Appendix

A Implementation Details

Training details. We train all the networks for 500 epochs with Adam optimizer. The learning rate is set to 0.001 for Darcy flow and 0.005 for Navier-Stokes. We use learning rate weight decay of 1e-4 for both Navier-Stokes and Darcy flow. The batch size is set to 32. In case of Darcy flow, we also use cosine annealing for learning rate scheduling. We run all our experiments on a combination of NVIDIA RTX A6000, NVIDIA GeForce RTX 2080 Ti and 3080 Ti. All networks can easily fit on a single NVIDIA RTX A6000, but training time varies between the networks.

For FNO-DEQ, we use Anderson solver [Anderson, 1965] to solve for the fixed point in the forward pass. The maximum number of Anderson solver steps is kept fixed at 32 for Dary Flow, and 16 for Navier Stokes. For the backward pass, we use phantom gradients [Geng et al., 2021] which are computed as:

$$u^\star = \tau G_\theta(u^\star, a) + (1-\tau)u^\star \tag{12}$$

where τ is a tunable damping factor and $u^\star$ is the fixed point computed using Anderson solver in the forward pass. This step can be repeated S times. We use $\tau = 0.5$ and $S = 1$ for Darcy Flow, and $\tau = 0.8$ and $S = 3$ for Navier-Stokes.

For the S-FNO-DEQ used in Table 1, we use Broyden's method [Broyden, 1965] to solve for the fixed point in the forward pass and use exact implicit gradients, computed through implicit function theorem as shown in Eq. (6), for the backward pass through DEQ. The maximum number of solver steps is fixed at 32.

For weight-tied networks, we repeatedly apply the FNO block to the input 12 times for Darcy flow, and 6 times for Navier-Stokes.

Network architecture details. The width of an FNO layer set to 32 across all the networks. Additionally, we retain only 12 Fourier modes in FNO layer, and truncate higher Fourier modes. We use the code provided by Li et al. [2020a] to replicate the results for FNO, and construct rest of the networks on top of this as described in Sec. 5.

B Datasets

B.1 Darcy Flow

As mentioned in Sec. 5 we use the dataset provided by Li et al. [2020a] for our experiments with steady-state Darcy-Flow.

All the models are trained on 1024 data samples and tested on 500 samples. The resolution of original images is 421×421 which we downsample to 85×85 for our experiments. For experiments with noisy inputs/observations, the variance of Gaussian noise that we add to PDEs are [0, 1e-9, 1e-8, 1e-7, 1e-6, 1e-5, 1e-4, 1e-3].

B.2 Steady-State Incompressible Fluid Navier-Stoke

$$\begin{aligned} u \cdot \nabla \omega &= \nu \Delta \omega + f, && x \in \Omega \\ \nabla \cdot u &= 0 && x \in \Omega \end{aligned}$$

To generate the dataset for steady-state Navier-Stokes, instead of solving the steady state PDE using steady-state solvers like the SIMPLE algorithm Patankar and Spalding [1983], we first choose the solution $\omega^\star := \nabla \times u^\star$ of the PDE and then generate the corresponding equation, i.e. calculate the corresponding force term $f = u^\star \cdot \nabla \omega^\star - \nu \Delta \omega^\star$.

To generate the solutions $\omega^\star$, we forward propagate a relatively simple initial distribution of ω_0 (sampled from a Gaussian random field) through a time-dependent Navier-Stokes equation in the vorticity form for a short period of time. This ensures our dataset contains solutions ω^* that are rich

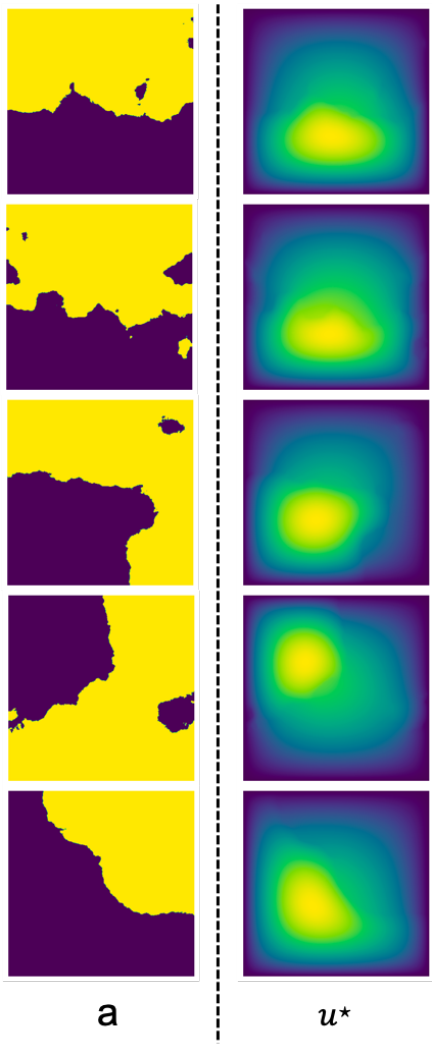


Figure 1: Samples from Darcy Flow

and complex. Precisely, recall the Navier-Stokes equations in their vorticity form:

$$\begin{aligned} \partial_t \omega(x,t) + u(x,t)\cdot\nabla\omega(x,t) &= \nu\Delta\omega(x,t) + g(x) \qquad x \in (0,2\pi)^2, t \in [0,T] \\ \nabla\cdot u(x,t) &= 0 \qquad x \in (0,2\pi)^2, t \in [0,T] \\ \omega(x,0) &= \omega_0(x) \qquad x \in (0,2\pi)^2 \end{aligned} \tag{13}$$

where $g(x) = \nabla \times \tilde{g}(x)$ and $\tilde{g}(x) = \sin(5x_1)\hat{x_2}$ is a divergence free forcing term and $x = (x_1, x_2)$ are the two coordinates of the input vector. We forward propagate the equations Equation 13 using a pseudo-spectral method using the functions provided in JAX-CFD [Kochkov et al., 2021, Dresdner et al., 2022] package. The initial vorticity ω_0 is sampled from a Gaussian random field $\mathcal{N}(0, (5^{3/2}(I + 25\Delta)^{-2.5}))$, which is then made divergence free. We forward propagate the Navier-Stokes equation in Equation 13 for time $T = 0.5$ with $dt = 0.002$ to get $\omega(1, x)$, which we choose as the solution to the steady-state PDE in Equation 10, i.e, $\omega^\star$ for Equation 10.

Subsequently, we use the stream function Ψ [Batchelor and Batchelor, 1967] to calculate $u = (\partial\Psi/\partial x_1, \partial\Psi/\partial x_2)$ by solving the Poisson equation $\Delta\Psi = \omega$ in the Fourier domain. Furthermore, since $f = u^\star \cdot \nabla\omega^\star - \nu\Delta\omega^\star$ we use the stream function to calculate (f_1, f_2), i.e., the different components of the force term.

We use 4500 training samples and 500 testing samples. The input to the network is the vector field $\tilde{f} = (f_1, f_2)$ and we learn a map that outputs the vorticity $\omega^\star$. The resolution of grid used to

generate the dataset is 256×256 which we downsample to 128×128 while training the models. For experiments with noisy inputs/observations, we consider two values of maximum variance of Gaussian noise: 1e-3 and 4e-3. The variances of the Gaussian noise that we add to the PDEs for the latter case are [0, 1e-9, 1e-8, 1e-7, 1e-6, 1e-5, 1e-4, 1e-3, 2e-3, 4e-3]. However, when conducting experiments with a variance of 1e-3, we exclude the last two values of variance from this list.

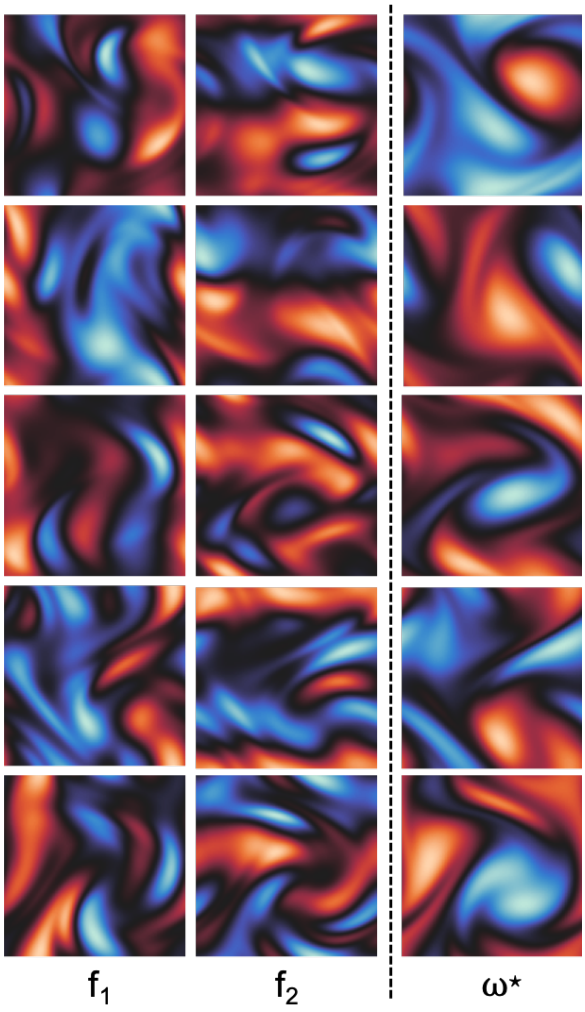


Figure 2: Samples from Steady-state Navier-Stokes dataset with viscosity 0.001. Each triplet visualizes the inputs f_1, f_2 and the ground truth output i.e. $\omega^\star$.

C Proof of Universal Approximation

The proof of the universal approximation essentially follows from the result on the universal approximation capabilities of FNO layers in Kovachki et al. [2021a], applied to $\mathcal{G}(v, f) = v - (Lv - f)$. For the sake of completeness, we reitarate the key steps.

For simplicity, we will assume that $d_u = d_v = d_f = 1$. (The results straightforwardly generalize.) We will first establish some key technical lemmas and introduce some notation and definitions useful for the proof for Theorem 1.

Definition 7. *An operator $T : L^2(\Omega; \mathbb{R}) \to L^2(\Omega; \mathbb{R})$ is continuous at $u \in L^2(\Omega; \mathbb{R})$ if for every $\epsilon > 0$, there exists a $\delta > 0$, such that for all $v \in L^2(\Omega)$ with $\|u - v\|_{L^2(\Omega)} \leq \delta$, we have $\|L(u) - L(v)\|_{L^2(\Omega)} \leq \epsilon$.*

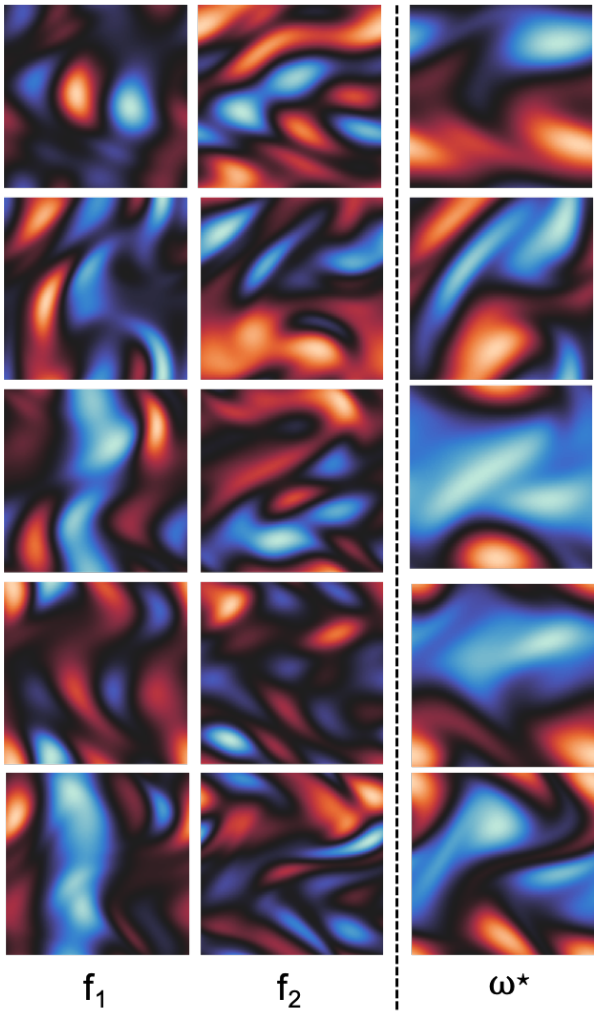


Figure 3: Samples from Steady-state Navier-Stokes dataset with viscosity 0.01. Each triplet visualizes the inputs f_1, f_2 and the ground truth output i.e. $\omega^\star$.

First, we approximate the infinite dimensional operator $\mathcal{G} : L^2(\Omega) \times L^2(\Omega) \to L^2(\Omega)$ by projecting the functions in $L^2(\Omega)$ to a finite-dimensional approximation $L^2_N(\Omega)$, and considering the action of the operator on this subspace. The linear projection we use is the one introduced in Equation 11. More precisely we show the following result,

Lemma 2. *Given a continuous operator $L : L^2(\Omega) \to L^2(\Omega)$ as defined in Equation 1, let us define an operator $\mathcal{G} : L^2(\Omega) \times L^2(\Omega) \to L^2(\Omega)$ as $\mathcal{G}(v, f) := v - (L(v) - f)$. Then, for every $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all v, f in any compact set $K \subset L^2(\Omega)$, the operator $\mathcal{G}_N = \Pi_N \mathcal{G}(\Pi_N v, \Pi_N f)$ is an ϵ-approximation of $\mathcal{G}(v, f)$, i.e., we have,*

$$\sup_{v,f \in K} \|\mathcal{G}(v, f) - \mathcal{G}_N(v, f)\|_{L^2(\Omega)} \leq \epsilon.$$

Proof. Note that for an $\epsilon > 0$ there exists an $N = N(\epsilon, d)$ such that for all $v \in K$ we have

$$\sup_{v \in K} \|v - \Pi_N v\|_{L^2(\Omega)} \leq \epsilon.$$

Therefore, using the definition of $\mathcal{G}_N$ we can bound the $L^2(\Omega)$ norm of the difference between $\mathcal{G}$ and $\mathcal{G}_N$ as follows,

$$\begin{aligned}
&\|\mathcal{G}(v,f) - \Pi_N\mathcal{G}(v_n,f_n)\|_{L^2(\Omega)} \\
&\leq \|\mathcal{G}(v,f) - \Pi_N\mathcal{G}(v,f)\|_{L^2(\Omega)} + \|\Pi_N\mathcal{G}(v,f) - \Pi_N\mathcal{G}(\Pi_N v,\Pi_N f)\|_{L^2(\Omega)} \\
&\leq \underbrace{\|\mathcal{G}(v,f) - \Pi_N\mathcal{G}(v,f)\|_{L^2(\Omega)}}_{I} + \underbrace{\|\mathcal{G}(v,f) - \mathcal{G}(\Pi_N v,\Pi_N f)\|_{L^2(\Omega)}}_{II}
\end{aligned}$$

We first bound the term I as follows:

$$\begin{aligned}
&\|\mathcal{G}(v,f) - \Pi_N\mathcal{G}(v,f)\|_{L^2(\Omega)} \\
&= \|v - (L(v) - f) - \Pi_N\left(v - (L(v) - f)\right)\|_{L^2(\Omega)} \\
&= \|v - \Pi_N v\|_{L^2(\Omega)} + \|f - \Pi_N f\|_{L^2(\Omega)} + \|L(v) - \Pi_N L(v)\|_{L^2(\Omega)} \\
&= \epsilon + \epsilon + \|L(v) - \Pi_N L(v)\|_{L^2(\Omega)} \qquad (14)
\end{aligned}$$

Since L is continuous, for all compact sets $K \subset L^2(\Omega)$, $L(K)$ is compact as well. This is because: (1) for any $u \in K$, $\|L(u)\|_{L^2(\Omega)}$ is finite; (2) for any $v \in K$, $\|L(v)\|_{L^2(\Omega)} \leq \|L(u)\|_{L^2(\Omega)} + C\|u - v\|_{L^2(\Omega)}$. Therefore, for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$\sup_{v\in K} \|L(v) - \Pi_N L(v)\|_{L^2(\Omega)} \leq \epsilon.$$

Substituting the above result in Equation 14, we have

$$\|\mathcal{G}(v,f) - \Pi_N\mathcal{G}(v,f)\|_{L^2(\Omega)} \leq 3\epsilon. \qquad (15)$$

Similarly, for all $v \in K$ where K is compact, we can bound Term II as following,

$$\begin{aligned}
&\|\mathcal{G}(v,f) - \mathcal{G}(\Pi_N v,\Pi_N f)\|_{L^2(\Omega)} \\
&\leq \|v - (L(v) - f) - \Pi_N v - (L(\Pi_N v) - \Pi_N f)\|_{L^2(\Omega)} \\
&\leq \|v - \Pi_N v\|_{L^2(\Omega)} + \|f - \Pi_N f\|_{L^2(\Omega)} + \|L(v) - L(\Pi_N v)\|_{L^2(\Omega)} \\
&\leq \epsilon + \epsilon + \|L(v) - L(\Pi_N v)\|_{L^2(\Omega)}. \qquad (16)
\end{aligned}$$

Now, since $v \in K$ and $L : L^2(\Omega) \to L^2(\Omega)$ is a continuous operator, there exists a modulus of continuity (an increasing real valued function) $\alpha \in [0,\infty)$, such that for all $v \in K$, we have

$$\|L(v) - L(\Pi_N v)\|_{L^2(\Omega)} \leq \alpha\left(\|v - \Pi_N v\|_{L^2(\Omega)}\right)$$

Hence for every $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that,

$$\alpha(\|v - \Pi_N v\|_{L^2(\Omega)}) \leq \epsilon.$$

Plugging these bounds in Equation 16, we get,

$$\|\mathcal{G}(v,f) - \mathcal{G}(\Pi_N v,\Pi_N f)\|_{L^2(\Omega)} \leq 3\epsilon. \qquad (17)$$

Therefore, combining Equation 15 and Equation 17 then for $\epsilon > 0$, there exists an $N \in \mathbb{N}$, such that for all $v, f \in K$ we have

$$\sup_{v,f\in K} \|\mathcal{G}(v,f) - \Pi_N\mathcal{G}(v_n,f_n)\|_{L^2(\Omega)} \leq 6\epsilon. \qquad (18)$$

Taking $\epsilon' = 6\epsilon$ proves the claim. □

Proof of Theorem 1. For Lemma 2 we know that there exists a finite dimensional projection for the operator $\mathcal{G}$, defined as $\mathcal{G}_N(v,f)$ such that for all $v, f \in L^2(\Omega)$ we have

$$\|\mathcal{G}(v,f) - \mathcal{G}_N(v,f)\|_{L^2(\Omega)} \leq \epsilon.$$

Now using the definition of $\mathcal{G}_N(v,f)$ we have

$$\begin{aligned}
\mathcal{G}_N(v,f) &= \Pi_N\mathcal{G}(\Pi_N v,\Pi_N f) \\
&= \Pi_N v - (\Pi_N L(\Pi_N v) - \Pi_N f)
\end{aligned}$$

From Kovachki et al. [2021a], Theorem 2.4 we know that there exists an FNO network G_{θ^L} of the form defined in Equation 2 such that for all $v \in K$, where K is a compact set, there exists an ϵ^L we have

$$\sup_{v \in K} \|\Pi_N L(\Pi_N v) - G_{\theta^L}\|_{L^2(\Omega)} \leq \epsilon^L \tag{19}$$

Finally, note that from Lemma D.1 in Kovachki et al. [2021a], we have that for any $v \in K$, there exists an FNO layers $G_{\theta^f} \in L^2(\Omega)$ and $G_{\theta^v} \in L^2(\Omega)$ defined in Equation 3 such that

$$\sup_{v \in K} \|\Pi_N v - G_{\theta^v}\|_{L^2(\Omega)} \leq \epsilon^v \tag{20}$$

and

$$\sup_{f \in K} \|\Pi_N f - G_{\theta^f}\|_{L^2(\Omega)} \leq \epsilon^f \tag{21}$$

for $\epsilon^v > 0$ and $\epsilon^f > 0$.

Therefore there exists an $\tilde{\epsilon} >$ such that there is an FNO network $G_\theta : L^2(\Omega) \times L^2(\Omega) \to L^2(\Omega)$ where $\theta := \{\theta^L, \theta^v, \theta^f\}$ such that

$$\sup_{v \in K, f \in L^2(\Omega)} \|\mathcal{G}_N(v, f) - G_\theta(v, f)\|_{L^2(\Omega)} \leq \tilde{\epsilon} \tag{22}$$

Now, since we know that $u^\star$ is the fixed point of the operator $\mathcal{G}$ we have from Lemma 2 and Equation 22,

$$\begin{aligned}\|\mathcal{G}(u^\star, f) - G_\theta(u^\star, f)\|_{L^2(\Omega)} &\leq \|u^\star - \mathcal{G}_N(u^\star, f)\|_{L^2(\Omega)} + \|\mathcal{G}_N(u^\star, f) - G_\theta(u^\star, f)\|_{L^2(\Omega)} \\ &\leq \tilde{\epsilon} + \epsilon.\end{aligned}$$

□

D Fast Convergence for Newton Method

Definition 8 (Frechet Derivative in $L^2(\Omega)$)**.** *For a continuous operator $F : L^2(\Omega) \to L^2(\Omega)$, the Frechet derivative at $u \in L^2(\Omega)$ is a linear operator $F'(u) : L^2(\Omega) \to L^2(\Omega)$ such that for all $v \in L^2(\Omega)$ we have*

$$\lim_{\|v\|_{L^2(\Omega)} \to 0} \frac{\|F(u+v) - F(u) - F'(u)(v)\|_{L^2(\Omega)}}{\|v\|_{L^2(\Omega)}} = 0.$$

Lemma 3. *Given the operator $L : L^2(\Omega) \to L^2(\Omega)$ with Frechet derivative L', such that for all $u, v \in L^2(\Omega)$, we have $\|L'(u)(v)\|_{L^2(\Omega)} \geq \lambda \|v\|_{L^2(\Omega)}$, then $L'(u)^{-1}$ exists and we have, for all $v_1, v_2 \in L^2(\Omega)$:*

1. $\|L'(u)^{-1}(v_1)\|_{L^2(\Omega)} \leq \frac{1}{\lambda}\|v_1\|_{L^2(\Omega)}$.
2. $\|v_1 - v_2\|_{L^2(\Omega)} \leq \frac{1}{\lambda}\|L(v_1) - L(v_2)\|_{L^2(\Omega)}$

Proof. Note that for all $u, v' \in L^2(\Omega)$ we have,

$$\|L'(u)v'\|_{L^2(\Omega)} \geq \lambda \|v'\|_{L^2(\Omega)}$$

Taking $v = L'(u)^{-1}(v')$, we have

$$\begin{aligned}\|L'(u)\left(L'(u)^{-1}(v)\right)\|_{L^2(\Omega)} &\geq \lambda \|L^{-1}(u)(v)\|_{L^2(\Omega)} \\ \implies \frac{1}{\lambda}\|v\|_{L^2(\Omega)} &\geq \|L^{-1}(u)(v)\|_{L^2(\Omega)}.\end{aligned}$$

For part 2, note that there exists a $c \in [0, 1]$ such that

$$\|L(v_1) - L(v_2)\|_{L^2(\Omega)} \geq \inf_{c \in [0,1]} \|L'(cv_1 + (1-c)v_2)\|_2 \|v_1 - v_2\|_{L^2(\Omega)} \geq \lambda \|v_1 - v_2\|_{L^2(\Omega)}.$$

□

We now show the proof for Lemma 4. The proof is standard and can be found in Faragó and Karátson [2002], however we include the complete proof here for the sake of completeness.

We restate the Lemma here for the convenience of the reader.

Lemma 4 (Faragó and Karátson [2002], Chapter 5)**.** *Consider the PDE defined Definition 2, such that $d_u = d_v = d_f = 1$. such that $L'(u)$ defines the Frechet derivative of the operator L. If for all $u, v \in L^2(\Omega; \mathbb{R})$ we have $\|L'(u)v\|_{L^2(\Omega)} \geq \lambda \|v\|_{L^2(\Omega)}$ [7] and $\|L'(u) - L'(v)\|_{L^2(\Omega)} \leq \Lambda \|u - v\|_{L^2(\Omega)}$ for $0 < \lambda \leq \Lambda < \infty$, then for the Newton update, $u_{t+1} \leftarrow u_t - L'(u_t)^{-1}\left(L(u_t) - f\right)$, with $u_0 \in L^2(\Omega; \mathbb{R})$, there exists an $\epsilon > 0$, such that $\|u_T - u^\star\|_{L^2(\Omega)} \leq \epsilon$ if [8] $T \geq \log\left(\log\left(\frac{1}{\epsilon}\right) / \log\left(\frac{2\lambda^2}{\Lambda \|L(u_0) - f\|_{L^2(\Omega)}}\right)\right)$.*

Proof of Lemma 4. Re-writing the updates in Lemma 4 as,

$$u_{t+1} = u_t + p_t \tag{23}$$

$$L'(u_t)p_t = -(L(u_t) - f) \tag{24}$$

Now, upper bounding $L(u_{t+1}) - f$ for all $x \in \Omega$ we have,

$$\begin{aligned}
&L(u_{t+1}(x)) - f(x) \\
&= L(u_t(x)) - f(x) + \int_0^1 \left(L'(u_t(x) + t(u_{t+1}(x) - u_t(x)))\right)(u_{t+1}(x) - u_t(x))dt \\
&= L(u_t(x)) - f(x) + L'(u_t(x))p_t(x) + \int_0^1 \left(L'(u_t(x) + t(u_{t+1}(x) - u_t(x))) - L'(u_t(x))\right)p_t(x)dt \\
&= \int_0^1 \left(L'(u_t(x) + t(u_{t+1}(x) - u_t(x))) - L'(u_t(x))\right)p_t(x)dt
\end{aligned}$$

where we use Equation 24 in the final step.

Taking $L^2(\Omega)$ norm on both sides and using the fact that $\|L'(u) - L'(v)\|_{L^2(\Omega)} \leq \Lambda \|u - v\|_{L^2(\Omega)}$, we have

$$\|L(u_{t+1}) - f\|_{L^2(\Omega)} \leq \int_0^1 \Lambda t \|u_{t+1} - u_t\|_{L^2(\Omega)} \|p_t\|_{L^2(\Omega)} dt$$

Noting that for all $x \in \Omega$, we have $u_{t+1} - u_t = p_t$, and using the fact that for all u, v $\|L'(u)^{-1}v\|_{L^2(\Omega)} \leq \frac{1}{\lambda}\|v\|_{L^2(\Omega)}$ we have, $\|L'(u_t)p_t\|_{L^2(\Omega)} \leq \frac{1}{\lambda}\|p_t\|_{L^2(\Omega)}$

$$\begin{aligned}
\|L(u_{t+1}) - f\|_{L^2(\Omega)} &\leq \int_0^1 \Lambda t \|u_{t+1} - u\|_{L^2(\Omega)} \|p_t\|_{L^2(\Omega)} dt \\
&\leq \Lambda/2 \|p_t\|^2_{L^2(\Omega)} \\
&\leq \Lambda/2 \| - L'(u_t)^{-1}(L(u_t) - f)\|^2_{L^2(\Omega)} \\
&\leq \frac{\Lambda}{2\lambda^2} \|L(u_t) - f)\|^2_{L^2(\Omega)}
\end{aligned}$$

where we use the result from Lemma 3 in the last step.

Therefore we have

$$\begin{aligned}
&\|L(u_{t+1}) - f\|_{L^2(\Omega)} \leq \left(\frac{\Lambda}{2\lambda^2}\right)^{2^t - 1} (L(u_0) - f)^{2^t} \\
\implies &\|L(u_{t+1}) - f\|_{L^2(\Omega)} \leq \left(\frac{\Lambda}{2\lambda^2}\right)^{2^t - 1} (L(u_0) - L(u^\star))^{2^t} \\
\implies &\|u_{t+1} - u^\star\|_{L^2(\Omega)} \leq \frac{1}{\lambda}\left(\frac{\Lambda}{2\lambda^2}\right)^{2^t - 1} \|L(u_0) - L(u^\star)\|^{2^t}_{L^2(\Omega)}.
\end{aligned}$$

[7] We note that this condition is different from the condition on the inner-product in the submitted version of the paper, which had. $\langle L'(u), v\rangle_{L^2(\Omega)} \geq \lambda \|v\|_{L^2(\Omega)}$.

[8] We note that this rate is different from the one in the submitted version of the paper.

Therefore, if

$$\frac{\Lambda}{2\lambda^2}\|L(u_0) - L(u^\star)\|_{L^2(\Omega)} \leq 1,$$

then we have

$$\|u_{t+1} - u^\star\|_{L^2(\Omega)} \leq \epsilon,$$

for

$$T \geq \log\left(\log\left(\frac{1}{\epsilon}\right) / \log\left(\frac{2\lambda^2}{\Lambda\|L(u_0) - f\|_{L^2(\Omega)}}\right)\right).$$

□

E Additional experimental results

We provide additional results for Navier-Stokes equation for noisy inputs and observations in Table 4 and Table 5. For these experiments, the maximum variance of Gaussian noise added to inputs and observations is 0.004. We observe that weight-tied FNO and FNO-DEQ outperform non-weight-tied architectures.

Architecture	Parameters	#Blocks	Test error ↓		
			$\sigma^2_{\max} = 0$	$(\sigma^2_{\max})^i = 0.004$	$(\sigma^2_{\max})^t = 0.004$
FNO	2.37M	1	0.184 ± 0.002	0.238 ± 0.008	0.179 ± 0.004
FNO	4.15M	2	0.162 ± 0.024	0.196 ± 0.011	0.151 ± 0.010
FNO	7.71M	4	0.157 ± 0.012	0.216 ± 0.002	0.158 ± 0.009
FNO++	2.37M	1	0.199 ± 0.001	0.255 ± 0.002	0.197 ± 0.004
FNO++	4.15M	2	0.154 ± 0.005	0.188 ± 0.006	0.157 ± 0.006
FNO++	7.71M	4	0.151 ± 0.003	0.184 ± 0.008	0.147 ± 0.004
FNO-WT	2.37M	1	0.151 ± 0.007	0.183 ± 0.026	0.129 ± 0.018
FNO-DEQ	2.37M	1	$\mathbf{0.128 \pm 0.004}$	$\mathbf{0.159 \pm 0.005}$	$\mathbf{0.121 \pm 0.015}$

Table 4: **Results on incompressible Steady-State Navier-Stokes (viscosity=0.001)**: clean data (Col 4), noisy inputs (Col 5) and noisy observations (Col 6) with max variance of added noise being $(\sigma^2_{\max})^i$ and $(\sigma^2_{\max})^t$, respectively. Reported test error has been averaged on three different runs with seeds 0, 1, and 2.
‡ indicates that the network diverges during training for one of the seeds.

Architecture	Parameters	#Blocks	Test error ↓		
			$\sigma^2_{\max} = 0$	$(\sigma^2_{\max})^i = 0.004$	$(\sigma^2_{\max})^t = 0.004$
FNO	2.37M	1	0.181 ± 0.005	0.207 ± 0.003	0.178 ± 0.008
FNO	4.15M	2	0.138 ± 0.007	0.163 ± 0.003	0.137 ± 0.006
FNO	7.71M	4	0.152 ± 0.006	0.203 ± 0.055	0.151 ± 0.008
FNO++	2.37M	1	0.188 ± 0.002	0.217 ± 0.001	0.187 ± 0.005
FNO++	4.15M	2	0.139 ± 0.004	0.170 ± 0.005	0.138 ± 0.005
FNO++	7.71M	4	0.130 ± 0.005	0.168 ± 0.007	0.126 ± 0.007
FNO-WT	2.37M	1	0.099 ± 0.007	0.159 ± 0.029	0.123 ± 0.023
FNO-DEQ	2.37M	1	$\mathbf{0.088 \pm 0.006}$	$\mathbf{0.104 \pm 0.001}$	$\mathbf{0.116 \pm 0.005}$

Table 5: **Results on incompressible Steady-State Navier-Stokes (viscosity=0.01)**: clean data (Col 4), noisy inputs (Col 5) and noisy observations (Col 6) with max variance of added noise being $(\sigma^2_{\max})^i$ and $(\sigma^2_{\max})^t$, respectively. Reported test error has been averaged on three different runs with seeds 0, 1, and 2.
‡ indicates that the network diverges during training for one of the seeds.

Convergence analysis of fixed point. We report variations in test error, absolute residual $\|G_\theta(\mathbf{z}_t) - \mathbf{z}_t\|_2$, and relative residual $\frac{\|G_\theta(\mathbf{z}_t)-\mathbf{z}_t\|_2}{\|\mathbf{z}_t\|_2}$ with an increase in the number of solver steps while solving for the fixed point in FNO-DEQ, for both Darcy Flow (See Table 6) and Steady-State Navier Stokes (See Table 7). We observe that all these values decrease with increase in the number of fixed point solver iterations and eventually saturate once we have a reasonable estimate of the fixed point. We observe that increasing the number of fixed point solver iterations results in a better estimation of the fixed point. For steady state PDEs, we expect the test error to reduce as the estimation of the fixed point improves. Furthermore, at inference time we observe that the test error improves (i.e. reduces) with increase in the number of fixed point solver iterations even though the FNO-DEQ is trained with fewer solver steps. For Navier-Stokes with viscosity 0.01, at inference time we get a test MSE loss of 0.0744 with 48 solver steps from 0.0847 when used with 24 solver steps.

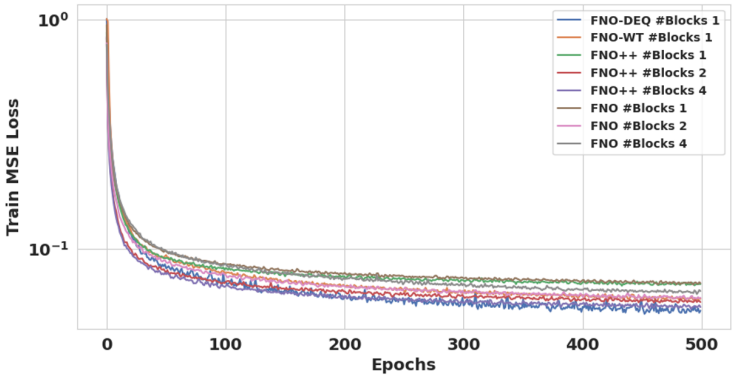
This further bolsters the benefits of DEQs (and weight-tied architectures in general) for training neural operators for steady-state PDEs. Moreover, performance saturates after a certain point once we have a reasonable estimate of the fixed point, hence showing that more solver steps stabilize to the same solution.

Solver steps	Absolute residual ↓	Relative residual ↓	Test Error ↓
2	212.86	0.8533	0.0777
4	18.166	0.0878	0.0269
8	0.3530	0.00166	0.00567
16	0.00239	1.13e-5	0.00566
32	0.000234	1.1e-6	0.00566

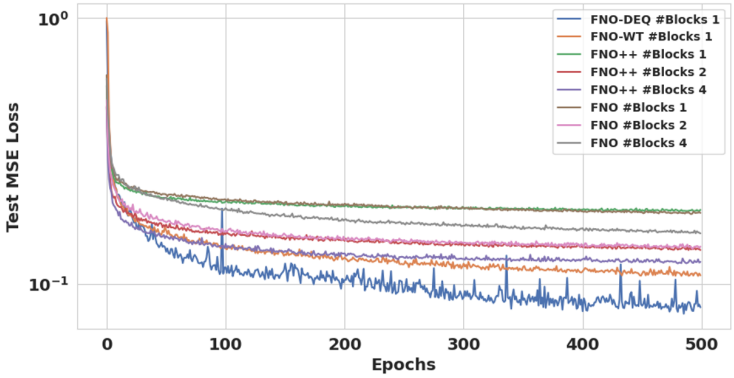
Table 6: Convergence analysis of fixed point for noiseless Darcy Flow: The test error, absolute residual $\|G_\theta(\mathbf{z}_t) - \mathbf{z}_t\|_2$ and relative residual $\frac{\|G_\theta(\mathbf{z}_t)-\mathbf{z}_t\|_2}{\|\mathbf{z}_t\|_2}$ decrease with increase in the number of fixed point solver iterations. The performance saturates after a certain point once we have a reasonable estimate of the fixed point. We consider the noiseless case, where we do not add any noise to inputs or targets.

Solver steps	Absolute residual ↓	Relative residual ↓	Test Error ↓
4	544.16	0.542	0.926
8	397.75	0.408	0.515
16	150.33	0.157	0.147
24	37.671	0.0396	0.0847
48	5.625	0.0059	0.0744
64	3.3	0.0034	0.0746

Table 7: Convergence analysis of fixed point for noiseless incompressible Steady-State Navier-Stokes with viscosity=0.01: The test error, absolute residual $\|G_\theta(\mathbf{z}_t) - \mathbf{z}_t\|_2$ and relative residual $\frac{\|G_\theta(\mathbf{z}_t)-\mathbf{z}_t\|_2}{\|\mathbf{z}_t\|_2}$ decrease with increase in the number of fixed point solver iterations. The performance saturates after a certain point once we have a reasonable estimate of the fixed point. We consider the noiseless case, where we do not add any noise to inputs or targets.



(a) Training Loss Curve



(b) Test Loss Curve

Figure 4: Training and Test Loss Curves for Steady-State Navier-Stokes with viscosity 0.01. The x axis is the number of epochs and y axis is the MSE loss in $\log$ scale. Note that while all the models converge to approximately the same MSE loss value while training, DEQs and weight-tied networks get a better test loss in fewer epochs.