Draining Fictitious Knots: Restoring Distance-Awareness Guarantees for High-Dimensional Spline Networks

Masoud Ataei
Electrical and Computer Engg.
University of Maine
Orono, ME, USA
masoud.ataei@maine.edu

Mohammad Javad Khojasteh
Electrical and Microelectronic Engg.
Rochester Institute of Technology
Rochester, NY, USA
mjkeme@rit.edu

Vikas Dhiman
Electrical and Computer Engg.
University of Maine
Orono, ME, USA
vikas.dhiman@maine.edu

***Abstract*—Kolmogorov-Arnold Networks (KANs) with spline activations have recently shown promise for interpretable function approximation. Distance-Aware Error for Kolmogorov Networks (DAREK) introduces a computationally efficient bottom-up approach to uncertainty quantification by equipping KANs with distance-aware error bounds; yet, in high-dimensional settings, the theoretical guarantees can be weakened by the emergence of fictitious knots. Inspired by the Kolmogorov-Arnold representation theorem, DAREK adopts a componentwise formulation in which each input dimension is treated separately; as a result, induced knot locations may appear in the combined input space without corresponding to actual training data. These fictitious knots mislead the DAREK uncertainty estimator into reporting low uncertainty far from any real observation, violating the distance-awareness guarantee. We identify this failure mode precisely, characterize its geometric structure, and propose a drainage uncertainty mechanism that restores distance-awareness by constructing a monotonically decreasing uncertainty path from any fictitious knot region toward the nearest real knot. The proposed drainage method provides a practical heuristic correction that mitigates the fictitious-knot failure mode while restoring theoretical distance-awareness in high-dimensional settings. Experiments on a 2D synthetic benchmark and a 100-dimensional face dataset show that drainage raises sampled distance-awareness (SDA) from 85% to 98-99%, matching Gaussian processes at lower computational cost.**

***Index Terms*—Uncertainty quantification, neural networks, spline neural networks, Kolmogorov-Arnold networks (KAN).**

I. Introduction

Splines are smooth piecewise polynomials that have long been used for function approximation in control and signal processing [1], [2]. They have also been studied as adaptive activation functions in neural networks [3]–[5]. Their recent use in Kolmogorov-Arnold Networks (KANs) [6] has renewed interest in spline-based neural architectures due to their expressivity and interpretability. Beyond these properties, distance-aware uncertainty is crucial in safety-critical applications such as autonomous navigation, where learned models should behave conservatively away from the training distribution [7], [8]. Fig. 1 shows a toy example where a standard classifier is uncertain only near its decision boundary —whether linear (left) or closed and nonlinear (middle)— and remains confident (low uncertainty) elsewhere, including far from the *training* data, causing a distant test point (star) to be labeled confidently. A distance-aware model (right) instead reports high uncertainty away from the training points, correctly flagging the same distant test point as unreliable.

Distance-aware uncertainty has traditionally been provided by Gaussian processes (GPs) [9] with kernels such as radial basis function (RBF), but their computational complexity grows cubically in the number of training samples. Deep ensembles [10] and Monte Carlo dropout (MC-D) [11] are cheaper, stochastic uncertainty estimators, but are not reliably distance-aware far from training data [8]. Deterministic single-forward-pass methods such as SNGP [8] and DUE [12] recover distance-awareness by pairing a distance-preserving feature extractor, enforced through spectral normalization or a bi-Lipschitz constraint, with a Gaussian-process output layer. However, these methods require architectural modification and end-to-end training, and their distance-awareness holds with respect to distances measured in the *learned feature space*; DAREK, by contrast, defines distance-awareness directly in the input space. Verification methods such as CROWN [13] instead bound the output of a fixed neural network over a prescribed input perturbation set using convex relaxations; they address robustness of the learned model, whereas DAREK bounds approximation error relative to the unknown target as a function of distance from observed data. Recently, DAREK [7] introduced a deterministic worst-case uncertainty framework for spline neural networks (SNNs), providing an efficient, sampling-free alternative that bounds each spline's error and propagates the resulting uncertainty through the network composition. Moreover, K-DAREK extends DAREK by applying distance-aware error bounds to Kurkova-Kolmogorov networks [14], [15], combining MLP components with spline-based structures to improve efficiency and stability while leveraging their expressive power [16].

Inspired by the Kolmogorov-Arnold representation theorem [17], [18], DAREK formulates distance-awareness compo-

This work is supported by the National Science Foundation under Grant No. 2218063 and the Gleason Endowment at RIT.

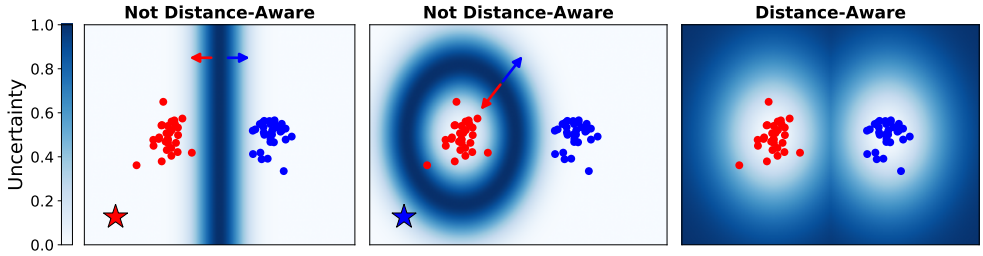

Fig. 1. Toy binary classification. **Left)** and **middle)** Non-distance-aware models are confident (low uncertainty) far from the data. Models are uncertain only near the decision boundary. A linear separator (left) and a closed nonlinear one (middle). Consequently, the distant query (star) lands in a confident region despite being far from training data. **Right)**, a distance-aware model raises uncertainty away from the training clusters.

nentwise, treating each input coordinate through separate univariate spline bounds. Although each component-level bound is distance-aware, its bottom-up propagation through summation and composition need not remain distance-aware in the full input space. This creates a geometric failure mode: coordinate-wise knots induce a grid of apparent knot locations, most of which are not actual training samples. We call these artificial locations *fictitious knots*. Because the propagated uncertainty can vanish at fictitious knots just as it does at real knots, DAREK may assign low uncertainty to points that are far from all training observations, thereby weakening its high-dimensional distance-awareness guarantee.

In this paper, we characterize the geometry of fictitious knots and introduce a drainage-based uncertainty mechanism to restore distance-awareness by linking fictitious-knot regions to nearby real training knots. Specifically, we

- identify and formalize the *fictitious knot* problem in high-dimensional KANs, showing how componentwise spline processing can break joint distance-awareness;
- propose drainage uncertainty, which restores a monotone uncertainty path from fictitious knot regions to the nearest real knot;
- bound the drainage-region volume and characterize its dependence on drainage thickness and input dimension;
- validate the estimator on a 2D synthetic benchmark and a 100-dimensional task, raising sampled distance-awareness from roughly 85% to 98–99% and matching a GP at lower computational cost; and
- release the implementation and experiment code [1].

Drainage provides a practical heuristic fix for fictitious knots, restoring theoretical distance-awareness while leaving the development of joint high-dimensional uncertainty bounds beyond componentwise methods for future work.

II. Background

In this section, we review the distance-awareness condition and the metric used to evaluate it, introduce spline neural networks, and summarize the worst-case error bounds underlying DAREK [19]. We use the notation m for the number of knots, k for the order of the spline, and n for the input dimension. Let $f : [a, b] \to \mathbb{R}$ be a scalar function, $\mathcal{T} = \{\tau_1, \ldots, \tau_m\}$ an ordered set of m distinct knots, and $\mathcal{D}_f(\mathcal{T}) = \{(\tau_i, f(\tau_i))\}_{i=1}^m$ the values observed at those knots. Let the input space $\mathcal{X} \subset \mathbb{R}^n$ and the finite set of training inputs $\mathcal{X}_\mathcal{D} \subset \mathcal{X}$. Knots are selected from the training inputs, so $\mathcal{T} \subset \mathcal{X}_\mathcal{D}$. A piecewise polynomial $\hat{f}$ of order k over knots $\mathcal{T}$ is a continuous function whose restriction to each interval $[\tau_j, \tau_{j+1})$ is a polynomial of order-k; we write $\mathcal{P}_{k,j}[\mathcal{D}_{\hat{f}}(\mathcal{T})]$ for the j-th piece. We denote the kth derivative of a function f by $f^{(k)}$.

A. Distance-awareness

A distance-aware uncertainty estimator is a model that reports lower confidence for query points that lie farther from the training data, as defined mathematically below.

Definition 1 (Input distance-awareness): Let $\hat{y} = \hat{f}(\mathbf{x})$ approximate a target $y = f(\mathbf{x})$ from inputs $x \in \mathcal{X}_\mathcal{D}$, and let $u_{\hat{f}}(\mathbf{x}) : \mathcal{X} \to \mathbb{R}^+$ be an uncertainty estimator that bounds the true values in the interval $y \in [\hat{f}(\mathbf{x}) - u_{\hat{f}}(\mathbf{x}), \hat{f}(\mathbf{x}) + u_{\hat{f}}(\mathbf{x})]$ for every $\mathbf{x} \in \mathcal{X}$. Then $u_{\hat{f}}(\mathbf{x})$ is **distance-aware** if it increases monotonically with the test point's distance from the training data under some distance function $d(\mathbf{x}, \mathcal{X}_\mathcal{D})$.

We use a specific distance function, which measures distance to the single closest training point (or knot) τ^* with a set distance $d(., \mathcal{X}_\mathcal{D})$ defined as

$$d(\mathbf{x}, \mathcal{X}_\mathcal{D}) = \min_{\tau \in \mathcal{T}} d_u(\mathbf{x}, \tau) = d_u(\mathbf{x}, \tau^*). \tag{1}$$

We call $d_u(\mathbf{x}, \tau^*)$ the *inducing distance* of the uncertainty estimator $u_{\hat{f}}$. It need not be Euclidean; it can be a geodesic distance on the data manifold. When both the uncertainty estimator and the inducing distance are differentiable, distance-awareness can be written as

$$[\nabla_\mathbf{x} u_{\hat{f}}(\mathbf{x})]^\top \nabla_\mathbf{x} d_u(\mathbf{x}, \tau^*) \geq 0 \quad \forall \mathbf{x} \in \mathcal{X}, \tag{2}$$

meaning that uncertainty increases with increasing distance.

Checking (2) at every point over $\mathcal{X}$ is intractable, so we estimate it from samples, fixing the inducing distance to be Euclidean for comparability across estimators, as follows.

Definition 2 (Sampled Distance-Awareness (SDA)): For a test point $\mathbf{x}_t$ drawn uniformly from $\mathcal{X}_{\text{test}}$, SDA is defined as

$$\text{SDA} = \frac{1}{N} \sum_{\mathbf{x}_t \sim \mathcal{X}_{\text{test}}} \mathbb{1}\left[\nabla_\mathbf{x} u_{\hat{f}}(\mathbf{x}_t)^\top (\mathbf{x}_t - \boldsymbol{\tau}^*) \geq 0\right], \tag{3}$$

where $\mathbb{1}[.]$ is the indicator function. SDA measures the fraction of sampled directions along which uncertainty correctly rises away from the nearest knot.

[1] https://github.com/Masoud-Ataei/Fictitious-Knots

B. Error of a spline network

Here we review the spline network we use (KANs) and the worst-case error bounds of the DAREK approach that our correction builds on.

Spline: A spline of order k is a piecewise polynomial of the same order that is $\mathcal{C}^{(k-1)}$ times continuous.

Spline networks: In KAN, the scalar weights of an MLP are replaced with learnable univariate spline activations. A two-layer KAN mapping $\mathbb{R}^{n_1} \to \mathbb{R}$ is $\mathrm{KAN}_2(\mathbf{x}) = \sum_{i=1}^{n_2} \phi_{2,1,i}\big(\sum_{j=1}^{n_1} \phi_{1,i,j}(x_j)\big)$, where each $\phi_{l,i,j}$ is a kth-order spline that connects input j to output i in layer l. Each spline in the KAN implementation is constructed as $\phi_{l,i,j}(x_j) = \boldsymbol{\alpha}_{l,i,j}^{\top} \mathbf{b}_{k,\tau_{l,j}}(x_j)$, where $\mathbf{b}_{k,\tau_{l,j}}$ are B-spline bases. lth layer of KAN can be written as $\mathbf{h}_l(\mathbf{x}) = (\sum_{j=1}^{n_1} \phi_{l,i,j}(x_j))_{i=1}^{n_l}$ and an L-layer KAN is the composition

$$\mathrm{KAN}_L(\mathbf{x}) = \mathbf{h}_L \circ \cdots \circ \mathbf{h}_1(\mathbf{x}), \qquad \mathbf{h}_l : \mathbb{R}^{n_l} \to \mathbb{R}^{n_{l+1}}. \quad (4)$$

Choosing the knots as a subset of the training inputs, $\mathcal{T} \subset \mathcal{X}_{\mathcal{D}}$, as proposed in DAREK [7], fixes the basis functions by the data, so only the coefficients $\alpha_{l,i,j}$ are learned, making the KAN a *semi-parametric* model. Here n_l is the dimension of layer l and $n_1 = n$ is the input dimension.

DAREK [7] computes a worst-case error bound for an SNN at a test point, using m selected training inputs as the knots of the spline network. We review the DAREK results needed here; complete derivations and proofs are available in [19]. DAREK additionally makes the following Lipschitz-continuity assumption:

Assumption 1 (kth-order Lipschitz continuity): For a $(k-1)$-times differentiable $f : [a, b] \to \mathbb{R}$, the kth-order Lipschitz constant L_f^k satisfies $|f^{(k-1)}(x) - f^{(k-1)}(y)| \le L_f^k\, d(x, y)$ for all $x \neq y$, where $f^{(k)}$ is the kth derivative of function f.

Theorem 1 (Newton interpolation error bound [7]): Let $f \in C^{k+1}$ on $[a, b]$ be $(k+1)$th-order Lipschitz with constant L_f^{k+1}. The kth-order Newton piecewise-polynomial fit $\mathcal{P}_{k,j}[\mathcal{D}_f(\mathcal{T})]$ through the knots satisfies, for $x \in [\tau_j, \tau_{j+1})$,

$$|f(x) - \mathcal{P}_{k,j}[\mathcal{D}_f](x)| \le \frac{L_f^{k+1}}{(k+1)!} \Big| \prod_{i=1}^{k+1} (x - \tau_i^{(j)}) \Big| := \bar{u}_f(x; \mathcal{T}), \quad (5)$$

which we denote $\bar{u}_f(x; \mathcal{T})$ as interpolation error bound and $\tau_i^{(j)}$ are the $k+1$ knots close by x.

The proof is provided in [19]. We restate the result here to establish the notation used throughout the remainder of this paper.

Proposition 1 (Distance-awareness of the interpolation-error bound [19]): Let $f : [a, b] \to \mathbb{R}$ be the scalar function with knots $\mathcal{T} = \{\tau_1, \ldots, \tau_m\}$. On each interval $[\tau_j, \tau_{j+1})$, the bound $\bar{u}_f(\cdot; \mathcal{T})$ of (5) vanishes at both knots, is strictly positive in the interior, and has a unique maximum $x_j^\star \in (\tau_j, \tau_{j+1})$. It is strictly increasing on $(\tau_j, x_j^\star)$ and strictly decreasing on $(x_j^\star, \tau_{j+1})$.

Error bound with linear fit (EBL): In practice, the trained network does not pass exactly through the knots, so there is a residual $e_j^f(x) := f(x) - \hat{f}_{[j]}(x)$ whose values at the knots, $\mathcal{D}_{e_j^f}(\tau_{1:m})$, are known. We account for this residual by adding a piecewise-linear interpolation of the absolute knot errors, $|\mathcal{P}_{1,j}[\mathcal{D}_{e_j^f}](x)| = [\tau_j, \tau_{j+1}]|e_j^f|(x - \tau_j) + |e_j^f|$, where the slope is the divided difference $[\tau_j, \tau_{j+1}]|e_j^f| = (|e_{j+1}^f| - |e_j^f|)/(\tau_{j+1} - \tau_j)$ and refer to this approach as *error bound with linear fit* (EBL).

$$u_f(x; \boldsymbol{\tau}_{1:m}) := \bar{u}_f(x; \boldsymbol{\tau}_{1:m}) + |\mathcal{P}_{1,j}[\mathcal{D}_{e_j^f}](x)| \quad \text{(EBL)}. \quad (6)$$

III. Problem Statement

In this section, we formalize the failure mode that motivates the paper. We first define real and fictitious knots (Sec. III-A), then demonstrate that the bottom-up DAREK bound vanishes at every fictitious knot, and, as a consequence, need not remain distance-aware once the input dimension exceeds one (Sec. III-B).

A. Real and fictitious knots

In KAN-based architectures, a separate univariate spline is applied to each input coordinate. Consider $\mathcal{T}$ with m selected knots from N training samples ($m \le N$), used as the knots of a single layer's spline. The spline acting on coordinate j therefore takes its knots from the projection of the selected points onto that axis, $\mathcal{T}_j = \{\boldsymbol{\tau}_{1,j}, \ldots, \boldsymbol{\tau}_{m,j}\} \subset \mathbb{R}$, for $j \in \{1, \ldots, n\}$. Because the coordinates are processed independently, the first-layer interpolation-error term $u(x) = \sum_{j=1}^{n} \bar{u}_j(x_j)$ depends on x only through the n separate proximities of each x_j to the axis knots, never through the joint position of x. Since each $\bar{u}_j \ge 0$ vanishes exactly at its knots (Theorem 1), u vanishes precisely on the grid $\mathcal{G} = \mathcal{T}_1 \times \mathcal{T}_2 \times \cdots \times \mathcal{T}_n$, where its cardinality is $|\mathcal{G}| = m^n$.

For instance, consider $n = 2$ and two selected knots $\boldsymbol{\tau}_1 = (-1, 0)$ and $\boldsymbol{\tau}_2 = (1, 2)$. Their per-axis projections are $\mathcal{T}_1 = \{-1, 1\}$ and $\mathcal{T}_2 = \{0, 2\}$, so the grid is $\mathcal{G} = \mathcal{T}_1 \times \mathcal{T}_2 = \{(-1, 0), (-1, 2), (1, 0), (1, 2)\}$, with $|\mathcal{G}| = 2^2 = 4$ vertices, even though only the two points $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2$ were ever selected. Of these four grid vertices, $\mathcal{R} = \{(-1, 0), (1, 2)\}$ are *real knots*, while $\mathcal{F} = \{(-1, 2), (1, 0)\}$ are *fictitious knots*, arising only from recombining coordinates across axes. Fig. 2 shows a 5×5 grid produced by $m = 5$ knots in $n = 2$ dimensions.

Definition 3 (Real and fictitious knots): A grid vertex $g \in \mathcal{G}$ is a *real knot* if it is one of the selected knot points, $g \in \mathcal{R} := \mathcal{T}$, and a *fictitious knot* otherwise, $g \in \mathcal{F} := \mathcal{G} \setminus \mathcal{R}$.

Real knots are the points the knot selection intended to mark; fictitious knots are the spurious coincidences created by recombining the per-axis projections. Each selected point contributes its projection to every $\mathcal{T}_j$, so $\mathcal{T} \subseteq \mathcal{G}$ and $|\mathcal{R}| = m$, while $|\mathcal{G}| = m^n$. The fraction of grid vertices that are real is therefore $\frac{|\mathcal{R}|}{|\mathcal{G}|} = \frac{m}{m^n} \xrightarrow[n \to \infty]{} 0$ $(m \ge 2)$, so in high dimensions the apparent knot grid is almost entirely fictitious.

The interpolation-error term cannot distinguish the two cases: $\bar{u}_j$ vanishes at every axis knot, so $u(g) = 0$ for every $g \in \mathcal{G}$, real or fictitious. The distance to the data $d(x, \mathcal{X}_D) = \min_{\tau \in \mathcal{R}} \|x - \tau\|$ vanishes only on $\mathcal{R}$. A fictitious knot can lie arbitrarily far from every training input while still receiving zero interpolation error. Consequently, the

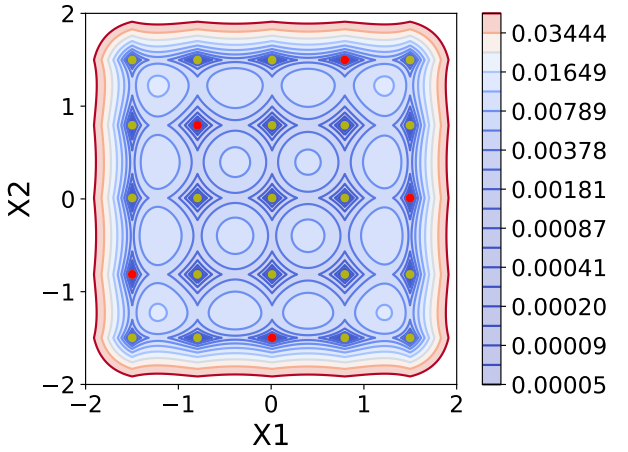

Fig. 2. Interpolation-error bound $\bar{u}$ of (5). Red circles (•) are real knots $\mathcal{R}$ and olive circles (•) are fictitious knots $\mathcal{F}$. Near-zero uncertainty appears at every vertex of the 5×5 grid, not just at the five real knots: $\bar{u}$ vanishes across all of $\mathcal{G}$, so fictitious vertices receive near-zero uncertainty.

theoretical distance-awareness guarantee can weaken in high-dimensional settings, where the bound may decrease to near-zero uncertainty at fictitious knots even though such locations can be far from the training data. We call this the **fictitious-knot problem**. Note that this is a theoretical weakening of the guarantee, not necessarily a large practical effect: our experiment with DAREK still observes SDA around 84-87% (Sec. V, Table I) even without correction; the drainage method closes this remaining gap and formally restores the guarantee.

Equivalently, the per-coordinate "nearest knot" of Theorem 1 is not a point but the hyperplane $\{x : x_j = \tau\}$; the n such hyperplanes intersect in the grid $\mathcal{G}$, and the bottom-up bound treats a query as close to the data whenever it is near one hyperplane in *each* coordinate at once, irrespective of its joint location.

B. The violation at fictitious knots

The collapse identified in Sec. III-A ($u \equiv 0$ on the whole grid $\mathcal{G}$) is not merely a loss of tightness; it creates regions where uncertainty can decrease while distance from the data increases at fictitious knots, as stated in Lemma 1.

Lemma 1 (Fictitious-knot violation): Let $u(x) = \sum_{j=1}^{n} \bar{u}_j(x_j)$ be the first-layer interpolation-error bound, and let $g \in \mathcal{F}$ be a fictitious knot whose nearest real knot is τ^*, so $d(g, \mathcal{X}_D) = \|g - \tau^*\| > 0$. Then on the approach to g along the segment from τ^*, the uncertainty u strictly decreases while $d(\cdot, \mathcal{X}_D)$ strictly increases, so $[\nabla_x u(x)]^\top \nabla_x d(x, \mathcal{X}_D) < 0$ on a neighborhood of g, violating (2). Hence, the DAREK bound is distance-aware component-wise, but the propagated bound in the full input space for $n > 1$ is not.

Proof 1: Since g has nearest real knot τ^* and Voronoi cells are convex polytopes, the segment $\{x(t) = \tau^* + t(g - \tau^*) : t \in [0, 1]\}$ lies in $\mathcal{V}(\tau^*)$, where $d(x(t), \mathcal{X}_D) = \|x(t) - \tau^*\| = t\,\|g - \tau^*\|$ increases in t. By Definition 3, both endpoints lie on $\mathcal{G}$, so $u(\tau^*) = u(g) = 0$, while $u > 0$ on the open segment because the product from (5) is strictly positive away from knots. Thus g is an isolated minimizer of u and u is strictly decreasing as $x(t) \to g$. Therefore $\nabla_x u$ points back toward τ^* (the direction of increasing u), whereas $\nabla_x d(\cdot, \mathcal{X}_D) = (x - \tau^*)/\|x - \tau^*\|$ points away from τ^*; the two gradients are opposed and their inner product is negative.

Composition across layers is expected to propagate rather than repair the violation, since each layer adds the previous layers' summed error; we do not analyze deeper-layer composition separately (see Limitations). The violation stays confined to the geometrically small fictitious-knot regions characterized in Sec. III-A; DAREK's overall SDA remains above chance because those regions are a small fraction of the input space. We correct the estimate directly in Sec. IV-A.

IV. Methodology

With the problem defined in Sec. III, we now introduce a drainage-based correction, prove that it restores distance-awareness under an explicit condition on its gain (Theorem 2), and bound the fraction of the input space the correction modifies (Corollary 1).

A. Drainage Estimator

We reshape the uncertainty so that a monotone path to the data is restored, without retraining the network or altering its knots. This is achieved by introducing a linear "drainage" path, which smoothly redirects uncertainty toward the nearest real knot. Let $\tau^*(x)$ be the nearest real knot to x, τ^+ the nearest fictitious knot to τ^*, and $\Delta = \|\tau^* - \tau^+\|$ the spacing from a real knot to its nearest grid neighbor; we first define the drainage term abstractly.

Definition 4 (Drainage uncertainty): A nonnegative function $u_d : \mathcal{X} \to \mathbb{R}^+$ is a **drainage uncertainty** for the real knots $\mathcal{R}$ if it is expressed through the inducing distance $d(x, \mathcal{X}_D)$ and satisfies

$$\text{(D1)}\; u_d(\tau) = 0 \quad \forall \tau \in \mathcal{R}, \tag{7}$$

$$\text{(D2)}\; u_d \text{ is strictly increasing in } d(x, \mathcal{X}_D). \tag{8}$$

Property (D2) makes u_d distance-aware by construction: writing $u_d(x) = \psi\big(d(x, \mathcal{X}_D)\big)$ with $\psi' > 0$ gives $[\nabla_x u_d]^\top \nabla_x d = \psi' \|\nabla_x d\|^2 \geq 0$. Two natural choices are the linear ramp and a saturating (RBF-type) form, $u_d^{\text{lin}}(x) = \eta \frac{d(x, \mathcal{X}_D)}{\Delta}$, $u_d^{\text{rbf}}(x) = \eta\Big(1 - e^{-d(x, \mathcal{X}_D)^2 / 2\ell^2}\Big)$, where $\Delta = \|\tau^* - \tau^+\|$ is the spacing from a real knot to its nearest grid neighbor and $\eta, \ell > 0$. The linear form rises at a constant rate set by a single gain η; the saturating form connects to the kernels of distance-aware Gaussian processes, trading the gain for a length-scale ℓ.

Definition 5 (Drainage-corrected estimator): Given a drainage uncertainty u_d, the drainage-corrected estimator u_{da} is

$$u_{\text{da}}(x) = \begin{cases} u(x), & |x_j - \tau_j^*(x)| < \epsilon \quad \forall j, \\ \max\{\, u(x),\, u_d(x) \,\}, & \text{otherwise.} \end{cases} \tag{9}$$

The first case $|x_j - \tau_j^*(x)| < \epsilon \quad \forall j$ corresponds to the genuine box around the nearest real knot. Inside this box, the principled DAREK bound is used unchanged; elsewhere, the estimate may be lifted by drainage. By construction, $u_{\mathrm{da}} \geq u$ everywhere, so the worst-case enclosure $y \in [\hat{f} \pm u_{\mathrm{da}}]$ is preserved (drainage never reports less than DAREK), and $u_d \to 0$ as $x \to \tau^*$, so the lift vanishes at the data and the correction is not over-conservative near training points.

Remark 1: The combination uses $\max$ rather than replacing u: a query may satisfy the partial-proximity condition yet have small u_d, and replacing u there could report less than the original DAREK bound. Taking the maximum keeps the larger of the DAREK bound and the distance-aware ramp.

In the remainder, we instantiate $u_d = u_d^{\mathrm{lin}}$, chosen for its single interpretable gain. Theorem 2 below gives an explicit sufficient threshold condition on this gain for restoring distance-awareness.

B. Restored distance-awareness

Drainage restores the distance-awareness condition (2) when its gain is sufficiently large. To state this condition, within a Voronoi cell $\mathcal{V}(\tau^*)$, let $U_{\max} = \sup_{y \in \mathcal{V}(\tau^*)} u(y)$ denote the largest value of the DAREK bound in that cell.

Theorem 2 (Restored distance-awareness): Suppose

$$\text{(C1)} \quad \epsilon < \tfrac{1}{2} \min_j(\text{adjacent-knot gap on axis } j), \tag{10}$$

$$\text{(C2)} \quad \eta \geq \frac{\Delta\, U_{\max}}{\epsilon}. \tag{11}$$

Then for every $x \in \mathcal{V}(\tau^*)$ the segment from x to τ^* is a path along which u_{da} is monotonically non-increasing, ending at $u_{\mathrm{da}}(\tau^*) = u(\tau^*)$. Consequently u_{da} satisfies the distance-awareness condition (2) on $\mathcal{V}(\tau^*)$, and hence on all of $\mathcal{X}$.

Proof 2: Parametrize the segment by $s = \|y - \tau^*\|$, so $d(y, \mathcal{X}_D) = s$ and $u_d(y) = (\eta/\Delta)s$ is linear and increasing. By (C1) the only grid knot in the ϵ-box is τ^*, so for $s < \epsilon$ the genuine box gives $u_{\mathrm{da}} = u$, which rises from $u(\tau^*)$ as a single isolated knot (non-decreasing in s). Leaving the box forces some coordinate to differ from τ^* by at least ϵ, so $s \geq \epsilon$; then by (C2), $u_d(y) = (\eta/\Delta)s \geq (\eta/\Delta)\epsilon \geq U_{\max} \geq u(y)$, hence $u_{\mathrm{da}} = \max\{u, u_d\} = u_d = (\eta/\Delta)s$, again increasing in s. Thus u_{da} is non-decreasing in s, i.e. non-increasing toward τ^*; since $d(\cdot, \mathcal{X}_D) = s$, (2) holds. Repeating per cell gives the global claim.

Remark 2 (The gain is conservative): Condition (C2) bounds drainage against the worst case $U_{\max}$ over the whole cell and is sufficient, not necessary; in practice, much smaller gains restore distance-awareness, as the ablation of Sec. V confirms. A single global $\Delta = \min_{\tau \in \mathcal{R}} \|\tau - \tau^+(\tau)\|$, the minimum grid spacing over all real knots, may be used in place of the per-knot Δ to make u_d continuous across Voronoi boundaries, at the cost of potentially more conservative drainage values in cells with larger per-knot spacing.

Remark 3 (Computation): Drainage adds negligible cost to a DAREK forward pass. Per real knot, τ^+ and Δ are precomputed once, and each query costs $\mathcal{O}(n)$ beyond the nearest-knot lookup. The correction requires no additional model evaluations.

C. Volume of Drainage Region

Although the drainage uncertainty restores distance awareness, the drainage term is not itself derived as a worst-case error bound and may increase the combined bound in the region where it applies. We therefore want the drainage region to be small relative to the rest of the input space. In this section, we quantify this proportion as a function of the thickness ϵ, the number of knots m, and the input dimension n.

Corollary 1 (Ratio of Drainage to Input Volume): Let each input dimension span an interval of length a, with m knots per dimension, drainage thickness ϵ around each knot, and non-overlapping neighborhoods of radius ϵ around the knots. The drainage ratio per dimension becomes $r = 2m\epsilon/a \leq 1$. Let the drainage region be the set of points for which at least one coordinate lies within ϵ of a knot on its corresponding axis; then the fraction of the input volume occupied by the drainage region is $\Delta S = 1 - (1 - r)^n$.

Proof 3: A point lies outside the drainage region if and only if, for every coordinate, it lies farther than ϵ from all m knots on the corresponding axis. Along each coordinate, the knot neighborhoods occupy a total length of $2m\epsilon$, leaving a fraction $1 - 2m\epsilon/a = 1 - r$ outside these neighborhoods, and the fraction lying outside the knot neighborhoods in every coordinate is $(1 - r)^n$. Therefore, the fraction inside the drainage region is $\Delta S = 1 - (1 - r)^n$.

Remark 4 (A thickness–dimension trade-off): For any **fixed** relative thickness $r > 0$, $\Delta S = 1 - (1 - r)^n \to 1$ as $n \to \infty$. Thus, in high dimensions, the union of the knot slabs occupies nearly all of the input domain. This product-volume effect is distinct from the combinatorial growth of the m^n fictitious knots. To keep the drainage region at a target volume fraction r_0, one may choose $r = 1 - (1 - r_0)^{1/n}$, or, equivalently, $\epsilon = a/(2m)[1 - (1 - r_0)^{1/n}] = \mathcal{O}\big(a/(mn)\big)$. However, this reduction in thickness comes at the cost of a larger sufficient gain in (C2), since its lower bound is inversely proportional to ϵ. Thus, the choice of ϵ represents a trade-off between the volume of the drainage region and the gain required to guarantee distance-awareness.

V. Experiments

We evaluate drainage through four studies: a 2D synthetic benchmark that visualizes the fictitious-knot failure mode and its correction; a 100-dimensional face bounding-box task that tests the method in a high-dimensional setting; an ablation over the drainage thickness ϵ, gain η, and knot count m; and a computational-cost comparison against DAREK, ensembles, and Gaussian processes. The uncertainty quality is reported as the sampled distance-awareness (SDA) of Definition 2.

A. 2D synthetic benchmark

Setup: The input domain is $\mathcal{X} = [-2, 2]^2$ and the target is the smooth multimodal function $f(x_1, x_2) = \sin(x_1)\cos(x_2)$.

We draw $N = 1000$ training points uniformly and select $m = 5$ of them as spline knots (real knots). Each of these knots contributes to each coordinate. Because DAREK analyzes the interpolation error separately along each coordinate, the five values on the first axis combine with the five values on the second axis, producing a grid of $5 \times 5 = 25$ apparent knots (Definition 3). Five of these knots preserve the original coordinate pairings and are real knots; the remaining 20 combine coordinates from different training points and are fictitious knots. DK1 and DK2 are one-layer and two-layer DAREK models, and ENS1 and ENS2 are ensembles of 5 one-layer and two-layer KANs, respectively. All models use cubic ($k = 3$) B-spline activations and grid size 5. The approximator is DK1 unless otherwise specified. We use the linear ramp u_d^{lin} with gain $\eta = 0.03$ and box half-width $\epsilon = 0.05$ so that the conditions in Theorem 2 hold. SDA is estimated from 2000 points drawn uniformly on $\mathcal{X}$.

Uncertainty collapse: Fig. 2 plots the $\bar{u}$ (5) with selected contour levels. Concentric wells appear at *every* vertex of the 5×5 grid $\mathcal{G}$, not only at the five real knots $\mathcal{R}$ (red). $\bar{u}$ vanishes on all of $\mathcal{G}$, exactly as Definition 3 predicts. A query sitting at a fictitious vertex therefore receives near-zero uncertainty despite being far from any training point as formalized in Lemma 1.

Draining uncertainty: Fig. 3 shows the three terms side by side. The DAREK panel (left) has low values across the whole interior grid, high only at the far corners. The drainage panel (middle), u_d^{lin}, is zero at the real knots and rises with distance to the nearest one. The combined estimate (right), $u_{\text{da}} = \max\{u, u_d\}$ of (9), keeps the principled DAREK bound inside the ϵ-box around each real knot and lifts the fictitious minima everywhere else, so the spurious wells disappear.

Distance-awareness violations: Fig. 4 visualizes the distance-awareness violation regions of the interpolation error (5), drainage error, and combined error (9) for the two different sets of knots. The small dots are test points and their color shows the uncertainty. The red circles indicate selected knots, and the highlights indicate regions where violations occur. We evaluate the condition (2) on each test point. These regions concentrate at the fictitious knots, confirming that the failure is of geometric origin rather than a sampling artifact.

Monotone path to real knots: Fig. 5 overlays streamlines of $-\nabla u_{\text{da}}$ on the combined uncertainty field. Every streamline terminates at a real knot (the low-uncertainty wells), so from an arbitrary query following decreasing uncertainty leads to actual training data, which is guaranteed by Theorem 2.

Compared with RBF: Fig. 6 normalizes the combined estimate and a RBF-kernel GP to $[0, 1]$. Both are low at the real knots and rise away from them; drainage reproduces the GP's distance-aware bowl structure, differing mainly in the residual slab pattern left by the per-axis grid.

Quantitative result: On this benchmark, the SDA of the uncorrected DAREK variants is $\approx$84–87% (Table I, DK1/DK2), reflecting the directions near fictitious knots where uncertainty wrongly decreases outward. Adding drainage raises SDA to 99.5–99.8%, matching the GP 100% while the other baselines (deep ensembles ENS1/ENS2) stay near random chance 50%.

B. Face bounding box

We evaluate drainage on a high-dimensional face task. The task is to predict the bounding box (coordinates of the face) from an image on the Celebrity Attribute (CelebA) dataset [20]. We use $8k$ training and $1k$ test images, and reduce each image to a 100-dimensional input via PCA feature projection. All KANs use cubic ($k = 3$) splines with 15 knots.

Because drainage only lifts the uncertainty estimate and never alters the network output, RMSE and IoU remain unchanged and are not the focus of this paper; the comparison focuses on SDA alone. As reported in Table I (100D row), DAREK attains 95.7–96.3% SDA, and drainage raises it to 98.4–99.9%, matching the GP (99.55%) while the ensembles remain near random chance ($\approx$ 50%). Drainage adds only $\mathcal{O}(n)$ per query, so it preserves DAREK's order-of-magnitude speed advantage over the cubic-cost GP.

TABLE I
SAMPLED DISTANCE-AWARENESS (SDA) FOR DIFFERENT MODELS FOR THE 2D AND 100D DATASETS.

Model	DK1	DK2	ENS1	ENS2	GP	DK1 + drainage	DK2 + drainage
2D	84.30	87.25	44.85	52.68	100.0	99.50	99.75
100D	96.27	95.67	52.09	50.15	99.55	99.85	98.36

C. Ablation and Computational Cost

In Table II, we study how the drainage parameters, thickness ϵ and the slope η, and the knot count m affect SDA, the drainage volume ratio (ΔS), and the SDA of interpolation error in the 2D experiment. Without drainage, the interpolation error achieves SDA values between 74.3% and 89.0%. Introducing the drainage term consistently increases SDA, reaching 100% for all tested models (m=5, 7, 9) under appropriate parameter settings. The greatest improvement occurs for $m = 7$, where SDA increases from 74.3% without drainage to 100% under appropriate drainage parameters. The results further indicate that increasing η improves SDA when violations are present, but the benefit quickly saturates in this experiment when $\eta = 0.1$. Moreover, at an adequate slope ($\eta \geq 0.1$), the smallest tested drainage region ($\epsilon = 0.025$) suffices to attain 100% SDA for every model order, corresponding to drainage volume ratios of only 0.121, 0.167, and 0.212 for $m = 5, 7, 9$, respectively. These findings suggest that the residual violations are confined to small regions that drainage can target without substantially enlarging the corrected volume.

Finally, we compare wall-clock cost across methods as the number of training/inference points N grows from 50 to 1000 (Fig. 7), with model capacity held fixed at m=5 real knots. DAREK and Drain reuse the same trained model — drainage is a post-hoc correction applied at inference (Eq. 9) — so their fit-time curves coincide exactly, confirming drainage adds no training overhead. The GP, trained here on the full N-point set, shows fit time increasing markedly with N, while DAREK, Drain, and the ensemble remain nearly flat since their capacity does not scale with N; at inference, DAREK/Drain's per-query

TABLE II

ABLATION ON THE DRAINAGE PARAMETERS. THE BASELINE DAREK DISTANCE-AWARENESS ($\text{SDA}_{\text{INT.}}$ (5), NO DRAINAGE) DEPENDS ONLY ON m; $\text{SDA}_{\text{COMBINED}}$ (%) REPORTS THE FULL CORRECTION OF EQ. (9) FOR FOUR VALUES OF THE SLOPE η. THE DRAINAGE VOLUME RATIO ΔS DEPENDS ONLY ON (m, ϵ) AND THE TEST LOSS ONLY ON m. BOLD MARKS THE SMALLEST ΔS THAT ATTAINS SDA $= 100\%$.

				$\text{SDA}_{\text{combined}}$ (%) at $\eta =$			
m	$\text{SDA}_{\text{int.}}$	ϵ	ΔS	0.01	0.03	0.1	0.3
5	89.0	0.025	0.121	98.50	99.80	**100.0**	100.0
		0.05	0.234	98.80	99.90	100.0	100.0
		0.10	0.438	99.10	99.90	100.0	100.0
		0.20	0.750	99.10	99.80	99.90	99.90
7	74.3	0.025	0.167	95.70	99.70	**100.0**	100.0
		0.05	0.319	96.10	99.70	100.0	100.0
		0.10	0.578	96.70	99.80	100.0	100.0
		0.20	0.910	98.30	99.90	99.90	99.90
9	77.0	0.025	0.212	**100.0**	100.0	100.0	100.0
		0.05	0.399	100.0	100.0	100.0	100.0
		0.10	0.698	100.0	100.0	100.0	100.0
		0.20	0.990	100.0	100.0	100.0	100.0

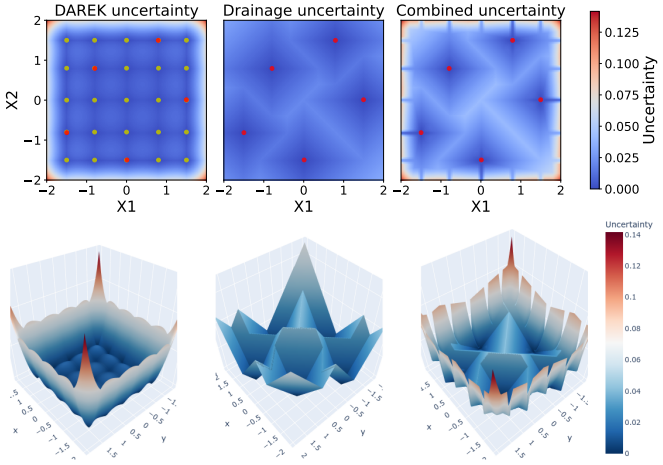

Fig. 3. DAREK uncertainty. **(left)** the original interpolation-error bound (5); **(middle)** the linear drainage uncertainty; and **(right)** the combined estimate (9). Red and olive markers denote real and fictitious knots, respectively; the second row shows the corresponding 3D surfaces.

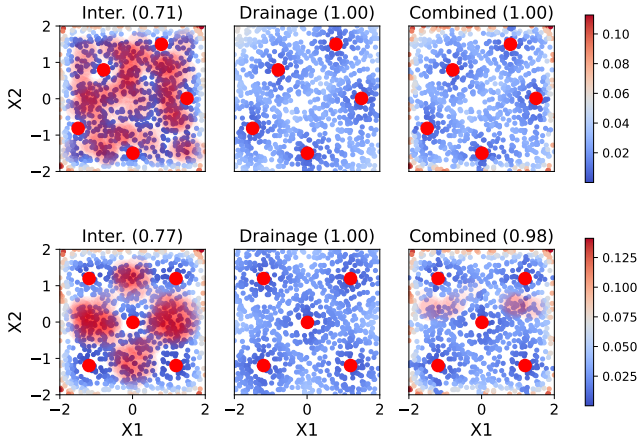

Fig. 4. Distance-awareness violations for two knot configurations. Red dots mark the real knots, shaded regions indicate violations, and parentheses report SDA (3). Drainage substantially reduces the violations of the interpolation-error bound.

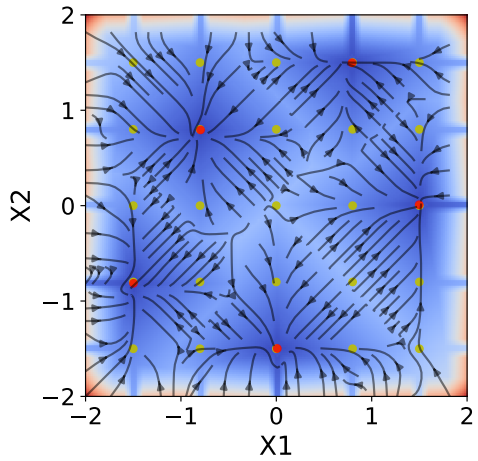

Fig. 5. Streamlines of $-\nabla u_{\text{da}}$ over the combined uncertainty field. Every streamline descends to a real knot, visualizing the monotone path to the training data guaranteed by Theorem 2.

cost stays close to constant, consistent with querying a fixed set of m knots regardless of dataset size.

VI. CONCLUSION

We identified and formalized the **fictitious-knot problem** in high-dimensional spline networks, showing that it arises because a KAN processes each coordinate independently; the knots of its per-axis splines recombine into a grid $\mathcal{G}$ of m^n apparent knots, of which only the fraction m^{1-n} are real knots while the remaining points are fictitious knots. The bottom-up DAREK bound vanishes on this entire grid, so it may report low uncertainty at fictitious knots that lie arbitrarily far from any data, and may therefore violate distance-awareness as the input dimension increases. To repair this, we introduced **drainage uncertainty**, a geometric correction that routes the uncertainty along a monotonically decreasing path toward the nearest real knot, restoring a distance-aware estimate while leaving the trained network untouched. We showed that the corrected region occupies a fraction $1-(1-r)^n$ of the input space, and characterized the trade-off between the drainage thickness and input dimension. On a 2-D synthetic benchmark and a 100-dimensional face task, drainage raised sampled distance-awareness from roughly 85% to 98–99%, matching a Gaussian process at a fraction of the cost.

Limitations. Drainage restores the geometry of the uncertainty estimate, not a tighter worst-case bound: the drainage term u_d is a linear ramp rather than a worst-case error bound, so while the combined estimate remains a valid enclosure (it never reports less than the underlying DAREK bound), its tightness in the drainage regions is not guaranteed. The volume guarantee, moreover, is not free in the limit: for any **fixed**

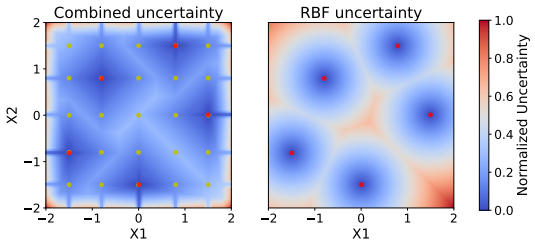

Fig. 6. Combined drainage estimate (left) vs. a RBF-kernel GP (right), each normalized to $[0, 1]$. Both are low at the real knots and rise away from them; drainage reproduces the GP's distance-aware structure without the spurious interior minima of raw DAREK.

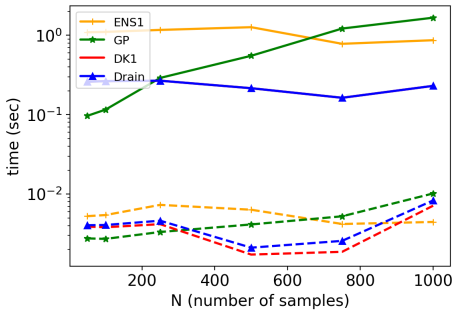

Fig. 7. Wall-clock time vs. number of training/inference points N (50–1000) at fixed model capacity ($m = 5$). Solid lines show training time and dashed lines show inference time. GP training time grows markedly with N, while DAREK, Drain, and the KAN ensemble remain essentially flat.

relative thickness $r > 0$, the drained fraction $1 - (1 - r)^n$ tends to one as $n \to \infty$, as a consequence of the product geometry. Keeping this fraction below a fixed target requires the thickness to shrink with dimension, $\epsilon = \mathcal{O}(a/(mn))$, which in turn increases the sufficient gain required by (C2). The restored-monotonicity guarantee depends on the gain η exceeding a sufficient threshold that scales with the worst-case DAREK uncertainty on each cell, and is therefore conservative; in practice (Sec. V-C) far smaller gains suffice. Finally, our analysis targets the first layer, where the knots are tied directly to the input data, and distance-awareness is measured in input space; composition across deeper layers propagates the corrected estimate but is not separately analyzed here, and the sampled metric (SDA) averages over uniformly drawn test points, which can understate a failure that is concentrated near the fictitious-knot grid.

Future work. Several directions follow naturally. The ramp could be replaced by a Lipschitz-based extension anchored at the real knots, yielding a drainage term that is itself a valid worst-case bound. Pairing drainage with conformal calibration would add distribution-free coverage on top of its geometric monotonicity, combining the per-point distance-awareness that conformal prediction lacks with the marginal guarantees it provides. Adapting the thickness ϵ and gain η per region, rather than globally, would tighten the correction further. A further direction is to connect these distance-aware approximation-error bounds with verification methods such as CROWN, which bound the output variation of a fixed learned network over prescribed input regions, to study how the two complementary guarantees can be combined.

REFERENCES

[1] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal processing magazine*, vol. 16, no. 6, pp. 22–38, 1999.

[2] M. Egerstedt and C. Martin, *Control theoretic splines: optimal control, statistics, and path planning*. Princeton University Press, 2009.

[3] S. Guarnieri, F. Piazza, and A. Uncini, "Multilayer feedforward networks with adaptive spline activation function," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 672–683, 1999.

[4] B. Igelnik and N. Parikh, "Kolmogorov's spline network," *IEEE Transactions on Neural Networks*, vol. 14, no. 4, pp. 725–733, 2003.

[5] P. Bohra, J. Campos, H. Gupta, S. Aziznejad, and M. Unser, "Learning activation functions in deep (spline) neural networks," *IEEE Open Journal of Signal Processing*, vol. 1, pp. 295–309, 2020.

[6] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljacic, T. Y. Hou, and M. Tegmark, "KAN: Kolmogorov–arnold networks," in *ICLR*, 2025.

[7] M. Ataei, M. J. Khojasteh, and V. Dhiman, "DAREK-distance aware error for Kolmogorov networks," in *ICASSP*, 2025, pp. 1–5.

[8] J. Liu, Z. Lin, S. Padhy, D. Tran, T. Bedrax Weiss, and B. Lakshminarayanan, "Simple and principled uncertainty estimation with deterministic deep learning via distance awareness," *NeurIPS*, vol. 33, pp. 7498–7512, 2020.

[9] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. The MIT Press, 2006.

[10] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *NeurIPS*, vol. 30, 2017.

[11] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *ICML*, vol. 48. PMLR, 20–22 Jun 2016, pp. 1050–1059.

[12] J. Van Amersfoort, L. Smith, A. Jesson, O. Key, and Y. Gal, "On feature collapse and deep kernel learning for single forward pass uncertainty," *arXiv:2102.11409*, 2021.

[13] H. Zhang, T.-W. Weng, P.-Y. Chen, C.-J. Hsieh, and L. Daniel, "Efficient neural network robustness certification with general activation functions," *Advances in neural information processing systems*, vol. 31, 2018.

[14] V. Kůrková, "Kolmogorov's theorem and multilayer neural networks," *Neural networks*, vol. 5, no. 3, pp. 501–506, 1992.

[15] J. D. Toscano, L.-L. Wang, and G. E. Karniadakis, "KKANs: Kurkova-Kolmogorov-Arnold networks and their learning dynamics," *Neural Networks*, vol. 191, p. 107831, 2025.

[16] M. Ataei, V. Dhiman, and M. J. Khojasteh, "K-DAREK: Distance aware error for Kurkova Kolmogorov networks," in *Asilomar Conference on Signals, Systems and Computers*, 2025.

[17] J. Schmidt-Hieber, "The Kolmogorov–Arnold representation theorem revisited," *Neural networks*, vol. 137, pp. 119–126, 2021.

[18] S. Morris, "Hilbert 13: Are there any genuine continuous multivariate real-valued functions?" *Bulletin of the American Mathematical Society*, vol. 58, no. 1, pp. 107–118, 2021.

[19] M. Ataei, M. J. Khojasteh, and V. Dhiman, "Distance-aware error for Spline networks: A bottom-up approach to uncertainty," *arXiv:2501.04757v2*, 2026.

[20] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3730–3738.