Deep Neural Networks are Reproducing Kernel Chains

Tjeerd Jan Heeringa, Len Spek, Christoph Brune

MIA, Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

Mathematics of Imaging & AI

AI & Mathematics

Introduction

A key open question is

What is the appropriate function space for deep neural networks?

Desirable properties:

- Based on kernel spaces
- Is consistent
- Works for all common activations functions
- Duality between data and weights

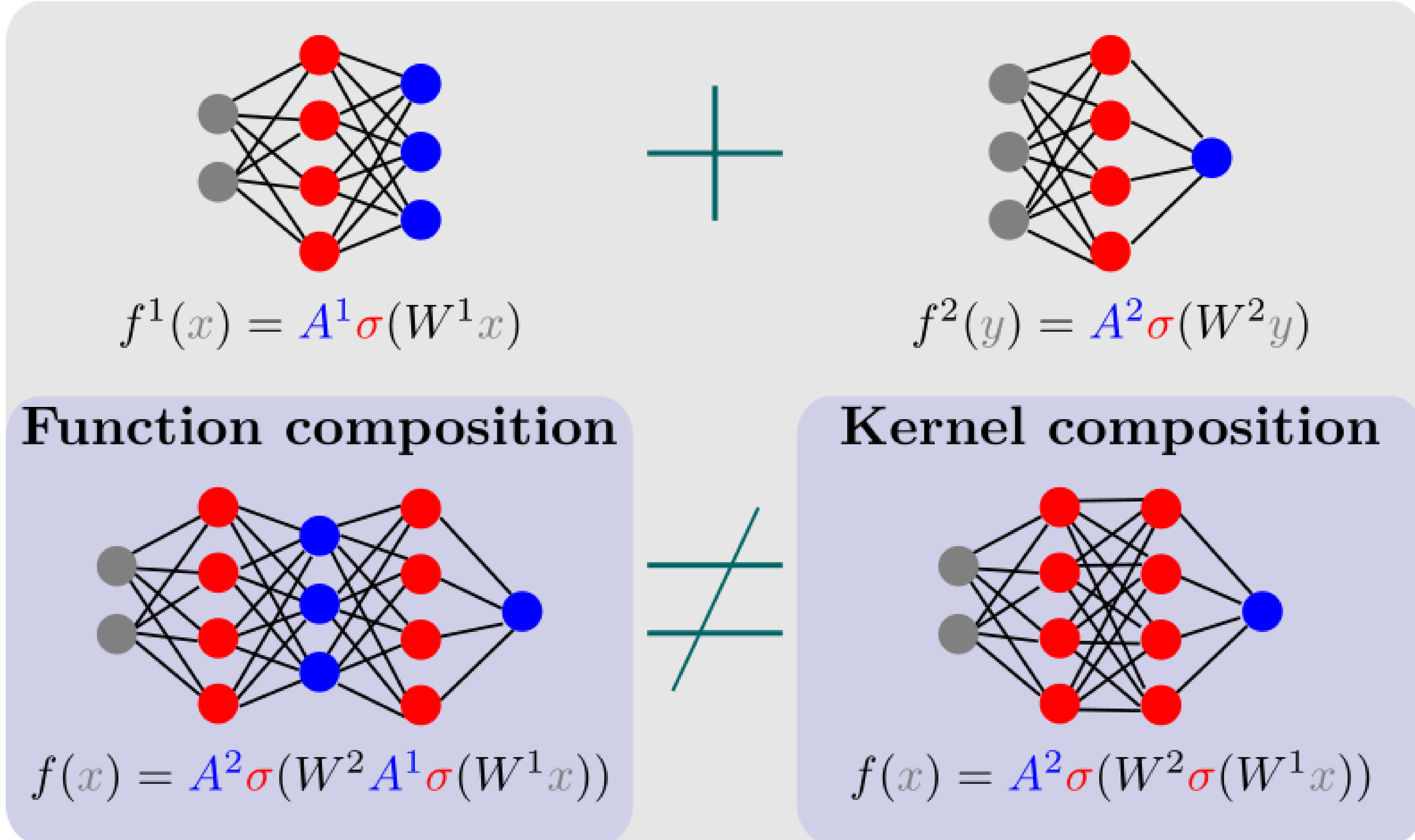
Candidate spaces fall into three categories

- Neural Tree spaces
- Hierarchical Hilbert spaces
- Bottlenecked spaces

No kernel structure

Lack of norm

Rank constraints



RKBS

A **Reproducing Kernel Banach space (RKBS)** $\mathcal{B}$ is Banach space of functions over X with bounded function evaluations, i.e.

$$f(x) = 0 \quad \forall x \in X \implies f = 0$$
$$\forall x \in X\ \exists C_x > 0: \quad |f(x)| \le C_x \|f\|_{\mathcal{B}} \qquad \forall f \in \mathcal{B}$$

Every pair of RKBS defines a kernel. If

- $\mathcal{B}$ is a RKBS over X
- $\mathcal{B}^\circ$ is a RKBS over Ω
- $K: X \times \Omega \longrightarrow \mathbb{R}$

and they satisfy

$$f(x) = \langle K(x,\cdot)|f\rangle \qquad \forall f \in \mathcal{B}, x \in X$$
$$g(w) = \langle g|K(\cdot,w)\rangle \qquad \forall g \in \mathcal{B}^\circ, w \in \Omega$$

then K is the **Reproducing kernel** of the **RKBS pair** $\mathcal{B}, \mathcal{B}^\circ$.

Chain RKBS

If

- $\mathcal{B}^1, \mathcal{B}^{1\circ}$ is a RKBS pair over X, Ω^1 with kernel K^1
- $\tilde{\mathcal{B}}^2, \tilde{\mathcal{B}}^{2\circ}$ is a RKBS pair over $\mathcal{B}^{1\circ}, \Omega^2$ with kernel K^1

and

$$Ah(x) = \langle \tilde{K}^2(K^1(x,\cdot))|h\rangle$$

then

$$\mathcal{B}^2 = \{Ah \mid h \in \tilde{\mathcal{B}}^2\}$$
$$\mathcal{B}^{2\circ} = \{q \in \tilde{\mathcal{B}}^{2\circ} \mid \langle q|h\rangle = 0 \quad \forall h \in \mathcal{N}(A)\}$$

is the **chain RKBS (cRKBS)** formed using the **link RKBS** pair $\tilde{\mathcal{B}}^2, \tilde{\mathcal{B}}^{2\circ}$ and the **initial RKBS** $\mathcal{B}^1, \mathcal{B}^{1\circ}$.

Theorem (Consistency)

$\mathcal{B}^2, \mathcal{B}^{2\diamond}$ is a RKBS pair over X, Ω^2 with kernel $\tilde{K}^2(K^1(x,\cdot))$.

Neural RKBS

Primal $\mathcal{B}$

$$f(x) = \int_\Omega \sigma(\langle v|x\rangle + b)d\mu(v,b)$$
$$\|f\|_{\mathcal{B}} = \inf_\mu \|\mu\|$$

Dual $\mathcal{B}^\diamond$

$$g(v,b) = \int_X \sigma(\langle v|x\rangle + b)d\rho(x)$$
$$\|g\|_{\mathcal{B}^\diamond} = \sup_{v,b} |g(v,b)|$$

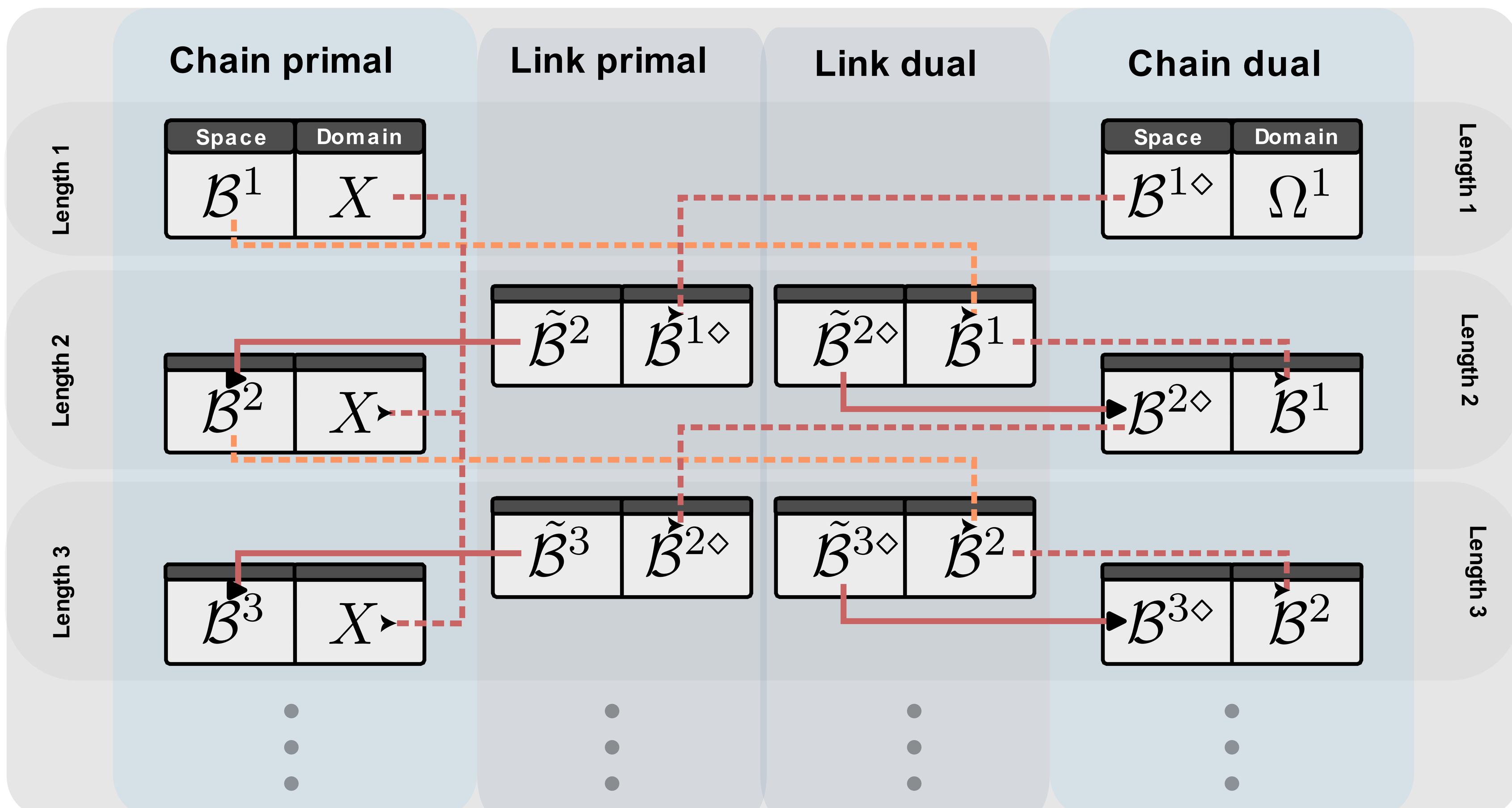
$$\langle f|g\rangle = \int_X \int_\Omega \sigma(\langle v|x\rangle + b)d\mu(v,b)d\rho(x)$$

Neural Chain RKBS

Neural cRKBS $\mathcal{B}^\ell, \mathcal{B}^{\ell\circ}$ have functions of the form

$$f^\ell(x) = \int_{\mathcal{B}^\ell} \sigma(f^{\ell-1}(x) + b)d\mu(f^{\ell-1}, b)$$

$$g^\ell(f^{\ell-1}, b) = \int_X \sigma(f^{\ell-1}(x) + b)d\rho(x)$$



Lemma

Let $\mathcal{B}, \mathcal{B}^\diamond$ be an adjoint pair of RKBS with domains X, Ω and kernel K. If $X = \{x_1, \ldots, x_N\}$, then $\mathcal{B}^\diamond = \mathrm{span}\{K_{x_n} \mid n = 1, \ldots, N\}$ and there exists N elements $w_j \in \Omega$ such that $\{K_{w_j} \mid j = 1, \ldots N\}$ forms a basis of $\mathcal{B}$, where $N = \dim \mathcal{B}^\diamond$.

Theorem (Deep neural networks are neural cRKBS)

If f is neural network of depth L, then $f \in \mathcal{B}^L$ with $\mathcal{B}^L$ a neural cKRBS of length L.

Theorem (Finite neural cRKBS are Deep neural networks)

If $X = \{x_1, \ldots, x_N\}$, then every function f in the neural cRKBS $\mathcal{B}^L$ of length L can be represented by a deep neural network with at most N neurons per layer. Moreover, all function share the same hidden layers.

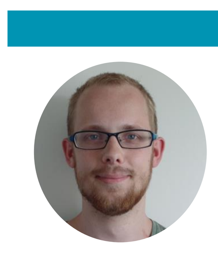
Conclusion

The answer to the question is

Deep neural networks are neural cRKBS

Desirable properties:

- Based on kernel spaces ✓
- Is consistent ✓
- Works for all common activations functions ✓
- Duality between data and weights ✓



Contact

Tjeerd Jan Heeringa

t.j.heeringa@utwente.nl

UNIVERSITY OF TWENTE.