How Deep Are Deep GPs, Really?

A Sharp Threshold and a Non-Gaussian Limit for Compositional GPs

Mark Kozdoba
Technion, IIT

Shie Mannor
Technion, IIT and NVIDIA

Overview

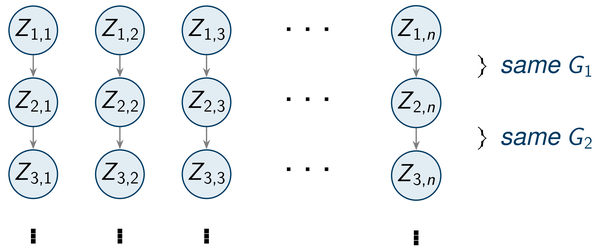
Priors in deep Bayesian models are built by **composing random functions**. We take each random function to be a Gaussian process (GP) and ask what the prior becomes as the depth L grows.

Previously, only *degenerate* limits were known. We show that *non-degenerate* limits also exist — genuinely **new prior objects**.

The compositional prior

Consider the process $\bar{Z}_i = (Z_{i,1}, \ldots, Z_{i,n})$, $\bar{Z}_i \in (\mathbb{R}^d)^n$, with

$$\bar{Z}_{i+1} = G_i \bar{Z}_i = (G_i Z_{i,1}, \ldots, G_i Z_{i,n}).$$



$G_1, G_2, \ldots$ are i.i.d. $\mathbb{R}^d$-valued Gaussian processes with the RBF kernel $k(x, x') = \exp(-\|x - x'\|^2/(2r^2))$.

The generative model

- The composition above can be viewed as a **generative** prior model (Damianou & Lawrence, 2013).
- With larger depth L, we might expect to be able to express more complex distributions.
- **What happens when we take $L \to \infty$?**

Known results

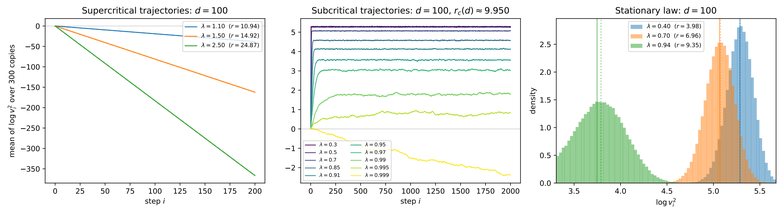
- **Dunlop et al. (2018):** for $r > \sqrt{d}$ the chain **synchronises** — a degenerate limit, collapsing onto constant functions. *Not useful as a prior.*
  - Rich functions stay *available*, but only constants are *likely*.
- **Infinite-width NN priors** collapse to a *single* GP. ReLU synchronises; tanh has a non-synchronised phase. The limit is **always a GP**.

Results

- **Sharp threshold** $r_c(d) = \sqrt{2}\, e^{\psi(d/2)/2} \sim \sqrt{d}$ (ψ = digamma):
  - For $r > r_c(d)$ the chain synchronises almost surely at an explicit exponential rate. This sharpens Dunlop et al.'s $\sqrt{d}$.
  - For $r < r_c(d)$ the chain does not synchronise.
- **Convergence.** For every $r < r_c(d)$ and every $n \geq 2$, the chain $\bar{Z}_i$ converges in total variation to a *unique* stationary law $\pi_{\bar{Z}}$ on $(\mathbb{R}^d)^n$. Until now the only known depth limit was collapse-to-constants.
- **The limit is non-Gaussian** with non-vanishing dependence between coordinates. Each marginal is $\bar{Z}_{\infty,t} \sim \mathcal{N}(0, I_d)$, yet no pair $(\bar{Z}_{\infty,s}, \bar{Z}_{\infty,t})$ is jointly Gaussian, and the dependence does not wash out as $d \to \infty$.

Dichotomy illustration

Consider r in the form $r = \lambda\, r_c(d)$, $\lambda > 0$. Set $v_i = \|Z_{i,s} - Z_{i,t}\|$.



$d = 100$. **Left:** above r_c, $\log v_i^2$ decays (synchronisation). **Middle:** below r_c it equilibrates. **Right:** the stationary law of $\log v_i^2$.

Exchangeability

Because the same G_i acts on all coordinates, $\pi_{\bar{Z}}$ is **exchangeable**.

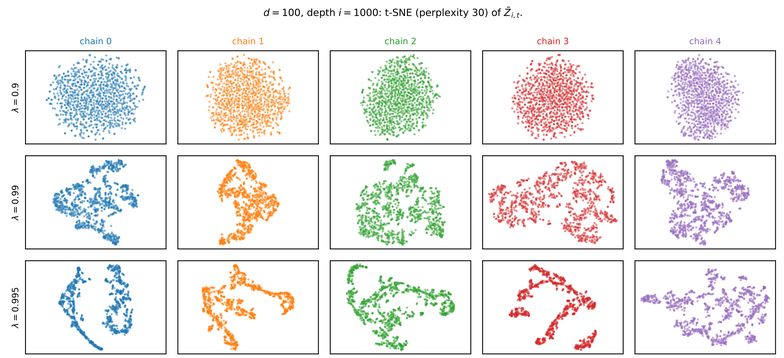
- By de Finetti it is a mixture of i.i.d. products:

$$\pi_{\bar{Z}} = \int_{\mathcal{P}(\mathbb{R}^d)} \overbrace{M \otimes M \otimes \cdots \otimes M}^{n \text{ times}}\, dQ(M).$$

- M is a random *directing measure* on $\mathbb{R}^d$ and $Q = \mathrm{Law}(M)$; conditional on M the coordinates are i.i.d. M.
- An *explicit generative prior* form.
- **What are the distributions M?**

Samples from the prior

M can exhibit complex behaviours.



Per-chain t-SNE of $\bar{Z}_{\infty,t} \in \mathbb{R}^{100}$ at depth 1000 ($n = 1000$); by de Finetti each panel is $\approx$ an i.i.d. sample from one draw of M ($r = \lambda\, r_c(d)$; rows $\lambda = 0.90,\ 0.99,\ 0.995$, columns five i.i.d. chains).

- **Small λ** (top): each $M \approx \mathcal{N}(0, I_d)$ with near-independent coordinates — a featureless blob, alike across chains. **As $\lambda \uparrow 1$**: the M turn **non-Gaussian**, dependent, and structured, differing chain-to-chain.
- The window of visible structure is narrow and shrinks with dimension — it lives at $1 - \lambda \asymp 1/d$ — so the **theory** ($r_c(d)$ and the dependence deficit) locates the right λ interval per d.
- **Not a t-SNE artefact:** the same chain-specific structure appears under linear PCA.

Paper, code, contact

Paper

Code

TECHNION Israel Institute of Technology

(RL)² Reinforcement Learning Research Labs

arXiv 2606.08218 · github.com/komarkdev/deep-gp-thresholds · markk@technion.ac.il

Supported by the Israel Science Foundation (grant No. 2198/25).