A Bregman Proximal Viewpoint on Neural Operators

A. Mezidi[1,*] J. Patracone[1,*] S. Salzo[2,*] A. Habrard[1,3]
M. Pontil[4,5] R. Emonet[1,3] M. Sebban[1]

[1] Université Jean Monnet, Hubert Curien Lab, France
[2] DIAG, Sapienza University of Rome, Italy
[3] Institut Universitaire de France (IUF)
[4] Computational Statistics and Machine Learning, IIT, Genova, Italy
[5] Department of Computer Science, UCL, London, United Kingdom
[*] Equal contribution

Context: Neural Operators

Initial Velocity | Predicted Velocity

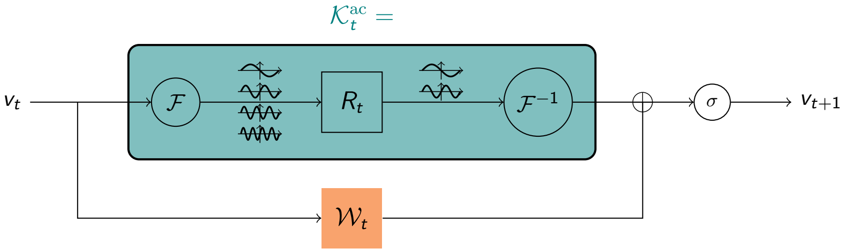
$a \in \mathcal{A} = L^p(D, \mathbb{R}^n)$ | $u \in \mathcal{U} = L^p(D, \mathbb{R}^n)$

Learn mappings between function spaces

Each operator layer is traditionally defined as:

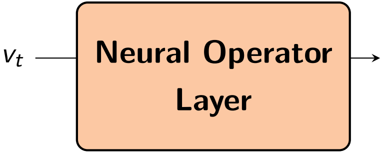
$$v_{t+1} = \sigma(\mathcal{K}_t^{\mathrm{ac}}(v_t) + \mathcal{W}_t) \quad \text{where} \quad \mathcal{W}_t = W_t v_t + b_t$$

Example: Fourier Neural Operator Layer



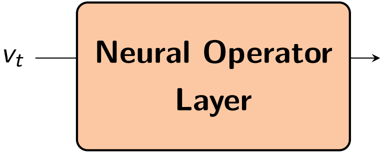
where:

- $\mathcal{K}_t^{\mathrm{ac}}$: integral kernel operator
- σ: pointwise activation operator
- $\mathcal{W}_t, R_t$: linear operators



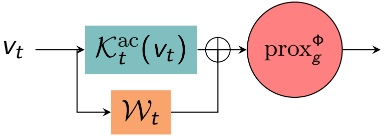
$$v_t \longrightarrow v_{t+1} = \mathcal{F}(v_t, \mathcal{K}_t^{\mathrm{ac}}, \mathcal{W}_t, \sigma)$$

Bregman Proximal Viewpoint



$$v_{t+1} = \operatorname*{argmin}_{w \in \mathcal{V}} \left\{ -\langle w, \mathcal{K}_t^{\mathrm{ac}}(v_t) + \mathcal{W}_t \rangle + g(w) + D_\Phi(w, v_t) \right\}$$

Bregman Proximal Viewpoint

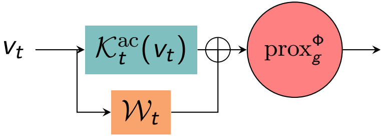


$$v_{t+1} = \underset{w \in \mathcal{V}}{\mathrm{argmin}} \left\{ -\langle w, \mathcal{K}_t^{\mathrm{ac}}(v_t) + \mathcal{W}_t \rangle + g(w) + D_{\Phi}(w, v_t) \right\}$$
$$= \mathrm{prox}_g^{\Phi}\left(\nabla\Phi(v_t) + \mathcal{K}_t^{\mathrm{ac}}(v_t) + \mathcal{W}_t\right)$$

Definition (Bregman Proximity Operator)

Let $g \in \Gamma_0(\mathcal{V})$ and Φ be a convex integral functional. The *Bregman proximity operator* is:

$$\mathrm{prox}_g^{\Phi} : \mathcal{V}^* \to \mathcal{V}, \quad v^* \mapsto \mathrm{argmin}\left\{\langle \cdot\,, -v^* \rangle + \Phi + g\right\}$$



$$= \underset{w \in \mathcal{V}}{\operatorname{argmin}} \left\{ -\langle w, \mathcal{K}_t^{\mathrm{ac}}(v_t) + \mathcal{W}_t \rangle + g(w) + D_\Phi(w, v_t) \right\}$$
$$= \mathrm{prox}_g^{\Phi}\left(\nabla \Phi(v_t) + \mathcal{K}_t^{\mathrm{ac}}(v_t) + \mathcal{W}_t\right)$$

Definition (Bregman Proximity Operator)

Let $g \in \Gamma_0(\mathcal{V})$ and Φ be a convex integral functional. The *Bregman proximity operator* is:

$$\mathrm{prox}_g^{\Phi} : \mathcal{V}^* \to \mathcal{V}, \quad v^* \mapsto \operatorname{argmin} \left\{ \langle \cdot, -v^* \rangle + \Phi + g \right\}$$

How to link prox_g^{Φ} to activation functions ?

Activation Functions as Proximity Operators

Prior Works

Combettes & Pesquet (2020)

$$\mathrm{prox}^{\frac{1}{2}\|\cdot\|^2}_{\Psi - \frac{1}{2}\|\cdot\|^2} = \sigma$$

- *Euclidean distance*
- *Finite dimension*

Activation Functions as Proximity Operators

Prior Works

Combettes & Pesquet (2020)

$$\operatorname{prox}^{\frac{1}{2}\|\cdot\|^2}_{\Psi - \frac{1}{2}\|\cdot\|^2} = \sigma$$

- *Euclidean distance*
- *Finite dimension*

Frecon et al. (2022)

$$\operatorname{prox}^{\Psi}_{0} = \sigma$$

- *Bregman distance*
- *Finite dimension*

Activation Functions as Proximity Operators

Prior Works

Combettes & Pesquet (2020)

$$\mathrm{prox}_{\Psi - \frac{1}{2}\|\cdot\|^2}^{\frac{1}{2}\|\cdot\|^2} = \sigma$$

- *Euclidean distance*
- *Finite dimension*

Frecon et al. (2022)

$$\mathrm{prox}_{0}^{\Psi} = \sigma$$

- *Bregman distance*
- *Finite dimension*

Our Contribution

$$\mathrm{prox}_{0}^{\Psi} = \sigma \quad \textbf{and} \quad \mathrm{prox}_{\Psi - \frac{1}{2}\|\cdot\|^2}^{\frac{1}{2}\|\cdot\|^2} = \sigma$$

- **Euclidean and Bregman distances**
- **Extended to function spaces**

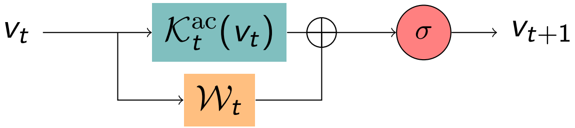
Classical Neural Operators

Proposition: Recovering Classical Neural Operator

Let $\mathcal{V} = L^2(D, \mathbb{R}^n)$ and D_Φ be the **Euclidean distance**:
Then:

$$v_{t+1} = \mathrm{prox}^{\frac{1}{2}\|\cdot\|^2}_{\Psi - \frac{1}{2}\|\cdot\|^2}\left(\mathcal{K}_t^{\mathrm{ac}}(v) + \mathcal{W}_t v\right) = \sigma\left(\mathcal{K}_t^{\mathrm{ac}}(v) + \mathcal{W}_t\right)$$

$\Rightarrow$ **Neural Operator Layer**

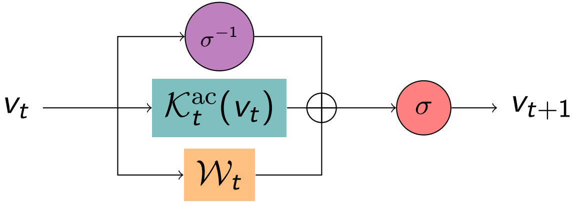


Proposition: Designing Bregman Neural Operators

Let $\mathcal{V} = L^p(D, \mathbb{R}^n)$ and D_Φ be the **Bregman distance**:
Then:

$$v_{t+1} = \mathrm{prox}_0^{\Psi}\left(\nabla\Psi_t(v) + \mathcal{K}_t^{\mathrm{ac}}(v) + \mathcal{W}_t\right)$$
$$= \sigma(\sigma^{-1}(v) + \mathcal{K}_t^{\mathrm{ac}}(v) + \mathcal{W}_t)$$

$\Rightarrow$ **Bregman Neural Operator Layer**



Novel architecture: Adds a "skip" branch with the inverse activation σ^{-1}

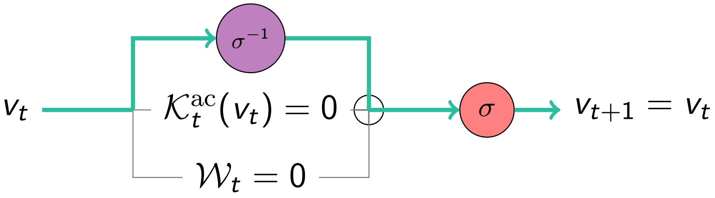
Bregman Neural Operators

Proposition: Designing Bregman Neural Operators

Let $\mathcal{V} = L^p(D, \mathbb{R}^n)$ and D_Φ be the **Bregman distance**:
Then:

$$v_{t+1} = \mathrm{prox}_0^{\Psi}\left(\nabla\Psi_t(v) + \mathcal{K}_t^{\mathrm{ac}}(v) + \mathcal{W}_t\right)$$
$$= \sigma(\sigma^{-1}(v) + \mathcal{K}_t^{\mathrm{ac}}(v) + \mathcal{W}_t)$$

$\Rightarrow$ **Bregman Neural Operator Layer**



When $\mathcal{K}_t^{\mathrm{ac}} = 0$ and $\mathcal{W}_t = 0$: $v_{t+1} = \sigma(\sigma^{-1}(v_t)) = v_t$ **Layer reduces to identity**

Universal Approximation Result

Theorem (Universal Approximation for Bregman Neural Operators)

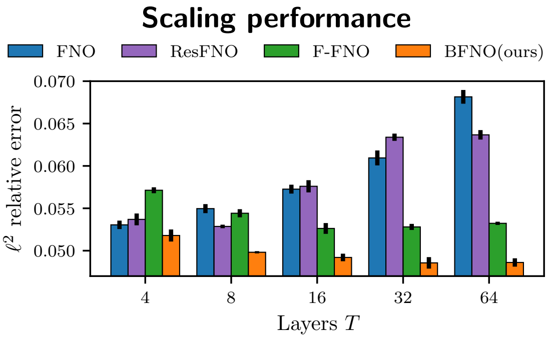
Let σ be sigmoidal. For any compact $K \subset \mathcal{A}$ and $\varepsilon > 0$, there exists a Bregman neural operator $\mathcal{N}_\theta$ such that:

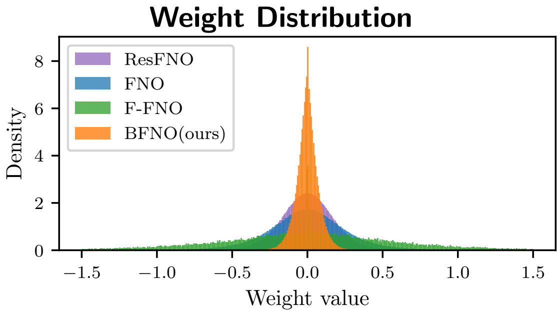
$$\sup_{u \in K} \|\mathcal{G}(u) - \mathcal{N}_\theta(u)\|_{\mathcal{U}} \leq \varepsilon$$

Significance

Bregman neural operators have the same expressivity as classical neural operators

Experimental Results





- **Bregman Fourier Neural Operator** demonstrates superior scaling with depth compared to other models.
- **BFNO** exhibits a Laplace-like weight distribution, indicating implicit regularization.

Summary: A Bregman Proximal Viewpoint on Neural Operators

- **Theoretical Framework**: Neural operator layers as solutions of optimization problems
- **Unification**: Classical neural operators as special case with Euclidean distance
- **Novel Architecture**: **Bregman Neural Operators** with added σ^{-1} skip connection
- **Theoretical Guarantees**: Universal approximation results
- **Empirical Results**: Better depth scaling on PDE tasks and implicit regularization

Impact: Opens new research directions at intersection of optimization and neural operators

Thank you!

Questions?