SUPPLEMENTARY MATERIAL: SPECTRAL-BIAS AND KERNEL-TASK ALIGNMENT IN PHYSICALLY INFORMED NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

A GAUSSIAN PROCESS REGRESSION (GPR) FORMULA

Focusing on linear differential operators, and taking the prior on f to be a Gaussian process (GP) with kernel, K. As shown in Ref. Raissi et al. (2017), or in Ref. Pang & Karniadakis (2020) section 14.3.3 for Laplacian operator, the average of $f(\boldsymbol{x}_*)$ under the multivariate Gaussian distribution in Eqs. (3-4) of the main text is given by

$$\langle f_* \rangle_{P(f|X_n)} = \vec{k}^T \left(K_{\text{PINN}} + \tilde{\mathrm{I}} \right)^{-1} \boldsymbol{y}, \tag{1}$$

where

$$\boldsymbol{y} = (\boldsymbol{\phi}, \boldsymbol{g})^T, \tag{2}$$

the GP kernel is divided into blocks,

$$K = \begin{bmatrix} K(X_{n_\Omega}, X_{n_\Omega}) & K(X_{n_{\partial\Omega}}, X_{n_\Omega}) \\ K(X_{n_{\partial\Omega}}, X_{n_\Omega}) & K(X_{n_{\partial\Omega}}, X_{n_{\partial\Omega}}) \end{bmatrix} = \begin{bmatrix} K_{\Omega\Omega} & K_{\partial\Omega\Omega} \\ K_{\Omega\partial\Omega} & K_{\partial\Omega\partial\Omega} \end{bmatrix}, \tag{3}$$

where $X_{n_\Omega}, X_{n_{\partial\Omega}}$ are the data-points on the domain, Ω and on the boundary, $\partial\Omega$, respectively. We also define

$$\tilde{\mathrm{I}} = \begin{pmatrix} \sigma_\Omega^2 \mathrm{I} & 0 \\ 0 & \sigma_{\partial\Omega}^2 \mathrm{I} \end{pmatrix}. \tag{4}$$

where

$$K_{\text{PINN}} = \begin{pmatrix} LK_{\Omega,\Omega}L^\dagger & LK_{\Omega,\partial\Omega} \\ K_{\partial\Omega,\Omega}L^\dagger & K_{\partial\Omega,\partial\Omega} \end{pmatrix} \text{ and } \vec{k} = \begin{pmatrix} \vec{k}_\Omega L^\dagger \\ \vec{k}_{\partial\Omega} \end{pmatrix}, \tag{5}$$

where $\left(\vec{k}_\Omega\right)_\mu = K\left(\boldsymbol{x}_*, \boldsymbol{x}_{\mu\in[1,n_\Omega]}\right)$ and $\left(\vec{k}_{\partial\Omega}\right)_\mu = K\left(\boldsymbol{x}_*, \boldsymbol{x}_{\mu\in[n_\Omega+1, n_\Omega+n_{\partial\Omega}]}\right)$. We denote by L the linear operator acting in the forward direction and $L^\dagger$ the same operator acting "backwards" so that $KL^\dagger$ is K acted upon by L on its second argument.

B DERIVATION OF THE NEURALLY-INFORMED EQUATION (NIE)

In this section, we derive our main result, the Neurally Informed Equation, Eq. (5) in the main text. We start by analyzing posterior distribution in function space and show that it can be written in terms of an effective energy function "action" for a general operator. This provides our first result. We then focus on a linear differential operator and use variational calculus to derive the NIE.

B.1 DERIVATION OF THE EFFECTIVE ACTION

Using the Bayesian perspective, the posterior distribution in the function space can be written as follows:

$$\begin{aligned} p(f|X_n) &= \mathcal{N}(\mathcal{L}(X_n), \sigma^2 = 1)p_0(f|X) \\ &= \prod_{\mu=1}^{n_\Omega} \mathcal{N}(L[f](\boldsymbol{x}_\mu) - \phi(\boldsymbol{x}_\mu), \sigma_\Omega^2) \prod_{\nu=1}^{n_{\partial\Omega}} \mathcal{N}(f(\boldsymbol{x}_\nu) - g(\boldsymbol{x}_\nu), \sigma_{\partial\Omega}^2) p_0(f|X_n), \end{aligned} \tag{6}$$

where the distribution of the output of the network is given by averaging over the final distribution of the weight $p_0(f|X_n) = \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) \prod_{\mu} \delta\left(f(\boldsymbol{x}_\mu) - \hat{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_\mu)\right)$, and $\mathcal{L}(X_n)$ is the training loss in Eq. (2) of the main text. Adopting a statistical physics viewpoint, we can write the posterior distribution as $p(f|X_n) = \frac{1}{\mathcal{Z}[X_n]} e^{-\mathcal{S}[f|X_n]}$, where

$$\mathcal{S}[f|X_n] = \mathcal{S}_{\mathrm{PDE}}[f|X_n] + \mathcal{S}_0[f|X_n] \tag{7}$$

where,

$$\mathcal{S}_{\mathrm{PDE}}[f|X_n] = \mathcal{S}_{\Omega}[f|X_n] + \mathcal{S}_{\partial\Omega}[f|X_n] \tag{8}$$

$$= \frac{1}{2\sigma_{\Omega}^2} \sum_{\mu=1}^{n_\Omega} \left(L[f](\boldsymbol{x}_\mu) - \phi(\boldsymbol{x}_\mu)\right)^2 + \frac{1}{2\sigma_{\partial\Omega}^2} \sum_{\nu=1}^{n_{\partial\Omega}} \left(f(\boldsymbol{x}_\nu) - g(\boldsymbol{x}_\nu)\right)^2 \tag{9}$$

$$\mathcal{S}_0[f|X_n] = \frac{1}{2} \sum_{\mu=1}^{n} \sum_{\nu=1}^{n} f_\mu [K^{-1}]_{\mu\nu} f_\nu + \frac{1}{2} \log|K| + \frac{n}{2} \left(3\log 2\pi + 2\log \sigma_{\partial\Omega}\sigma_{\Omega}\right), \tag{10}$$

and $\mathcal{Z}[X_n] = \int e^{-\mathcal{S}[f|X_n]} df$ is the partition function.

To understand generalization, we are interested in computing ensemble averages over many datasets. In the following, we use our freedom to choose the prior on the function to be on an infinite dataset, i.e. in the reproduction kernel Hilbert space (RKHS), with the kernel being the continuum version of Eq. (3) Williams & Rasmussen (2006). Utilizing a statistical mechanics approach, we want to compute the free energy, $\mathbb{E}_{\mathcal{D}_n}\left[\log \mathcal{Z}[X_n]\right]$ which can also be viewed as the generating function of the process. For this purpose, we employ the replica trick:

$$\mathbb{E}_{\mathcal{D}_n}\left[\log \mathcal{Z}[X_n]\right] = \lim_{p \to 0} \frac{\partial \log \mathbb{E}_{\mathcal{D}_n}\left[\mathcal{Z}^p[X_n]\right]}{\partial p}$$

where

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}_n}\left[\mathcal{Z}^p[X]\right] &= \mathbb{E}_{\mathcal{D}_n}\left[\int \prod_{\alpha}^{p} e^{-\mathcal{S}[f_\alpha|X_n]} df_\alpha\right] \\
&= \int \mathbb{E}_{\mathcal{D}_n}\left[e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\mathrm{PDE}}[f_\alpha|X_n]}\right] \prod_{\alpha}^{p} p_0(f_\alpha) df_\alpha \\
&= \langle \mathbb{E}_{\mathcal{D}_n}\left[e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\mathrm{PDE}}[f_\alpha|X_n]}\right] \rangle_{0,p} = \langle \left(\mathbb{E}_{\boldsymbol{x}}\left[e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\mathrm{PDE}}[f_\alpha|\boldsymbol{x}]}\right]\right)^n \rangle_{0,p}
\end{aligned} \tag{11}$$

where $\langle ... \rangle_{0,p}$ denotes expectation over the p replicated prior distribution, $p_0(f)$, where α denotes the index of the replica modes of the system, in addition to the above we use the fact that the samples are i.i.d. The above object is still hard to analyze due to the coupling between the replica modes. To make progress, we follow a similar approach as in Cohen et al. (2021); Malzahn & Opper (2001). We transform into a "grand canonical" partition function, meaning that we treat the dataset size n as a random variable drawn from a Poisson distribution which we average over in the end.The idea is that for a large enough data set size, the dominating term in the grand canonical partition function $\bar{G}_p$ is a good estimator of the entire free energy. The grand canonical partition function is given by

$$\begin{aligned}
&\bar{G}_p = \mathbb{E}_n\left[\mathbb{E}_{\mathcal{D}_n}\left[\mathcal{Z}^p[X_n]\right]\right] \\
&= \langle \sum_{u_1=0}^{\infty} \frac{\bar{\eta}_{\Omega}^{u_1} e^{-\bar{\eta}_\Omega}}{u_1!} \mathbb{E}_{\mathcal{D}_{u_1}}\left[e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\Omega}[f_\alpha|X_{u_1}]}\right] \sum_{u_2=0}^{\infty} \frac{\bar{\eta}_{\partial\Omega}^{u_2} e^{-\bar{\eta}_{\partial\Omega}}}{u_2!} \mathbb{E}_{\mathcal{D}_{u_2}}\left[e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\partial\Omega}[f_\alpha|X_{u_2}]}\right] \rangle_{0,p} \\
&= \langle \sum_{u_1=0}^{\infty} \frac{\bar{\eta}_{\Omega}^{u_1} e^{-\bar{\eta}_\Omega}}{u_1!} \left(\mathbb{E}_{x}\left[e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\Omega}[f_\alpha|\boldsymbol{x}]}\right]\right)^{u_1} \sum_{u_2=0}^{\infty} \frac{\bar{\eta}_{\partial\Omega}^{u_2} e^{-\bar{\eta}_{\partial\Omega}}}{u_2!} \left(\mathbb{E}_{\boldsymbol{x}}\left[e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\partial\Omega}[f_\alpha|\boldsymbol{x}]}\right]\right)^{u_2} \rangle_{0,p} \\
&= \int \prod_{\alpha}^{p} e^{-\bar{\eta} + \bar{\eta}_\Omega \mathbb{E}_{\boldsymbol{x}}\left[e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\Omega}[f_\alpha|\boldsymbol{x}]}\right] + \bar{\eta}_{\partial\Omega} \mathbb{E}_{\boldsymbol{x}}\left[e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\partial\Omega}[f_\alpha|\boldsymbol{x}]}\right]} p_0(f_\alpha) df_\alpha \\
&= \int e^{-\sum_{\alpha=1}^{p} \mathcal{S}_{\mathrm{eff}}[f_\alpha]} \prod_{\alpha}^{p} \mathcal{Z}_\alpha^{-1} df_\alpha
\end{aligned} \tag{12}$$

where in the third transition, we assume that if the ith point is in the bulk then $\boldsymbol{x}_i \sim \mu_\Omega(\boldsymbol{x})$, otherwise, $\boldsymbol{x}_i \sim \mu_{\partial\Omega}(\boldsymbol{x})$, in addition, all points are chosen to be independent and identically distributed. With a slight abuse of notation, we take the number of data points in the bulk $n_\Omega \sim \mathrm{Pois}(\bar{\eta}_\Omega)$, and on the boundary $n_{\partial\Omega} \sim \mathrm{Pois}(\bar{\eta}_{\partial\Omega})$ as independent Poisson-distributed random variables. We denote by $\bar{\eta} = \bar{\eta}_\Omega + \bar{\eta}_{\partial\Omega}$, where $\bar{\eta}_\Omega$ and $\bar{\eta}_{\partial\Omega}$, the expectation of the Poisson variables on the bulk and boundary, are the true deterministic number of points in the bulk and on the boundary, respectively. The effective action is then defined as:

$$\mathcal{S}_{\text{eff}}[f] = \bar{\eta} - \bar{\eta}_\Omega \mathbb{E}_{\boldsymbol{x}}\left[e^{-\mathcal{S}_\Omega[f|\boldsymbol{x}]}\right] - \bar{\eta}_{\partial\Omega}\mathbb{E}_{\boldsymbol{x}}\left[e^{-\mathcal{S}_{\partial\Omega}[f|\boldsymbol{x}]}\right] + \frac{1}{2}\int_\Omega\int_\Omega f(\boldsymbol{x})[K^{-1}](\boldsymbol{x},\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{y}d\boldsymbol{x}. \quad (13)$$

Next, Taylor expanding the above exponent, we obtain the following effective action in first-order

$$\mathcal{S}_{\text{eff}}[f] = \bar{\eta}_\Omega \mathbb{E}_{\boldsymbol{x}}\left[\mathcal{S}_\Omega[f|\boldsymbol{x}]\right] - \bar{\eta}_{\partial\Omega}\mathbb{E}_{\boldsymbol{x}}\left[\mathcal{S}_{\partial\Omega}[f|\boldsymbol{x}]\right] + \frac{1}{2}\int_\Omega\int_\Omega f(\boldsymbol{x})[K^{-1}](\boldsymbol{x},\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y} + O(\bar{\eta}S^2_{\text{PDE}}), \quad (14)$$

where $\sigma^2 = \sigma^2_\Omega + \sigma^2_{\partial\Omega}$. Note that, the first-order action decouples the replica modes, which drastically simplifies the analysis . Taking higher-order corrections into account would lead to a tighter prediction. Taking the standard GP limit of infinite sample size proportional to the noise level and using the fact that the square distance between the estimate and the target decreases as well with n, these higher-order terms decrease to zero asymptotically in n. Figure (1) of the main text shows that the estimated output derived using the first-order action provides a good prediction for a large amount of data. For an analysis of the higher-order terms in the context of regression, see Cohen et al. (2021). Using variational calculus, this action provides the network's average prediction at any point $\boldsymbol{x}$ in the domain for a general nonlinear operator. Intuitively, one would expect that the effective action follows from taking the continuum limit of $\mathcal{S}[f|X_n]$. Yet doing so is subtle, due to the appearance of separate bulk and boundary measures in the continuum limit and the fact that the operator K^{-1} (unlike K) is measure-dependent. More precisely, the inversion operation depends on the whole measure of points both on the boundary and on the bulk $\mu(\boldsymbol{x})$ meaning that $\int[K]^{-1}(\boldsymbol{z},\boldsymbol{x})K(\boldsymbol{x},\boldsymbol{y})\mu(\boldsymbol{x})d\boldsymbol{x} = \delta_{\boldsymbol{z},\boldsymbol{y}}/\mu(y)$.

In the next section, we apply the calculus of variations to derive the neurally-informed equation (NIE) for the estimator from the effective action.

B.2 Derivation of the NIE using variational calculus

In principle, one could apply variational calculus to derive from the effective action an NIE for a general non-linear operator. However, since there are infinitely many ways for the operator to be nonlinear (e.g. $L = L(1, f, f_x, f_{xx}, .., f^2, f_x f, ...)$), and this is very much a problem dependent. We explicitly demonstrate the derivation for a general linear operator, L. The operator L is applied to the scalar function $f(\boldsymbol{x})$, with $\boldsymbol{x} \in \mathbb{R}^d$. The corresponding effective action is

$$\mathcal{S}_{\text{eff}}[f] = \frac{1}{2}\eta_\Omega \int_\Omega \left(Lf(\boldsymbol{x}) - \phi(\boldsymbol{x})\right)^2 d\boldsymbol{x} + \frac{1}{2}\eta_{\partial\Omega}\int_{\partial\Omega}\left(f(\boldsymbol{x}) - g(\boldsymbol{x})\right)^2 d\boldsymbol{x} + \frac{1}{2}\int_\Omega\int_\Omega f(\boldsymbol{x})[K^{-1}](\boldsymbol{x},\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{y}d\boldsymbol{x}. \quad (15)$$

Suppose that f_0 is a minimizer, and $h : \Omega \to \mathbb{R}$ is a variation function,

$$
\begin{aligned}
\delta\mathcal{S}_{\text{eff}}[f_0] = \frac{1}{2}\frac{d}{d\epsilon}S_{\text{eff}}(f_0+\epsilon h, Lf_0+\epsilon Lh))\ |_{\epsilon=0} \\
= \eta_\Omega \int_\Omega \left(Lf_0(\boldsymbol{x}) + \epsilon Lh(\boldsymbol{x}) - \phi(\boldsymbol{x})\right) Lh(\boldsymbol{x}) d\boldsymbol{x} \\
+ \eta_{\partial\Omega} \int_{\partial\Omega} \left(f_0(\boldsymbol{x}) - \epsilon h(\boldsymbol{x}) - g(\boldsymbol{x})\right) h(\boldsymbol{x}) d\boldsymbol{x} \\
+ \frac{1}{2}\int_\Omega h(\boldsymbol{x}) \int_\Omega [K^{-1}](\boldsymbol{x},\boldsymbol{y}) \left(f_0(\boldsymbol{y}) + \epsilon h(\boldsymbol{y})\right) d\boldsymbol{x} d\boldsymbol{y} \\
+ \frac{1}{2}\int_\Omega \left(f_0(\boldsymbol{x}) + \epsilon h(\boldsymbol{x})\right) \int_\Omega [K^{-1}](\boldsymbol{x},\boldsymbol{y})] h(\boldsymbol{y}) d\boldsymbol{y} d\boldsymbol{x}\ |_{\epsilon=0} \\
= \eta_\Omega \int_\Omega d\boldsymbol{x} \left(Lf_0(\boldsymbol{x}) - \phi(\boldsymbol{x})\right) Lh(\boldsymbol{x}) \\
+ \eta_{\partial\Omega} \int_{\partial\Omega} \left(f_0(\boldsymbol{x}) - g(\boldsymbol{x})\right) h(\boldsymbol{x}) d\boldsymbol{x} \\
+ \int_\Omega \int_\Omega [K^{-1}](\boldsymbol{x},\boldsymbol{y}) f_0(\boldsymbol{x}) h(\boldsymbol{y}) d\boldsymbol{y} d\boldsymbol{x} \quad (16)
\end{aligned}
$$

Next, we use our freedom how to parameterize the variation to simplify the boundary terms. Specifically, we consider a variation of the form

$$h(\boldsymbol{x}) = (K \star \tilde{h})(\boldsymbol{x}) = \int_\Omega K(\boldsymbol{x},\boldsymbol{y}) h(\boldsymbol{y}) d\boldsymbol{y},$$

where $\star$ denote the convolution operator on the bulk measure and $\tilde{h} : \Omega \to \mathbb{R}$. Provided that the kernel K is invertible, we stress that any variation can be presented in this manner. Substituting into Eq. equation 16

$$
\begin{aligned}
\delta\mathcal{S}_{\text{eff}}[f_0] = \eta_\Omega \int_\Omega \int_\Omega \left(Lf_0(\boldsymbol{x}) - \phi(\boldsymbol{x})\right) [LK](\boldsymbol{x},\boldsymbol{y}) \tilde{h}(\boldsymbol{y}) d\boldsymbol{y} d\boldsymbol{x} \\
+ \eta_{\partial\Omega} \int_\Omega d\boldsymbol{y} \int_{\partial\Omega} d\boldsymbol{x} (f_0(\boldsymbol{x}) - g(\boldsymbol{x})) K(\boldsymbol{x},\boldsymbol{y}) \tilde{h}(\boldsymbol{y}) + \int_\Omega f_0(\boldsymbol{x}) \tilde{h}(\boldsymbol{x}) d\boldsymbol{x}. \quad (17)
\end{aligned}
$$

In the notation, LK, the operator L is defined to act on the first coordinate of the operator K. Differentiating with respect to $\tilde{h}$ gives

$$
\begin{aligned}
\eta_\Omega \int_\Omega \left(Lf_0(\boldsymbol{y}) - \phi(\boldsymbol{y})\right) [LK](\boldsymbol{y},\boldsymbol{x}) d\boldsymbol{y} \\
+ \eta_{\partial\Omega} \int_{\partial\Omega} (f_0(\boldsymbol{y}) - g(\boldsymbol{y})) K(\boldsymbol{y},\boldsymbol{x}) d\boldsymbol{y} + f_0(\boldsymbol{x}) = 0. \quad (18)
\end{aligned}
$$

We note in passing that the above derivation can also be done for a nonlinear operator, by allowing the functional derivative to also act on the operator L.

C Generating some NNGP kernels

The kernel used in the toy model could be generated using the following random neural network acting on a one-dimensional input x

$$f(x) = \sum_{c=1}^{C} a_c \cos(w_c x) \quad (19)$$

with

$$
\begin{aligned}
a_c &\sim N(0, \sigma_a^2/C) \quad (20) \\
w_c &\sim N(0, \sigma_w^2)
\end{aligned}
$$

where we will soon take $\sigma_w^2 = 1/l^2$ and $\sigma_a^2 = \frac{1}{\sqrt{2\pi l^2}}$. The NNGP kernel is then

$$K(x,y) \equiv \langle f(x)f(y)\rangle_{a,w} = \sigma_a^2 \frac{1}{\sqrt{2\pi\sigma_w^2}} \int dw e^{-\frac{w^2}{2\sigma_w^2}} \cos(wx)\cos(wy) \tag{21}$$
$$= \frac{\sigma_a^2}{2\sqrt{2\pi\sigma_w^2}} \int dw e^{-\frac{w^2}{2\sigma_w^2}} [e^{iw(x+y)} + e^{iw(x-y)}] = \frac{\sigma_a^2}{2}[e^{-\frac{\sigma_w^2(x+y)^2}{2}} + e^{-\frac{\sigma_w^2(x-y)^2}{2}}]$$

hence, as mentioned, choosing $\sigma_w^2 = 1/l^2$ and $\sigma_a^2 = 1/\sqrt{2\pi l^2}$ reproduces the desired NNGP kernel. This means that training such a neural network with weight decay proportional to l^2 on the w_c weights, PINN loss, and using Langevin type training, samples from the GP posterior we used in our toy example.

D Green's function for the toy model

We first focus on inverting the bulk operator on the l.h.s. $(K^{-1} - \eta_\Omega \delta(x-y)\partial_y^2)$, which now derive. Consider the following set of basis functions for the positive half-interval,

$$\langle x|k > 0\rangle = \frac{\sqrt{2}}{\sqrt{\pi}}\cos(kx) = \frac{1}{\sqrt{2\pi}}[e^{ikx} + e^{-ikx}] \tag{22}$$

Notably

$$\langle k|k'\rangle = \int_0^\infty dx \langle k|x\rangle\langle x|k'\rangle \tag{23}$$
$$= \int_0^\infty dx \frac{2}{\pi}\cos(kx)\cos(k'x) = \int_{-\infty}^\infty dx \frac{1}{\pi}\cos(kx)\cos(k'x)$$
$$= \int_{-\infty}^\infty dx \frac{1}{4\pi}[e^{ikx} + e^{-ikx}][e^{ik'x} + e^{-ik'x}] = \delta(|k| - |k'|)$$

Consider $K(x,y)$ on this basis

$$K|k\rangle = \int_0^\infty dy K(x,y)\frac{\sqrt{2}}{\sqrt{\pi}}\cos(ky) = \int_{-\infty}^\infty dy K(x,y)\frac{1}{\sqrt{2\pi}}\cos(ky) \tag{24}$$
$$= \int_{-\infty}^\infty dy K(x,y)\frac{1}{2\sqrt{2\pi}}[e^{iky} + e^{-iky}] = 2e^{-(kl)^2/2}\frac{1}{\sqrt{2\pi}}\cos(kx) = e^{-(kl)^2/2}|k\rangle$$

Transforming the bulk operator to this "k-space"

$$\langle k|[K^{-1} - \eta_\Omega\partial_x^2]|k'\rangle = \delta(k-k')\left[e^{(kl)^2/2} + \eta_\Omega k^2\right]. \tag{25}$$

We thus find the following expression for the Green function which is defined here as $[K^{-1} + \eta_\Omega L^T L]G(x,y) = \delta(x-y)$,

$$G(x,x') = \frac{2}{\pi}\int_0^\infty dk dk' \cos(kx)\cos(k'x')\delta(k-k')\left[e^{+(kl)^2/2} + \eta_\Omega k^2\right]^{-1} \tag{26}$$
$$= \frac{2}{\pi}\int_0^{+\infty} dk \cos(kx)\cos(kx')\left[e^{+(kl)^2/2} + \eta_\Omega k^2\right]^{-1}$$
$$= \frac{1}{\pi}\int_{-\infty}^{+\infty} dk \cos(kx)\cos(kx')\left[e^{+(kl)^2/2} + \eta_\Omega k^2\right]^{-1}$$
$$= \frac{1}{4\pi}\int_{-\infty}^{+\infty} dk[e^{ikx} + e^{-ikx}][e^{ikx'} + e^{-ikx'}]\left[e^{+(kl)^2/2} + \eta_\Omega k^2\right]^{-1}.$$

This integral can be evaluated numerically or via contour integration. Considering the latter, one finds an infinite set of simple poles. As $|x - x'|$ grows or for very large η_Ω, we find numerically that a single pole dominates the result and yields $G(x,x') \approx \frac{\kappa}{2}\left[e^{-\kappa|x-x'|} + e^{-\kappa|x+x'|}\right]$ where $\kappa = \frac{1}{\sqrt{l^2/2+\eta_\Omega}}$ for $\kappa l \ll 1$. While this approximation can be systematically improved by accounting for additional poles, in the numerics carried below we simply calculate this integral numerically.

E DERIVING THE $Q_n[\phi, g]$ FIGURE OF MERIT

Here, we derive the $Q_n[\phi, g]$ figure of merit, which also exposes some hidden algebraic relations between the GPR formula and the neurally-informed equation.

First, we note that one can re-rewrite the neurally-informed equation as an operator equation without performing any integration by parts, namely

$$\int_\Omega [K^{-1} + \eta_{\partial\Omega}\delta\delta + \eta_\Omega L^\dagger L]_{\boldsymbol{x}\boldsymbol{y}} f(\boldsymbol{y}) d\boldsymbol{y} = \eta_{\partial\Omega}\delta_{\partial\Omega}(\boldsymbol{x}) g(\boldsymbol{x}) + \eta_\Omega L^\dagger \phi(\boldsymbol{x}) \tag{27}$$

where $\delta_{\partial\Omega}(\boldsymbol{x}) = \int_{\partial\Omega} \delta(\boldsymbol{x} - \boldsymbol{z}) d\boldsymbol{z}$, $[\delta\delta]_{\boldsymbol{x}\boldsymbol{y}} = \int_{\partial\Omega} \delta(\boldsymbol{x} - \boldsymbol{z})\delta(\boldsymbol{z} - \boldsymbol{y}) d\boldsymbol{z}$, and $[fL^\dagger]_{\boldsymbol{x}} = [Lf]_{\boldsymbol{x}}$, which is then consistent with $\int_\Omega [L^\dagger L]_{\boldsymbol{x}\boldsymbol{y}} f(\boldsymbol{y}) d\boldsymbol{y} = \int_\Omega L^\dagger_{\boldsymbol{x}} \delta(\boldsymbol{x} - \boldsymbol{y}) L_{\boldsymbol{y}} f(\boldsymbol{y}) d\boldsymbol{y}$. Noting that the operator on the l.h.s. is a sum of positive semi-definite operators and that K^{-1} is positive definite (indeed K is generically semi-definite and bounded), we invert the operator on the left-hand side and obtain

$$\begin{aligned} f(\boldsymbol{x}) &= \eta_{\partial\Omega} \int_{\partial\Omega} \left[[K^{-1} + \eta_\Omega L^\dagger L + \eta_{\partial\Omega}\delta\delta]^{-1}\right]_{\boldsymbol{x}\boldsymbol{z}} g(\boldsymbol{z}) d\boldsymbol{z} \\ &+ \eta_\Omega \int_\Omega \left[[K^{-1} + \eta_\Omega L^\dagger L + \eta_{\partial\Omega}\delta\delta]^{-1}\right]_{\boldsymbol{x}\boldsymbol{y}} [L^\dagger \phi]_{\boldsymbol{y}} d\boldsymbol{y} \end{aligned} \tag{28}$$

where the inverse is taken with respect to the bulk measure. Next, we define the following operator,

$$\begin{aligned} \hat{K}_{\boldsymbol{x}\boldsymbol{y}} &\equiv \left[[K^{-1} + \eta_{\partial\Omega}\delta\delta]^{-1}\right]_{\boldsymbol{x}\boldsymbol{y}} \\ &= K(\boldsymbol{x}, \boldsymbol{y}) - \int_{\partial\Omega}\int_{\partial\Omega} K(\boldsymbol{x}, \boldsymbol{z}_1)[K + \eta_{\partial\Omega}^{-1}]^{-1}(\boldsymbol{z}_1, \boldsymbol{z}_2) K(\boldsymbol{z}_2, \boldsymbol{y}) d\boldsymbol{z}_1 d\boldsymbol{z}_2 \end{aligned} \tag{29}$$

where the second transition is since the operator K is assumed invertible, a Woodbury-type manipulation can be applied. Note also that the inverse after the second inequality is w.r.t. the boundary measure and $\eta_{\partial\Omega} = n_{\partial\Omega}/\sigma^2_{\partial\Omega}$. If the boundary is a single point, obtaining $\hat{K}$ is again straightforward, since the operator inverse becomes just a simple algebraic inverse. Substituting Eq. equation 29

$$[K^{-1} + \eta_{\partial\Omega}\delta\delta + \eta_\Omega L^\dagger L]^{-1} = [\hat{K}^{-1} + \eta_\Omega L^\dagger L]^{-1}. \tag{30}$$

Next, we perform a similar manipulation to the one leading to $\hat{K}$ namely

$$\begin{aligned} [\hat{K}^{-1} + \eta_\Omega L^\dagger L]^{-1} &= \hat{K}[1 + \eta_\Omega L^\dagger L \hat{K}]^{-1} \\ &= \hat{K} - \hat{K} L^\dagger \left(\eta_\Omega^{-1} + L\hat{K}L^\dagger\right)^{-1} L\hat{K} \end{aligned} \tag{31}$$

To obtain the result in the main text, we apply the operator L on Eq. equation 28, i.e. Lf. It has two contributions, we first start with the source term contribution

$$\begin{aligned} &\eta_\Omega L \int_\Omega \left[[\hat{K}^{-1} + \eta_\Omega L^\dagger L]^{-1}\right]_{\boldsymbol{x}\boldsymbol{y}} [L^\dagger \phi]_{\boldsymbol{y}} d\boldsymbol{y} \\ &= \eta_\Omega L \left[\hat{K} - \hat{K} L^\dagger \left(\eta_\Omega^{-1} + L\hat{K}L^\dagger\right)^{-1} L\hat{K}\right] L^\dagger \phi \\ &= \eta_\Omega \left[(L\hat{K}L^\dagger) - (L\hat{K}L^\dagger)\left(\eta_\Omega^{-1} + (L\hat{K}L^\dagger)\right)^{-1} (L\hat{K}L^\dagger)\right] \phi \\ &= \eta_\Omega \left[\eta_\Omega^{-1}\left(\eta_\Omega^{-1} + (L\hat{K}L^\dagger)\right)^{-1} (L\hat{K}L^\dagger)\right] \phi \\ &= \left(\eta_\Omega^{-1} + (L\hat{K}L^\dagger)\right)^{-1} (L\hat{K}L^\dagger)\phi, \end{aligned} \tag{32}$$

where to simplify the notation in places where there is no indication of position the position is $\boldsymbol{x}$. Similarly, the boundary contribution,

$$\begin{aligned}
&\eta_{\partial\Omega} L \int_{\partial\Omega} \left[[\hat{K}^{-1} + \eta_{\Omega} L^{\dagger} L]^{-1} \right]_{\boldsymbol{xz}} g(\boldsymbol{z}) d\boldsymbol{z} \\
&= \eta_{\partial\Omega} L \left[\hat{K} - \hat{K} L^{\dagger} \left(\eta_{\Omega}^{-1} + L \hat{K} L^{\dagger} \right)^{-1} L \hat{K} \right] g \\
&= \eta_{\partial\Omega} \left[(L\hat{K}) - (L\hat{K}L^{\dagger}) \left(\eta_{\Omega}^{-1} + (L\hat{K}L^{\dagger}) \right)^{-1} (L\hat{K}) \right] g \\
&= \eta_{\partial\Omega} \left[1 - (L\hat{K}L^{\dagger}) \left(\eta_{\Omega}^{-1} + (L\hat{K}L^{\dagger}) \right)^{-1} \right] L\hat{K} g \\
&= \frac{\eta_{\partial\Omega}}{\eta_{\Omega}} \left(\eta_{\Omega}^{-1} + (L\hat{K}L^{\dagger}) \right)^{-1} L\hat{K} g \\
&= \frac{\eta_{\partial\Omega}}{\eta_{\Omega}} \int_{\partial\Omega} \left[\left(\eta_{\Omega}^{-1} + (L\hat{K}L^{\dagger}) \right)^{-1} L\hat{K} \right]_{\boldsymbol{xz}} g(\boldsymbol{z}) d\boldsymbol{z}
\end{aligned} \tag{33}$$

Consequently, $Lf - \phi$ is

$$\begin{aligned}
Lf(\boldsymbol{x}) - \phi &= \left[\left(\eta_{\Omega}^{-1} + (L\hat{K}L^{\dagger}) \right)^{-1} (L\hat{K}L^{\dagger}) - 1 \right] \phi + \frac{\eta_{\partial\Omega}}{\eta_{\Omega}} \left(\eta_{\Omega}^{-1} + (L\hat{K}L^{\dagger}) \right)^{-1} L\hat{K} g \\
&= -\eta_{\Omega}^{-1} \left(\eta_{\Omega}^{-1} + L\hat{K}L^{\dagger} \right)^{-1} \phi + \frac{\eta_{\partial\Omega}}{\eta_{\Omega}} \left(\eta_{\Omega}^{-1} + (L\hat{K}L^{\dagger}) \right)^{-1} L\hat{K} g
\end{aligned} \tag{34}$$

we thus find that $\eta_{\partial\Omega} L\hat{K} g$ acts as an additional source term. While it may seem it diverges with $\eta_{\partial\Omega}$ we recall that $\hat{K}$ goes to zero at this limit for arguments on the boundary, hence its contribution is finite and the overall η_{Ω}^{-1} ensures this quantity $Lf - \phi$ goes to zero. Changing the basis to the eigenfunction basis of $L\hat{K}L^{\dagger}$ leads to the spectral bias result of the main text.

Last, we note in passing that $\hat{K}$ as the interpretation of the dataset-averaged posterior covariance, given that one introduced only boundary points and fixes them to zero Cohen et al. (2021). Thus, as $n_{\partial\Omega} \to \infty$, $\hat{K}$ involving any boundary point, is zero.

From a practical point of view, one can obtain an estimate for $\hat{K}$ at small, $\eta_{\partial\Omega}$. A straightforward expansion of this quantity to its leading order is then,

$$\begin{aligned}
\hat{K}_{\boldsymbol{xy}} = K(\boldsymbol{x}, \boldsymbol{y}) &- \eta_{\partial\Omega} \int_{\partial\Omega} K(\boldsymbol{x}, \boldsymbol{z}) K(\boldsymbol{z}, \boldsymbol{y}) d\boldsymbol{z} \\
&- \eta_{\partial\Omega}^{2} \int_{\partial\Omega} K(\boldsymbol{x}, \boldsymbol{z}) K(\boldsymbol{z}, \boldsymbol{z}') K(\boldsymbol{z}', \boldsymbol{y}) d\boldsymbol{z} d\boldsymbol{z}' + O(\eta_{\partial\Omega}^{3})
\end{aligned} \tag{35}$$

which can then be evaluated analytically in some cases.

REFERENCES

Omry Cohen, Or Malka, and Zohar Ringel. Learning curves for overparametrized deep neural networks: A field theory perspective. *Physical Review Research*, 3(2):023034, 2021.

Dörthe Malzahn and Manfred Opper. A variational approach to learning curves. *Advances in neural information processing systems*, 14, 2001.

Guofei Pang and George Em Karniadakis. Physics-informed learning machines for partial differential equations: Gaussian processes versus neural networks. *Emerging Frontiers in Nonlinear Science*, pp. 323–343, 2020.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Machine learning of linear differential equations using gaussian processes. *Journal of Computational Physics*, 348:683–693, 2017.

Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.