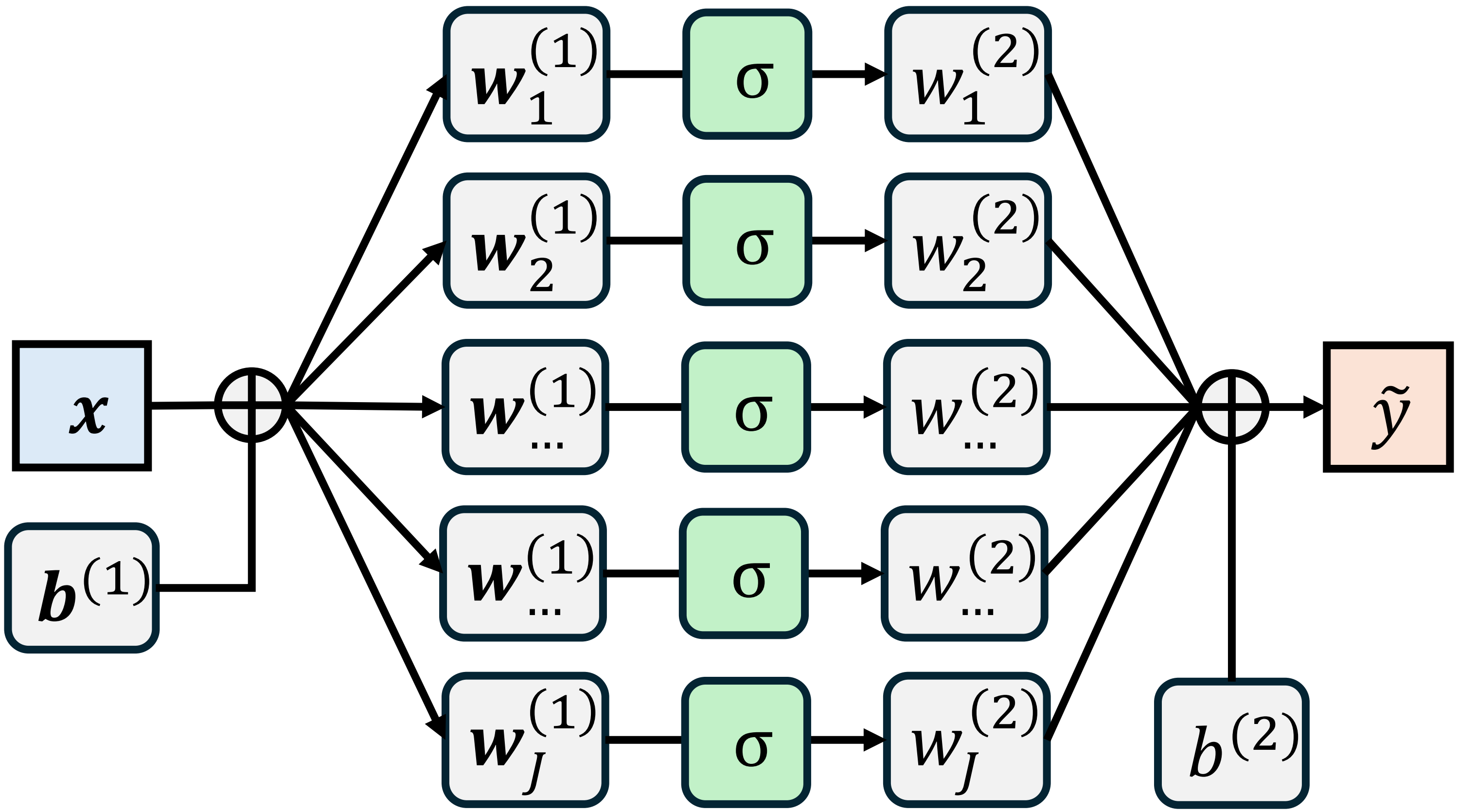
# Infinite Neural Operators: Gaussian processes on functions

*Daniel Augusto de Souza*[1], Yuchen Zhu[1], Harry Jake Cunningham[1], Yuri Saporito[2], Diego Mesquita[2], and Marc Peter Deisenroth[1]

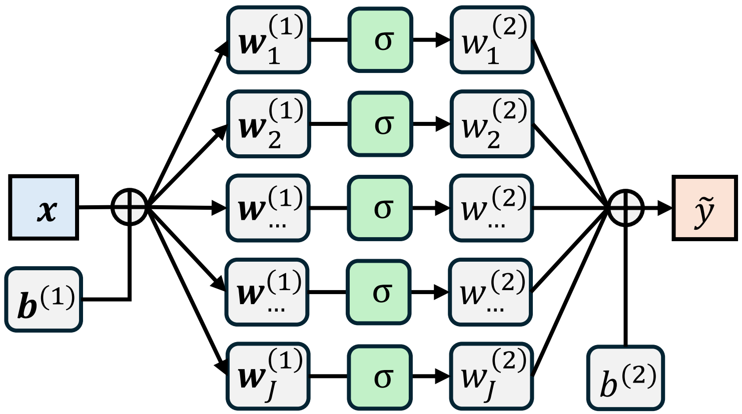
[1]University College London, [2]Fundação Getulio Vargas

# Background

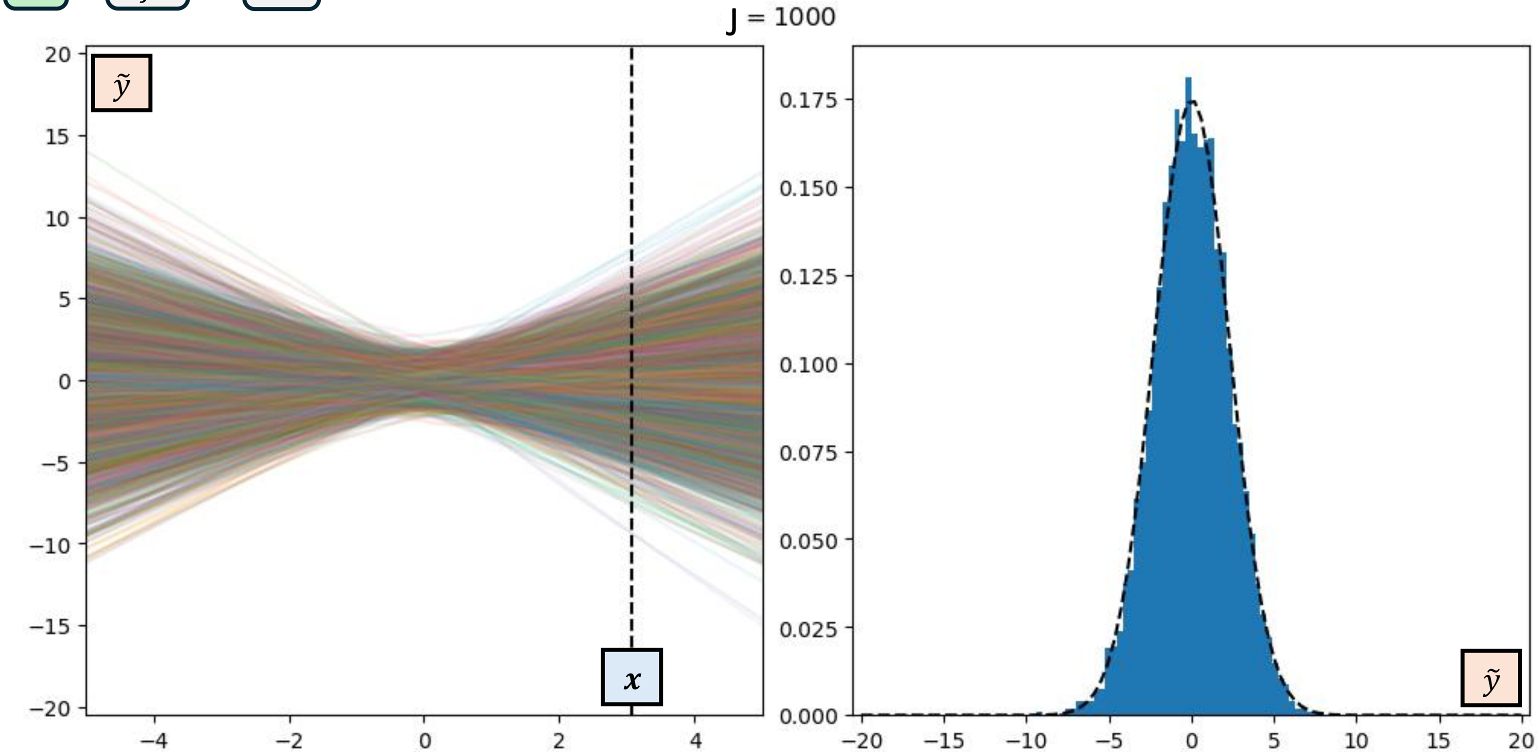
## Neural networks and their infinite limits



# Background

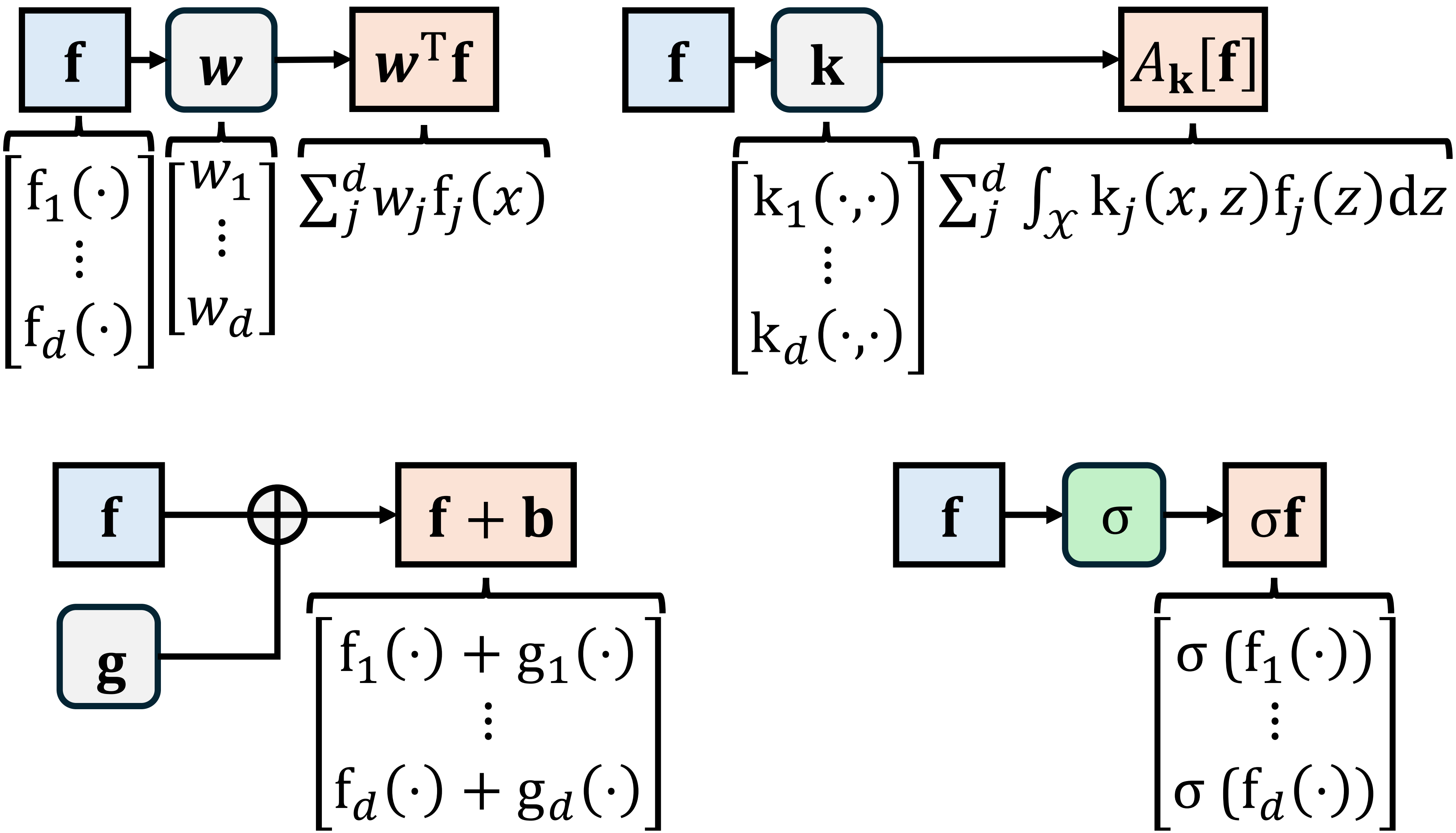



$$w_{ij}^{(1)} \sim N(0,1/J)$$
$$w_j^{(2)} \sim N(0,1/J)$$



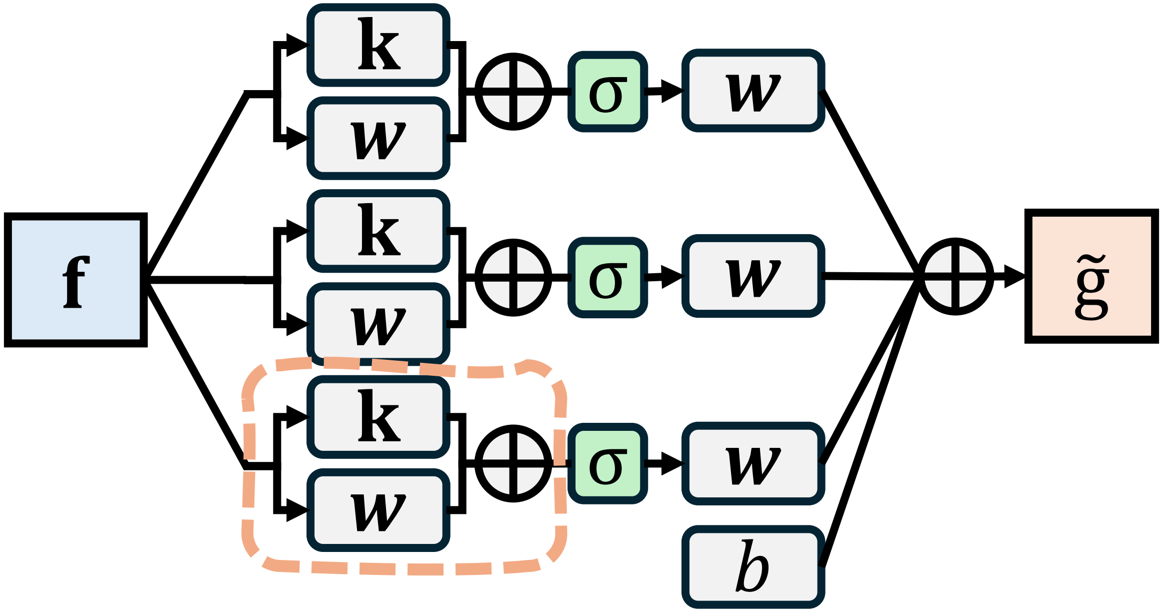
# Background

## Neural operators



# Background

## Neural operators




**Neural operator layer**

$$\sum_j^d \int_{\chi} k_j(x,z) f_j(z) dz + w_j f_j(x)$$

# Background

## Probability in Hilbert spaces

- Unlike with neural networks, the operation $A_{\mathbf{k}}[\mathbf{f}](\mathrm{x})$ can fail when $\int_{\mathcal{X}} \mathbf{k}(x,z)^{\mathrm{T}} \mathbf{f}(z) \mathrm{d}z = \pm\infty$.
- Assume $\mathbf{f}$ and $\mathbf{k}$ belong to a Lebesgue space:
  - $\mathbf{f} \in L^2(\mathcal{X}; \mathbb{R}^d)$:
  - $\mathbf{k} \in L^2(\mathcal{X} \times \mathcal{X}; \mathbb{R}^d)$:
- $\mathrm{F}(\omega): \Omega \to L^2(\mathcal{X}; \mathbb{R}^d)$ is a random element in $L^2$, if, for all $\mathbf{g} \in L^2(\mathcal{X}; \mathbb{R}^d)$, $\langle \mathbf{g}, \mathrm{F}(\cdot) \rangle_{L^2} \in \mathbb{R}$ is a random variable.

# Background

## Probability in Hilbert spaces

- For a set $\mathcal{A}$, an ***operator-valued covariance function*** $\mathbf{C}[a,b](x_1,x_2)\colon \mathcal{A}\times\mathcal{A}\to\mathcal{X}\times\mathcal{X}\to\mathbb{R}^{d\times d}$ such that:
  - **Symmetry:** $\mathbf{C}[a,b](x_1,x_2)=\mathbf{C}[b,a](x_2,x_1)^{\mathrm{T}}$
  - **Positive semi-definite:**
    For any $\{(a_i,\mathbf{h}_i)\}_{i=1}^N\subset\mathcal{A}\times L^2\left(\mathcal{X};\mathbb{R}^d\right)$ and $\left\{\alpha_{ij}\right\}_{i,j=1}^N\subset\mathbb{R}$

$$\sum_{i,j=1}^N \alpha_{ij}\left\langle \mathbf{h}_j, A_{C[a_i,a_j]}[\mathbf{h}_i]\right\rangle > 0$$

- Every operator-valued covariance function $\mathbf{C}$ determines a centered Gaussian process $B\colon\mathcal{A}\to L^2\left(\mathcal{X};\mathbb{R}^d\right)$:
  - For any $\{(a_i,\mathbf{h}_i)\}_{i=1}^N$, $[\langle\mathbf{h}_1,B(a_1)\rangle,\dots,\langle\mathbf{h}_N,B(a_N)\rangle]^{\mathrm{T}}\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$
    with $\boldsymbol{\Sigma}_{ij}=\left\langle\mathbf{h}_j,A_{C[a_i,a_j]}[\mathbf{h}_i]\right\rangle$.

# Infinite limits of neural operators

## Lemma (Compositionality of covariance functions)

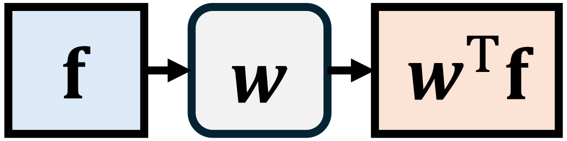
- Let $B_1 : L^2(\mathcal{X}; \mathbb{R}^d) \rightarrow L^2(\mathcal{X}; \mathbb{R}^J)$ be a random operator and $B_2 : L^2(\mathcal{X}; \mathbb{R}^J) \rightarrow L^2(\mathcal{X})$ be a centered function-valued Gaussian process. If the following assumptions hold:
  1. For all $\mathbf{f} \in L^2(\mathcal{X}; \mathbb{R}^d)$ and $\mathbf{x} \in \mathcal{X}$, each component of $B_1[\mathbf{f}](\mathbf{x}) \in \mathbb{R}^J$ is independent and identically distributed such that the covariance function $\mathbf{C}_{B_1}[\mathbf{f}, \mathbf{g}] = \mathrm{c}_{B_1}[\mathbf{f}, \mathbf{g}] \cdot \boldsymbol{I}_J$;
  2. The covariance function of $B_2$ can be expressed, for all $\mathbf{f}, \mathbf{g} \in L^2(\mathcal{X}; \mathbb{R}^J)$ as $\mathrm{c}_{B_2}[\mathbf{f}, \mathbf{g}] = \mathrm{c}_{B_2}\left[{}^1/_J\, \mathbf{g}^\top \mathbf{f}\right]$ and the function $h \mapsto c_{B_2}[h]$ is a continuous map from $L^2(\mathcal{X} \times \mathcal{X})$ to itself.

  Then, $B_2 \circ B_1$ converges in distribution to a function-valued Gaussian process as $J \rightarrow \infty$, and

  $c_{B_2 \circ B_1}[\mathbf{f}_1, \mathbf{f}_2] = c_{B_2}\left[c_{B_1}[\mathbf{f}_1, \mathbf{f}_2]\right].$

# Infinite limits of neural operators

## Lemma (Compositionality of covariance functions)



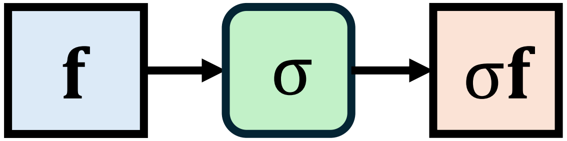

- For the operator $\boldsymbol{w}^{\mathrm{T}}[\cdot]$, it becomes a centered Gaussian process by placing the i.i.d. distribution $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_d)$, giving rise to the covariance function:

$$\mathrm{c}_w[\mathbf{f}_1, \mathbf{f}_2](\boldsymbol{x}_1, \boldsymbol{x}_2) = \sigma^2 \mathbf{f}_2^{\mathrm{T}}(\boldsymbol{x}_2)\mathbf{f}_1(\boldsymbol{x}_1)$$

- This operator satisfies condition 2 of the lemma, as it depends only on $\mathbf{f}_2^{\mathrm{T}}\mathbf{f}_1(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathbf{f}_2^{\mathrm{T}}(\boldsymbol{x}_2)\mathbf{f}_1(\boldsymbol{x}_1)$ and it is homogenous as: $\alpha \mathrm{c}_w[\mathbf{f}_1, \mathbf{f}_2] = \mathrm{c}_w\left[\alpha \mathbf{f}_2^{\mathrm{T}}\mathbf{f}_1\right]$.

# Infinite limits of neural operators

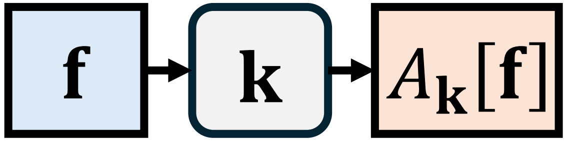
## Lemma (Compositionality of covariance functions)




- For the operator $\sigma[\mathbf{f}]$, this is not a well-defined operator for any activation function $\sigma(\cdot)$. However, the popular ReLU, ELU, tanh, and sigmoid activations satisfy the sufficient condition of being linearly bounded.
- For a GP $B\colon L^2\left(\mathcal{X};\mathbb{R}^d\right) \to L^2(\mathcal{X};\mathbb{R}^J)$ with i.i.d. outputs, $\sigma \circ B\colon L^2\left(\mathcal{X};\mathbb{R}^d\right) \to L^2(\mathcal{X};\mathbb{R}^J)$ is not a GP, but:
  - $\mathbf{C}_{(\sigma\circ B)}[\mathbf{f}_1,\mathbf{f}_2] = \mathrm{c}_{(\sigma\circ B)}[\mathbf{f}_1,\mathbf{f}_2] \cdot \boldsymbol{I}_J$, thus, outputs are still i.i.d, satisfying condition 1 of the lemma.
  - $\mathrm{c}_{(\sigma\circ B)}[\mathbf{f}_1,\mathbf{f}_2]$ can be computed from $\mathrm{c}_B$ and the dual kernel (Han et al., 2022) of $\sigma(\cdot)$.

# Infinite limits of neural operators

## Lemma (Compositionality of covariance functions)




- As the kernel integral operator $A_{\mathbf{k}}$ is linear, it also becomes a function-value GP when the kernel $\mathbf{k} \in L^2\left(\mathcal{X} \times \mathcal{X}; \mathbb{R}^d\right)$ follows a GP distribution $\mathbf{k} \sim \mathrm{GP}(\mathbf{0}, \mathrm{c}_{\mathbf{k}} \cdot \boldsymbol{I}_d)$, giving rise to the covariance function:

$$\mathrm{c}_{A_{\mathbf{k}}}[\mathbf{f}_1, \mathbf{f}_2](\boldsymbol{z}_1, \boldsymbol{z}_2) = \iint_{\mathcal{X}} \mathrm{c}_{\mathbf{k}}(\boldsymbol{z}_1, \boldsymbol{x}_1, \boldsymbol{z}_2, \boldsymbol{x}_2) \mathbf{f}_2^{\mathrm{T}} \mathbf{f}_1(\boldsymbol{x}_1, \boldsymbol{x}_2) \mathrm{d}\boldsymbol{x}_1 \mathrm{d}\boldsymbol{x}_2$$

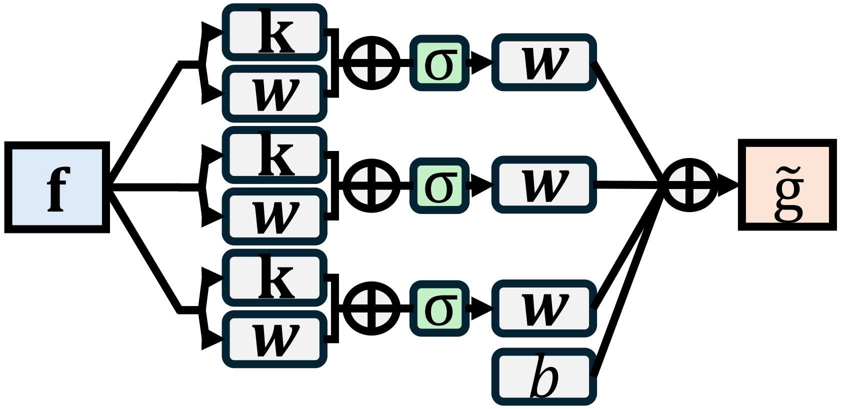
  Again, this covariance function satisfies condition 2 of our lemma.

# Infinite limits of neural operators

## Theorem (Infinite-width neural operators are Gaussian processes)

- Given the lemma, we recursively take each layer's width $J_\ell$ to infinite.
- $\boldsymbol{w}^{(\ell)} \sim \mathcal{N}\left(\mathbf{0}, \sigma_\ell^2 / J_{\ell-1}\, \boldsymbol{I}_{J_\ell}\right)$; $\mathbf{k}^{(\ell)} \sim \mathrm{GP}\left(\mathbf{0}, c_{\mathrm{k}^{(\ell)}} / J_{\ell-1}\, \boldsymbol{I}_{J_\ell}\right)$.
- The covariance function of the NO is obtained by composition following lemma:

$$\mathrm{c}_\infty[\mathbf{f}_1, \mathbf{f}_2] = c_{H_L}\left[c_{H_{L-1}}[\cdots]\right]$$



# Infinite limits of neural operators

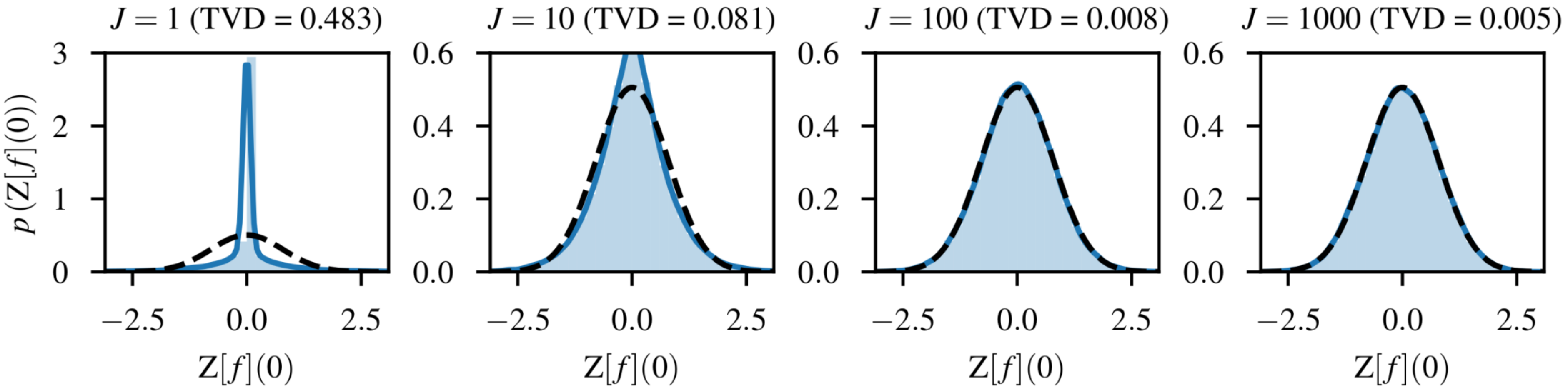
## Fourier infinite neural operator

- Three assumptions are imposed into $\mathbf{k}$: periodicity $\mathcal{X} = \mathbb{T}^{d_x}$, shift-invariance $\mathbf{k}(\boldsymbol{z} - \boldsymbol{x})$, and band-limitedness.
- $\mathrm{c}_{A_\mathbf{k}}[\mathbf{f}_1, \mathbf{f}_2](\boldsymbol{z}, \boldsymbol{z}') =$
$$\sigma_\mathbf{k}^2 (2\pi)^{2d_x} \sum_{\boldsymbol{s}\in\boldsymbol{B}} \mathrm{FS}_{-\boldsymbol{s}}[\mathbf{f}_2]^\mathrm{T} \mathrm{FS}_{\boldsymbol{s}}[\mathbf{f}_1] \exp\left(i\boldsymbol{s}^\mathrm{T}(\boldsymbol{z} - \boldsymbol{z}')\right),$$
where $\boldsymbol{B} = \{-B, \dots, B\}^{d_x}$ and $\mathrm{FS}_{\boldsymbol{s}}$ is the $\boldsymbol{s}$-th Fourier series coeff.

## Matérn infinite neural operators

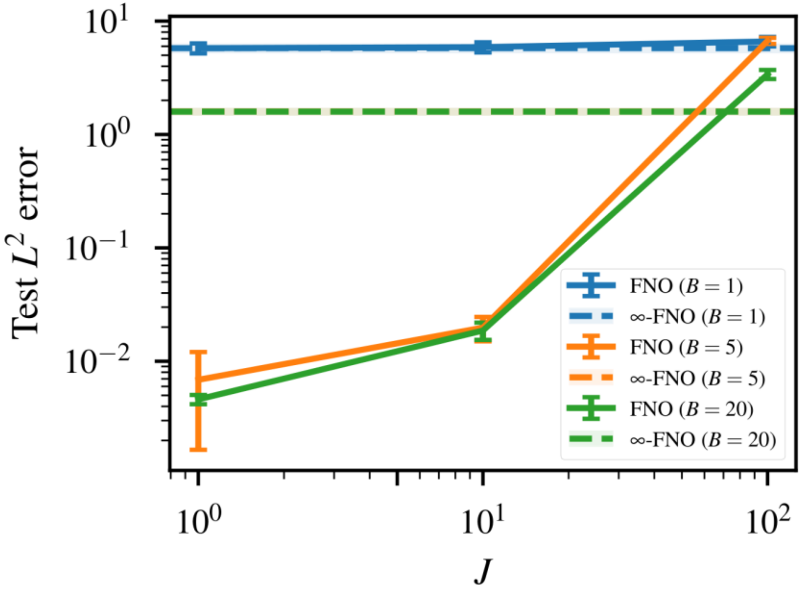
- If we only impose periodicity $\mathcal{X} = \mathbb{T}^{d_x}$, we can place a toroidal Matérn GP prior on $\mathbf{k}(\mathbf{z}, \mathbf{x})$.
- $\mathrm{c}_{A_\mathbf{k}}[\mathbf{f}_1, \mathbf{f}_2](\boldsymbol{z}, \boldsymbol{z}') =$
$$\sum_{\boldsymbol{s}\in\mathbb{Z}^{d_x}} \exp\left(i\boldsymbol{s}^\mathrm{T}(\boldsymbol{z} - \boldsymbol{z}')\right) \hat{\mathrm{c}}_Z(\boldsymbol{s}) \sum_{\boldsymbol{s}\in\mathbb{Z}^{d_x}} \mathrm{FS}_{-\boldsymbol{s}}[\mathbf{f}_2]^\mathrm{T} \mathrm{FS}_{\boldsymbol{s}}[\mathbf{f}_1]\, \hat{\mathrm{c}}_X(\boldsymbol{s})$$
where $\hat{c}(\boldsymbol{s}) = \prod_{j=1}^{d_x} \left({}^{2\nu}/_{\ell^2} + s_j\right)^{-\nu-1/2}$ is the spectral density.

# Experimental validation

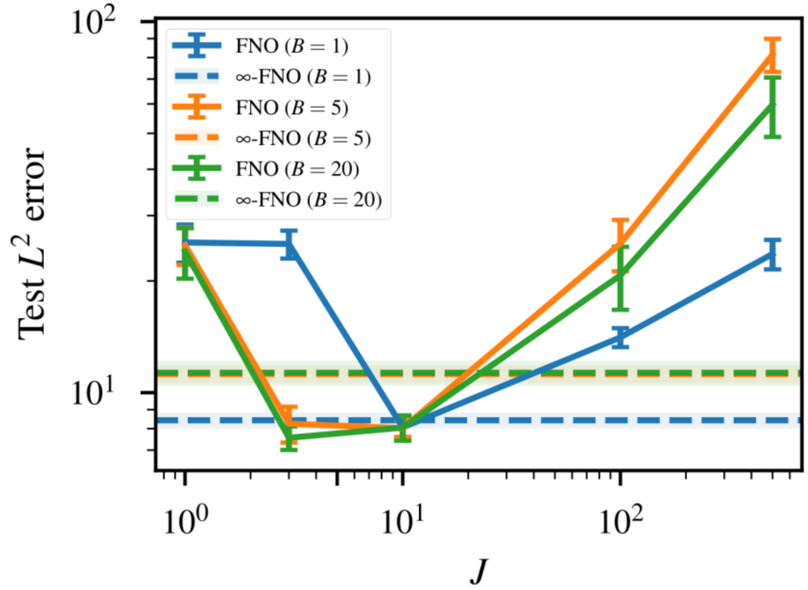



A density estimation of the empirical distribution of the output of increasing channel dimension compared to the infinite width distribution. On top of each plot, we show the total variation distance of the empirical distribution against the infinite width distribution.

# Experimental validation




(a) Synthetic data. ∞-FNO with $B = 5$ (▪▪) and $B = 20$ (▪▪) overlap.




(b) 1D Burgers' equation ($\nu = 0.002$). ∞-FNO with $B = 5$ (▪▪) and $B = 20$ (▪▪) overlap.

Results for the regression experiments. Mean and std. of test $L^2$ loss as a function of width $J$ for different band-limits $B$.

# Conclusion

**Future work.**

- Extending the neural tangent kernel framework to Hilbert space values, thus investigating SGD-trained infinite NOs.
- Deriving covariance functions for other architectures, such as the graph neural operator.

**Limitations.**

- Our implementation scales cubically in both the evaluation grid size and the number of training functions. Existing literature can be incorporated to reduce this cost.
- Only considered the toroidal case, thus we can't incorporate other boundary conditions.

Code, poster and more!

spectral.space

# Thank you!