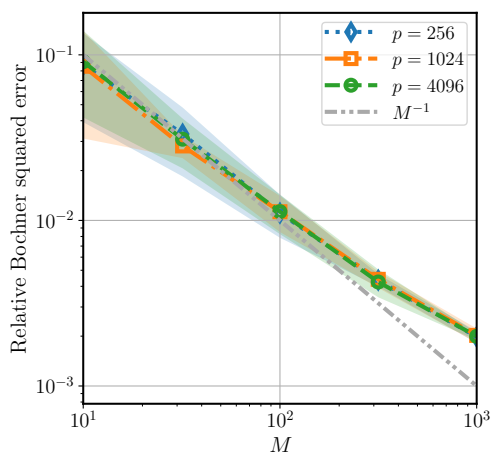
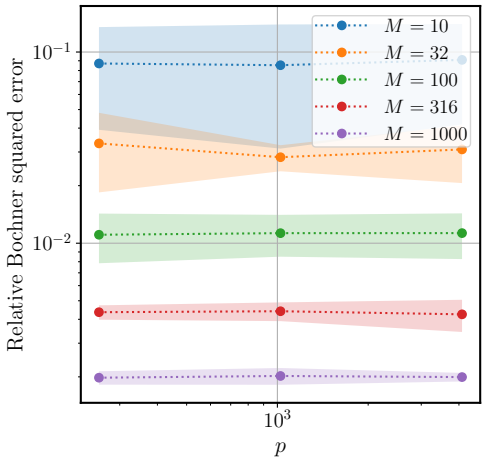
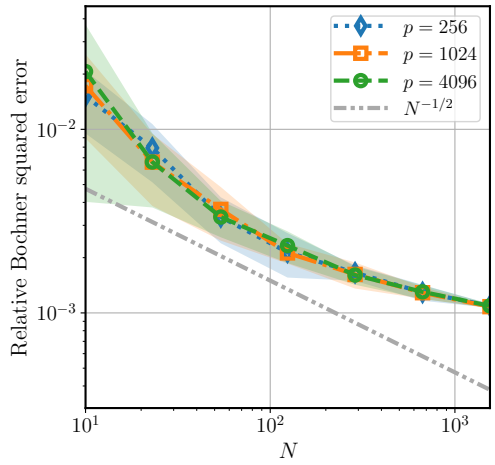


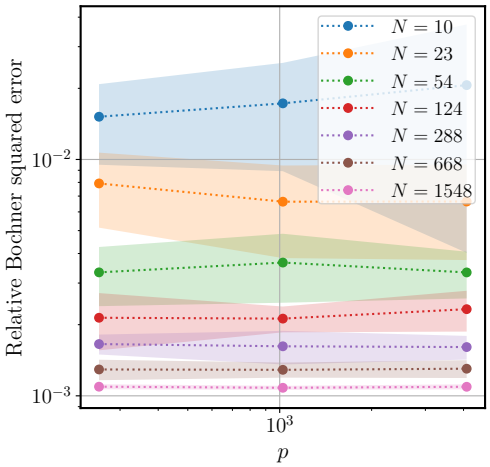
(a) Error versus number of random features M



(b) Error versus discretized output dimension p



(c) Error versus number of training data N



(d) Error versus discretized output dimension p

Figure 1: *RFM for Burgers' equation solution operator.* A RFM with M features (i.e., $\varphi(\cdot;\theta_m)$ with $\{\theta_m\}_{m=1,\dots,M} \sim \mu^{\otimes M}$) is trained on N input-output pairs $\{(u_n, \mathcal{G}(u_n))\}_{n=1,\dots,N}$ according to the RF ridge regression algorithm. The ground truth map $\mathcal{G}\colon L^2(\mathbb{T};\mathbb{R}) \to L^2(\mathbb{T};\mathbb{R})$ is a nonlinear operator defined by $u^{(0)}(\cdot) \mapsto u(\cdot,1)$, where $u = \{u(x,t)\}_{x,t}$ solves the partial differential equation

$$\begin{cases} \frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) &= 10^{-1}\frac{\partial^2 u}{\partial x^2} \quad \text{in} \quad \mathbb{T}\times(0,\infty) \\ u(\cdot,0) &= u^{(0)} \quad \text{in} \quad \mathbb{T}\,. \end{cases}$$

Here, $T \simeq (0,2\pi)_{\text{per}}$ is the 1D torus which comes with periodic boundary conditions. The initial conditions $u_n \sim \nu$ are sampled iid from a Matérn Gaussian process according to [KLLABSA23, Sec. 6.3, p. 32].

The random features are defined in a similar way to those in [NS21, Sec. 3.1, p. 15],

$$\varphi(u^{(0)};\theta) = \text{ELU}(\mathcal{F}^{-1}\{1_{(|k|\le k_{\max})}\chi_k \cdot (\mathcal{F}u^{(0)})_k \cdot (\mathcal{F}\theta)_k\}_{k\in\mathbb{Z}})\,,$$

where μ is also a Matérn Gaussian measure. Here $\mathcal{F}$ maps a function to its Fourier series coefficients, and $\mathcal{F}^{-1}$ expresses a Fourier coefficient sequence as a function expanded in the Fourier basis. The filter χ is given by [NS21, Eqn. 3.6]. The feature map φ lifts the notion of hidden neuron in neural network architectures to function space.

Figure 1a shows the decay of the relative average squared error (500 test pairs) as M increases, for fixed $N = 1548$. The regularization is chosen as $\lambda = 7\cdot 10^{-4}/M$. The error closely follows the rate $O(M^{-1})$ until it begins to saturate at larger M. This is either due to $\mathcal{G}$ not belonging to the RKHS of (φ,μ) or the finite data error dominating. As predicted by the theory, Figure 1b shows that the error does not depend on the discretized output dimension $p = \dim(\mathbb{R}^p)$, where $\mathbb{R}^p \approx \mathcal{X} = \mathcal{Y} = L^2(\mathbb{T};\mathbb{R})$ is discretized on a p-point grid. Figures 1c and 1d display similar behavior as N is varied (now with $\lambda = 3\cdot 10^{-6}/\sqrt{N}$ and fixed $M = 10^4$). All shaded bands denote two empirical standard deviations from the empirical mean of the error computed over 10 different models, each with iid choice of the features and training data indices.