ONE-SHOT CONDITIONAL SAMPLING: MMD MEETS NEAREST NEIGHBORS

Anonymous authors
Paper under double-blind review

ABSTRACT

How can we generate samples from a conditional distribution that we never fully observe? This question arises across a broad range of applications in both modern machine learning and classical statistics, including image post-processing in computer vision, approximate posterior sampling in simulation-based inference, and conditional distribution modeling in complex data settings. In such settings, compared with unconditional sampling, additional feature information can be leveraged to enable more adaptive and efficient sampling. Building on this, we introduce Conditional Generator using MMD (CGMMD), a novel framework for conditional sampling. Unlike many contemporary approaches, our method frames the training objective as a simple, adversary-free direct minimization problem. A key feature of CGMMD is its ability to produce conditional samples in a single forward pass of the generator, enabling practical one-shot sampling with low test-time complexity. We establish rigorous theoretical bounds on the loss incurred when sampling from the CGMMD sampler, and prove convergence of the estimated distribution to the true conditional distribution. In the process, we also develop a uniform concentration result for nearest-neighbor based functionals, which may be of independent interest. Finally, we show that CGMMD performs competitively on synthetic tasks involving complex conditional densities, as well as on practical applications such as image denoising and image super-resolution.

1 INTRODUCTION

A fundamental problem in statistics and machine learning is to model the relationship between a response $\boldsymbol{Y} \in \mathcal{Y}$ and a predictor $\boldsymbol{X} \in \mathcal{X}$. Classical regression methods [Hastie et al., 2009; Koenker & Bassett Jr, 1978], typically summarize this relationship through summary statistics, which are often insufficient for many downstream tasks that require the knowledge of the entire conditional law. Access to the full conditional distribution enables quantification of uncertainty associated with prediction [Castillo & Randrianarisoa, 2022], uncovers latent structure [Mimno et al., 2015], supports dimension reduction [Reich et al., 2011], and graphical modeling [Chen et al., 2024]. In modern scientific applications, it provides a foundation for simulation-based inference [Cranmer et al., 2020] across various domains, including computer vision [Gupta et al., 2024], neuroscience [von Krause et al., 2022], and the physical sciences [Hou et al., 2024; Mastandrea et al., 2024].

Classical approaches such as distributional regression and conditional density estimation [Rosenblatt, 1969; Fan et al., 1996; Hothorn et al., 2014] model the full conditional distribution directly but often rely on strong assumptions and offer limited flexibility. In contrast, recent advances in generative models like Generative Adversarial Networks (GANs) [Zhou et al., 2023; Mirza & Osindero, 2014; Odena et al., 2017], Variational Autoencoders (VAEs) [Harvey et al., 2021; Doersch, 2016; Mishra et al., 2018], and diffusion models [Rombach et al., 2022; Saharia et al., 2022; Zhan et al., 2025] provide more flexible, assumption lean alternatives for conditional distribution learning across applications in vision, language, and scientific simulation. A more detailed discussion of related work, background, and connections to simulation-based inference is provided in Section A.

GANs, introduced by Goodfellow et al. [2014] as a two-player minimax game optimizing the Jensen–Shannon divergence [Fuglede & Topsoe, 2004], are a widely adopted class of generative models, known for their flexibility and empirical success. However, training remains delicate and unstable, even in the unconditional setting [Arjovsky & Bottou, 2017; Salimans et al., 2016]. As Ar-

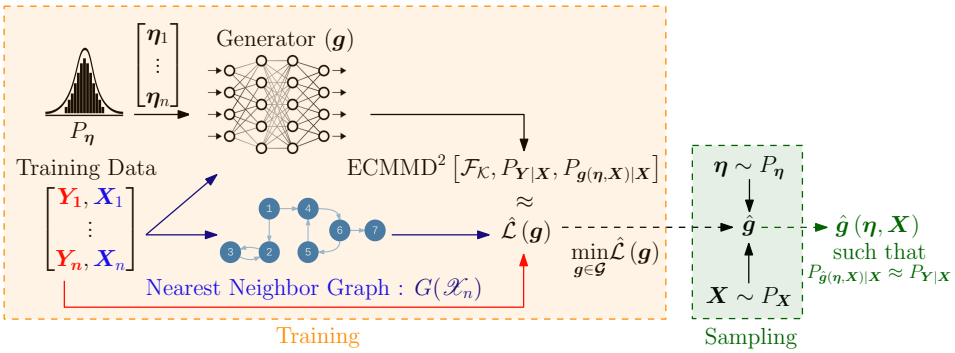

Figure 1: Schematic overview of CGMMD: Given training data $(\boldsymbol{Y}_1, \boldsymbol{X}_1), \ldots, (\boldsymbol{Y}_n, \boldsymbol{X}_n)$, the samples $\mathscr{X}_n = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ and auxiliary noise $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n$ are passed through the generator $\boldsymbol{g}$ to produce samples $\boldsymbol{g}(\boldsymbol{\eta}_1, \boldsymbol{X}_1), \ldots, \boldsymbol{g}(\boldsymbol{\eta}_n, \boldsymbol{X}_n)$. These outputs are compared with the observed $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ values using a nearest-neighbor ($G(\mathscr{X}_n)$) based estimate of the ECMMD discrepancy (see (1.2)) between true and generated conditional distributions. Edges are color-coded to highlight the dependence of each section on the corresponding inputs. After training, sampling is immediate: for any new input $\boldsymbol{X}$, independently generate new $\boldsymbol{\eta} \sim P_{\boldsymbol{\eta}}$, the trained model $\hat{\boldsymbol{g}}$ then produces $\hat{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{X})$ as the conditional output. Each component is described in greater details in Section 2 and Section 3.

jovsky & Bottou [2017] point out, the generator and target distributions often lie on low-dimensional manifolds that do not intersect, rendering divergences like Jensen–Shannon or KL constant or infinite and thus providing no useful gradient. To address this, alternative objectives based on Integral Probability Metrics (IPMs) [Müller, 1997], such as the Wasserstein distance [Villani et al., 2008] and Maximum Mean Discrepancy (MMD) [Gretton et al., 2012], have been proposed for more stable training in unconditional sampling using GANs.

Building on the success of MMD-GANs [Li et al., 2015; Dziugaite et al., 2015; Bińkowski et al., 2018; Huang et al., 2022b], we propose an MMD-based loss using nearest neighbors to quantify discrepancies between conditional distributions. While MMD has been used in conditional generation, to the best of our knowledge we are the first to provide sharp theoretical guarantees for MMD based conditional sampling, offering a principled foundation for training conditional generators. Initially developed for two-sample testing by Gretton et al. [2012], MMD has since seen broad adoption across the statistical literature [Gretton et al., 2007; Fukumizu et al., 2007; Chwialkowski et al., 2016; Sutherland et al., 2016]. It quantifies the discrepancy between two probability distributions as the maximum difference in expectations over functions f drawn from the unit ball of a Reproducing Kernel Hilbert Space (RKHS) defined on $\mathcal{Y}$ [Aronszajn, 1950]. Formally, let $\mathcal{Y}$ be a separable metric space equipped with $\mathcal{B}_\mathcal{Y}$, the sigma-algebra generated by the open sets of $\mathcal{Y}$. Let $\mathcal{P}(\mathcal{Y})$ be the collection of all probability measures on $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$. Then for any $P_{\boldsymbol{Y}}, P_{\boldsymbol{Z}} \in \mathcal{P}(\mathcal{Y})$,

$$\mathrm{MMD}(\mathcal{F}_\mathcal{K}, P_{\boldsymbol{Y}}, P_{\boldsymbol{Z}}) := \sup_{f\in\mathcal{F}_\mathcal{K}} \mathbb{E}[f(\boldsymbol{Y})] - \mathbb{E}[f(\boldsymbol{Z})], \tag{1.1}$$

where $\mathcal{F}_\mathcal{K}$ is the unit ball of a reproducing kernel Hilbert space (RKHS) $\mathcal{K}$ on $\mathcal{Y}$.

1.1 Conditional Generator using Maximum Mean Discrepancy (CGMMD)

To extend MMD to the conditional setting, we employ the expected conditional MMD (ECMMD) from Chatterjee et al. [2024] (also see Huang et al. [2022b]), which naturally generalizes the MMD distance to a discrepancy between conditional distributions. Formally, for $\boldsymbol{X} \sim P_{\boldsymbol{X}}$, conditional distributions $P_{\boldsymbol{Y}|\boldsymbol{X}}$ and $P_{\boldsymbol{Z}|\boldsymbol{X}}$ supported on $\mathcal{Y}$, the squared ECMMD can be defined as,

$$\mathrm{ECMMD}^2(\mathcal{F}_\mathcal{K}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{Z}|\boldsymbol{X}}) := \mathbb{E}_{\boldsymbol{X}\sim P_{\boldsymbol{X}}}\left[\mathrm{MMD}^2(\mathcal{F}_\mathcal{K}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{Z}|\boldsymbol{X}})\right]. \tag{1.2}$$

We discuss simplified formulations of this measure later in Section 2.1. By Chatterjee et al. [2024, Proposition 2.3], ECMMD is indeed a strict scoring rule, meaning that $\mathrm{ECMMD}^2(\mathcal{F}_\mathcal{K}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{Z}|\boldsymbol{X}}) = 0$ if and only if $P_{\boldsymbol{Y}|\boldsymbol{X}} = P_{\boldsymbol{Z}|\boldsymbol{X}}$ almost surely. This property establishes ECMMD as a principled and reliable tool for comparing conditional distributions.

Instead of estimating the target conditional distribution $P_{\boldsymbol{Y}|\boldsymbol{X}}$ directly, we follow the generative approach from Zhou et al. [2023] and Song et al. [2025]. By the noise outsourcing lemma (see Lemma

2.1), the problem of nonparametric conditional density estimation can be reformulated as a generalized nonparametric regression problem. In particular, for a given predictor value $\boldsymbol{X} = \boldsymbol{x}$, our goal is to learn a conditional generator $\boldsymbol{g}(\boldsymbol{\eta}, \boldsymbol{x})$, where $\boldsymbol{\eta}$ is drawn from a simple reference distribution (e.g., Gaussian or uniform). The generator is trained so that $\boldsymbol{g}(\boldsymbol{\eta}, \boldsymbol{x})$ approximates the conditional distribution of $\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x}$ for all $\boldsymbol{x}$. Discrepancy between the true conditional distribution $P_{\boldsymbol{Y}|\boldsymbol{X}}$ and the model distribution $P_{\boldsymbol{g}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}$ is measured using the squared ECMMD. Once training is complete, conditional sampling becomes a one-shot procedure: draw $\boldsymbol{\eta}$ from the reference distribution and sample $\boldsymbol{g}(\boldsymbol{\eta}, \boldsymbol{x})$. In this way, the generator provides an explicit and efficient representation of the conditional distribution of $\boldsymbol{Y} \mid \boldsymbol{X}$. We refer to $\boldsymbol{g}(\boldsymbol{\eta}, \boldsymbol{x})$ as the Conditional Generator using Maximum Mean Discrepancy, or CGMMD for short. We provide the schematic overview of the method in Figure 1. Now, we turn to the main contributions of our proposed method.

1.2 Main Contributions

Our main contributions are summarized below.

- **Direct Minimization.** Similar to MMD-GANs in the unconditional setting, CGMMD avoids adversarial min-max optimization and instead enables direct minimization of a well-defined loss, offering a more straightforward and tractable alternative to GAN-based training [Zhou et al., 2023; Song et al., 2025; Ramesh et al., 2022]. This design helps avoid common issues in conditional GANs, such as mode collapse and unstable min–max dynamics.
- **One-shot Sampling.** While diffusion models have demonstrated remarkable success in generating high-quality and diverse samples, their iterative denoising procedure [Ho et al., 2020] makes sampling computationally expensive and time-consuming. In contrast, CGMMD enables efficient one-shot sampling, i.e., conditional samples are obtained in a single forward pass of the generator. Specifically, to sample from $\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x}$, one simply draws $\boldsymbol{\eta}$ from a simple reference distribution (e.g., Gaussian or uniform) and evaluates $\hat{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{x})$, where $\hat{\boldsymbol{g}}$ is a solution of (3.2).
- **Theoretical Guarantees.** We provide rigorous theoretical guarantees for CGMMD. Theorem 4.1 gives a non-asymptotic finite-sample bound on the error of the conditional sampler $\hat{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{x})$, and Corollary 4.1 establishes convergence to the true conditional distribution as the sample size increases. Together, these results provide strong theoretical justification for CGMMD.

  To the best of our knowledge, this is the first application of tools from uniform concentration of nonlinear functionals, nearest neighbor methods, and generalization theory to conditional generative modeling. In the process, we also establish a general uniform concentration result for a broad class of nearest-neighbor-based functionals (Appendix G), which may be of independent interest.
- **Numerical Experiments.** Finally, we provide experiments on both synthetic and real data (mainly in image post-processing tasks) to evaluate the performance of CGMMD and compare it with existing approaches in the literature. Overall, our proposed approach performs reliably across different settings and often matches or exceeds the alternative approaches in more challenging cases.

2 Technical Background

In this section, we introduce the necessary concepts and previous works required to understand our proposed framework, CGMMD. To that end, we begin with the necessary formalism.

Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces, that is, complete separable metric spaces equipped with the corresponding Borel-sigma algebras $\mathcal{B}_{\mathcal{X}}$ and $\mathcal{B}_{\mathcal{Y}}$ respectively. Let $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ be the collection of all probability measures defined on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ and $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$ respectively. Recalling the RKHS $\mathcal{K}$ defined on $\mathcal{Y}$ from (1.1), the Riesz representation theorem [Reed & Simon, 1980, Therorem II.4] guarantees the existence of a positive definite kernel $\mathsf{K} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ such that for every $\boldsymbol{y} \in \mathcal{Y}$, the feature map $\phi_{\boldsymbol{y}} \in \mathcal{K}$ satisfies $\mathsf{K}(\boldsymbol{y}, \cdot) = \phi_{\boldsymbol{y}}(\cdot)$ and $\mathsf{K}(\boldsymbol{y}_1, \boldsymbol{y}_2) = \langle \phi_{\boldsymbol{y}_1}, \phi_{\boldsymbol{y}_2} \rangle_{\mathcal{K}}$.

The definition of feature maps can now be extended to embed any distribution $P \in \mathcal{P}(\mathcal{Y})$ into $\mathcal{K}$. In particular, for $P \in \mathcal{P}(\mathcal{Y})$ we can define the kernel mean embedding μ_P as $\langle f, \mu_P \rangle_{\mathcal{K}} = \mathbb{E}_{Y \sim P}[f(Y)]$. Moreover, by the canonical form of the feature maps, it follows that $\mu_P(t) := \mathbb{E}_{Y \sim P}[K(Y, t)]$ for all $t \in \mathcal{Y}$. Henceforth, we make the following assumptions on the kernel K.

Assumption 2.1. The kernel $\mathsf{K} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is positive definite and satisfies the following:

1. The kernel K is bounded, that is $\|\mathsf{K}\|_\infty < K$ for some $K > 0$ and Lipschitz continuous.

2. The kernel mean embedding $\mu : \mathcal{P}(\mathcal{Y}) \to \mathcal{K}$ is a one-to-one (injective) function. This is also known as the *characteristic kernel* property [Sriperumbudur et al., 2011].

Assumption 2.1 ensures that the mean embedding $\mu_P \in \mathcal{K}$ (see Lemma 3 in Gretton et al. [2012] and Lemma 2.1 in Park & Muandet [2020]), and that MMD defines a metric on $\mathcal{P}(\mathcal{Y})$. While these properties can be guaranteed under weaker conditions on the kernel K, we adopt the above assumption for technical convenience. With the above notations the MMD (recall (1.1)) can be equivalently expressed as $\mathrm{MMD}^2(\mathcal{F}_\mathcal{K}, P_{\boldsymbol{Y}}, P_{\boldsymbol{Z}}) = \|\mu_{P_{\boldsymbol{Y}}} - \mu_{P_{\boldsymbol{Z}}}\|^2_\mathcal{K}$ (see Lemma 4 from Gretton et al. [2012]) where $\|\cdot\|_\mathcal{K}$ is the norm induced by the inner product $\langle\cdot,\cdot\rangle_\mathcal{K}$. In the following, we express the ECMMD in an equivalent form and leverage it to obtain a consistent empirical estimator.

2.1 ECMMD: Representation via Kernel Embeddings

Recalling the definition of ECMMD from (1.2), we note that it admits an equivalent formulation. In particular, for distributions $P_{\boldsymbol{Y}|\boldsymbol{X}}$ and $P_{\boldsymbol{Z}|\boldsymbol{X}}$ (which exists by Klenke [2008, Theorem 8.37]), define the conditional mean embeddings $\mu_{P_{\boldsymbol{Y}|\boldsymbol{X}}}(t) := \mathbb{E}[\mathsf{K}(\boldsymbol{Y}, t) \mid \boldsymbol{X}]$ and $\mu_{P_{\boldsymbol{Z}|\boldsymbol{X}}}(t) := \mathbb{E}[\mathsf{K}(\boldsymbol{Z}, t) \mid \boldsymbol{X}]$ for all $t \in \mathcal{Y}$. Under Assumption 2.1, the conditional mean embeddings are indeed well defined by Park & Muandet [2020, Lemma 3.2]. Consequently, $\|\mu_{P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}} - \mu_{P_{\boldsymbol{Z}|\boldsymbol{X}=\boldsymbol{x}}}\|^2_\mathcal{K}$ is the squared MMD metric between the conditional distributions for a particular value of $\boldsymbol{X} = \boldsymbol{x}$. Averaging this quantity over the marginal distribution of $\boldsymbol{X}$ yields the squared ECMMD distance:

$$\mathrm{ECMMD}^2(\mathcal{F}_\mathcal{K}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{Z}|\boldsymbol{X}}) = \mathbb{E}_{\boldsymbol{X}\sim P_{\boldsymbol{X}}}\left[\|\mu_{P_{\boldsymbol{Y}|\boldsymbol{X}}} - \mu_{P_{\boldsymbol{Z}|\boldsymbol{X}}}\|^2_\mathcal{K}\right] \tag{2.1}$$

However, to use ECMMD as a loss function for estimating the conditional sampler, we require a consistent estimator of the expression in (2.1). To that end, the well-known *kernel trick* enables a more tractable reformulation of ECMMD, making it amenable to estimation from observed data. By [Chatterjee et al., 2024, Proposition 2.4], the squared ECMMD admits the tractable form

$$\mathrm{ECMMD}^2(\mathcal{F}_\mathcal{K}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{Z}|\boldsymbol{X}}) = \mathbb{E}\big[\mathsf{K}(\boldsymbol{Y}, \boldsymbol{Y}') + \mathsf{K}(\boldsymbol{Z}, \boldsymbol{Z}') - \mathsf{K}(\boldsymbol{Y}, \boldsymbol{Z}') - \mathsf{K}(\boldsymbol{Z}, \boldsymbol{Y}')\big], \tag{2.2}$$

where $(\boldsymbol{Y}, \boldsymbol{Y}', \boldsymbol{Z}, \boldsymbol{Z}', \boldsymbol{X})$ is generated by first sampling $\boldsymbol{X} \sim P_{\boldsymbol{X}}$, then drawing $(\boldsymbol{Y}, \boldsymbol{Z})$ and $(\boldsymbol{Y}', \boldsymbol{Z}')$ independently from $P_{\boldsymbol{Y}|\boldsymbol{X}} \times P_{\boldsymbol{Z}|\boldsymbol{X}}$.

2.2 ECMMD: Consistent Estimation using Nearest Neighbors

Towards estimating the ECMMD, we leverage the equivalent expression from (2.2). By the tower property of conditional expectations, (2.2) can be further expanded as,

$$\mathrm{ECMMD}^2(\mathcal{F}_\mathcal{K}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{Z}|\boldsymbol{X}}) = \mathbb{E}\left[\mathbb{E}\big[\mathsf{K}(\boldsymbol{Y}, \boldsymbol{Y}') + \mathsf{K}(\boldsymbol{Z}, \boldsymbol{Z}') - \mathsf{K}(\boldsymbol{Y}, \boldsymbol{Z}') - \mathsf{K}(\boldsymbol{Z}, \boldsymbol{Y}') \mid \boldsymbol{X}\big]\right].$$

To estimate ECMMD, we observe that it involves averaging a conditional expectation over the distribution $P_{\boldsymbol{X}}$. Given observed samples $\{(\boldsymbol{Y}_i, \boldsymbol{Z}_i, \boldsymbol{X}_i) : 1 \le i \le n\}$ drawn from the joint distribution $P_{\boldsymbol{Y}\boldsymbol{Z}\boldsymbol{X}} = P_{\boldsymbol{Y}|\boldsymbol{X}} \times P_{\boldsymbol{Z}|\boldsymbol{X}} \times P_{\boldsymbol{X}}$, we proceed by first estimating the inner conditional expectation given $\boldsymbol{X} = \boldsymbol{X}_i$, and then averaging these estimates over the observed values $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. To estimate the inner conditional expectation given $\boldsymbol{X} = \boldsymbol{X}_i$, one can, in principle, average the inner function over sample indices whose corresponding predictors are 'close' to $\boldsymbol{X}_i$. A natural way to quantify such proximity is through nearest-neighbor graphs. Formally we construct the estimated ECMMD as follows.

Fix $k = k_n \ge 1$ and let $G(\mathscr{X}_n)$ be the directed $k-$nearest neighbor graph on $\mathscr{X}_n = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$. Moreover let $N_{G(\mathscr{X}_n)}(i) := \{j \in [n] : \boldsymbol{X}_i \to \boldsymbol{X}_j \text{ is an edge in } G(\mathscr{X}_n)\}$ for all $i \in [n]$. Now the $k-$NN based estimator of ECMMD can be defined as,

$$\widehat{\mathrm{ECMMD}}^2\left(\mathcal{F}_\mathcal{K}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{Z}|\boldsymbol{X}}\right) := \tfrac{1}{n}\textstyle\sum_{i=1}^n \tfrac{1}{k_n} \sum_{j \in N_{G(\mathscr{X}_n)}(i)} \mathsf{H}\left(\boldsymbol{W}_i, \boldsymbol{W}_j\right) \tag{2.3}$$

where $\boldsymbol{W}_i = (\boldsymbol{Y}_i, \boldsymbol{Z}_i)$ for all $i \in [n]$ and $\mathsf{H}\left(\boldsymbol{W}_i, \boldsymbol{W}_j\right) = \mathsf{K}\left(\boldsymbol{Y}_i, \boldsymbol{Y}_j\right) - \mathsf{K}\left(\boldsymbol{Y}_i, \boldsymbol{Z}_j\right) - \mathsf{K}\left(\boldsymbol{Z}_i, \boldsymbol{Y}_j\right) + \mathsf{K}\left(\boldsymbol{Z}_i, \boldsymbol{Z}_j\right)$ for all $1 \le i, j \le n$. Chatterjee et al. [2024, Theorem 3.2] shows that under mild conditions, this estimator is consistent for the oracle ECMMD. We exploit this nearest-neighbor construction to define the CGMMD objective in Section 3.

2.3 Generative Representation of Conditional Distribution

As outlined in Section 1.1, conditional density estimation can be reformulated as a generalized nonparametric regression problem. Suppose $(\boldsymbol{Y}, \boldsymbol{X}) \in \mathcal{X} \times \mathcal{Y}$ follows some joint distribution

Algorithm 1: CGMMD Training

Input: Training dataset $\{(\boldsymbol{Y}_i, \boldsymbol{X}_i)\}_{i=1}^n$. Conditional generator $\boldsymbol{g} = \boldsymbol{g}_\theta$ with initial parameters θ. Auxillary Kernel function H (see (2.3)). Noise distribution $P_{\boldsymbol{\eta}}$. Learning rate α, epochs E, batch size B and number of nearest neighbors k_B.

Output: Trained generator parameters $\hat{\theta}$.

Sample $\{\boldsymbol{\eta}_i : 1 \le i \le n\} \sim P_{\boldsymbol{\eta}}$.

for *epoch = 1* **to** E **do**

  for *each* $I \subseteq [n]$ *of size* B **do**

    $\mathscr{X}_I \leftarrow \{\boldsymbol{X}_i\}_{i \in I}$;

    $G(\mathscr{X}_I) \leftarrow k_B$-Nearest Neighbor graph on $\mathscr{X}_I$;

    $N_{G(\mathscr{X}_I)}(i) \leftarrow$ neighbors of $\boldsymbol{X}_i$ in $G(\mathscr{X}_I), \boldsymbol{g}_i \leftarrow \boldsymbol{g}_\theta(\boldsymbol{\eta}_i, \boldsymbol{X}_i), \boldsymbol{W}_{i,\boldsymbol{g}} \leftarrow (\boldsymbol{Y}_i, \boldsymbol{g}_i)\, \forall i \in I$;

    $\hat{\mathcal{L}}_{\text{batch}} \leftarrow \frac{1}{Bk_B} \sum_{i \in I} \sum_{j \in N_{G(\mathscr{X}_I)}(i)} \mathsf{H}(\boldsymbol{W}_{i,\boldsymbol{g}}, \boldsymbol{W}_{j,\boldsymbol{g}})$;

    $\theta \leftarrow \theta - \alpha \nabla_\theta \hat{\mathcal{L}}_{batch}$.

return trained parameters $\hat{\theta} \leftarrow \theta$.

$P_{\boldsymbol{YX}}$, and we observe n independent samples $\{(\boldsymbol{Y}_1, \boldsymbol{X}_1), \ldots, (\boldsymbol{Y}_n, \boldsymbol{X}_n)\}$ from $P_{\boldsymbol{YX}}$. Our goal is to generate samples from the unknown conditional distribution $P_{\boldsymbol{Y}|\boldsymbol{X}}$. The *noise outsourcing lemma* (see Kallenberg, Theorem 5.10 and Zhou et al. [2023, Lemma 2.1]) formally connects conditional distribution estimation with conditional sample generation. For completeness, we state it below.

Lemma 2.1 (Noise Outsourcing Lemma)**.** Suppose $(\boldsymbol{Y}, \boldsymbol{X}) \sim P_{\boldsymbol{YX}}$. Then, for any $m \ge 1$, there exist a random vector $\boldsymbol{\eta} \sim P_{\boldsymbol{\eta}} = \mathrm{N}(\boldsymbol{0}_m, \boldsymbol{I}_m)$ and a Borel-measurable function $\bar{\boldsymbol{g}} : \mathbb{R}^m \times \mathcal{X} \to \mathcal{Y}$ such that $\boldsymbol{\eta}$ is generated independent of $\boldsymbol{X}$ and $(\boldsymbol{Y}, \boldsymbol{X}) = (\bar{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{X}), \boldsymbol{X})$ almost surely.

Moreover, by Zhou et al. [2023, Lemma 2.2], $(\boldsymbol{Y}, \boldsymbol{X}) \overset{d}{=} (\bar{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{X}), \boldsymbol{X})$ if and only if $\bar{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{x}) \sim P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$ for every $\boldsymbol{x} \in \mathcal{X}$. This identifies $\bar{\boldsymbol{g}}$ as a conditional generator. Consequently, to draw from $P_{\boldsymbol{Y}|\boldsymbol{X}}$, we sample $\boldsymbol{\eta} \sim \mathrm{N}(\boldsymbol{0}_m, \boldsymbol{I}_m)$ and output $\bar{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{X})$.

This perspective places conditional density estimation firmly within the realm of generative modeling. The task reduces to: given n independent samples from $P_{\boldsymbol{YX}}$, learn the conditional generator $\bar{\boldsymbol{g}}$. Zhou et al. [2023]; Ramesh et al. [2022]; Song et al. [2025]; Liu et al. [2021] leveraged this idea to develop a GAN-based (respectively Wasserstein-GAN) framework for conditional sampling. In contrast, our approach follows a similar path but replaces the potentially unstable min–max optimization of GANs with a principled minimization objective based on ECMMD discrepancy. The precise formulation is given in the following section.

3 ECMMD Based Objective for CGMMD

Building on the generative representation of conditional distributions and the ECMMD discrepancy introduced earlier, our goal is to learn a conditional generator $\bar{\boldsymbol{g}}$ by minimizing the ECMMD distance between the true conditional distribution $\boldsymbol{Y} \mid \boldsymbol{X}$ and the generated conditional distribution $\bar{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{X}) \mid \boldsymbol{X}$. We restrict our attention to a parameterized function class $\mathcal{G}$, as solving this unconstrained minimization problem over all measurable functions is intractable. To that end, we begin by defining the population objective

$$\mathcal{L}(\boldsymbol{g}) := \mathrm{ECMMD}^2\left[\mathcal{F}_{\mathcal{K}}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{g}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}\right] = \mathbb{E}_{\boldsymbol{X}\sim P_{\boldsymbol{X}}}\left[\|\mu_{P_{\boldsymbol{Y}|\boldsymbol{X}}} - \mu_{P_{\boldsymbol{g}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}}\|_{\mathcal{K}}^2\right].$$

The target generator is then given by $\boldsymbol{g}^\star \in \arg\min_{\boldsymbol{g}\in\mathcal{G}} \mathcal{L}(\boldsymbol{g})$. Since the oracle objective $\mathcal{L}(\cdot)$ is not directly available, we employ the estimation strategy outlined in Section 2.2 to construct a consistent empirical approximation of $\mathcal{L}(\boldsymbol{g})$. Given n independent samples $(\boldsymbol{Y}_1, \boldsymbol{X}_1), \ldots, (\boldsymbol{Y}_n, \boldsymbol{X}_n) \sim P_{\boldsymbol{YX}}$ and independent draws of noise variables $\boldsymbol{\eta}_1 \ldots, \boldsymbol{\eta}_n \sim P_{\boldsymbol{\eta}}$, we define the empirical objective,

$$\hat{\mathcal{L}}(\boldsymbol{g}) := \widehat{\mathrm{ECMMD}}^2\left(\mathcal{F}_{\mathcal{K}}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{g}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}\right) = \frac{1}{nk_n} \sum_{i=1}^n \sum_{j \in N_{G(\mathscr{X}_n)}(i)} \mathsf{H}(\boldsymbol{W}_{i,\boldsymbol{g}}, \boldsymbol{W}_{j,\boldsymbol{g}}) \quad (3.1)$$

where H is defined from (2.3) and $\boldsymbol{W}_{i,\boldsymbol{g}} := (\boldsymbol{Y}_i, \boldsymbol{g}(\boldsymbol{\eta}_i, \boldsymbol{X}_i))$ for all $1 \le i \le n$. Our estimate of the conditional generator is then defined as

$$\hat{\boldsymbol{g}} \in \arg\min_{\boldsymbol{g}\in\mathcal{G}} \hat{\mathcal{L}}(\boldsymbol{g}). \quad (3.2)$$

With the framework now in place, we emphasize that CGMMD offers substantial flexibility to practitioners. In our experiments, we restrict $\boldsymbol{\mathcal{G}}$ to deep neural networks, i.e., $\mathcal{G} =$

$\{\boldsymbol{g}_\theta : \mathbb{R}^m \times \mathcal{X} \to \mathcal{Y} \mid \theta \in \mathbb{R}^{\mathcal{S}}\}$ where $\mathcal{S}$ is the total number of parameters of the neural network $\boldsymbol{g}_\theta$. Here, (3.2) reduces to solving $\hat{\theta} \in \arg\min_{\theta \in \mathbb{R}^{\mathcal{S}}} \hat{\mathcal{L}}(\boldsymbol{g}_\theta)$. A corresponding pseudo-code is provided in Algorithm 1. In practice, the user may tailor the method by selecting the kernel K, the function class $\mathcal{G}$, number of neighbors k_n, and the manner in which the auxiliary noise variable $\boldsymbol{\eta}$ is incorporated into $\boldsymbol{g}(\cdot, \boldsymbol{x})$. We discuss some of these potential choices as well as refinements to the CGMMD objective when $P_{\boldsymbol{X}}$ has discrete support in Appendix D.

4 Analysis and Convergence Guarantees

In this section, we analyze the error of estimating the true conditional sampler $\bar{\boldsymbol{g}}$ (see Lemma 2.1). This section is further divided into two parts. In Section 4.1 we begin by deriving a finite-sample bound on the error arising from replacing the true conditional sampler $\bar{\boldsymbol{g}}$ with its empirical estimate $\hat{\boldsymbol{g}}$. As a further contribution in Section 4.2, we establish the convergence of the conditional distribution induced by the empirical sampler to the true conditional distribution. For clarity and ease of exposition, we present simplified versions of the assumptions and main results here, while deferring the complete statements and proofs to Appendix E.

4.1 Non-Asymptotic Error Bounds

For the estimated empirical sampler $\hat{\boldsymbol{g}}$ defined in (3.2) the estimation error can be defined as (recall Definition 1.2),

$$\mathcal{L}(\hat{\boldsymbol{g}}) = \mathrm{ECMMD}^2\left[\mathcal{F}, P_{\bar{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}, P_{\hat{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}\right] = \mathbb{E}\left[\left\|\mu_{P_{\bar{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}} - \mu_{P_{\hat{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}}\right\|_{\mathcal{K}}^2 \mid \hat{\boldsymbol{g}}\right], \quad (4.1)$$

where the expectations are taken over the randomness of $\boldsymbol{\eta}$ and $\boldsymbol{X}$ keeping the empirical sampler $\hat{\boldsymbol{g}}$ fixed. In other words, the estimation error evaluates the squared ECMMD between the conditional distributions of $\bar{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{X})$ and $\hat{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{X})$ given $\boldsymbol{X}$. In the following, we will provide non-asymptotic bounds on the estimation error $\mathcal{L}(\hat{\boldsymbol{g}})$. To that end, for the rest of the article, we assume $\mathcal{Y} \subseteq \mathbb{R}^p$ for some $p \geq 1$ and we begin by rigorously defining the class of functions $\boldsymbol{\mathcal{G}}$.

Details of $\boldsymbol{\mathcal{G}}$: Let $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{G}}_{\mathcal{H},\mathcal{W},\mathcal{S},\mathcal{B}}$ be the set of ReLU neural networks $\boldsymbol{g} : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^p$ with depth $\mathcal{H}$, width $\mathcal{W}$, size $\mathcal{S}$ and $\|\boldsymbol{g}\|_\infty \leq \mathcal{B}$. In particular, $\mathcal{H}$ denotes the number of hidden layers and $(w_0, w_2, \ldots, w_{\mathcal{H}})$ denotes the width of each layer, where $w_0 = d + m$ and $w_{\mathcal{H}} = p$ denotes the input and output dimension, respectively. We take $\mathcal{W} = \max\{w_0, w_1, \ldots, w_{\mathcal{H}}\}$. Finally, size $\mathcal{S} = \sum_{i=1}^{\mathcal{H}} w_i\,(w_{i-1} + 1)$ refers to the total number of parameters of the network. To establish the error bounds, we make the following assumption about the parameters of $\boldsymbol{\mathcal{G}}$.

Assumption 4.1. The network parameters of $\boldsymbol{\mathcal{G}}$ satisfies $\mathcal{B} \geq 1$ and $\mathcal{H}, \mathcal{W} \to \infty$ such that,

$$\frac{\mathcal{H}\mathcal{W}}{(\log n)^{\frac{d+m}{2}}} \xrightarrow{n\to\infty} \infty \quad \text{and} \quad \frac{\mathcal{B}^2\mathcal{H}\mathcal{S}\log\mathcal{S}\log n}{n} \xrightarrow{n\to\infty} 0.$$

The imposed conditions require that the neural network's size grows with the sample size, specifically that the product of its depth and width increases with n. These assumptions are flexible enough to accommodate a wide range of architectures, but a key constraint is that the network size must remain smaller than the sample size. This arises from the use of empirical process theory [Van Der Vaart & Wellner, 1996; Bartlett et al., 2019] to control the stochastic error in the estimated generator. Similar conditions appear in recent work on conditional sampling [Zhou et al., 2023; Liu et al., 2021; Song et al., 2025] and in convergence analyses for deep nonparametric regression [Schmidt-Hieber, 2020; Kohler & Langer, 2019; Nakada & Imaizumi, 2020]. We also make the following technical assumptions.

Assumption 4.2. The following conditions on $P_{\boldsymbol{Y}\boldsymbol{X}}$, the kernel K, the true conditional sampler $\bar{\boldsymbol{g}}$ and the class $\boldsymbol{\mathcal{G}}$ holds.

1. $P_{\boldsymbol{X}}$ is supported on $\mathcal{X} \subseteq \mathbb{R}^d$ for some $d > 0$ and $\|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_2$ has a continuous distribution for $\boldsymbol{X}_1, \boldsymbol{X}_2 \sim P_{\boldsymbol{X}}$.
2. Moreover $\boldsymbol{X} \sim P_{\boldsymbol{X}}$ is sub-gaussian, that is [1], $\mathbb{P}\left(\|\boldsymbol{X}\|_2 > t\right) \lesssim \exp\left(-t^2\right)$ for all $t > 0$.

[1] We use the notation $a \lesssim_\theta b$ to imply $a \leq C_\theta b$ for some constant $C_\theta > 0$ depending on the parameter θ. In particular $a \lesssim b$ implies $a \leq Cb$ for some universal constant $C > 0$. Henceforth take $\boldsymbol{\theta} = (d, m, p, \mathsf{K})$.

3. The target conditional sampler $\bar{\boldsymbol{g}} : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^p$ is uniformly continuous with $\|\bar{\boldsymbol{g}}\|_\infty \leq 1$.
4. For any $\boldsymbol{g} \in \boldsymbol{\mathcal{G}}$ consider $h_{\boldsymbol{g}}(\boldsymbol{x}) = \mathbb{E}\left[\mathsf{K}(\boldsymbol{Y}, \cdot) - \mathsf{K}\left(\boldsymbol{g}\left(\boldsymbol{\eta}, \boldsymbol{X}\right), \cdot\right) | \boldsymbol{X} = x\right]$ and assume that $|\langle h_{\boldsymbol{g}}(\boldsymbol{x}), h_{\boldsymbol{g}}(\boldsymbol{x}_1) - h_{\boldsymbol{g}}(\boldsymbol{x}_2)\rangle| \lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2$, for all $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$ where the constant is independent of $\boldsymbol{g}$.

The first two assumptions are standard in the nearest neighbor literature and have been studied in the context of conditional independence testing using nearest neighbor-based methods [Huang et al., 2022a; Deb et al., 2020; Azadkia & Chatterjee, 2021; Borgonovo et al., 2025; Dasgupta & Kpotufe, 2014]. The first, concerning uniqueness in nearest neighbor selection, can be relaxed via tie-breaking schemes (see Section 7.3 in [Deb et al., 2020]), though we do not pursue this direction. The second, on the tail behavior of the predictor $\boldsymbol{X}$, can be weakened to include heavier-tailed distributions, such as those satisfying sub-Weibull conditions [Vladimirova et al., 2020] (also see (E.1)). The third assumption is mainly for technical convenience; similar conditions appear in prior work on neural network-based conditional sampling [Zhou et al., 2023; Song et al., 2025; Liu et al., 2021]. Its uniform continuity condition can also be relaxed to continuity (see Appendix E).

Remark 4.1. Assumption 4.2.4 is arguably the most critical in our analysis. It quantifies the sensitivity of the conditional mean embeddings to changes in the predictor $\boldsymbol{X}$, and is essential for establishing concentration of the nearest-neighbor-based ECMMD estimator (see (2.3)) around its population counterpart. Similar assumptions have been used in prior work on nearest neighbor methods [Huang et al., 2022a; Deb et al., 2020; Azadkia & Chatterjee, 2021; Dasgupta & Kpotufe, 2014]. As noted in Azadkia & Chatterjee [2021, Section 4], omitting such regularity conditions can lead to arbitrarily slow convergence rates. While the locally lipschitz-type condition can be relaxed, for example to Hölder continuity upto polynomial factors (see (E.2)) it remains a key assumption for our theoretical guarantees. We further elaborate on this assumption in Appendix F.

Under the above assumptions, we are now ready to present our main theorem on the error incurred by using the empirical sampler $\hat{\boldsymbol{g}}$.

Theorem 4.1 (Simpler version of Theorem E.1)**.** Adopt Assumption 2.1, Assumption 4.1 and Assumption 4.2. Moreover take $\omega_{\bar{\boldsymbol{g}}}(r) := \sup\left\{\|\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{y})\|_2 : \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^p, \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq r\right\}$ to be the optimal modulus of continuity of the true conditional sampler $\bar{\boldsymbol{g}}$. Let $k_n = o\left(n^\gamma\right)$ for some $0 < \gamma < 1$. Then for any $\delta > 0$, with probability at least $1 - \delta$,

$$\mathcal{L}\left(\hat{\boldsymbol{g}}\right) \lesssim_{\boldsymbol{\theta}} \frac{\operatorname{poly}\log(n)}{n^{\frac{1-\gamma}{d}}} + \sqrt{\frac{\mathcal{B}^2 \mathcal{H} \mathcal{S} \log \mathcal{S} \log n}{n}} + \omega_{\bar{\boldsymbol{g}}}\left(\frac{2\sqrt{\log n}}{(\mathcal{H}\mathcal{W})^{\frac{1}{d+m}}}\right) + \sqrt{\frac{\log\left(1/\delta\right)}{n}}.$$

The first two terms capture the stochastic error from the uniform concentration of the empirical loss around the population ECMMD objective. The third term reflects approximation error from estimating the true conditional sampler $\bar{\boldsymbol{g}}$ using neural networks in $\boldsymbol{\mathcal{G}}$. While we defer the proof of this result and its generalization to Appendix B.1 and Appendix E, respectively, we highlight the main novelty of our analysis here. Specifically, it integrates tools from recent advances in uniform concentration for non-linear functionals [Maurer & Pontil, 2019; Ni & Huo, 2024], nearest neighbor methods [Azadkia & Chatterjee, 2021; Deb et al., 2020], and generalization theory, including neural network approximation of smooth functions [Shen et al., 2020; Zhang et al., 2022]. To our knowledge, this is the first application of these techniques to conditional generative modeling with nonparametric nearest neighbor objectives. Additionally, we establish a uniform concentration result for a broad class of nearest-neighbor-based functionals (Appendix G), which may be of independent interest.

4.2 Convergence of the Empirical Sampler

As outlined earlier, in this section, we leverage the bound established in Theorem 4.1 to demonstrate the convergence of the conditional distribution identified by the estimated sampler $\hat{\boldsymbol{g}}(\boldsymbol{\eta}, \boldsymbol{X})$ to the true conditional distribution.

While Theorem 4.1 provides a finite-sample quantitative guarantee on the loss incurred by using the estimated sampler in place of the true sampler $\boldsymbol{g}$, we now show that the conditional distribution induced by $\hat{\boldsymbol{g}}$ converges to the true conditional distribution. Furthermore, we strengthen this result by establishing convergence in terms of characteristic functions as well. By a classical result by Bochner (see Theorem H.1) every continuous positive definite function ψ is associated with a finite

non-negative Borel measure Λ_ψ. With this notation, we have the following convergence result with proof given in Appendix B.2.

Corollary 4.1. Suppose the assumptions from Theorem 4.1 hold. Then,

$$\mathbb{E}\left[\mathrm{MMD}^2\left[\mathcal{F}, P_{\hat{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}, P_{\bar{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}\right]\right] \longrightarrow 0. \tag{4.2}$$

Moreover, if the kernel $\mathsf{K}(\boldsymbol{x},\boldsymbol{y}) = \psi(\boldsymbol{x}-\boldsymbol{y})$ for some bounded, lipschitz continuous positive definite function ψ. Then,

$$\mathbb{E}\left[\int \left(\phi_{\hat{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}(\boldsymbol{t}) - \phi_{\bar{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}(\boldsymbol{t})\right)^2 \mathrm{d}\Lambda_\psi(\boldsymbol{t})\right] \longrightarrow 0 \tag{4.3}$$

where $\phi_{\hat{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}$ and $\phi_{\bar{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}$ are the characteristic functions of the conditional distributions $P_{\hat{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}$ and $P_{\bar{\boldsymbol{g}}(\boldsymbol{\eta},\boldsymbol{X})|\boldsymbol{X}}$ respectively.

The above results demonstrate the efficacy of CGMMD. In particular, they show that the conditional distribution learned by the conditional sampler in CGMMD closely approximates the true conditional distribution.

5 Numerical Experiments

We begin our empirical study with toy examples of bivariate conditional sample generation, then move to practical applications such as image denoising and super-resolution on MNIST [Yann, 2010], CelebHQ [Karras et al., 2018], and STL10 [Coates et al., 2011]. We compare CGMMD with the methods in Zhou et al. [2023] and Song et al. [2025] on synthetic data. Moreover, to assess test-time complexity, we compare CGMMD with a diffusion model using classifier-free guidance [Ho & Salimans, 2022]. Due to space constraints, only selected results are shown here; full details appear in Appendix C.

5.1 Synthetic experiment: conditional bivariate sampling

In this section, we compare our proposed CGMMD with two baseline approaches: the GCDS [Zhou et al., 2023], a vanilla GAN framework, and a Wasserstein-based modification, WGAN (trained with pure Wasserstein loss) [Song et al., 2025].
We consider a synthetic setup with $\boldsymbol{X} \sim \mathrm{N}(0,1)$, $\boldsymbol{U} \sim \mathrm{Unif}[0,2\pi]$, and $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2 \overset{\text{iid}}{\sim} \mathrm{N}(0,\sigma^2)$. The response variables are $\boldsymbol{Y}_1 = 2\boldsymbol{X} + \boldsymbol{U}\sin(2\boldsymbol{U}) + \boldsymbol{\varepsilon}_1, \boldsymbol{Y}_2 = 2\boldsymbol{X} + \boldsymbol{U}\cos(2\boldsymbol{U}) + \boldsymbol{\varepsilon}_2$, and our goal is to generate conditional samples from $(\boldsymbol{Y}_1, \boldsymbol{Y}_2) \mid \boldsymbol{X}$ at varying noise levels (σ). All three methods use the same two-hidden-layer feed-forward ReLU generator with noise $\boldsymbol{\eta}$ concatenated to the generator input, and are evaluated at noise levels $\sigma \in \{0.2, 0.4, 0.6\}$.

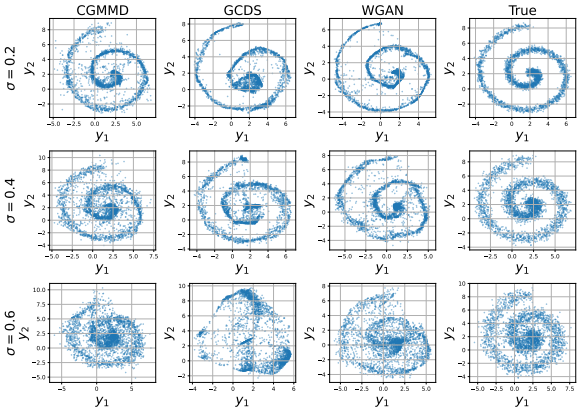

Figure 2: Comparison of conditional generators on the Helix benchmark at $\boldsymbol{X} = 1$.

At low noise ($\sigma = 0.2$), all three methods recover the helix structure well. As the noise level rises, however, CGMMD maintains the overall curvature, in particular at the 'eye' (the center of the helix), while the reconstructions from GCDS and WGAN degrade noticeably (See Figure 2). In this regard we have noticed that without ℓ_1 regularisation WGAN training is often unstable. We also explore an additional conditional bivariate setting (which imitates circular structure), with qualitatively similar results deferred to Appendix C.1.

5.2 Real Data Analysis: image super-resolution and denoising

In this section, we evaluate the performance of CGMMD across two tasks: image super-resolution and image denoising. For this, we use the MNIST and CelebHQ datasets.

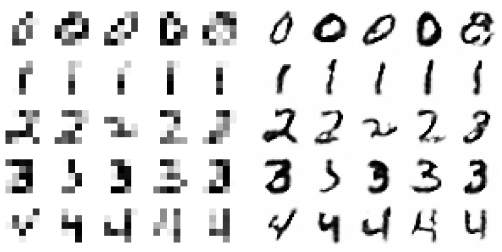

Figure 3: Low and high resolution images for MNIST digits $\{0, 1, 2, 3, 4\}$.

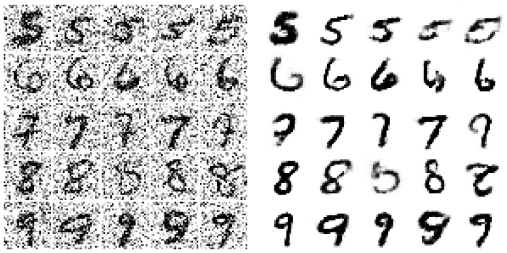

Figure 4: Noisy and denoised MNIST digits $\{5, 6, 7, 8, 9\}$ at $\sigma = 0.5$.

Super-Resolution. We now implement CGMMD for 4X image super-resolution task using MNIST. Given a 7×7 low-resolution input, the model aims to reconstruct the original 28×28 image, treating this as a conditional generation problem: producing a high-resolution image from a low-resolution one. In Figure 3 we show that CGMMD accurately reconstructs the high-resolution images (right panel) from the low-resolution inputs (left panel), and they closely match the ground-truth digits. Additional results and details are given in Appendix C.2

Image Denoising. We evaluate CGMMD on the image denoising task using the MNIST (28×28 iamges) and CelebHQ ($3 \times 64 \times 64$ images) datasets. In this task, the inputs are images (digits for MNIST and facial images for CelebHQ) corrupted with additive Gaussian noise ($\sigma = 0.5, 0.25$ for MNIST and CelebHQ respectively). We can indeed formulate this as a conditional generation problem. In Figure 4, the left 5 columns represent the noisy digit images while the right 5 columns are the clean images reconstructed using CGMMD.Additional experiments and details are given in Appendix C.2.
For the CelebHQ experiment, Figure 5 shows original images (left), noisy inputs (middle), and denoised outputs produced by CGMMD (right). The results demonstrate that our model effectively reconstructs clean facial images from noisy inputs and preserves quality even under high noise levels. Additional denoised images and details are given in Appendix C.3.

Figure 5: CelebHQ denoising using CGMMD at $\sigma = 0.25$.

Comparison with Conditional Diffusion Model. In Table 1, we compare CGMMD with a diffusion model using classifier-free guidance [Ho & Salimans, 2022] on the MNIST image denoising task ($\sigma = 0.9$). The diffusion model produces better reconstructions, but it comes at a much higher computational cost. As shown in the last column of Table 1, generating a single image takes about 5.42×10^{-2} seconds with the diffusion model, whereas CGMMD requires only 5.6×10^{-4} seconds. In other words, *our method is about* $100\times$ *faster*, while still delivering reasonable image quality. This efficiency makes CGMMD attractive for applications where fast conditional sampling is critical.

Table 1: Comparison of CGMMD with conditional diffusion model for MNIST image denoising.

Model	PSNR	SSIM	FID	Inception Score	Generation Time (seconds/ batch)	Generation Time (seconds/ image)
Diffusion Model	13.326	0.861	1.32×10^{-3}	2.07	6.94	5.42×10^{-2}
CGMMD	8.922	0.718	8×10^{-3}	2.411	7.21×10^{-2}	5.6×10^{-4}

Ethics Statement

As this study is purely exploratory and theoretical, relying solely on simulated and benchmark datasets, we do not anticipate any significant ethical concerns.

Reproducibility Statement

To facilitate reproducibility, we include all theoretical results, corresponding proofs, assumptions, and discussions of potential limitations in the main text and Supplementary Materials. All relevant codes are also provided in the Supplementary Materials.

LLM Usage Statement

The authors recognize the use of LLMs for polishing and improving the clarity of the manuscript.

References

Martin Anthony and Peter L Bartlett. *Neural network learning: Theoretical foundations*. cambridge university press, 2009.

Martin Arjovsky and Leon Bottou. Towards principled methods for training generative adversarial networks. In *International Conference on Learning Representations*, 2017.

Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

Mona Azadkia and Sourav Chatterjee. A simple measure of conditional dependence. *The Annals of Statistics*, 49(6):3070–3102, 2021.

Ricardo Baptista, Bamdad Hosseini, Nikola B Kovachki, and Youssef M Marzouk. Conditional sampling with monotone gans: From generative models to likelihood-free inference. *SIAM/ASA Journal on Uncertainty Quantification*, 12(3):868–900, 2024.

Peter L Bartlett, Nick Harvey, Christopher Liaw, and Abbas Mehrabian. Nearly-tight vc-dimension and pseudodimension bounds for piecewise linear neural networks. *Journal of Machine Learning Research*, 20(63):1–17, 2019.

Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*, 2018.

Emanuele Borgonovo, Alessio Figalli, Promit Ghosal, Elmar Plischke, and Giuseppe Savaré. Convexity and measures of statistical association. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, pp. qkaf018, 2025.

Stéphane Boucheron, Gábor Lugosi, and Olivier Bousquet. Concentration inequalities. In *Summer school on machine learning*, pp. 208–240. Springer, 2003.

Ismaël Castillo and Thibault Randrianarisoa. Optional pólya trees: Posterior rates and uncertainty quantification. *Electronic Journal of Statistics*, 16(2):6267–6312, 2022.

Anirban Chatterjee and Bhaswar B Bhattacharya. Boosting the power of kernel two-sample tests. *Biometrika*, 112(1):asae048, 2025.

Anirban Chatterjee, Ziang Niu, and Bhaswar B Bhattacharya. A kernel-based conditional two-sample test using nearest neighbors (with applications to calibration, regression curves, and simulation-based inference). *arXiv preprint arXiv:2407.16550*, 2024.

Jie Chen, Hua Mao, Yuanbiao Gou, Zhu Wang, and Xi Peng. Conditional distribution learning on graphs. *arXiv preprint arXiv:2411.15206*, 2024.

Xiaohong Chen, Oliver Linton, and Peter M Robinson. The estimation of conditional densities. *Asymptotics in Statistics and Probability: Papers in Honor of George Gregory Roussas*, pp. 71–84, 2000.

Kacper Chwialkowski, Heiko Strathmann, and Arthur Gretton. A kernel test of goodness of fit. In *International conference on machine learning*, pp. 2606–2615. PMLR, 2016.

Adam Coates, Andrew Y Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 215–223, 2011.

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.

Sanjoy Dasgupta and Samory Kpotufe. Optimal rates for k-nn density and mode estimation. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/a5549f3f66cedf4204ffe35552e5b59c-Paper.pdf.

Nabarun Deb, Promit Ghosal, and Bodhisattva Sen. Measuring association on topological spaces using kernels and geometric graphs. *arXiv preprint arXiv:2010.01768*, 2020.

Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. *Advances in neural information processing systems*, 28, 2015.

Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

Gintare Karolina Dziugaite, Daniel M Roy, and Zoubin Ghahramani. Training generative neural networks via maximum mean discrepancy optimization. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pp. 258–267, 2015.

Jianqing Fan and Tsz Ho Yim. A crossvalidation method for estimating conditional densities. *Biometrika*, 91(4):819–834, 2004.

Jianqing Fan, Qiwei Yao, and Howell Tong. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, 83(1):189–206, 1996.

Jerome H Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*, 3 (3):209–226, 1977.

Bent Fuglede and Flemming Topsoe. Jensen-shannon divergence and hilbert space embedding. In *International symposium onInformation theory, 2004. ISIT 2004. Proceedings.*, pp. 31. IEEE, 2004.

Kenji Fukumizu, Arthur Gretton, Xiaohai Sun, and Bernhard Schölkopf. Kernel measures of conditional dependence. *Advances in neural information processing systems*, 20, 2007.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Arthur Gretton, Kenji Fukumizu, Choon Teo, Le Song, Bernhard Schölkopf, and Alex Smola. A kernel statistical test of independence. *Advances in neural information processing systems*, 20, 2007.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The journal of machine learning research*, 13(1):723–773, 2012.

Parul Gupta, Munawar Hayat, Abhinav Dhall, and Thanh-Toan Do. Conditional distribution modelling for few-shot image synthesis with diffusion models. In *Proceedings of the Asian Conference on Computer Vision*, pp. 818–834, 2024.

Peter Hall and Qiwei Yao. Approximating conditional distribution functions using dimension reduction. 2005.

William Harvey, Saeid Naderiparizi, and Frank Wood. Conditional image generation by conditioning variational auto-encoders. *arXiv preprint arXiv:2102.12037*, 2021.

Trevor Hastie, Robert Tibshirani, Jerome Friedman, et al. The elements of statistical learning, 2009.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Torsten Hothorn, Thomas Kneib, and Peter Bühlmann. Conditional transformation models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):3–27, 2014.

Jiamin Hou, Azadeh Moradinezhad Dizgah, ChangHoon Hahn, Michael Eickenberg, Shirley Ho, Pablo Lemos, Elena Massara, Chirag Modi, Liam Parker, and Bruno Régaldo-Saint Blancard. Cosmological constraints from the redshift-space galaxy skew spectra. *Physical Review D*, 109 (10):103528, 2024.

Zhen Huang, Nabarun Deb, and Bodhisattva Sen. Kernel partial correlation coefficient—a measure of conditional dependence. *Journal of Machine Learning Research*, 23(216):1–58, 2022a.

Ziyi Huang, Henry Lam, and Haofeng Zhang. Evaluating aleatoric uncertainty via conditional generative models. *arXiv preprint arXiv:2206.04287*, 2022b.

Rob J Hyndman, David M Bashtannyk, and Gary K Grunwald. Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, 5(4):315–336, 1996.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.

Rafael Izbicki and Ann B Lee. Nonparametric conditional density estimation in a high-dimensional regression setting. *Journal of Computational and Graphical Statistics*, 25(4):1297–1316, 2016.

Ariel Jaffe, Yuval Kluger, George C Linderman, Gal Mishne, and Stefan Steinerberger. Randomized near-neighbor graphs, giant components and applications in data science. *Journal of applied probability*, 57(2):458–476, 2020.

Olav Kallenberg. *Foundations of modern probability*, volume 2. Springer.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations (ICLR)*, 2018.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646–1654, 2016.

Achim Klenke. *Probability theory: a comprehensive course*. Springer, 2008.

Lucas Kock and Nadja Klein. Truly multivariate structured additive distributional regression. *Journal of Computational and Graphical Statistics*, pp. 1–13, 2025.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.

Michael Kohler and Sophie Langer. On the rate of convergence of fully connected very deep neural network regression estimates. *arXiv preprint arXiv:1908.11133*, 2019.

Yujia Li, Kevin Swersky, and Rich Zemel. Generative moment matching networks. In *International conference on machine learning*, pp. 1718–1727. PMLR, 2015.

Zhao Lincheng and Liu Zhijun. Strong consistency of the kernel estimators of conditional density function. *Acta Mathematica Sinica*, 1(4):314–318, 1985.

Julia Linhart, Alexandre Gramfort, and Pedro Luiz Coelho Rodrigues. Validation diagnostics for sbi algorithms based on normalizing flows. In *NeurIPS 2022-the 36th conference on Neural Information Processing Systems-Machine Learning and the Physical Sciences workshop*, pp. 1–7, 2022.

Shiao Liu, Xingyu Zhou, Yuling Jiao, and Jian Huang. Wasserstein generative learning of conditional distribution. *arXiv preprint arXiv:2112.10039*, 2021.

Gael M Martin, David T Frazier, and Christian P Robert. Approximating bayes in the 21st century. *Statistical Science*, 39(1):20–45, 2024.

Radha Mastandrea, Benjamin Nachman, and Tilman Plehn. Constraining the higgs potential with neural simulation-based inference for di-higgs production. *Physical Review D*, 110(5):056004, 2024.

Andreas Maurer and Massimiliano Pontil. Uniform concentration and symmetrization for weak interactions. In *Conference on Learning Theory*, pp. 2372–2387. PMLR, 2019.

David Mimno, David M Blei, and Barbara E Engelhardt. Posterior predictive checks to quantify lack-of-fit in admixture models of latent population structure. *Proceedings of the National Academy of Sciences*, 112(26):E3441–E3450, 2015.

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

Ashish Mishra, Shiva Krishna Reddy, Anurag Mittal, and Hema A Murthy. A generative model for zero shot learning using conditional variational autoencoders. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 2188–2196, 2018.

Takeru Miyato and Masanori Koyama. cgans with projection discriminator. *arXiv preprint arXiv:1802.05637*, 2018.

Alfred Müller. Integral probability metrics and their generating classes of functions. *Advances in applied probability*, 29(2):429–443, 1997.

Ryumei Nakada and Masaaki Imaizumi. Adaptive approximation and generalization of deep neural network with intrinsic dimensionality. *Journal of Machine Learning Research*, 21(174):1–38, 2020.

Yijin Ni and Xiaoming Huo. A uniform concentration inequality for kernel-based two-sample statistics. *arXiv preprint arXiv:2405.14051*, 2024.

Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, pp. 2642–2651. PMLR, 2017.

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.

Junhyung Park and Krikamol Muandet. A measure-theoretic approach to kernel conditional mean embeddings. *Advances in neural information processing systems*, 33:21247–21259, 2020.

Poornima Ramesh, Jan-Matthis Lueckmann, Jan Boelts, Álvaro Tejero-Cantero, David S Greenberg, Pedro J Goncalves, and Jakob H Macke. Gatsbi: Generative adversarial training for simulation-based inference. In *The 10th International Conference on Learning Representations (ICLR 2022)*. OpenReview. net, 2022.

Michael Reed and Barry Simon. *Methods of modern mathematical physics: Functional analysis*, volume 1. Gulf Professional Publishing, 1980.

Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis. In *International conference on machine learning*, pp. 1060–1069. Pmlr, 2016.

Brian J Reich, Howard D Bondell, and Lexin Li. Sufficient dimension reduction via bayesian mixture modeling. *Biometrics*, 67(3):886–895, 2011.

Yong Ren, Jun Zhu, Jialian Li, and Yucen Luo. Conditional generative moment-matching networks. *Advances in Neural Information Processing Systems*, 29, 2016.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

Robert A Rigby and D Mikis Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 54(3):507–554, 2005.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Murray Rosenblatt. Conditional probability density and regression estimators. *Multivariate analysis II*, 25:31, 1969.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

Johannes Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function. *The Annals of Statistics*, 48(4):1875 – 1897, 2020. doi: 10.1214/19-AOS1875. URL https://doi.org/10.1214/19-AOS1875.

Antonin Schrab, Benjamin Guedj, and Arthur Gretton. Ksd aggregated goodness-of-fit test. *Advances in Neural Information Processing Systems*, 35:32624–32638, 2022.

Antonin Schrab, Ilmun Kim, Mélisande Albert, Béatrice Laurent, Benjamin Guedj, and Arthur Gretton. Mmd aggregated two-sample test. *Journal of Machine Learning Research*, 24(194):1–81, 2023.

Zuowei Shen, Haizhao Yang, and Shijun Zhang. Deep network approximation characterized by number of neurons. *Communications in Computational Physics*, 28(5):1768–1811, 2020. ISSN 1991-7120. doi: https://doi.org/10.4208/cicp.OA-2020-0149. URL https://global-sci.com/article/79740/deep-network-approximation-characterized-by-number-of-neurons.

Le Song, Jonathan Huang, Alex Smola, and Kenji Fukumizu. Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proceedings of the 26th annual international conference on machine learning*, pp. 961–968, 2009.

Shanshan Song, Tong Wang, Guohao Shen, Yuanyuan Lin, and Jian Huang. Wasserstein generative regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, pp. qkaf053, 08 2025. ISSN 1369-7412. doi: 10.1093/jrsssb/qkaf053. URL https://doi.org/10.1093/jrsssb/qkaf053.

Bharath K Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert RG Lanckriet. Hilbert space embeddings and metrics on probability measures. *The Journal of Machine Learning Research*, 11:1517–1561, 2010.

Bharath K Sriperumbudur, Kenji Fukumizu, and Gert RG Lanckriet. Universality, characteristic kernels and rkhs embedding of measures. *Journal of Machine Learning Research*, 12(7), 2011.

Masashi Sugiyama, Ichiro Takeuchi, Taiji Suzuki, Takafumi Kanamori, Hirotaka Hachiya, and Daisuke Okanohara. Least-squares conditional density estimation. *IEICE Transactions on Information and Systems*, 93(3):583–594, 2010.

Danica J Sutherland, Hsiao-Yu Tung, Heiko Strathmann, Soumyajit De, Aaditya Ramdas, Alex Smola, and Arthur Gretton. Generative models and model criticism via optimized maximum mean discrepancy. *arXiv preprint arXiv:1611.04488*, 2016.

Aad W Van Der Vaart and Jon A Wellner. Weak convergence. In *Weak convergence and empirical processes: with applications to statistics*, pp. 16–28. Springer, 1996.

Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2008.

Mariia Vladimirova, Stéphane Girard, Hien Nguyen, and Julyan Arbel. Sub-weibull distributions: Generalizing sub-gaussian and sub-exponential properties to heavier tailed distributions. *Stat*, 9 (1):e318, 2020.

Mischa von Krause, Stefan T Radev, and Andreas Voss. Mental speed is high until age 60 as revealed by analysis of over a million participants. *Nature human behaviour*, 6(5):700–708, 2022.

Holger Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.

LeCun Yann. Mnist handwritten digit database. *ATT Labs.*, 2010.

Andrew Zammit-Mangion, Matthew Sainsbury-Dale, and Raphaël Huser. Neural methods for amortized inference. *Annual Review of Statistics and Its Application*, 12, 2024.

Zheyuan Zhan, Defang Chen, Jian-Ping Mei, Zhenghe Zhao, Jiawei Chen, Chun Chen, Siwei Lyu, and Can Wang. Conditional image synthesis with diffusion models: A survey, 2025. URL https://arxiv.org/abs/2409.19365.

Shijun Zhang, Zuowei Shen, and Haizhao Yang. Deep network approximation: Achieving arbitrary accuracy with fixed number of neurons. *Journal of Machine Learning Research*, 23(276):1–60, 2022.

Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 286–301, 2018.

Xingyu Zhou, Yuling Jiao, Jin Liu, and Jian Huang. A deep generative approach to conditional sampling. *Journal of the American Statistical Association*, 118(543):1837–1848, 2023.