GATSBI: GENERATIVE ADVERSARIAL TRAINING FOR SIMULATION-BASED INFERENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Simulation-based inference (SBI) refers to statistical inference on stochastic models for which we can generate samples, but not compute likelihoods. Like SBI algorithms, generative adversarial networks (GANs) do not require explicit likelihoods. We study the relationship between SBI and GANs, and introduce GATSBI, an adversarial approach to SBI. GATSBI reformulates the variational objective in an adversarial setting to learn implicit posterior distributions. Inference with GATSBI is amortised across observations, works in high-dimensional posterior spaces and supports implicit priors. We evaluate GATSBI on two SBI benchmark problems and on two high-dimensional simulators. On a model for wave propagation on the surface of a shallow water body, we show that GATSBI can return well-calibrated posterior estimates even in high dimensions. On a model of camera optics, it infers a high-dimensional posterior given an implicit prior, and performs better than a state-of-the-art SBI approach. We also show how GATSBI can be extended to perform sequential posterior estimation to focus on individual observations. Overall, GATSBI opens up opportunities for leveraging advances in GANs to perform Bayesian inference on high-dimensional simulation-based models.

1 INTRODUCTION

Hypothesis-making in many scientific disciplines relies on stochastic simulators that—unlike expressive statistical models such as neural networks—have domain-relevant, interpretable parameters. Finding the parameters θ of a simulator that reproduce the observed data x_{o} constitutes an inverse problem. In order to use these simulators to formulate further hypotheses and experiments, one needs to obtain uncertainty estimates for the parameters and allow for multi-valuedness, i.e., different candidate parameters accounting for the same observation. These requirements are met by Bayesian inference, which attempts to approximate the posterior distribution $p(\theta|x_{\mathrm{o}})$. While a variety of techniques exist to calculate posteriors for scientific simulators which allow explicit likelihood calculations $p(x|\theta)$, inference on black-box simulators with intractable likelihoods a.k.a. 'simulation-based inference' (Cranmer et al., 2020), poses substantial challenges. In traditional, so-called Approximate Bayesian Computation (ABC) approaches to SBI (Beaumont et al., 2002, 2009; Marjoram et al., 2003; Sisson et al., 2007), one samples parameters θ from a prior $\pi(\theta)$ and accepts only those parameters for which the simulation output $x \sim p(x|\theta)$ is close to the observation $d(x, x_{\mathrm{o}}) < \epsilon$. With increasing data dimensionality N, this approach incurs an exponentially growing simulation expense ('curse of dimensionality'); it also requires suitable choices of distance function d and acceptance threshold ϵ.

Recent SBI algorithms (e.g., Greenberg et al., 2019; Gutmann and Corander, 2015; Hermans et al., 2020; Lueckmann et al., 2017; Meeds and Welling, 2014; Papamakarios and Murray, 2016; Radev et al., 2020; Thomas et al., 2021) draw on advances in machine learning and are often based on Gaussian Processes (Rasmussen and Williams, 2006) or neural networks for density estimation (e.g. using normalizing flows; Papamakarios et al., 2021). While this has led to substantial improvements over classical approaches, and numerous applications in fields such as cosmology (Cole et al., 2021), neuroscience (Gonçalves et al., 2020) or robotics (Muratore et al., 2021), statistical inference for high-dimensional parameter spaces is an open challenge (Cranmer et al., 2020; Lueckmann et al., 2021): This prevents the application of these methods to many real-world scientific simulators.

In machine learning, generative adversarial networks (GANs, Goodfellow et al., 2014) have emerged as powerful tools for learning generative models of high-dimensional data, for instance, images (Isola

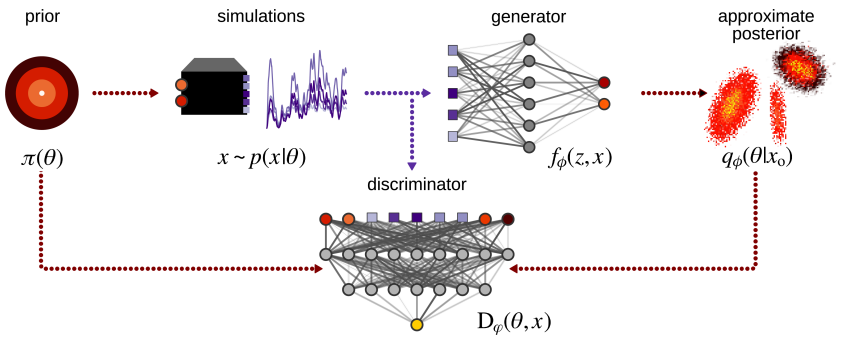

Figure 1: **GATSBI uses a conditional generative adversarial network (GAN) for simulation-based inference (SBI)**. We sample parameters θ from a prior $\pi(\theta)$, and use them to generate synthetic data x from a black-box simulator. The GAN generator learns an implicit approximate posterior $q_\phi(\theta|x)$, i.e., it learns to generate posterior samples θ' given data x. The discriminator is trained to differentiate between θ' and θ, conditioned on x.

et al., 2016; Radford et al., 2015; Zhu et al., 2017). Using a minimax game between a 'generator' and a 'discriminator', GANs can generate samples which look uncannily similar to the data they have been trained on. Like SBI algorithms, adversarial training is 'likelihood-free', i.e., it does not require explicit evaluation of a density function, and also relies on comparing empirical and simulated data. This raises the question of whether GANs and SBI solve the same problem, and in particular whether and how adversarial training can help scale SBI to high-dimensional regimes.

Here, we address these questions and make the following contributions: First, we show how one can use adversarial training to learn an implicit posterior density for SBI. We refer to this approach as GATSBI (Generative Adversarial Training for SBI)[1]. Second, we show that GATSBI can be formulated as an adversarial variational inference algorithm (Huszár, 2017; Makhzani et al., 2015; Mescheder et al., 2017), thus opening up a potential area of exploration for SBI approaches. Third, we show how it can be extended to refine posterior estimates for specific observations, i.e., sequential posterior estimation. Fourth, on two low-dimensional SBI benchmark problems, we show that GATSBI's performance is comparable to common SBI algorithms. Fifth, we illustrate GATSBI on high-dimensional inference problems which are out of reach for most SBI methods: In a fluid dynamics problem from earth science, we learn an implicit posterior over 100 parameters. Furthermore, we show that GATSBI (in contrast to most SBI algorithms) can be used on problems in which the *prior* is only specified implicitly: by recasting image-denoising as an SBI problem, we learn a posterior over (784-dimensional) images. Finally, we discuss the opportunities and limitations of GATSBI relative to current SBI algorithms, and chart out avenues for future work.

2 METHODS

2.1 SIMULATION-BASED INFERENCE AND ADVERSARIAL NETWORKS

Simulation-based inference (SBI) targets stochastic simulators with parameters θ that we can use to generate samples x. In many scientific applications, simulators are 'black-box': we cannot evaluate the likelihood $p(x|\theta)$, take derivatives through the simulator, or access its internal random numbers. The goal of SBI is to infer the parameters θ, given empirically observed data x_o i.e., to obtain the posterior distribution $p(\theta|x_\mathrm{o})$. In most applications of existing SBI algorithms, the simulator is a mechanistic forward model depending on a low (~ 10) number of parameters θ.

Generative adversarial networks (GANs) (Goodfellow et al., 2014) consist of two networks trained adversarially. The generator $f_\phi(z)$ produces samples x from latent variables z; the discriminator D_ψ acts as a similarity measure between the distribution of generated x and observations x_o from a distribution $p_\mathrm{data}(x_\mathrm{o})$. The two networks are pitted against each other: the generator attempts to produce samples that would fool the discriminator; the discriminator is trained to tell apart generated

[1]Note that the name is shared by recently proposed, and entirely unrelated, algorithms (Min et al., 2021; Yu et al., 2020)

and ground-truth samples. This takes the form of a minimax game $\min_\phi \max_\psi L(\phi, \psi)$ between the networks, with cross-entropy as the value function:

$$L(\phi, \psi) = \mathbb{E}_{p_{\text{data}}(x_\text{o})} \log \text{D}_\psi(x_\text{o}) + \mathbb{E}_{p(z)} \log(1 - \text{D}_\psi(f_\phi(z))).$$

At the end of training, if the networks are sufficiently expressive, samples from f_ϕ are hard to distinguish from the ground-truth samples, and the generator approximates the observed data distribution $p_{\text{data}}(x_\text{o})$. GANs can also be used to perform (implicit) *conditional* density estimation, by making the generator and discriminator conditional on external covariates t (Mirza and Osindero, 2014). Note that the GAN formulation does not evaluate the intractable distribution $p_{\text{data}}(x_\text{o})$ or $p_{\text{data}}(x_\text{o}|t)$ at any point. Thus, GANs and SBI appear to be solving similar problems: They are both 'likelihood-free' i.e., they do not evaluate the target distribution, but rather exclusively work with samples from it.

Nevertheless, GANs and SBI differ in multiple ways: GANs use neural networks, and thus are always differentiable, whereas the numerical simulators in SBI may not be. Furthermore, the latent variables for the GAN generator are typically accessible, while in SBI, one may not have access to the simulator's internal random variables. Finally, in GANs, the goal is to match the distribution of the generated samples x to that of the observed data x_o, and parameters of the generator are learnt as point estimates. In contrast, the goal of SBI is to learn a distribution over parameters θ of a simulator consistent with both the observed data x_o *and* prior knowledge about the parameters. However, conditional GANs *can* be used to perform conditional density estimation. How can we repurpose them for SBI? We propose an approach using adversarial training for SBI, by employing GANs as implicit density estimators rather than generative models of data.

2.2 Generative Adversarial Training for SBI: GATSBI

Our approach leverages conditional GANs for SBI. We train a deep generative neural network f_ϕ with parameters ϕ, i.e., f takes as inputs observations x and noise z from a fixed source with distribution $p(z)$, so that $f_\phi(x, z)$ deterministically transforms z to generate θ. This generative network f is an implicit approximation $q_\phi(\theta|x)$ of the posterior distribution, i.e.,

$$\left.\begin{array}{l} z \sim p(z) \\ \theta = f_\phi(z, x) \end{array}\right\} \Rightarrow \theta \sim q_\phi(\theta|x). \tag{1}$$

We train f adversarially: We sample many parameters θ from the prior, and for each θ, simulate $x \sim p(x|\theta)$ i.e., we sample tuples $(\theta, x) \sim \pi(\theta)p(x|\theta) = p(\theta|x)p(x)$. These samples from the joint constitute our training data. We define a discriminator D, parameterised by ψ, to differentiate between samples θ drawn from the approximate posterior $q_\phi(\theta|x)$ and the true posterior $p(\theta|x)$. The discriminator is conditioned on x. We calculate the cross-entropy of the discriminator outputs, and maximize it with respect to the discriminator parameters ψ and minimize it with respect to the generator parameters ϕ i.e., $\min_\phi \max_\psi L(\phi, \psi)$, with

$$\begin{aligned} L(\phi, \psi) &= \mathbb{E}_{\pi(\theta)p(x|\theta)p(z)} \Big[\log \text{D}_\psi(\theta, x) + \log \big(1 - \text{D}_\psi(f_\phi(z, x), x)\big) \Big] \\ &= \mathbb{E}_{p(x)} \Big[\mathbb{E}_{p(\theta|x)} \Big(\log \text{D}_\psi(\theta, x) \Big) + \mathbb{E}_{q(\theta|x)} \Big(\log \big(1 - \text{D}_\psi(\theta, x)\big) \Big) \Big] \end{aligned} \tag{2}$$

This conditional GAN value function targets the desired posterior $p(\theta|x)$:

Proposition 1. *Given a fixed generator f_ϕ, the discriminator D_{ψ^*} maximizing equation 2 satisfies*

$$\text{D}_{\psi^*}(\theta, x) = \frac{p(\theta|x)}{p(\theta|x) + q_\phi(\theta|x)}, \tag{3}$$

and the corresponding loss function for the generator parameters is the Jensen-Shannon divergence (JSD) between the true and approximate posterior,

$$L_{\psi^*}(\phi) = 2\,\text{JSD}(p(\theta|x)||q_\phi(\theta|x)) - \log 4.$$

Using sufficiently flexible networks, the GATSBI generator trained with the cross-entropy loss converges in the limit of infinitely many samples to the true posterior, since the JSD is minimised if and only if $q_\phi(\theta|x) = p(\theta|x)$. The proof directly follows Proposition 1 and 2 from Goodfellow et al. (2014) and is given in detail in App. A.1. GATSBI thus performs *amortised inference*, i.e., learns a single inference network which can be used to perform inference rapidly and efficiently for a wide range of potential observations, without having to re-train the network for each new observation. The training steps for GATSBI are described in App. B.

2.3 RELATING GATSBI TO EXISTING SBI APPROACHES

There have been previous attempts at using GANs to fit black-box scientific simulators to empirical data. However, unlike GATSBI, these approaches do not perform Bayesian inference: Kim et al. (2020) and Louppe et al. (2017) train a discriminator to differentiate between samples from the black-box simulator and empirical observations. They use the discriminator to adjust a proposal distribution for simulator parameters until the resulting simulated data and empirical data match. However, neither approach returns a posterior distribution representing the balance between a prior distribution and a fit to empirical data. Similarly, Jethava and Dubhashi (2017) use two generator networks, one to approximate the prior and one to learn summary statistics, as well as two discriminators. This scheme yields a generator for parameters leading to realistic simulations, but does not return a posterior over parameters. Parikh et al. (2020) train a conditional GAN (c-GAN), as well as two additional GANs (r-GAN and t-GAN) within an active learning scheme, in order to solve 'population of models' problems (Britton et al., 2013; Lawson et al., 2018). While c-GAN exhibits similarities with GATSBI, the method does not do Bayesian inference, but rather solves a constrained optimization problem. Adler and Öktem (2018) use Wasserstein-GANs (Arjovsky et al., 2017) to solve Bayesian inverse problems in medical imaging. While their set-up similar to GATSBI, the Wasserstein metric imposes stronger conditions for the generator to converge to the true posterior (App. Sec. A.3).

Several SBI approaches train discriminators to circumvent evaluation of the intractable likelihood (e.g., Cranmer et al., 2015; Gutmann et al., 2018; Pham et al., 2014; Thomas et al., 2021). Hermans et al. (2020) train density-ratio estimators by discriminating samples $(\theta, x) \sim \pi(\theta)p(x|\theta)$ from $(\theta, x) \sim \pi(\theta)p(x)$, which can be used to sample from the posterior. Unlike GATSBI, this requires a potentially expensive step of MCMC sampling. A closely related approach (Durkan et al., 2020; Greenberg et al., 2019) directly targets the posterior, but requires a density estimator that can be evaluated—unlike GATSBI, where the only restriction on the generator is that it remains differentiable.

Finally, ABC and synthetic likelihood methods have been developed to tackle problems in high-dimensional parameter spaces (>100-dimensions) although these require model assumptions regarding the Gaussianity of the likelihood (Ong et al., 2018), or low-dimensional summary statistics (Kousathanas et al., 2015; Rodrigues et al., 2019), in addition to MCMC sampling.

2.4 RELATION TO ADVERSARIAL INFERENCE

GATSBI reveals a direct connection between SBI and adversarial variational inference. Density estimation approaches for SBI usually use the forward Kullback-Leibler divergence (D_{KL}). The reverse D_{KL} can offer complementary advantages (e.g. by promoting mode-seeking rather than mode-covering (Turner and Sahani, 2011)), but is hard to compute and optimise for conventional density estimation approaches. We here show how GATSBI, by using an approach similar to adversarial variational inference, performs an optimization that also finds a minimum of the reverse D_{KL}.

First, we note that density estimation approaches for SBI involve maximising the approximate log posterior over samples simulated from the prior. This is equivalent to minimising the *forward* D_{KL}:

$$\begin{aligned} L_{\mathrm{fwd}}(\phi) &= \mathbb{E}_{p(x)} D_{\mathrm{KL}}(p(\theta|x)||q_\phi(\theta|x)) \\ &= -\mathbb{E}_{p(x)p(\theta|x)} \log q_\phi(\theta|x) + \mathbb{E}_{p(x)p(\theta|x)} \log p(\theta|x). \end{aligned} \tag{4}$$

rather than reverse D_{KL}:

$$\begin{aligned} L_{\mathrm{rev}}(\phi) &= \mathbb{E}_{p(x)} D_{\mathrm{KL}}(q_\phi(\theta|x)||p(\theta|x)) \\ &= \mathbb{E}_{p(x)q_\phi(\theta|x)} \log \frac{q_\phi(\theta|x)}{p(\theta|x)}, \end{aligned} \tag{5}$$

This is because it is feasible to compute L_{fwd} with an intractable likelihood: it only involves evaluating the approximate posterior under *samples* from the true joint distribution $p(\theta, x) = p(x|\theta)\pi(\theta)$. The problem with the reverse D_{KL}, however, is that it requires evaluating the true posterior $p(\theta|x)$ under samples from the approximate posterior $q_\phi(\theta|x)$, which is impossible with an intractable likelihood. This is a problem also shared by variational-autoencoders (VAEs, Kingma and Welling, 2013) but which can be solved using adversarial inference.

A VAE typically consists of an encoder network $q_\phi(u|x)$ approximating the posterior over latents u given data x, and a decoder network approximating the likelihood $p_\alpha(x|u)$. The parameters ϕ and α

of the two networks are optimised by maximising the Evidence Lower Bound (ELBO):

$$\begin{aligned}\mathrm{ELBO}(\alpha, \phi) &= -\mathbb{E}_{p(x)}\left[D_{\mathrm{KL}}(q_\phi(u|x)||p(u)) + \mathbb{E}_{q_\phi(u|x)}[\log p_\alpha(x|u)]\right] \\ &= -\mathbb{E}_{p(x)q_\phi(u|x)} \log \frac{q_\phi(u|x)p(x)}{p(u)p_\alpha(x|u)} + \text{const.} \\ &= -D_{\mathrm{KL}}(q_\phi(u|x)p(x)||p(u)p_\alpha(x|u)) + \text{const.} \end{aligned} \tag{6}$$

Note that the D_{KL} is minimised when maximising the ELBO. If any of the distributions used to compute the ratio in the D_{KL} in equation 6 is intractable, one can do adversarial inference, i.e., the ratio can be approximated using a discriminator (see Proposition 1) and $q_\phi(u|x)$ is recast as the generator. Various adversarial inference approaches approximate different ratios, depending on which of the distributions is intractable (see App. Table 1 for a non-exhaustive comparison of different discriminators and corresponding losses), and their connection to GANs has been explored extensively (Chen et al., 2018; Hu et al., 2017; Mohamed and Lakshminarayanan, 2016; Srivastava et al., 2017).

Recasting SBI into the VAE framework, the intractable distribution is the likelihood $p_\alpha(x|\theta)$,[2] and α is a constant, i.e., the simulator only has parameters to do inference over. Accordingly, $p(x)$ is defined as the marginal likelihood. In this setting, the objective for SBI becomes the reverse D_{KL}, and we can use an adversarial method, e.g. GATSBI, to minimise it for learning the posterior $q_\phi(\theta|x)$. There is an additional subtlety at play here: GATSBI's generator loss, given an optimal discriminator, is not the reverse D_{KL}, but the JSD (see Proposition 1). However, following the proof for Proposition 3 in Mescheder et al. (2017), one can show that the approximate posterior that minimises the JSD *also* maximises the ELBO (see App. A.3). Thus, GATSBI is an adversarial inference method that can be used to perform SBI using the *reverse* D_{KL}.

Note that the converse is not true: not all adversarial inference methods can be adapted for SBI. Specifically, some adversarial inference methods require explicit evaluation of the likelihood. Of those methods listed in Table 1, only likelihood-free variational inference (LFVI) (Tran et al., 2017) is explicitly used to perform inference for a simulator with intractable likelihood. LFVI is defined as performing inference over latents u and global parameters β shared across multiple observations x, which are sampled from an empirical distribution $p'(x) \neq p(x)$, i.e., it learns $q_\phi(u, \beta|x)$. However, if $p'(x) = p(x)$, and β is constant, i.e., the simulator does not have tunable parameters (but only parameters one does inference over) then LFVI and GATSBI become (almost) equivalent: LFVI differs by a single term in the loss function, and we empirically found the methods to yield similar results (see App. Fig. 6). We discuss LFVI's relationship to GATSBI in more detail in App. A.3.

2.5 Sequential GATSBI

The GATSBI formulation in the previous sections allows us to learn an amortised posterior $p(\theta|x)$ for *any* observation x. However, one is often interested in good posterior samples for a particular experimental observation x_{o}—and not for all possible observations x. This can be achieved by training the density estimator using samples θ from a proposal prior $\tilde{\pi}(\theta)$ instead of the usual prior $\pi(\theta)$. The proposal prior ideally produces samples θ that are localised around the modes of $p(\theta|x_{\mathrm{o}})$ and can guide the density estimator towards inferring a posterior for x_{o} using less training data. To this end, sequential schemes using proposal distributions to generate training data over several rounds of inference (e.g. by using the approximate posterior from previous rounds as the proposal prior for subsequent rounds) can be more sample efficient (Lueckmann et al., 2021).

Sequential GATSBI uses a scheme similar to other sequential SBI methods (e.g., Greenberg et al., 2019; Hermans et al., 2020; Lueckmann et al., 2017; Papamakarios and Murray, 2016; Papamakarios et al., 2019): the generator from the current round of posterior approximation becomes the proposal distribution in the next round. However, using samples from a proposal prior $\tilde{\pi}(\theta)$ to train the GAN would lead to the GATSBI generator learning a *proposal posterior* $\tilde{p}(\theta|x)$ rather than the true posterior $p(\theta|x)$. Previous sequential methods solve this problem either by post-hoc correcting the learned density (Papamakarios and Murray, 2016), by adjusting the loss function (Lueckmann et al., 2017), or by reparametrizing the posterior network (Greenberg et al., 2019). While these approaches might work for GATSBI (see App. A.4 for an outline), we propose a different approach: at training time, we

[2]Note the difference in notation: the parameters θ correspond to the latents u of a VAE, and the simulator in SBI to the decoder of the VAE.

sample the latents z—that the generator transforms into parameters θ—from a *corrected* distribution $p_t(z)$, so that the samples θ are implicitly generated from an approximate proposal posterior $\tilde{q}_\phi(\theta|x)$:

$$\left.\begin{array}{l} z \sim p_t(z) \\ \theta = f_\phi(z, x) \end{array}\right\} \Rightarrow \theta \sim \tilde{q}_\phi(\theta|x), \tag{7}$$

where $p_t(z) = p(z)\ (\omega(f_\phi(z,x),x))^{-1}$, $\omega(\theta, x) = \frac{\pi(\theta)}{\tilde{\pi}(\theta)} \frac{\tilde{p}(x)}{p(x)}$, and $\tilde{p}(x)$ is the marginal likelihood under samples from the proposal prior. Since $\tilde{\pi}(\theta)$, $p(x)$ and $\tilde{p}(x)$ are intractable, we approximate $\omega(\theta, x)$ using ratio density estimators for $\frac{\pi(\theta)}{\tilde{\pi}(\theta)}$ and $\frac{\tilde{p}(x)}{p(x)}$. This ensures that at inference time, the GATSBI generator transforms samples from $p(z)$ into an approximation of the true posterior $p(\theta|x)$ with highest accuracy at $x = x_\mathrm{o}$ (see App. A.4, B). Azadi et al. (2019); Che et al. (2020) use similar approaches to improve GAN training. Note that this scheme may be used with other inference algorithms evaluating a loss under approximate posterior samples, e.g. LFVI (see Sec. 2.4).

3 Results

We first investigate GATSBI on two common benchmark problems in SBI (Lueckmann et al., 2021). We then show how GATSBI infers posteriors in a high-dimensional regime where most state-of-the-art SBI algorithms are inadequate, as well as in a problem with an implicit prior (details in App. D.1).

3.1 Benchmark problems

We selected two benchmark problems with low-dimensional posteriors on which state-of-the-art SBI algorithms have been tested extensively. We used the benchmarking-tools and setup from Lueckmann et al. (2021). Briefly, we compared GATSBI against five SBI algorithms: classical rejection ABC (Tavaré et al., 1997, REJ-ABC), sequential Monte Carlo ABC (Beaumont et al., 2009, SMC-ABC) and the single-round variants of flow-based neural likelihood estimation (Papamakarios et al., 2019, NLE), flow-based neural posterior estimation (Greenberg et al., 2019; Papamakarios and Murray, 2016, NPE) and neural ratio estimation (Durkan et al., 2020; Hermans et al., 2020, NRE). GANs have previously been shown to be poor density estimators for low-dimensional distributions (Zaheer et al., 2017). Thus, while we include these examples to provide empirical support that GATSBI works, we do not expect it to perform as well as state-of-the-art SBI algorithms on these tasks. We compared samples from the approximate posterior against a reference posterior using a classifier trained on the two sets of samples (classification-based two-sample test, C2ST). A C2ST accuracy of 0.5 (chance level) corresponds to perfect posterior estimation.

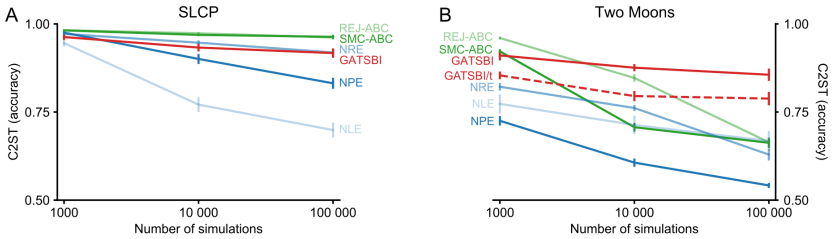

Figure 2: **GATSBI performance on benchmark problems.** Mean C2ST score ($\pm$ standard error of the mean) for 10 observations each. A. The classification score for GATSBI (red) on SCLP decreases with increasing simulation budget, and is comparable to NRE. It outperforms rejection ABC and SMC-ABC, but has worse performance than NPE and NLE. B. The classification score for GATSBI (red) on the two-moons model decreases with increasing simulation budget, is comparable to REJ-ABC and SMC-ABC for simulation budgets of 1k and 10k, and is outperformed by all other methods for a 100k simulation budget. However, GATSBI's classification score improves when its architecture and optimization parameters are tuned to the two-moons model (red, dashed).

"Simple Likelihood Complex Posterior" (SLCP) is a challenging SBI problem designed to have a simple likelihood and a complex posterior (Papamakarios et al., 2019). The prior $\pi(\theta)$ is a 5-dimensional uniform distribution and the likelihood for the 8-dimensional x is a Gaussian whose

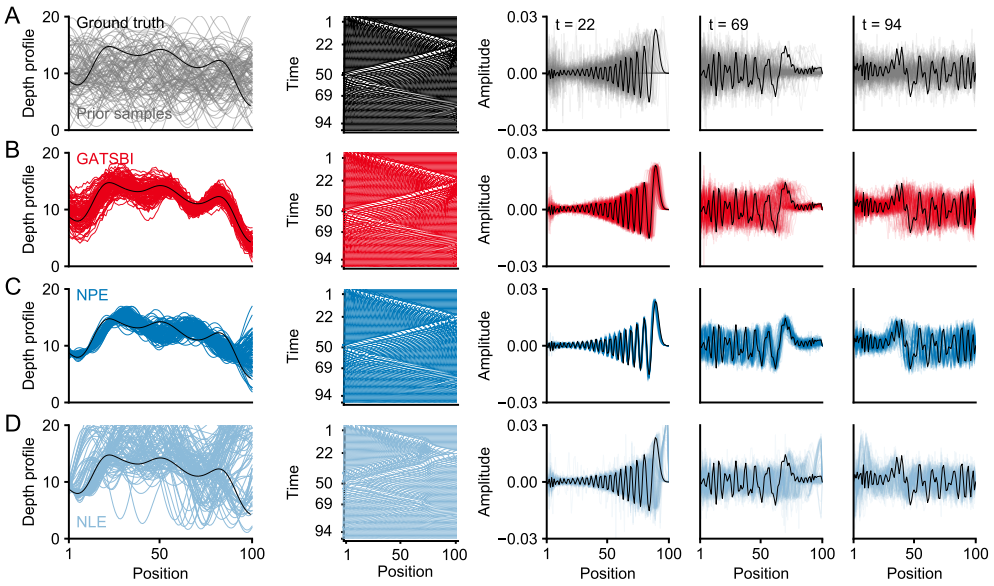

Figure 3: **Shallow water model inference with GATSBI, NPE and NLE.** A. Ground truth, observation and prior samples. Left: ground-truth depth profile and prior samples. Middle: surface wave simulated from ground-truth profile as a function of position and time. Right: wave amplitudes at three different fixed times for ground-truth depth profile (black), and waves simulated from multiple prior samples (gray). B. GATSBI inference. Left: posterior samples (red) versus ground-truth depth profile (black). Middle: surface wave simulated from a single GATSBI posterior sample. Right: wave amplitudes for multiple GATSBI posterior samples, at three different times (red). Ground-truth waves in black. C. NPE inference. Panels as in B. D. NLE inference.

mean and variance are nonlinear functions of θ. This induces a posterior over θ with four symmetrical modes and vertical cut-offs (see App. Fig. 7), making it a challenging inference problem. We trained all algorithms on a budget of 1k, 10k and 100k simulations. The GATSBI C2ST scores were on par with NRE, better than REJ-ABC and SMC-ABC, but below the NPE and NLE scores (Figure 2A).

The **"Two-Moons Model"** (Greenberg et al., 2019) has a simple likelihood and a bimodal posterior, where each mode is crescent-shaped (see App. Fig. 8). Both x and θ are two-dimensional. We initially trained GATSBI with the same architecture and settings as for the SLCP problem. While the classification score for GATSBI (see Figure 2B) was on par with that of REJ-ABC and SMC-ABC for 1k simulations, performance did not significantly improve when increasing training budget. When we tuned the GAN architecture and training hyperparameters specifically for this model, we found a performance improvement. Lueckmann et al. (2021) showed that performance can be significantly improved with sequential inference schemes. We found that although using sequential schemes with GATSBI leads to small performance improvements, it still did not ensure performance on-par with the flow-based methods (see App. Fig. 9). Moreover, since it had double the number of hyperparameters as amortised GATSBI, sequential GATSBI also required more hyperparameter tuning.

3.2 High-dimensional inference problems

We showed above that, on low-dimensional examples, GATSBI does not perform as well as the best SBI methods. However, we expected that it would excel on problems on which GANs excel: high-dimensional and structured parameter spaces, in which suitably designed neural networks can be used as powerful classifiers. We apply GATSBI to two problems with $\sim$100-dimensional parameter spaces, one of which also has an implicit prior.

Shallow Water Model The 1D shallow water equations describe the propagation of an initial disturbance across the surface of a shallow basin with an irregular depth profile and bottom friction (Figure 3A). They are relevant to oceanographers studying the dynamics on the continental sea shelf

(Backhaus, 1983; Holton, 1973). We discretised the partial differential equations on a 100-point spatial grid, obtaining a simulator parameterised by the depth profile of the basin, $\theta \in \mathbb{R}^{100}$ (substantially higher than most previous SBI-applications, which generally are ≈ 10 dim). The resulting amplitude of the waves evolving over a total of 100 time steps constitutes the 10k-dimensional raw observation from which we aim to infer the depth profile. The observation is taken into the Fourier domain, where both the real and imaginary parts receive additive noise and are concatenated, entering the inference pipeline as an array $x \in \mathbb{R}^{20k}$. Our task was to estimate $p(\theta|x)$, i.e., to infer the basin depth profile from a noisy Fourier transform of the surfaces waves. We trained GATSBI with 100k depth profiles sampled from the prior and the resulting surface wave simulations. For comparison, we trained NPE and NLE on the same data, using the same embedding net as the GATSBI discriminator to reduce the high dimensionality of the observations $x \in \mathbb{R}^{20k}$ to $\mathbb{R}^{100}$.

Samples from the GATSBI posterior captured the structure of the underlying ground-truth depth profile (Figure 3B, left). In comparison, only NPE, but not NLE, captured the true shape of the depth profile (Figure 3C, D, left). In addition, the inferred posterior means for GATSBI and NPE, but not NLE, were correlated with the ground-truth parameters (0.88 ± 0.10 for GATSBI and 0.91 ± 0.09 for NPE, mean $\pm$ standard deviation across 1000 different observations). Furthermore, while GATSBI posterior predictive samples did not follow the ground-truth observations as closely as NPE's, they captured the structure and were not as phase-shifted as NLE's (Figure 3 B-D). This is expected since wave-propagation speed depends strongly on depth profile, and thus incorrect inference of the depth parameters leads to temporal shifts between observed and inferred waves. Finally, simulation-based calibration (SBC) (Talts et al., 2020) shows that both NPE and GATSBI provide well-calibrated posteriors on this challenging, 100 dimensional inference problem (clearly, neither is perfectly calibrated, Figure 4). This realistic simulator example illustrates that GATSBI can learn posteriors with a dimensionality far from feasible for most SBI algorithms. It performs similar to NPE, previously reported to be a powerful approach on high-dimensional inference problems (Lueckmann et al., 2021).

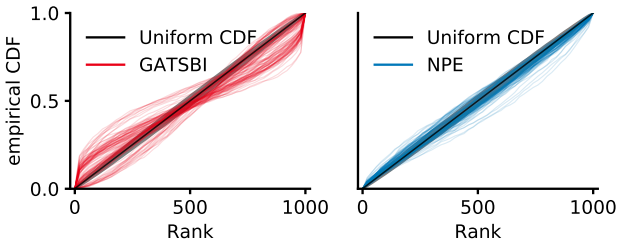

Figure 4: SBC results on shallow water model. Empirical cdf of SBC ranks, one line per posterior dimension (red, GATSBI; blue, NPE). In gray, region showing 99% of deviation expected under the ideal uniform cdf (black).

Noisy Camera Model Finally, we demonstrate a further advantage of GATSBI over many SBI algorithms: GATSBI (like LFVI) can be applied in cases in which the prior distribution is only specified through an implicit model. We illustrate this by considering inference over images from a noisy and blurred camera model. We interpret the model as a black-box simulator. Thus, the clean images θ are samples from an implicit prior over images $\pi(\theta)$, and the simulated sensor activations are observations x. We can then recast the task of denoising the images as an inference problem (we note that our goal is *not* to advance the state-of-the-art for super-resolution algorithms, for which multiple powerful approaches are available, e.g. Kim et al., 2016; Zhang et al., 2018).

Since images are typically high-dimensional objects, learning a posterior over images is already a difficult problem for many SBI algorithms – either due to the implicit prior or the curse of dimensionality making MCMC sampling impracticable. However, the implicit prior poses no challenge for GATSBI since it is trained only with samples.

We chose the EMNIST dataset (Cohen et al., 2017) with 800k 28×28-dimensional images as the implicit prior. This results in an inference problem in a 784-dimensional parameter space—much higher dimensionality than in typical SBI applications. The GATSBI generator was trained to produce clean images, given only the blurred image as input, while the discriminator was trained with clean images from the prior and the generator, and the corresponding blurred images. GATSBI indeed recovered crisp sources from blurred observations: samples from the GATSBI posterior given blurred observations accurately matched the underlying clean ground-truth images (see Figure 5). In contrast, NPE did not produce coherent posterior samples, and took substantially longer to train than GATSBI (NLE and NRE are not applicable due to the implicit prior).

Figure 5: **Camera model inference.** Top: ground-truth parameter samples from the implicit prior and corresponding blurred camera model observations. Bottom: mean and standard deviation (SD) of inferred GATSBI and NPE posterior samples for matching observation from top.

4 Discussion

We proposed a new method for simulation-based inference in stochastic models with intractable likelihoods. We used conditional GANs to fit a neural network functioning as an implicit density, allowing efficient, scalable sampling. We clarified the relationship between GANs and SBI, and showed how our method is related to adversarial VAEs that learn implicit posteriors over latent variables. We showed how GATSBI can be extended for sequential estimation, fine-tuning inference to a particular observation at the expense of amortisation. In contrast to conventional density-estimation based SBI-methods, GATSBI targets the *reverse* D_{KL}—by thus extending both amortised and sequential SBI, we made it possible to explore the advantages of the forward and reverse D_{KL}.

We found GATSBI to work adequately on two low-dimensional benchmark problems, though it did not outperform neural network-based methods. However, its real potential comes from the ability of GANs to learn generative models in high-dimensional problems. We demonstrated this by using GATSBI to infer a well-calibrated posterior over a 100-dimensional parameter space for the shallow water model, and a 784-dimensional parameter space for the camera model. This is far beyond what is typically attempted with SBI: indeed we found that GATSBI outperformed NPE on the camera model, and performed similarly to NPE and substantially outperformed NLE on the shallow water model (both SBI approaches were previously shown to be more powerful than others in high dimensions; Lueckmann et al., 2021). Moreover, we showed that GATSBI can learn posteriors in a model in which the prior is only defined implicitly (i.e., through a database of samples, not a parametric model).

Despite the advantages that GATSBI confers over other SBI algorithms in terms of flexibility and scalability, it comes with computational and algorithmic costs. GANs are notoriously difficult to train—they are highly sensitive to hyperparameters (as demonstrated in the two-moons example) and problems such as mode collapse are common. Moreover, training time is significantly higher than in other SBI algorithms. In addition, since GATSBI encodes an implicit posterior, i.e., a density from which we can sample but cannot evaluate, it yields posterior samples but not a parametric model. Thus, GATSBI is unlikely to be the method of choice for low-dimensional problems. Similarly, sequential GATSBI requires extensive hyperparameter tuning in order to produce improvements over amortised GATSBI, with the additional cost of training a classifier to approximate the correction factor. Hence, sequential GATSBI might be useful in applications where the computational cost of training is offset by having an expensive simulator. We note that GATSBI might benefit from recent improvements on GAN training speed and sample efficiency (Sauer et al., 2021).

Overall, we expect that GATSBI will be particularly impactful on problems in which the parameter space is high-dimensional and has GAN-friendly structure, e.g. images. Spatially structured models and data abound in the sciences, from climate modelling to ecology and economics. High-dimensional parameter regimes are common, but inaccessible for current SBI algorithms. Building on the strengths of GANs, we expect that GATSBI will help close this gap and open up new application-domains for SBI, and new directions for building SBI methods employing adversarial training.

ETHICS STATEMENT

Since this is a purely exploratory and theoretical work based on simulated data, we do not foresee major ethical concerns. However, we also note that highly realistic GANs have a potential of being exploited for applications with negative societal impacts e.g. deep fakes (Tolosana et al., 2020).

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we provide detailed information about simulated data, model set up, training details and experimental results in the Appendix. In order to avoid de-anonymization issues, we will only release code upon publication.

REFERENCES

Jonas Adler and Ozan Öktem. Deep bayesian inversion, 2018.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 06–11 Aug 2017.

Samaneh Azadi, Catherine Olsson, Trevor Darrell, Ian Goodfellow, and Augustus Odena. Discriminator Rejection Sampling. *arXiv:1810.06758 [cs, stat]*, 2019. arXiv: 1810.06758.

Jan O Backhaus. A semi-implicit scheme for the shallow water equations for application to shelf sea modelling. *Continental Shelf Research*, 2(4):243–254, 1983.

Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.

Mark A Beaumont, Jean-Marie Cornuet, Jean-Michel Marin, and Christian P Robert. Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990, 2009.

Oliver J Britton, Alfonso Bueno-Orovio, Karel Van Ammel, Hua Rong Lu, Rob Towart, David J Gallacher, and Blanca Rodriguez. Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. *Proceedings of the National Academy of Sciences*, 110(23):E2098–E2105, 2013.

Tong Che, Ruixiang Zhang, Jascha Sohl-Dickstein, Hugo Larochelle, Liam Paull, Yuan Cao, and Yoshua Bengio. Your GAN is Secretly an Energy-based Model and You Should use Discriminator Driven Latent Sampling. *arXiv:2003.06060 [cs, stat]*, 2020.

Liqun Chen, Shuyang Dai, Yunchen Pu, Erjin Zhou, Chunyuan Li, Qinliang Su, Changyou Chen, and Lawrence Carin. Symmetric variational autoencoder and connections to adversarial learning. In *International Conference on Artificial Intelligence and Statistics*, pages 661–669. PMLR, 2018.

Gregory Cohen, Saeed Afshar, Jonathan Tapson, and André van Schaik. EMNIST: an extension of MNIST to handwritten letters, 2017.

Alex Cole, Benjamin Kurt Miller, Samuel J. Witte, Maxwell X. Cai, Meiert W. Grootes, Francesco Nattino, and Christoph Weniger. Fast and credible likelihood-free cosmology with truncated marginal neural ratio estimation, 2021.

Kyle Cranmer, Juan Pavez, and Gilles Louppe. Approximating likelihood ratios with calibrated discriminative classifiers. *arXiv preprint arXiv:1506.02169*, 2015.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

Conor Durkan, Iain Murray, and George Papamakarios. On contrastive learning for likelihood-free inference. In *Proceedings of the 36th International Conference on Machine Learning*, volume 98 of *Proceedings of Machine Learning Research*. PMLR, 2020.

Pedro J Gonçalves, Jan-Matthis Lueckmann, Michael Deistler, Marcel Nonnenmacher, Kaan Öcal, Giacomo Bassetto, Chaitanya Chintaluri, William F Podlaski, Sara A Haddad, Tim P Vogels, David S. Greenberg, and Jakob H. Macke. Training deep neural density estimators to identify mechanistic models of neural dynamics. *eLife*, 2020.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2404–2414. PMLR, 2019.

Michael U. Gutmann and Jukka Corander. Bayesian optimization for likelihood-free inference of simulator-based statistical models, 2015.

Michael U. Gutmann, Ritabrata Dutta, Samuel Kaski, and Jukka Corander. Likelihood-free inference via classification. *Statistics and Computing*, 28(2):411–425, 2018.

Joeri Hermans, Volodimir Begy, and Gilles Louppe. Likelihood-free MCMC with approximate likelihood ratios. In *Proceedings of the 37th International Conference on Machine Learning*, volume 98 of *Proceedings of Machine Learning Research*. PMLR, 2020.

James R Holton. An introduction to dynamic meteorology. *American Journal of Physics*, 41(5): 752–754, 1973.

Zhiting Hu, Zichao Yang, Ruslan Salakhutdinov, and Eric P Xing. On unifying deep generative models. *arXiv preprint arXiv:1706.00550*, 2017.

Ferenc Huszár. Variational inference using implicit distributions. *arXiv preprint arXiv:1702.08235*, 2017.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2016.

Vinay Jethava and Devdatt Dubhashi. Easy high-dimensional likelihood-free inference. *arXiv preprint arXiv:1711.11139*, 2017.

Dongjun Kim, Weonyoung Joo, Seungjae Shin, and Il-Chul Moon. Adversarial likelihood-free inference on black-box generator. *arXiv preprint arXiv:2004.05803*, 2020.

J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Athanasios Kousathanas, Christoph Leuenberger, Jonas Helfer, Mathieu Quinodoz, Matthieu Foll, and Daniel Wegmann. Likelihood-free inference in high-dimensional models, 2015.

Brodie AJ Lawson, Christopher C Drovandi, Nicole Cusimano, Pamela Burrage, Blanca Rodriguez, and Kevin Burrage. Unlocking data sets by calibrating populations of models to data density: A study in atrial electrophysiology. *Science Advances*, 4(1):e1701676, 2018.

Gilles Louppe, Joeri Hermans, and Kyle Cranmer. Adversarial variational optimization of non-differentiable simulators. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1438–1447, 2017.

Jan-Matthis Lueckmann, Pedro J Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H Macke. Flexible statistical inference for mechanistic models of neural dynamics. In *Advances in Neural Information Processing Systems 30*, pages 1289–1299. Curran Associates, Inc., 2017.

Jan-Matthis Lueckmann, Jan Boelts, David Greenberg, Pedro Goncalves, and Jakob Macke. Benchmarking simulation-based inference. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pages 343–351, 2021.

Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.

P Marjoram, J Molitor, V Plagnol, and S Tavare. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26), 2003.

Edward Meeds and Max Welling. Gps-abc: Gaussian process surrogate approximate bayesian computation, 2014.

Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. Adversarial variational Bayes: Unifying variational autoencoders and generative adversarial networks. *34th International Conference on Machine Learning, ICML 2017*, 5:3694–3707, 2017.

Cheol-Hui Min, Jinseok Bae, Junho Lee, and Young Min Kim. Gatsbi: Generative agent-centric spatio-temporal object interaction, 2021.

Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

Shakir Mohamed and Balaji Lakshminarayanan. Learning in implicit generative models. *arXiv preprint arXiv:1610.03483*, 2016.

Fabio Muratore, Fabio Ramos, Greg Turk, Wenhao Yu, Michael Gienger, and Jan Peters. Robot learning from randomized simulations: A review. *arXiv preprint arXiv:2111.00956*, 2021.

Victor M.-H. Ong, David J. Nott, Minh-Ngoc Tran, Scott A. Sisson, and Christopher C. Drovandi. Likelihood-free inference in high dimensions with synthetic likelihood. *Computational Statistics & Data Analysis*, 128:271–291, 2018. ISSN 0167-9473.

George Papamakarios and Iain Murray. Fast ϵ-free inference of simulation models with Bayesian conditional density estimation. In *Advances in Neural Information Processing Systems 29*, pages 1028–1036. Curran Associates, Inc., 2016.

George Papamakarios, David Sterratt, and Iain Murray. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89 of *Proceedings of Machine Learning Research*, pages 837–848. PMLR, 2019.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

Jaimit Parikh, James Kozloski, and Viatcheslav Gurev. Integration of AI and mechanistic modeling in generative adversarial networks for stochastic inverse problems. *arXiv preprint arXiv:2009.08267*, 2020.

Kim Cuc Pham, David J Nott, and Sanjay Chaudhuri. A note on approximating ABC-MCMC using flexible classifiers. *STAT*, 3(1):218–227, 2014.

Stefan T Radev, Ulf K Mertens, Andreas Voss, Lynton Ardizzone, and Ullrich Köthe. Bayesflow: Learning complex stochastic models with invertible neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

CE. Rasmussen and CKI. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, 2006.

G. S. Rodrigues, D. J. Nott, and S. A. Sisson. Likelihood-free approximate gibbs sampling, 2019.

Axel Sauer, Kashyap Chitta, Jens Müller, and Andreas Geiger. Projected gans converge faster. *Bridging Theory and Practice Workshop at Neural Information Processing Systems*, 2021.

Scott A Sisson, Yanan Fan, and Mark M Tanaka. Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007.

Akash Srivastava, Lazar Valkov, Chris Russell, Michael U Gutmann, and Charles Sutton. Veegan: Reducing mode collapse in gans using implicit variational learning. *arXiv preprint arXiv:1705.07761*, 2017.

Sean Talts, Michael Betancourt, Daniel Simpson, and Aki Vehtari. Validating Bayesian Inference Algorithms with Simulation-Based Calibration. pages 1–26, 2020.

Simon Tavaré, David J. Balding, R. C. Griffiths, and Peter Donnelly. Inferring coalescence times from DNA sequence data. *Genetics*, 145(2), 1997.

Owen Thomas, Ritabrata Dutta, Jukka Corander, Samuel Kaski, and Michael U. Gutmann. Likelihood-Free Inference by Ratio Estimation. *Bayesian Analysis*, pages 1–31, 2021.

Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia. Deepfakes and beyond: A survey of face manipulation and fake detection, 2020.

Dustin Tran, Rajesh Ranganath, and David Blei. Hierarchical implicit models and likelihood-free variational inference. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Richard Eric Turner and Maneesh Sahani. *Two problems with variational expectation maximisation for time series models*, page 104–124. Cambridge University Press, 2011.

Kevin Yu, Harnaik Dhami, Kartik Madhira, and Pratap Tokekar. Gatsbi: An online gtsp-based algorithm for targeted surface bridge inspection, 2020.

Manzil Zaheer, Chun-Liang Li, Barnabás Póczos, and Ruslan Salakhutdinov. GAN connoisseur: Can GANs learn simple 1d parametric distributions? *Bridging Theory and Practice Workshop at Neural Information Processing Systems*, 2017.

Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Computer Vision – ECCV 2018*, pages 294–310. Springer International Publishing, 2018.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.