EHT Collaboration extracted, among other quantities, the following characteristic ring features: diameter, width, orientation, asymmetry, and fractional central brightness. It found the ring diameter and orientation angle to be most consistent across imaging methods, with the other quantities varying depending on the imaging pipeline. We focus on these characteristic features and analyze the effect of the prior on them.

The **diameter** d indicates the full size of the ring and is calculated based on the distance between the peak brightness and the ring center. The **width** w indicates the thickness of the ring itself. The **orientation** angle n, measured east of north, roughly indicates the radial position of most of the brightness. The azimuthal **asymmetry** A, a measure of brightness asymmetry throughout the ring, roughly indicates the magnitude of brightness at the measured orientation. The **fractional central brightness** f_C indicates how bright the interior of the ring is compared to the ring itself and can be considered an inverse brightness contrast ratio. We use the `REx` feature extraction algorithm [62] implemented in the `eht-imaging` library. The parameters d and f_C are measured the same way as in Akiyama et al. [7], but the other features have slight differences (we default to the `REx` implementation rather than the equations provided by Akiyama et al. [7]). Appendix A.2 contains the exact equations we used to compute these features. Figure 3.9 helps visualize some of the features.

3.5.1 Ring features of simulated-data reconstructions

We performed feature extraction on images from simulated data, focusing on the source images with ring-like structure: Crescent, Ring, GRMHD 1, and GRMHD 2. We analyzed the recovered posteriors under all the score-based priors but excluded the non-ring-like samples from the CelebA posteriors for the Ring and GRMHD 1 data (see Figure 3.2, which shows that the second mode of the Ring posterior and the first mode of the GRMHD 1 posterior do not have a ring-like structure). Figure 3.10 shows all the extracted features and their error bars. In the following paragraphs, we discuss each feature and its dependence (or lack thereof) on the prior.

Diameter All priors recover the diameter of the source image (within one standard deviation from the mean extracted diameter). The only exception is that the RIAF

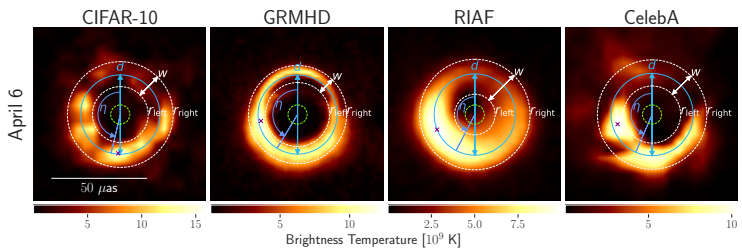


Figure 3.9: Visualization of extracted ring features of M87* images. Shown is a random sample (centered by `REx`) from each posterior for April 6. Parameter d is the ring diameter; w is the ring width. Parameters r_{left} and r_{right} delimit the radial FWHM used to estimate the orientation angle and asymmetry. Parameter η is the orientation angle measured east of north. The purple cross marks the location of peak brightness. Asymmetry A and fractional central brightness f_C are not visualized. The green dashed circle demarcates the inner disk of radius 5 μas used to define f_C.

prior leads to a larger diameter for GRMHD 1. A possible explanation is that the RIAF model is too strong of a prior for these data, as it must account for all the flux in the image with a thick crescent (as evidenced by the relatively high extracted diameter and width of the RIAF-reconstructed images). The RIAF prior also leads to the highest χ^2 values (Table 3.2), further evidence that it has difficulty fitting the GRMHD 1 observations. On the other hand, the RIAF prior correctly recovers the diameter and has lower χ^2 values for the Ring and Crescent, two objects that can be well-approximated with a RIAF model. Even though the diameter should be well-constrained by the measurements, our results demonstrate how a strong-enough prior may recover a ring structure but with an incorrect diameter.

The CIFAR-10 prior gives the most accurate mean diameter, although with greater uncertainty than the GRMHD and RIAF priors. This makes sense, as the CIFAR-10 prior, in making the weakest assumptions, is most flexible with the image reconstruction. The CelebA prior exhibits significantly high uncertainty for GRMHD 2, perhaps because its recovered images have the least ring-like structure (Figure 3.2). The GRMHD prior accurately recovers the diameter across all these ring-like data and with relatively little uncertainty.

Width The ring width varies significantly with the prior, which supports previous findings that the width is less well-constrained by the measurements than the diameter is [7, 9]. The GRMHD prior recovers the thinnest rings as a result of being trained on GRMHD images that exhibit thin rings.

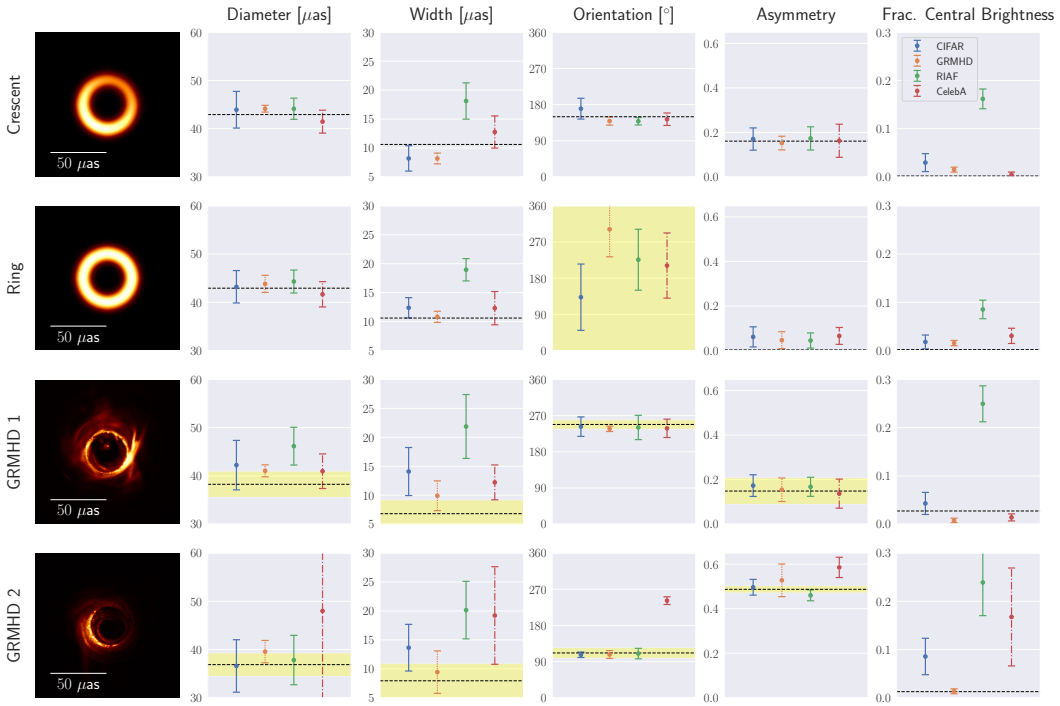


Figure 3.10: Extracted ring features of images reconstructed from simulated data. The dotted line indicates the measured quantity of the ground-truth image, and the shaded yellow region indicates the std. dev. as computed by REx (the orientation of the Ring image was manually set to span 0° to 360°). For the score-based priors (CIFAR, GRMHD, RIAF, CelebA), the mean was computed across the measured quantities of 128 posterior samples. The std. dev. for all features except fractional central brightness (f_C) was computed as the square root of the average variance across these samples. Since REx does not provide uncertainty for f_C of a single image, the f_C error bars were computed as the std. dev. of the f_C values of all posterior samples. All choices of prior roughly recover the target diameter, orientation, and asymmetry. The only exceptions are that the RIAF prior over-estimates the diameter of GRMHD 1 (due to the difficulty of fitting GRMHD data to a strong RIAF prior) and that the CelebA prior over-biases the orientation of GRMHD 2 (due to the CelebA prior struggling to produce a strong ring structure with such data). The width and fractional central brightness are particularly prior-dependent, with the GRMHD prior providing consistently small values for these features. In contrast, the RIAF prior prefers large widths and high f_C values. Please see Figure 3.2 for image samples.

Orientation All priors recover the orientation angle of the source image, except the CelebA prior given GRMHD 2 data. Like the high diameter uncertainty, this may be due to the reconstructed images having relatively weak ring structure. As can be seen in the two samples from this posterior in Figure 3.2, there is actually some brightness outside of the ring-like area. These results further highlight that strong and incorrect assumptions in the prior may inhibit correct recovery of ring features that should be constrained by the observations. Besides this extreme case of applying a CelebA prior to GRMHD data, the ring orientation appears to be independent of the prior.

Asymmetry The brightness asymmetry appears to be fairly robust to the prior. Again, the exception is the CelebA prior applied to the GRMHD 2 data, which can be explained by the weak ring structure present in the CelebA-recovered images. We note that for the Ring data, all priors produce a slight asymmetry even though the ground-truth object has no asymmetry, which is unavoidable with most imaging algorithms [7].

Fractional central brightness This feature varies extremely with the prior, and the original M87* imaging work also found that f_C is not well-constrained by the data [7]. The GRMHD prior recovers the lowest fractional central brightness for the GRMHD observations. Along with smaller ring widths, these low f_C values indicate that using a well-matched GRMHD prior on GRMHD observations gives the benefit of sharper images than could be obtained with a weaker prior like CIFAR-10.

3.5.1.1 Discussion

To summarize our results on simulated data, we find that the recovered ring diameter, orientation angle, and asymmetry are fairly robust to the image prior, while the width and fractional central brightness are tied to the prior. However, it is possible to overly bias the diameter with an overly-biased prior. In particular, the RIAF prior applied to observations of a GRMHD simulation with flux extending beyond the ring may be too constraining and cause over-estimation of the diameter. Reassuringly, priors that are less constraining or more accurate dependably recover the diameter, orientation, and asymmetry. On the other hand, features like the ring width and fractional central brightness can be adjusted by imposing different priors.

3.5.2 Ring Features of M87* Reconstructions

Figure 3.11 and Table 3.4 show the results of feature extraction for the M87* image reconstructions. Table 3.4 includes the results of using the `eht-imaging` algorithm with fiducial parameters [7] for reference, although `eht-imaging` is not directly comparable to our score-based priors since it utilizes hand-crafted regularizers as a proxy for a prior and visibility amplitudes instead of closure amplitudes. Figure 3.9 visualizes the ring features on April 6 reconstructions. We assumed ring structure in all the posterior samples under the different score-based priors. As shown in Figures 3.6 and 3.8, the CelebA images have the weakest ring structure, which may have caused higher variance in the extracted features.

The CIFAR-10 prior recovers a mean diameter of 41.2 μas to 42.5 μas across the four observations days. Accounting for error bars, all score-based priors agree on a range of possible diameters. The diameters recovered by our score-based priors are consistent with the diameter recovered by `eht-imaging`, although with slightly higher means and larger error bars. Like with the simulated data, the RIAF prior leads to relatively high diameters. Combined with the fact that the RIAF prior has the highest χ^2 values in Table 3.3, this suggests that the RIAF model is likely too constraining for M87*.

The ring width depends on the prior, with the GRMHD prior resulting in the lowest width (about 9 μas). The RIAF prior causes the largest width (about 20 μas), which is only slightly larger than the width recovered by `eht-imaging` (about 16 μas, as listed in Table 3.4).

The orientation angle is consistent across the priors, roughly ranging from 150° on April 5 to 170° on April 11. Like previous work, we find that both the diameter and orientation angle have an upward trend over the observation days as brightness shifts to the southwest [7, 9].

The amount of asymmetry is also roughly consistent across the priors. Interestingly, there is the most discrepancy between the priors in the April 10 reconstructions. April 10 is the day with fewest observations, which may cause the brightness asymmetry to be more flexible under the data.

As expected, the fractional central brightness varies significantly with the prior. As with the simulated data, the GRMHD prior achieves the greatest brightness contrast.

Overall, the conclusions regarding the effects of priors on the M87* reconstructions are consistent with the conclusions of previous analyses [7, 9] and our simulated-data

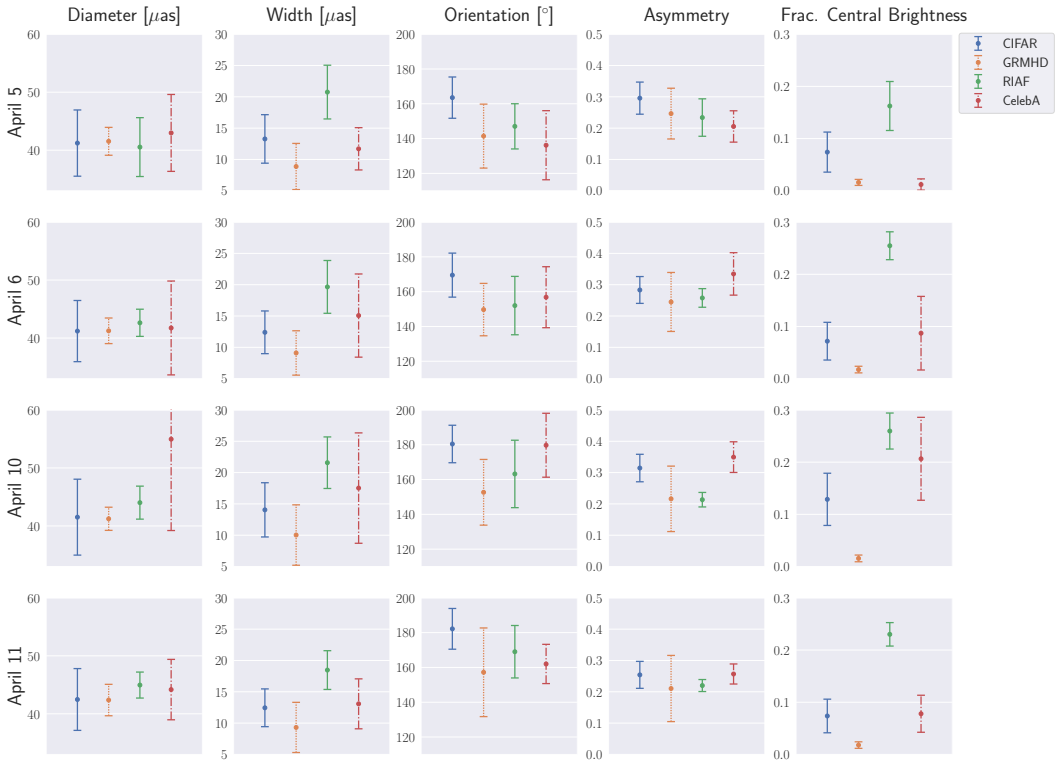


Figure 3.11: Extracted ring features of M87* images. The means and error bars were computed the same way as for Figure 3.10. The different priors (CIFAR, GRMHD, RIAF, CelebA) all agree on the diameter, orientation, and asymmetry up to error bars. There is some disagreement in asymmetry for April 10, which is the day with fewest observations. We note a slight upward trend in diameter and orientation angle over the observation days. The width and fractional central brightness change with the prior, with the GRMHD prior providing the thinnest rings and most brightness contrast. Please see Figure 3.6 for image samples.

analysis. The diameter, orientation, and asymmetry are robust to the prior, although some priors result in greater variability than others. There is a slight upward bias in diameter from the RIAF prior, due to wider rings that account for all the flux in one RIAF model. The most prior-dependent features are the width and fractional central brightness.

Feature	Date	`eht-imaging`*	CIFAR-10	GRMHD	RIAF	CelebA
Diameter [μas]	April 5	39.3 ± 1.6	41.2 ± 5.7	41.5 ± 2.4	40.5 ± 5.1	43.0 ± 6.6
	April 6	39.5 ± 1.5	41.2 ± 5.3	41.2 ± 2.2	42.6 ± 2.3	41.7 ± 8.1
	April 10	40.5 ± 1.3	41.5 ± 6.5	41.2 ± 2.0	44.0 ± 2.8	55.0 ± 15.8
	April 11	41.1 ± 1.2	42.5 ± 5.3	42.4 ± 2.7	45.0 ± 2.2	44.2 ± 5.2
Width [μas]	April 5	16.3 ± 1.5	13.3 ± 3.9	8.9 ± 3.7	20.8 ± 4.3	11.7 ± 3.4
	April 6	16.2 ± 1.0	12.4 ± 3.4	9.1 ± 3.5	19.7 ± 4.2	15.1 ± 6.6
	April 10	15.7 ± 1.3	14.1 ± 4.3	10.0 ± 4.8	21.6 ± 4.1	17.5 ± 8.8
	April 11	15.6 ± 0.9	12.4 ± 3.0	9.3 ± 4.0	18.5 ± 3.1	13.1 ± 4.0
Orientation [°]	April 5	149.0 ± 4.0	163.5 ± 11.9	141.4 ± 18.4	147.0 ± 13.0	136.2 ± 19.9
	April 6	151.2 ± 3.2	169.5 ± 12.6	149.7 ± 15.1	152.0 ± 16.8	156.8 ± 17.5
	April 10	171.1 ± 3.4	180.4 ± 10.8	152.7 ± 18.9	163.2 ± 19.4	179.7 ± 18.4
	April 11	167.5 ± 3.1	182.2 ± 11.7	157.2 ± 25.5	169.0 ± 15.1	162.0 ± 11.3
Asymmetry	April 5	0.25 ± 0.01	0.30 ± 0.05	0.25 ± 0.08	0.23 ± 0.06	0.21 ± 0.05
	April 6	0.24 ± 0.02	0.28 ± 0.04	0.24 ± 0.09	0.26 ± 0.03	0.33 ± 0.07
	April 10	0.23 ± 0.00	0.31 ± 0.04	0.22 ± 0.10	0.21 ± 0.02	0.35 ± 0.05
	April 11	0.20 ± 0.01	0.25 ± 0.04	0.21 ± 0.11	0.22 ± 0.02	0.26 ± 0.03
F. C. B.	April 5	0.07	0.07 ± 0.04	0.02 ± 0.01	0.16 ± 0.05	0.01 ± 0.01
	April 6	0.07	0.07 ± 0.04	0.02 ± 0.01	0.25 ± 0.03	0.09 ± 0.07
	April 10	0.04	0.13 ± 0.05	0.02 ± 0.01	0.26 ± 0.03	0.21 ± 0.08
	April 11	0.04	0.07 ± 0.03	0.02 ± 0.01	0.23 ± 0.02	0.08 ± 0.04

Table 3.4: Extracted ring features of M87* images, grouped by feature and date. Each row shows the mean and standard deviation of the extracted value for one feature on one observation date. *`eht-imaging` is a hand-crafted imaging algorithm and not directly comparable to the data-driven priors.

Part II

Physics-based Priors

Chapter 4

4D BLACK-HOLE TOMOGRAPHY WITH PHYSICS CONSTRAINTS

Despite our advances in black-hole imaging, 2D images are an incomplete snapshot of the ultimate subject of interest: the dynamic 3D scene around a black hole. An image is a complicated 2D projection of the 3D emissivity around the black hole and thus may obscure important 3D structures. Moreover, a static image does not capture the dynamic nature of the radiation-emitting material, which moves, appears, and disappears over time. Visualizing the dynamic 3D emissivity field near a black hole would push the boundaries of scientific knowledge by revealing previously-unseen parts of the universe and informing new physics models.

In this chapter we consider the task of reconstructing the time-dependent 3D emissivity field given only image-plane measurements. This can be framed as a single-viewpoint 4D tomography problem, which is severely ill-posed. Unlike in traditional tomography problems, which assume multiple viewpoints and a static source, we only have access to the imaging plane of our viewpoint from Earth, and our source evolves over time. Furthermore, the velocity field that dictates how matter moves around the black hole is not perfectly known and becomes more variable as it gets closer to the black hole. Another distinction from traditional tomography is that light does not follow straight paths due to how black holes warp spacetime. The forward model in our setting includes non-linear ray-tracing that accounts for gravitational lensing, and it partially depends on the unknown velocity field.

Previously, Levis et al. [202] proposed an approach based on hard physics constraints to reconstruct dynamic 3D emission given EHT measurements. The approach, known as BH-NeRF, reconstructs an initial 3D emissivity field and propagates it according to a Keplerian velocity model. It does not account for new emission over the time of imaging, and it does not allow for any deviations from Keplerian dynamics. The method was used to perform the first 3D recovery of the gas around a black hole [203], specifically a flare near Sgr A*. However, it was only effective in that setting because the flare was sufficiently far away from the black hole so that its dynamics could be described as Keplerian, and it was suspected that no new flares appeared during the observation time used for imaging. In contrast, we target emission much closer to the black hole, where hot spots of emission appear and disappear much more rapidly, and the velocity fluctuates much more from the Keplerian model.

We propose an approach based on soft physics constraints that simultaneously reconstructs the dynamic 3D emissivity field and the 3D velocity field. We represent both fields as coordinate-based neural networks and optimize them according to a

physics-informed loss. Specifically, a dynamics loss term ensures the dynamics of the estimated emissivity are consistent with the estimated velocity, a velocity term regularizes the estimated velocity according to a predefined (e.g., Keplerian) velocity model, and the data-fit term incorporates gravitational lensing and velocity-dependent redshift effects in the forward model. By reconstructing a 4D (i.e., time-dependent 3D) field, we are able to account for new emissions. By enforcing physics as a soft constraint, we are robust to differences between the assumed velocity field and the real-world velocity field, which may include turbulence, radial infall and outfall, and sub-Keplerian speeds.

In Section 4.1, we discuss related work in tomographic imaging and computational methods. Section 4.2 covers physics background needed to understand this chapter. We then present our approach in Section 4.3 and results in Section 4.4. In Section 4.5 we discuss future directions for this work.

4.1 Related Work

4.1.1 Tomography

Tomography is a type of inverse problem that aims to recover an object or structure from its lower-dimensional projections, typically obtained as line integrals. A common tomographic problem is reconstructing a 3D image from 2D projections at multiple viewpoints, as in computed tomography [178] for medical imaging and 3D scene reconstruction from photos [229]. These settings, however, benefit from linear ray-tracing, multiple viewpoints, and the assumption of a static source. In our case, we work with curved light paths, a single viewpoint, and a dynamic source. Essentially, we aim to solve a 4D tomography problem by recovering a time-dependent 3D image.

Previous efforts in many scientific domains have dealt with problems related to non-linear light paths. For example, imaging underwater environments [336], imaging through the Earth's atmosphere [201], and imaging cosmological objects [158] are difficult due to refraction, scattering, and gravitational lensing, respectively. We deal with curved light paths due to strong gravitational lensing, which is the effect of curved spacetime around a black hole. In the context of astronomical imaging, deprojecting galaxies [262, 354, 353] from 2D images is a common problem, but the galaxies are usually parameterized as static, simple shapes. Previous work recovered hot spots of emission around a black hole using EHT data [303, 203], but they assumed a simplified geometric model [303] or a simplified dynamics model [202] for the hot spots. In contrast, we use a time-dependent neural network to capture complicated features and dynamics.

4.1.2 Coordinate-based neural fields

Instead of representing a volume with a discretized representation such as a mesh or voxel grid, we can represent it with a coordinate-based neural network. Coordinate-based neural fields are parameter-efficient, continuous representations that use a multilayer perceptron (MLP) to map from the coordinates of a point in space to the value of the field at that point. They also benefit from the implicit regularization of a neural network that leads to smoothness in the output field. They are commonly used to represent solutions to inverse problems, where the weights of the MLP are optimized to minimize a data-fit loss. For example, a neural radiance field (NeRF) [229] is trained to represent a 3D scene that agrees with a sparse set of 2D views, thus solving a 2D-to-3D tomography problem in which linear ray-tracing is done to form 2D images. Similar approaches have been developed for scientific imaging tasks,

such as cryo-electron microscopy [356], MRI [298], and computed tomography [292].

Coordinate-based neural fields have also been applied to dynamic scenes. A straightforward approach is to add a time dimension to the input coordinates [130, 335, 207, 129]. Another approach is to use one neural representation for the scene at the first time frame and a second neural representation for the time-dependent deformation [253, 305, 247, 246]. Our work combines both approaches by estimating both a time-dependent emissivity field and a velocity field describing its dynamics. When measurements are sparse, the main challenge is finding a temporal prior to describe how points in space change across time. Previous methods regularized the deformation to be sparse or rigid [247, 305], while others imposed models highly specialized to the object being reconstructed [202, 207, 249, 126]. We propose a way to impose a soft constraint on the velocity field based on known physics.

We note that Gaussian splatting [182, 217] has emerged as an alternative efficient and continuous representation. This approach represents a 3D density distribution as a set of Gaussians with learnable position, scale, and orientation. Its main benefit over coordinate-based neural networks is that rendering can be made extremely efficient. However, the number of Gaussians needs to pre-defined, so it is unsuitable for our setting, in which hot spots of emission may appear and disappear over time.

4.1.3 BH-NeRF

Our work is inspired by BH-NeRF [202], which was used to obtain the first 3D reconstruction of the gas around a black hole [203]. BH-NeRF represents the initial 3D emissivity with a coordinate-based neural field, denoted as $e_0(\mathbf{x})$. It then propagates the initial condition according to Keplerian orbital dynamics. The weights of e_0 are optimized to minimize a data-fit loss with respect to time-dependent EHT measurements. For a given time t, it propagates e_0 according to Keplerian orbital velocities to that time, passes it through the EHT forward model, and compares the estimated measurements with the observed measurements at time t.

BH-NeRF encounters two major limitations: (1) it cannot account for velocity that is different from the assumed velocity model, and (2) it cannot account for emission that appears and disappears over time. For (1), the Keplerian model assumes that particles stay on circular orbits around the black hole. It does not account for infall of material into the black hole, and it does not account for effects of turbulence or dragging. Velocity becomes less and less Keplerian as we get closer to the black

hole, where high-energy effects and geodesic effects causing gas to fall into the black hole become more prevalent. Furthermore, imposing the assumed velocity model as a strong constraint makes it difficult to overcome modeling errors. For (2), because BH-NeRF only estimates the initial 3D emissivity, it cannot account for new emission appearing in the volume over time. Our work addresses these limitations, recovering a time-dependent emissivity field and a velocity field. We impose velocity as a soft constraint, meaning our method is robust to errors in the assumed velocity model. Since our emissivity field is time-dependent, it can capture new emission.

4.2 Black-hole Emission Physics

Emission near a black hole can come from either inflowing plasma in an accretion disk or outflowing plasma from a jet [2]. The plasma radiates, emitting photons that travel through space. Compact bright emission spots, known as "flares" or "hot spots," arise when magnetic reconnection or overdensities of mass or magnetic field strength create bursts of radiation [52].

How fluid moves and how its emission appears to an observer both depend on the complicated physics near a black hole. A mathematical object known as a *spacetime metric* captures the geometry of the warped spacetime around a black hole. It is represented as a *metric tensor* $g_{\mu\nu}$, which allows us to define distances between points in spacetime, which are denoted as x^μ.

Preliminaries In general relativity, we work with *contravariant* (index-up) vectors a^μ and *covariant* (index-down) vectors a_μ. To lower indices, we multiply a contravariant vector by the spacetime metric:

$$a_\mu = g_{\mu\nu} a^\nu. \tag{4.1}$$

Conversely, to raise indices, we multiply a covariant vector by the inverse metric:

$$b^\mu = g^{\mu\nu} b_\nu. \tag{4.2}$$

Here $\mu, \nu \in (0, 1, 2, 3)$ are indices. We can compute dot products between pairs of covariant and contravariant vectors:

$$a \cdot b = a^\mu b_\mu = a_\mu b^\mu = \sum_{i=0}^{3} a_i b^i. \tag{4.3}$$

4.2.1 Fluid velocity

We denote the spherical velocity vector as v^i:

$$v^i = \left(v^r, v^\theta, v^\phi\right) = \left(\frac{\mathrm{d}r}{\mathrm{d}t}, \frac{\mathrm{d}\theta}{\mathrm{d}t}, \frac{\mathrm{d}\phi}{\mathrm{d}t}\right), \tag{4.4}$$

where (r, θ, ϕ) are spherical coordinates, with r denoting the radius, θ denoting the polar angle, and ϕ denoting the azimuthal angle. In general relativity, we often work with the four-velocity vector u^μ, which represents velocity relative to the proper time τ:

$$u^\mu = \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}, \frac{\mathrm{d}r}{\mathrm{d}\tau}, \frac{\mathrm{d}\theta}{\mathrm{d}\tau}, \frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right) = u^t \left(1, v^r, v^\theta, v^\phi\right), \tag{4.5}$$

since $u^t = \mathrm{d}t/\mathrm{d}\tau$ by definition. Note that t and τ are two different parameters relating to a notion of time. The parameter t is the coordinate time, or the time that the observer at radius $r \to \infty$ measures. The parameter τ is the proper time, which is what the particle moving with velocity u^μ measures. In our work, we estimate the coordinate three-velocity v^i. The four-velocity u^μ can be directly derived from v^i via Equation 4.5, with

$$u^t = \sqrt{\frac{-1}{g_{tt} + 2g_{ti}v^i + g_{ij}v^i v^j}} \tag{4.6}$$

to satisfy the normalization condition $u^\mu u_\mu = -1$. Note that for an arbitrary choice of components in v^i, the square root in Equation 4.6 may be undefined. In Section 4.2.1.2, we introduce another frame of reference that allows for arbitrary three-velocities and is thus more numerically stable.

4.2.1.1 Velocity model

We would like a velocity model that accurately represents the dynamics near a black hole. The simplest model is Keplerian orbital dynamics, which is only reasonable outside of the innermost stable circular orbit (ISCO). Inside the ISCO, matter plunges inward along geodesics, and even outside of the ISCO, motion is not exactly Keplerian. For a more general velocity model, we turn to the velocity model proposed by Cárdenas-Avendaño, Lupsasca, and Zhu [60] for adaptive analytical ray-tracing (AART), which accounts for sub-Keplerian velocity and infall.

The AART velocity model mixes two velocities: (1) a sub-Keplerian velocity denoted by u^μ_{subkep} and (2) an infall velocity denoted by u^μ_{infall}. The velocity u^μ_{subkep} is based on the Cunningham [87] model of Keplerian dynamics, which includes infall inside the ISCO. The velocity u^μ_{infall} only represents infall due to geodesics, assuming a particle that starts with zero velocity at radius infinity. The AART velocity depends on a fixed spin a and mass M, and it is axially symmetric, meaning in the equatorial plane it only depends on the radius r. The exact formulae for u^μ_{subkep} and u^μ_{infall} can be found in Appendix F of [63]. In the rest of this section, we will refer

to the following common abbreviations in the Kerr metric:

$$\Delta = r^2 + a^2 - 2Mr, \tag{4.7}$$

$$\Sigma = r^2 + a^2 \cos^2\theta, \tag{4.8}$$

$$\Xi = \left(r^2 + a^2\right)^2 - a^2\Delta \sin^2\theta, \tag{4.9}$$

$$\Omega = \frac{2Mar}{(r^2 + a^2)^2 - a^2\Delta \sin^2\theta}. \tag{4.10}$$

We define $\Omega_{\text{subkep}} = \frac{u^{\phi}_{\text{subkep}}(r)}{u^{t}_{\text{subkep}}(r)}$ and $\Omega_{\text{infall}} = \frac{u^{\phi}_{\text{infall}}(r)}{u^{t}_{\text{infall}}(r)}$. The four-velocity components of the general velocity model are given by

$$u^t(r) = \frac{1 + \frac{r^2(u^r)^2}{\Delta_r}}{1 - (r^2 + a^2)\Omega^2 - \frac{2}{r}(1 - a\Omega)^2}, \tag{4.11}$$

$$u^r(r) = \beta_r u^r_{\text{subkep}} + (1 - \beta_r) u^r_{\text{infall}}, \tag{4.12}$$

$$u^\theta(r) = 0, \tag{4.13}$$

$$u^\phi(r) = u^t(r) \cdot \left(\beta_\phi \Omega_{\text{subkep}} + (1 - \beta_\phi)\Omega_{\text{infall}}\right). \tag{4.14}$$

The parameters $\beta_\phi, \beta_r \in [0, 1]$ represent the relative magnitude of the sub-Keplerian velocity in the angular and radial velocities, respectively. Notice that $\beta_\phi = \beta_r = 1$ corresponds to a purely sub-Keplerian model, whereas $\beta_\phi = \beta_r = 0$ corresponds to pure infall. The sub-Keplerianity parameter $\xi = (0, 1]$ represents the ratio of the fluid angular momentum to the Keplerian value. Therefore, for a fixed spin, mass, and direction of orbit (i.e., prograde or retrograde), this general velocity model depends only on the three parameters $(\xi, \beta_r, \beta_\phi)$.

4.2.1.2 Normal observer frame

In our work, we represent the velocity field as a coordinate-based neural network, which may estimate unreasonable velocities that would cause the equation for u^t (Equation 4.6) to be undefined. To avoid numerical errors, we instead estimate velocities in the *normal observer* frame. In the normal observer frame, we denote the three-velocity as $\tilde{u}^i$. The conversion from $\tilde{u}^i$ to u^μ is given by

$$u^t = \frac{\gamma}{\alpha}, \tag{4.15}$$

$$u^i = \tilde{u}^i - \left(\frac{\gamma}{\alpha}\right)\beta^i, \tag{4.16}$$

where i is an index into (r, θ, ϕ). The lapse α and shift vector β^i are defined as

$$\alpha = \sqrt{\frac{1}{-g^{tt}}}, \tag{4.17}$$

$$\beta^i = -\frac{g^{ti}}{g^{tt}}. \tag{4.18}$$

The Lorentz factor $\gamma \geq 1$ can be computed from the normal-observer $\tilde{u}^i$ as

$$\gamma = \sqrt{1 + g_{ij}\tilde{u}^i\tilde{u}^j}. \tag{4.19}$$

The condition for the four-velocity to be physical is $\gamma > 1$. In Boyer-Lindquist coordinates in the Kerr spacetime, we have that

$$\gamma^2 = 1 + \frac{\Sigma}{\Delta}(u^r)^2 + \Sigma\left(u^\theta\right)^2 + \frac{\Xi \sin^2\theta}{\Sigma}\left(u^\phi\right)^2. \tag{4.20}$$

By inspection, we can see that $\gamma^2 > 1$ for all u^i. There may be numerical instabilities when $\Delta \to 0$, which does happen at the horizon.

The conversion can be done in the other direction by solving for $\tilde{u}^i$ given u^μ in Equation 4.16. Recall that u^μ can be determined from v^i by computing u^t according to Equation 4.6.

4.2.2 Image formation

The light captured on the image plane is the result of a complicated process that depends on two main ingredients: the emissivity field and the velocity field. The emissivity determines how much intensity is integrated along light paths, and the velocity influences how much the intensity gets boosted along the way.

Emissivity The emissivity e is essentially the light intensity in 3D space. More specifically, it quantifies the total electromagnetic energy per unit volume, per unit time, per unit frequency, and per unit solid angle emitted in a region of space. It depends on many characteristics, including the particle density, temperature, and magnetic field strength of the plasma. We drop the dependence on the frequency and direction and denote the emissivity field as a function of time and space: $e(t, \mathbf{x})$, where $\mathbf{x} = (x, y, z)$.

Redshift Due to the Doppler effect, the frequency of light observed on the image plane is different from that of the original emission. The velocity field dictates how

the intensity of the light is boosted (in the case of redshift) or dimmed (in the case of blueshift) as it travels to the observer. In other words, a source moving towards us near the speed of light appears brighter, whereas a source moving away appears dimmer. Assuming a photon energy of $E = 1$, the location-dependent redshift factor g is computed as

$$g = \frac{E}{-k_\mu u^\mu} = \frac{1}{-k_\mu u^\mu}, \tag{4.21}$$

where k_μ is the photon momentum vector. It is related to the derivative of the photon position $x^\mu(\tau)$ with respect to the Mino time τ:

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \frac{\Sigma}{E} k^\mu = \Sigma k^\mu = \Sigma g^{\mu\nu} k_\nu \tag{4.22}$$

assuming $E = 1$. For a given spin and mass of the black hole, the photon momentum at every point in spacetime is fixed.

4.2.2.1 Radiative transfer and geodesic ray-tracing

Radiative transfer describes how light propagates through a medium, considering factors such as emission, absorption, and scattering. Two additional complicating factors arise when working with curved spacetime: (1) curved light paths and (2) redshift. To make an image, we essentially have to integrate a general relativistic radiative transfer (GRRT) equation along the curved light paths, known as *geodesics*.

We make two simplifying assumptions in the GRRT equation that have also been made in previous work [202]. First, we assume that the attenuation of light due to absorption and scattering is negligible, as is the case for EHT images [8]. Second, we do not account for emission changing at relativistic speeds as the light ray propagates at a finite speed [66, 96, 230]. With these assumptions, we can simply integrate the emissivity along the ray paths that end at the image plane. Supposing a discretized image plane with $N \times N$ pixels, let $\Gamma_n = (t(s), \mathbf{x}(s))$ denote the geodesic that ends at the n-th pixel, where s is the distance along the geodesic. The observed intensity at the n-th pixel at time t can be computed as

$$p_n(t) = \int_{\Gamma_n} g^2(\mathbf{x}) e(t, \mathbf{x}) \mathrm{d}s. \tag{4.23}$$

4.3 Method

4.3.1 Representing the emissivity and velocity fields

We represent both the emissivity and velocity fields with coordinate-based neural networks. We denote the estimated emissivity and velocity by $e(t, \mathbf{x})$ and $\tilde{u}^i(\mathbf{x})$, respectively. In particular,

$$e(t, \mathbf{x}) = e(t, \mathbf{x}; \boldsymbol{\theta}_e) = \mathrm{MLP}\left(\gamma\left([t, \mathbf{x}]^\top\right); \boldsymbol{\theta}_e\right), \tag{4.24}$$

$$\tilde{u}^i(\mathbf{x}) = \tilde{u}^i(\mathbf{x}; \boldsymbol{\theta}_v) = \mathrm{MLP}\left(\gamma(\mathbf{x}); \boldsymbol{\theta}_v\right), \tag{4.25}$$

where $\boldsymbol{\theta}_e$ and $\boldsymbol{\theta}_v$ denote the parameters of the neural networks. The positional encoding γ has been shown to improve the representation of high-dimensional features [298]. It is defined as

$$\gamma(\mathbf{x}) = \left[\sin(\mathbf{x}), \cos(\mathbf{x}), \ldots, \sin\left(2^{L-1}\mathbf{x}\right), \cos\left(2^{L-1}\mathbf{x}\right)\right]^\top, \tag{4.26}$$

where the degree L determines the bandwidth of the interpolation kernel [172] underlying the MLP (i.e., higher L allows for higher-frequency representations). Note that $\tilde{u}^i$ is the three-vector in the normal observer frame defined in Section 4.2.1.2. Figure 4.1 illustrates the optimization procedure for both networks, which we detail in the following subsection.

4.3.2 Optimization

The loss function combines three terms: a data-fit loss, a dynamics loss, and velocity regularization. We write it as follows:

$$\mathcal{L}\left(e, \tilde{u}^i\right) = \lambda_{\mathrm{data}}\mathcal{L}_{\mathrm{data}}\left(e, \tilde{u}^i\right) + \lambda_{\mathrm{dyn}}\mathcal{L}_{\mathrm{dyn}}\left(e, \tilde{u}^i\right) + \lambda_{\mathrm{reg}}\mathcal{L}_{\mathrm{reg}}\left(\tilde{u}^i\right). \tag{4.27}$$

The **data-fit loss** encourages the emissivity network to fit the observed image-plane video. Note that the forward model depends on the estimated velocity through the redshift factor (Equation 4.21), so this loss term depends on both networks.

$$\mathcal{L}_{\mathrm{data}}\left(e, \tilde{u}^i\right) = \sum_{i=1}^{N_t} \left\|\mathbf{I}\left(e(t_i); \tilde{u}^i\right) - \mathbf{I}_{\mathrm{obs}}(t_i)\right\|_2^2, \tag{4.28}$$

where N_t is the number of time frames, $\mathbf{I}_{\mathrm{obs}}$ is the observed image-plane video, and $\mathbf{I}$ is the forward model that ray-traces the 3D emissivity onto the 2D image plane. More specifically, for an $N \times N$ image grid,

$$\mathbf{I}\left(e(t_i); \tilde{u}^i\right) = [p_1(e(t_i); \tilde{u}^i), \ldots, p_{N^2}(e(t_i); \tilde{u}^i)], \tag{4.29}$$

$$p_n\left(e(t_i); \tilde{u}^i\right) = \int_{\Gamma_n} g^2(\mathbf{x}; \tilde{u}^i) e(t, \mathbf{x}) \mathrm{d}s, \tag{4.30}$$

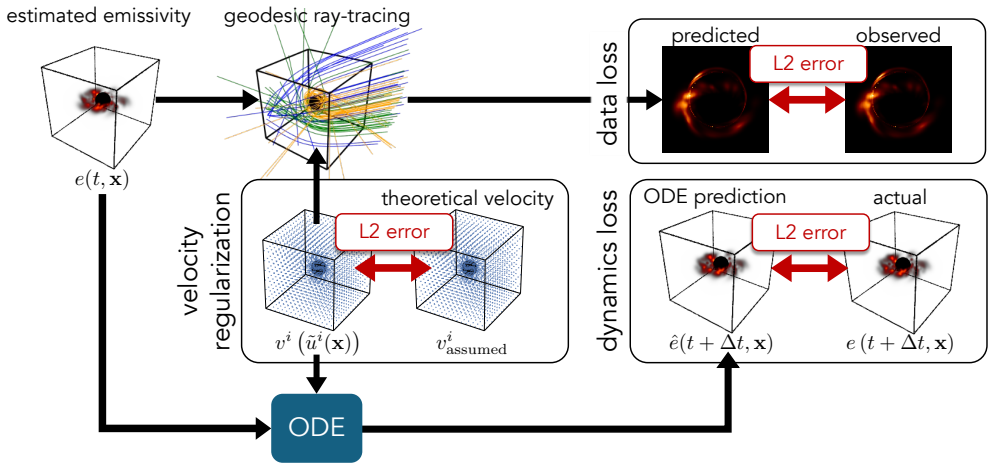


Figure 4.1: Method overview. For every time t, we optimize the emissivity network $e(t, \mathbf{x})$ and velocity network $\tilde{u}^i(\mathbf{x})$ according to a data-fit loss, dynamics loss, and velocity regularization. The data-fit loss makes sure that the estimated emission video agrees with the observed emission video. It depends on the velocity network through the redshift boosting factor. The dynamics loss connects both networks by ensuring that the dynamics within the emissivity network agrees with the estimated velocity model. Velocity regularization supervises the velocity network with an assumed theoretical velocity model. All the loss terms are fully differentiable with respect to the network weights.

where $g\left(\cdot\,; \tilde{u}^i\right)$ is the redshift factor computed via Equation 4.21 with the u^μ that is derived from $\tilde{u}^i$, and $\Gamma_n = ((t(s), \mathbf{x}(s))$ is the curved light path ending at the n-th pixel.

The **dynamics loss** connects the emissivity and velocity networks by imposing a soft velocity constraint on the emissivity dynamics. We define it as

$$\mathcal{L}_{\text{dyn}}\left(e, \tilde{u}^i\right) = \mathbb{E}_{t\sim\mathcal{U}([0,T])} \left\| e(t+\Delta t) - \hat{e}(t+\Delta t) \right\|_2^2, \tag{4.31}$$

where Δt is a small time interval, and T is the total observation time. The propagated emissivity as predicted by the velocity network is denoted by $\hat{e}(t+\Delta t)$, and it is given by

$$\hat{e}(t+\Delta t) = e(t, \mathbf{x} - \hat{\mathbf{x}}(\Delta t)), \tag{4.32}$$

where

$$\hat{\mathbf{x}} = \int_0^{\Delta t} v^i\left(\tilde{u}^i(\mathbf{x}(t)\right) \mathrm{d}t. \tag{4.33}$$

In words, $\hat{e}(t+\Delta t)$ propagates the predicted emissivity at time t forward in time by Δt, using the prediction of the velocity network. Equation 4.31 compares this

to the actual emissivity field at time $t + \Delta t$, encouraging the time evolution in the emissivity network to agree with the velocity network.

We incorporate an assumed theoretical velocity model via **velocity regularization**:

$$\mathcal{L}_{\text{reg}}(\tilde{u}^i) = \left\| v^i\left(\tilde{u}^i\right) - v^i_{\text{assumed}} \right\|_2^2, \tag{4.34}$$

where we compute v^i_{assumed} from the AART velocity model detailed in Section 4.2.1.1, and $v^i\left(\tilde{u}^i\right)$ is the three-velocity derived from the output of the velocity network in the normal observer frame.

4.3.3 Simulating data

We parameterize hot spots in the same way as Levis et al. [202], where a hot spot is either a Gaussian blob with a certain position and scale or a tube with a certain position, length, and scale. We inject hot spots at different times throughout the simulation, which lasts for a total time T. We consider two types of simulations, which we refer to as "simple" and "GRMHD-style."

The simple approach is to define a total number of hot spots N_{hotspot} and introduce the hot spots at evenly spaced time intervals. For each hot spot, we randomly sample a radius $r \sim \mathcal{U}([r_{\min}, r_{\max}])$ and scale $\sigma \sim \mathcal{U}([\sigma_{\min}, \sigma_{\max}])$. The hot spot has probability $1/2$ of being either a Gaussian blob or a tube. If it is a Gaussian blob, then we randomly sample an azimuthal position $\phi \sim \mathcal{U}([0, 2\pi])$. If it is a tube, then we randomly sample an arc length $\ell \sim \mathcal{U}([0, \pi/5])$ and starting azimuthal position $\phi_0 \sim \mathcal{U}([0, 2\pi - \ell])$.

We developed a "GRMHD-style" approach such that the video would look similar to a ray-traced GRMHD simulation [329]. At any given time, there is a certain probability of a hot spot appearing in any location in the equatorial plane. The probability of a hot spot appearing at a position (r, ϕ) is equal to $p_r p_\phi$, where

$$p_r = \begin{cases} 0, & r < r_{\min} \\ 1/r^2 & r \geq r_{\min} \end{cases},$$

and $p_\phi = 0.03$. Note that our model reflects the phenomenon that more hot-spot activity occurs closer to the black hole. We discretize the time domain into N_t time frames, and at each frame, we flip a biased coin at each location in a discretized grid on the equatorial plane to determine whether a hot spot appears there. Each hot spot is a Gaussian blob with scale $\sigma = \text{clip}\left(\frac{1}{1+e^{-(r-6)}}, \sigma_{\min}, \sigma_{\max}\right)$.

Both the simple and GRMHD-style simulation approaches output a random set of hot spots with certain injection times. We solve an ODE involving the injected hot spots and the AART velocity (Section 4.2.1.1) to determine the emissivity field at every time t. To form the video, we perform geodesic ray-tracing at every time frame according to Equation 4.23 with the AART redshift factor.

4.4 Results

In this section we present results validating our method on simulated emission data.

4.4.1 Implementation

In our experiments, we used MLPs with 4 layers with ReLU activations. Each layer was 256 units wide in the emissivity network and 128 units wide in the velocity network. We set the position-encoding degree as $L = 3$. We used the Adam optimizer [185] with a polynomial learning-rate schedule going from 10^{-4} to 10^{-6} over 200K iterations and a batch size of 6. We set $\lambda_{\text{data}} = 1$, $\lambda_{\text{dyn}} = 0.1$, and $\Delta t = 0.01$. Rather than use a fixed regularization weight, we started with a regularization weight of 1 and exponentially decayed it to $\lambda_{\text{reg}} = 10^{-6}$ during optimization using the following weight-annealing function:

$$\lambda(i) = \lambda_{\text{final}} + (\lambda_{\text{init}} - \lambda_{\text{final}}) \cdot \exp(-d \cdot i), \tag{4.35}$$

where i is the iteration, and we set the decay rate to $d = 0.001$.

For the experiments presented in this section, we assumed a spin of $a = 0.2$, inclination angle of $\theta_{\text{o}} = 60°$, and FOV of 20 M. We randomly sampled hot spots with $r_{\text{min}} = 7.5$ and $r_{\text{max}} = 8$ and $\sigma_{\text{min}} = 0.25$ and $\sigma_{\text{max}} = 0.75$. We propagated them according to the AART velocity model with $\xi = 0.7$ and $\beta_r = \beta_\phi = 0.9$. On the measurement side, we discretized the image plane as a 128×128 pixel grid and took 64 equally-spaced frames across 1 hour of observation. We took the result after 180K optimization iterations. PSNR was calculated per-frame after normalizing both the estimated and true emissivities to the range $[0, 1]$.

Parameterization of velocity network The true velocity model is axially symmetric, and we can enforce this constraint on the velocity network by making it only depend on the radius r. All the results in this section are based on an axially-symmetric velocity network, except in Figure 4.3, in which we compare velocity networks that depend on r, (r, θ), and (x, y, z).

4.4.2 Results with velocity mismatch

We performed an experiment in which we assumed a velocity with $\xi = 1$ and $\beta_r = \beta_\phi = 1$ for the velocity regularization. Figure 4.2 shows emissivity reconstructions for the simple hot-spot simulation with a true velocity where $\xi = 0.7$ and $\beta_r = \beta_\phi = 0.9$. When we do not include any velocity term in the optimization (i.e., $\lambda_{\text{dyn}} = \lambda_{\text{reg}} = 0$ and the only objective is data-fitting), the reconstruction includes

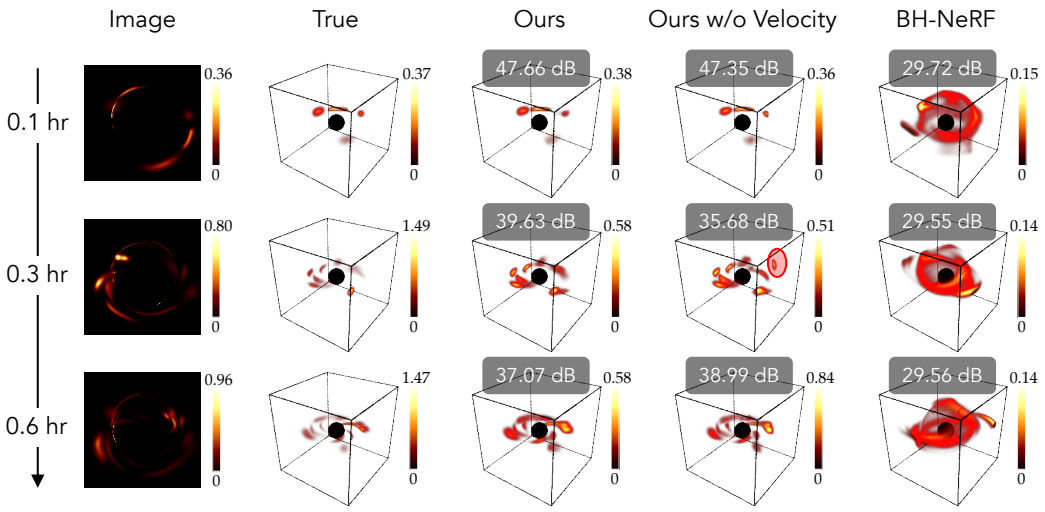


Figure 4.2: Emissivity reconstructions with a mismatch in the assumed velocity model. When not incorporating any velocity constraint, the emissivity network introduces spurious hot spots (the red oval highlights an example) to fit the image-plane measurements. BH-NeRF cannot overcome the velocity mismatch because it enforces a Keplerian model as a hard constraint.

some spurious regions of emissivity that appear and disappear quickly. This is because when the hot spots do not need to propagate according to a velocity model, the emissivity network can easily introduce spurious hot spots to fit the data. We note that the method without velocity supervision may suffer more with sparser measurements.

We are also able to recover the true velocity even with an incorrect assumed velocity model. Figure 4.3 compares the recovered radial velocity v^r and azimuthal velocity v^ϕ in the equatorial plane ($\theta = \pi/2$) to the true velocities. It accurately captures the increasingly negative radial velocity as the radius r gets smaller, and it accurately captures the dropoff in azimuthal velocity as r gets bigger. In contrast, BH-NeRF assumes a fixed Keplerian velocity model that has zero radial velocity everywhere and an inaccurate dropoff in azimuthal velocity. We note that our recovered velocity is less accurate in regions of low emission signal, such as where there is little emissivity to begin with at large radius and where there is less signal that gets propagated to the image plane at small radius.

Figure 4.3 also compares reconstructed velocities assuming different degrees of freedom in the velocity network. As expected, the velocity network that only depends on r has the most accurate velocity recovery, followed by the one that depends on (r, θ) and then the one that depends on (x, y, z). Table 4.1 verifies that

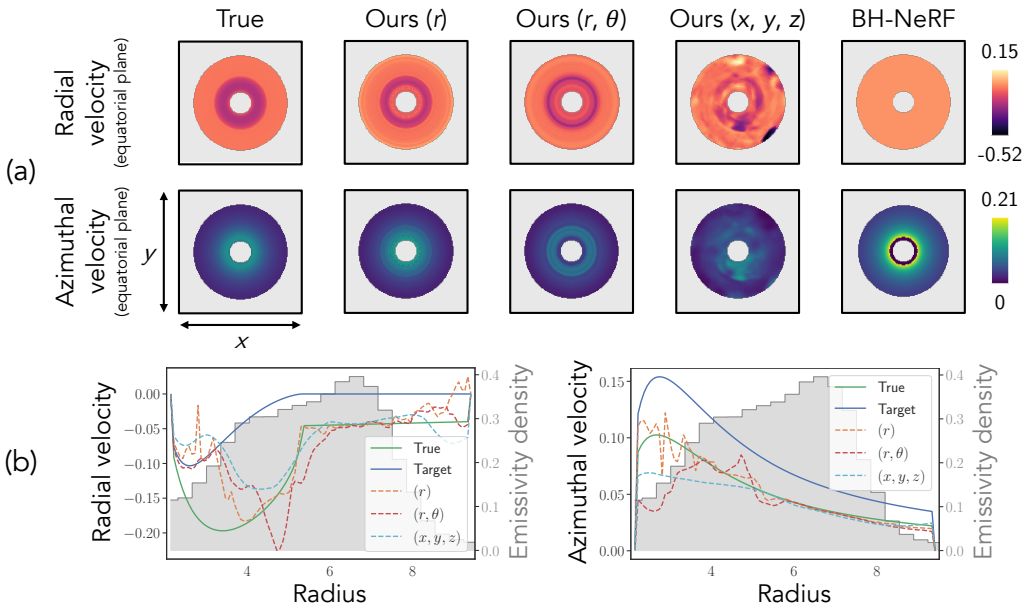


Figure 4.3: Velocity reconstructions with a mismatch in the assumed velocity model. Here we compare three types of velocity networks, parameterized by r, (r, θ), and (x, y, z), respectively. (a) shows the radial velocity v^r and azimuthal velocity v^ϕ in the equatorial plane ($\theta = \pi/2$). The BH-NeRF velocity is fixed and assumes $v^r = 0$ everywhere and overestimates v^ϕ near the black hole. (b) plots v^r and v^ϕ (averaged azimuthally) versus r. The blue curve shows the assumed velocity, and the green curve shows the true velocity. Despite being regularized according to the incorrect assumed velocity, our method is able to recover a velocity closer to the true one.

the RMSE of the estimated velocity in the equatorial plane increases as the number of degrees of freedom increases from 1 to 3 in the velocity network. In terms of PSNR, the estimated emissivity does not change much between 1 and 2 DOFs, but having 3 DOFs in the velocity network results in the least-accurate emissivity reconstruction. It makes sense that enforcing a correct constraint on the velocity network leads to better reconstruction overall.

GRMHD-style emissivity Ultimately, we would like for the method to work on real data. We tested our method on a simulation based on our GRMHD-style hot-spot model with the same velocity mismatch as before. The results are shown in Figure 4.4. The PSNR of the reconstruction is overall lower than that of the simple hot-spot reconstruction. This may because there are more and smaller hot spots in this simulation that are difficult to represent with a coordinate-based neural network. Still, our preliminary results serve as a proof-of-concept for applying our method to real GRMHD ray-traced images and potentially real EHT data.

Velocity MLP coordinates	e PSNR (↑)	v^r $(\theta = \pi/2)$ RMSE (↓)	v^ϕ $(\theta = \pi/2)$ RMSE (↓)
(r)	41.58	0.042	0.007
(r, θ)	41.67	0.046	0.022
(x, y, z)	40.60	0.052	0.015

Table 4.1: Comparing 1, 2, and 3 DOFs in the velocity network. The PSNR of the full 4D emissivity field does not change much between the (r) and (r, θ) parameterizations, but it decreases when the velocity network depends on three DOFs. The RMSE of the v^r and v^ϕ curves (which are plotted in Figure 4.3(b)) increases with the number of DOFs in the velocity network.

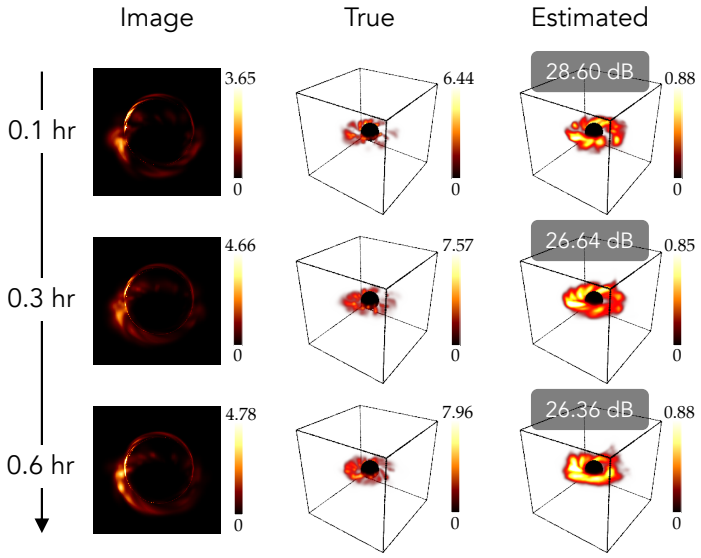


Figure 4.4: Reconstruction of GRMHD-like emissivity field.

4.5 Discussion

In this chapter we proposed and validated a physics-informed approach to the highly ill-posed problem of 4D black-hole emission tomography. Ultimately, the goal is to use this type of method to do advance our scientific understanding of the environments near real black holes. Future work involves extending this approach to real EHT VLBI measurements, which brings about a couple more challenges: (1) EHT measurements depend non-linearly on image-plane videos, making the optimization landscape more complex, and (2) EHT measurements are sparse in both the spatial-frequency and time domains. These challenges mean that the measurements contain less information, so we might need to place more weight on the velocity regularization to inform a reasonable solution.

We have already seen how our approach can be used to estimate the true velocity field around a black hole, but it might be possible to estimate other physical characteristics of the black hole such as its spin. Our optimization objective is fully differentiable with respect to the spin associated with the assumed velocity model, so we could treat the spin as an unknown parameter that we wish to optimize in conjunction with the emissivity and velocity fields. Another approach would be to take the estimated velocity field and find the spin that best explains it, assuming that the estimated velocity field does not overfit the assumed one.

A long-term goal is to use our method to inform the design of future EHT arrays. One way to do this is to treat the locations of a fixed set of EHT stations as parameters and optimize them to achieve the best 4D reconstruction. In summary, our physics-informed approach offers a lot of potential for analyzing the physical phenomena near black holes, from imaging them in the first place to discovering characteristics of the black holes themselves.

Chapter 5

VISUAL VIBRATION TOMOGRAPHY

We now turn our attention from astronomy to materials science, for which physics-based imaging methods also prove useful.

The subtle motions of objects around us are clues to their physical properties. Among such properties are stiffness and density, which dictate how an object will respond to environmental forces. As humans, we can vaguely characterize how stiff or heavy a material is, such as when we infer that a rubber basketball will bounce higher than a ceramic bowling ball by tapping on its surface. Most engineering applications, however, require a greater level of detail, such as when an aeronautical engineer must faithfully simulate how an airplane wing will react to wind turbulence. In computer vision and graphics, a full characterization of an object's material properties allows one to faithfully simulate its behavior. These scenarios require non-destructive testing (NDT) to obtain physical properties of the object without altering it.

We propose *visual vibration tomography* (VVT), a method to estimate material properties of an object directly from vibration signals extracted from monocular video. Much of NDT has focused on measuring vibrations to identify the presence of defects in structures with a known geometry. However, NDT tools are not generally used to determine the precise spatial distribution of physical properties in objects with a heterogeneous interior structure.

We show that we can measure vibrations as sub-pixel motion in 2D video and then use this motion to constrain 3D material-property estimation. Videos have several advantages over existing NDT techniques: while contact sensors and laser vibrometers take point measurements, videos offer spatially dense measurements of surface vibrations. While laser vibrometers are expensive and specialized, cameras are ubiquitous and general-purpose. While existing image-based techniques require stereo cameras for 3D motion tracking, our method shows that in many cases, a monocular view is all you need.

Our motivating insight is that, under fixed geometry, an object's material properties determine its motion. The inverse direction is also true: motion determines material properties up to a scaling factor. If the motion is small, it can be decomposed into independent modes at natural frequencies, lending itself to a concise mathematical equation linking modes and material properties. This link lays the foundation for our physics-constrained optimization approach. The key challenge of our task is to deal with incomplete and 2D (as opposed to full-field) modes. Despite these challenges, we show that we can estimate material properties from image-space motion and recover full-field modes. Our work demonstrates how a physical equation (in this

case, a fundamental equation from solid mechanics) is all we need to constrain a severely ill-posed inverse problem.

In this chapter, we first review related work in Section 5.1 and the theoretical relationship between modes and material properties in Section 5.2. In Section 5.3, we show how to extract image-space modes from video and recover material properties. In Section 5.4, we demonstrate our approach on simulated data of 3D geometries and discuss the effects of damping and model mismatch. Finally, in Section 5.5, we present proof-of-concept experiments on real data, showing that we are able to image the shape of unseen material inhomogeneities on drum heads and the presence of a defect in a real 3D Jello cube. These experiments demonstrate promise for the future of the approach in more challenging environments.

5.1 Related Work

5.1.1 Material analysis from images and video

In computer vision, scene understanding is an important goal that includes, among many tasks, characterization of materials. Previous work has estimated material categories [210, 33, 269] and surface properties [272, 157] from images. In contrast to static images, videos have been used to estimate material properties, although these are often restricted to specific object categories, such as fabrics [35, 37, 325, 228, 47] and trees [324]. Other work has inferred material properties from 3D point clouds [323, 184] and known external forces [337], but such measurements are harder to obtain than a 2D video. "Visual vibrometry" [91, 92] uses a video's motion spectrum to estimate stiffness and damping of fabrics and rods. This is a promising step towards a general approach for estimating material properties, but it is restricted to homogenized properties. In a similar vein, others have used video data to identify structural modes [70, 340, 149]. Davis et al. [90] and Davis, Chen, and Durand [89] demonstrated how to visualize image-space modes and use them for plausible simulation.

5.1.2 Traditional NDT

NDT is an umbrella term for any technique that collects data of a material or structure without damaging it. Usually, the goal is to identify defects or material inconsistencies that would change the expected behavior of the object. Laser vibrometry [104] and digital image correlation (DIC) [76] are popular tools for measuring surface displacements. Laser vibrometry has been used to examine the integrity of building structures [235, 259] and materials [108, 69]. DIC also has been used to identify defects in 2D structures [286, 308, 112, 333, 151]. Both laser vibrometry [220, 223] and DIC [145, 304] can be used for modal analysis, which involves identifying modal frequencies and shapes of a structure. While usually regarded as a verification tool rather than a means to directly infer material properties, recovered modal information has been used to solve for homogenized material properties [91, 125]. However, to our knowledge, modal analysis has not been used to solve the more challenging inverse problem of quantifying the heterogeneous properties addressed in this work.

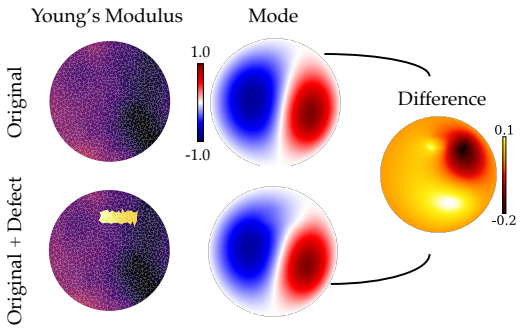


Figure 5.1: Small changes in material properties affect motion. Here a small region of a circular membrane becomes stiffer from "Original" to "Original + Defect." This change appears as a slight change in the mode shown. We propose using small changes in observed modal motion to recover the locations and shapes of defects.

5.2 Background

5.2.1 Modal analysis

Every object has resonant, or natural, frequencies. At each *resonant frequency*, the object vibrates periodically in a particular shape, called a *mode*. The vibration of a linear elastic object can be decomposed into independent modes.

In the finite element method (FEM), we model an object as a mesh, composed of elements that each take on material-property values. The mechanical properties that determine an object's vibration are Young's modulus (E), Poisson's ratio (ν), and density (ρ). The parameters E and ν define the stiffness of connections between vertices, while ρ defines the mass distribution. In this discretized model, the $n \times n$ stiffness matrix $\boldsymbol{K}$ describes the stiffness between each pair of n total DOFs, and the $n \times n$ mass matrix $\boldsymbol{M}$ describes the mass concentrated between each pair of DOFs. A mode $\boldsymbol{u}$ and frequency ω are an eigenvector-eigenvalue solution of the generalized eigenvalue problem:

$$\boldsymbol{K}\boldsymbol{u} = \omega^2 \boldsymbol{M}\boldsymbol{u}. \tag{5.1}$$

As Figure 5.1 illustrates, a small change in material properties (within a fixed geometry) results in a small change in modal motion. As most solid materials have Poisson's ratio ≈ 0.3 [251], the principal properties affecting motion are Young's modulus, which determines $\boldsymbol{K}$, and density, which determines $\boldsymbol{M}$. *Our method is based on the insight that mode shapes on the surface of an object may reveal internal spatial inhomogeneities in these properties.*

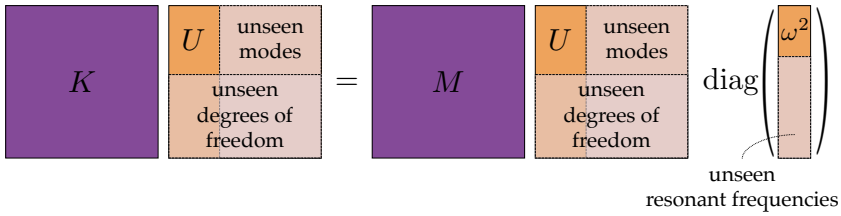


Figure 5.2: Visualizing the challenge of monocular material estimation. The generalized eigenvalue equation (Equation 5.1) defines the relationship between $\boldsymbol{K}$, $\boldsymbol{M}$ and $\boldsymbol{U}$, $\boldsymbol{\omega^2}$. The matrix $\boldsymbol{U}$ has columns corresponding to modes and rows corresponding to DOFs. The vector $\boldsymbol{\omega^2}$ contains associated eigenvalues. We would like to solve for $\boldsymbol{K}$ and $\boldsymbol{M}$ given partial information about $\boldsymbol{U}$ and ω^2.

5.2.2 Challenge of monocular material estimation

We begin by setting up a simplified version of the inverse problem. Assuming we perfectly measure all modes $\boldsymbol{u}$ and frequencies ω, then by Equation 5.1, we have the following minimization problem:

$$\boldsymbol{K}^*, \boldsymbol{M}^* = \underset{\boldsymbol{K},\boldsymbol{M}}{\arg\min} \|\boldsymbol{K}\boldsymbol{U} - \boldsymbol{M}\boldsymbol{U}\boldsymbol{\Lambda}\|_2^2, \tag{5.2}$$

where $\boldsymbol{U}$ is the matrix whose columns are modes $\boldsymbol{u}$, and $\boldsymbol{\Lambda}$ is the diagonal matrix containing eigenvalues ω^2. For a known geometry, this is a convex problem with respect to $\boldsymbol{K}$ and $\boldsymbol{M}$. However, Equation 5.2 requires that we have access to all 3D modes and frequencies. In contrast, we will be working with experimentally-observed, image-space modes, incurring the following challenges:

1. *Unseen degrees of freedom (DOFs).* We typically only observe a fraction of an object. For example, when observing a 3D cube with a monocular camera, one can see at most three of its sides, projected onto two directions of motion. Consider an $8 \times 8 \times 8$ cubic mesh, which has $(8 + 1)^3 = 729$ vertices. With three directions of motion, it has $3 \times 729 = 2187$ total DOFs. But a single monocular view of three sides of the cube can only observe 217 vertices, moving in two directions of motion, amounting to $2 \times 217 = 434$ image-space DOFs. This alone limits us to observing *fewer than 20%* of the full-field DOFs for an $8 \times 8 \times 8$ cube.

2. *Unseen modes.* Theoretically, for discrete meshes, there are as many modes as there are DOFs. However, we can only capture modes at frequencies below the Nyquist sampling rate of the camera, which is FPS$/2$.

3. *Noise.* Aside from camera noise, there is noise from motion extraction, particularly in non-textured regions.

Due to limited data, the problem of solving for $\boldsymbol{K}$ and $\boldsymbol{M}$ (Equation 5.2) is ill-posed.[1] As Figure 5.2 shows, observed data typically accounts for a tiny fraction of the matrices involved.

[1] For a known geometry and complete mode and eigenvalue information, $\boldsymbol{K}$ and $\boldsymbol{M}$ are fully determined up to a scaling factor.

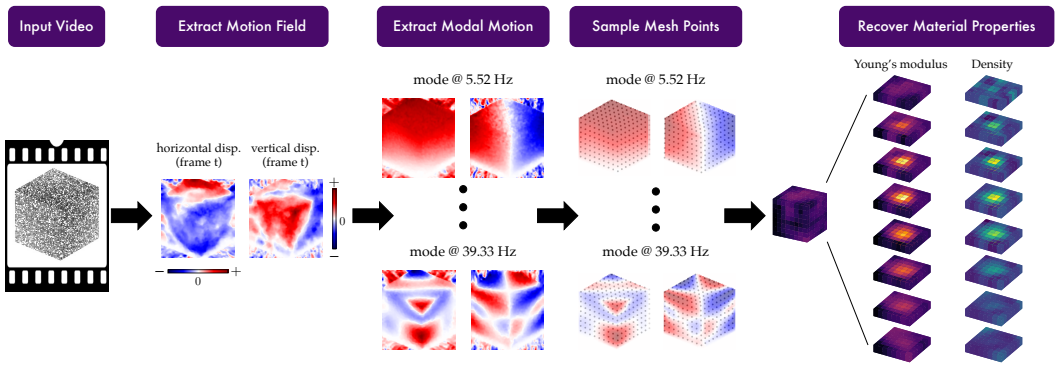


Figure 5.3: VVT method overview. Starting with a video showing vibration of an object, we extract the motion fields across time and then decompose this motion into image-space modes. From the image-space modes sampled at visible mesh points, we are able to recover a voxelized volume of the Young's modulus and density throughout the interior of the object.

5.3 Method

Our aim is to use motion features from a video to estimate material properties. This involves two stages: (1) motion extraction and image-space mode identification, and (2) solving for material properties that best match the observed image-space modes. The input is a video of an object, of which a mesh is known, and we assume that it is vibrating under linear elasticity (i.e., small motion). The output is two 3D images showing voxelized Young's modulus and density values throughout the object. Figure 5.3 shows an overview of the pipeline.

5.3.1 Extracting image-space modes from video

Motion extraction Since our approach relies on small, often imperceptible, motions, we need a way to extract sub-pixel motions from video. To quantify the displacements, we use the phase-based approach of Wadhwa et al. [321], which computes local phase shifts in a complex steerable pyramid [274, 275, 252]. This method has the advantage over other tracking methods (e.g., optical flow) of being robust to tiny motion, down to 0.001 pixel [91]. The phase shifts are converted to pixel displacements using the approach proposed by Fleet and Jepson [123]. To increase SNR, we filter out outlier pixels (i.e., top 1% of displacement magnitudes) and then apply an amplitude-weighted Gaussian blur. The result of this step is a motion field for each frame, which quantifies the horizontal and vertical displacements at each pixel relative to the first frame.

Identifying image-space modes Modes are simply periodic motions occurring at particular frequencies, so we would expect them to appear as peaks in the power spectrum of motion amplitude. As is done in previous work that extracts image-space modes [89, 90, 92], we perform a discrete Fourier transform on the motion fields to analyze them in frequency space. To make this more concrete, let $\boldsymbol{d}_x(t) = d_x(t,x,y)$ and $\boldsymbol{d}_y(t) = d_y(t,x,y)$ be the horizontal and vertical displacements, respectively, at each pixel (x,y) at frame t. The 1D FFT across time of these displacement fields results in complex-valued $\widehat{\boldsymbol{d}}_x(\ell) = \widehat{d}_x(\ell,x,y)$ and $\widehat{\boldsymbol{y}}_y(\ell) = \widehat{d}_y(\ell,x,y)$, corresponding to frequencies $f_\ell = (\text{FPS}\cdot\ell/T)$ Hz for $\ell \in [1,T]$, where T is the total number of frames. The motion power at frequency f_ℓ is then defined as $\left\|\left[\widehat{\boldsymbol{d}}_x(\ell), \widehat{\boldsymbol{d}}_y(\ell)\right]\right\|_2^2$. A peak ℓ^* in the power spectrum ideally corresponds to a natural frequency f_{ℓ^*} and image-space mode defined by the horizontal-displacement image $\boldsymbol{d}_x^* = d_x^*(x,y) = \text{Re}\left(\widehat{\boldsymbol{d}}_x\left(\ell^*,x,y\right)\right)$ and vertical-displacement image $\boldsymbol{d}_y^* = d_y^*(x,y) = \text{Re}\left(\widehat{\boldsymbol{d}}_y\left(\ell^*,x,y\right)\right)$.

Sampling image-space modes at mesh vertices To approximate the 3D-to-2D projection matrix, a user manually identifies the pixel locations of several "reference" mesh vertices, and $\boldsymbol{P}$ is the projection matrix that best maps the corresponding mesh coordinates to the image. Using $\boldsymbol{P}$, we map all of the mesh vertices from their 3D coordinates to 2D image coordinates. We then sample each image-space mode at the pixel locations of *visible* mesh vertices. For mode j, we construct a vector $\boldsymbol{\gamma}_j$ that contains the horizontal and vertical displacements of each mesh vertex at the corresponding natural frequency. Supposing we observe q' out of q mesh vertices, the vector $\boldsymbol{\gamma}_j$ has the form

$$\boldsymbol{\gamma}_j = \left[d_x^*(x_1,y_1), d_y^*(x_1,y_1), \ldots, d_x^*(x_{q'},y_{q'}), d_y^*(x_{q'},y_{q'}), 0, \ldots, 0\right]^\intercal \in \mathbb{R}^{2q}, \tag{5.3}$$

where $d_x^*(x_i,y_i)$ and $d_y^*(x_i,y_i)$ are the horizontal and vertical (pixel) displacements, respectively, of vertex i. Unseen vertices are assigned displacements of 0, and for notational clarity, we position them at the end of the vector.

5.3.2 Estimating material properties

The matrices $\boldsymbol{K}$ and $\boldsymbol{M}$ are functions of Young's modulus and density. While typically expressed as *global* matrices, they can be decomposed into *local* matrices, which scale linearly with local material properties. As a result, $\boldsymbol{K}$ and $\boldsymbol{M}$ can each be written as a weighted sum of "unit" local matrices. Specifically, we voxelize the volume containing the mesh so that each voxel contains a sub-collection of mesh

elements. Given Young's modulus w_e and density v_e for each voxel, we express the global matrices as

$$\boldsymbol{K} = \sum_{e=1}^{m} w_e \boldsymbol{K}_e \quad \text{and} \quad \boldsymbol{M} = \sum_{e=1}^{m} v_e \boldsymbol{M}_e, \tag{5.4}$$

where $\boldsymbol{K}_e$ and $\boldsymbol{M}_e$ are "unit" local stiffness and mass matrices, which we assemble using `FEniCS` [16], and m is the number of voxels. This allows us to represent $\boldsymbol{K}$ and $\boldsymbol{M}$ as functions of vectors $\boldsymbol{w}, \boldsymbol{v} \in \mathbb{R}^m$.

5.3.2.1 Optimization Formulation

Data-matching objective Suppose we have k modal observations, where $\widehat{\gamma}_i$ and $\widehat{\omega}_i$ are the i-th observed image-space mode and natural frequency, respectively. We would like to determine the voxel-wise Young's modulus values $\boldsymbol{w}$ and density values $\boldsymbol{v}$ that, when assembled into global stiffness and mass matrices, result in 3D modes $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ that agree with $\widehat{\gamma}_1, \ldots, \widehat{\gamma}_k$ when projected onto image-space. Since we do not know the full-field 3D modes, we need to include them as decision variables. Intuitively, the data-matching objective is to minimize $\|\boldsymbol{P}\boldsymbol{u}_i - \widehat{\gamma}_i\|$ for each i.

Regularization To make the solution well-defined, we choose to minimize the total squared variation (TSV) of $\boldsymbol{w}$ and $\boldsymbol{v}$, which encourages spatial smoothness. Moreover, since we are estimating both stiffness and mass, the objective function can become arbitrarily low if we do not constrain the range of material-property values; this is because scaling $\boldsymbol{K}$ and $\boldsymbol{M}$ by a factor of s still satisfies the generalized eigenvalue equation: $(s\boldsymbol{K})\boldsymbol{u} = \omega^2(s\boldsymbol{M})\boldsymbol{u}$. To resolve this ambiguity, we choose to minimize the deviation of $\boldsymbol{w}$ from a mean value $\bar{w}$. Regardless of $\bar{w}$, the relative differences in $\boldsymbol{w}^*$, $\boldsymbol{v}^*$ will not change, and for defect characterization, we generally only care about relative changes. The resulting optimization problem is written as

$$\begin{aligned} \boldsymbol{w}^*, \boldsymbol{v}^* = \underset{\substack{\boldsymbol{w},\boldsymbol{v}\in\mathbb{R}^m \\ \boldsymbol{K},\boldsymbol{M}\in\mathbb{R}^{n\times n} \\ \boldsymbol{u}_i\in\mathbb{R}^n, i=1,\ldots,k}}{\arg\min} & \left\{ \frac{\alpha_u}{2k} \sum_{i=1}^{k} \|\boldsymbol{P}\mu_i - \widehat{\gamma}_i\|_2^2 \right. \\ & \left. + \frac{\alpha_w}{2m} \left\|\nabla^2 \boldsymbol{w}\right\|_2^2 + \frac{\alpha_v}{2m} \left\|\nabla^2 \boldsymbol{v}\right\|_2^2 + \left(\sum_{e=1}^{m} \frac{w_e}{m} - \bar{w} \right)^2 \right\} \\ \text{s.t. } \boldsymbol{K} = \sum_{e=1}^{m} w_e \boldsymbol{K}_e, \quad & \boldsymbol{M} = \sum_{e=1}^{m} v_e \boldsymbol{M}_e, \quad \boldsymbol{K}u_i = \widehat{\omega}_i^2 \boldsymbol{M}\boldsymbol{u}_i,\ i = 1, \ldots, k, \end{aligned} \tag{5.5}$$

where α_u, α_w, and α_v are hyperparameters that balance the objective terms.

5.3.2.2 Optimization strategy

As defined in Equation 5.6, we approximately solve Equation 5.5 via a dual formulation of the problem. The eigen-constraints in Equation 5.5 are too strict to enforce directly, so we incorporate them as quadratic penalties in the dual problem. The weight of each penalty term is a dual variable, y_i, and we apply dual ascent to gradually increase these penalty weights.

$$\begin{aligned} \boldsymbol{w}^*, \boldsymbol{v}^* = \underset{\substack{\boldsymbol{w},\boldsymbol{v}\in\mathbb{R}^m \\ \boldsymbol{K},\boldsymbol{M}\in\mathbb{R}^{n\times n} \\ \boldsymbol{u}_i\in\mathbb{R}^n, i=1,\ldots,k}}{\arg\min} & \left\{ \frac{1}{2k}\sum_{i=1}^{k} y_i \left\| \boldsymbol{K}\boldsymbol{u}_i - \widehat{\omega}_i^2 \boldsymbol{M}\boldsymbol{u}_i \right\|_2^2 \right. \\ & \left. + \frac{\alpha_u}{2k}\sum_{i=1}^{k} \|\boldsymbol{P}\boldsymbol{u}_i - \widehat{\boldsymbol{\gamma}}_i\|_2^2 + \frac{\alpha_w}{2m}\left\|\nabla^2 \boldsymbol{w}\right\|_2^2 + \frac{\alpha_v}{2m}\left\|\nabla^2 \boldsymbol{v}\right\|_2^2 + \left(\sum_{e=1}^{m}\frac{w_e}{m} - \bar{w}\right)^2 \right\} \\ \text{s.t. } \boldsymbol{K} = \sum_{e=1}^{m} w_e \boldsymbol{K}_e, \quad & \boldsymbol{M} = \sum_{e=1}^{m} v_e \boldsymbol{M}_e. \end{aligned} \tag{5.6}$$

Equation 5.6 is a non-convex problem, but it is quadratic with respect to $\boldsymbol{w}, \boldsymbol{v}$ for fixed $\boldsymbol{u}_i$, and it is quadratic with respect to $\boldsymbol{u}_i$ for fixed $\boldsymbol{w}, \boldsymbol{v}$. Our procedure is to iteratively compute the closed-form solution for $\boldsymbol{U} = [\boldsymbol{u}_1 \ldots \boldsymbol{u}_k]$ and then $\boldsymbol{z} = [\boldsymbol{w}^\intercal, \boldsymbol{v}^\intercal]^\intercal$, thereby minimizing the objective function at each step. We update the dual variables according to

$$y_i^{t+1} = y_i^t + \eta \left\| \boldsymbol{K}^{t+1}\boldsymbol{u}_i^{t+1} - \widehat{\omega}_i^2 \boldsymbol{M}^{t+1}\boldsymbol{u}_i^{t+1} \right\|_2, \tag{5.7}$$

where $\eta > 0$ is the dual-variable update rate. Once the decision variables have converged, we output the minimizing solution $\boldsymbol{z}^* = [\boldsymbol{w}^{*\intercal}, \boldsymbol{v}^{*\intercal}]^\intercal$. Here $\boldsymbol{w}^*$ and $\boldsymbol{v}^*$ are the voxel-wise estimated Young's modulus and density values.

5.4 Simulated Experiments

In this section, we test our approach on the simulated vibration of 3D cubes with "defects," and we discuss the practical concerns of complex geometries, model mismatch, and damping.

5.4.1 Creating synthetic data

Cube model We model a cube as a $10 \times 10 \times 10$ hexahedral mesh, similar to a voxel grid. Each of the 1000 voxels is assigned a Young's modulus and density that correspond to either the primary material or a defect material. The material properties are chosen to resemble Jello and clay, respectively ($E_{\text{jello}} = 9000\,\text{Pa}$, $\rho_{\text{jello}} = 1270\,\text{kg}\,\text{m}^{-3}$, $E_{\text{clay}} = 5 \times 10^6\,\text{Pa}$, $\rho_{\text{clay}} = 7620\,\text{kg}\,\text{m}^{-3}$). We set a homogeneous Poisson's ratio of $\nu = 0.3$ [251].

Vibration animation Once the cube's mesh and material properties have been defined, we run a transient analysis in COMSOL [83], a commercial FEM software. The analysis calculates the cube's deformation over time given an initial condition. We choose an initial condition that mimics "plucking" a corner of the cube (e.g., an initial displacement vector of $(0.5\,\text{cm}, 0.5\,\text{cm}, 0.5\,\text{cm})$ of the top-front corner) and keep the bottom surface fixed. The resulting simulation represents free vibration with a Dirichlet boundary condition. The simulation is 6 seconds long at 2000 FPS. From the calculated displacements, we create an animation of the cube deforming over time by plotting the motion of random points on the surface of the cube with `matplotlib` [168].

5.4.2 Implementation and evaluation details

Mode selection We use `scipy`'s [319] peak-finder to automatically identify peaks in the log-power spectrum of motion amplitude, as described in Section 5.3.1. For a given simulation, this leads to around 20 to 30 selected peaks. Most peaks correspond to either a true mode or a linear combination of true modes whose frequencies fall in the same FFT frequency bin. However, a few peaks do not correspond to a true mode; we include these false modes in the synthetic results to best mimic analysis of real videos.

Inference cube mesh We infer on an $8 \times 8 \times 8$ hexahedral mesh. Since the simulations are done with a $10 \times 10 \times 10$ mesh, our results indicate robustness to a slight mesh mismatch. In the results presented, the simulation model and inference

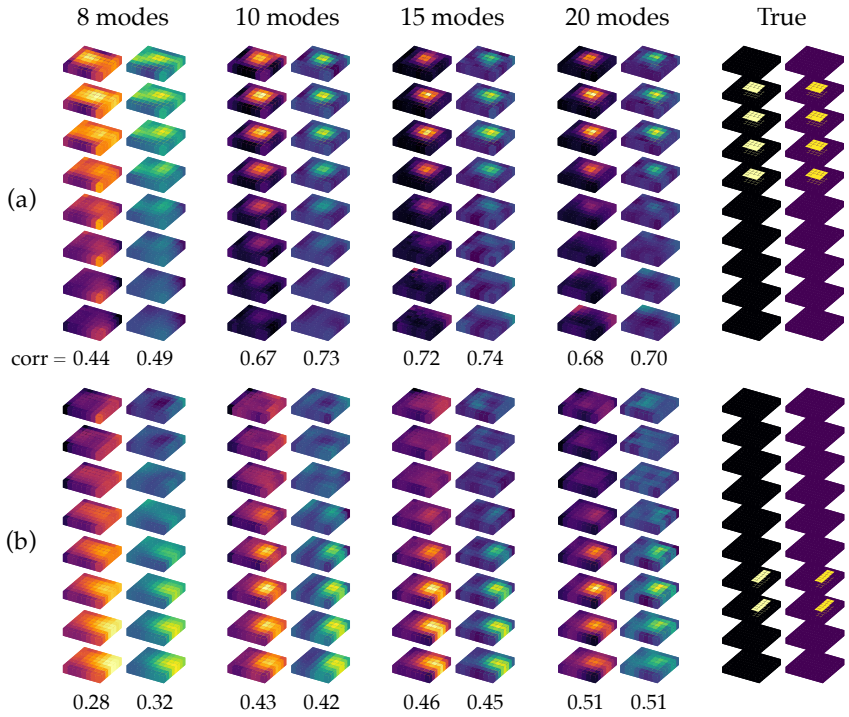


Figure 5.4: Reconstruction on two synthetic cubes with different defects. The given motion-extracted image-space modes range from the 8 to 20 lowest extracted modes. Normalized correlation generally increases as the number of modes increases. (b) is more challenging because the defect is smaller and closer to the bottom of the cube, where there is no motion.

model use linear elements.

Hyperparameters For every presented result from simulated data, $\alpha_w = 10^{-10}$, $\alpha_v = 10^{-7}$, and $\bar{w} = 9000$. Keeping these hyperparameters fixed, we ran a hyperparameter search on a dataset of 12 cubes with various defects to identify good values for α_u (Equation 5.6) and η (Equation 5.7). After testing all combinations of $\alpha_u \in \{1, 10, 100, 1000\}$ and $\eta \in \{0.1, 0.5, 1, 2, 5, 10\}$, we determined to set $\alpha_u = 10$ if the number of input modes is ≥ 10; otherwise, $\alpha_u = 1$. The dual variables y_i are always initialized to 1, with $\eta = 1$. The decision variables w and v are initialized to homogeneous values of 9000 Pa and 1270 kg m^{-3}, respectively (the true values of the primary material).

Evaluation Our method recovers *relative changes* in material properties (see Section 5.3.2.1). As such, the normalized 3D images of estimated Young's modulus and density should match the true normalized properties. We use normalized correlation between the estimated image and ground-truth image as the reconstruction score.

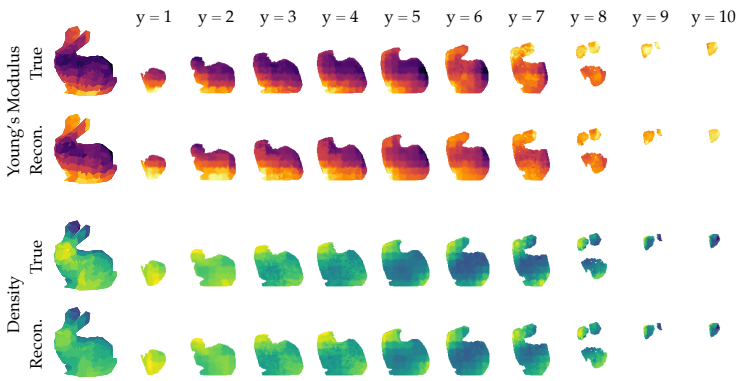


Figure 5.5: Reconstruction for the Stanford Bunny from (true) image-space modes. Slices along the y-axis are shown.

5.4.3 Results

Figure 5.4 shows results for two different cubes with defects appearing at different locations. These results are obtained from noisy, motion-extracted image-space modes. As more modes are observed, the inverse problem becomes better constrained, sharpening the image of the interior defect. Also note that a defect near the top of the cube is easier to identify than one near the bottom. This is because the base of the cube is fixed and thus provides less motion signal.

Complex geometry To demonstrate the approach on a more complex geometry, Figure 5.5 shows a volumetric reconstruction of material properties for the Stanford Bunny [309]. We voxelize the volume containing the tetrahedral mesh[2] of the bunny into an $8 \times 8 \times 8$ grid and match each mesh element to the nearest voxel, resulting in about 21 elements per voxel. 20 true image-space modes of the monocular view of the bunny shown are used for this reconstruction.

We next consider some challenges that may arise with real-world data: geometric mismatch and damping. More investigations into model mismatch are provided in the supplementary material.

5.4.3.1 Geometric mismatch

Figure 5.6 shows what happens when the dimensions of the inference mesh do not match the cube's true dimensions. We gradually increase the length of the inferred cube geometry in the x-direction from 1 to 1.4 times the true length. Scaling the length in one direction results in a gradually degrading estimate of the defect size.

[2]The bunny surface mesh is from `https://www.thingiverse.com/thing:151081`, and tetrahedralization is done with TetWild [164].

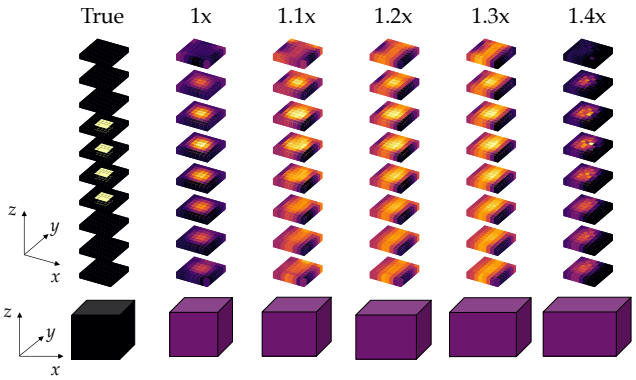


Figure 5.6: Geometric model mismatch. From 10 motion-extracted image-space modes, we infer on a mesh of various incorrect geometries, extending the inferred geometry width by a multiple of the true width. Results on Young's modulus show that some geometric mismatch can be accommodated.

However, even with 30% geometric error, we are still able to distinguish that there is a defect located in the central region of the cube.

5.4.3.2 Damping

Real-world objects exhibit various types of damping, which can affect both the frequencies and relative phases of their modes. To simulate damping, we incorporate Rayleigh damping into our synthetic cubes. Our damping parameters were estimated following the procedure outlined by Davis et al. [91], who fit a Lorentzian curve to a peak in the motion power spectrum to estimate the damping ratio. We find that Jello cubes exhibit significant damping: from a real video of one, we estimated critical damping ratios of 0.01749 at 12.5 Hz and 0.01999 at 15.5 Hz.

Through realistic simulations in COMSOL, we find that damping poses the additional challenge of fewer observable modes. We can increase the number of observed modes by extracting modes from multiple simulations with different initial conditions. For example, "plucking" the top-back corner of a cube will cause slightly different modal expression than "plucking" its top-front corner. Figure 5.7 shows reconstruction for damped cubes. From two different plucking conditions, we are able to extract between 7 to 15 modes and use these modes to coarsely reconstruct the defect.

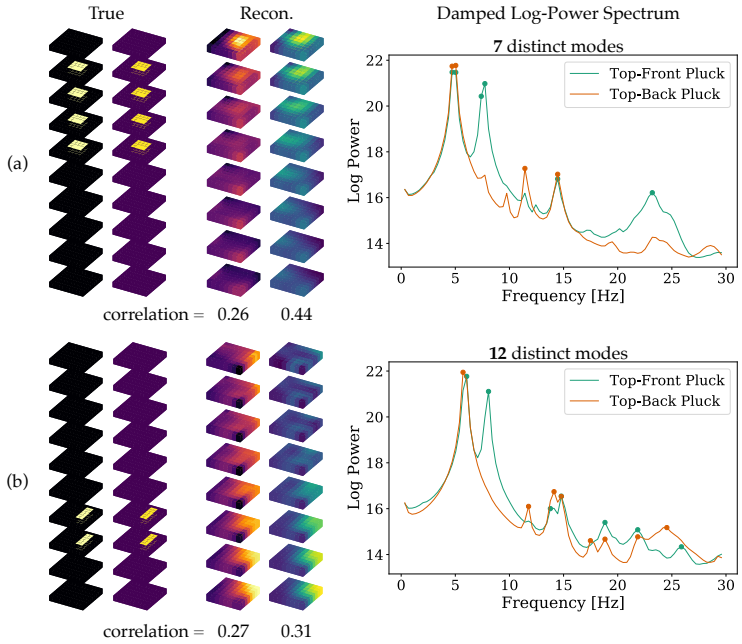


Figure 5.7: Reconstructions from two simulated damped cubes. Animations (3 seconds at 2000 FPS) of two different forcings are done: (1) a small initial displacement of the cube's top-front corner ("Top-Front Pluck") and (2) a small initial displacement of its top-back corner ("Top-Back Pluck"). Modes (marked as dots on the line plot) are selected based on the log-power spectrum of motion amplitude. Asymmetry plays a role in determining how many distinct modes are observable. As a cube becomes more asymmetric in its material-property distribution, its repeated eigenfrequencies become more separated. Since the defect in (b) is more off-center than the defect in (a), more distinct modes are identifiable in those simulations. In (a), with only 7 observed image-space modes, the reconstruction quality is consistent with Figure 5.4, which shows only a coarse defect reconstruction when given 8 modes. (Note: in (b), although the number of observed modes is > 10, we show the reconstruction for $\alpha_u = 1$ instead of 10.)

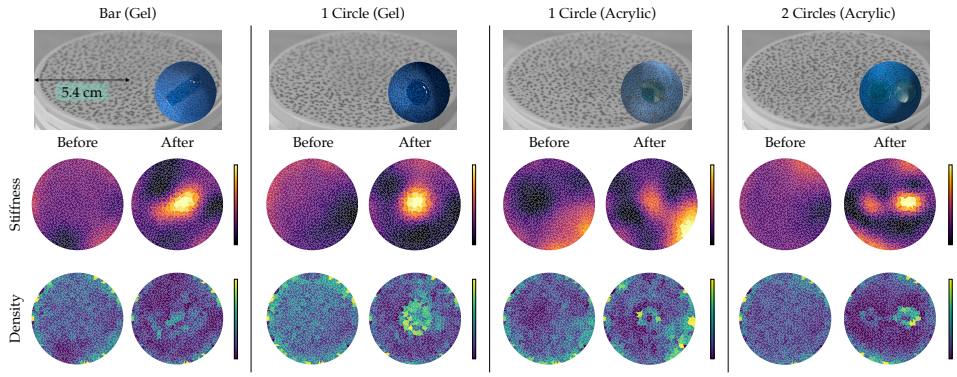


Figure 5.8: Reconstruction from real videos of drums. The defects shown are a gel bar, gel circle, acrylic circle, and two acrylic circles, applied to the underside of the drum head. For each defect, we recorded a video of the drum pre- and post-defect. One cannot see the defect in a video frame, but after applying our method, we were able to image the defects as changes in stiffness and density. For each type of defect, the "Before" and "After" material properties are plotted with the same normalized colormap.

5.5 Real-world Experiments

To demonstrate the potential of our approach in the real world, we applied it to real videos of drum heads and Jello cubes. With drum heads, we achieve reconstructions that allow one to discern distinct defect shapes, providing a proof-of-concept for defect discovery and characterization using our approach. The damping of 3D Jello cubes poses a challenge for extracting enough image-space modes for high-fidelity defect reconstruction; nonetheless, we are able to identify heterogeneity in the cube.

5.5.1 Real drums

We tested our method on a dataset of real drum heads, each altered with a defect beneath the surface. The defects were created from two materials: nail hardening gel (painted beneath the surface) or acrylic plastic circles (glued onto the bottom of the surface). Although all DOFs of the 2D membrane are visible in the video, solving for material properties is still ill-posed because we observe a limited number of projected modes (see Figure 5.9).

Results Figure 5.8 shows estimated Young's moduli and densities for various drum heads, before and after defects were included. For both materials, the defect appears as a bright region in stiffness. Interestingly, gel and acrylic appear differently in their density estimations. For gel defects, there is a bright, filled region in the density map that corresponds to a higher mass from the defect. For acrylic defects, this

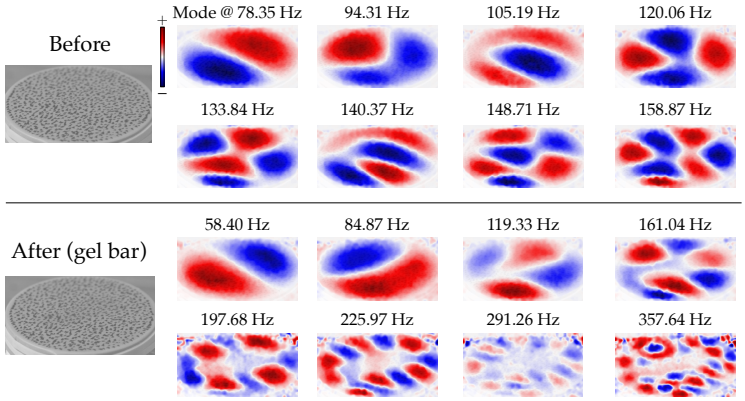


Figure 5.9: Extracted image-space modes from real videos of a drum, before and after a defect was introduced. The defect shown here is a gel rectangle, which was painted on the bottom of the drum head. Only vertical motion is shown.

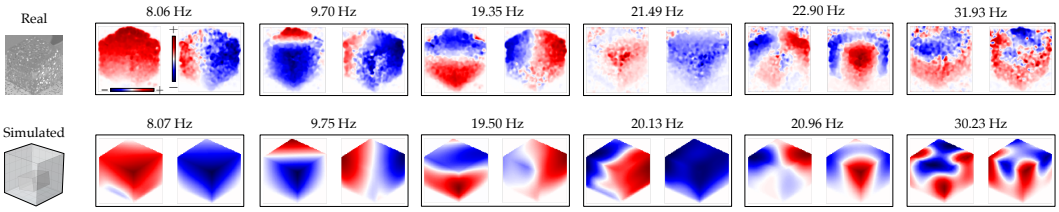


Figure 5.10: Extracted image-space modes from real videos of a Jello cube with an interior clay defect ("Real"). The true image-space modes identified from a COMSOL simulation of a cube with a defect are shown for comparison ("Simulated"). Each observed image-space mode has a corresponding simulated mode that appears similar in both image-space and eigenfrequency.

change only appears on the edges of the defect. This is possibly because the acrylic circles are much stiffer than gel, which bends along with the rubber membrane. These results indicate that our proposed approach could be used to identify not just the presence of a defect, but also its shape.

5.5.2 Real cubes

To gain further insight into practical challenges, we conducted an experiment on a real Jello cube with an interior clay defect. This object is more challenging than the drum membrane in two respects: (1) the high damping of Jello, perhaps due to its water content, and (2) the large proportion of unseen DOFs in the cube geometry. The cube had dimensions 4.9 cm×4.7 cm×4.5 cm, while the rectangular clay defect was of size 2.2 cm × 2.9 cm × 1.4 cm. We recorded three videos of the cube under different initial deformation conditions (e.g., in one video, we lifted and then quickly released the top-front corner of the cube). Multiple videos allowed us to identify

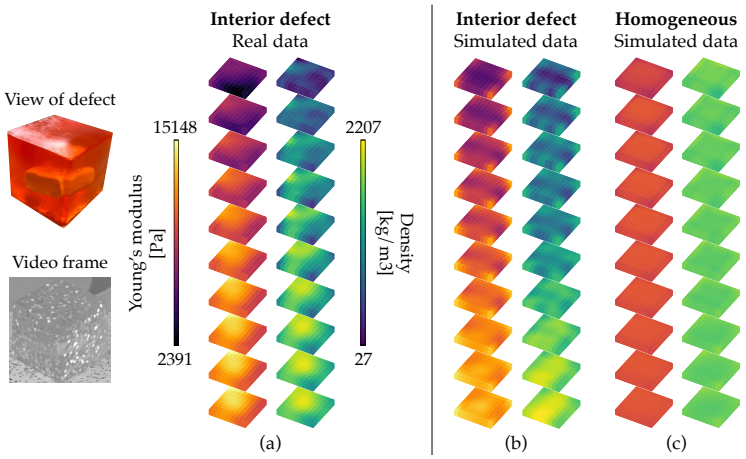


Figure 5.11: Reconstruction from real data vs. reconstructions from simulated data of the defect cube and a homogeneous cube. All reconstructions use 6 image-space modes and the same hyperparameters and are plotted with the same colormaps. As Figure 5.10 shows, there is a one-to-one correspondence between the modes given for (a) and the modes given for (b). (a) is more similar to (b) than to (c), indicating that with 6 modes, we can differentiate between a cube with a defect and a homogeneous one.

more unique modes and average duplicate ones.

We created two COMSOL models that would be comparable to the real Jello cube: one simulated cube had a clay defect and the other did not. The Young's modulus values of the Jello and clay were set so that the natural frequencies would agree with those observed. The Rayleigh damping parameters were estimated following the method mentioned in Section 5.4.3.2. As illustrated in Figure 5.10, the COMSOL image-space modes of the simulated cube with a defect appear similar to those captured from the real Jello cube.

Results Figure 5.11 shows the result of our approach applied to real video data of the Jello cube. The reconstruction is obtained using six unique, motion-extracted image-space modes. As expected based on our findings in Figures 5.4 and 5.7, we are able to recover only a large-scale estimation of material properties with six constraining modes. Still, it is very promising to have identified inhomogeneities in a real 3D object with our method. We further compare this real-data reconstruction to reconstructions obtained from simulated data of a homogeneous cube and one with a defect, showing that we achieve a solution that resembles the solution for the simulated defect cube more than it resembles the solution for the simulated homogeneous cube (Figure 5.11).

5.5.3 Experiment details

5.5.3.1 Drum experiment

Drum construction The drums were constructed by fixing a thin rubber sheet over a 4"×4" PVC adaptor with a rubber band. We tested defects of two materials: nail hardening gel and acrylic plastic circles. For each defect, we recorded a video of the homogeneous drum before the defect was applied for comparison. We drew a speckle pattern on the drum head for texture.

Vibration-capture setup Figure 5.12 shows a schematic of the setup. We taped the drum onto an optical table, with the high-speed camera standing on the same optical table. The excitation source was a PreSonus Sceptre S8 loudspeaker, which sat on a platform separate from the optical table and was pointed at the drum. For each video, we recorded the drum head's vibration in response to a 3.5-second linear frequency sweep (50 Hz to 1000 Hz) played by the speaker.

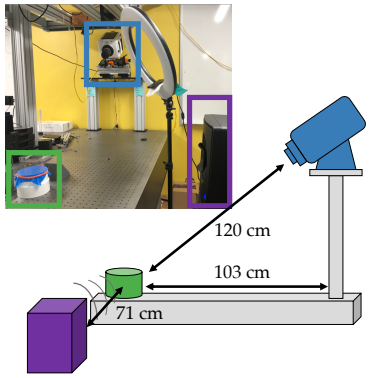


Figure 5.12: Experimental setup for real drums. Vibrations were induced by a loudspeaker and recorded with a high-speed camera.

Video capture Our camera was a Phantom V1610 high-speed camera. Each video was captured at 6000 FPS at an image resolution of 288×384. To reduce noise, we averaged every two frames for a resulting temporal frequency of 3000 FPS. Note that in Figure 5.9, the drums vibrate at frequencies below 120 Hz. While we chose to first demonstrate our approach using a high-speed camera, where compression and camera noise are less challenging, many modal frequencies can be captured on a consumer camera.

Extracting image-space modes We found that in real videos, some level of manual selection was necessary to verify peaks in the motion amplitude spectrum as

modal motion. For instance, spurious camera motion would often appear as spikes in the spectrum. Verification was done by visually inspecting the magnified motion in the video at the frequency in question (following the method proposed in [321]). We believe that in the future this step could be automated. The number of extracted modes ranged from 12 to 31, depending on the video.

Inference details We modeled each drum as a triangular membrane mesh with 1530 linear elements and inferred material properties on a 20×20 pixel grid. In the presented results, the hyperparameter values are $\alpha_u = 10^{12}$, $\eta = 1$, $\alpha_w = 0.1$, $\alpha_v = 0.1$, and $\bar{w} = 10^6$. The parameters $\boldsymbol{w}$ and $\boldsymbol{v}$ are initialized to uniform values of 1×10^6 Pa and $1000\,\mathrm{kg\,m^{-3}}$, respectively, and reflect the estimated stiffness and density of latex.

5.5.3.2 Jello cube experiment

Our inference model was a $10 \times 10 \times 10$ hexadral mesh with linear elements. The optimization hyperparameters were $\alpha_u = 0.1, \eta = 1, \alpha_w = 10^{-10}, \alpha_v = 10^{-8}$, and $\bar{w} = 10,000$. The parameters $\boldsymbol{w}$ and $\boldsymbol{v}$ were initialized to 10 000 Pa and $1500\,\mathrm{kg\,m^{-3}}$, which are the estimated Young's modulus and measured density values of Jello.

5.6 Discussion

We have shown that it is possible to recover spatially-varying material properties of 3D objects from monocular video, even in regions unseen in the image. This can be done by decomposing 2D surface motion into image-space modes, and then solving for the Young's modulus and density values that agree with the observed modes. We demonstrated our method on synthetic and real-world data of objects ranging from 2D drum heads to a 3D bunny.

Our results highlight that monocular videos are a simple, yet powerful, source of data for understanding the physical properties of objects around us. We believe that videos are a promising domain for further research into non-destructive testing, turning everyday visual sensors into tools for material characterization.

5.6.1 Limitations

Our method assumes that materials are isotropic and linear elastic. Linear elasticity is only satisfied if the object's motion is small. Further, we assume that the geometry is, at least roughly, known ahead of time (see Figure 5.6).

For now, we have validated our method with a high-speed camera. We have not yet demonstrated the approach with consumer-grade cameras that bring additional challenges such as image compression and noise. Generally, the hardware required depends on the amplitude and frequencies of the modes. For large structures that vibrate below 100 Hz [225, 193, 264], a smartphone camera theoretically provides enough temporal frequency. Objects that vibrate more quickly require high-speed cameras. Tricks such as temporal aliasing via a strobe may expand the capabilities of a camera. We have shown one example of attaining simulation quality on a real cube. Further work needs to be done to achieve consistent results across a variety of objects, with additional camera views offering one simple possible solution.

The primary challenge with applying this technique to real-world objects is capturing enough image-space modes to recover interior defects with high fidelity. Damping causes a reduction in the number of modes that can be extracted. We demonstrated that for damped Jello cubes, we could still recover some information from only six image-space modes (Section 5.5.2). Even so, in many objects, damping will pose a more significant challenge. In the future, acquiring more modal observations could be solved by exciting modes through mechanical vibration tables.

Chapter 6

VISUAL SURFACE WAVE ELASTOGRAPHY

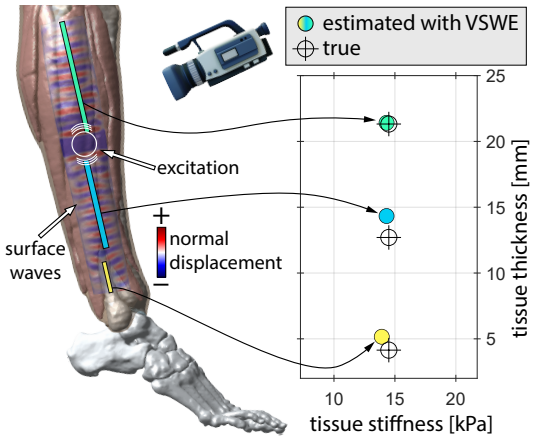


Figure 6.1: Estimating subsurface tissue properties of a human leg with VSWE. From a video of surface wave motion, VSWE estimates the thickness and stiffness of the soft tissue layer. In this example, we are able to recover the three different thicknesses of the three regions highlighted on the leg. We also recover the stiffness of the leg tissue, which does not change across the leg.

In the previous chapter, we analyzed global modes of surface motion to infer interior material structure. In this chapter, we will see how the surface motion in a local region of a large, complicated structure (e.g., a human body) can help us infer the local depth and stiffness of the material.

How waves propagate on the surface of a medium reveals information about its subsurface properties. For example, by watching how ocean waves evolve and break as they near the shore, one can infer the rises and dips of the seafloor below [328]. Waves can also be observed on the surfaces of biological systems. Imagine applying a massage gun to your calf. The ripples on your skin convey information about the underlying layers of fat, muscle, and bone. In fact there is a well-defined relationship linking the thickness and stiffness of each layer to the wave propagation behavior. We propose *visual surface wave elastography* (VSWE), a physics-based method to estimate the thickness and stiffness of a medium from a video of waves on its surface.

We are primarily motivated by the task of biological tissue characterization, which has broad applications including at-home health monitoring. For example, tumors [162, 197], musculoskeletal degeneration [315, 198, 318], and liver disease [135, 291] often lead to changes in tissue thickness or stiffness. However, existing techniques for elastography (i.e., measuring tissue stiffness) require specialized ultrasound [330, 296] and sometimes magnetic resonance [222] devices, along with medical experts to operate the equipment [32, 57]. We show that it is possible to obtain coarse estimates of thickness and stiffness given surface waves measured with

just a video camera. Another application is human computer interaction (HCI), where visually inferring subsurface changes in muscle stiffness may unlock new modes of gesture recognition. In biomechanics, many regions of the body are often modeled as a series of layers of biological tissue, such as skin, muscle, or bone [20, 134]. Similarly, in this work, we model the medium as a layer of soft tissue with unknown thickness and stiffness atop a layer of bone. We leverage the mathematical relationship between the geometric and mechanical properties of the soft layer and the wave propagation behavior of the medium.

The wave propagation behavior of a material is compactly described by a mathematical object known as a *dispersion relation*. Under some common biomechanical assumptions in our layer model (discussed further in Section 6.2.1), the thickness and stiffness of the soft layer fully determine the dispersion relation. The main idea of our method is to find the thickness and stiffness values that lead to a dispersion relation that matches the dispersion relation extracted from surface waves observed in the video. Similarly to VVT in Chapter 5, VSWE demonstrates the effectiveness of fitting to a physics model to solve an ill-posed material-characterization problem.

In this chapter, we first discuss related work on video-based and wave-based material characterization in Section 6.1 and provide background on wave mechanics in Section 6.2. We then describe in Section 6.3 our method, which consists of taking a video of surface waves, extracting the dispersion relation from the video, and then solving an optimization problem to find the best-fitting thickness and stiffness parameters. In Section 6.4, we validate our method on real and simulated data, including real videos of gelatin-based phantoms that mimic biological tissue and realistic simulations of a human leg. Section 6.6 provides details about our experiments.

6.1 Related Work

6.1.1 Video-based material characterization

Videos encode rich information about the environment and physical objects. For example, previous work leveraged surface vibrations [90] or sound [120, 73] in video to recover unknown information. For material characterization tasks, the way an object moves in a video provides useful information about its physical properties. Videos have been used to estimate material parameters of specific object classes, such as fabrics [35, 228, 47], rods [91, 92], and trees [324], whose dynamics are entirely visible. In contrast, VVT [119] infers *subsurface* material properties. Specifically, it recovers the spatially-varying stiffness and density throughout a 3D object with known geometry by analyzing its global vibrational modes.

Surface waves have been under-utilized for video-based material characterization, even though they can also be observed in video and contain useful information about underlying physical properties. Our work shows how to leverage surface waves in video to estimate geometric and mechanical properties. A surface-wave-based approach is beneficial when one wishes to analyze local structure without having to model the global structure, which may be complicated. In the case of biological tissue characterization, for example, it would be impractical to model the 3D vibrational modes of the entire human anatomy in order to estimate local tissue properties. Our surface-wave-based approach circumvents the need to model a complex geometry and solve for its global modes by targeting local regions where we can analyze wave modes of a simpler geometry.

6.1.2 Wave-based material characterization

In general, wave-based imaging relies on an understanding of how physical characteristics of interest affect wave propagation. For example, magnetic resonance imaging [150] uses knowledge of how radio-frequency waves are absorbed and re-emitted by different types of tissues in the body. Ultrasound techniques, which are pervasive in non-destructive testing [177], medical imaging [23], and wearable technology [166], use high-frequency mechanical bulk waves to characterize tissue and locate features. Among tissue characterizaton techniques, transient elastography [330], shear wave elastography [296], and magnetic resonance elastography [222] leverage knowledge of bulk wave physics to estimate the elasticity of tissues and organs. However, such methods often require not only high-end, expensive equipment, but also trained medical specialists [32, 57], making regular screening infeasible. In contrast, our work leverages surface waves, which are generally less

expensive to observe than their bulk counterparts.

Although very different fields from our application of interest, geophysics and seismic imaging provide insights into the possibility of using surface waves to infer subsurface features [245, 271, 142, 208, 236]. Inspired by the success of surface-wave methods in seismology, we propose to harness surface waves observed in video to infer subsurface tissue properties. Whereas previous work [36] suggested an approach using sensors sparsely placed on the skin, we leverage dense visual data. Vision-based tissue characterization would enable the next generation of health monitoring systems that take advantage of the ubiquity of visual sensors.

6.2 Background

6.2.1 Dispersion relations

Just as sound waves can be expressed as the combination of simple harmonic modes, waves traveling through any medium can be expressed as the combination of wave modes of different spatial and temporal frequencies. A *dispersion relation* compactly describes wave propagation by defining all the possible wave modes of a medium. Specifically, it defines the *wavevector* and *frequency* of each possible wave mode, where the wavevector and frequency indicate the spatial and temporal rates of oscillation, respectively.

Mathematically, the dispersion relation can be determined by solving the harmonic elastic wave equation (a PDE) subject to phase-shifted periodic boundary conditions. Let $\mathbf{x} = [x, y, z]^\intercal$ denote the spatial location and $\boldsymbol{u}(\mathbf{x}) = [u(\mathbf{x}), v(\mathbf{x}), w(\mathbf{x})]^\intercal$ denote the displacement at $\mathbf{x}$. Assuming an isotropic linear-elastic material, the harmonic elastic wave equation can be written as

$$\omega^2 \mathcal{M}(\boldsymbol{u})(\mathbf{x}) = \mathcal{K}(\boldsymbol{u})(\mathbf{x}) \qquad \forall \mathbf{x} \in \Omega. \tag{6.1}$$

Note that Equation 6.1 is the infinite-dimensional version of Equation 5.1 used for VVT. The linear operators $\mathcal{M}$ and $\mathcal{K}$ represent the mass and stiffness of the medium. The mass operator $\mathcal{M}$ depends on the density field $\rho(\mathbf{x})$. The stiffness operator $\mathcal{K}$ depends on the elastic modulus $E(\mathbf{x})$ and Poisson's ratio $\nu(\mathbf{x})$. The eigenvector solution $\boldsymbol{u}(\mathbf{x})$ and eigenfrequency ω describe the shape and temporal frequency, respectively, of the wave mode. The solutions also depend on the domain Ω, which in our case is parameterized by the thickness and length of the tissue layer.

When solving Equation 6.1, boundary conditions are necessary to impose real-life assumptions of the physical system. In our setting, we assume that Ω is a finite subregion of the full wave medium (e.g., a section of a leg). Furthermore, to simplify our analysis, we target waves traveling in one direction (i.e., along the x direction). This leads us to apply 1D periodic boundary conditions on the domain $x \in [0, a]$ to impose the assumption that the wave medium continues past the boundary of Ω.

Specifically, we apply the Bloch-Floquet [124, 41] (a.k.a. phase-shifted) periodic boundary conditions, which define a boundary condition for every wavenumber $\gamma \in [0, \pi/a]$. We note that in general γ is a wavevector, but in the case of 1D wave analysis, it is a wavenumber that indicates the number of wavelengths per unit length. For a given γ, the boundary condition is defined as

$$\boldsymbol{u}(x = a, y, z) = \boldsymbol{u}(x = 0, y, z) e^{\mathbf{i}\gamma a}. \tag{6.2}$$

In words, the solution at one end ($x = 0$) only differs from the solution at the other end ($x = a$) by the phase shift γa.

To compute a dispersion relation, we solve Equation 6.1 subject to Equation 6.2 for $\gamma \in [0, \pi/a]$. For each γ, we solve a generalized eigenvalue problem and obtain multiple $(\omega, \boldsymbol{u}(\mathbf{x}))$ solution pairs. The dispersion relation is exactly the set of all ω solutions for every γ.

6.2.2 Assumptions for tissue characterization

For biological tissue characterization, we can model the medium as a layer of soft tissue atop a hard bone layer, where the top layer has uniform thickness T. We note that throughout this work, we refer to the elastic modulus of the tissue as its stiffness. We make several assumptions to reduce the complexity of the problem:

1. The soft layer has a uniform stiffness $E(\mathbf{x}) = E$.
2. The density and Poisson's ratio of the soft tissue are known. We set $\rho = 1\,\mathrm{g\,cm^{-3}}$ [346] and $\nu = 0.45$ [174].
3. The bone is much stiffer than the soft tissue. This means neither the thickness nor the exact stiffness of the bone layer matters, and we can model it as motionless.

These assumptions are common in biomechanical analysis [20, 134]. Under these assumptions, the stiffness operator $\mathcal{K}$ in Equation 6.1 only depends on the elastic modulus E (a.k.a. stiffness) of the soft tissue, and the spatial domain Ω on which we solve Equation 6.1 only depends on the thickness T. This mean we can fully determine the dispersion relation based on T and E, so we denote the dispersion relation as

$$\mathfrak{D}(T, E) = \{\omega_i(\gamma)\}_{i=1}^{N} \quad \forall \gamma \in [0, \pi/a], \tag{6.3}$$

where N is the number of branches, or the number of eigenvalue solutions computed for Equation 6.1.

The dispersion relation is only concerned with the spatial (γ) and temporal (ω) frequencies of the mode shapes rather than the mode shapes $\boldsymbol{u}(\mathbf{x})$ themselves. From waves that express themselves, even in part, on the surface, we can extract spatial and temporal frequency information to obtain a dispersion relation. Figure 6.2 shows how the dispersion relation changes with the thickness and stiffness of the

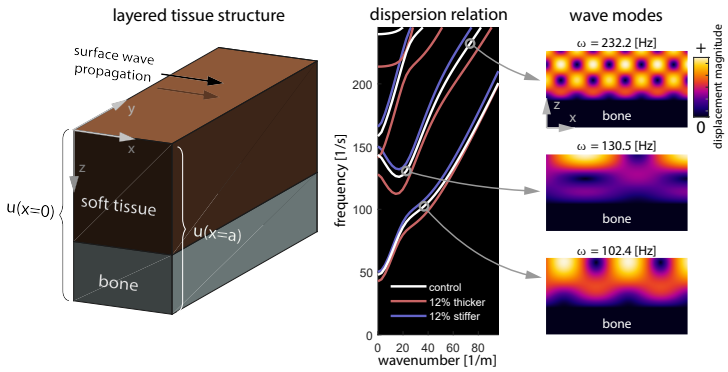


Figure 6.2: The dispersion relation depends on both the thickness and the stiffness of the soft tissue. We show how the dispersion relation changes when the soft tissue is made slightly thicker or slightly stiffer, with examples of the wave modes that are embedded in each dispersion relation. Wave mode dynamics occurring beneath the surface affect their expression on the surface.

soft tissue layer. *Our method leverages the sensitivity of the dispersion relation to small changes in these parameters.*

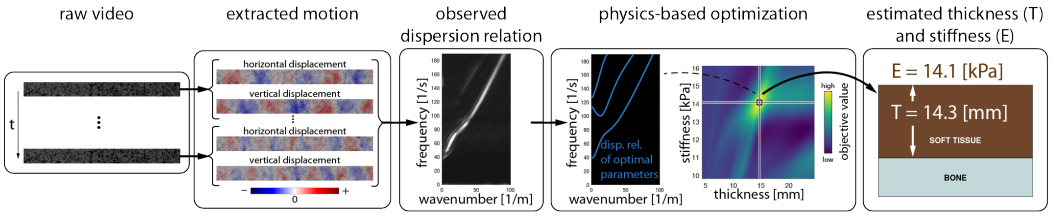


Figure 6.3: VSWE method overview. Given a video of the surface of the medium of interest, we first extract motion fields in image-space. From these we extract the observed dispersion relation, which dictates the spatial and temporal frequencies of waves that travel through the medium. We solve a physics-based optimization problem to estimate the thickness and stiffness values that lead to a theoretical dispersion relation that best agrees with the observed dispersion relation.

6.3 Method

This section describes the computational method of VSWE. The input is a video of motion on the surface of the medium of interest, and the output is the estimated thickness and stiffness of the medium. In between there are two broad stages: (1) extracting a dispersion relation from the video and (2) solving for the tissue properties that best agree with the observed dispersion relation. Figure 6.3 presents an overview of the pipeline.

6.3.1 Extracting a dispersion relation from video

6.3.1.1 Motion extraction

The first step is to quantify the image-space displacements in the video. Similarly to VVT in Chapter 5, we use phase-based motion processing [321] to obtain the image-space horizontal displacement $d_x(t, x, y)$ and vertical displacement $d_y(t, x, y)$ at each pixel (x, y) and video frame t, where the displacements are relative to the first frame. Note that x and y denote image-space coordinates here.

6.3.1.2 Estimating the dispersion relation

We process the image-space surface displacements into a dispersion relation via the fast Fourier transform (FFT). We assume that the waves are traveling in the horizontal direction in image-space. For each row of pixels in the $d_x(t, x, y)$ video, we take a 2D FFT, transforming the space dimension x to the wavenumber dimension γ and the time dimension t to the frequency dimension ω. That is, for the row where $y = y_i$, we convert the $d_x(t, x, y = y_i)$ signal to the complex-valued signal $\widehat{\boldsymbol{d}}_x(i) = \widehat{d}_x(i)(\gamma, \omega)$, whose magnitude gives the dispersion relation. To improve the

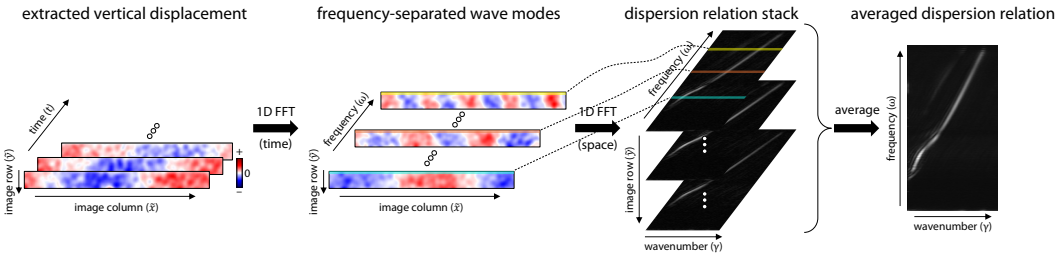


Figure 6.4: Obtaining a dispersion relation from image-space motion. Here we demonstrate with the vertical displacements taken from a real video of a gelatin sample. Taking a 1D FFT across time for every pixel allows us to separate wave modes by frequency (ω). Taking another 1D FFT across space (in the $\tilde{x}$ direction) decomposes the wave modes into spatial frequencies (γ). The result is a 2D FFT representing the dispersion relation. We average the dispersion relations across image rows and both displacement directions ($\tilde{u}$ and $\tilde{v}$).

SNR, we average across all the rows and both displacement directions, resulting in the observed dispersion relation $\boldsymbol{D}_{\text{obs}}$:

$$\boldsymbol{D}_{\text{obs}} := \frac{1}{2H} \sum_{i=1}^{H} \left(\left| \widehat{\boldsymbol{d}}_x(i) \right| + \left| \widehat{\boldsymbol{d}}_y(i) \right| \right), \tag{6.4}$$

where H is the number of rows in the image. Figure 6.4 illustrates this process, showing how the 2D FFT pulls out spatial and temporal frequency content.

The main challenge is that the observed dispersion relation likely does not fully agree with the true theoretical dispersion relation of the medium. First of all, it is incomplete: wave modes that do not prominently express on the surface or require large amounts of energy to propagate may not appear strongly in the video, and frequencies above the Nyquist sampling rate of the video cannot be captured. Second of all, the extracted motion may be noisy due to camera noise or spurious motion (from, e.g., lights flickering or rigid-body motion). VSWE has some margin of error due to the discrepancy between the observed and true dispersion relations, although we show in our experiments in Section 6.4.1 that it is sensitive to 5% changes in the true parameters.

6.3.2 Estimating the thickness and stiffness

The next stage is to find the geometric and mechanical parameters that best explain the observed dispersion relation. We do this via physics-based simulation and optimization. For a given hypothesized thickness T and stiffness E, we employ FEM to numerically compute the dispersion relation. We wrote specialized FEM

code to do so efficiently.[1] The goal is to maximize an objective function that rewards similarity between $\mathfrak{D}(T, E)$ and $\boldsymbol{D}_{\text{obs}}$.

After testing various objective functions, we found that treating dispersion relations as images and using SSIM [327] to work best. That is, we aim to maximize the SSIM between the images of the observed and proposed dispersion relations. The observed dispersion relation $\boldsymbol{D}_{\text{obs}}$, since it is derived via a 2D FFT, is already represented as an image. However, the physics-based dispersion relation $\mathfrak{D}(T, E)$ is computed as a set of curves. We transform $\mathfrak{D}(T, E)$ into an image $\boldsymbol{D}_{\text{hyp}}(T, E)$ by assigning intensities with a Gaussian kernel based on the distance from each (γ, ω) pixel to the curves. Then we solve the following optimization problem:

$$T^*, E^* = \arg\max_{T,E} \text{SSIM}\left(\boldsymbol{D}_{\text{hyp}}(T, E), \boldsymbol{D}_{\text{obs}}\right). \tag{6.5}$$

We solve Equation 6.5 with a grid search over possible thickness and stiffness values, although more efficient approaches can be taken for computationally demanding settings.

[1] Our code is available at `https://github.com/aco8ogren/tissue-dispersion`.

6.4 Experiments

We validate our approach on real and simulated data. We demonstrate remarkable accuracy on real gelatin samples, and we demonstrate recovering spatially-varying thickness and stiffness across a simulated 3D human leg.

6.4.1 Simulated plane strain

We created a two-layer tissue model in COMSOL [83] that would allow us to simulate samples with ground-truth geometric and mechanical parameters. We set stiffness values on the order of 10 kPa, which is a reasonable range for soft tissue [107]. We simulated the model's response to a chirp excitation signal applied to the leftmost side of the surface of the sample. The simulations were done with the plane-strain assumption, which allows for modeling 3D dynamics with only two dimensions by assuming that strain is constant in the z direction. We used the simulated horizontal and vertical displacements to obtain dispersion relations.

We tested the sensitivity of our method by applying slight changes to the true thickness and stiffness of the simulated sample. That is, for a certain thickness and stiffness, we can simulate samples assuming those parameters as well as perturbations of those parameters and assess whether VSWE picks up on these slight changes. As an experiment, we simulated samples at 5 distinct thicknesses and 4 distinct moduli. For each thickness-stiffness pair, we looked at 5% and 10% perturbations of either parameter. This led to 9 clusters with 9 simulated samples each. As Figure 6.5 shows, within each cluster, the estimated parameters track with changes in the true parameters, even when the true parameters have changed by only 5%.

6.4.2 Real videos of gelatin-based phantoms

To test our method on real data, we created gelatin-based phantoms (i.e., samples mimicking biological tissue) of varying thicknesses. We poured varying amounts of gelatin (1000 mL, 1100 mL, and 1500 mL) into the same-sized container, leading to three different thicknesses. We set the samples in the refrigerator for about 24 hours. Once they were set, we sprinkled garlic powder onto the samples to create texture for motion extraction.

For each set sample, we measured the thickness with calipers, and we used rheometry to obtain ground-truth stiffness values. To obtain videos, we applied a shaker at one end of the sample to excite waves with a chirp signal and recorded the surface of the sample with a high-speed camera at 600 FPS (each video was about four seconds long). Figure 6.6 shows pictures of the experiment setup.

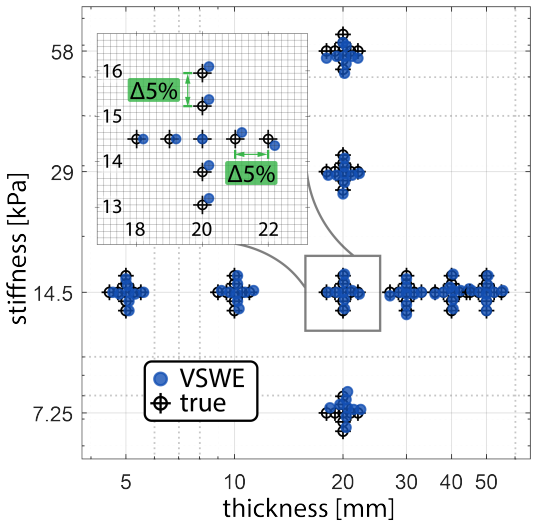


Figure 6.5: VSWE sensitivity analysis. We applied VSWE to plane-strain simulations with a range of true thickness and stiffness values, along with slight perturbations of those values. In each cluster, we perturbed the true parameters by ±5% and ±10% from the central values and found that VSWE was sensitive to these changes.

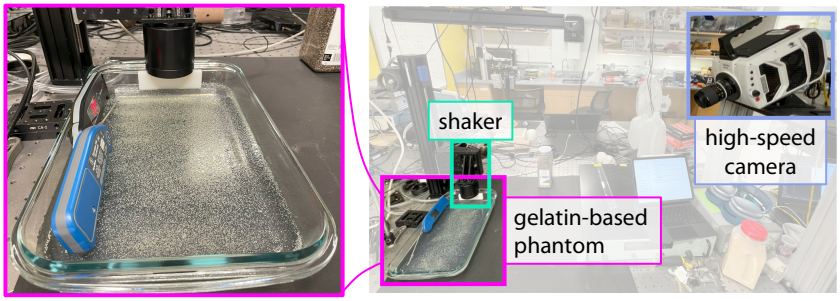


Figure 6.6: Real-world experiment setup. A shaker was applied on one side of the gelatin-based phantom to excite waves in the medium. A high-speed camera, zoomed in on the phantom's surface, captured videos of the response to the shaker. As the left picture shows, we took temperature recordings with a thermometer.

For each of the three samples, we took many rheometry measurements and videos over the course of about one hour after removing the sample from the refrigerator. In total we obtained about 60 videos per sample over time, each with corresponding ground-truth parameters. Figure 6.7 shows the inferred parameters for each sample and each point in time. Figure 6.7(a) shows that our method clearly identifies the three different thicknesses of the samples. As a sample spends more time out of the refrigerator, its temperature increases and its stiffness decreases. Figure 6.7(b) shows that the estimated stiffness decreases accordingly with temperature.

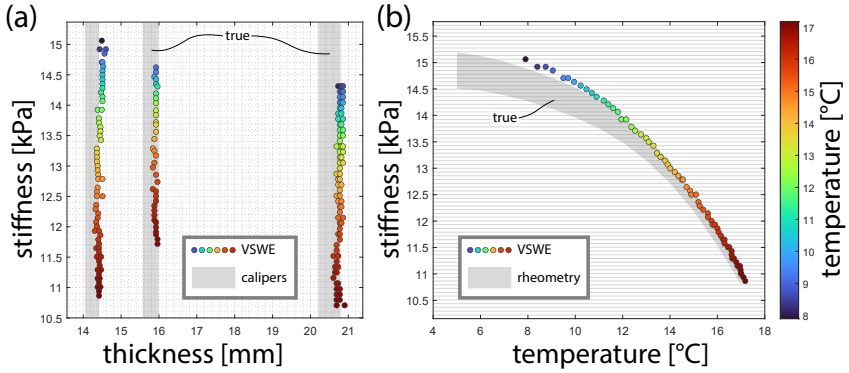


Figure 6.7: **(a) Estimated thickness and stiffness for three phantoms over a range of temperatures.** The phantoms were produced with three different volumes (1000 mL, 1100 mL, 1500 mL) of gelatin but set in the same-sized container, thus generating three different thicknesses. The thickness of each sample was measured with calipers. We produced a confidence interval of the ground-truth thickness by taking the 0.2 and 0.8 quantiles of multiple calipers measurements. The VSWE-estimated thickness consistently falls within the confidence interval. Note that while each phantom was polymerized according to the same approximate recipe, slight variations in the recipe, along with slightly different polymerization durations, led to some variation in the stiffness across samples. The temperature changes occurred as the samples spent more time at room temperature after being removed from the fridge. **(b) Estimated stiffness values for the thinnest sample over the same range of temperatures.** The ground-truth elastic modulus was measured with rheometry over a similar temperature range. Because the modulus of hydrogels can depend on both temperature and frequency [363], the rheometry measurements show a different stiffness range at each temperature. Note that the rheometry measurements were only taken between 10 Hz to 100 Hz due to instrument limitations, while the excitation signal ranged from 40 Hz to 200 Hz. This may explain some of the discrepancy between VSWE and rheometry, especially for the colder (and thus stiffer) samples which exhibit a stronger expression of higher-frequency wave dynamics. Regardless, in each case, VSWE estimates the stiffness extremely well, within 1.2% error of the rheometry range.

6.4.3 3D human leg with spatially-varying thickness

As another step towards realism, we tested VSWE on a simulated female human leg. The anatomical geometry of the leg was obtained as an STL file from the dataset of Andreassen et al. [19], who created 3D models from the National Library of Medicine's Visible Human Project [300]. We simulated the leg's response to a chirp excitation applied on the leg in COMSOL, running a full 3D physics simulation without any assumptions besides linear elasticity. For the sake of computational feasibility, we ran the simulation on the lower half of the leg, which we refer to as the calf region. To obtain dispersion relations, we considered the simulated displacements in two directions: one tangent to the leg's surface and one normal to the leg's surface.

We applied a sweeping window across the upper calf region, estimating the thickness and modulus at each window location. Figure 6.8 shows how, as the window sweeps around the leg, the inferred thickness changes. We computed the true thickness by taking the distance from the point on the skin to the nearest point on the bone. This results in a thickness distribution within the window, since the thickness changes slightly in the lengthwise direction of the leg, as well. Figure 6.8 shows the inferred thickness and the distribution of ground-truth thicknesses for each location of the sweeping window. We find that the estimated thickness roughly agrees with the true thickness distribution. There appears to be slightly more error near the edges of the simulated domain, which may be due to boundary effects. While the physics-based simulation in COMSOL is well-modeled in the interior of the domain, there may be artifacts near the boundary of the domain, making it harder to infer the correct tissue parameters near the boundary.

We also applied VSWE to three different windows in the lower calf region. Rather than perform a sweep in this region, we chose to focus on three distinct windows since there are three subregions in the lower calf area that have notably different thicknesses, due to the tissue structure changing a lot near the ankle. As Figure 6.8 shows, the inferred thickness agrees with the true thickness distribution for each subregion. Our method also recovered the constant stiffness throughout the calf, as shown in Figure 6.1.

6.4.3.1 Objective function ablation

As mentioned in Section 6.3.2, we find that SSIM performs best for the optimization objective function. This choice was crucial for obtaining acceptable results with

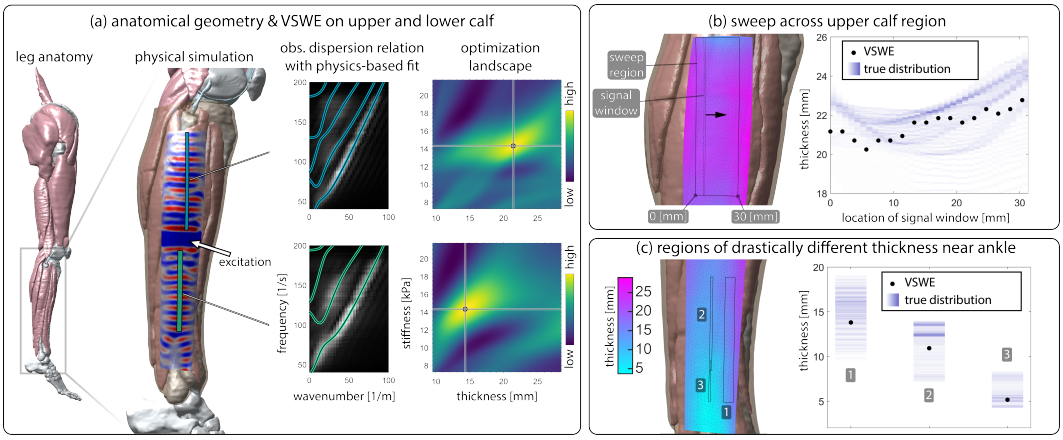


Figure 6.8: 3D anatomical inference. Panel **(a)** shows the realistic 3D leg anatomy (taken from the Visible Human Project [300]). We ran a full 3D simulation of the leg's response to a chirp excitation applied on the skin. The observed dispersion relation, along with the optimization landscape and hypothesized dispersion relation of the optimal parameters, is shown for two regions: one on the upper calf and one on the lower calf near the ankle. Panel **(b)** shows results of sweeping an observation window across a region of the upper calf. A ground-truth thickness distribution was obtained by computing the distance from the skin to the nearest point on the bone across the observation window. The VSWE-estimated thickness as the observation window slides from left to right tracks with the changing thickness of the leg. Panel **(c)** shows results for three distinct regions near the ankle. The VSWE estimations reflect the drastically different thicknesses of these regions.

the 3D leg. Figure 6.9 shows the optimization landscape for different choices of the objective function, demonstrated on both the upper calf simulation and a real video of gelatin. We compare SSIM to the following objective functions: a curve-based objective function, the negative MSE between the images, and the PSNR between the images. The curve-based loss function goes through all the curves (i.e., $\{\omega_i(\gamma)\}_{i=1}^{N} \forall \gamma \in [0, \pi/a]$) in $\mathfrak{D}(T, E)$ and integrates the values of the observed dispersion relation $\boldsymbol{D}_{\text{obs}}$ along the curves. Figure 6.9 shows that SSIM leads to the sharpest loss landscape and the most accurate estimated parameters.

6.5 Discussion

6.5.1 Practical considerations

VSWE's performance depends on many factors, including the types of waves supported by the tissue, the quality of the observation, and the quality of the physics model. One consideration is the resolution of the FEM mesh used to model physics, as the element size determines the modeling accuracy and the smallest modeled wavelength. Figure 6.10 shows that as the element size decreases, the estimated

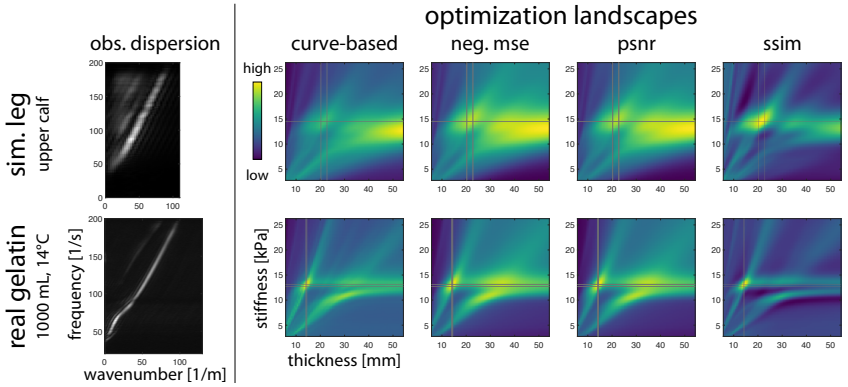


Figure 6.9: Ablation of optimization objective function. The optimization landscape is shown for various objective functions, applied in two different scenarios. Each landscape is visualized with its own color map limits, but bright yellow and dark blue indicate high (good) and low (bad) objective values, respectively. SSIM gives the sharpest optimization landscape. For the simulated upper calf, all the objective functions besides SSIM lead to the wrong optimal parameters (cross hairs indicate target parameters). Additionally, note how the simulated upper calf leads to a blurrier dispersion relation and hence blurrier optimization landscapes than the real gelatin sample, which has a simple geometry.

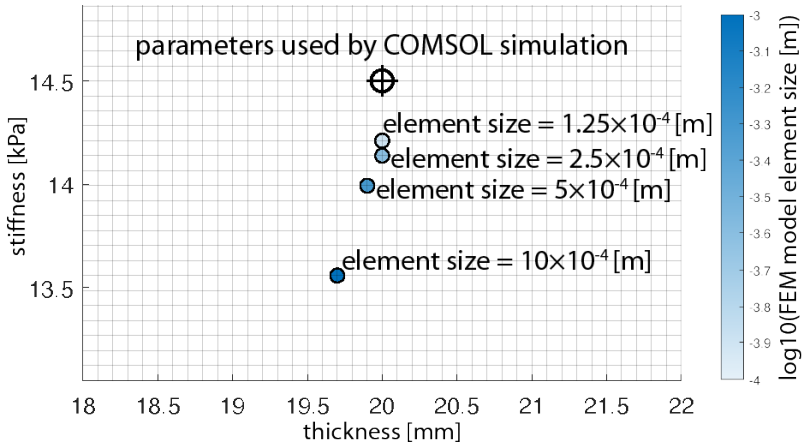


Figure 6.10: Ablation of FEM mesh resolution. As the mesh becomes finer, the VSWE-estimated parameters approach the true parameters used for the COMSOL plane-strain simulation.

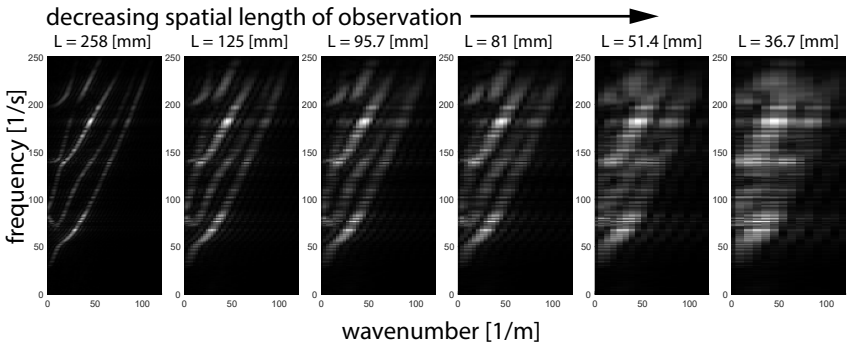


Figure 6.11: Ablation of spatial extent of observation. Keeping all else fixed, decreasing the spatial length L of the observation window degrades the quality of the observed dispersion relation. A degraded dispersion relation in turn leads to a fuzzy optimization landscape (see Figure 6.9 for an example).

parameters approach the true parameters.

Another consideration is that the spatial length L of the observation window determines the largest wavelength and the number of wavelengths that can be observed. Figure 6.11 shows how, keeping all else fixed, the FFT degrades as the observed domain shrinks, which makes inference more challenging.

6.5.2 Characteristic numbers

In our ablation studies in Figure 6.11 and Figure 6.10, we kept all other parameters of the problem fixed. By changing other parameters (e.g., increasing the image resolution, changing the range of frequencies in the excitation signal, or changing the tissue thickness), we can overcome certain limitations. This leads us to introduce *characteristic numbers*.

We can characterize the physical system and inference problem with characteristic numbers (a.k.a. dimensionless numbers or $\pi-$groups). Some characteristic numbers reflect the observability of wave modes, such as $\pi_1 := \gamma L$, $\pi_2 := \text{PPM}/\gamma$, $\pi_3 := \omega\tau$, and $\pi_4 := \text{FPS}/\omega$, where PPM (pixels per meter) is the spatial sampling rate, FPS is the temporal sampling rate, and τ is the total observation time. Increasing π_{1-4} improves VSWE's performance by improving the quality of the FFT dispersion.[2]

Other characteristic numbers reflect the numerical accuracy of the FEM model used to fit the observation, such as $\pi_5 := \frac{1}{\gamma e}$, where e is the FEM element size. Increasing π_5 improves performance by ensuring that physics is well-modeled in the

[2]Remark: ω and γ are roughly coupled through the wavespeed $c \sim \sqrt{E/\rho} \sim \omega/\gamma$ of the tissue, meaning that π_{1-2} and π_{3-4} are related via the material properties.

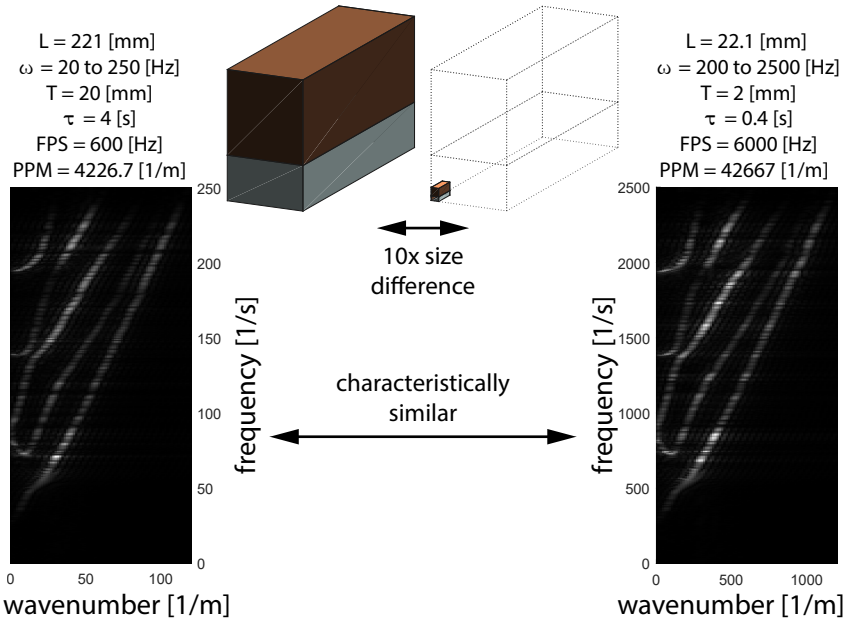


Figure 6.12: Leveraging similitude of characteristic numbers enables the application of VSWE to extremely different parameter ranges. In this simulated example, we show that by preserving the values of π_{1-6}, we see a similar observed signal from a system with parameters 10× different than the primary ranges studied in this paper. In particular, we rescaled the size of the observation window, the thickness of the tissue, the spatial sampling rate, the total observation time, the frequency range of the excitation, and the temporal sampling rate. Note that the observation window on the right is smaller than that of the most degraded FFT in Figure 6.11, yet the FFT here has not degraded.

parameter-fitting process.

Other characteristic numbers reflect the actual physics that are possible in the system. For example, $\pi_6 := \gamma T$ conveys the "shallowness" of the wave, measuring how significantly the stiff foundation affects the dynamics of a surface wave. It is beneficial for π_6 to be neither too large nor too small. If π_6 is too large, then waves behave similarly to how they would in an infinitely-thick medium, making it difficult to pinpoint the thickness. If π_6 is too small, waves above a certain wavelength will not be permitted to exist, making it necessary to capture smaller wavelengths.[3]

Understanding these characteristic numbers is important for extending the use of VSWE to scenarios and applications with vastly different parameter ranges than those studied in this paper, whether the objects are much larger or smaller, stiffer or softer, denser or lighter, or exhibit dynamics at much higher or lower wavenumbers

[3]A small π_6 explains why ocean waves break near the shore. It is also the reason that a higher-frequency excitation was helpful for the very thin tissue near the ankle in Figure 6.8.

or frequencies. In general, by preserving the values of these characteristic numbers (e.g., by increasing FPS in coordination with ω), we can preserve the performance of VSWE.

Figure 6.12 provides an example of how to take advantage of characteristic numbers. Specifically, we overcome the challenge of a smaller observed domain (as illustrated in Figure 6.11) by rescaling other parameters to preserve the relevant characteristic numbers and thus maintain the quality of the observed dispersion relation. In general, one should look to characteristic numbers as guidelines for adjusting system and/or inference parameters in order to maintain the performance of VSWE across a variety of scenarios.

6.6 Experiment Details

In this section we give details about the runtime of VSWE, the optimization objective functions that we tested, and details about the simulated experiments and real-world experiments.

6.6.1 Runtime analysis

The primary computational cost of VSWE stems from calculating dispersion relations to fit observed data. To address this efficiently, we precomputed dispersion relation datasets over grids of stiffness and thickness values suitable for soft tissue characterization. These datasets can be reused for multiple fitting scenarios, eliminating the need for repeated dispersion calculations. Although employing more efficient optimization methods than grid search would substantially reduce computational cost, we provide representative runtimes below to illustrate the expense associated with computing dispersion relations in this study.

In each dispersion dataset, there are several parameters that have bearing on the cost to compute the dataset. For the computation of a dispersion dataset containing 12 eigenvalue branches and 60 wavenumber values, on a grid of 41 stiffness values (E) $\times$ 41 (T) thickness values, the runtime for various numbers of elements corresponding to different element sizes e are tabulated below. Computations were performed on an Intel Xeon CPU E5-2663 v3 with 20 cores, and all computations used less than 64 GB of memory. While we found $e = 0.5, N_{\text{ele}} = 240$ to be sufficient for reasonable performance in our settings, using smaller e with higher N_{ele} may be desirable if modeling higher frequency wave modes.

- $e = 0.5$ mm, $N_{\text{ele}} = 240$ had runtime of 157 s
- $e = 0.25$ mm, $N_{\text{ele}} = 960$ had runtime of 470 s
- $e = 0.125$ mm, $N_{\text{ele}} = 3840$ had runtime of 3169 s

6.6.2 Optimization objective functions

Here we define the objective functions we considered when developing our approach. Recall that $\boldsymbol{D}_{\text{obs}}$ is the observed FFT-derived dispersion relation. For a given hypothesized (T, E) pair, $\mathfrak{D}(T, E)$ is the physics-based dispersion relation, and $\boldsymbol{D}_{\text{hyp}}(T, E)$ is the image version of $\mathfrak{D}(T, E)$.

Curve-based objective function We define a simple and intuitive curve-based objective function f_{curve} that assigns points for the observed dispersion relation

having high magnitude at the points that exist in the hypothesized physics-based dispersion relation $\mathfrak{D}(T, E) = \{\omega(\gamma)\}_{i=1}^{N} \forall \gamma \in [0, \pi/a]$:

$$f_{\text{curve}}\left(\mathfrak{D}(T, E), \boldsymbol{D}_{\text{obs}}\right) = \sum_{i=1}^{N} \int_{0}^{\pi/a} \boldsymbol{D}_{\text{obs}}\left(\gamma, \omega_i(\gamma)\right) \mathrm{d}\gamma. \tag{6.6}$$

Image-based objective functions To convert the dispersion relations from curve format to image format, we assign pixel values with a Gaussian kernel based on the distance from the pixel to any point on any of the curves. More precisely, the value of a pixel located at point $(\gamma_{\text{im}}, \omega_{\text{im}})$ is given by

$$v(\omega_{\text{im}}, \gamma_{\text{im}}) = \exp\left(\frac{-d_{\text{min}}^2}{2\sigma^2}\right), \tag{6.7}$$

where

$$d_{\text{min}}(\omega_{\text{im}}, \gamma_{\text{im}}) = \min_{\gamma_c, i}\left(\sqrt{(\omega_{\text{im}} - \omega_c^i(\gamma_c))^2 + (\gamma_{\text{im}} - \gamma_c)^2}\right), \tag{6.8}$$

and $\omega_c^i(\gamma_c)$ denotes the dispersion relation in curve form on any band i, and σ is a parameter that controls the width of the curve once converted to image form.

We use standard definitions of MSE, PSNR, and SSIM [327].

6.6.3 Simulated experiments

6.6.3.1 Sensitivity experiments (Figure 6.5)

The simulated experiments in Figure 6.5 were simulated on a 30 cm domain and used an observation window of 29.5 cm (the remaining 0.5 cm region was where the input excitation was applied). The simulated duration was 1 s and was equivalent to the observed duration. Displacements were sampled at 1280 positions and 600 times, equating to sampling rates of $\sim 4336\,\text{px/m}$ and 600 FPS.

6.6.3.2 Mesh resolution ablation experiments (Figure 6.10)

The simulated experiments for the mesh size ablation study in Figure 6.10 were simulated on a 60 cm domain (unrealistic, but we chose this to be large so that domain size was not a confounding factor) and used an observation window of 59.5 cm (the remaining 0.5 cm region was where the input excitation was applied). The simulated duration was 1 s and was equivalent to the observed duration. Displacements were sampled at 1280 positions and 600 times, equating to sampling rates of $\sim 2150\,\text{px/m}$ and 600 FPS.

6.6.3.3 Observation length ablation experiments (Figure 6.11)

The simulated experiments for the ablation study where the spatial length of the observed surface was ablated in Figure 6.11 were simulated on a 30 cm domain and used observation windows of varying sizes. Even though the spatial extent of the observation changed, the spatial distance between sample points was maintained constant at $\sim 4336\,\mathrm{px/m}$. The simulated duration was 1 s and was equivalent to the observed duration. Displacements were sampled at 600 FPS.

6.6.3.4 Anatomical experiments (Figure 6.8)

The operational parameters of the simulated experiments in the anatomical section were not as uniform as those from the plane-strain section, mainly due to the more complicated nature of the geometry and the fact that simulation of the thin region near the ankle required its own separate simulation due to computational cost reasons. We provide details for each scenario and for the two transient simulations.

6.6.3.5 Anatomical simulations

Here we give the details of the two anatomical simulations.

Common between the two Many of the parameters are shared between the two simulations. Both models prescribed $E = 14.5\,\mathrm{kPa}$, $\rho = 1000\,\mathrm{kg\,m^{-3}}$, and $\nu = 0.45$ everywhere in the soft tissue, and they treated the bone interface as a fixed boundary. Both models used low-reflecting boundary conditions on a segmented subsection of the calf to avoid modeling the entire domain. Both models used a quadratic tetrahedral finite element discretization, though with different element sizes. Both models used an excitation region that spanned laterally across the outer surface of the leg, with a length of 1.5 cm in the direction of wave propagation.

Simulation 1 simulated both the upper-calf and lower-calf regions to demonstrate simultaneous inferences of these significantly different regions from a single excitation. The prescribed chirp excitation was a smoothed chirp sweeping from 25 Hz to 200 Hz. This simulation used a target finite element (spatial) discretization size of 3 mm. We used a physical timestep of 800 Hz and solved for a duration of 2 s. The length of the simulated domain was about 27 cm with approximate cross-section dimensions of about $(2\,\mathrm{cm}$ to $5\,\mathrm{cm}) \times (2\,\mathrm{cm}$ to $5\,\mathrm{cm})$ with significant variance along

the length. The model contained $\sim 180,000$ elements and took 42 h 45 min to solve on an Intel Xeon CPU E5-2663 with 20 cores.

Simulation 2 specifically simulated the lower-calf region near the ankle to demonstrate inference in a region of the body where the soft tissue is much thinner and varies significantly. The prescribed chirp excitation was a smoothed chirp sweeping from 40 Hz to 400 Hz. This simulation used a target finite element (spatial) discretization size of 2 mm. We used a physical timestep of 800 Hz and solved for a duration of 0.5 s. This timestep was likely a bit too coarse for the upper end of the excitation frequency (400 Hz) in this simulation. The model used low-reflecting boundary conditions on a segmented subsection of the lower-calf to avoid modeling the entire domain. The length of the simulated domain was $\sim$ 11 cm with approximate cross-section dimensions of about (1.5 cm to 4 cm) $\times$ (1.5 cm to 4 cm) with significant variance along the length. The model contained $\sim 130,000$ elements and took 5 h to solve on an Intel Xeon CPU E5-2663 with 20 cores.

The main difference between simulations 1 and 2 is that simulation 2 modeled a thinner region of the body than simulation 1, so the waves in this region were smaller in both space and time (i.e., higher wavenumbers and frequencies). Thus finer spatial and temporal discretizations were needed, but over shorter spatial and temporal observation windows. Additionally, the excitation frequency range was approximately doubled.

6.6.3.6 Anatomical VSWE inferences

Figure 6.8(a) In this panel, two inferences are presented—one inference in the upper calf and one in the lower calf. Both used simulation 1. In Figure 6.8(a), the upper calf inference used the following parameters:

- $\sim$ 11.3 cm spatial observation window
- 452 spatial sample points
- $\sim$ 4000 px/m spatial sample rate
- 2 s temporal observation window
- 1601 temporal sample points
- 800 FP temporal sample rate

In Figure 6.8(a), the lower calf inference used the following parameters:

- $\sim$ 9.3 cm spatial observation window

- 373 spatial sample points
- $\sim$ 4000 px/m spatial sample rate
- 2 s temporal observation window
- 1601 temporal sample points
- 800 FPS temporal sample rate

Figure 6.8(b) This panel presents a series of inferences as an observation window sweeps laterally across the upper calf, making an inference at each location in the sweep. All inferences in this sweep used simulation 1. In Figure 6.8(b), the inferences in the upper calf sweep used the following parameters:

- $\sim$ 11.6 cm spatial observation window
- 463 spatial sample points
- $\sim$ 4000 px/m spatial sample rate
- 2 s temporal observation window
- 1601 temporal sample points
- 800 FPS temporal sample rate

Figure6.8(c) Three inferences are presented in three different regions near the ankle of significantly different thickness. All inferences in this panel used simulation 2. In Figure 6.8(c), the inference on region **1** used the following parameters:

- $\sim$ 9.1 cm spatial observation window
- 730 spatial sample points
- $\sim$ 8000 px/m spatial sample rate
- 0.5 s temporal observation window
- 401 temporal sample points
- 800 FPS temporal sample rate

In Figure 6.8(c), the inference on region **2** used the following parameters:

- $\sim$ 6.9 cm spatial observation window
- 549 spatial sample points
- $\sim$ 8000 px/m spatial sample rate
- 0.5 s temporal observation window
- 401 temporal sample points
- 800 FPS temporal sample rate

In Figure 6.8(c), the inference on region **3** used the following parameters:

- $\sim$ 3 cm spatial observation window
- 240 spatial sample points
- $\sim$ 8000 px/m spatial sample rate
- 0.5 s temporal observation window
- 401 temporal sample points
- 800 FPS temporal sample rate

6.6.4 Real-world experiments

6.6.4.1 Real-world VSWE gelatin experiments

Camera We used a Phantom V1610 high-speed camera. We operated it at 600 FPS and a resolution of 96×1280 pixels (cropping was later applied before inputting the videos to the VSWE pipeline). For our setup, this spatial resolution equates to a spatial sampling rate of $\sim$ 3840 px/m. Each video was taken in grayscale at a bit-depth of 16 and then downsampled to a bit-depth of 8 before entering the VSWE pipeline (for computational efficiency). Each video was originally $\sim$ 8 s long but trimmed to $\sim$ 4 s before entering the VSWE pipeline.

Geometry The gelatin was polymerized in a glass baking dish with an interior dimension of $\sim$ 30 cm in length and 19 cm in width. Waves traveled in the 30 cm direction, originating near one end of the glass baking dish and propagating lengthwise toward the other end. The excitation head attached to the shaker was $\sim$ 10 cm in width. The camera was positioned $\sim$ 1.2 m from the gelatin surface, and the viewing angle of the camera was at a $\sim 30^\circ$ angle with the surface.

Temperature measurements The gelatin was polymerized in a refrigerator $\sim 8\,^\circ$C for $\sim 24 \pm 3$ h. After being removed from the refrigerator, the gelatin was immediately textured with garlic powder and subsequently positioned in the camera's line of sight with the shaker applied to its surface. A Thermoworks Thermapen One thermometer was used to measure temperature. The thermomenter was positioned such that the probe was approximately halfway between the upper and lower surface of the gelatin, a few centimeters from the edge of the glass baking dish. As often as the hardware would allow (some amount of time is needed to write the video to disk), a video was taken. For each video, the current time and temperature were recorded.

6.6.4.2 Rheometry

Rheometry was used as the ground-truth measurement for the gelatin stiffness. Here, we give more details on the rheometry setup. A rheometer is an expensive piece of bench-top equipment that is meant to measure the stiffnesses and viscosities of soft solids and fluids. These measured values are often dependent on temperature, excitation frequency, humidity, and a variety of other factors, so high-end rheometers often have built-in systems to control these variables.

Rheometers use material samples that are shaped like disks. Between this fact and the price tag of a rheometer, they are unfit for the *in vivo* measurement tasks that VSWE aims to perform, but they are helpful for benchmarking against a well-established material measurement method in a controlled environment.

Polymerization All gelatin in this work was polymerized at a weight concentration of 4%. We polymerized our gelatin samples directly on the rheometer with gelatin taken directly from the gelatin samples prepared for the real-world VSWE gelatin experiments. The gelatin samples were taken when it was in liquid state, just before setting the remainder to polymerize in the refrigerator for the VSWE experiment. To replicate the polymerization environment as exactly as possible, we set the rheometer gelatin and the VSWE gelatin to polymerize at the same time (give or take a few minutes) and took the VSWE videos concurrently with the rheometer measurements. During polymerization, we set the rheometer stage to the same temperature as the refrigerator. This way, we could ensure that both samples had the exact same amount of time to polymerize and at the same temperature.

Stiffness measurements The rheometer measures the stiffness of the gelatin by twisting the disk-shaped sample in a shearing motion, back and forth, at a prescribed frequency. The deformation applied is small, so as not to damage the sample. In our experiments we applied 7% shear strain. This measurement gives the *shear modulus* G of the gelatin, which can then be converted to an elastic modulus E by assuming a value for the Poisson's ratio ν. We assume $\nu = 0.45$, which is reasonable for a hydrogel material like gelatin.

Sweep over frequency and temperature This stiffness measurement is repeated over a grid of frequencies and temperatures, since frequency and temperature both affect how stiffly the gelatin behaves. For each gelatin sample, the stiffness was measured over the ranges of $[10, 100]$ Hz and $[5, 17]$ °C. These ranges were chosen

to emulate the frequencies and temperatures observed in some preliminary VSWE experiments; however, we should note that the rheometer was limited to not exceed 100 Hz, whereas the VSWE videos observed surface waves up to at least $\sim$ 200 Hz. The rheometry measurements indicate that the gelatin stiffens as excitation frequency increases, though we have no ground-truth stiffness measurements from rheometry in the range of [100, 200] Hz. This mismatch between frequency ranges may explain why the VSWE inferences appear to be at the upper bound, or slightly above, the rheometry measurements.

6.6.4.3 Excitation system

To excite the surface of the gelatin, we uploaded chirp waveforms to an Agilent signal generator. The signal generator passed this signal to a B&K amplifier, which then passed the signal to a B&K shaker. Finally, the shaker mechanically applied the excitation to the surface of the gelatin.

Part III

Merging Data-driven and Physics-based Priors

Chapter 7

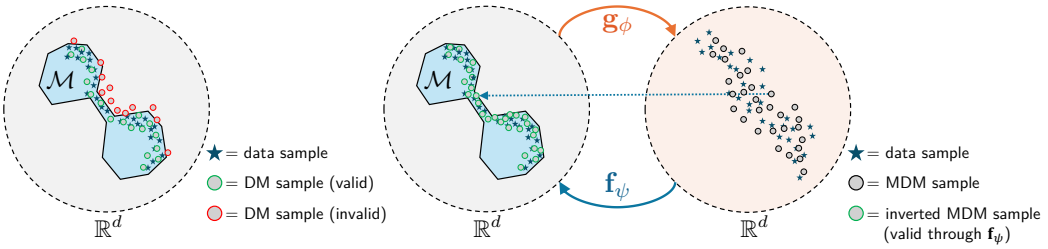
NEURAL APPROXIMATE MIRROR MAPS FOR CONSTRAINED GENERATIVE MODELS

Until this chapter, we have focused on data-driven or physics-based methods in isolation. However, many if not most scientific imaging problems would benefit from both types of priors. We now explore the intersection between data-driven and physics-based priors by looking at a method for obtaining physics-constrained generative models.

Many data distributions follow a rule that is not visually obvious. For example, videos of fluid flow obey a partial differential equation (PDE), but a human may find it difficult to discern whether a video agrees with the prescribed PDE. We can characterize such distributions as constrained distributions. Theoretically, a generative model trained on a constrained image distribution should satisfy the constraint, but in practice due to learning and sampling errors [88], it may generate visually-convincing images that break the rules. Ensuring constraint satisfaction in spite of such errors would make generative models more reliable, including as data-driven priors for solving inverse problems.

We focus on diffusion models, for which previous approaches for incorporating constraints either restrict the type of constraint or do not scale well. Equivariant [240], Riemannian [44, 165], reflected [215, 121], log-barrier [121], and mirror [211] diffusion models are all restricted to certain types of constraints, such as symmetry groups [240], Riemannian manifolds [44, 165], or convex sets [121, 211]. Generally speaking, these restrictions are in the service of guaranteeing a hard constraint, whereas a soft constraint offers flexibility and may be sufficient in many scenarios. One could, for example, introduce a guidance term to encourage constraint satisfaction during sampling [139, 24, 352], but so far there does not exist a principled framework to do so. Previous work suggested imposing a soft constraint during training [88, 312] by estimating the clean image at every noisy diffusion step and evaluating its constraint satisfaction. However, approximation [105] of the clean image from the intermediate noisy image is crude at high noise levels and thus unsuitable for constraints that are sensitive to approximation error and noise, such as a PDE constraint. Of course it is possible to penalize invalid generated samples during or after training [167], but this approach relies on computationally-expensive simulation steps. Instead, we aim for a flexible approach to impose general constraints by construction.

Our goal is to find an invertible function that maps constrained images into an unconstrained space so that a regular generative model can be trained in the unconstrained space and automatically satisfy the constraint through the inverse function.



(a) Regular diffusion model (b) Mirror diffusion model with NAMM

Figure 7.1: NAMM conceptual illustration. (a) Despite being trained on a data distribution constrained to $\mathcal{M}$, a regular diffusion model (DM) may generate samples that violate the constraint. (b) We propose to learn a neural approximate mirror map (NAMM) that entails a forward map $\boldsymbol{g}_\phi$ and inverse map $\boldsymbol{f}_\psi$. The forward map transforms the constrained space into an unconstrained ("mirror") space. Once $\boldsymbol{g}_\phi$ and $\boldsymbol{f}_\psi$ are learned, a mirror diffusion model (MDM) can be trained on the pushforward of the data distribution through $\boldsymbol{g}_\phi$ and its samples mapped back to the constrained space through $\boldsymbol{f}_\psi$.

We propose *neural approximate mirror maps* (NAMMs), which bring the flexibility of soft constraints into the principled framework of mirror diffusion models [211]. A mirror diffusion model (MDM) allows for training a completely unconstrained diffusion model in a "mirror" space defined by a mirror map. Unconstrained samples from the diffusion model are mapped back to the constrained space via an inverse mirror map. However, invertible mirror maps are challenging or impossible to derive in closed form for general constraints. We address this by jointly optimizing two networks to approximate a mirror map and its inverse.

A NAMM encompasses a (forward) mirror map $\boldsymbol{g}_\phi$ and its approximate inverse map $\boldsymbol{f}_\psi$. They are trained so that $\boldsymbol{f}_\psi \approx \boldsymbol{g}_\phi^{-1}$, and $\boldsymbol{f}_\psi$ maps unconstrained points to the constrained space (see Figure 7.1 for a conceptual illustration). Our method works for any constraint that has a differentiable function to quantify the distance from an image to the constraint set. We parameterize $\boldsymbol{g}_\phi$ as the gradient of a strongly input-convex neural network (ICNN) [17] to satisfy invertibility. We train the NAMM with a cycle-consistency loss [359] to ensure $\boldsymbol{f}_\psi \approx \boldsymbol{g}_\phi{}^{-1}$ and train the inverse map with a constraint loss to ensure $\boldsymbol{f}_\psi(\tilde{\mathbf{x}})$ is close to the constraint set for all $\tilde{\mathbf{x}}$ that we are interested in (we define this formally in Section 7.2.1). An MDM can be trained on the pushforward of the data distribution through $\boldsymbol{g}_\phi$, and its generated samples can be mapped to the constraint set via $\boldsymbol{f}_\psi$. Although not inherently restricted to diffusion models, our approach maintains the many advantages of diffusion models, including expressive generation, simulation-free training [283], and tractable computation of probability densities [211]. One can also adapt existing PnPDP algorithms for the

mirror space and enforce constraints with the inverse map.

This chapter presents a promising avenue towards incorporating physics constraints into data-driven priors in a flexible and principled manner. In Section 7.1, we review background and related work on constrained generative models. We then describe our proposed method in Section 7.2. We present our experimental results in Section 7.3, showing improved constraint satisfaction for various physics-based, geometric, and semantic constraints. We also discuss ablation studies and adapt a popular diffusion-based inverse solver to solve constrained inverse problems, in particular data assimilation with PDE constraints. Sections 7.4 and 7.5 provide details about our experiments and demonstrated constraints, respectively.

7.1 Background

7.1.1 Constrained generative models

Explicitly incorporating a known constraint into a generative model poses benefits such as data efficiency [128, 26], generalization capabilities [192], and feasibility of samples [137]. Some methods leverage equivariant neural networks [266, 133, 301] for symmetry [15, 240, 190, 285, 49, 256, 265, 192, 227, 97, 338, 161, 341] but do not generalize to other types of constraints or generative models [190, 285, 49, 256, 265, 192, 227, 240, 338, 161, 341, 97, 15]. Previous methods for constrained diffusion models [44, 165, 121, 215, 211] make strong assumptions about the constraint, such as being characterized as a Riemannian manifold [44, 165], having a well-defined reflection operator [215] or projection operator [75], or corresponding to a convex constraint set [121, 211]. Fishman et al. [122] proposed a diffusion model that incorporates Metropolis-Hastings steps to work with general constraints, but impractically high rejection rates may occur with constraints that are challenging to satisfy, such as a PDE constraint.

An alternative approach is to introduce a soft constraint penalty when training the generative model [128, 221, 68, 137, 88, 312]. However, evaluating the constraint loss of generated samples during training may be prohibitively expensive. Instead, one could add constraint-violating training examples [137], but it is often difficult to procure useful negative examples. In contrast, our approach does not alter the training objective of the generative model.

7.1.2 Mirror maps

For any convex constraint set $C \subseteq \mathbb{R}^d$, one can define a *mirror map* that maps from C to $\mathbb{R}^d$. This is done by defining a *mirror potential* $\phi : C \to \mathbb{R}$ that is continuously-differentiable and strongly-convex [56, 297]. The mirror map is the gradient $\nabla\phi : C \to \mathbb{R}^d$ [211]. Every mirror map has an inverse $(\nabla\phi)^{-1} : \mathbb{R}^d \to C$, which, unlike the forward mirror map, is not necessarily the gradient of a strongly-convex function [357, 297]. Mirror maps have been used for constrained optimization [29] and sampling [163, 205, 211]. Although true mirror maps exist only for convex constraints, we seek to generalize the concept to learn approximate mirror maps to handle non-convex constraints. Recent work suggested learned mirror maps for convex optimization [297] and reinforcement learning [14] but did not tackle constrained generative modeling. Our work proposes a novel training objective for learning mirror maps with the goal of constraining generative models to satisfy arbitrary constraints.

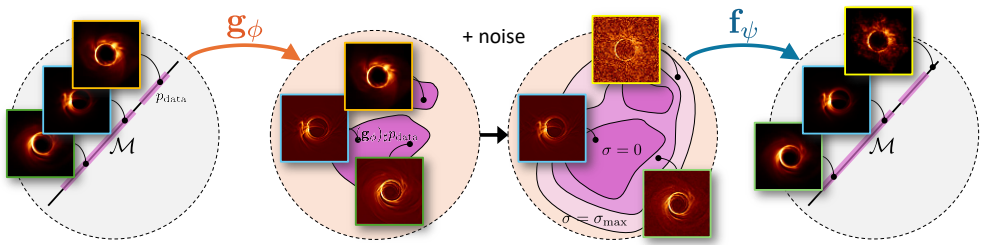


Figure 7.2: NAMM training illustration. Given data that lie on a constraint manifold $\mathcal{M}$ (e.g., the hyperplane of images with the same total brightness), we jointly train an approximate mirror map $\boldsymbol{g}_\phi$ and its approximate inverse $\boldsymbol{f}_\psi$. After mapping data $\boldsymbol{x} \sim p_{\text{data}}$ to the mirror space as $\boldsymbol{g}_\phi(\boldsymbol{x})$, we perturb them with additive Gaussian noise whose standard deviation can be anywhere between 0 and $\sigma_{\max}$. The inverse map $\boldsymbol{f}_\psi$ is trained to map these perturbed samples back onto $\mathcal{M}$.

7.2 Method

We now describe our method. We focus on diffusion models, but any generative model can be trained in the learned mirror space (see Section 7.3.4 for results with a VAE). We denote the constrained image distribution by p_{data} and the (not necessarily convex) constraint set by $\mathcal{M} \subseteq \mathbb{R}^d$. Images in the constrained and mirror spaces are denoted by $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$, respectively. The pushforward of the data distribution p_{data} through a mirror map $\boldsymbol{g}_\phi$ is denoted by $(\boldsymbol{g}_\phi)_\sharp p_{\text{data}}$.

7.2.1 Learning the forward and inverse mirror maps

Let $\boldsymbol{g}_\phi$ and $\boldsymbol{f}_\psi$ be the neural networks modeling the forward and inverse mirror maps, respectively, where ϕ and ψ are their parameters. We formulate the following learning problem:

$$\phi^*, \psi^* = \arg\min_{\phi,\psi} \left\{ \mathcal{L}_{\text{cycle}}(\boldsymbol{g}_\phi, \boldsymbol{f}_\psi) + \lambda_{\text{constr}} \mathcal{L}_{\text{constr}}(\boldsymbol{g}_\phi, \boldsymbol{f}_\psi) + \lambda_{\text{reg}} \mathcal{R}(\boldsymbol{g}_\phi) \right\}, \quad (7.1)$$

where $\mathcal{L}_{\text{cycle}}$ encourages $\boldsymbol{g}_\phi$ and $\boldsymbol{f}_\psi$ to be inverses of each other, $\mathcal{L}_{\text{constr}}$ encourages $\boldsymbol{f}_\psi$ to map unconstrained points back to the constraint set, and $\mathcal{R}$ is a regularization term to ensure a unique solution for the maps. Here $\lambda_{\text{constr}}, \lambda_{\text{reg}} \in \mathbb{R}_{>0}$ are scalar hyperparameters.

A true inverse mirror map satisfies cycle consistency and constraint satisfaction on all of $\mathbb{R}^d$, so ideally $\boldsymbol{f}_\psi(\tilde{\boldsymbol{x}}) = \boldsymbol{g}_\phi^{-1}(\tilde{\boldsymbol{x}})$ and $\boldsymbol{f}_\psi(\tilde{\boldsymbol{x}}) \in \mathcal{M}$ for all $\tilde{\boldsymbol{x}} \in \mathbb{R}^d$. But since it would be computationally infeasible to optimize $\boldsymbol{f}_\psi$ over all possible points in $\mathbb{R}^d$, we instead optimize it over distributions that we would expect the inverse map to face in practice in the context of generative models. That is, we only need $\boldsymbol{f}_\psi$ to be valid for samples from an MDM trained on $(\boldsymbol{g}_\phi)_\sharp p_{\text{data}}$, which we refer

to as the *mirror distribution*. To make $\boldsymbol{f}_\psi$ robust to learning/sampling error of the MDM, we consider a sequence of noisy distributions in the mirror space, each corresponding to adding Gaussian noise to samples from $(\boldsymbol{g}_\phi)_\sharp p_{\text{data}}$, which, for a maximum perturbation level $\sigma_{\max}$, we denote by

$$\left((\boldsymbol{g}_\phi)_\sharp p_{\text{data}} * \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})\right)_{\sigma \in [0, \sigma_{\max}]}. \tag{7.2}$$

We train $\boldsymbol{f}_\psi$ to be a valid inverse mirror map only for points from these noisy mirror distributions. Since we do not know *a priori* how much noise the MDM samples will contain, we consider all possible noise levels up to $\sigma_{\max}$ for robustness.

We define a **cycle-consistency loss** [359] that covers the forward and inverse directions and evaluates the inverse direction for the entire sequence of distributions defined in Equation 7.2:

$$\begin{aligned}\mathcal{L}_{\text{cycle}}(\boldsymbol{g}_\phi, \boldsymbol{f}_\psi) := \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}} \Bigg[& \left\| \boldsymbol{x} - \boldsymbol{f}_\psi \left(\boldsymbol{g}_\phi(\boldsymbol{x}) \right) \right\|_1 \\ & + \int_0^{\sigma_{\max}} \mathbb{E}_{\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})} \left[\left\| \boldsymbol{g}_\phi(\boldsymbol{x}) + \sigma \boldsymbol{z} - \boldsymbol{g}_\phi \left(\boldsymbol{f}_\psi(\boldsymbol{g}_\phi(\boldsymbol{x}) + \sigma \boldsymbol{z}) \right) \right\|_1 \right] \Bigg]. \end{aligned} \tag{7.3}$$

Let $\ell_{\text{constr}} : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be a differentiable *constraint distance* that measures the distance from an input image to the constraint set. We define the following **constraint loss** to encourage $\boldsymbol{f}_\psi$ to map points from the noisy mirror distributions (Equation 7.2) to the constraint set:

$$\mathcal{L}_{\text{constr}}(\boldsymbol{g}_\phi, \boldsymbol{f}_\psi) := \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}} \left[\int_0^{\sigma_{\max}} \mathbb{E}_{\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})} \left[\ell_{\text{constr}} \left(\boldsymbol{f}_\psi(\boldsymbol{g}_\phi(\boldsymbol{x}) + \sigma \boldsymbol{z}) \right) \right] \mathrm{d}\sigma \right]. \tag{7.4}$$

To ensure a unique solution, we **regularize** $\boldsymbol{g}_\phi$ to be close to the identity function:

$$\mathcal{R}(\boldsymbol{g}_\phi) := \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}} \left[\left\| \boldsymbol{x} - \boldsymbol{g}_\phi(\boldsymbol{x}) \right\|_1 \right]. \tag{7.5}$$

We use Monte-Carlo to approximate the expectations in the objective over the noisy mirror distributions with $\sigma \sim \mathcal{U}([0, \sigma_{\max}])$ and solve Equation 7.1 with SGD.

Architecture We parameterize $\boldsymbol{g}_\phi$ as the gradient of an ICNN following the implementation of Tan et al. [297]. For convex constraints, this satisfies the theoretical requirement that $\boldsymbol{g}_\phi$ be the gradient of a strongly-convex function. Even for non-convex constraints, this choice brings practical benefits, as we discuss in Section 7.3.3. We note that $\boldsymbol{g}_\phi$ is not a true mirror map since $\mathcal{M}$ is not assumed to be convex, and $\boldsymbol{g}_\phi$ is defined on all of $\mathbb{R}^d$ instead of just on $\mathcal{M}$. We parameterize $\boldsymbol{f}_\psi$ as a ResNet-based CNN similar to the one used in CycleGAN [359].

7.2.2 Learning the mirror diffusion model

Similarly to Liu et al. [211], we train an MDM on the mirror distribution $(\boldsymbol{g}_\phi)_\sharp p_{\text{data}}$ and map its samples to the constrained space through $\boldsymbol{f}_\psi$. In particular, we train a score model $\boldsymbol{s}_{\boldsymbol{\theta}}$ with the following denoising score matching objective in the learned mirror space (defined as the range of $\boldsymbol{g}_\phi$):

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_t \Bigg\{ \lambda(t)\, \mathbb{E}_{\tilde{\boldsymbol{x}}(0)\sim(\boldsymbol{g}_\phi)_\sharp p_{\text{data}}}\, \mathbb{E}_{\tilde{\boldsymbol{x}}(t)|\tilde{\boldsymbol{x}}(0)} \Big[\Big\| \tilde{\boldsymbol{s}}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}}(t), t) - \nabla_{\tilde{\boldsymbol{x}}(t)} \log p_{0t}(\tilde{\boldsymbol{x}}(t) \mid \tilde{\boldsymbol{x}}(0)) \Big\|_2^2 \Big] \Bigg\}, \tag{7.6}$$

where $\tilde{\boldsymbol{x}}(0) \sim (\boldsymbol{g}_\phi)_\sharp p_{\text{data}}$ is obtained as $\tilde{\boldsymbol{x}}(0) := \boldsymbol{g}_\phi(\boldsymbol{x}(0))$ for $\boldsymbol{x}(0) \sim p_{\text{data}}$. Here p_{0t} denotes the transition kernel from $\tilde{\boldsymbol{x}}(0)$ to $\tilde{\boldsymbol{x}}(t)$ under the diffusion SDE, and $\lambda(t) \in \mathbb{R}_{>0}$ is a time-dependent weight. To sample new images, we sample $\tilde{\boldsymbol{x}}(T) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, run reverse diffusion in the mirror space, and map the resulting $\tilde{\boldsymbol{x}}(0)$ to the constrained space via $\boldsymbol{f}_\psi$.

7.2.3 Finetuning the inverse mirror map

The inverse map $\boldsymbol{f}_\psi$ is trained with samples from the noisy mirror distributions in Equation 7.2, but we ultimately wish to evaluate $\boldsymbol{f}_\psi$ with samples from the MDM. To reduce the distribution shift, it may be helpful to finetune $\boldsymbol{f}_\psi$ with MDM samples. We generate a training dataset of samples $\tilde{\boldsymbol{x}}$ from the MDM and then finetune the inverse map to deal with such samples specifically. In the following finetuning objective, we replace $\tilde{\boldsymbol{x}} \sim (\boldsymbol{g}_\phi)_\sharp p_{\text{data}}$ with $\tilde{\boldsymbol{x}} \sim p_{\boldsymbol{\theta}}$, where $p_{\boldsymbol{\theta}}$ is the distribution of MDM samples in the mirror space:

$$\begin{aligned} \psi^* = \arg\min_{\psi} \Bigg\{ & \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}} \left\| \boldsymbol{x} - \boldsymbol{f}_\psi(\boldsymbol{g}_\phi(\boldsymbol{x})) \right\|_1 \\ & + \mathbb{E}_{\tilde{\boldsymbol{x}}\sim p_{\boldsymbol{\theta}}} \Bigg[\int_0^{\sigma_{\max}} \mathbb{E}_{\boldsymbol{z}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})} \left[\left\| \tilde{\boldsymbol{x}} + \sigma\boldsymbol{z} - \boldsymbol{g}_\phi(\boldsymbol{f}_\psi(\tilde{\boldsymbol{x}} + \sigma\boldsymbol{z})) \right\|_1 + \lambda_{\text{constr}} \ell_{\text{constr}} \left(\boldsymbol{f}_\psi(\tilde{\boldsymbol{x}} + \sigma\boldsymbol{z}) \right) \right] \mathrm{d}\sigma \Bigg] \Bigg\}. \end{aligned} \tag{7.7}$$

Finetuning essentially tailors $\boldsymbol{f}_\psi$ to the MDM. The original objective assumes that the MDM will sample Gaussian-perturbed images from the mirror distribution, but in reality it samples from a slightly different distribution. As the ablation study in Section 7.3.3 shows, finetuning is not an essential component of the method; we suggest it as an optional step for when it is critical to optimize the constraint distance.

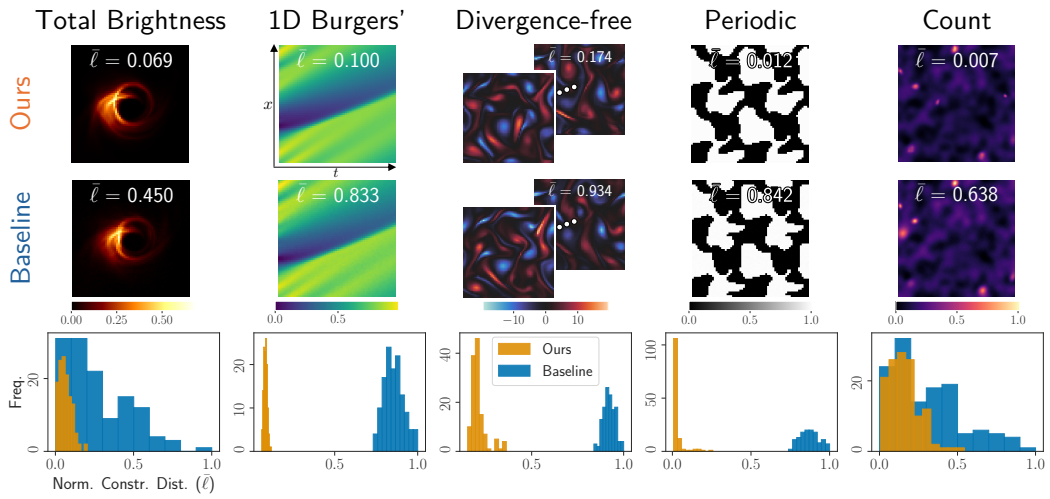


Figure 7.3: Improved constraint satisfaction with NAMMs. Samples from our approach are nearly indistinguishable from baseline samples, yet there is a significant difference in their distances from the constraint set. The baseline is a DM trained on the original constrained dataset. Our approach is to train a NAMM and then an MDM in the mirror space induced by $\boldsymbol{g}_\phi$. Samples are obtained by sampling from the MDM and then passing samples through $\boldsymbol{f}_\psi$. The histograms show normalized constraint distances $\bar{\ell}$ of 128 samples (normalized so that each constraint has a maximum of 1 across the samples from both methods). Our results are from the finetuned NAMM. For each constraint, we made sure that the DM was trained for at least as long as the NAMM, MDM, and finetuned NAMM combined.

7.3 Results

We present experiments with constraints ranging from physics-based to semantic. For the considered examples, our method achieves from 38% to as much as 96% improvement in constraint satisfaction upon a vanilla DM trained on the same data (see Table 7.1). The following paragraphs introduce the demonstrated constraints ℓ. For each we consider an image dataset for which the constraint is physically meaningful.

Total brightness In astronomical imaging, even if a source's structure is unknown *a priori*, its total brightness, or total flux, is often well constrained [7]. We define $\ell_{\text{flux}}(\boldsymbol{x})$ as the absolute difference between $\sum_{i=1}^{d} \boldsymbol{x}_i$ and the true total brightness. We demonstrate with a dataset of 64×64 images of black-hole simulations [329] whose pixel values sum to 120. While this constraint is a simple warmup example, generic diffusion models perform surprisingly poorly on it.

1D Burgers' We consider Burgers' equation [25, 58] for a 1D viscous fluid, representing the discretized solution as an $n_x \times n_t$ image $\boldsymbol{x}$, where n_x and n_t are the numbers of grid points in space and time, respectively. The distance $\ell_{\text{burgers}}(\boldsymbol{x})$ compares each 1D state in the image to the PDE solver's output given the previous state (based on Crank-Nicolson time-discretization [86, 183]). The dataset consists of 64×64 images of Crank-Nicolson solutions with Gaussian random fields as initial conditions.

Divergence-free A time-dependent 2D velocity field $\boldsymbol{u} = \boldsymbol{u}(t, x, y)$ is called *divergence-free* or *incompressible* if $\nabla \cdot \boldsymbol{u} = 0$. We define the constraint distance ℓ_{div} as the ℓ^1-norm of the divergence and demonstrate this constraint with 2D Kolmogorov flows [67, 42, 261]. We represent the trajectory of the 2D velocity, discretized in space-time, as a two-channel (for both velocity components) image $\boldsymbol{x}$ with the states appended sequentially. We used `jax-cfd` [191] to generate trajectories of eight 64×64 states and appended them in a 2×4 pattern to create 128×256 images.

Periodic We consider images $\boldsymbol{x}$ that are periodic tilings of a unit cell. This type of symmetry appears in materials science, such as when constructing metamaterials out of unit cells [241]. We use a distance function ℓ_{periodic} that compares all pairs of tiles in the image and computes the average ℓ^1-norm of their differences. We created a dataset of 64×64 images (composed of 32×32 unit cells tiled in a 2×2 fashion) using code from Ogren et al. [241].

Count Generative models can sometimes generate an incorrect number of objects [243]. We formulate a differentiable count constraint by relying on a CNN to estimate the count of a particular object in an image $\boldsymbol{x}$. Note that using a neural network leads to a non-analytical and highly non-convex constraint. Letting $f_{\text{CNN}} : \mathbb{R}^d \to \mathbb{R}$ be the trained counting CNN, we use the distance function $\ell_{\text{count}}(\boldsymbol{x}) := |f_{\text{CNN}}(\boldsymbol{x}) - \bar{c}|$ for a target count $\bar{c}$. The dataset consists of 128×128 simulated images of exactly eight (8) radio galaxies with background noise [84].

7.3.1 Improved constraint satisfaction and training efficiency

First and foremost, we verify that our approach leads to better constraint satisfaction than a vanilla diffusion model (DM). We evaluate constraint satisfaction by computing the average constraint distance of generated samples. Since the con-

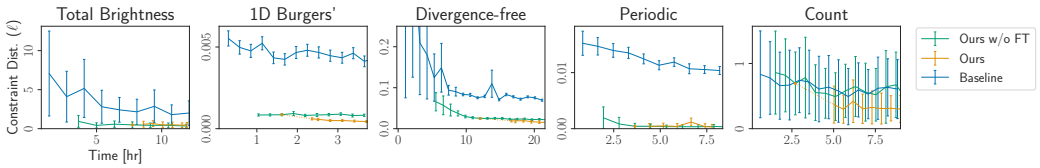
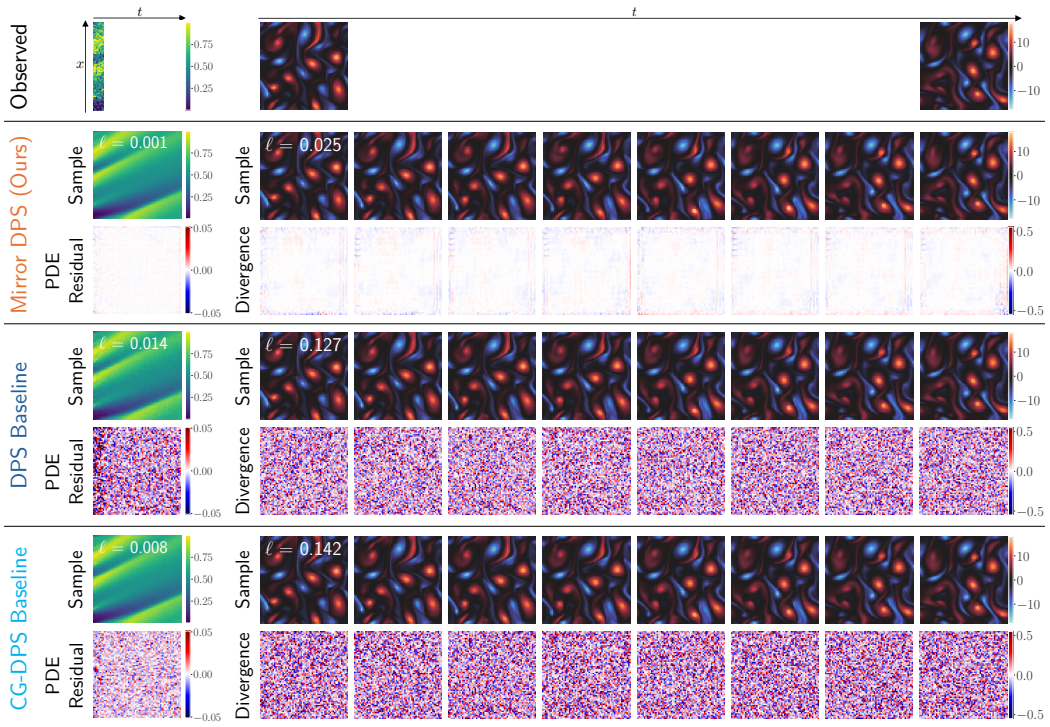


Figure 7.4: NAMM training efficiency. For each method, we clocked the total compute time during training (ignoring validation and I/O operations) and here plot the mean ± std. dev. of the constraint distances ℓ of 128 generated samples at each checkpoint. The MDM training curve ("Ours w/o FT") is offset by the time it took to train the NAMM. The finetuning curve ("Ours") is offset by the time it took to train the NAMM and MDM and generate finetuning data. For most constraints, the DM has consistently higher constraint distance without any sign of converging to the same performance as that of the MDM. For the count constraint, the MDM performs on par with the DM, but finetuning noticeably accelerates constraint satisfaction. Each run was done on the same hardware (4× A100 GPUs).

straint distance is non-negative, an average constraint distance of 0 implies that the constraint is satisfied almost surely.

For each constraint, we trained a NAMM on the corresponding dataset and then trained an MDM on the pushforward of the dataset through the learned $\boldsymbol{g}_\phi$. We show results from a finetuned NAMM, but as shown in Section 7.3.3, finetuning is often not necessary. The baseline DM was trained on the original dataset. Figure 7.3 highlights that MDM samples inverted through $\boldsymbol{f}_\psi$ are much closer to the constraint set than DM samples despite being visually indistinguishable. For the total brightness, 1D Burgers', divergence-free, and periodic constraints, there is a significant gap between our distribution of constraint distances and the baseline's. The gap is smaller for the count constraint, which may be due to difficulties in identifying and learning the mirror map for a highly non-convex constraint.

Furthermore, our approach achieves better constraint satisfaction in less training time. In Figure 7.4, we plot constraint satisfaction as a function of compute time, comparing our approach (with and without finetuning) to the DM. Accounting for the time it takes to train the NAMM, our MDM achieves much lower constraint distances than the DM for the three physics-based constraints and the periodic constraint, often reaching a level that the DM struggles to achieve. For the count constraint, we find that finetuning is essential for improving constraint satisfaction, and it is more time-efficient to finetune the inverse map than to continue training the MDM.



(a) 1D Burgers' (b) Kolmogorov flow (divergence-free)

Figure 7.5: Data assimilation. We used the same finetuned NAMM, MDM, and DM checkpoints as in Figure 7.3. (a) Given noisy observations of the first eight states, we sampled possible full trajectories of a 1D Burgers' system. Our solutions have smaller deviation from the PDE than samples obtained with DPS, even those of constraint-guided DPS (CG-DPS). (b) The task is to infer the full Kolmogorov flow from noisy observations of the first and last states. Our solution has significantly less divergence.

7.3.2 Solving constrained inverse problems with mirror DPS

As discussed in Section 2.2, many PnPDP methods have been proposed to use a pretrained diffusion model to sample images from the posterior distribution $p(\boldsymbol{x} \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{x})p(\boldsymbol{x})$, given measurements $\boldsymbol{y} \in \mathbb{R}^m$ and a diffusion-model prior $p(\boldsymbol{x})$. Most PnPDP methods can be adapted to the mirror space by evaluating the measurement likelihood on inverted mirror images $\boldsymbol{f}_\psi(\tilde{\boldsymbol{x}})$ instead of on images $\boldsymbol{x}$ in the original space. We demonstrate this approach with DPS [79].

In particular, we apply mirror DPS to data assimilation, an inverse problem that aims to recover the hidden state of a dynamical system given imperfect observations of the state. In Figure 7.5, we show results for data assimilation of a 1D Burgers' system and a divergence-free Kolmogorov flow given a few noisy state observations, which can be essentially formulated as a denoise-and-inpaint problem. For each case,

we used mirror DPS with the corresponding NAMM-based MDM. We include two baselines: (1) vanilla DPS with the DM and (2) constraint-guided DPS (CG-DPS) with the DM. The latter incorporates the constraint distance as an additional likelihood term. As Figure 7.5 shows, our approach leads to notably less constraint violation (i.e., less deviation from the PDE or less divergence) than both baselines.

7.3.3 Ablation studies

Finetuning Table 7.1 shows the improvement in constraint satisfaction after finetuning $\boldsymbol{f}_\psi$ while also verifying that the generated distribution stays close to the true data distribution. We use maximum mean discrepancy (MMD) [141] and Kernel Inception Distance (KID) [38] as measures of distance between distributions. MMD evaluates distance in a feature space defined by a Gaussian kernel, and KID uses Inception v3 [295] features. Finetuning does not notably change the distribution-matching accuracy of the MDM and in some cases improves it while improving constraint satisfaction. Compared to a vanilla DM, our approach before and after finetuning does not lead to significantly different MMD and even gives better KID while significantly improving constraint distance.

		Total Brightness	1D Burgers'	Divergence-free	Periodic	Count
Ours	CD (↓)	**0.49** ± 0.35	**0.04** ± 0.01	**1.57** ± 0.31	**0.04** ± 0.06	**0.35** ± 0.27
	MMD (↓)	0.1023 ± 0.0131	0.1291 ± 0.0096	0.0781 ± 0.0023	0.0758 ± 0.0058	0.1978 ± 0.0062
	KID (↓)	**0.0027** ± 0.0008	**0.0022** ± 0.0004	0.0058 ± 0.0006	**0.0014** ± 0.0005	**0.0570** ± 0.0014
Baseline	CD (↓)	2.20 ± 1.61	0.45 ± 0.03	6.96 ± 0.27	1.04 ± 0.08	0.56 ± 0.44
	MMD (↓)	**0.0956** ± 0.0099	**0.0621** ± 0.0091	**0.0595** ± 0.0026	**0.0533** ± 0.0043	**0.1276** ± 0.0066
	KID (↓)	0.0462 ± 0.0016	0.2308 ± 0.0028	**0.0053** ± 0.0007	0.0064 ± 0.0004	0.1084 ± 0.0015

Table 7.1: Effect of NAMM finetuning. Constr. dist. (CD) = $100\lambda_{\text{constr}}\ell$. The improvements in mean CD are (left to right, comparing "Ours" to "Baseline"): 78%, 91%, 77%, 96%, 38% for five problems. For all metrics, mean ± std. dev. is estimated with 10000 samples. In terms of MMD/KID, finetuning does not significantly impact distribution-matching accuracy but improves constraint distance. DM baseline results are shown for comparison. According to MMD, the baseline gives better distribution-matching accuracy; according to KID, our approach captures the true data distribution better.

Constraint loss There are two hyperparameters for the constraint loss in Equation 7.1: $\sigma_{\max}$ determines how much noise to add to samples from the mirror distribution, and λ_{constr} is the weight of the constraint loss. Intuitively, a higher $\sigma_{\max}$ means that the inverse map $\boldsymbol{f}_\psi$ must map larger regions of $\mathbb{R}^d$ back to the constraint set, making its learning objective more challenging. We would expect $\boldsymbol{g}_\phi$ to cooperate by maintaining a reasonable SNR in the noisy mirror distributions. Figure 7.6 shows

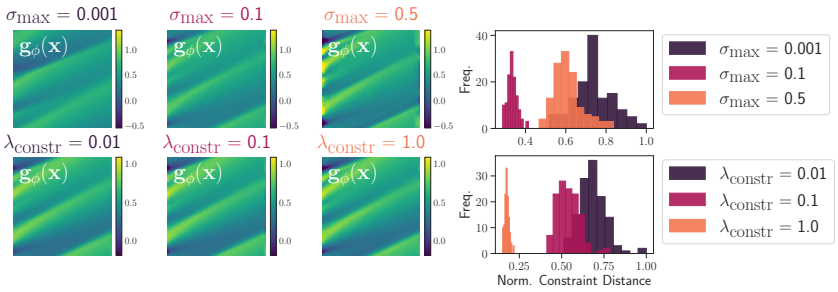


Figure 7.6: Effect of $\sigma_{\max}$ and λ_{constr}, demonstrated with 1D Burgers'. **First row:** as $\sigma_{\max}$ increases (keeping $\lambda_{\text{constr}} = 1.0$), the mirror image $\boldsymbol{g}_\phi(\boldsymbol{x})$ for $\boldsymbol{x} \sim p_{\text{data}}$ increases in magnitude to maintain a similar SNR. Histograms show constraint distances of 128 inverted MDM samples, normalized to have a maximum of 1 across samples from all three settings. Decreasing $\sigma_{\max}$ from 0.5 to 0.1 improves the constraint distances, but further lowering $\sigma_{\max}$ to 0.001 causes them to go back up. This indicates a tradeoff between robustness and performance of $\boldsymbol{f}_\psi$. **Second row:** as λ_{constr} increases (keeping $\sigma_{\max} = 0.1$), $\boldsymbol{g}_\phi(\boldsymbol{x})$ does not change as much as when increasing $\sigma_{\max}$, but the constraint distances decrease (with a tradeoff in cycle consistency). For all three settings, the same number of NAMM and MDM epochs was used as in Figure 7.3 but without finetuning.

how increasing $\sigma_{\max}$ causes $\boldsymbol{g}_\phi(\boldsymbol{x})$ for $\boldsymbol{x} \sim p_{\text{data}}$ to have larger magnitudes so that the added noise will not hide the signal. However, setting $\sigma_{\max}$ too high can worsen constraint satisfaction, perhaps due to the challenge of mapping a larger region of $\mathbb{R}^d$ back to the constraint set. On the flip side, setting $\sigma_{\max}$ too low can worsen constraint satisfaction because of poor robustness of $\boldsymbol{f}_\psi$. Meanwhile, increasing λ_{constr} for the same $\sigma_{\max}$ leads to lower constraint distance, although there is a tradeoff between constraint distance and cycle-consistency inherent in the NAMM objective (Equation 7.1).

Mirror map parameterization Figure 7.7 compares parameterizing $\boldsymbol{g}_\phi$ as the gradient of an ICNN versus as a ResNet-based CNN. We demonstrate how the mirror space changes when parameterizing the forward map as a ResNet-based CNN. The mirror space becomes less regularized, leading to worse constraint satisfaction of the MDM, perhaps because the MDM struggles to learn a less-regularized mirror space. Thus, even when the constraint is non-convex and there are no theoretical reasons to use an ICNN, it may still be practically favorable.

7.3.4 NAMMs for constrained VAEs

Our approach is compatible with any generative model, not just diffusion models. Once a NAMM is trained, any generative model can be trained in the learned mirror

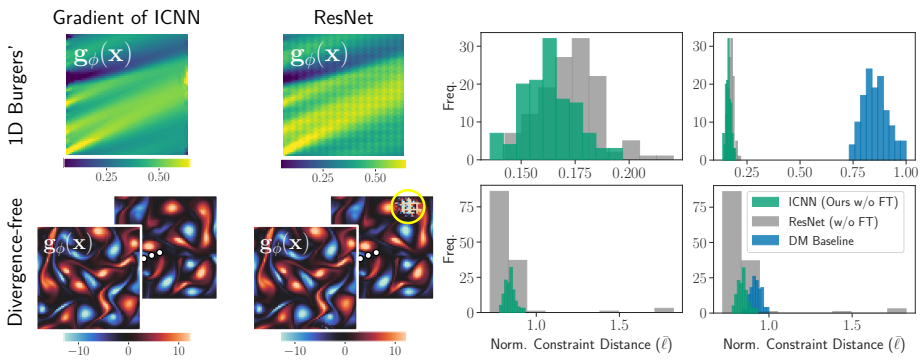


Figure 7.7: Architecture of $\boldsymbol{g}_\phi$: gradient of ICNN vs. ResNet-based CNN. Both approaches preserve visual structure in the mirror space, but the ResNet causes irregularities, such as the patch circled in yellow. The histograms show the normalized constraint distances $\bar{\ell}$ of 128 inverted MDM samples (without finetuning). DM histograms from Figure 7.3 are shown for comparison. An ICNN leads to better constraint satisfaction with fewer outliers. We trained the NAMM and MDM for the same number of epochs as in Figure 7.3 without finetuning. We found that even with finetuning, a ResNet-based forward map leads to worse constraint satisfaction or noticeably worse visual quality of generated samples.

space and its samples mapped back to the constrained space via the learned inverse mirror map. In this experiment, we apply our approach to training a VAE that satisfies the divergence-free constraint, comparing a VAE trained in the learned mirror space ("MVAE") to a VAE trained in the original data space.

We trained a VAE in the mirror space induced by the learned mirror map that was trained for the divergence-free constraint (without finetuning). We call this the MVAE approach. As a baseline, we trained the same VAE architecture on the original divergence-free data without transformation. We note that the same data are used to train both the MVAE and VAE; the only difference is that the MVAE is trained in the mirror space, while the VAE is trained in the original space. The training procedure was otherwise the same for both the MVAE and VAE.

Figure 7.8 shows samples from the MVAE and VAE. We ensured that the total training time of the MVAE did not exceed that of the VAE (on the same hardware, 4× A100 GPUs). The VAE was trained for 3500 epochs, and the MVAE was trained for 600 epochs (following 100 epochs of NAMM training). Both approaches produce visually similar samples, yet the images of the divergence field and histograms for this constraint distance show that the MVAE leads to overall better constraint satisfaction. Furthermore, in terms of MMD and KID, the MVAE distribution is even closer to the true data distribution.

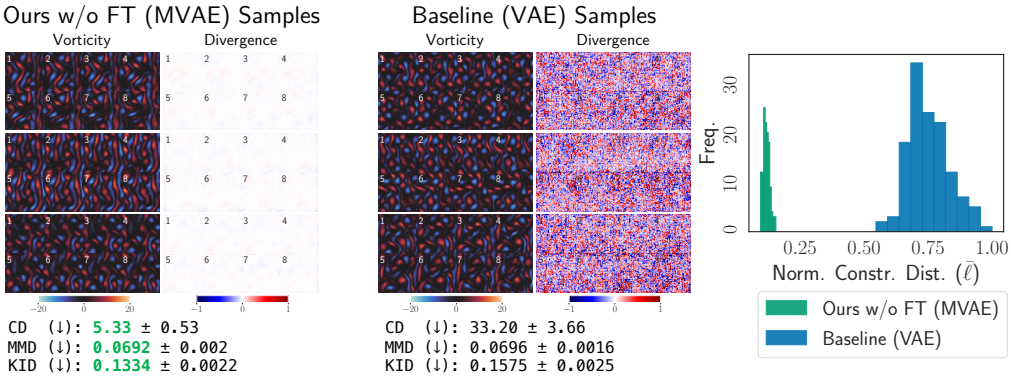


Figure 7.8: Mirror VAE (MVAE) vs. VAE, demonstrated on the divergence-free constraint. Baseline samples are obtained from a VAE trained in the original space, while MVAE samples are obtained by sampling from a VAE trained in the mirror space and then mapping those samples back to the original space via the learned inverse mirror map. Three samples are shown (vorticity on the left and divergence on the right) for each method. (Recall that each image consists of eight state snapshots; here we have labeled the number of each snapshot.) The vorticity fields show that the visual statistics of both generated distributions are extremely similar, but the corresponding divergence fields are drastically different. The MVAE samples are much closer to satisfying 0-divergence everywhere. As further evidence, the histograms show that normalized constraint distances of MVAE samples are significantly lower. We also report the mean ± std. dev. constraint distance (CD), computed as $100\lambda_{\text{constr}}\ell$, as well as the MMD and KID. All three metrics were estimated with 10000 generated and true samples. The MVAE leads to improved constraint satisfaction and distribution-matching accuracy compared to a vanilla VAE. This experiment demonstrates how a NAMM can be used to constrain generative models besides diffusion models.

7.4 Experiment Details

7.4.1 Implementation details

MDM score model For training the score model s_θ in the learned mirror space, we followed the implementation of Song et al. [282]. We used the NCSN++ architecture with 64 filters in the first layer and the VP SDE with $\beta_{\text{min}} = 0.1$ and $\beta_{\text{max}} = 20$. Training was done using the Adam optimizer with a learning rate of 0.0002 and gradient norm clipping with a threshold of 1.

NAMM For g_ϕ, we followed the implementation of the gradient of a strongly-convex ICNN of Tan et al. [297], configuring the ICNN to be 0.9-strongly convex. Following the settings of CycleGAN [359], f_ψ was implemented as a ResNet-based generator with 6 residual blocks and 32 filters in the last convolutional layer. For all constraints except the divergence-free constraint, we had the ResNet-based generator output the residual image (i.e., $f_\psi(\tilde{x}) = \text{ResNet}(\tilde{x}) + \tilde{x}$). We found that for the divergence-free constraint, a non-residual-based inverse map (i.e., $f_\psi(\tilde{x}) = \text{ResNet}(\tilde{x})$) achieves better constraint loss. The NAMM was trained using Adam optimizer with a learning rate of 0.001 for the divergence-free constraint and a learning rate of 0.0002 for all other constraints.

Table 7.2 shows the hyperparameter choices for each constraint. The regularization weight λ_{reg} in the NAMM objective (Equation 7.1) was fixed at 0.001. We used 3 ICNN layers for images 64×64 or smaller and 2 ICNN layers for images 128×128 or larger for the sake of efficiency. These hyperparameter values do not need to be heavily tuned, as we chose these settings through a coarse parameter search (e.g., trying $\lambda_{\text{constr}} = 0.01$ or $\lambda = 1$ to see which would lead to reasonable loss curves).

	Num. ICNN layers	σ_{max}	λ_{constr}
Total Brightness	3	0.1	0.01
1D Burgers'	3	0.1	1
Divergence-free	2	0.5	1
Periodic	3	0.1	1
Count	2	0.1	0.01

Table 7.2: NAMM hyperparameter values for each constraint in our experiments.

The main results shown in Figure 7.3 were taken from the finetuned NAMM, ensuring that the total training time of the NAMM, MDM, and finetuning did not exceed the total training time of the baseline DM. While we kept track of the

validation loss, this was not used to determine stopping time. We found that the NAMM training and MDM training were not prone to overfitting, so we chose the total number of epochs based on observing a reasonable level of convergence of the loss curves. We found that some overfitting is possible during finetuning but did not perform early stopping. All results were obtained from unseen test data because we fed random samples from the MDM into the inverse map and made sure not to use the same random seed as the one used to generate finetuning data. For all constraints, we generated 12800 training examples from the MDM for finetuning.

Table 7.3 details the exact number of training epochs for each constraint. Figure 7.4 in Section 7.3 compares the constraint distances of our method without finetuning, our method with finetuning, and the baseline DM as a function of compute time.

	NAMM epochs (before FT)	MDM epochs	FT epochs	DM epochs
Total Brightness	30	200	1000	450
1D Burgers'	100	300	700	1500
Divergence-free	100	700	500	2000
Periodic	50	300	1000	1000
Count	50	300	700	1500

Table 7.3: Number of training epochs of the NAMM, MDM, finetuning, and DM used for the results in Figure 7.3. These were chosen so that our method (including the NAMM training, MDM training, finetuning data generation, and finetuning) did not take longer to train than the DM.

Mirror map parameterization ablation study For the comparison of parameterizing the mirror map as the gradient of an ICNN versus as a ResNet-based generator (Figure 7.7), we used a ResNet-based generator that outputs the residual image. This means that the inverse mirror map was parameterized as a residual-based network ($\boldsymbol{f}_\psi(\tilde{\boldsymbol{x}}) = \text{ResNet}(\tilde{\boldsymbol{x}}; \psi) + \tilde{\boldsymbol{x}}$), and so was the ResNet-based forward mirror map ($\boldsymbol{g}_\phi(\boldsymbol{x}) = \text{ResNet}(\boldsymbol{x}; \phi) + \boldsymbol{x}$).

VAE We used a convolutional autoencoder architecture consisting of five convolutional layers with GELU activation functions in the encoder and decoder. Our implementation is borrowed from the autoencoder tutorial of Lippe [209]. We set the number of latent dimensions to 128 and the number of features in the first layer of the encoder to 64. For training, we followed the β-VAE training objective [153], which consists of two terms: one to increase the likelihood of training data under

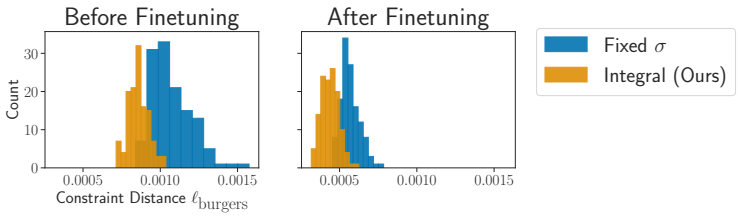


Figure 7.9: Fixed σ vs. integrating over $[0, \sigma_{\text{max}}]$. The NAMM objective involves optimizing $\boldsymbol{f}_\psi$ over the sequence of noisy mirror distributions defined in Equation 7.2. We compare this approach of integrating over $\sigma \in [0, \sigma_{\text{max}}]$ to setting a fixed noise standard deviation $\sigma = \sigma_{\text{max}}$ in the context of the 1D Burgers' constraint (here $\sigma_{\text{max}} = 0.1$). Both before and after finetuning, the constraint distances ℓ_{burgers} of inverted MDM samples are smaller if the NAMM was trained with varying σ. The histograms show the constraint distances of 128 samples from each method.

the VAE probabilistic model and one to minimize the KL divergence from the latent distribution to a Gaussian prior. The latter is weighted by a scalar $\beta > 0$. For our purposes, the maximum-likelihood term corresponds to an MSE reconstruction loss, and we set $\beta = 0.001$. We used Adam with a learning rate of 0.0002.

7.4.2 Data assimilation with mirror DPS

Following the original DPS [79], we use a hyperparameter $\zeta \in \mathbb{R}_{>0}$ to re-weight the time-dependent measurement likelihood. Higher values of ζ impose greater measurement consistency, but setting ζ too high can cause instabilities and artifacts. The data assimilation results in Figure 7.5 used $\zeta = 0.1$ and constraint-guidance weight equal to 200.

7.4.3 Measures of distance between distributions

7.4.3.1 MMD

The maximum mean discrepancy (MMD) between two distributions is computed by embedding both distributions into a reproducing kernel Hilbert space (RKHS) and using samples to estimate the resulting distance. We use the popular Gaussian radial basis function (RBF) kernel to construct the RKHS, setting the length scale σ as

$$\sqrt{\text{median}\left(\left\{\left\|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\right\|_2^2\right\}\right)/2}, \tag{7.8}$$

i.e., the square root of half the median of the pairwise squared Euclidean distances in the dataset $\left\{\boldsymbol{x}^{(i)}\right\}_{i=1}^n$. This is a popular choice in previous work [51] and has been theoretically and empirically justified [131].

Our MMD implementation is based on the code provided for the work of Sutherland et al. [294] [293]. We estimated mean and standard deviation empirically with 50 random subsets of $1,000$ samples from each dataset. The length scale was estimated (Equation 7.8) for each subset using the samples in the true subset. In total, the generated and true datasets contained 10000 samples each. Held-out test images were used as true samples for all constraints, except for total brightness (due to a lack of test images in the dataset, training images were used for this constraint only).

7.4.3.2 KID

The Kernel Inception Distance (KID) between two distributions is based on the Inception v3 features evaluated for samples from both distributions. Following standard practice, we use the $2,048$-dimensional final average pooling features. The KID is computed as the squared MMD (using a polynomial kernel) between the two embedded distributions. It has several advantages over the Fréchet Inception Distance (FID) [152], including being unbiased and more sample-efficient [38]. We note, however, that the Inception network was trained on natural images, so both KID and FID are not perfect metrics for the types of data we consider in this work, such as physics-based simulation outputs.

KID evaluation is based on the `gan-metrics-pytorch` repository [111], using the same Inception v3 weights as those used in the official TensorFlow implementation of FID [152]. We evaluated KID with the same samples that were used for MMD. Since the Inception network takes RGB images as input, we represented the samples as grayscale images converted to RGB. For all the constraint datasets except for the divergence-free Kolmogorov flows, we clipped the image pixel values to $[0, 1]$ before converting them to RGB. For the divergence-free data, we clipped the values in the vorticity images to $[-20, 20]$ and then rescaled this range to $[0, 1]$.

7.4.3.3 FID

Although FID is a biased finite-sample estimator and heavily depends on the number of samples, it is a popular metric for evaluating generative models. Table 7.4 reports FID values in addition to the MMD and KID values in Table 7.1. We find that the rankings of our method before finetuning, our method after finetuning, and the baseline method are consistent with the rankings when using KID in Table 7.1. Again we note that FID and KID are based on Inception features that were tuned

to natural images, so they are not the most reliable measures of distance between distributions for these particular image datasets.

	Total Brightness	1D Burgers'	Divergence-free	Periodic	Count
Ours w/o FT	5.02	3.36	**2.21**	2.26	**24.18**
Ours	**4.58**	**2.04**	3.97	**1.99**	37.65
Baseline	42.25	140.70	3.41	6.42	69.76

Table 7.4: FID values (↓) associated with the comparisons in Table 7.1. According to FID, our approach consistently outperforms the baseline DM approach in matching the true distribution. Finetuning even sometimes improves FID while improving constraint satisfaction (see Table 7.1 for constraint distances).

Our FID implementation is borrowed from the `pytorch-fid` codebase [231]. Per standard practice, we estimated FID with 50000 generated samples and 50000 true samples. As was the case with MMD and KID, held-out test images were used as true samples, except that training images were used for the total brightness data. We pre-processed the samples for the Inception network in the same way we did for KID.

7.5 Constraint Details

This section provides details about each demonstrated constraint and its corresponding dataset.

Total brightness The total brightness, or total flux, of a discrete image $\boldsymbol{x} \in \mathbb{R}^d$ is simply the sum of its pixel values: $V(\boldsymbol{x}) := \sum_{i=1}^{d} \boldsymbol{x}_i$. We use the constraint distance function

$$\ell_{\text{flux}}(\boldsymbol{x}) := \left| V(\boldsymbol{x}) - \bar{V} \right|,$$

where $\bar{V} \in \mathbb{R}_{\geq 0}$ is the target total brightness. The dataset used for this constraint contains images from general relativistic magneto-hydrodynamic (GRMHD) simulations [329] of Sgr A* with a fixed field of view. The images (originally 400×400) were resized to 64×64 pixels and rescaled to have a total flux of 120. The dataset consists of $100,000$ training images and 100 validation images.

1D Burgers' Burgers' equation [25, 58] is a nonlinear PDE that is a useful model for fluid mechanics. We consider the equation for a viscous fluid $u = u(t, x)$ in one-dimensional space:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2}, \tag{7.9}$$

where $u(0, x)$ is some initial condition $u_0(x)$, and $\nu \in \mathbb{R}_{\geq 0}$ is the viscosity coefficient. We use Crank-Nicolson [86] to discretize and approximately solve Equation 7.9 by representing the solution on an $n_x \times n_t$ grid, where n_x is the spatial discretization, and n_t is the number of snapshots in time. Given an $n_x \times n_t$ image, we wish to verify that it could be a solution to Equation 7.9 with the Crank-Nicolson discretization. Letting $\boldsymbol{x} \in \mathbb{R}^{n_x \times n_t}$ denote the 2D image, we formulate the following distance function for evaluating agreement with the Crank-Nicolson solver:

$$\ell_{\text{burgers}}(\boldsymbol{x}) := \frac{1}{n_t - 1} \sum_{t=0}^{n_t - 2} \left\| \boldsymbol{x}[:, t+1] - f_{\text{C-N}}(\boldsymbol{x}[:, t]) \right\|_1,$$

where $f_{\text{C-N}} : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ outputs the snapshot at the next time using Crank-Nicolson, and Pythonic notation is used for simplicity. Note that a finite-differences loss as proposed for PINNs [254] would also work, but then our data would have non-negligible constraint distances since Crank-Nicolson solutions do not strictly follow a low-order finite-differences approximation.

Using a Crank-Nicolson solver [86] implemented with Diffrax [183], we numerically solved the 1D Burgers' equation (Equation 7.9) with viscosity coefficient $\nu = 0.5$.

The initial conditions were sampled from a Gaussian process based on a Matérn kernel with smoothness parameter 1.5 and length scale equal to 1.0. We discretized the spatiotemporal domain into a 64×64 grid covering the spatial extent $x \in [0, 10]$ and time interval $t \in [0, 8]$. We ran Crank-Nicolson with a time step of $\Delta t = 0.025$ and saved every fifth step for a total of 64 snapshots. We followed this process to create our 1D Burgers' dataset of $10,000$ training images and $1,000$ validation images.

Divergence-free The study of fluid dynamics often involves incompressible, or divergence-free, fluids. Letting $\boldsymbol{u} = \boldsymbol{u}(x, y, t)$ be the time-dependent trajectory of a 2D velocity field, the divergence-free constraint says that $\nabla \cdot \boldsymbol{u} = \frac{\partial \boldsymbol{u}_x}{\partial x} + \frac{\partial \boldsymbol{u}_y}{\partial y} = 0$. We assume an $n_x \times n_y$ spatial grid and represent trajectories as two-channel (for the two velocity components) images showing each $n_x \times n_y$ snapshot for a total of n_t snapshots. Such an image $\boldsymbol{x}$ has a corresponding image of the divergence field $\mathrm{div}(\boldsymbol{x})$, which has the same size as $\boldsymbol{x}$ and represents the divergence of the trajectory. We formulate the following distance function that penalizes non-zero divergence:

$$\ell_{\mathrm{div}}(\boldsymbol{x}) := \|\mathrm{div}(\boldsymbol{x})\|_1 .$$

We created a dataset of Kolmogorov flows, which satisfy a Navier-Stokes PDE, to demonstrate the divergence-free constraint. The Navier-Stokes PDEs are ubiquitous in fields including fluid dynamics, mathematics, and climate modeling and have the following form:

$$\frac{\partial \boldsymbol{u}}{\partial t} = -\boldsymbol{u}\nabla\boldsymbol{u} + \frac{1}{Re}\nabla^2\boldsymbol{u} - \frac{1}{\rho}\nabla p + \mathbf{f}\nabla \cdot \boldsymbol{u} = 0,$$

where $\boldsymbol{u} = \boldsymbol{u}(x, y, t)$ is the 2D velocity field at spatial location (x, y) and time t, Re is the Reynolds number, ρ is the density, p is the pressure field, and $\mathbf{f}$ is the external forcing. Following Kochkov et al. [191] and Rozet and Louppe [261], we set $Re = 10^3$, $\rho = 1$, and $\mathbf{f}$ corresponding to Kolmogorov forcing [67, 42] with linear damping. We consider the spatial domain $[0, 2\pi]^2$ with periodic boundary conditions and discretize it into a 64×64 uniform grid. We used `jax-cfd` [191] to randomly sample divergence-free, spectrally filtered initial conditions and then solve the Navier-Stokes equations with the forward Euler integration method with $\Delta t = 0.01$ time units. We saved a snapshot every 20 time units for a total of 8 snapshots in the time interval $[3, 4.6]$. We represent the solution as a two-channel 128×256 image showing the snapshots in left-to-right order. In total, the dataset consists of $10,000$ training images and $1,000$ validation images.

Periodic Assuming the constraint that every image is a periodic tiling of n_{tiles} unit cells, we formulate the following constraint distance for a given image $\boldsymbol{x}$:

$$\ell_{\text{periodic}}(\boldsymbol{x}) := \sum_{i=1}^{n_{\text{tiles}}} \frac{1}{n_{\text{tiles}}} \sum_{j=1}^{n_{\text{tiles}}} \left\| \boldsymbol{t}_i(\boldsymbol{x}) - \boldsymbol{t}_j(\boldsymbol{x}) \right\|_1 ,$$

which compares each pair of tiles, where $\boldsymbol{t}_i(\boldsymbol{x})$ denotes the i-th tile in the image. For our experiments, we consider 32×32 unit cells that are tiled in a 2×2 pattern to create 64×64 images. Using the unit-cell generation code of Ogren et al. [241], we created a dataset of $30,000$ training images and 300 validation images.

Count For the count constraint, we rely on a CNN to estimate the count of a particular object. Letting $f_{\text{CNN}} : \mathbb{R}^d \to \mathbb{R}$ be the trained counting CNN, we turn to the following constraint distance function for a target count $\bar{c}$:

$$\ell_{\text{count}}(\boldsymbol{x}) := |f_{\text{CNN}}(\boldsymbol{x}) - \bar{c}| .$$

We demonstrate this constraint with astronomical images that contain a certain number of galaxies. In particular, we simulated 128×128 images of radio galaxies with background noise [84], each of which has exactly eight ($\bar{c} = 8$) galaxies with an SNR ≥ 15 dB. The dataset consists of $10,000$ training images and $1,000$ validation images.

To train the counting CNN, we created a mixed dataset with images of $6, 7, 8, 9$, or 10 galaxies that includes $10,000$ training images and $1,000$ validation images for each of the five labels. The CNN architecture was adapted from a simple MNIST classifier [1] with two convolutional layers followed by two dense layers with ReLU activations. The CNN was trained to minimize the mean squared error between the real-valued estimated count and the ground-truth count.

Chapter 8

CONCLUSION

As long as sensors fall short of capturing reality, priors are necessary to produce images of reality. Even digital cameras, which many people believe to be near-perfect sensors of the visual world, use an RGB color filter that captures just one primary color in each pixel. This means that our digital photos depend on implicit assumptions to interpolate the missing two colors in each pixel. As the gap between sensors and reality grows even larger, as is the case in many scientific applications, even stronger assumptions are required to fill in the holes in the data. However, more holes in the data also means that more features in the image come from prior assumptions rather than real observations. It is important to develop methods to responsibly harness assumptions to identify reliable image features.

One way to responsibly harness imaging assumptions is to evaluate images under a variety of priors rather than a single one. There are two main benefits of reconstructing images under multiple different priors: (1) we can identify image features that are consistent across the priors and therefore reliable for scientific interpretation, and (2) we offer people the option of choosing a preferred image based on which prior they most believe. In Chapter 2, we proposed taking a fully data-driven approach, leveraging score-based priors to flexibly and scalably incorporate a wider range of image assumptions simply by changing the training dataset for the score-based prior. Previously, with handcrafted regularizers and simple priors, it would have been infeasible to cover such a wide range of priors. In Chapter 3, we demonstrated our approach by reconstructing images from the real EHT observations of M87* under different assumptions of varying strengths, including a prior trained on images of everyday objects, a prior trained on human faces, a prior trained black-hole fluid-flow simulations, and a prior trained on a geometric model of black-hole accretion disks. We found that some image features were robust to the prior, including the appearance of a ring, the diameter of the ring, and the location of the bright spot. Other details of the image were left to the interpretation of the prior. Although there is no way to say which image most closely reflects reality (likely all of them are far from reality), with our approach, people can imagine M87* under different assumptions and choose to believe the picture that most closely aligns with their prior beliefs, such as believing the GRMHD-based reconstruction if they fully trust

the GRMHD model. At the same time, image features that persist regardless of the prior should be indisputable.

Another way to responsibly harness imaging assumptions is by using priors that are themselves indisputable. Physics-based priors are a great example of this since they are rooted in well-verified physics equations and models. In Chapters 5 and 6, we saw how our visual elastography methods, visual vibration tomography (VVT) and visual surface wave elastography (VSWE), relied on solid mechanics theory to infer subsurface material properties from limited surface measurements. The solutions sought by VVT and VSWE satisfy equations that are generally accepted under the assumption of linear elasticity. Even when the physics model is not known for sure, it is possible to use assumed physics as a starting point and arrive at a physics-informed solution. Chapter 4 on black-hole emission tomography provided an example of this. We proposed a physics-informed approach that recovers both the 4D emissivity field and 3D velocity field from single-viewpoint measurements, whereby the velocity field is guided by an assumed velocity model but ultimately describes the dynamics of the estimated emissivity field. This type of approach can be used to arrive at more accurate velocity models and in turn more accurate physics priors.

To make data-driven priors themselves more reliable, we saw in Chapter 7 how to embed physics constraints into them. Many inverse problems, such as data assimilation, are inherently constrained by some underlying physics equation but would still benefit from the flexibility and scalability of data-driven priors. We proposed neural approximate mirror maps (NAMMs) for flexibly imposing constraints, including physics-based ones, on generative models as a means to obtaining physics-constrained data-driven priors. We demonstrated how our method can be used to solve inverse problems in a learned mirror space and obtain solutions that satisfy soft physics constraints via a learned inverse mirror map. As we aim to image more and more evasive scientific phenomena, it will be crucial to cleverly combine data-driven and physics-based assumptions.

The vision of this thesis is to expand scientific knowledge beyond the limits imposed by physical sensors by putting more power in the prior. Given that priors are often unavoidable in the imaging process, we might as well use them to our advantage. This thesis aims to provide scientists with the tools to impose whatever assumptions they desire in order to both reliably interpret image reconstructions and update their own assumptions.

8.0.1 Future directions

With the continued advancement of generative artificial intelligence (AI), we will inevitably aim to image more and more phenomena beyond the reach of physical sensors, requiring more methods to ensure responsible use of the rich priors afforded by generative AI. Although our work on score-based priors establishes a paradigm shift from imaging under a single prior to imaging under multiple priors, there remain challenges to fully realize this paradigm shift. One such challenge is to rigorously discern which image features arise from the prior versus from the measurements. Our study of extracted ring features in Chapter 3 required manually estimating parameters under a specialized geometric model of the image. Instead, it may be possible to automatically identify image features and categorize them as coming from the prior or from the measurements. There are existing learning-based techniques to identify large-scale image features that explain the uncertainty in the posterior [237], but it would be beneficial to further decompose the posterior uncertainty into a prior component and a measurement component. Such a method for uncertainty visualization would help scientists better understand which parts of the image can be reliably interpreted and which parts should be reserved for imagination.

Furthermore, there remain computational bottlenecks to applying multiple priors. Using a single score-based prior requires gathering a training dataset for the score-based diffusion model, training the score-based diffusion model, and then running the inference algorithm. Training the diffusion model and running the inference algorithm, which in our case requires optimizing a variational distribution, both take a significant amount of time and computational resources. Creating a training dataset can also be time- and resource-intensive, for example if it requires running physics simulations. One promising step is to engineer software that automates and expedites the process of creating training data, training the diffusion model with an automatic early-stopping criterion, and applying the resulting prior to a chosen inference algorithm. We could also imagine creating a meta posterior: a distribution of posteriors under different priors. This would require users to define a set of priors and offer a way to analyze uncertainty under the chosen set of priors. It may also offer insights into which prior is most suited to the imaging task at hand, such as by finding the prior that leads to the highest posterior probability. These are just a few examples of ways to automatically reconstruct images and analyze uncertainty under multiple priors.

Even with advances in generative AI, there will always be a place for physics-based

priors. Physics knowledge is especially useful when there are limited data and computational resources for training a fully data-driven prior. However, a challenge is overcoming uncertainty in the chosen physics model. Our VVT and VSWE approaches require a pre-defined physics equation, but it would be useful to be able to update this equation to account for mismatches between the assumed physics and reality (e.g., if the material is not linear-elastic or if the Poisson's ratio is not exactly correct). Our approach to black-hole emission tomography provides one example of a way to simultaneously find the desired solution and update the assumed physics model. A further benefit of this type of approach, which we did not explore in this thesis, is that it is differentiable and thus allows for estimating unknown physical parameters, such as the spin of the black hole. We can estimate such parameters by jointly optimizing them along with the emissivity field and velocity field. By approaching imaging problems in a physics-informed way, we can better update our own understanding of physics.

The most promising avenue for many scientific imaging applications is to combine data-driven and physics-based approaches. We believe that NAMMs are applicable to many problems that entail physics constraints. For example, 3D protein imaging with cryo-electron microscopy [356] would benefit from symmetry constraints; specifically, we could train a data-driven prior on 3D protein structures in a learned mirror space to satisfy rotational symmetries and solve the inverse problem of 3D protein reconstruction from multiple image-plane projections in the learned mirror space. Seismic traveltime tomography [176] would benefit from velocity constraints; given traveltimes from multiple sources, we could inpaint the traveltime fields with the constraint that the velocity fields must be consistent across all the sources and derive the velocity field from the inpainted traveltimes. The inverse design of materials would benefit from manufacturability constraints; we could train a black-box neural network to estimate how structurally sound and manufacturable a design is and perform inverse design in a learned mirror space that maximizes manufacturability. In all of these applications, the data-driven aspect allows for sophisticated reconstructions while the physics-based aspect ensures physical plausibility of the reconstructions.

Indeed, the intersection of data-driven and physics-based approaches offers exciting opportunities to advance both AI and science. Many generative AI methods are currently fully data-driven and depend on seeing enough training data to learn implicit physics. Instead, we can incorporate physics constraints in these methods to im-

prove the data efficiency, generalizability, and physical plausibility of generative AI. For example, current methods for video generation, although impressive, still show artifacts of implausible physical motion [212]. These artifacts can be overcome by imposing basic physics principles, such as Newton's laws of motion. This type of approach provides an example of "Science for AI." In the other direction, "AI for Science," we have already discussed how a physics-informed machine-learning approach can help us update physics models. We can also consider how to discover entirely new physics through physics-informed AI. Drawing inspiration from previous methods for equation discovery [110], it may be possible to simultaneously train a NAMM and learn the underlying physics constraint. The long-term vision of this thesis is to push the frontiers of AI and science by blending physics-based and data-driven methods to solve challenging imaging problems. Combining the worlds of machine learning and physics will allow computational imaging to better aid scientific discovery.

Appendix A

EHT IMAGING BACKGROUND

A.1 VLBI Data Products

In this appendix, we provide background on the data products obtained with VLBI. In VLBI, a network of radio telescopes collects spatial-frequency measurements of the sky's image. We denote the source image as $I(x, y)$, where (x, y) are 2D spatial coordinates. Each pair of telescopes is called a *baseline* and provides a Fourier measurement called a *visibility*. The van Cittert-Zernike Theorem [81, 347] states that the ideal visibility v_{ij}^* measured by the baseline $\boldsymbol{b}_{ij}$ between telescopes i and j is a single (u, v) measurement on the complex 2D Fourier plane [302]:

$$v_{ij}^* := \tilde{I}(u, v) = \int \int I(x, y) e^{-2\pi \mathrm{i}(xu+yv)} \mathrm{d}x \mathrm{d}y. \tag{A.1}$$

(Here i is used to denote the imaginary unit to avoid confusion with the telescope index i.) The coordinates (u, v) (measured in wavelengths) are the projected baseline orthogonal to the line of sight. An array of N_s telescopes, or stations, has $\binom{N_s}{2}$ independent baselines, each providing a visibility at each point in time.

In practice, ideal visibilities are corrupted due to multiple factors: (1) baseline-dependent thermal noise, (2) station-dependent gain errors, and (3) station-dependent phase errors. Baseline-dependent **thermal noise** is modeled as a Gaussian random variable $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2)$, where σ_{ij} is based on the system equivalent flux density (SEFD) of each telescope: $\sigma_{ij} \propto \sqrt{\mathrm{SEFD}_i + \mathrm{SEFD}_j}$. The **station-dependent gain error** g_i arises from each telescope i using its own time-dependent 2×2 Jones matrix [146]. The **station-dependent phase error** ϕ_i arises from atmostpheric turbulence that causes light to travel at different velocities toward each telescope [154, 194, 299]. Other sources of corruption, including polarization leakage and bandpass errors, may introduce baseline-dependent errors, but they are slow-varying and assumed to be removable with *a priori* calibration [65]. The measured visibility of baseline $\boldsymbol{b}_{ij}$ can be written as

$$v_{ij} = g_i g_j e^{\mathrm{i}(\phi_i - \phi_j)} v_{ij}^* + \varepsilon_{ij}, \tag{A.2}$$

where all systematic errors (i.e., those besides thermal noise) are wrapped into station-dependent gain/phase errors.

A.1.1 Closure quantities

Station-dependent errors are difficult to remove owing to the absence of corroborating information from other stations. Calibrating the measured visibilities calls for an iterative self-calibration process that introduces many *a priori* assumptions and

becomes infeasible at high telescope frequencies [65]. An alternative avenue is to use closure quantities that are unchanged by station-dependent errors.

Closure phases [175] are robust to station-dependent phase errors. They arise from a data product known as the complex *bispectrum*, which is formed by multiplying the three baselines within a triangle of telescopes i, j, k:

$$v_{ij}v_{jk}v_{ki} = \left(g_i g_j e^{\mathrm{i}(\phi_i-\phi_j)}v^*_{ij} + \varepsilon_{ij}\right)\left(g_j g_k e^{\mathrm{i}(\phi_j-\phi_k)}v^*_{jk} + \varepsilon_{jk}\right)\left(g_k g_i e^{\mathrm{i}(\phi_k-\phi_j)}v^*_{ki} + \varepsilon_{ki}\right) \tag{A.3}$$

$$= g^2_{ijk}e^{\mathrm{i}(\phi_i-\phi_j)}e^{\mathrm{i}(\phi_j-\phi_k)}e^{\mathrm{i}(\phi_k-\phi_i)}v^*_{ij}v^*_{jk}v^*_{ki} + \varepsilon_{ijk} \tag{A.4}$$

$$= g^2_{ijk}v^*_{ij}v^*_{jk}v^*_{ki} + \varepsilon_{ijk}, \tag{A.5}$$

where ε_{ijk} is the thermal noise in the measured bispectrum. Importantly, Equation equation A.5 does not include any phase errors. Thus the closure phase is given by the phase of the bispectrum and is robust to phase corruption. While the total number of triplets in the telescope array is $\binom{N_s}{3}$, the total number of linearly independent closure phases is $N_{\mathrm{cp}} = \binom{N_s-1}{2}$. To understand this number, see that the set of independent closure phases can be formed by selecting one station as a reference and then creating the set of all triangles that contain that station. The resulting closure phases are independent in that no one closure phase can be formed as a linear combination of other closure phases.

Closure amplitudes [310] address the issue of station-dependent gain errors. A closure amplitude arises from a combination of four telescopes i, j, k, ℓ:

$$\frac{v_{ij}v_{kl}}{v_{ik}v_{jl}} = \frac{\left(g_i g_j e^{\mathrm{i}(\phi_i-\phi_j)}v^*_{ij} + \varepsilon_{ij}\right)\left(g_k g_\ell e^{\mathrm{i}(\phi_k-\phi_\ell)}v^*_{k\ell} + \varepsilon_{k\ell}\right)}{\left(g_i g_k e^{\mathrm{i}(\phi_i-\phi_k)}v^*_{ik} + \varepsilon_{ik}\right)\left(g_j g_\ell e^{\mathrm{i}(\phi_j-\phi_\ell)}v^*_{j\ell} + \varepsilon_{j\ell}\right)} \tag{A.6}$$

$$= \frac{g_i g_j g_k g_\ell e^{\mathrm{i}(\phi_i-\phi_j)}e^{\mathrm{i}(\phi_k-\phi_\ell)}v^*_{ij}v^*_{k\ell}}{g_i g_j g_k g_\ell e^{\mathrm{i}(\phi_i-\phi_k)}e^{\mathrm{i}(\phi_j-\phi_\ell)}v^*_{ik}v^*_{j\ell}} + \varepsilon_{ijk\ell} \tag{A.7}$$

$$= e^{2\mathrm{i}(\phi_k-\phi_j)}\frac{v^*_{ij}v^*_{k\ell}}{v^*_{ik}v^*_{j\ell}} + \varepsilon_{ijk\ell}, \tag{A.8}$$

where $\varepsilon_{ijk\ell}$ is the thermal noise in the closure amplitude. Equation A.8 does not depend on station-dependent gain errors, so the amplitude $\left|\frac{v_{ij}v_{kl}}{v_{ik}v_{jl}}\right|$ is taken as the closure amplitude. In our work, we use the log of the closure amplitude. The number of linearly independent log closure amplitudes is $N_{\mathrm{logca}} = \binom{N_s(N_s-3)}{2}$.

A.2 Ring Feature Extraction

We used the `REx` feature extraction algorithm [62] to compute the characteristic features in Section 3.5. The diameter and fractional central brightness follow the same formulae as in Akiyama et al. [7]. Except for the fractional central brightness, which is not included in `REx`, all features were computed exactly according to the latest implementation of `REx`[1] (as of 2023 October). The `REx` implementation corresponds to slightly different equations for computing features than those given in Akiyama et al. [7]. In this appendix, we will note any differences from the equations used in Akiyama et al. [7].

`REx` first pre-processes the image by blurring it with a 2 μas FWHM Gaussian and re-gridding it to 160×160 pixels. It identifies the ring center based on the image thresholded to 5% of the peak brightness, and then it computes characteristic features. The characteristic features, which we define in the following paragraphs, are all computed based on radial–angular profiles $I(r, \theta)$ of the centered image, where $I(r, \theta)$ is the brightness at radius r and azimuthal angle θ from the measured center. The profiles are interpolated over the domains $r \in [0, 50]$ μas and $\theta \in [0, 2\pi]$ radians.

The diameter d is measured as twice the mean radial distance of the peak brightness:

$$d = 2 \left\langle r_{\rm pk}(\theta) \right\rangle_{\theta \in [0,2\pi]} := 2 \left\langle \arg\max_r I(r, \theta) \right\rangle_{\theta \in [0,2\pi]}, \tag{A.9}$$

where $\langle \cdot \rangle_{\theta \in [0,2\pi]}$ denotes the mean over the domain $\theta \in [0, 2\pi]$. The uncertainty of the diameter is given as the corresponding standard deviation. This equation for d exactly agrees with Equation (18) in Akiyama et al. [7].

The width w is measured as the mean FWHM of radial slices:

$$w = \langle \mathrm{FWHM}\left(I(r, \theta)\right) \rangle_{\theta \in [0,2\pi]}, \tag{A.10}$$

where $\mathrm{FWHM}(\cdot)$ evaluates the FWHM of a 1D radial profile. The uncertainty is computed as the corresponding standard deviation. Note that this is slightly different from the expression for w given in Equation (20) in Akiyama et al. [7], which first subtracts the mean flux outside the ring before estimating the width.

To measure the orientation angle, `REx` first estimates the FWHM of the mean radial profile with zero-mean outside flux, defined as

$$\mathrm{FWHM}\left(\bar{I}(r) - I_{\rm floor}\right),$$

[1] `https://github.com/achael/eht-imaging/blob/main/ehtim/features/rex.py`

where $\bar{I}(r) := \langle I(r,\theta)\rangle_{\theta\in[0,2\pi]}$ and $I_{\rm floor} := \langle I(r=50,\theta)\rangle_{\theta\in[0,2\pi]}$. Let $r_{\rm left}$ and $r_{\rm right}$ denote the minimum and maximum radii, respectively, of this FWHM. The orientation angle is computed by finding the phase of the first angular mode of each normalized angular profile $I(r,\theta)$ at fixed r throughout the FWHM and then taking the circular mean:

$$\eta = \left\langle \angle \left[\frac{\int_0^{2\pi} I(r,\theta) e^{i\theta} \mathrm{d}\theta}{\int_0^{2\pi} I(r,\theta)\mathrm{d}\theta} \right] \right\rangle_{r\in[r_{\rm left}, r_{\rm right}]} . \tag{A.11}$$

The uncertainty of η is the corresponding circular standard deviation. The only difference between this equation and Equation (21) in Akiyama et al. [7] comes from $r_{\rm left}$ and $r_{\rm right}$. In Akiyama et al. [7], $r_{\rm in} = (d-w)/2$ and $r_{\rm out} = (d+w)/2$ are used instead.

The azimuthal asymmetry A is measured as the mean normalized amplitude of the same first angular modes:

$$A = \left\langle \left| \frac{\int_0^{2\pi} I(r,\theta) e^{i\theta} \mathrm{d}\theta}{\int_0^{2\pi} I(r,\theta)\mathrm{d}\theta} \right| \right\rangle_{r\in[r_{\rm left}, r_{\rm right}]} , \tag{A.12}$$

and the uncertainty of A is the corresponding standard deviation. Once again, the only difference from Equation (22) in Akiyama et al. [7] is that $r_{\rm left}$ and $r_{\rm right}$ are used instead of $r_{\rm in}$ and $r_{\rm out}$.

We define the fractional central brightness $f_{\rm C}$ (which is not included in `REx`) as the ratio of an "interior" mean flux to the mean flux along the ring:

$$f_{\rm C} = \frac{\langle I(r,\theta)\rangle_{r\in[0,5],\theta\in[0,2\pi]}}{\langle I(d/2,\theta)\rangle_{\theta\in[0,2\pi]}}. \tag{A.13}$$

Here the inside is defined as the inner disk of radius 5 μas, and the outside is defined as the region with radius larger than the measured radius $d/2$. There is no uncertainty quantification for $f_{\rm C}$. This definition of $f_{\rm C}$ exactly agrees with Equation (23) in Akiyama et al. [7].

Appendix B

OTHER FORWARD MODELS

In this appendix, we describe the forward models of the inverse problems explored in Chapter 2: accelerated MRI, denoising, and reconstruction from low spatial frequencies ("deblurring"). These tasks have linear forward models such that

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\boldsymbol{y}}^2 \boldsymbol{I}), \tag{B.1}$$

with the corresponding log-likelihood function

$$\log p(\boldsymbol{y} \mid \boldsymbol{x}) \propto -\frac{1}{2\sigma_{\boldsymbol{y}}^2} \left\| \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \right\|_2^2 . \tag{B.2}$$

B.1 Accelerated MRI

Accelerated MRI collects sparse spatial-frequency measurements in κ-space of an underlying anatomical image. As the acceleration rate increases, the number of measurements decreases. The forward model can be written as

$$\boldsymbol{y} = \boldsymbol{M} \odot \mathcal{F}(\boldsymbol{x}^*) + \boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\boldsymbol{y}}^2 \boldsymbol{I}), \tag{B.3}$$

where $\boldsymbol{x} \in \mathbb{C}^D$ and $\boldsymbol{y} \in \mathbb{C}^M$. The operator $\mathcal{F}$ denotes the 2D Fourier transform, and $\boldsymbol{M} \in \{0, 1\}^D$ is a binary sampling mask that reduces the number of non-zero measurements to $M << D$. Often $\sigma_{\boldsymbol{y}}$ is assumed to be small (e.g., corresponding to an SNR of at least 30 dB). We use Poisson-disc sampling [313] for the sampling mask. 16×-acceleration, for example, corresponds to a sampling mask with only $1/16$ nonzero elements.

Experimental setup In our experiments, we assumed that $|\sigma_{\boldsymbol{y}}|$ is 0.05% of the DC (zero-frequency) amplitude. This corresponds to a maximum SNR of 40 dB. The only exception is for comparison to baselines, since baseline methods do not account for measurement noise. In this case, we let $|\sigma_{\boldsymbol{y}}| = 0.1\%$ of the DC amplitude along the horizontal direction of the true image, which amounts to a very low level of noise.

B.2 Denoising

The denoising forward model is simply

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\boldsymbol{y}}^2 \boldsymbol{I}), \tag{B.4}$$

where $\boldsymbol{x} \in \mathbb{R}^D$, and $\sigma_{\boldsymbol{y}}$ determines the level of i.i.d. Gaussian noise added to the clean image to get $\boldsymbol{y} \in \mathbb{R}^D$.

Experimental setup In our presented experiments on denoising, $\sigma_{\boldsymbol{y}} = 0.2$, which is 20% of the dynamic range.

B.3 Deblurring

We refer to the task of reconstruction from the lowest spatial frequencies as deblurring. The forward model is given by

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} + \boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\boldsymbol{y}}^2 \boldsymbol{I}), \tag{B.5}$$

where $\boldsymbol{x} \in \mathbb{C}^D$, $\boldsymbol{y} \in \mathbb{C}^M$, and $\boldsymbol{D} \in \mathbb{C}^{M \times D}$ performs a 2D DFT with only the first M DFT components.

Experimental setup In our presented experiments on deblurring, the measurements are the lowest 6.25% of the DFT components, and $|\sigma_{\boldsymbol{y}}| = 1$.

Appendix C

MECHANICS BACKGROUND

C.1 Linear Elasticity and the Elastic Wave Equation

Linear elasticity is the theory that dictates how a solid moves when subjected to external forces.

Momentum balance Within a solid body, material moves and displaces as a result of stresses exerted by neighboring pieces of material and external forces applied to the solid body. We can express this with Cauchy's momentum balance

$$\rho\, \ddot{\boldsymbol{u}} = \nabla \cdot \boldsymbol{\sigma} + \mathbf{f}, \tag{C.1}$$

where ρ is density, and the vectors $\boldsymbol{u}$, $\boldsymbol{\sigma}$, and $\mathbf{f}$ represent displacement, stress, and body forces (respectively) in each of the three spatial directions. Note ρ, $\boldsymbol{u}$, $\boldsymbol{\sigma}$, and $\mathbf{f}$ are all taken to be functions of space ($\mathbf{x}$). Cauchy's momentum balance is essentially the continuum mechanics version of Newton's second law.

Linear-elastic constitutive law In general, if a body of material displaces non-uniformly (such that $\boldsymbol{u}$ has a non-zero spatial gradient), stress ($\boldsymbol{\sigma}$) often develops within the material (and in many cases acts as a restorative force to return the body to some equilibrium configuration). The exact relationship between the spatial gradient of displacement (often written compactly as strain, $\boldsymbol{\varepsilon}$) and stress $\boldsymbol{\sigma}$ is called *a constitutive law*, which is different for different classes of materials. In our work, we restrict our studies to a *linear-elastic* constitutive law.

Linear elasticity proposes a constitutive law where the stress $\boldsymbol{\sigma}$ (like a pressure, or area-normalized force) that a piece of material feels is *linearly* proportional to the strain $\boldsymbol{\varepsilon}$ (the symmetrized spatial gradient of displacement) that the piece of material is experiencing. It can be written more concisely if instead of elastic modulus E and Poisson's ratio ν, we use the Lamé parameters $\mu := \frac{E}{2(1+\nu)}$ and $\lambda := \frac{E\nu}{(1+\nu)(1-2\nu)}$:

$$\boldsymbol{\sigma} = \lambda\, (\nabla \cdot \boldsymbol{u})\, \boldsymbol{I} + 2\mu\, \boldsymbol{\varepsilon}, \tag{C.2}$$

where $\boldsymbol{\varepsilon} := \frac{1}{2}\big(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^{\top}\big)$.

Elastic wave equation Substituting the linear-elastic constitutive law into Cauchy's momentum balance yields the elastic wave equation (a linear PDE problem where one traditionally solves for $\boldsymbol{u}$):

$$\rho\, \ddot{\boldsymbol{u}} = (\lambda + \mu)\, \nabla(\nabla \cdot \boldsymbol{u}) + \mu\, \nabla^2 \boldsymbol{u} + \mathbf{f}, \tag{C.3}$$

which describes how elastic waves travel through linear elastic materials.

C.2 Transient Analysis

Transient analysis means that time is being explicitly modeled. In transient simulations, we solve Equation C.3 for $\boldsymbol{u}(\mathbf{x}, t)$ by some choice of time-stepping scheme. Although these simulations are often more costly than their assumption-leveraging counterparts (e.g., harmonic analysis, described in Appendix C.3.1), they are able to model very general situations with few restrictions and with high fidelity.

In our VVT and VSWE work, all of the simulated experiments used transient analysis to represent the observed systems because we wanted the simulated material behavior to involve as few simplifying assumptions as possible and match as closely as possible with the behavior of real-world materials.

C.3 Bloch-Floquet Analysis

Bloch-Floquet (BF) analysis aims to reveal the natural wave modes of a system comprehensively, finding waves of every possible (γ, ω) frequency-wavevector pair. BF analysis applies a *time-harmonic assumption* and *BF periodic boundary conditions* (a spatially quasi-periodic assumption) to Equation C.3. Each of these assumptions is discussed in this section. In our VSWE experiments, all of the dispersion relations computed to fit the observed dispersion relations used BF analysis.

C.3.1 Time-harmonic assumption

BF analysis is a subcategory of *harmonic analysis*, and as such it employs the same time-harmonic assumption from harmonic analysis. The time-harmonic assumption says the solution is harmonic (sinusoidal) in time, written as

$$\boldsymbol{u}(\mathbf{x}, t) = \boldsymbol{u}_0(\mathbf{x}) e^{\mathbf{i}\omega t}, \tag{C.4}$$

where ω represents the time frequency of the harmonic solution, and $\boldsymbol{u}_0(\mathbf{x})$ is explicitly *not* a function of time. In general, assuming the displacement field is a harmonic function at frequency ω only makes sense when the external forces applied to the system are also harmonic functions at the same frequency. That is,

$$\mathbf{f}(\mathbf{x}, t) = \mathbf{f}_0(x) e^{\mathbf{i}\omega t}. \tag{C.5}$$

In cases where we are looking for natural modes (e.g., in VSWE we look for natural wave modes), we set external forces $\mathbf{f}_0 = \mathbf{0}$. However, there are other situations (e.g., in *frequency-domain* analysis, which is not used in this work) where non-zero forcing would be appropriate. Applying the harmonic assumption in Equation C.4

and assuming $\mathbf{f}_0 = \mathbf{0}$, Equation C.3 simplifies to

$$\mu\nabla^2\boldsymbol{u}_0 \;+\; (\lambda+\mu)\,\nabla\big(\nabla\cdot\boldsymbol{u}_0\big) \;+\; \rho\,\omega^2\,\boldsymbol{u}_0 \;=\; \mathbf{0}, \tag{C.6}$$

which is a linear eigenvalue PDE, where $\boldsymbol{u}_0(\mathbf{x})$ is the eigenfunction (representing the displacement field of the wave mode), ω^2 is the eigenvalue, and ω is the frequency (a.k.a. *eigenfrequency*) of the wave mode.

C.3.2 Bloch-Floquet boundary conditions

Alongside the time-harmonic assumption, BF analysis also adopts a *spatially* quasi-periodic assumption: that the solution field repeats from unit cell to unit cell, but with a phase shift. This is imposed through *BF periodic boundary conditions*:

$$\boldsymbol{u}(\mathbf{x}+\boldsymbol{a}_n) = \boldsymbol{u}(\mathbf{x})\,e^{\mathbf{i}\boldsymbol{\gamma}\cdot\boldsymbol{a}_n}, \qquad n = 1,\ldots,c, \tag{C.7}$$

where c is the spatial dimensionality of the problem ($c = 1, 2, 3$ is common for real-world problems), $\boldsymbol{a}_n$ is the size of the modeled material domain in each direction (commonly known as a *lattice vector*), and $\boldsymbol{\gamma}$ is the wavevector (as usual) which defines the lengthscale of Bloch-Floquet periodicity in each direction. An intuitive way to think about these boundary conditions is "the solution at one boundary is the same as the solution at the opposite boundary, but with a phase shift prescribed by the wavevector."

BIBLIOGRAPHY

[1] 8bitmp3. *8bitmp3/JAX-Flax-Tutorial-Image-Classification-with-Linen*. original-date: 2020-12-24T03:00:15Z. Sept. 25, 2024. URL: https://github.com/8bitmp3/JAX-Flax-Tutorial-Image-Classification-with-Linen (visited on 07/30/2025).

[2] Marek A. Abramowicz and P. Chris Fragile. "Foundations of Black Hole Accretion Disk Theory". In: *Living Reviews in Relativity* 16.1 (Jan. 14, 2013), p. 1. ISSN: 1433-8351. DOI: 10.12942/lrr-2013-1. URL: https://doi.org/10.12942/lrr-2013-1 (visited on 07/09/2025).

[3] Alexandre Adam et al. *Posterior samples of source galaxies in strong gravitational lenses with score-based priors*. 2022. DOI: https://doi.org/10.48550/arXiv.2211.03812. URL: https://arxiv.org/abs/2211.03812.

[4] Kazunori Akiyama et al. "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole". In: *The Astrophysical Journal Letters* 875.1 (Apr. 2019). Publisher: The American Astronomical Society, p. L1. ISSN: 2041-8205. DOI: 10.3847/2041-8213/ab0ec7. URL: https://dx.doi.org/10.3847/2041-8213/ab0ec7 (visited on 06/26/2025).

[5] Kazunori Akiyama et al. "First M87 Event Horizon Telescope Results. II. Array and Instrumentation". In: *The Astrophysical Journal Letters* 875.1 (Apr. 2019). Publisher: The American Astronomical Society, p. L2. ISSN: 2041-8205. DOI: 10.3847/2041-8213/ab0c96. URL: https://dx.doi.org/10.3847/2041-8213/ab0c96 (visited on 06/26/2025).

[6] Kazunori Akiyama et al. "First M87 Event Horizon Telescope Results. III. Data Processing and Calibration". In: *The Astrophysical Journal Letters* 875.1 (Apr. 2019). Publisher: The American Astronomical Society, p. L3. ISSN: 2041-8205. DOI: 10.3847/2041-8213/ab0c57. URL: https://dx.doi.org/10.3847/2041-8213/ab0c57 (visited on 06/26/2025).

[7] Kazunori Akiyama et al. "First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole". In: *The Astrophysical Journal Letters* 875.1 (Apr. 2019). Publisher: The American Astronomical Society, p. L4. ISSN: 2041-8205. DOI: 10.3847/2041-8213/ab0e85. URL: https://dx.doi.org/10.3847/2041-8213/ab0e85 (visited on 06/26/2025).

[8] Kazunori Akiyama et al. "First M87 Event Horizon Telescope Results. V. Physical Origin of the Asymmetric Ring". In: *The Astrophysical Journal Letters* 875.1 (Apr. 2019). Publisher: The American Astronomical Society, p. L5. ISSN: 2041-8205. DOI: 10.3847/2041-8213/ab0f43. URL: https://dx.doi.org/10.3847/2041-8213/ab0f43 (visited on 06/26/2025).

[9] Kazunori Akiyama et al. “First M87 Event Horizon Telescope Results. VI. The Shadow and Mass of the Central Black Hole”. In: *The Astrophysical Journal Letters* 875.1 (Apr. 2019). Publisher: The American Astronomical Society, p. L6. ISSN: 2041-8205. DOI: 10.3847/2041-8213/ab1141. URL: https://dx.doi.org/10.3847/2041-8213/ab1141 (visited on 06/26/2025).

[10] Kazunori Akiyama et al. “First Sagittarius A* Event Horizon Telescope Results. III. Imaging of the Galactic Center Supermassive Black Hole”. In: *The Astrophysical Journal Letters* 930.2 (May 2022). Publisher: The American Astronomical Society, p. L14. ISSN: 2041-8205. DOI: 10.3847/2041-8213/ac6429. URL: https://dx.doi.org/10.3847/2041-8213/ac6429 (visited on 06/27/2025).

[11] Kazunori Akiyama et al. “Imaging the Schwarzschild-radius-scale Structure of M87 with the Event Horizon Telescope Using Sparse Modeling”. In: *The Astrophysical Journal* 838.1 (Mar. 2017). Publisher: The American Astronomical Society, p. 1. ISSN: 0004-637X. DOI: 10.3847/1538-4357/aa6305. URL: https://dx.doi.org/10.3847/1538-4357/aa6305 (visited on 06/26/2025).

[12] Kazunori Akiyama et al. “SMILI: Sparse Modeling Imaging Library for Interferometry”. In: *Astrophysics Source Code Library* (2019), ascl–1904. DOI: https://doi.org/10.5281/zenodo.2616725.

[13] Kazunori Akiyama et al. “The persistent shadow of the supermassive black hole of M 87 - I. Observations, calibration, imaging, and analysis”. In: *Astronomy & Astrophysics* 681 (Jan. 1, 2024). Publisher: EDP Sciences, A79. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202347932. URL: https://www.aanda.org/articles/aa/abs/2024/01/aa47932-23/aa47932-23.html (visited on 06/25/2025).

[14] Carlo Alfano et al. *Learning mirror maps in policy mirror descent*. June 7, 2024. DOI: 10.48550/arXiv.2402.05187. arXiv: 2402.05187[stat]. URL: http://arxiv.org/abs/2402.05187 (visited on 06/27/2025).

[15] James Urquhart Allingham et al. “Learning Generative Models with Invariance to Symmetries”. In: *NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations*. 2022.

[16] Martin Alnæs et al. “The FEniCS Project Version 1.5”. In: *Archive of Numerical Software* 3.100 (Dec. 7, 2015). Number: 100. ISSN: 2197-8263. DOI: 10.11588/ans.2015.100.20553. URL: https://journals.ub.uni-heidelberg.de/index.php/ans/article/view/20553 (visited on 06/27/2025).

[17] Brandon Amos, Lei Xu, and J. Zico Kolter. “Input Convex Neural Networks”. In: *Proceedings of the 34th International Conference on Machine Learning*. International Conference on Machine Learning. ISSN: 2640-3498. PMLR,

July 17, 2017, pp. 146–155. URL: https://proceedings.mlr.press/v70/amos17b.html (visited on 06/27/2025).

[18] Brian D. O. Anderson. "Reverse-time diffusion equation models". In: *Stochastic Processes and their Applications* 12.3 (May 1, 1982), pp. 313–326. ISSN: 0304-4149. DOI: 10.1016/0304-4149(82)90051-5. URL: https://www.sciencedirect.com/science/article/pii/0304414982900515 (visited on 06/25/2025).

[19] Thor E. Andreassen et al. "Three Dimensional Lower Extremity Musculoskeletal Geometry of the Visible Human Female and Male". In: *Scientific Data* 10.1 (Jan. 18, 2023). Publisher: Nature Publishing Group, p. 34. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01905-2. URL: https://www.nature.com/articles/s41597-022-01905-2 (visited on 06/27/2025).

[20] J. W. Andrews, M. J. Adams, and T. D. Montenegro-Johnson. "A universal scaling law of mammalian touch". In: *Science Advances* 6.41 (Oct. 9, 2020). Publisher: American Association for the Advancement of Science, eabb6912. DOI: 10.1126/sciadv.abb6912. URL: https://www.science.org/doi/10.1126/sciadv.abb6912 (visited on 06/27/2025).

[21] Philipp Arras et al. "Variable structures in M87* from space, time and frequency resolved interferometry". In: *Nature Astronomy* 6.2 (Feb. 2022). Publisher: Nature Publishing Group, pp. 259–269. ISSN: 2397-3366. DOI: 10.1038/s41550-021-01548-0. URL: https://www.nature.com/articles/s41550-021-01548-0 (visited on 06/26/2025).

[22] Muhammad Asim et al. "Invertible generative models for inverse problems: mitigating representation error and dataset bias". In: *Proceedings of the International Conference on Machine Learning*. 2020.

[23] Danilo Avola et al. "Ultrasound Medical Imaging Techniques: A Survey". In: *ACM Comput. Surv.* 54.3 (Apr. 22, 2021), 67:1–67:38. ISSN: 0360-0300. DOI: 10.1145/3447243. URL: https://doi.org/10.1145/3447243 (visited on 06/27/2025).

[24] Arpit Bansal et al. "Universal Guidance for Diffusion Models". In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE Computer Society, June 1, 2023, pp. 843–852. ISBN: 979-8-3503-0249-3. DOI: 10.1109/CVPRW59228.2023.00091. URL: https://www.computer.org/csdl/proceedings-article/cvprw/2023/024900a843/1PBxrAeTzDq (visited on 06/27/2025).

[25] HARRY BATEMAN. "SOME RECENT RESEARCHES ON THE MOTION OF FLUIDS". In: *Monthly Weather Review* 43.4 (Apr. 1, 1915). Place: Boston MA, USA Publisher: American Meteorological Society, pp. 163–170. DOI: 10.1175/1520-0493(1915)43<163:SRROTM>2.0.CO;2. URL:

https://journals.ametsoc.org/view/journals/mwre/43/4/1520-0493_1915_43_163_srrotm_2_0_co_2.xml.

[26] Simon Batzner et al. "E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials". In: *Nature Communications* 13.1 (May 4, 2022). Publisher: Nature Publishing Group, p. 2453. ISSN: 2041-1723. DOI: 10.1038/s41467-022-29939-5. URL: https://www.nature.com/articles/s41467-022-29939-5 (visited on 06/27/2025).

[27] Amir Beck and Marc Teboulle. "A fast Iterative Shrinkage-Thresholding Algorithm with application to wavelet-based image deblurring". In: *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. ISSN: 2379-190X. Apr. 2009, pp. 693–696. DOI: 10.1109/ICASSP.2009.4959678. URL: https://ieeexplore.ieee.org/document/4959678 (visited on 07/28/2025).

[28] Amir Beck and Marc Teboulle. "Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems". In: *IEEE Transactions on Image Processing* 18.11 (Nov. 2009), pp. 2419–2434. ISSN: 1941-0042. DOI: 10.1109/TIP.2009.2028250. URL: https://ieeexplore.ieee.org/document/5173518 (visited on 07/28/2025).

[29] Amir Beck and Marc Teboulle. "Mirror descent and nonlinear projected subgradient methods for convex optimization". In: *Operations Research Letters* 31.3 (May 1, 2003), pp. 167–175. ISSN: 0167-6377. DOI: 10.1016/S0167-6377(02)00231-6. URL: https://www.sciencedirect.com/science/article/pii/S0167637702002316.

[30] Jens Behrmann et al. "Invertible Residual Networks". In: *Proceedings of the International Conference on Machine Learning*. Proceedings of Machine Learning Research. PMLR, 2019.

[31] Anthony J. Bell and Terrence J. Sejnowski. "The "independent components" of natural scenes are edge filters". In: *Vision Research* 37.23 (1997), pp. 3327–3338. ISSN: 0042-6989. DOI: https://doi.org/10.1016/S0042-6989(97)00121-1. URL: https://www.sciencedirect.com/science/article/pii/S0042698997001211.

[32] R. A. Bell. "Economics of MRI technology". In: *Journal of magnetic resonance imaging: JMRI* 6.1 (1996), pp. 10–25. ISSN: 1053-1807. DOI: 10.1002/jmri.1880060105.

[33] Sean Bell et al. "Material recognition in the wild with the Materials in Context Database". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 1063-6919. June 2015, pp. 3479–3487. DOI: 10.1109/CVPR.2015.7298970. URL: https://ieeexplore.ieee.org/document/7298970 (visited on 06/27/2025).

[34] Dimitri P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Google-Books-ID: j6LiBQAAQBAJ. Academic Press, May 10, 2014. 412 pp. ISBN: 978-1-4832-6047-1.

[35] Kiran S. Bhat et al. "Estimating cloth simulation parameters from video". In: *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*. SCA '03. Goslar, DEU: Eurographics Association, July 26, 2003, pp. 37–51. ISBN: 978-1-58113-659-3. (Visited on 06/27/2025).

[36] Marcin Białas and Bojan B. Guzina. "On the viscoelastic characterization of thin tissues via surface-wave sensing". In: *International Journal of Solids and Structures* 48.14 (July 1, 2011), pp. 2209–2217. ISSN: 0020-7683. DOI: 10.1016/j.ijsolstr.2011.03.025. URL: https://www.sciencedirect.com/science/article/pii/S0020768311001272.

[37] Bernd Bickel et al. "Design and fabrication of materials with desired deformation behavior". In: *ACM Trans. Graph.* 29.4 (July 26, 2010), 63:1–63:10. ISSN: 0730-0301. DOI: 10.1145/1778765.1778800. URL: https://dl.acm.org/doi/10.1145/1778765.1778800 (visited on 06/27/2025).

[38] Mikołaj Bińkowski et al. "Demystifying MMD GANs". In: *Proceedings of the International Conference on Learning Representations*. 2018. URL: https://openreview.net/forum?id=r1lUOzWCW.

[39] Lindy Blackburn et al. "Closure Statistics in Interferometric Data". In: *The Astrophysical Journal* 894.1 (May 2020). Publisher: The American Astronomical Society, p. 31. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ab8469. URL: https://dx.doi.org/10.3847/1538-4357/ab8469 (visited on 06/26/2025).

[40] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518 (Apr. 3, 2017). Publisher: ASA Website _eprint: https://doi.org/10.1080/01621459.2017.1285773, pp. 859–877. ISSN: 0162-1459. DOI: 10.1080/01621459.2017.1285773. URL: https://doi.org/10.1080/01621459.2017.1285773 (visited on 06/25/2025).

[41] Felix Bloch. "Über die Quantenmechanik der Elektronen in Kristallgittern". In: *Zeitschrift für Physik* 52.7 (July 1, 1929), pp. 555–600. ISSN: 0044-3328. DOI: 10.1007/BF01339455. URL: https://doi.org/10.1007/BF01339455 (visited on 06/27/2025).

[42] Guido Boffetta and Robert E. Ecke. "Two-Dimensional Turbulence". In: *Annual Review of Fluid Mechanics* 44 (Jan. 21, 2012). Publisher: Annual Reviews, pp. 427–451. ISSN: 0066-4189, 1545-4479. DOI: 10.1146/annurev-fluid-120710-101240. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-fluid-120710-101240 (visited on 06/28/2025).

[43] Ashish Bora et al. “Compressed sensing using generative models”. In: *Proceedings of the 34th International Conference on Machine Learning - Volume 70*. ICML’17. Sydney, NSW, Australia: JMLR.org, Aug. 6, 2017, pp. 537–546. (Visited on 07/03/2025).

[44] Valentin De Bortoli et al. “Riemannian Score-Based Generative Modelling”. In: *Advances in Neural Information Processing Systems*. Vol. 35. Oct. 31, 2022, pp. 2406–2422. URL: https://openreview.net/forum?id=oDRQGo8I7P (visited on 06/27/2025).

[45] Charles Bouman and Ken Sauer. “A generalized Gaussian image model for edge-preserving MAP estimation”. In: *IEEE Transactions on Image Processing* 2.3 (July 1993), pp. 296–310. ISSN: 1941-0042. DOI: 10.1109/83.236536. URL: https://ieeexplore.ieee.org/document/236536 (visited on 06/25/2025).

[46] Katherine L. Bouman et al. “Computational Imaging for VLBI Image Reconstruction”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016*. IEEE, 2016. DOI: 10.1109/CVPR.2016.105.

[47] Katherine L. Bouman et al. “Estimating the Material Properties of Fabric from Video”. In: *2013 IEEE International Conference on Computer Vision*. 2013 IEEE International Conference on Computer Vision. ISSN: 2380-7504. Dec. 2013, pp. 1984–1991. DOI: 10.1109/ICCV.2013.455. URL: https://ieeexplore.ieee.org/document/6751357 (visited on 06/27/2025).

[48] Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization*. Google-Books-ID: mYm0bLd3fcoC. Cambridge University Press, Mar. 8, 2004. 744 pp. ISBN: 978-0-521-83378-3.

[49] Denis Boyda et al. “Sampling using $\mathrm{SU}(N)$ gauge equivariant flows”. In: *Physical Review D* 103.7 (Apr. 20, 2021). Publisher: American Physical Society, p. 074504. DOI: 10.1103/PhysRevD.103.074504. URL: https://link.aps.org/doi/10.1103/PhysRevD.103.074504.

[50] Benjamin Boys et al. *Tweedie Moment Projected Diffusions For Inverse Problems*. Sept. 25, 2024. DOI: 10.48550/arXiv.2310.06721. arXiv: 2310.06721[stat]. URL: http://arxiv.org/abs/2310.06721 (visited on 07/03/2025).

[51] Francois-Xavier Briol et al. *Statistical Inference for Generative Models with Maximum Mean Discrepancy*. June 13, 2019. DOI: 10.48550/arXiv.1906.05944. arXiv: 1906.05944[stat]. URL: http://arxiv.org/abs/1906.05944 (visited on 07/30/2025).

[52] Avery E. Broderick and Abraham Loeb. “Imaging bright-spots in the accretion flow near the black hole horizon of Sgr A*”. In: *Monthly Notices of the Royal Astronomical Society* 363.2 (Oct. 21, 2005), pp. 353–362.

ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09458.x. URL: https://doi.org/10.1111/j.1365-2966.2005.09458.x (visited on 07/09/2025).

[53] Avery E. Broderick and Dominic W. Pesce. "Closure Traces: Novel Calibration-insensitive Quantities for Radio Astronomy". In: *The Astrophysical Journal* 904.2 (Nov. 2020). Publisher: The American Astronomical Society, p. 126. ISSN: 0004-637X. DOI: 10.3847/1538-4357/abbd9d. URL: https://dx.doi.org/10.3847/1538-4357/abbd9d (visited on 06/26/2025).

[54] Avery E. Broderick et al. "Evidence for Low Black Hole Spin and Physically Motivated Accretion Models from Millimeter-VLBI Observations of Sagittarius A*". In: *The Astrophysical Journal* 735.2 (June 2011). Publisher: The American Astronomical Society, p. 110. ISSN: 0004-637X. DOI: 10.1088/0004-637X/735/2/110. URL: https://dx.doi.org/10.1088/0004-637X/735/2/110 (visited on 06/26/2025).

[55] Steve Brooks et al. *Handbook of Markov Chain Monte Carlo*. 1st ed. New York: Chapman and Hall/CRC, May 9, 2011. 619 pp. ISBN: 978-0-429-13850-8.

[56] Sébastien Bubeck. "Convex Optimization: Algorithms and Complexity". In: *Found. Trends Mach. Learn.* 8.3 (Nov. 12, 2015), pp. 231–357. ISSN: 1935-8237. DOI: 10.1561/2200000050. URL: https://doi.org/10.1561/2200000050 (visited on 06/27/2025).

[57] Nathalie J. Bureau and Daniela Ziegler. "Economics of Musculoskeletal Ultrasound". In: *Current Radiology Reports* 4 (2016), p. 44. ISSN: 2167-4825. DOI: 10.1007/s40134-016-0169-5. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4914528/ (visited on 06/27/2025).

[58] J.M. Burgers. "A Mathematical Model Illustrating the Theory of Turbulence". In: *Advances in Applied Mechanics*. Ed. by Richard Von Mises and Theodore Von Kármán. Vol. 1. Elsevier, Jan. 1, 1948, pp. 171–199. DOI: 10.1016/S0065-2156(08)70100-5. URL: https://www.sciencedirect.com/science/article/pii/S0065215608701005.

[59] Emmanuel Candès and Justin Romberg. "Sparsity and incoherence in compressive sampling". In: *Inverse Problems* 23.3 (Apr. 2007), p. 969. ISSN: 0266-5611. DOI: 10.1088/0266-5611/23/3/008. URL: https://dx.doi.org/10.1088/0266-5611/23/3/008 (visited on 06/25/2025).

[60] Alejandro Cárdenas-Avendaño, Alexandru Lupsasca, and Hengrui Zhu. "Adaptive analytical ray tracing of black hole photon rings". In: *Physical Review D* 107.4 (Feb. 22, 2023). Publisher: American Physical Society, p. 043030. DOI: 10.1103/PhysRevD.107.043030. URL: https://link.aps.org/doi/10.1103/PhysRevD.107.043030 (visited on 07/10/2025).

[61] Gabriel Cardoso et al. "Monte Carlo guided Denoising Diffusion models for Bayesian linear inverse problems." In: The Twelfth International Conference on Learning Representations. Oct. 13, 2023. URL: https://openreview.net/forum?id=nHESwXvxWK (visited on 07/03/2025).

[62] Andrew Chael. "Simulating and Imaging Supermassive Black Hole Accretion Flows". In: American Astronomical Society Meeting Abstracts #235. Vol. 235. ADS Bibcode: 2020AAS...23541103C. Jan. 1, 2020, p. 411.03. URL: https://ui.adsabs.harvard.edu/abs/2020AAS...23541103C (visited on 06/27/2025).

[63] Andrew Chael et al. "Black Hole Polarimetry I. A Signature of Electromagnetic Energy Extraction". In: *The Astrophysical Journal* 958.1 (Nov. 14, 2023). Publisher: The American Astronomical Society, p. 65. ISSN: 0004-637X. DOI: 10.3847/1538-4357/acf92d. URL: https://dx.doi.org/10.3847/1538-4357/acf92d.

[64] Andrew A. Chael et al. "High-resolution Linear Polarimetric Imaging for the Event Horizon Telescope". In: *The Astrophysical Journal* 829.1 (Sept. 2016). Publisher: The American Astronomical Society, p. 11. ISSN: 0004-637X. DOI: 10.3847/0004-637X/829/1/11. URL: https://dx.doi.org/10.3847/0004-637X/829/1/11 (visited on 06/25/2025).

[65] Andrew A. Chael et al. "Interferometric Imaging Directly with Closure Phases and Closure Amplitudes". In: *The Astrophysical Journal* 857.1 (Apr. 2018). Publisher: The American Astronomical Society, p. 23. ISSN: 0004-637X. DOI: 10.3847/1538-4357/aab6a8. URL: https://dx.doi.org/10.3847/1538-4357/aab6a8 (visited on 06/25/2025).

[66] Chi-kwan Chan, Dimitrios Psaltis, and Feryal Özel. "GRay: A MASSIVELY PARALLEL GPU-BASED CODE FOR RAY TRACING IN RELATIVISTIC SPACETIMES". In: *The Astrophysical Journal* 777.1 (Oct. 2013). Publisher: The American Astronomical Society, p. 13. ISSN: 0004-637X. DOI: 10.1088/0004-637X/777/1/13. URL: https://dx.doi.org/10.1088/0004-637X/777/1/13 (visited on 07/09/2025).

[67] Gary J. Chandler and Rich R. Kerswell. "Invariant recurrent solutions embedded in a turbulent two-dimensional Kolmogorov flow". In: *Journal of Fluid Mechanics* 722 (2013). Edition: 2013/03/28 Publisher: Cambridge University Press, pp. 554–595. ISSN: 0022-1120. DOI: 10.1017/jfm.2013.122. URL: https://www.cambridge.org/core/product/78CC6B29A670F84CBC79D29408DC2674.

[68] Ming-Wei Chang, Lev Ratinov, and Dan Roth. "Guiding Semi-Supervision with Constraint-Driven Learning". In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. ACL 2007. Ed. by Annie Zaenen and Antal van den Bosch. Prague, Czech Republic: Association for Computational Linguistics, June 2007, pp. 280–287. URL: https://aclanthology.org/P07-1036/ (visited on 06/27/2025).

[69] Justin Chen, Robert Haupt, and O. Büyükoztürk. "The Acoustic-laser Vibrometry Technique for the Noncontact Detection of Discontinuities in Fiber Reinforced Polymer-retrofitted Concrete". In: *Materials Evaluation* 72 (Oct. 2014), pp. 1305–1313.

[70] Justin G. Chen et al. "Structural Modal Identification Through High Speed Camera Video: Motion Magnification". In: *Topics in Modal Analysis I, Volume 7*. Ed. by James De Clerck. Cham: Springer International Publishing, 2014, pp. 191–197. ISBN: 978-3-319-04753-9. DOI: 10.1007/978-3-319-04753-9_19.

[71] Ricky T. Q. Chen et al. "Neural ordinary differential equations". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS'18. Red Hook, NY, USA: Curran Associates Inc., Dec. 3, 2018, pp. 6572–6583. (Visited on 06/25/2025).

[72] Yongxin Chen et al. "Improved analysis for a proximal algorithm for sampling". In: *Proceedings of Thirty Fifth Conference on Learning Theory*. Conference on Learning Theory. ISSN: 2640-3498. PMLR, June 28, 2022, pp. 2984–3014. URL: https://proceedings.mlr.press/v178/chen22c.html (visited on 07/03/2025).

[73] Ziyang Chen, Shengyi Qian, and Andrew Owens. "Sound Localization from Motion: Jointly Learning Sound Direction and Camera Rotation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023*. IEEE, 2023.

[74] Jooyoung Choi et al. "ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2021, pp. 14347–14356. DOI: 10.1109/ICCV48922.2021.01410. URL: https://ieeexplore.ieee.org/document/9711284 (visited on 06/25/2025).

[75] Jacob K. Christopher, Stephen Baek, and Ferdinando Fioretto. "Constrained Synthesis with Projected Diffusion Models". In: *Advances in Neural Information Processing Systems*. The Thirty-eighth Annual Conference on Neural Information Processing Systems. Nov. 6, 2024. URL: https://openreview.net/forum?id=FsdB3I9Y24&referrer=%5Bthe%20profile%20of%20Ferdinando%20Fioretto%5D(%2Fprofile%3Fid%3D~Ferdinando_Fioretto1) (visited on 06/27/2025).

[76] T. C. Chu, W. F. Ranson, and M. A. Sutton. "Applications of digital-image-correlation techniques to experimental mechanics". In: *Experimental Mechanics* 25.3 (Sept. 1, 1985), pp. 232–244. ISSN: 1741-2765. DOI: 10.1007/BF02325092. URL: https://doi.org/10.1007/BF02325092 (visited on 06/27/2025).

[77] Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. "Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022, pp. 12413–12422.

[78] Hyungjin Chung and Jong Chul Ye. "Score-based diffusion models for accelerated MRI". In: *Medical Image Analysis* 80 (2022), p. 102479. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2022.102479. URL: https://www.sciencedirect.com/science/article/pii/S1361841522001268.

[79] Hyungjin Chung et al. "Diffusion Posterior Sampling for General Noisy Inverse Problems". In: *Proceedings of the Eleventh International Conference on Learning Representations*. The Eleventh International Conference on Learning Representations. Kigali, Rwanda, 2023. URL: https://openreview.net/forum?id=OnD9zGAGT0k.

[80] Hyungjin Chung et al. "Improving Diffusion Models for Inverse Problems using Manifold Constraints". In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. NIPS '22. Red Hook, NY, USA: Curran Associates Inc., Nov. 28, 2022, pp. 25683–25696. ISBN: 978-1-7138-7108-8. (Visited on 06/25/2025).

[81] Pieter Hendrik van Cittert. "Die Wahrscheinliche Schwingungsverteilung in Einer von Einer Lichtquelle Direkt Oder Mittels Einer Linse Beleuchteten Ebene". In: *Physica* 1.1 (Jan. 1, 1934), pp. 201–210. ISSN: 0031-8914. DOI: 10.1016/S0031-8914(34)90026-4. URL: https://www.sciencedirect.com/science/article/pii/S0031891434900264 (visited on 06/25/2025).

[82] B. G. Clark. "An efficient implementation of the algorithm 'CLEAN'". In: *Astronomy and Astrophysics* 89 (Sept. 1, 1980). Publisher: EDP ADS Bibcode: 1980A&A....89..377C, p. 377. ISSN: 0004-6361. URL: https://ui.adsabs.harvard.edu/abs/1980A&A....89..377C (visited on 06/25/2025).

[83] *COMSOL Multiphysics*. Version 5.5. URL: https://comsol.com.

[84] Liam Connor et al. "Deep Radio Interferometric Imaging with POLISH: DSA-2000 and weak lensing". In: *Monthly Notices of the Royal Astronomical Society* 514.2 (2022), pp. 2614–2626.

[85] Tim Cornwell, Robert Braun, and Daniel S. Briggs. "Deconvolution". In: *Synthesis Imaging in Radio Astronomy II, A Collection of Lectures from the Sixth NRAO/NMIMT Synthesis Imaging Summer School* 180 (1999), p. 151. URL: https://adsabs.harvard.edu/full/1999ASPC..180..151C.

[86] J. Crank and P. Nicolson. “A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 43.1 (1947). Edition: 2008/10/24 Publisher: Cambridge University Press, pp. 50–67. ISSN: 0305-0041. DOI: 10.1017/S0305004100023197. URL: https://www.cambridge.org/core/product/B3230893A53384D418228AB39D41A451.

[87] C. T. Cunningham. “The effects of redshifts and focusing on the spectrum of an accretion disk around a Kerr black hole.” In: *The Astrophysical Journal* 202 (Dec. 1, 1975). Publisher: IOP ADS Bibcode: 1975ApJ...202..788C, pp. 788–802. ISSN: 0004-637X. DOI: 10.1086/154033. URL: https://ui.adsabs.harvard.edu/abs/1975ApJ...202..788C (visited on 07/09/2025).

[88] Giannis Daras et al. “Consistent Diffusion Models: Mitigating Sampling Drift by Learning to be Consistent”. In: Thirty-seventh Conference on Neural Information Processing Systems. Nov. 2, 2023. URL: https://openreview.net/forum?id=GfZGdJHj27 (visited on 06/27/2025).

[89] Abe Davis, Justin G. Chen, and Frédo Durand. “Image-space modal bases for plausible manipulation of objects in video”. In: *ACM Trans. Graph.* 34.6 (Nov. 2, 2015), 239:1–239:7. ISSN: 0730-0301. DOI: 10.1145/2816795.2818095. URL: https://dl.acm.org/doi/10.1145/2816795.2818095 (visited on 06/27/2025).

[90] Abe Davis et al. “The visual microphone: passive recovery of sound from video”. In: *ACM Trans. Graph.* 33.4 (July 27, 2014), 79:1–79:10. ISSN: 0730-0301. DOI: 10.1145/2601097.2601119. URL: https://dl.acm.org/doi/10.1145/2601097.2601119 (visited on 06/27/2025).

[91] Abe Davis et al. “Visual vibrometry: Estimating material properties from small motions in video”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 1063-6919. June 2015, pp. 5335–5343. DOI: 10.1109/CVPR.2015.7299171. URL: https://ieeexplore.ieee.org/document/7299171 (visited on 06/27/2025).

[92] Abe Davis* et al. “Visual Vibrometry: Estimating Material Properties from Small Motions in Video”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4 (Apr. 2017), pp. 732–745. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2016.2622271. URL: https://ieeexplore.ieee.org/document/7728146 (visited on 06/27/2025).

[93] Juan Carlos De Los Reyes. *Numerical PDE-Constrained Optimization*. SpringerBriefs in Optimization. Cham: Springer International Publishing, 2015. ISBN: 978-3-319-13395-9. DOI: 10.1007/978-3-319-13395-9. URL: https://link.springer.com/10.1007/978-3-319-13395-9 (visited on 07/28/2025).

[94] Mauricio Delbracio and Peyman Milanfar. "Inversion by Direct Iteration: An Alternative to Denoising Diffusion for Image Restoration". In: *Transactions on Machine Learning Research* 2835 (2023). URL: https://openreview.net/forum?id=VmyFF5lL3F.

[95] Mauricio Delbracio, Hossein Talebei, and Pevman Milanfar. "Projected Distribution Loss for Image Enhancement". In: *2021 IEEE International Conference on Computational Photography (ICCP)*. 2021 IEEE International Conference on Computational Photography (ICCP). ISSN: 2472-7636. May 2021, pp. 1–12. DOI: 10.1109/ICCP51581.2021.9466271. URL: https://ieeexplore.ieee.org/document/9466271 (visited on 06/25/2025).

[96] Jason Dexter. "A public code for general relativistic, polarised radiative transfer around spinning black holes". In: *Monthly Notices of the Royal Astronomical Society* 462.1 (Oct. 11, 2016), pp. 115–136. ISSN: 0035-8711. DOI: 10.1093/mnras/stw1526. URL: https://doi.org/10.1093/mnras/stw1526 (visited on 07/09/2025).

[97] Neel Dey, Antong Chen, and Soheil Ghafurian. "Group Equivariant Generative Adversarial Networks". In: *Proceedings of the International Conference on Learning Representations*. 2021. URL: https://openreview.net/forum?id=rgFNuJHHXv.

[98] Sreemanti Dey et al. "Score-Based Diffusion Models for Photoacoustic Tomography Image Reconstruction". In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 2470–2474.

[99] Prafulla Dhariwal and Alex Nichol. "Diffusion Models Beat GANs on Image Synthesis". In: *Advances in Neural Information Processing Systems*. 2021.

[100] Laurent Dinh, David Krueger, and Yoshua Bengio. *NICE: Non-linear Independent Components Estimation*. 2014. DOI: https://doi.org/10.48550/arXiv.1410.8516. URL: https://arxiv.org/abs/1410.8516.

[101] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using RealNVP". In: *Proceedings of the International Conference on Learning Representations*. 2017. URL: https://openreview.net/forum?id=HkpbnH9lx.

[102] J. R. Dormand and P. J. Prince. "A family of embedded Runge-Kutta formulae". In: *Journal of Computational and Applied Mathematics* 6.1 (Mar. 1, 1980), pp. 19–26. ISSN: 0377-0427. DOI: 10.1016/0771-050X(80)90013-3. URL: https://www.sciencedirect.com/science/article/pii/0771050X80900133 (visited on 06/25/2025).

[103] Zehao Dou and Yang Song. "Diffusion Posterior Sampling for Linear Inverse Problem Solving: A Filtering Perspective". In: The Twelfth International Conference on Learning Representations. Oct. 13, 2023. URL: https://openreview.net/forum?id=tplXNcHZs1 (visited on 07/03/2025).

[104] F. Durst et al. “Principles and Practice of Laser-Doppler Anemometry”. In: *Journal of Applied Mechanics* 44.3 (Sept. 1, 1977), p. 518. ISSN: 0021-8936. DOI: 10.1115/1.3424128. URL: https://doi.org/10.1115/1.3424128 (visited on 06/27/2025).

[105] Bradley Efron. “Tweedie’s Formula and Selection Bias”. In: *Journal of the American Statistical Association* 106.496 (Dec. 1, 2011). Publisher: ASA Website, pp. 1602–1614. ISSN: 0162-1459. DOI: 10.1198/jasa.2011.tm11181. URL: https://doi.org/10.1198/jasa.2011.tm11181.

[106] Edward H. Egelman. “The Current Revolution in Cryo-EM”. In: *Biophysical Journal* 110.5 (Mar. 8, 2016), pp. 1008–1012. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2016.02.001. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4788751/ (visited on 07/01/2025).

[107] V. Egorov et al. “Soft tissue elastometer”. In: *Medical Engineering & Physics* 30.2 (Mar. 2008), pp. 206–212. ISSN: 1350-4533. DOI: 10.1016/j.medengphy.2007.02.007.

[108] Timothy James Emge. “Remote nondestructive evaluation of composite-steel interface by acoustic laser vibrometry”. Accepted: 2012-11-19T19:16:54Z. Thesis. Massachusetts Institute of Technology, 2012. URL: https://dspace.mit.edu/handle/1721.1/74902 (visited on 06/27/2025).

[109] Heino Falcke, Fulvio Melia, and Eric Agol. “Viewing the Shadow of the Black Hole at the GalacticCenter”. In: *The Astrophysical Journal* 528.1 (Dec. 7, 1999). Publisher: IOP Publishing, p. L13. ISSN: 0004-637X. DOI: 10.1086/312423. URL: https://iopscience.iop.org/article/10.1086/312423/meta (visited on 06/25/2025).

[110] U. Fasel et al. “Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 478.2260 (Apr. 13, 2022). Publisher: Royal Society, p. 20210904. DOI: 10.1098/rspa.2021.0904. URL: https://royalsocietypublishing.org/doi/full/10.1098/rspa.2021.0904 (visited on 08/19/2025).

[111] Abdul Fatir. *abdulfatir/gan-metrics-pytorch*. original-date: 2019-02-25T09:18:42Z. June 20, 2025. URL: https://github.com/abdulfatir/gan-metrics-pytorch (visited on 07/30/2025).

[112] J. Feiteira et al. “Monitoring crack movement in polymer-based self-healing concrete through digital image correlation, acoustic emission analysis and SEM in-situ loading”. In: *Materials & Design* 115 (Feb. 5, 2017), pp. 238–246. ISSN: 0264-1275. DOI: 10.1016/j.matdes.2016.11.050. URL: https://www.sciencedirect.com/science/article/pii/S0264127516314368 (visited on 06/27/2025).

[113] Berthy T. Feng. "EHT Black-Hole Imaging". Tutorial. Tutorial. ICCP Summer School. Toronto, ON, CA, July 19, 2025. URL: https://github.com/berthyf96/eht_imaging_tutorial.

[114] Berthy T. Feng, Ricardo Baptista, and Katherine L. Bouman. "Neural Approximate Mirror Maps for Constrained Diffusion Models". In: The Thirteenth International Conference on Learning Representations. Singapore, 2025. URL: https://openreview.net/forum?id=vgZDcUetWS.

[115] Berthy T. Feng and Katherine L. Bouman. "Seeing Beyond the Blur with Generative AI". In: *XRDS* 31.2 (Jan. 8, 2025), pp. 28–31. ISSN: 1528-4972. DOI: 10.1145/3703400. URL: https://dl.acm.org/doi/10.1145/3703400.

[116] Berthy T. Feng and Katherine L. Bouman. "Variational Bayesian Imaging with an Efficient Surrogate Score-based Prior". In: *Transactions on Machine Learning Research* (Mar. 2, 2024). ISSN: 2835-8856. URL: https://openreview.net/forum?id=db2pFKVcm1.

[117] Berthy T. Feng, Katherine L. Bouman, and William T. Freeman. "Event-horizon-scale Imaging of M87* under Different Assumptions via Deep Generative Image Priors". In: *The Astrophysical Journal* 975.2 (Nov. 2024). Publisher: The American Astronomical Society, p. 201. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ad737f. URL: https://dx.doi.org/10.3847/1538-4357/ad737f.

[118] Berthy T. Feng et al. "Score-Based Diffusion Models as Principled Priors for Inverse Imaging". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (ICCV)*. Paris, France: IEEE, 2023. DOI: 10.1109/ICCV51070.2023.00965.

[119] Berthy T. Feng et al. "Visual Vibration Tomography: Estimating Interior Material Properties from Monocular Video". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022*. ISSN: 2575-7075. IEEE, June 2022, pp. 16210–16219. DOI: 10.1109/CVPR52688.2022.01575. URL: https://ieeexplore.ieee.org/document/9880380.

[120] Chao Feng, Ziyang Chen, and Andrew Owens. "Self-Supervised Video Forensics by Audio-Visual Anomaly Detection". In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, June 1, 2023, pp. 10491–10503. ISBN: 979-8-3503-0129-8. DOI: 10.1109/CVPR52729.2023.01011. URL: https://www.computer.org/csdl/proceedings-article/cvpr/2023/012900k0491/1POQOW8pCuI (visited on 06/27/2025).

[121] Nic Fishman et al. "Diffusion Models for Constrained Domains". In: *Transactions on Machine Learning Research* 2835 (2023). URL: https://openreview.net/forum?id=xuWTFQ4VGO.

[122] Nic Fishman et al. “Metropolis sampling for constrained diffusion models”. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS ’23. Red Hook, NY, USA: Curran Associates Inc., Dec. 10, 2023, pp. 62296–62331. (Visited on 06/27/2025).

[123] David J. Fleet and Allan D. Jepson. “Computation of component image velocity from local phase information”. In: *International Journal of Computer Vision* 5.1 (Aug. 1, 1990), pp. 77–104. ISSN: 1573-1405. DOI: 10.1007/BF00056772. URL: https://doi.org/10.1007/BF00056772.

[124] Gaston Floquet. “Sur les équations différentielles linéaires à coefficients périodiques”. In: *Annales scientifiques de l’École Normale Supérieure, Serie 2* 12 (1883), pp. 47–88.

[125] Dora Foti et al. “Ambient vibration testing, dynamic identification and model updating of a historic tower”. In: *NDT & E International* 47 (Apr. 1, 2012), pp. 88–95. ISSN: 0963-8695. DOI: 10.1016/j.ndteint.2011.11.009. URL: https://www.sciencedirect.com/science/article/pii/S0963869511001745.

[126] Guy Gafni et al. “Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction”. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 2575-7075. June 2021, pp. 8645–8654. DOI: 10.1109/CVPR46437.2021.00854. URL: https://ieeexplore.ieee.org/document/9578714 (visited on 07/11/2025).

[127] Yarin Gal and Zoubin Ghahramani. *Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference*. ADS Bibcode: 2015arXiv150602158G. June 1, 2015. DOI: 10.48550/arXiv.1506.02158. URL: https://ui.adsabs.harvard.edu/abs/2015arXiv150602158G (visited on 06/25/2025).

[128] Kuzman Ganchev et al. “Posterior Regularization for Structured Latent Variable Models”. In: *Journal of Machine Learning Research* 11.67 (2010), pp. 2001–2049. ISSN: 1533-7928. URL: http://jmlr.org/papers/v11/ganchev10a.html (visited on 06/27/2025).

[129] Chen Gao et al. “Dynamic View Synthesis from Dynamic Monocular Video”. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2021, pp. 5692–5701. DOI: 10.1109/ICCV48922.2021.00566. URL: https://ieeexplore.ieee.org/document/9709972 (visited on 07/11/2025).

[130] Hang Gao et al. “Monocular dynamic view synthesis: a reality check”. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. NIPS ’22. Red Hook, NY, USA: Curran Associates

Inc., Nov. 28, 2022, pp. 33768–33780. ISBN: 978-1-7138-7108-8. (Visited on 07/10/2025).

[131] Damien Garreau, Wittawat Jitkrittum, and Motonobu Kanagawa. *Large sample analysis of the median heuristic*. Oct. 30, 2018. DOI: 10.48550/arXiv.1707.07269. arXiv: 1707.07269[math]. URL: http://arxiv.org/abs/1707.07269 (visited on 07/30/2025).

[132] Karl Gebhardt et al. "The Black Hole Mass in M87 from Gemini/NIFS Adaptive Optics Observations". In: *The Astrophysical Journal* 729.2 (Feb. 2011). Publisher: The American Astronomical Society, p. 119. ISSN: 0004-637X. DOI: 10.1088/0004-637X/729/2/119. URL: https://dx.doi.org/10.1088/0004-637X/729/2/119 (visited on 06/26/2025).

[133] Mario Geiger and Tess Smidt. *e3nn: Euclidean Neural Networks*. July 18, 2022. DOI: 10.48550/arXiv.2207.09453. arXiv: 2207.09453[cs]. URL: http://arxiv.org/abs/2207.09453 (visited on 06/27/2025).

[134] Freija Geldof et al. "Layer thickness prediction and tissue classification in two-layered tissue structures using diffuse reflectance spectroscopy". In: *Scientific Reports* 12.1 (Feb. 1, 2022). Publisher: Nature Publishing Group, p. 1698. ISSN: 2045-2322. DOI: 10.1038/s41598-022-05751-5. URL: https://www.nature.com/articles/s41598-022-05751-5 (visited on 06/27/2025).

[135] Penelope C. Georges et al. "Increased stiffness of the rat liver precedes matrix deposition: implications for fibrosis". In: *American Journal of Physiology. Gastrointestinal and Liver Physiology* 293.6 (Dec. 2007), G1147–1154. ISSN: 0193-1857. DOI: 10.1152/ajpgi.00032.2007.

[136] Mathieu Germain et al. "MADE: Masked Autoencoder for Distribution Estimation". In: *Proceedings of the 32nd International Conference on Machine Learning*. International Conference on Machine Learning. ISSN: 1938-7228. PMLR, June 1, 2015, pp. 881–889. URL: https://proceedings.mlr.press/v37/germain15.html (visited on 06/25/2025).

[137] Giorgio Giannone et al. *Learning from Invalid Data: On Constraint Satisfaction in Generative Models*. 2023. DOI: https://doi.org/10.48550/arXiv.2306.15166. URL: https://arxiv.org/abs/2306.15166.

[138] Ian Goodfellow et al. "Generative adversarial networks". In: *Commun. ACM* 63.11 (Oct. 22, 2020), pp. 139–144. ISSN: 0001-0782. DOI: 10.1145/3422622. URL: https://dl.acm.org/doi/10.1145/3422622 (visited on 06/25/2025).

[139] Alexandros Graikos et al. "Diffusion Models as Plug-and-Play Priors". In: *Advances in Neural Information Processing Systems 35*. Conference on Neural Information Processing Systems (NeurIPS). New Orleans, USA, Oct. 31, 2022. URL: https://openreview.net/forum?id=yhlMZ3iR7Pu (visited on 06/25/2025).

[140] Will Grathwohl et al. "FFJORD: Free-Form Continuous Dynamics for Scalable Reversible Generative Models". In: *Proceedings of the International Conference on Learning Representations*. 2019. URL: https://openreview.net/forum?id=rJxgknCcK7.

[141] Arthur Gretton et al. "A Kernel Two-Sample Test". In: *Journal of Machine Learning Research* 13.25 (2012), pp. 723–773. ISSN: 1533-7928. URL: http://jmlr.org/papers/v13/gretton12a.html (visited on 06/28/2025).

[142] N. Gucunski and R.D. Woods. "Numerical simulation of the SASW test". In: *Soil Dynamics and Earthquake Engineering* 11.4 (Jan. 1, 1992), pp. 213–227. ISSN: 0267-7261. DOI: 10.1016/0267-7261(92)90036-D. URL: https://www.sciencedirect.com/science/article/pii/026772619290036D.

[143] S.F. Gull and J. Skilling. "Maximum entropy method in image processing". In: *IEE Proceedings F (Communications, Radar and Signal Processing)* 131.6 (Oct. 1984). Publisher: The Institution of Engineering and Technology, pp. 646–659. DOI: 10.1049/ip-f-1.1984.0099. URL: https://digital-library.theiet.org/doi/10.1049/ip-f-1.1984.0099 (visited on 07/01/2025).

[144] Shivam Gupta et al. *Diffusion Posterior Sampling is Computationally Intractable*. Feb. 20, 2024. DOI: 10.48550/arXiv.2402.12727. arXiv: 2402.12727[cs]. URL: http://arxiv.org/abs/2402.12727 (visited on 07/04/2025).

[145] N. S. Ha, H. M. Vang, and N. S. Goo. "Modal Analysis Using Digital Image Correlation Technique: An Application to Artificial Wing Mimicking Beetle's Hind Wing". In: *Experimental Mechanics* 55.5 (June 1, 2015), pp. 989–998. ISSN: 1741-2765. DOI: 10.1007/s11340-015-9987-2. URL: https://doi.org/10.1007/s11340-015-9987-2.

[146] J. P. Hamaker, J. D. Bregman, and R. J. Sault. "Understanding radio polarimetry. I. Mathematical foundations." In: *Astronomy and Astrophysics Supplement Series* 117 (May 1, 1996). ADS Bibcode: 1996A&AS..117..137H, pp. 137–147. ISSN: 0365-0138. URL: https://ui.adsabs.harvard.edu/abs/1996A&AS..117..137H (visited on 07/05/2025).

[147] Paul Hand, Oscar Leong, and Vlad Voroninski. "Phase Retrieval Under a Generative Prior". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018. URL: https://papers.nips.cc/paper_files/paper/2018/hash/1bc2029a8851ad344a8d503930dfd7f7-Abstract.html (visited on 07/04/2025).

[148] Paul Hand and Vladislav Voroninski. "Global Guarantees for Enforcing Deep Generative Priors by Empirical Risk". In: *IEEE Transactions on Information Theory* 66.1 (Jan. 2020), pp. 401–418. ISSN: 1557-9654. DOI: 10.1109/TIT.2019.2935447. URL: https://ieeexplore.ieee.org/document/8801854 (visited on 07/04/2025).

[149] Yunus Emre Harmanci et al. “A Novel Approach for 3D-Structural Identification through Video Recording: Magnified Tracking”. In: *Sensors* 19.5 (Jan. 2019). Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 1229. ISSN: 1424-8220. DOI: 10.3390/s19051229. URL: https://www.mdpi.com/1424-8220/19/5/1229 (visited on 06/27/2025).

[150] Ray H. Hashemi, William G. Bradley, and Christopher J. Lisanti. *MRI: The Basics*. Google-Books-ID: v4LFgAHxNz4C. Lippincott Williams & Wilkins, 2010. 400 pp. ISBN: 978-1-60831-115-6.

[151] J. D. Helm. “Digital Image Correlation for Specimens with Multiple Growing Cracks”. In: *Experimental Mechanics* 48.6 (Dec. 1, 2008), pp. 753–762. ISSN: 1741-2765. DOI: 10.1007/s11340-007-9120-2. URL: https://doi.org/10.1007/s11340-007-9120-2 (visited on 06/27/2025).

[152] Martin Heusel et al. “GANs trained by a two time-scale update rule converge to a local nash equilibrium”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Red Hook, NY, USA: Curran Associates Inc., Dec. 4, 2017, pp. 6629–6640. ISBN: 978-1-5108-6096-4. (Visited on 07/29/2025).

[153] Irina Higgins et al. “β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework”. In: International Conference on Learning Representations. Feb. 6, 2017. URL: https://openreview.net/forum?id=Sy2fzU9gl (visited on 07/30/2025).

[154] R. A. Hinder. “Observations of Atmospheric Turbulence with a Radio Telescope at 5 GHz”. In: *Nature* 225.5233 (Feb. 1, 1970), pp. 614–617. ISSN: 1476-4687. DOI: 10.1038/225614a0. URL: https://doi.org/10.1038/225614a0.

[155] Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising Diffusion Probabilistic Models”. In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 6840–6851.

[156] Jonathan Ho et al. “Flow++: Improving Flow-Based Generative Models with Variational Dequantization and Architecture Design”. In: *Proceedings of the 36th International Conference on Machine Learning*. Proceedings of Machine Learning Research. PMLR, 2019, pp. 2722–2730. URL: https://proceedings.mlr.press/v97/ho19a.html.

[157] Yun-Xian Ho, Michael S. Landy, and Laurence T. Maloney. “How direction of illumination affects visually perceived surface roughness”. In: *Journal of Vision* 6.5 (Apr. 5, 2006), p. 8. ISSN: 1534-7362. DOI: 10.1167/6.5.8. URL: https://doi.org/10.1167/6.5.8 (visited on 06/27/2025).

[158] Henk Hoekstra and Bhuvnesh Jain. “Weak Gravitational Lensing and Its Cosmological Applications”. In: *Annual Review of Nuclear and Particle Science* 58 (Volume 58, 2008 Nov. 23, 2008). Publisher: Annual Reviews, pp. 99–123. ISSN: 0163-8998, 1545-4134. DOI: 10.1146/annurev.

nucl.58.110707.171151. URL: https://www.annualreviews.org/content/journals/10.1146/annurev.nucl.58.110707.171151 (visited on 07/11/2025).

[159] J. A. Högbom. "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines". In: *Astronomy and Astrophysics Supplement Series* 15 (June 1, 1974). ADS Bibcode: 1974A&AS...15..417H, p. 417. ISSN: 0365-01380004-6361. URL: https://ui.adsabs.harvard.edu/abs/1974A&AS...15..417H (visited on 06/25/2025).

[160] Jacob Hollingsworth et al. "Efficient sampling of constrained high-dimensional theoretical spaces with machine learning". In: *The European Physical Journal C* 81.12 (Dec. 26, 2021), p. 1138. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09941-9. URL: https://doi.org/10.1140/epjc/s10052-021-09941-9 (visited on 07/28/2025).

[161] Emiel Hoogeboom et al. "Equivariant Diffusion for Molecule Generation in 3D". In: *Proceedings of the International Conference on Machine Learning*. 2022, pp. 8867–8887.

[162] Kenneth Hoyt et al. "Tissue elasticity properties as biomarkers for prostate cancer". In: *Cancer Biomarkers: Section A of Disease Markers* 4.4 (2008), pp. 213–225. ISSN: 1574-0153. DOI: 10.3233/cbm-2008-44-505.

[163] Ya-Ping Hsieh et al. "Mirrored langevin dynamics". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS'18. Red Hook, NY, USA: Curran Associates Inc., Dec. 3, 2018, pp. 2883–2892. (Visited on 06/27/2025).

[164] Yixin Hu et al. "Tetrahedral meshing in the wild". In: *ACM Trans. Graph.* 37.4 (July 30, 2018), 60:1–60:14. ISSN: 0730-0301. DOI: 10.1145/3197517.3201353. URL: https://dl.acm.org/doi/10.1145/3197517.3201353 (visited on 06/27/2025).

[165] Chin-Wei Huang et al. "Riemannian Diffusion Models". In: *Advances in Neural Information Processing Systems*. Vol. 35. 2022, pp. 2750–2761.

[166] Hao Huang et al. "Emerging Wearable Ultrasound Technology". In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 71.7 (July 2024), pp. 713–729. ISSN: 1525-8955. DOI: 10.1109/TUFFC.2023.3327143.

[167] Yujia Huang et al. "Symbolic music generation with non-differentiable rule guided diffusion". In: *Proceedings of the 41st International Conference on Machine Learning*. Vol. 235. ICML'24. Vienna, Austria: JMLR.org, July 21, 2024, pp. 19772–19797. (Visited on 06/27/2025).

[168] John D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Computing in Science & Engineering* 9.3 (May 2007), pp. 90–95. ISSN: 1558-366X. DOI: 10.1109/MCSE.2007.55. URL: https://ieeexplore.ieee.org/document/4160265 (visited on 06/27/2025).

[169] M.F. Hutchinson. "A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines". In: *Communications in Statistics - Simulation and Computation* 19.2 (Jan. 1, 1990). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/03610919008812866, pp. 433–450. ISSN: 0361-0918. DOI: 10.1080/03610919008812866. URL: https://doi.org/10.1080/03610919008812866 (visited on 06/25/2025).

[170] A. Hyvärinen and E. Oja. "Independent component analysis: algorithms and applications". In: *Neural Networks* 13.4 (2000), pp. 411–430. ISSN: 0893-6080. DOI: https://doi.org/10.1016/S0893-6080(00)00026-5. URL: https://www.sciencedirect.com/science/article/pii/S0893608000000265.

[171] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. "Globally and locally consistent image completion". In: *ACM Trans. Graph.* 36.4 (July 20, 2017), 107:1–107:14. ISSN: 0730-0301. DOI: 10.1145/3072959.3073659. URL: https://doi.org/10.1145/3072959.3073659 (visited on 06/25/2025).

[172] Arthur Jacot, Franck Gabriel, and Clément Hongler. "Neural tangent kernel: convergence and generalization in neural networks". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS'18. Red Hook, NY, USA: Curran Associates Inc., Dec. 3, 2018, pp. 8580–8589. (Visited on 07/11/2025).

[173] Ajil Jalal et al. "Robust Compressed Sensing MRI with Deep Generative Priors". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 14938–14954. URL: https://proceedings.neurips.cc/paper/2021/hash/7d6044e95a16761171b130dcb476a43e-Abstract.html (visited on 06/25/2025).

[174] Yousef Javanmardi et al. "Quantifying cell-generated forces: Poisson's ratio matters". In: *Communications Physics* 4.1 (Nov. 4, 2021). Publisher: Nature Publishing Group, p. 237. ISSN: 2399-3650. DOI: 10.1038/s42005-021-00740-y. URL: https://www.nature.com/articles/s42005-021-00740-y (visited on 06/27/2025).

[175] R. C. Jennison. "A Phase Sensitive Interferometer Technique for the Measurement of the Fourier Transforms of Spatial Brightness Distributions of Small Angular Extent". In: *Monthly Notices of the Royal Astronomical Society* 118.3 (June 1, 1958), pp. 276–284. ISSN: 0035-8711. DOI: 10.1093/mnras/118.3.276. URL: https://doi.org/10.1093/mnras/118.3.276 (visited on 06/26/2025).

[176] Jun Hyeon Jo and Wansoo Ha. "Seismic Traveltime Tomography Using Deep Learning". In: *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023), pp. 1–11. ISSN: 1558-0644. DOI: 10.1109/TGRS.2023.3334283. URL: https://ieeexplore.ieee.org/document/10322754 (visited on 07/29/2025).

[177] Jitesh Jodhani et al. “Ultrasonic non-destructive evaluation of composites: A review”. In: *3rd Biennial International Conference on Future Learning Aspects of Mechanical Engineering (FLAME 2022)* 78 (Jan. 1, 2023), pp. 627–632. ISSN: 2214-7853. DOI: 10.1016/j.matpr.2022.12.055. URL: https://www.sciencedirect.com/science/article/pii/S2214785322074296.

[178] Avinash C. Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. _eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9780898719277. Society for Industrial and Applied Mathematics, 2001. DOI: 10.1137/1.9780898719277. URL: https://epubs.siam.org/doi/abs/10.1137/1.9780898719277.

[179] George Em Karniadakis et al. “Physics-informed machine learning”. In: *Nature Reviews Physics* 3.6 (June 2021). Publisher: Nature Publishing Group, pp. 422–440. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00314-5. URL: https://www.nature.com/articles/s42254-021-00314-5 (visited on 07/05/2025).

[180] Bahjat Kawar et al. “Denoising diffusion restoration models”. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. NIPS ’22. Red Hook, NY, USA: Curran Associates Inc., Nov. 28, 2022, pp. 23593–23606. ISBN: 978-1-7138-7108-8. (Visited on 06/25/2025).

[181] Tomohisa Kawashima et al. “A Jet-bases Emission Model of the EHT2017 Image of M87*”. In: *The Astrophysical Journal* 909.2 (Mar. 2021). Publisher: The American Astronomical Society, p. 168. ISSN: 0004-637X. DOI: 10.3847/1538-4357/abd5bb. URL: https://dx.doi.org/10.3847/1538-4357/abd5bb (visited on 06/26/2025).

[182] Bernhard Kerbl et al. “3D Gaussian Splatting for Real-Time Radiance Field Rendering”. In: *ACM Trans. Graph.* 42.4 (July 26, 2023), 139:1–139:14. ISSN: 0730-0301. DOI: 10.1145/3592433. URL: https://dl.acm.org/doi/10.1145/3592433 (visited on 07/11/2025).

[183] P Kidger. “On neural differential equations”. PhD thesis. University of Oxford, 2021.

[184] Meekyoung Kim et al. “Data-driven physics for human soft tissue animation”. In: *ACM Trans. Graph.* 36.4 (July 20, 2017), 54:1–54:12. ISSN: 0730-0301. DOI: 10.1145/3072959.3073685. URL: https://doi.org/10.1145/3072959.3073685 (visited on 06/27/2025).

[185] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *Proceedings of the International Conference on Learning Representations*. 2015.

[186] Diederik P. Kingma and Prafulla Dhariwal. “Glow: Generative Flow with Invertible 1x1 Convolutions”. In: *Advances in Neural Information Processing Systems*. Vol. 31. 2018.

[187] Diederik P. Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2013. DOI: https://doi.org/10.48550/arXiv.1312.6114. URL: https://arxiv.org/abs/1312.6114.

[188] Diederik P. Kingma et al. "Variational Diffusion Models". In: *Advances in Neural Information Processing Systems*. 2021.

[189] Polina Kirichenko, Pavel Izmailov, and Andrew Gordon Wilson. "Why normalizing flows fail to detect out-of-distribution data". In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS '20. Red Hook, NY, USA: Curran Associates Inc., Dec. 6, 2020, pp. 20578–20589. ISBN: 978-1-7138-2954-6. (Visited on 06/25/2025).

[190] Leon Klein, Andreas Krämer, and Frank Noé. "Equivariant flow matching". In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS '23. Red Hook, NY, USA: Curran Associates Inc., Dec. 10, 2023, pp. 59886–59910. (Visited on 06/27/2025).

[191] Dmitrii Kochkov et al. "Machine learning–accelerated computational fluid dynamics". In: *Proceedings of the National Academy of Sciences* 118.21 (May 25, 2021). Publisher: Proceedings of the National Academy of Sciences, e2101784118. DOI: 10.1073/pnas.2101784118. URL: https://www.pnas.org/doi/10.1073/pnas.2101784118 (visited on 06/28/2025).

[192] Jonas Köhler, Leon Klein, and Frank Noé. "Equivariant flows: exact likelihood generative learning for symmetric densities". In: *Proceedings of the 37th International Conference on Machine Learning*. Vol. 119. ICML'20. JMLR.org, July 13, 2020, pp. 5361–5370. (Visited on 06/27/2025).

[193] Ali R. Kolaini, Walter Tsuha, and Juan P. Fernandez. "Spacecraft Vibration Testing: Benefits and Potential Issues". In: European Conference on Spacecraft Structures, Materials and Environmental Testing. NTRS Author Affiliations: Jet Propulsion Lab., California Inst. of Tech. NTRS Report/Patent Number: JPL-CL-16-4234 NTRS Document ID: 20190026838 NTRS Research Center: Jet Propulsion Laboratory (JPL). Toulouse, Sept. 27, 2016. URL: https://ntrs.nasa.gov/citations/20190026838 (visited on 07/06/2025).

[194] Andrei Nikolaevich Kolmogorov et al. "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers". In: *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 434.1890 (Jan. 1997). Publisher: Royal Society, pp. 9–13. DOI: 10.1098/rspa.1991.0075. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1991.0075 (visited on 07/05/2025).

[195] Alex Krizhevsky and Geoffrey Hinton. "Learning Multiple Layers of Features from Tiny Images". In: *Technical report, University of Toronto* (2009).

URL: https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

[196] Kazuki Kuramochi et al. "Superresolution Interferometric Imaging with Sparse Modeling Using Total Squared Variation: Application to Imaging the Black Hole Shadow". In: *The Astrophysical Journal* 858.1 (May 2018). Publisher: The American Astronomical Society, p. 56. ISSN: 0004-637X. DOI: 10.3847/1538-4357/aab6b5. URL: https://dx.doi.org/10.3847/1538-4357/aab6b5 (visited on 06/25/2025).

[197] Sangwoo Kwon et al. "Comparison of Cancer Cell Elasticity by Cell Type". In: *Journal of Cancer* 11.18 (July 11, 2020), pp. 5403–5412. ISSN: 1837-9664. DOI: 10.7150/jca.45897. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7391204/ (visited on 06/27/2025).

[198] Lilian Lacourpaille et al. "Non-invasive assessment of muscle stiffness in patients with Duchenne muscular dystrophy". In: *Muscle & Nerve* 51.2 (Feb. 2015), pp. 284–286. ISSN: 1097-4598. DOI: 10.1002/mus.24445.

[199] Hugo Larochelle and Iain Murray. "The Neural Autoregressive Distribution Estimator". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, 2011, pp. 29–37. URL: https://proceedings.mlr.press/v15/larochelle11a.html.

[200] Yin Tat Lee, Ruoqi Shen, and Kevin Tian. "Structured Logconcave Sampling with a Restricted Gaussian Oracle". In: *Proceedings of Thirty Fourth Conference on Learning Theory*. Conference on Learning Theory. ISSN: 2640-3498. PMLR, July 21, 2021, pp. 2993–3050. URL: https://proceedings.mlr.press/v134/lee21a.html (visited on 07/03/2025).

[201] Aviad Levis, Yoav Y. Schechner, and Anthony B. Davis. "Multiple-Scattering Microphysics Tomography". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 1063-6919. July 2017, pp. 5797–5806. DOI: 10.1109/CVPR.2017.614. URL: https://ieeexplore.ieee.org/document/8100097 (visited on 07/11/2025).

[202] Aviad Levis et al. "Gravitationally Lensed Black Hole Emission Tomography". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

[203] Aviad Levis et al. "Orbital polarimetric tomography of a flare near the Sagittarius A* supermassive black hole". In: *Nature Astronomy* 8.6 (June 1, 2024), pp. 765–773. ISSN: 2397-3366. DOI: 10.1038/s41550-024-02238-3. URL: https://doi.org/10.1038/s41550-024-02238-3.

[204] Axel Levy et al. "Amortized inference for heterogeneous reconstruction in cryo-EM". In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. NIPS '22. Red Hook, NY, USA: Curran

Associates Inc., Nov. 28, 2022, pp. 13038–13049. ISBN: 978-1-7138-7108-8. (Visited on 07/01/2025).

[205] Ruilin Li et al. "The Mirror Langevin Algorithm Converges with Vanishing Bias". In: *Proceedings of the International Conference on Algorithmic Learning Theory*. PMLR, 2022, pp. 718–742.

[206] Xiang Li et al. *Decoupled Data Consistency with Diffusion Purification for Image Restoration*. June 8, 2025. DOI: 10.48550/arXiv.2403.06054. arXiv: 2403.06054[eess]. URL: http://arxiv.org/abs/2403.06054 (visited on 07/03/2025).

[207] Zhengqi Li et al. "Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 2575-7075. June 2021, pp. 6494–6504. DOI: 10.1109/CVPR46437.2021.00643. URL: https://ieeexplore.ieee.org/document/9578364 (visited on 07/11/2025).

[208] Chih-Ping Lin, Chun-Hung Lin, and Chih-Jung Chien. "Dispersion analysis of surface wave testing – SASW vs. MASW". In: *Journal of Applied Geophysics* 143 (Aug. 1, 2017), pp. 223–230. ISSN: 0926-9851. DOI: 10.1016/j.jappgeo.2017.05.008. URL: https://www.sciencedirect.com/science/article/pii/S0926985116305067.

[209] Phillip Lippe. *Welcome to the UvA Deep Learning Tutorials! — UvA DL Notebooks v1.2 documentation*. 2024. URL: https://uvadlc-notebooks.readthedocs.io/en/latest/ (visited on 07/30/2025).

[210] Ce Liu et al. "Exploring features in a Bayesian framework for material recognition". In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. ISSN: 1063-6919. June 2010, pp. 239–246. DOI: 10.1109/CVPR.2010.5540207. URL: https://ieeexplore.ieee.org/document/5540207 (visited on 06/27/2025).

[211] Guan-Horng Liu et al. "Mirror Diffusion Models for Constrained and Watermarked Generation". In: *Advances in Neural Information Processing Systems*. Vol. 36. 2024.

[212] Yixin Liu et al. *Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models*. Apr. 17, 2024. DOI: 10.48550/arXiv.2402.17177. arXiv: 2402.17177[cs]. URL: http://arxiv.org/abs/2402.17177 (visited on 08/19/2025).

[213] Ziwei Liu et al. "Deep Learning Face Attributes in the Wild". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2015*. Santiago, Chile: IEEE, 2015. DOI: 10.1109/ICCV.2015.425.

[214] Will Lockhart and Samuel E. Gralla. “How narrow is the M87* ring – II. A new geometric model”. In: *Monthly Notices of the Royal Astronomical Society* 517.2 (Dec. 2022), pp. 2462–2470. DOI: https://doi.org/10.1093/mnras/stac2743.

[215] Aaron Lou and Stefano Ermon. “Reflected diffusion models”. In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. ICML’23. Honolulu, Hawaii, USA: JMLR.org, July 23, 2023, pp. 22675–22701. (Visited on 06/27/2025).

[216] Ru-Sen Lu et al. “Imaging the Supermassive Black Hole Shadow and Jet Base of M87 with the Event Horizon Telescope”. In: *The Astrophysical Journal* 788.2 (May 2014). Publisher: The American Astronomical Society, p. 120. ISSN: 0004-637X. DOI: 10.1088/0004-637X/788/2/120. URL: https://dx.doi.org/10.1088/0004-637X/788/2/120 (visited on 06/26/2025).

[217] Jonathon Luiten et al. “Dynamic 3D Gaussians: Tracking by Persistent Dynamic View Synthesis”. In: *2024 International Conference on 3D Vision (3DV)*. 2024 International Conference on 3D Vision (3DV). ISSN: 2475-7888. Mar. 2024, pp. 800–809. DOI: 10.1109/3DV62453.2024.00044. URL: https://ieeexplore.ieee.org/document/10550869 (visited on 07/11/2025).

[218] J. -P. Luminet. “Image of a spherical black hole with thin accretion disk.” In: *Astronomy and Astrophysics* 75 (May 1, 1979). ADS Bibcode: 1979A&A....75..228L, pp. 228–235. ISSN: 0004-6361. URL: https://ui.adsabs.harvard.edu/abs/1979A&A....75..228L (visited on 06/25/2025).

[219] H. M\{{\textbackslash}"u\}ller and A. P. Lobanov. “Multiscale and multidirectional very long baseline interferometry imaging with CLEAN”. In: *Astronomy & Astrophysics* 672 (Apr. 1, 2023). Publisher: EDP Sciences, A26. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202244664. URL: https://www.aanda.org/articles/aa/abs/2023/04/aa44664-22/aa44664-22.html (visited on 06/25/2025).

[220] William N. MacPherson et al. “Multipoint laser vibrometer for modal analysis”. In: *Applied Optics* 46.16 (June 1, 2007). Publisher: Optica Publishing Group, pp. 3126–3132. ISSN: 2155-3165. DOI: 10.1364/AO.46.003126. URL: https://opg.optica.org/ao/abstract.cfm?uri=ao-46-16-3126 (visited on 06/27/2025).

[221] Gideon S. Mann and Andrew McCallum. “Simple, robust, scalable semi-supervised learning via expectation regularization”. In: *Proceedings of the 24th international conference on Machine learning*. ICML ’07. New York, NY, USA: Association for Computing Machinery, June 20, 2007, pp. 593–600. ISBN: 978-1-59593-793-3. DOI: 10.1145/1273496.1273571. URL: https://doi.org/10.1145/1273496.1273571 (visited on 06/27/2025).

[222] Yogesh K Mariappan, Kevin J Glaser, and Richard L Ehman. "Magnetic Resonance Elastography: A Review". In: *Clinical anatomy (New York, N.Y.)* 23.5 (July 2010), pp. 497–511. ISSN: 0897-3806. DOI: 10.1002/ca.21006. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3066083/ (visited on 06/27/2025).

[223] Milena Martarelli, Gian Marco Revel, and Claudio Santolini. "Automated Modal Analysis by Scanning Laser Vibrometry: Problems and Uncertainties Associated with the Scanning System Calibration". In: *Mechanical Systems and Signal Processing* 15.3 (May 1, 2001), pp. 581–601. ISSN: 0888-3270. DOI: 10.1006/mssp.2000.1336. URL: https://www.sciencedirect.com/science/article/pii/S0888327000913360.

[224] Lia Medeiros et al. "The Image of the M87 Black Hole Reconstructed with PRIMO". In: *The Astrophysical Journal Letters* 947.1 (Apr. 2023). Publisher: The American Astronomical Society, p. L7. ISSN: 2041-8205. DOI: 10.3847/2041-8213/acc32d. URL: https://dx.doi.org/10.3847/2041-8213/acc32d (visited on 06/26/2025).

[225] T.J. Memory, D.P. Thambiratnam, and G.H. Brameld. "Free vibration analysis of bridges". In: *Engineering Structures* 17.10 (Dec. 1, 1995), pp. 705–713. ISSN: 0141-0296. DOI: 10.1016/0141-0296(95)00037-8. URL: https://www.sciencedirect.com/science/article/pii/0141029695000378.

[226] Jianwei Miao et al. "Extending X-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes". In: *Annual Review of Physical Chemistry* 59 (2008), pp. 387–410. ISSN: 0066-426X. DOI: 10.1146/annurev.physchem.59.032607.093642.

[227] Laurence I. Midgley et al. "SE(3) equivariant augmented coupling flows". In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS '23. Red Hook, NY, USA: Curran Associates Inc., Dec. 10, 2023, pp. 79200–79225. (Visited on 06/27/2025).

[228] E. Miguel et al. "Data-Driven Estimation of Cloth Simulation Models". In: *Comput. Graph. Forum* 31.2 (May 1, 2012), pp. 519–528. ISSN: 0167-7055.

[229] Ben Mildenhall et al. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis". In: *Computer Vision – ECCV 2020*. Ed. by Andrea Vedaldi et al. Cham: Springer International Publishing, 2020, pp. 405–421. ISBN: 978-3-030-58452-8.

[230] M Mościbrodzka and C F Gammie. "IPOLE – semi-analytic scheme for relativistic polarized radiative transport". In: *Monthly Notices of the Royal Astronomical Society* 475.1 (Mar. 21, 2018), pp. 43–54. ISSN: 0035-8711. DOI: 10.1093/mnras/stx3162. URL: https://doi.org/10.1093/mnras/stx3162 (visited on 07/09/2025).

[231] mseitzer. *mseitzer/pytorch-fid*. original-date: 2018-02-10T12:33:06Z. July 28, 2025. URL: https://github.com/mseitzer/pytorch-fid (visited on 07/30/2025).

[232] Nalewajko, Krzysztof, Sikora, Marek, and Różańska, Agata. "Orientation of the crescent image of M 87*". In: *A&A* 634 (2020), A38. DOI: 10.1051/0004-6361/201936586. URL: https://doi.org/10.1051/0004-6361/201936586.

[233] Eric Nalisnick et al. "Do Deep Generative Models Know What They Don't Know?" In: *Proceedings of the International Conference on Machine Learning*. 2019. URL: https://openreview.net/forum?id=H1xwNhCcYm.

[234] Ramesh Narayan and Rajaram Nityananda. "Maximum entropy image restoration in astronomy". In: *Annual Review of Astronomy and Astrophysics* 24 (1986), pp. 127–170. DOI: 10.1146/annurev.aa.24.090186.001015.

[235] Hani H. Nassif, Mayrai Gindy, and Joe Davis. "Comparison of laser Doppler vibrometer with contact sensors for monitoring bridge deflection and vibration". In: *NDT & E International*. Structural Faults and Repair 38.3 (Apr. 1, 2005), pp. 213–218. ISSN: 0963-8695. DOI: 10.1016/j.ndteint.2004.06.012. URL: https://www.sciencedirect.com/science/article/pii/S0963869504001033 (visited on 06/27/2025).

[236] Soheil Nazarian and Kenneth H. Stokoe. *Evaluation of moduli and thicknesses of pavement systems by spectral-analysis-of-surface-waves method*. Publication Title: Texas Univ., Austin Report ADS Bibcode: 1983uta..rept.....N. Dec. 1, 1983. URL: https://ui.adsabs.harvard.edu/abs/1983uta..rept.....N (visited on 06/27/2025).

[237] Elias Nehme, Omer Yair, and Tomer Michaeli. "Uncertainty quantification via neural posterior principal components". In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS '23. Red Hook, NY, USA: Curran Associates Inc., Dec. 10, 2023, pp. 37128–37141. (Visited on 07/28/2025).

[238] Rodrigo Nemmen. "The Spin of M87*". In: *The Astrophysical Journal Letters* 880.2 (July 2019). Publisher: The American Astronomical Society, p. L26. ISSN: 2041-8205. DOI: 10.3847/2041-8213/ab2fd3. URL: https://dx.doi.org/10.3847/2041-8213/ab2fd3 (visited on 06/26/2025).

[239] Alex Nichol and Prafulla Dhariwal. "Improved Denoising Diffusion Probabilistic Models". In: *Proceedings of the International Conference on Machine Learning*. PMLR, 2021, pp. 8162–8171.

[240] Chenhao Niu et al. "Permutation invariant graph generation via score-based generative modeling". In: *Proceedings of the International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 4474–4484.

[241] Alexander C. Ogren et al. “Gaussian process regression as a surrogate model for the computation of dispersion relations”. In: *Computer Methods in Applied Mechanics and Engineering* 420 (Feb. 15, 2024), p. 116661. ISSN: 0045-7825. DOI: 10.1016/j.cma.2023.116661. URL: https://www.sciencedirect.com/science/article/pii/S0045782523007843.

[242] Alexander C. Ogren et al. “Visual Surface Wave Elastography: Revealing Subsurface Physical Properties via Visible Surface Waves”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2025.

[243] Roni Paiss et al. “Teaching CLIP to Count to Ten”. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2023 IEEE/CVF International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2023, pp. 3147–3157. DOI: 10.1109/ICCV51070.2023.00294. URL: https://ieeexplore.ieee.org/document/10376915 (visited on 06/28/2025).

[244] George Papamakarios, Theo Pavlakou, and Iain Murray. “Masked Autoregressive Flow for Density Estimation”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/6c1da886822c67822bcf3679d04369fa-Paper.pdf.

[245] Choon Park. “MASW for geotechnical site investigation”. In: *The Leading Edge* 32.6 (June 2013). Publisher: Society of Exploration Geophysicists, pp. 656–662. ISSN: 1070-485X. DOI: 10.1190/tle32060656.1. URL: https://library.seg.org/doi/10.1190/tle32060656.1 (visited on 06/27/2025).

[246] Keunhong Park et al. “HyperNeRF: a higher-dimensional representation for topologically varying neural radiance fields”. In: *ACM Trans. Graph.* 40.6 (Dec. 10, 2021), 238:1–238:12. ISSN: 0730-0301. DOI: 10.1145/3478513.3480487. URL: https://dl.acm.org/doi/10.1145/3478513.3480487 (visited on 07/11/2025).

[247] Keunhong Park et al. “Nerfies: Deformable Neural Radiance Fields”. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2021, pp. 5845–5854. DOI: 10.1109/ICCV48922.2021.00581. URL: https://ieeexplore.ieee.org/document/9711476 (visited on 07/11/2025).

[248] Deepak Pathak et al. “Context Encoders: Feature Learning by Inpainting”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016, pp. 2536–2544.

[249] Sida Peng et al. “Animatable Neural Radiance Fields for Modeling Dynamic Human Bodies”. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2021, pp. 14294–14303. DOI: 10.1109/ICCV48922.2021.01405. URL: https://ieeexplore.ieee.org/document/9710330 (visited on 07/11/2025).

[250] B.T. Polyak. “Newton’s method and its use in optimization”. In: *European Journal of Operational Research* 181.3 (Sept. 16, 2007), pp. 1086–1096. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2005.06.076. URL: https://www.sciencedirect.com/science/article/pii/S0377221706001469.

[251] Yuriy M. Poplavko. “Chapter 2 - Mechanical properties of solids”. In: *Electronic Materials*. Ed. by Yuriy M. Poplavko. Elsevier, Jan. 1, 2019, pp. 71–93. ISBN: 978-0-12-815780-0. DOI: 10.1016/B978-0-12-815780-0.00002-5. URL: https://www.sciencedirect.com/science/article/pii/B9780128157800000025 (visited on 06/27/2025).

[252] Javier Portilla and Eero P. Simoncelli. “A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients”. In: *International Journal of Computer Vision* 40.1 (Oct. 1, 2000), pp. 49–70. ISSN: 1573-1405. DOI: 10.1023/A:1026553619983. URL: https://doi.org/10.1023/A:1026553619983 (visited on 06/27/2025).

[253] Albert Pumarola et al. “D-NeRF: Neural Radiance Fields for Dynamic Scenes”. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 2575-7075. June 2021, pp. 10313–10322. DOI: 10.1109/CVPR46437.2021.01018. URL: https://ieeexplore.ieee.org/document/9578753 (visited on 07/11/2025).

[254] M. Raissi, P. Perdikaris, and G. E. Karniadakis. “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations”. In: *Journal of Computational Physics* 378 (Feb. 1, 2019), pp. 686–707. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2018.10.045. URL: https://www.sciencedirect.com/science/article/pii/S0021999118307125 (visited on 07/05/2025).

[255] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. “Stochastic Backpropagation and Approximate Inference in Deep Generative Models”. In: *Proceedings of the 31st International Conference on Machine Learning*. Vol. 32. Proceedings of Machine Learning Research. Beijing, China: PMLR, 2014, pp. 1278–1286. URL: https://proceedings.mlr.press/v32/rezende14.html.

[256] Danilo Jimenez Rezende et al. *Equivariant Hamiltonian Flows*. Sept. 30, 2019. DOI: 10.48550/arXiv.1909.13739. arXiv: 1909.13739[stat]. URL: http://arxiv.org/abs/1909.13739 (visited on 06/27/2025).

[257] Alan E. E. Rogers, Sheperd S. Doeleman, and James M. Moran. “Fringe Detection Methods for Very Long Baseline Arrays”. In: *The Astronomical Journal* 109 (Mar. 1, 1995). Publisher: IOP ADS Bibcode: 1995AJ....109.1391R, p. 1391. ISSN: 0004-6256. DOI: 10.1086/117371. URL: https://ui.adsabs.harvard.edu/abs/1995AJ....109.1391R (visited on 06/26/2025).

[258] Yaniv Romano, Michael Elad, and Peyman Milanfar. “The Little Engine That Could: Regularization by Denoising (RED)”. In: *SIAM Journal on Imaging Sciences* 10.4 (2017). _eprint: https://doi.org/10.1137/16M1102884, pp. 1804–1844. DOI: 10.1137/16M1102884. URL: https://doi.org/10.1137/16M1102884.

[259] N. B. Roozen et al. “Determining radiated sound power of building structures by means of laser Doppler vibrometry”. In: *Journal of Sound and Vibration* 346 (June 23, 2015), pp. 81–99. ISSN: 0022-460X. DOI: 10.1016/j.jsv.2015.02.029. URL: https://www.sciencedirect.com/science/article/pii/S0022460X15001637 (visited on 06/27/2025).

[260] Litu Rout et al. “Solving Linear Inverse Problems Provably via Posterior Sampling with Latent Diffusion Models”. In: Thirty-seventh Conference on Neural Information Processing Systems. Nov. 2, 2023. URL: https://openreview.net/forum?id=XKBFdYwfRo (visited on 07/03/2025).

[261] François Rozet and Gilles Louppe. “Score-based data assimilation”. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS ’23. Red Hook, NY, USA: Curran Associates Inc., Dec. 10, 2023, pp. 40521–40541. (Visited on 06/28/2025).

[262] G. B. Rybicki. “Deprojection of Galaxies - how much can BE Learned”. In: Structure and Dynamics of Elliptical Galaxies. Vol. 127. ADS Bibcode: 1987IAUS..127..397R. Jan. 1, 1987, p. 397. DOI: 10.1007/978-94-009-3971-4_41. URL: https://ui.adsabs.harvard.edu/abs/1987IAUS..127..397R (visited on 07/11/2025).

[263] Chitwan Saharia et al. “Image Super-Resolution via Iterative Refinement”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.4 (Apr. 2023), pp. 4713–4726. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2022.3204461. URL: https://ieeexplore.ieee.org/document/9887996 (visited on 06/25/2025).

[264] Regina Augusta Sampaio and Remo Magalhães de Souza. “Vibration Analysis of a Residential Building”. In: *MATEC Web of Conferences* 24 (2015). Publisher: EDP Sciences, p. 09007. ISSN: 2261-236X. DOI: 10.1051/matecconf/20152409007. URL: https://www.matec-conferences.org/articles/matecconf/abs/2015/05/matecconf_evaces2015_09007/matecconf_evaces2015_09007.html (visited on 07/06/2025).

[265] Víctor Garcia Satorras et al. "E(n) Equivariant Normalizing Flows". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 4181–4192. URL: https://proceedings.neurips.cc/paper/2021/hash/21b5680d80f75a616096f2e791affac6-Abstract.html (visited on 06/27/2025).

[266] Víctor Garcias Satorras, Emiel Hoogeboom, and Max Welling. "E(n) Equivariant Graph Neural Networks". In: *Proceedings of the International Conference on Machine Learning*. PMLR, 2011, pp. 9323–9332.

[267] A. van der Schaaf and J. H. van Hateren. "Modelling the Power Spectra of Natural Images: Statistics and Information". In: *Vision Research* 36.17 (Sept. 1, 1996), pp. 2759–2770. ISSN: 0042-6989. DOI: 10.1016/0042-6989(96)00002-8. URL: https://www.sciencedirect.com/science/article/pii/0042698996000028 (visited on 06/27/2025).

[268] F. R. Schwab. "Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry". In: *The Astronomical Journal* 89 (July 1, 1984). Publisher: IOP ADS Bibcode: 1984AJ.....89.1076S, pp. 1076–1081. ISSN: 0004-6256. DOI: 10.1086/113605. URL: https://ui.adsabs.harvard.edu/abs/1984AJ.....89.1076S (visited on 06/25/2025).

[269] Gabriel Schwartz and Ko Nishino. "Visual Material Traits: Recognizing Per-Pixel Material Context". In: *2013 IEEE International Conference on Computer Vision Workshops*. 2013 IEEE International Conference on Computer Vision Workshops. Dec. 2013, pp. 883–890. DOI: 10.1109/ICCVW.2013.121. URL: https://ieeexplore.ieee.org/document/6755990 (visited on 06/27/2025).

[270] U. J. Schwarz. "Mathematical-statistical Description of the Iterative Beam Removing Technique (Method CLEAN)". In: *Astronomy and Astrophysics* 65 (Apr. 1, 1978). ADS Bibcode: 1978A&A....65..345S, p. 345. ISSN: 0004-6361. URL: https://ui.adsabs.harvard.edu/abs/1978A&A....65..345S (visited on 06/25/2025).

[271] Nikolai M. Shapiro et al. "High-Resolution Surface-Wave Tomography from Ambient Seismic Noise". In: *Science* 307.5715 (Mar. 11, 2005). Publisher: American Association for the Advancement of Science, pp. 1615–1618. DOI: 10.1126/science.1108339. URL: https://www.science.org/doi/10.1126/science.1108339 (visited on 06/27/2025).

[272] Lavanya Sharan et al. "Image statistics for surface reflectance perception". In: *JOSA A* 25.4 (Apr. 1, 2008). Publisher: Optica Publishing Group, pp. 846–865. ISSN: 1520-8532. DOI: 10.1364/JOSAA.25.000846. URL: https://opg.optica.org/josaa/abstract.cfm?uri=josaa-25-4-846 (visited on 06/27/2025).

[273] Martin Shepherd. “Difmap: Synthesis Imaging of Visibility Data”. In: *Astrophysics Source Code Library* (2011), ascl–1103.

[274] E.P. Simoncelli and W.T. Freeman. “The steerable pyramid: a flexible architecture for multi-scale derivative computation”. In: *Proceedings., International Conference on Image Processing*. ., International Conference on Image Processing. Vol. 3. Oct. 1995, 444–447 vol.3. DOI: 10.1109/ICIP.1995.537667. URL: https://ieeexplore.ieee.org/document/537667 (visited on 06/27/2025).

[275] E.P. Simoncelli et al. “Shiftable multiscale transforms”. In: *IEEE Transactions on Information Theory* 38.2 (Mar. 1992), pp. 587–607. ISSN: 1557-9654. DOI: 10.1109/18.119725. URL: https://ieeexplore.ieee.org/document/119725 (visited on 06/27/2025).

[276] John Skilling. “The Eigenvalues of Mega-dimensional Matrices”. In: *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*. Ed. by J. Skilling. Dordrecht: Springer Netherlands, 1989, pp. 455–466. ISBN: 978-94-015-7860-8. DOI: 10.1007/978-94-015-7860-8_48. URL: https://doi.org/10.1007/978-94-015-7860-8_48 (visited on 06/25/2025).

[277] Jascha Sohl-Dickstein et al. “Deep Unsupervised Learning using Nonequilibrium Thermodynamics”. In: *Proceedings of the 32nd International Conference on Machine Learning*. International Conference on Machine Learning. ISSN: 1938-7228. PMLR, June 1, 2015, pp. 2256–2265. URL: https://proceedings.mlr.press/v37/sohl-dickstein15.html (visited on 06/25/2025).

[278] Bowen Song et al. “Solving Inverse Problems with Latent Diffusion Models via Hard Data Consistency”. In: The Twelfth International Conference on Learning Representations. Oct. 13, 2023. URL: https://openreview.net/forum?id=j8hdRqOUhN (visited on 07/03/2025).

[279] Jiamin Song, Chenlin Meng, and Stefano Ermon. “Denoising Diffusion Implicit Models”. In: *Proceedings of the International Conference on Learning Representations*. 2021. URL: https://openreview.net/forum?id=St1giarCHLP.

[280] Jiaming Song et al. “Pseudoinverse-Guided Diffusion Models for Inverse Problems”. In: *Proceedings of the Eleventh International Conference on Learning Representations*. International Conference on Learning Representations (ICLR). Kigali, Rwanda, Sept. 29, 2022. URL: https://openreview.net/forum?id=9_gsMA8MRKQ (visited on 06/25/2025).

[281] Yang Song and Stefano Ermon. “Generative modeling by estimating gradients of the data distribution”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. 1067. Red Hook, NY, USA: Curran Associates Inc., Dec. 8, 2019, pp. 11918–11930. (Visited on 06/25/2025).

[282] Yang Song et al. “Maximum Likelihood Training of Score-Based Diffusion Models”. In: *Proceedings of the 35th International Conference on Neural Information Processing Systems*. NIPS ’21. Red Hook, NY, USA: Curran Associates Inc., Dec. 6, 2021, pp. 1415–1428. ISBN: 978-1-7138-4539-3. (Visited on 06/25/2025).

[283] Yang Song et al. “Score-Based Generative Modeling through Stochastic Differential Equations”. In: *Proceedings of the Ninth International Conference on Learning Representations*. International Conference on Learning Representations (ICLR). 2021. URL: https://openreview.net/forum?id=PxTIG12RRHS.

[284] Yang Song et al. “Solving Inverse Problems in Medical Imaging with Score-Based Generative Models”. In: NeurIPS 2021 Workshop on Deep Learning and Inverse Problems. Oct. 19, 2021. URL: https://openreview.net/forum?id=4rFAhgrA0lA (visited on 06/25/2025).

[285] Yuxuan Song et al. “Equivariant flow matching with hybrid probability transport for 3D molecule generation”. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS ’23. Red Hook, NY, USA: Curran Associates Inc., Dec. 10, 2023, pp. 549–568. (Visited on 06/27/2025).

[286] Emanuela Speranzini and Stefano Agnetti. “The technique of digital image correlation to identify defects in glass structures”. In: *Structural Control and Health Monitoring* 21.6 (2014). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/stc.1629, pp. 1015–1029. ISSN: 1545-2263. DOI: 10.1002/stc.1629. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/stc.1629 (visited on 06/27/2025).

[287] He Sun. *HeSunPU/DPI*. original-date: 2020-12-02T20:52:51Z. Apr. 20, 2025. URL: https://github.com/HeSunPU/DPI (visited on 07/30/2025).

[288] He Sun and Katherine L. Bouman. “Deep Probabilistic Imaging: Uncertainty Quantification and Multi-modal Solution Characterization for Computational Imaging”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 2021. DOI: https://doi.org/10.1609/aaai.v35i3.16366.

[289] He Sun et al. “α-deep Probabilistic Inference (α-DPI): Efficient Uncertainty Quantification from Exoplanet Astrometry to Black Hole Feature Extraction”. In: *The Astrophysical Journal* 932.2 (June 2022). Publisher: The American Astronomical Society, p. 99. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ac6be9. URL: https://dx.doi.org/10.3847/1538-4357/ac6be9 (visited on 06/25/2025).

[290] Jing-An Sun et al. *Align-DA: Align Score-based Atmospheric Data Assimilation with Multiple Preferences*. May 28, 2025. DOI: 10.48550/arXiv.

2505.22008. arXiv: 2505.22008[physics]. URL: http://arxiv.org/abs/2505.22008 (visited on 07/28/2025).

[291] Shan Sun et al. "Biomechanics and functionality of hepatocytes in liver cirrhosis". In: *Journal of Biomechanics* 47.9 (June 27, 2014), pp. 2205–2210. ISSN: 1873-2380. DOI: 10.1016/j.jbiomech.2013.10.050.

[292] Yu Sun et al. "CoIL: Coordinate-Based Internal Learning for Tomographic Imaging". In: *IEEE Transactions on Computational Imaging* 7 (2021), pp. 1400–1412. ISSN: 2333-9403. DOI: 10.1109/TCI.2021.3125564. URL: https://ieeexplore.ieee.org/document/9606601 (visited on 07/11/2025).

[293] Danica J. Sutherland. *djsutherland/opt-mmd*. original-date: 2016-11-17T22:23:29Z. Apr. 8, 2025. URL: https://github.com/djsutherland/opt-mmd (visited on 07/30/2025).

[294] Danica J. Sutherland et al. "Generative Models and Model Criticism via Optimized Maximum Mean Discrepancy". In: International Conference on Learning Representations. Mar. 30, 2021. URL: https://openreview.net/forum?id=HJWHIKqgl (visited on 07/30/2025).

[295] Christian Szegedy et al. "Going deeper with convolutions". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Journal Abbreviation: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 7, 2015, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594. URL: http://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594.

[296] Mihra S. Taljanovic et al. "Shear-Wave Elastography: Basic Physics and Musculoskeletal Applications". In: *Radiographics: A Review Publication of the Radiological Society of North America, Inc* 37.3 (2017), pp. 855–870. ISSN: 1527-1323. DOI: 10.1148/rg.2017160116.

[297] Hong Ye Tan et al. "Data-Driven Mirror Descent with Input-Convex Neural Networks". In: *SIAM Journal on Mathematics of Data Science* 5.2 (June 30, 2023). Publisher: Society for Industrial and Applied Mathematics, pp. 558–587. DOI: 10.1137/22M1508613. URL: https://epubs.siam.org/doi/10.1137/22M1508613 (visited on 06/27/2025).

[298] Matthew Tancik et al. "Fourier features let networks learn high frequency functions in low dimensional domains". In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS '20. Red Hook, NY, USA: Curran Associates Inc., Dec. 6, 2020, pp. 7537–7547. ISBN: 978-1-7138-2954-6. (Visited on 07/10/2025).

[299] G. I. Taylor. "The Spectrum of Turbulence". In: *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences* 164.919 (Jan. 1997). Publisher: Royal Society, pp. 476–490. DOI: 10.1098/rspa.

1938.0032. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1938.0032 (visited on 07/05/2025).

[300] *The Visible Human Project*. Publisher: U.S. National Library of Medicine. URL: https://www.nlm.nih.gov/research/visible/visible_human.html (visited on 06/27/2025).

[301] Nathaniel Thomas et al. *Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds*. May 18, 2018. DOI: 10.48550/arXiv.1802.08219. arXiv: 1802.08219[cs]. URL: http://arxiv.org/abs/1802.08219 (visited on 06/27/2025).

[302] A. Richard Thompson, James M. Moran, and George W. Swenson. *Interferometry and Synthesis in Radio Astronomy*. Astronomy and Astrophysics Library. Cham: Springer International Publishing, 2017. ISBN: 978-3-319-44429-1. DOI: 10.1007/978-3-319-44431-4. URL: http://link.springer.com/10.1007/978-3-319-44431-4 (visited on 06/25/2025).

[303] Paul Tiede et al. "Spacetime Tomography Using the Event Horizon Telescope". In: *The Astrophysical Journal* 892.2 (Apr. 2020). Publisher: The American Astronomical Society, p. 132. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ab744c. URL: https://dx.doi.org/10.3847/1538-4357/ab744c (visited on 07/11/2025).

[304] F. Trebuňa and M. Hagara. "Experimental modal analysis performed by high-speed digital image correlation system". In: *Measurement* 50 (Apr. 1, 2014), pp. 78–85. ISSN: 0263-2241. DOI: 10.1016/j.measurement.2013.12.038. URL: https://www.sciencedirect.com/science/article/pii/S0263224113006593.

[305] Edgar Tretschk et al. "Non-Rigid Neural Radiance Fields: Reconstruction and Novel View Synthesis of a Dynamic Scene From Monocular Video". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2021, pp. 12939–12950. DOI: 10.1109/ICCV48922.2021.01272. URL: https://ieeexplore.ieee.org/document/9709900 (visited on 07/11/2025).

[306] Brian L. Trippe et al. "Diffusion Probabilistic Modeling of Protein Backbones in 3D for the motif-scaffolding problem". In: The Eleventh International Conference on Learning Representations. Sept. 29, 2022. URL: https://openreview.net/forum?id=6TxBxqNME1Y (visited on 07/03/2025).

[307] Ch. Tsitouras. "Runge–Kutta pairs of order 5(4) satisfying only the first column simplifying assumption". In: *Computers & Mathematics with Applications* 62.2 (July 1, 2011), pp. 770–775. ISSN: 0898-1221. DOI: 10.1016/j.camwa.2011.06.002. URL: https://www.sciencedirect.com/science/article/pii/S0898122111004706 (visited on 06/25/2025).

[308] Shih-Heng Tung, Ming-Hsiang Shih, and Wen-Pei Sung. "Development of digital image correlation method to analyse crack variations of masonry wall". In: *Sadhana* 33.6 (Dec. 1, 2008), pp. 767–779. ISSN: 0973-7677. DOI: 10.1007/s12046-008-0033-2. URL: https://doi.org/10.1007/s12046-008-0033-2 (visited on 06/27/2025).

[309] Greg Turk and Marc Levoy. "Zippered polygon meshes from range images". In: *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. SIGGRAPH '94. New York, NY, USA: Association for Computing Machinery, July 24, 1994, pp. 311–318. ISBN: 978-0-89791-667-7. DOI: 10.1145/192161.192241. URL: https://dl.acm.org/doi/10.1145/192161.192241 (visited on 06/27/2025).

[310] R. Q. Twiss, A. W. L. Carter, and A. G. Little. "Brightness distribution over some strong radio sources at 1427 Mc/s". In: *The Observatory* 80 (Aug. 1, 1960). ADS Bibcode: 1960Obs....80..153T, pp. 153–159. ISSN: 0029-7704. URL: https://ui.adsabs.harvard.edu/abs/1960Obs....80..153T (visited on 06/26/2025).

[311] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. "Deep Image Prior". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018*. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018, pp. 9446–9454. DOI: https://doi.org/10.1007/s11263-020-01303-4.

[312] Rishi Upadhyay et al. "Enhancing Diffusion Models with 3D Perspective Geometry Constraints". In: *ACM Trans. Graph.* 42.6 (Dec. 5, 2023), 237:1–237:15. ISSN: 0730-0301. DOI: 10.1145/3618389. URL: https://dl.acm.org/doi/10.1145/3618389 (visited on 06/27/2025).

[313] Muhammad Usman and Philipp G. Batchelor. "Optimized Sampling Patterns for Practical Compressed MRI". In: *SAMPTA'09, International Conference on Sampling Theory and Applications*. Ed. by Laurent Fesquet and Bruno Torrésani. Marseille, France, May 2009, Poster session. URL: https://hal.science/hal-00453565 (visited on 07/29/2025).

[314] Aäron Van Den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. "Pixel recurrent neural networks". In: *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*. ICML'16. New York, NY, USA: JMLR.org, June 19, 2016, pp. 1747–1756. (Visited on 06/25/2025).

[315] N. H. Varady and A. J. Grodzinsky. "Osteoarthritis year in review 2015: mechanics". In: *Osteoarthritis and Cartilage* 24.1 (Jan. 2016), pp. 27–35. ISSN: 1522-9653. DOI: 10.1016/j.joca.2015.08.018.

[316] 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 IEEE Global Conference on Signal and Information Processing. Dec. 2013, pp. 945–948. DOI: 10.1109/GlobalSIP.2013.6737048. URL: https://ieeexplore.ieee.org/document/6737048 (visited on 06/25/2025).

[317] Vincent, F. H. et al. "Geometric modeling of M87* as a Kerr black hole or a non-Kerr compact object". In: *A&A* 646 (2021), A37. DOI: 10.1051/0004-6361/202037787. URL: https://doi.org/10.1051/0004-6361/202037787.

[318] Kelley M. Virgilio et al. "Multiscale models of skeletal muscle reveal the complex effects of muscular dystrophy on tissue mechanics and damage susceptibility". In: *Interface Focus* 5.2 (Apr. 6, 2015), p. 20140080. ISSN: 2042-8898. DOI: 10.1098/rsfs.2014.0080. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4342948/ (visited on 06/27/2025).

[319] Pauli Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature Methods* 17.3 (Mar. 2020). Publisher: Nature Publishing Group, pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2. URL: https://www.nature.com/articles/s41592-019-0686-2 (visited on 06/27/2025).

[320] Maxime Vono, Nicolas Dobigeon, and Pierre Chainais. "Split-and-Augmented Gibbs Sampler—Application to Large-Scale Inference Problems". In: *Trans. Sig. Proc.* 67.6 (Mar. 1, 2019), pp. 1648–1661. ISSN: 1053-587X. DOI: 10.1109/TSP.2019.2894825. URL: https://doi.org/10.1109/TSP.2019.2894825 (visited on 07/03/2025).

[321] Neal Wadhwa et al. "Phase-based video motion processing". In: *ACM Trans. Graph.* 32.4 (July 21, 2013), 80:1–80:10. ISSN: 0730-0301. DOI: 10.1145/2461912.2461966. URL: https://dl.acm.org/doi/10.1145/2461912.2461966 (visited on 06/27/2025).

[322] Jonelle L. Walsh et al. "The M87 Black Hole Mass from Gas-dynamical Models of Space Telescope Imaging Spectrograph Observations". In: *The Astrophysical Journal* 770.2 (May 2013). Publisher: The American Astronomical Society, p. 86. ISSN: 0004-637X. DOI: 10.1088/0004-637X/770/2/86. URL: https://dx.doi.org/10.1088/0004-637X/770/2/86 (visited on 06/26/2025).

[323] Bin Wang et al. "Deformation capture and modeling of soft objects". In: *ACM Trans. Graph.* 34.4 (July 27, 2015), 94:1–94:12. ISSN: 0730-0301. DOI: 10.1145/2766911. URL: https://doi.org/10.1145/2766911 (visited on 06/27/2025).

[324] Bohan Wang, Yili Zhao, and Jernej Barbič. "Botanical materials based on biomechanics". In: *ACM Trans. Graph.* 36.4 (July 20, 2017), 135:1–135:13. ISSN: 0730-0301. DOI: 10.1145/3072959.3073655. URL: https://dl.acm.org/doi/10.1145/3072959.3073655 (visited on 06/27/2025).

[325] Huamin Wang, James F. O'Brien, and Ravi Ramamoorthi. "Data-driven elastic models for cloth: modeling and measurement". In: *ACM SIGGRAPH 2011 papers*. SIGGRAPH '11. New York, NY, USA: Association for Computing Machinery, July 25, 2011, pp. 1–12. ISBN: 978-1-4503-0943-1. DOI: 10.1145/1964921.1964966. URL: https://doi.org/10.1145/1964921.1964966 (visited on 06/27/2025).

[326] Yinhuai Wang, Jiwen Yu, and Jian Zhang. "Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model". In: *Proceedings of the International Conference on Learning Representations*. The Eleventh International Conference on Learning Representations. 2023.

[327] Zhou Wang et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE Transactions on Image Processing* 13.4 (Apr. 2004), pp. 600–612. ISSN: 1941-0042. DOI: 10.1109/TIP.2003.819861. URL: https://ieeexplore.ieee.org/document/1284395 (visited on 06/27/2025).

[328] Paul Webb. *Introduction to Oceanography*. Pages 231–249. Roger Williams University, 2021. 19 pp. URL: https://rwu.pressbooks.pub/webboceanography/ (visited on 06/27/2025).

[329] George N. Wong et al. "PATOKA: Simulating Electromagnetic Observables of Black Hole Accretion". In: *The Astrophysical Journal Supplement Series* 259.2 (2022), p. 64.

[330] Vincent Wai-Sun Wong and Henry Lik-Yuen Chan. "Transient elastography". In: *Journal of Gastroenterology and Hepatology* 25.11 (Nov. 2010), pp. 1726–1731. ISSN: 1440-1746. DOI: 10.1111/j.1440-1746.2010.06437.x.

[331] Michael M. Woolfson. *An Introduction to X-ray Crystallography*. 2nd ed. Cambridge University Press, 1997. ISBN: 978-0-511-62255-7.

[332] Luhuan Wu et al. "Practical and asymptotically exact conditional sampling in diffusion models". In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS '23. Red Hook, NY, USA: Curran Associates Inc., Dec. 10, 2023, pp. 31372–31403. (Visited on 07/03/2025).

[333] ZhiMin Wu et al. "An experimental investigation on the FPZ properties in concrete using digital image correlation technique". In: *Engineering Fracture Mechanics* 78.17 (Dec. 1, 2011), pp. 2978–2990. ISSN: 0013-7944. DOI: 10.1016/j.engfracmech.2011.08.016. URL: https://www.

sciencedirect.com/science/article/pii/S0013794411003146 (visited on 06/27/2025).

[334] Zihui Wu et al. "Principled Probabilistic Imaging using Diffusion Models as Plug-and-Play Priors". In: The Thirty-eighth Annual Conference on Neural Information Processing Systems. Nov. 6, 2024. URL: https://openreview.net/forum?id=Xq9HQf7VNV (visited on 07/03/2025).

[335] Wenqi Xian et al. "Space-time Neural Irradiance Fields for Free-Viewpoint Video". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 2575-7075. June 2021, pp. 9416–9426. DOI: 10.1109/CVPR46437.2021.00930. URL: https://ieeexplore.ieee.org/document/9577376 (visited on 07/11/2025).

[336] Jinhui Xiong and Wolfgang Heidrich. "In-the-Wild Single Camera 3D Reconstruction Through Moving Water Surfaces". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2021, pp. 12538–12547. DOI: 10.1109/ICCV48922.2021.01233. URL: https://ieeexplore.ieee.org/document/9710376 (visited on 07/11/2025).

[337] Hongyi Xu et al. "Interactive Material Design Using Model Reduction". In: *ACM Trans. Graph.* 34.2 (Mar. 2, 2015), 18:1–18:14. ISSN: 0730-0301. DOI: 10.1145/2699648. URL: https://doi.org/10.1145/2699648 (visited on 06/27/2025).

[338] Minkai Xu et al. "GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation". In: *Proceedings of the International Conference on Learning Representations*. 2022. URL: https://openreview.net/forum?id=PzcvxEMzvQC.

[339] Xingyu Xu and Yuejie Chi. "Provably Robust Score-Based Diffusion Posterior Sampling for Plug-and-Play Image Reconstruction". In: The Thirty-eighth Annual Conference on Neural Information Processing Systems. Nov. 6, 2024. URL: https://openreview.net/forum?id=SLnsoaY4u1 (visited on 07/03/2025).

[340] Yongchao Yang et al. "Blind, simultaneous identification of full-field vibration modes and large rigid-body motion of output-only structures from digital video measurements". In: *Engineering Structures* 207 (Mar. 15, 2020), p. 110183. ISSN: 0141-0296. DOI: 10.1016/j.engstruct.2020.110183. URL: https://www.sciencedirect.com/science/article/pii/S0141029618321370 (visited on 06/27/2025).

[341] Jason Yim et al. "SE(3) diffusion model with application to protein backbone generation". In: *Proceedings of the 40th International Conference on*

Machine Learning. Vol. 202. ICML'23. Honolulu, Hawaii, USA: JMLR.org, July 23, 2023, pp. 40001–40039. (Visited on 06/27/2025).

[342] Tianwei Yin et al. "End-to-End Sequential Sampling and Reconstruction for MRI". In: *Proceedings of Machine Learning Research*. Machine Learning for Health (ML4H). Vol. 158. PMLR, 2021, pp. 261–279.

[343] Jiahui Yu et al. "Free-Form Image Inpainting with Gated Convolution". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019*. IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea, 2019.

[344] Feng Yuan, Haiyang Wang, and Hai Yang. "The Accretion Flow in M87 is Really MAD". In: *The Astrophysical Journal* 924.2 (Jan. 2022). Publisher: The American Astronomical Society, p. 124. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ac4714. URL: https://dx.doi.org/10.3847/1538-4357/ac4714 (visited on 06/26/2025).

[345] Jure Zbontar et al. *fastMRI: An Open Dataset and Benchmarks for Accelerated MRI*. Dec. 11, 2019. DOI: 10.48550/arXiv.1811.08839. arXiv: 1811.08839[cs]. URL: http://arxiv.org/abs/1811.08839 (visited on 07/01/2025).

[346] Babette S. Zemel. "Chapter 19 - Body composition during growth and development". In: *Human Growth and Development (Third Edition)*. Ed. by Noël Cameron and Lawrence M. Schell. Boston: Academic Press, Jan. 1, 2022, pp. 517–545. ISBN: 978-0-12-822652-0. DOI: 10.1016/B978-0-12-822652-0.00018-3. URL: https://www.sciencedirect.com/science/article/pii/B9780128226520000183 (visited on 06/27/2025).

[347] F. Zernike. "The concept of degree of coherence and its application to optical problems". In: *Physica* 5.8 (Aug. 1, 1938), pp. 785–795. ISSN: 0031-8914. DOI: 10.1016/S0031-8914(38)80203-2. URL: https://www.sciencedirect.com/science/article/pii/S0031891438802032 (visited on 06/26/2025).

[348] Bingliang Zhang et al. "Improving Diffusion Inverse Problem Solving with Decoupled Noise Annealing". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2025. URL: https://openaccess.thecvf.com/content/CVPR2025/papers/Zhang_Improving_Diffusion_Inverse_Problem_Solving_with_Decoupled_Noise_Annealing_CVPR_2025_paper.pdf.

[349] Bingliang Zhang et al. *STeP: A Framework for Solving Scientific Video Inverse Problems with Spatiotemporal Diffusion Priors*. June 10, 2025. DOI: 10.48550/arXiv.2504.07549. arXiv: 2504.07549[cs]. URL: http://arxiv.org/abs/2504.07549.

[350] Jinwei Zhang et al. “Extending LOUPE for K-Space Under-Sampling Pattern Optimization in Multi-coil MRI”. In: *Machine Learning for Medical Image Reconstruction: Third International Workshop, MLMIR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings*. Berlin, Heidelberg: Springer-Verlag, Oct. 8, 2020, pp. 91–101. ISBN: 978-3-030-61597-0. DOI: 10.1007/978-3-030-61598-7_9. URL: https://doi.org/10.1007/978-3-030-61598-7_9 (visited on 06/25/2025).

[351] Kaihao Zhang et al. “Deep Image Deblurring: A Survey”. In: *International Journal of Computer Vision* 130.9 (Sept. 1, 2022), pp. 2103–2130. ISSN: 1573-1405. DOI: 10.1007/s11263-022-01633-5. URL: https://doi.org/10.1007/s11263-022-01633-5 (visited on 06/25/2025).

[352] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. “Adding Conditional Control to Text-to-Image Diffusion Models”. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2023 IEEE/CVF International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2023, pp. 3813–3824. DOI: 10.1109/ICCV51070.2023.00355. URL: https://ieeexplore.ieee.org/document/10377881 (visited on 06/27/2025).

[353] Brandon Zhao et al. “Revealing the 3D Cosmic Web through Gravitationally Constrained Neural Fields”. In: The Thirteenth International Conference on Learning Representations. Oct. 4, 2024. URL: https://openreview.net/forum?id=Ax0i933gtp (visited on 07/11/2025).

[354] Brandon Zhao et al. “Single View Refractive Index Tomography with Neural Fields”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2024.

[355] Hongkai Zheng et al. “InverseBench: Benchmarking Plug-and-Play Diffusion Priors for Inverse Problems in Physical Sciences”. In: The Thirteenth International Conference on Learning Representations. Oct. 4, 2024. URL: https://openreview.net/forum?id=U3PBITXNG6.

[356] Ellen D. Zhong et al. “CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks”. In: *Nature Methods* 18.2 (Feb. 1, 2021), pp. 176–185. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01049-4. URL: https://doi.org/10.1038/s41592-020-01049-4.

[357] Xingyu Zhou. *On the Fenchel Duality between Strong Convexity and Lipschitz Continuous Gradient*. Mar. 17, 2018. DOI: 10.48550/arXiv.1803.06573. arXiv: 1803.06573[math]. URL: http://arxiv.org/abs/1803.06573 (visited on 06/27/2025).

[358] Huaisheng Zhu, Teng Xiao, and Vasant G. Honavar. “DSPO: Direct Score Preference Optimization for Diffusion Model Alignment”. In: The Thirteenth International Conference on Learning Representations. Oct. 4, 2024.

URL: https://openreview.net/forum?id=xyfb9HHvMe (visited on 07/28/2025).

[359] Jun-Yan Zhu et al. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017 IEEE International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2017, pp. 2242–2251. DOI: 10.1109/ICCV.2017.244. URL: https://ieeexplore.ieee.org/document/8237506 (visited on 06/27/2025).

[360] Yuanzhi Zhu et al. "Denoising Diffusion Models for Plug-and-Play Image Restoration". In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). ISSN: 2160-7516. June 2023, pp. 1219–1229. DOI: 10.1109/CVPRW59228.2023.00129. URL: https://ieeexplore.ieee.org/document/10208800 (visited on 07/03/2025).

[361] Daniel Zoran and Yair Weiss. "From learning models of natural image patches to whole image restoration". In: *2011 International Conference on Computer Vision*. 2011 International Conference on Computer Vision. ISSN: 2380-7504. Nov. 2011, pp. 479–486. DOI: 10.1109/ICCV.2011.6126278. URL: https://ieeexplore.ieee.org/document/6126278 (visited on 06/25/2025).

[362] Daniel Zoran and Yair Weiss. "Natural Images, Gaussian Mixtures and Dead Leaves". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/e97ee2054defb209c35fe4dc94599061-Paper.pdf.

[363] S. Zörner et al. "Measurement of the elasticity modulus of soft tissues". In: *Journal of Biomechanics* 43.8 (May 28, 2010), pp. 1540–1545. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2010.01.035. URL: https://www.sciencedirect.com/science/article/pii/S0021929010000734.