Physics Enhanced Deep Surrogates for the diffusion equation

Eike Mueller

July 14, 2025

1 Problem specification

Consider the domain $\Omega \subset \mathbb{R}^d$ where $d = 1, 2$. Let the function $f : \Omega \to \mathbb{R}$ be fixed. For a given function $\alpha : \Omega \to \mathbb{R}$, we want to find the function $u : \Omega \to \mathbb{R}$ which satisfies

$$-\nabla\left(\exp[\alpha(x)]\nabla u(x)\right) = f(x) \qquad \text{for all } x \in \Omega \tag{1}$$

subject to suitable boundary conditions on $\partial\Omega$. In a more abstract sense, let V be the space of real-valued functions defined on Ω. We want to find the map $\mathcal{F} : V \to V$ with $\mathcal{F} : \alpha \mapsto u(\alpha)$ where $u = u(\alpha)$ is defined by (1) together with the boundary conditions on u.

1.1 One dimensional problem

Let $\Omega = [0, L]$ be the interval of length L. In $d = 1$ dimension (1) becomes

$$-\frac{d}{dx}\left(\exp[\alpha(x)]\frac{du}{dx}(x)\right) = f(x) \qquad \text{for all } 0 < x < 1 \tag{2}$$

and we choose the boundary conditions

$$\begin{aligned} u(x=0) &= 0 \qquad \text{(homogeneous Dirichlet)}\\ \frac{du}{dx}(x=L) &= -g\exp[-\alpha(L)] \qquad \text{(von Neumann)} \end{aligned} \tag{3}$$

1.2 Two-dimensional problem

The problem can be extended to two dimensions by considering the domain $\Omega = [0, L] \times [0, L]$ and seeking $u(x)$ such that

$$-\frac{\partial}{\partial x_0}\left(\exp[\alpha(x)]\frac{\partial u}{\partial x_0}(x)\right) - \frac{\partial}{\partial x_1}\left(\exp[\alpha(x)]\frac{\partial u}{\partial x_1}(x)\right) = f(x) \qquad \text{for all } x = (x_0, x_1) \in \Omega \tag{4}$$

subject to the boundary conditions

$$\begin{aligned} u(x_0 = 0, x_1) &= 0 \qquad \text{(homogeneous Dirichlet)}\\ \frac{\partial u}{\partial x_1}(x_0 = L, x_1) &= -g^{(\text{right})}(x_1)\exp[-\alpha(x_0 = L, x_1)]\\ \frac{\partial u}{\partial x_0}(x_0, x_1 = 0) &= g^{(\text{bottom})}(x_0)\exp[-\alpha(x_0, x_1 = 0)] \qquad \text{(von Neumann)}\\ \frac{\partial u}{\partial x_0}(x_0, x_1 = L) &= -g^{(\text{top})}(x_0)\exp[-\alpha(x_0, x_1 = L)] \end{aligned} \tag{5}$$

1.3 Physical interpretation

Note that by setting $K(x) := \exp[\alpha(x)]$, (1) can be written as $-\nabla\cdot(K\nabla u) = f$. So physically the problem is about finding the solution of the diffusion equation for a given spatially varying diffusion coefficient $K(x) > 0$ which is parametrised by its logarithm $\alpha(x)$.

Imagine, for example, that u is a temperature field. Then the heat flux is given by

$$\phi = -K\nabla u \tag{6}$$

where K is the spatially varying thermal conductivity. If f describes the sources and sinks of heat, then the (stationary) conservation law is

$$\nabla\cdot\phi = f. \tag{7}$$

Dirichlet boundary conditions correspond to fixing the temperature whereas von Neumann boundary conditions correspond to fixing the heat flux.

1.4 Quantity of interest

Usually we are not really interested in predicting $u(x)$ itself for a given $\alpha(x)$ but instead want to compute some functional Q of $u(x)$

$$\mathcal{Q} : V \to \mathbb{R}^\beta. \tag{8}$$

In other words, we are interested in the map $\mathcal{Q}\circ\mathcal{F} : V \to \mathbb{R}^\beta$. $\mathcal{Q}$ is also called the "Quantity of Interest" (short QoI). Some examples are:

- Average field ($\beta = 1$):

$$\mathcal{Q}(u) = \int_\Omega u(x)\;dx \tag{9}$$

- Sampled field:

$$\mathcal{Q}(u) = (u(\xi_1), u(\xi_2), \ldots, u(\xi_\beta)) \tag{10}$$

  for some sample points $\xi_1, \xi_2, \ldots, \xi_\beta \in \Omega$.

- Higher moments of field:

$$\mathcal{Q}(u) = \left(\int_\Omega u(x)^{\nu_1}\;dx, \int_\Omega u(x)^{\nu_2}\;dx, \ldots, \int_\Omega u(x)^{\nu_\beta}\;dx\right) \tag{11}$$

  for some powers $\nu_1, \nu_2, \ldots, \nu_\beta \in \mathbb{N}$.

Note that $\mathcal{Q}$ is linear in the first two cases but non-linear in the last case.

2 Discretisation

2.1 One-dimensional problem

Let $\Omega = [0, L]$ be the interval of length L. Split Ω into $n = m$ sub-intervals of size $h = L/m$ to obtain a grid $\Omega_h := \bigcup_{j=0}^{m-1} I_j^{(h)}$ where $I_j^{(h)} = [jh, (j+1)h] = [x_j^{(h)}, x_{j+1}^{(h)}]$ for $j = 0, 1, \ldots, m-1$ is the j-th sub-interval or cell of the grid. We call the points $x_j^{(h)} = jh$ for $j = 0, 1, \ldots, m$ the vertices of the grid.

Let $\alpha^{(h)} \in \mathbb{R}^{\widetilde{n}}$ with $\widetilde{n} = m + 1$ be the vector with $\alpha_j^{(h)} \approx \alpha(x_j^{(h)})$ for $j = 0, 1, \ldots, m$ and $u^{(h)} \in \mathbb{R}^n$ be the vector with

$$u_j^{(h)} \approx \frac{1}{h}\int_{I_j^{(h)}} u(x)\;dx \qquad \text{for } j = 0, 1, \ldots, m-1, \tag{12}$$

with an analogous definition of the vector $f^{(h)} \in \mathbb{R}^n$. The definition in (12) says that the components of the vector $u^{(h)}$ approximate the volume averages of the solution in the grid cells. To discretise (2) with the finite volume method, integrate the equation over cell $I_j^{(h)}$ and use the boundary conditions in (3):

$$
\begin{aligned}
hf_j^{(h)} &:= \int_{I_j^{(h)}} f(x)\,dx = -\int_{I_j^{(h)}} \frac{d}{dx}\left(\exp[\alpha(x)]\frac{du}{dx}(x)\right)\,dx \\
&= \exp[\alpha(x_j^{(h)})]\frac{du}{dx}(x_j^{(h)}) - \exp[\alpha(x_{j+1}^{(h)})]\frac{du}{dx}(x_{j+1}^{(h)}) \\
&\approx \begin{cases}
\exp[\alpha_j^{(h)}]\frac{2u_j^{(h)}}{h} - \exp[\alpha_{j+1}^{(h)}]\frac{u_{j+1}^{(h)}-u_j^{(h)}}{h} & \text{for } j=0 \\
\exp[\alpha_j^{(h)}]\frac{u_j^{(h)}-u_{j-1}^{(h)}}{h} - \exp[\alpha_{j+1}^{(h)}]\frac{u_{j+1}^{(h)}-u_j^{(h)}}{h} & \text{for } 0<j<m-1 \\
\exp[\alpha_j^{(h)}]\frac{u_j^{(h)}-u_{j-1}^{(h)}}{h} + g & \text{for } j=m-1
\end{cases}
\end{aligned}
\tag{13}
$$

This results in the following linear system:

$$
A^{(h)}u^{(h)} = \widetilde{f}^{(h)} := f^{(h)} - \frac{1}{h}\widehat{g}^{(h)}
\tag{14}
$$

where $\widehat{g}^{(h)} = (0,0,\ldots,0,g) \in \mathbb{R}^n$ is the vector which is zero everywhere except for its final entry. The $n \times n$ matrix $A^{(h)} = A^{(h)}(\alpha^{(h)})$ is given by

$$
A^{(h)}(\alpha^{(h)}) = \frac{1}{h^2}\begin{pmatrix}
2K_0+K_1 & -K_1 & 0 & 0 & 0 & \cdots & 0 \\
-K_1 & K_1+K_2 & -K_2 & 0 & 0 & \cdots & 0 \\
0 & -K_2 & K_2+K_3 & -K_3 & 0 & \cdots & 0 \\
0 & 0 & -K_3 & \ddots & & & \\
0 & 0 & 0 & & & & \vdots \\
\vdots & \vdots & \vdots & & \ddots & -K_{n-2} & 0 \\
 & & & & -K_{n-2} & K_{n-2}+K_{n-1} & -K_{n-1} \\
0 & 0 & 0 & \cdots & 0 & -K_{n-1} & K_{n-1}
\end{pmatrix}
\tag{15}
$$

where, for simplicity, we defined the diffusion coefficients as

$$
K_j^{(h)} = K_j^{(h)}(\alpha_j^{(h)}) := \exp[\alpha_j^{(h)}].
\tag{16}
$$

Note that by solving (14) for $u^{(h)}$ for a given $\alpha^{(h)}$ we can construct an approximation $\mathcal{F}_h$ of the map $\mathcal{F}_h$:

$$
\mathcal{F}_h : \mathbb{R}^{\widetilde{n}} \to \mathbb{R}^n, \qquad \alpha^{(h)} \mapsto u^{(h)} = u^{(h)}(\alpha^{(h)}) = \text{ the solution of (14) for given } \alpha^{(h)}.
\tag{17}
$$

2.2 Two-dimensional problem

Let $\Omega = [0,1]\times[0,1]$ be the unit square. Split Ω into $n = m^2$ cells of size $h \times h$ with $h = L/m$ to obtain a grid $\Omega_h := \bigcup_{j=0}^{m-1}\bigcup_{k=0}^{m-1} I_{jk}^{(h)}$ where $I_{jk}^{(h)} = [jh,(j+1)h]\times[kh,(k+1)h] = [x_j^{(h)}, x_{j+1}^{(h)}]\times[x_k^{(h)}, x_{k+1}^{(h)}]$ is the cell of the grid with integer coordinates $j,k = 0,1,\ldots,m-1$. The centre of cell $I_{jk}^{(h)}$ is given by $(x_{j+\frac{1}{2}}^{(h)}, x_{k+\frac{1}{2}}^{(h)})$. We call the points $x_{jk}^{(h)} = (jh,kh)$ for $j,k = 0,1,\ldots,m$ the vertices of the grid. Note that the grid has $\widetilde{n} = (m+1)^2$ vertices.

Let $\alpha^{(h)} \in \mathbb{R}^{(m+1)\times(m+1)}$ be the tensor with $\alpha_{jk}^{(h)} \approx \alpha(x_j^{(h)}, x_k^{(h)})$ for $j,k = 0,1,\ldots,m$ and $u^{(h)} \in \mathbb{R}^{m\times m}$ be the tensor with

$$
u_{jk}^{(h)} \approx \frac{1}{h^2}\int_{I_{jk}^{(h)}} u(x)\,dx \qquad \text{for } j,k = 0,1,\ldots,m-1,
\tag{18}
$$

with an analogous definition of the tensor $f^{(h)} \in \mathbb{R}^{m\times m}$. The definition in (18) says that the components of the matrix $u^{(h)}$ approximate the volume averages of the solution in the grid cells. To discretise (1) with the finite volume method, integrate the equation over cell $I_{jk}^{(h)}$ and use the boundary conditions in (5):

$$\begin{aligned} h^2 f_{jk}^{(h)} := \int_{I_{jk}^{(h)}} f(x)\;dx &= -\int_{I_{jk}^{(h)}} \frac{\partial}{\partial x_0}\left(\exp[\alpha(x)]\frac{\partial u}{\partial x_0}(x)\right)\;dx_0\;dx_1 \\ &\quad - \int_{I_{jk}^{(h)}} \frac{\partial}{\partial x_1}\left(\exp[\alpha(x)]\frac{\partial u}{\partial x_1}(x)\right)\;dx_0\;dx_1 \\ &=: J_{jk}^{(\text{left})} - J_{jk}^{(\text{right})} + J_{jk}^{(\text{bottom})} - J_{jk}^{(\text{top})} \end{aligned} \tag{19}$$

with

$$\begin{aligned} J_{jk}^{(\text{left})} &= \int_{x_k^{(h)}}^{x_{k+1}^{(h)}} \exp[\alpha(x_j^{(h)},x_1)]\frac{\partial u}{\partial x_0}(x_j^{(h)},x_1)\;dx_1 \approx h\begin{cases} K^{(h)}_{j,k+\frac{1}{2}}\frac{2u^{(h)}_{jk}}{h} & \text{for } j=0\\ K^{(h)}_{j,k+\frac{1}{2}}\frac{u^{(h)}_{jk}-u^{(h)}_{j-1,k}}{h} & \text{for } 0<j\le m-1\end{cases}\\ J_{jk}^{(\text{right})} &= \int_{x_k^{(h)}}^{x_{k+1}^{(h)}} \exp[\alpha(x_{j+1}^{(h)},x_1)]\frac{\partial u}{\partial x_0}(x_{j+1}^{(h)},x_1)\;dx_1 \approx h\begin{cases} K^{(h)}_{j+1,k+\frac{1}{2}}\frac{u^{(h)}_{j+1,k}-u^{(h)}_{jk}}{h} & \text{for } 0\le j<m-1\\ -g_k^{(\text{right})} & \text{for } j=m-1\end{cases}\\ J_{jk}^{(\text{bottom})} &= \int_{x_j^{(h)}}^{x_{j+1}^{(h)}} \exp[\alpha(x_0,x_k^{(h)})]\frac{\partial u}{\partial x_1}(x_0,x_k^{(h)})\;dx_0 \approx h\begin{cases} K^{(h)}_{j+\frac{1}{2},k}\frac{u^{(h)}_{jk}-u^{(h)}_{j,k-1}}{h} & \text{for } 0<k\le m-1\\ -g_j^{(\text{bottom})} & \text{for } k=0\end{cases}\\ J_{jk}^{(\text{top})} &= \int_{x_j^{(h)}}^{x_{j+1}^{(h)}} \exp[\alpha(x_0,x_{k+1}^{(h)})]\frac{\partial u}{\partial x_1}(x_0,x_{k+1}^{(h)})\;dx_0 \approx h\begin{cases} K^{(h)}_{j+\frac{1}{2},k+1}\frac{u^{(h)}_{j,k+1}-u^{(h)}_{jk}}{h} & \text{for } 0\le k<m-1\\ -g_j^{(\text{top})} & \text{for } k=m-1\end{cases} \end{aligned} \tag{20}$$

where we defined the diffusion coefficients

$$K^{(h)}_{j,k+\frac{1}{2}} := \exp\left[\frac{1}{2}\left(\alpha^{(h)}_{j,k}+\alpha^{(h)}_{j,k+1}\right)\right], \qquad K^{(h)}_{j+\frac{1}{2},k} := \exp\left[\frac{1}{2}\left(\alpha^{(h)}_{j,k}+\alpha^{(h)}_{j+1,k}\right)\right] \tag{21}$$

and set

$$g_k^{(\text{right})} := \int_{x_k^{(h)}}^{x_{k+1}^{(h)}} g^{(\text{right})}(x_1)\;dx_1,\quad g_j^{(\text{bottom})} := \int_{x_j^{(h)}}^{x_{j+1}^{(h)}} g^{(\text{bottom})}(x_0)\;dx_0,\quad g_j^{(\text{top})} := \int_{x_j^{(h)}}^{x_{j+1}^{(h)}} g^{(\text{top})}(x_0)\;dx_0. \tag{22}$$

Let $\nu : [0,1,\dots,m-1]\times[0,1,\dots,m-1] \to [0,1,\dots,n-1]$ be the function which maps the tuple j,k to a linear index with $\nu(j,k) := jm+k$. With this, we get the linear system

$$A^{(h)}u^{(h)} = \widetilde{f}^{(h)} := f^{(h)} - \frac{1}{h}\widehat{g}^{(h)} \tag{23}$$

where in a slight abuse of notation we write $u^{(h)}$, $\alpha^{(h)}$ both for the two-dimensional tensors and the corresponding flattened vectors with the same entries. The non-zero entries of the $n\times n$ matrix $A^{(h)} =$

$A^{(h)}(\alpha^{(h)})$ are given by

$$
\begin{aligned}
A^{(h)}_{\nu(j,k)\nu(j,k)} &= 2\Theta_0^1(j)\Theta_0^m(k)\frac{1}{h^2}K^{(h)}_{j,k+\frac{1}{2}} + \Theta_1^m(j)\Theta_0^m(k)\frac{1}{h^2}K^{(h)}_{j,k+\frac{1}{2}} + \Theta_0^{m-1}(j)\Theta_0^m(k)\frac{1}{h^2}K^{(h)}_{j+1,k+\frac{1}{2}} \\
&\quad + \quad \Theta_0^m(j)\Theta_1^m(k)\frac{1}{h^2}K^{(h)}_{j+\frac{1}{2},k} + \Theta_0^m(j)\Theta_0^{m-1}(k)\frac{1}{h^2}K^{(h)}_{j+\frac{1}{2},k+1} \\
A^{(h)}_{\nu(j,k)\nu(j-1,k)} &= -\Theta_1^m(j)\Theta_0^m(k)\frac{1}{h^2}K^{(h)}_{j,k+\frac{1}{2}} \\
A^{(h)}_{\nu(j,k)\nu(j+1,k)} &= -\Theta_0^{m-1}(j)\Theta_0^m(k)\frac{1}{h^2}K^{(h)}_{j+1,k+\frac{1}{2}} \\
A^{(h)}_{\nu(j,k)\nu(j,k-1)} &= -\Theta_0^m(j)\Theta_1^m(k)\frac{1}{h^2}K^{(h)}_{j+\frac{1}{2},k} \\
A^{(h)}_{\nu(j,k)\nu(j,k+1)} &= -\Theta_0^m(j)\Theta_0^{m-1}(k)\frac{1}{h^2}K^{(h)}_{j+\frac{1}{2},k+1}
\end{aligned}
\tag{24}
$$

where we defined the following function for integer i:

$$
\Theta_a^b(i) = \begin{cases} 1 & \text{for } a \le i < b \\ 0 & \text{otherwise} \end{cases}. \tag{25}
$$

The non-zero entries of the vector $\widehat{g}^{(h)} \in \mathbb{R}^n$ are given by

$$
\widehat{g}^{(h)}_{\nu(m-1,k)} = g_k^{(\text{right})}, \qquad \widehat{g}^{(h)}_{\nu(j,0)} = g_j^{(\text{bottom})}, \qquad \widehat{g}^{(h)}_{\nu(j,m-1)} = g_j^{(\text{top})}. \tag{26}
$$

Again, by solving (23) we can construct a solution map $\mathcal{F}(\alpha^{(h)}) = u^{(h)}$ as in (17).

2.3 Quantity of Interest.

We also need to construct an approximation $Q^{(h)}$ of the QoI in (8). For example, if the QoI is the average field, then $Q^{(h)} : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$
Q^{(h)}(u^{(h)}) = \begin{cases} \frac{1}{m}\sum_{j=0}^{m-1} u_j^{(h)} & \text{in } d = 1 \text{ dimensions} \\ \frac{1}{m^2}\sum_{j=0}^{m-1}\sum_{k=0}^{m-1} u_{jk}^{(h)} & \text{in } d = 2 \text{ dimensions.} \end{cases} \tag{27}
$$

If the QoI is the field sampled at some locations $\xi_1, \xi_2, \dots, \xi_\beta$, then $Q^{(h)} : \mathbb{R}^n \to \mathbb{R}^\beta$ is defined by

$$
Q^{(h)}(u^{(h)}) = \begin{cases} \left(u^{(h)}_{j_1}, u^{(h)}_{j_2}, \dots, u^{(h)}_{j_\beta}\right) & \text{with } \xi_t \in I^{(h)}_{j_t} \text{ for } t = 1,2,\dots,\beta. \quad \text{in } d = 1 \text{ dimensions} \\ \left(u^{(h)}_{j_1,k_1}, u^{(h)}_{j_2,k_2}, \dots, u^{(h)}_{j_\beta,k_\beta}\right) & \text{with } \xi_t \in I^{(h)}_{j_t,k_t} \text{ for } t = 1,2,\dots,\beta. \quad \text{in } d = 2 \text{ dimensions.} \end{cases} \tag{28}
$$

3 Solution procedure

3.1 Pure Numerical solution

1. Given $\alpha^{(h)} \in \mathbb{R}^{\widetilde{n}}$, compute $A^{(h)} = A^{(h)}(\alpha^{(h)})$.
2. Solve the linear system in (14) or (23) to obtain $u^{(h)} = u^{(h)}(\alpha^{(h)}) \in \mathbb{R}^n$.
3. Compute the quantity of interest $q^{(h)} = q^{(h)}(\alpha^{(h)}) = Q^{(h)}(u^{(h)}(\alpha^{(h)})) \in \mathbb{R}^\beta$

This method can be made very accurate by choosing $h \ll 1$, but it also becomes very expensive in this case. The computational bottleneck is the solution of the linear system $A^{(h)}u^{(h)} = \widetilde{f}^{(h)}$ in (14) or (23).

3.2 Pure ML solution

1. Generate a training data set $\{(\alpha^{(h)\ell}, q^{(h)\ell})\}_{\ell=1}^{N_{\text{samples}}}$ where for each pair $(\alpha^{(h)\ell}, q^{(h)\ell})$ we have that $q^{(h)\ell} = Q^{(h)}(u^{(h)\ell})$ and $u^{(h)\ell}$ satisfies (14) or (23), i.e.

$$A^{(h)}(\alpha^{(h)\ell})u^{(h)\ell} = \widetilde{f}^{(h)} \qquad \text{for } \ell = 1, 2, \ldots, N_{\text{samples}}. \tag{29}$$

2. Construct a neural network Φ_{NN} which is parametrised by some parameters θ

$$\Phi_{\text{NN}}(\cdot|\theta) : \mathbb{R}^{\widetilde{n}} \to \mathbb{R}^{\beta} \tag{30}$$

3. Train the neural network by minimising some loss function $L(\{(\alpha^{(h)\ell}, q^{(h)\ell})\}_{\ell=1}^{N_{\text{samples}}}|\theta)$. For example, one could minimise the average (squared) L_2 error:

$$\begin{aligned} L(\{(\alpha^{(h)\ell}, q^{(h)\ell})\}_{\ell=1}^{M}|\theta) &= \frac{1}{N_{\text{samples}}} \sum_{\ell=1}^{N_{\text{samples}}} L_2(\alpha^{(h)\ell}, q^{(h)\ell}|\theta) \\ L_2(\alpha^{(h)}, q^{(h)}|\theta) &:= \left(\Phi_{\text{NN}}(\alpha^{(h)}|\theta) - q^{(h)}\right)^2 \end{aligned} \tag{31}$$

If trained properly, Φ_{NN} with optimal parameters θ_{opt} will give us a good approximation of the quantity of interest for any $\alpha^{(h)} \in \mathbb{R}^n$:

$$\Phi_{\text{NN}}(\alpha^{(h)}|\theta_{\text{opt}}) \approx q^{(h)} \tag{32}$$

where $q^{(h)} = Q^{(h)}(u^{(h)})$ and $u^{(h)}$ is the solution of $A^{(h)}(\alpha^{(h)})u^{(h)} = \widetilde{f}^{(h)}$. The problem with this approach is that the neural network is a black box.

3.3 Hybrid approach: Physics Enhanced Deep Surrogates (PEDS)

The following approach is described in [1]:

1. Generate a training data set $\{(\alpha^{(h)\ell}, q^{(h)\ell})\}_{\ell=1}^{N_{\text{samples}}}$ as in Step 1 of the Pure ML approach in Section 3.2.

2. Construct a second grid Ω_H with grid spacing $H = L/M$ consisting of $N < n$ cells; in $d = 1$ we have that $n = m$, $N = M$ whereas in $d = 2$ this becomes $n = m^2$, $N = M^2$. We usually want to choose $H \gg h$.

3. Define the downsampling operator

$$R : \mathbb{R}^{\widetilde{n}} \to \mathbb{R}^{\widetilde{N}}, \qquad \alpha^{(h)} \mapsto \alpha^{(H)}_{\text{DS}} \tag{33}$$

which takes a high-resolution $\alpha^{(h)}$ and turns it into a low-fidelity downsampled $\alpha^{(H)}_{\text{DS}}$. If $H = \mu h$ where the scale parameter $\mu = H/h = m/M \in \mathbb{N}$ is an integer, then we can for example simply set

$$\begin{aligned} (\alpha^{(H)}_{\text{DS}})_j &= \alpha^{(h)}_{\mu\cdot j} \qquad \text{for } j = 0, 1, \ldots, M \quad \text{in } d = 1 \text{ dimensons}, \\ (\alpha^{(H)}_{\text{DS}})_{j,k} &= \alpha^{(h)}_{\mu\cdot j,\mu\cdot k} \qquad \text{for } j, k = 0, 1, \ldots, M \quad \text{in } d = 2 \text{ dimensons}. \end{aligned} \tag{34}$$

4. Construct a neural network Ψ_{NN} parametrised by some parameters θ (which have nothing to do with the NN parameters in Section 3.2)

$$\Psi_{\text{NN}}(\cdot|\theta) : \mathbb{R}^n \to \mathbb{R}^N. \tag{35}$$

Ψ_{NN} maps $\alpha^{(h)}$ to some generated $\alpha^{(H)}_{\text{NN}}$ on the low-resolution grid.

5. Define some function $A : \mathbb{R}^{\widetilde{N}} \times \mathbb{R}^{\widetilde{N}} \times [0,1] \to \mathbb{R}^{\widetilde{N}}$ for combining the downsampled $\alpha_{\text{DS}}^{(H)}$ and the generated $\alpha_{\text{NN}}^{(H)}$ into some $\alpha^{(H)} = A(\alpha_{\text{DS}}^{(H)}, \alpha_{\text{NN}}^{(H)}, w)$ where w is a learnable parameter. For example, one might choose one of the following two weighted means

$$\alpha_j^{(H)} = w(\alpha_{\text{NN}}^{(H)})_j + (1-w)(\alpha_{\text{DS}}^{(H)})_j \qquad \text{for } j = 0, 1, \dots, M \tag{36}$$

or

$$\alpha_j^{(H)} = \log\Big(w \exp[(\alpha_{\text{NN}}^{(H)})_j] + (1-w)\exp[(\alpha_{\text{DS}}^{(H)})_j]\Big) \qquad \text{for } j = 0, 1, \dots, M \tag{37}$$

in $d = 1$ dimensions with the obvious corresponding expressions in $d = 2$ dimensions.

6. For given neural network parameters θ and weight w define the function $\mathcal{F}_{\text{PEDS}}^{(h)}(\cdot|\theta, w) : \mathbb{R}^{\widetilde{n}} \to \mathbb{R}^{\widetilde{N}}$. This function maps a high-resolution $\alpha^{(h)}$ to the low fidelity solution $u^{(H)} = u^{(H)}(\alpha^{(h)})$ as follows:
   - Compute $\alpha_{\text{DS}}^{(H)}$ and $\alpha_{\text{NN}}^{(H)}$ and combine them into $\alpha^{(H)}$ as described above.
   - Compute the low-fidelity solution $u^{(H)} \in \mathbb{R}^N$ by solving $A^{(H)}(\alpha^{(H)})u^{(H)} = \widetilde{f}^{(H)}$

7. Find optimal θ_{opt} and w_{opt} by minimising some loss function $L(\{(\alpha^{(h)\ell}, q^{(h)\ell})\}_{\ell=1}^{N_{\text{samples}}}|\theta)$. Again, the natural choice is to minimise the average (squared) L_2 error:

$$\begin{aligned}
L(\{(\alpha^{(h)\ell}, q^{(h)\ell})\}_{\ell=1}^{M}|\theta) &= \frac{1}{N_{\text{samples}}} \sum_{\ell=1}^{N_{\text{samples}}} L_2(\alpha^{(h)\ell}, q^{(h)\ell}|\theta) \\
L_2(\alpha^{(h)}, q^{(h)}|\theta) &:= \Big(Q^{(H)}\Big(\mathcal{F}_{\text{PEDS}}(\alpha^{(h)}|\theta, w)\Big) - q^{(h)}\Big)^2
\end{aligned} \tag{38}$$

If trained properly, $\mathcal{F}_{\text{PEDS}}$ with optimal parameters θ_{opt}, w_{opt} will give us a good approximation of the quantity of interest for any $\alpha^{(h)} \in \mathbb{R}^n$:

$$Q^{(H)}\Big(\mathcal{F}_{\text{PEDS}}^{(h)}(\alpha^{(h)}|\theta_{\text{opt}}, w_{\text{opt}})\Big) \approx q^{(h)} \tag{39}$$

where $q^{(h)} = Q^{(h)}(u^{(h)})$ and $u^{(h)}$ is the solution of $A^{(h)}(\alpha^{(h)})u^{(h)} = \widetilde{f}^{(h)}$.

A Gradients

We need to back-propagate through the solution, i.e. the map $\alpha \mapsto u$ defined by

$$\begin{aligned}
A(\alpha)u &= \widetilde{f} \\
\Leftrightarrow \qquad u &= A(\alpha)^{-1}\widetilde{f}
\end{aligned} \tag{40}$$

where for simplicity we have dropped all superscripts relating to the grid spacing H.

A.1 One-dimensional problem

The chain rule gives:

$$\frac{\partial L}{\partial \alpha_i} = \sum_j \frac{\partial u_j}{\partial \alpha_i}\frac{\partial L}{\partial u_j} \tag{41}$$

which leads to

$$\begin{aligned}
\frac{\partial u_j}{\partial \alpha_i} &= \sum_k \left(\frac{\partial}{\partial \alpha_i} A(\alpha)^{-1}\right)_{jk} \widetilde{f}_k \\
&= -\Big(A(\alpha)^{-1} D_i(\alpha) A(\alpha)^{-1}\widetilde{f}\Big)_j \qquad \text{with } D_i(\alpha) := \frac{\partial A}{\partial \alpha_i}(\alpha).
\end{aligned} \tag{42}$$

The matrix $D_i(\alpha)$ contains mostly zero entries. For $i=0$ it has exactly one non-zero entry, namely $(D_0(\alpha))_{0,0} = 2h^{-2}\exp[\alpha_0]$. For $1 < i < N$ it has exactly four non-zero entries namely $(D_i(\alpha))_{i,i} = (D_i(\alpha))_{i-1,i-1} = -(D_i(\alpha))_{i,i-1} = -(D_i(\alpha))_{i-1,i} = h^{-2}\exp[\alpha_i]$.

Further, we can write:

$$\frac{\partial L}{\partial \alpha_i} = -w^\top D_i(\alpha) u \tag{43}$$

with

$$A(\alpha)w = \frac{\partial L}{\partial u}, \qquad A(\alpha)u = \widetilde{f} \tag{44}$$

A.2 Two-dimensional problem

In two dimensions we can write

$$\frac{\partial L}{\partial \alpha^{(h)}_{rs}} = -\sum_{j,k}\sum_{j',k'} w_{jk}(D_{rs}(\alpha^{(h)}))_{jk,j'k'}u_{j'k'} \tag{45}$$

where again u and w are obtained by solving the two-dimensional equivalent of (44). The sparse tensor $D(\alpha^{(h)})$ is given by

$$(D_{rs}(\alpha^{(h)}))_{jk,j'k'} = \frac{A(\alpha^{(h)})_{\nu(j,k)\nu(j',k')}}{\partial \alpha^{(h)}_{r,s}} \tag{46}$$

From (21) and (24) we can work out the non-zero entries of $D(\alpha^{(h)})$:

$$\begin{aligned}
(D_{jk})_{jk,jk} &= \frac{1}{2h^2}\left(2\Theta_0^1(j)\Theta_0^m(k)K^{(h)}_{j,k+\frac{1}{2}} + \Theta_1^m(j)\Theta_0^m(k)K^{(h)}_{j,k+\frac{1}{2}} + \Theta_0^m(j)\Theta_1^m(k)K^{(h)}_{j+\frac{1}{2},k}\right)\\
(D_{j+1,k})_{jk,jk} &= \frac{1}{2h^2}\left(\Theta_0^{m-1}(j)\Theta_0^m(k)K^{(h)}_{j+1,k+\frac{1}{2}} + \Theta_0^m(j)\Theta_1^m(k)K^{(h)}_{j+\frac{1}{2},k}\right)\\
(D_{j,k+1})_{jk,jk} &= \frac{1}{2h^2}\left(\Theta_0^1(j)\Theta_0^m(k)K^{(h)}_{j,k+\frac{1}{2}} + \Theta_1^m(j)\Theta_0^m(k)K^{(h)}_{j,k+\frac{1}{2}} + \Theta_0^m(j)\Theta_0^{m-1}(k)K^{(h)}_{j+\frac{1}{2},k+1}\right)\\
(D_{j+1,k+1})_{jk,jk} &= \frac{1}{2h^2}\left(\Theta_0^{m-1}(j)\Theta_0^m(k)K^{(h)}_{j+1,k+\frac{1}{2}} + \Theta_0^m(j)\Theta_0^{m-1}(k)K^{(h)}_{j+\frac{1}{2},k+1}\right)\\
(D_{j,k})_{jk,j-1,k} = (D_{j,k+1})_{jk,j-1,k} &= -\frac{1}{2h^2}\Theta_1^m(j)\Theta_0^m(k)K^{(h)}_{j,k+\frac{1}{2}}\\
(D_{j+1,k})_{jk,j+1,k} = (D_{j+1,k+1})_{jk,j+1,k} &= -\frac{1}{2h^2}\Theta_0^{m-1}(j)\Theta_0^m(k)K^{(h)}_{j+1,k+\frac{1}{2}}\\
(D_{jk})_{jk,j,k-1} = (D_{j+1,k})_{jk,j,k-1} &= -\frac{1}{2h^2}\Theta_0^m(j)\Theta_1^m(k)K^{(h)}_{j+\frac{1}{2},k}\\
(D_{j,k+1})_{jk,j,k+1} = (D_{j+1,k+1})_{jk,j,k+1} &= -\frac{1}{2h^2}\Theta_0^m(j)\Theta_0^{m-1}(k)K^{(h)}_{j+\frac{1}{2},k+1}
\end{aligned} \tag{47}$$

For each index pair (r,s) we have

$$\begin{aligned}
\frac{\partial L}{\partial \alpha^{(h)}_{rs}} = &-F^{(x)}_{r,s}\left(w_{r,s-1}u_{r,s-1} + w_{r-1,s-1}u_{r-1,s-1} - w_{r,s-1}u_{r-1,s-1} - w_{r-1,s-1}u_{r,s-1}\right)\\
&- F^{(x)}_{r,s+1}\left(w_{r,s}u_{r,s} + w_{r-1,s}u_{r-1,s} - w_{r,s}u_{r-1,s} - w_{r-1,s}u_{r,s}\right)\\
&- F^{(y)}_{r,s}\left(w_{r-1,s}u_{r-1,s} + w_{r-1,s-1}u_{r-1,s-1} - w_{r-1,s}u_{r-1,s-1} - w_{r-1,s-1}u_{r-1,s}\right)\\
&- F^{(y)}_{r+1,s}\left(w_{r,s}u_{r,s} + w_{r,s-1}u_{r,s-1} - w_{r,s}u_{r,s-1} - w_{r,s-1}u_{r,s}\right)\\
&- 2F^{(y,0)}_{r,s}w_{r,s-1}u_{r,s-1}\\
&- 2F^{(y,0)}_{r,s+1}w_{r,s}u_{r,s}
\end{aligned} \tag{48}$$

The coefficients $F^{(x)}_{r,s}$, $F^{(y)}_{r,s}$ and $F^{(y,0)}_{r,s}$ that appear in (48) are

$$
\begin{aligned}
F^{(x)}_{r,s} &:= \frac{1}{2h^2}\Theta_1^m(r)\Theta_1^{m+1}(s)K^{(h)}_{r,s-\frac{1}{2}} = \begin{cases}\frac{1}{2h^2}K^{(h)}_{r,s-\frac{1}{2}} & \text{for } 1\le r\le m-1,\ 1\le s\le m\\ 0 & \text{otherwise}\end{cases}\\
F^{(y)}_{r,s} &:= \frac{1}{2h^2}\Theta_1^{m+1}(r)\Theta_1^{m}(s)K^{(h)}_{r-\frac{1}{2},s} = \begin{cases}\frac{1}{2h^2}K^{(h)}_{r-\frac{1}{2},s} & \text{for } 1\le r\le m,\ 1\le s\le m-1\\ 0 & \text{otherwise}\end{cases}\\
F^{(y,0)}_{r,s} &:= \frac{1}{2h^2}\Theta_0^1(r)\Theta_1^{m+1}(s)K^{(h)}_{r,s-\frac{1}{2}} = \begin{cases}\frac{1}{2h^2}K^{(h)}_{r,s-\frac{1}{2}} & \text{for } r=0,\ 1\le s\le m\\ 0 & \text{otherwise}\end{cases}
\end{aligned}
\tag{49}
$$

B Distribution of diffusion coefficient

We assume that the logarithm $\alpha(x)$ of the diffusion coefficient is a Gaussian Random Field (GRF) which satisfies

$$
(-\Delta+\kappa^2)^{a/2}\alpha(x) = \mathcal{W}(x) \tag{50}
$$

where $\mathcal{W}$ is white noise with $\mathbb{E}[\mathcal{W}(x)]=0$, $\mathbb{E}[\mathcal{W}(x)\mathcal{W}(y)]=\delta(x-y)$. $\kappa^{-1} =: \Lambda$ is the correlation length. We have trivially that $\mathbb{E}[\alpha(x)]=0$. If the field was defined on the entire real line then its covariance would be given by the Matern function

$$
\mathrm{Cov}[\alpha(x)\alpha(y)] = \mathbb{E}[\alpha(x)\alpha(y)] = C_{\nu,\kappa}(\kappa|x-y|) := \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}\left(\kappa|x-y|\right)^{\nu}K_\nu(\kappa|x-y|) \tag{51}
$$

where $a=\nu+\frac{1}{2}$ and K_ν is the modified Bessel function of the second kind [2]. The marginal covariance is given by

$$
\sigma^2 = \sigma^2(\nu,\kappa) = \frac{\Gamma(\nu)}{\Gamma(\nu+\frac{1}{2})\sqrt{4\pi}\kappa^{2\nu}} \tag{52}
$$

Here we only consider the two cases $a=1$, $\nu=\frac{1}{2}$ and $a=2$, $\nu=\frac{3}{2}$ for which the Matern function $C_{\nu,\kappa}$ reduces to

$$
\begin{aligned}
C_{\nu=\frac{1}{2},\kappa}(\kappa|x-y|) &= \frac{1}{2\kappa}\exp[-\kappa|x-y|],\\
C_{\nu=\frac{3}{2},\kappa}(\kappa|x-y|) &= \frac{1}{4\kappa^3}\left(1+\kappa|x-y|\right)\exp[-\kappa|x-y|].
\end{aligned}
\tag{53}
$$

For the finite domain $\Omega=[0,L]$ we need to impose boundary conditions which we choose to be $\frac{d\alpha}{dx}(0) = \frac{d\alpha}{dx}(L)=0$. To generate samples of $\alpha(x)$ in practice, we divide the domain into n cells as above and sample the vector $\alpha^{(h)} = (\alpha(x_0),\alpha(x_1),\dots,\alpha(x_n))\in\mathbb{R}^{n+1}$ at the vertices x_j. A finite difference discretisation of the shifted Laplace operator $-\Delta+\kappa^2$ is given by the tridiagonal matrix

$$
Q^{(h)} = \frac{1}{h^2}\begin{pmatrix}
2+\kappa^2h^2 & -1 & 0 & 0 & 0 & \cdots & 0\\
-1 & 2+\kappa^2h^2 & -1 & 0 & 0 & \cdots & 0\\
0 & -1 & 2+\kappa^2h^2 & -1 & 0 & \cdots & 0\\
0 & 0 & -1 & \ddots & & & \\
0 & 0 & 0 & & & & \vdots\\
\vdots & \vdots & \vdots & & \ddots & -1 & 0\\
 & & & & -1 & 2+\kappa^2h^2 & -1\\
0 & 0 & 0 & \cdots & 0 & -1 & 2+\kappa^2h^2
\end{pmatrix}. \tag{54}
$$

Let $\xi^{(h)} = (\xi_0,\xi_1,\dots,\xi_n)\in\mathbb{R}^{n+1}$ be a vector of independent and identically distributed (i.i.d.) Gaussian random variables $\xi_j\sim\mathcal{N}(0,h^{-1})$ for $j=0,1,\dots,n$; we also write $\xi^{(h)}\sim\pi_0^{(h)}$. For $a=1$, $\nu=\frac{3}{2}$ we can

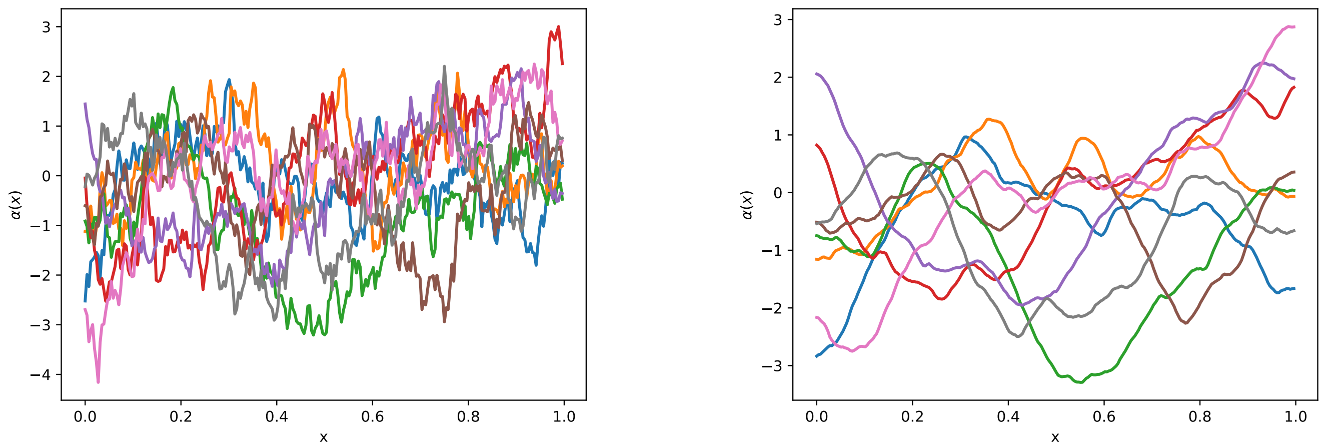

Figure 1: Random samples $\alpha(x)$ drawn from $\pi^{(h)}(\cdot|\nu,\kappa)$ for $a=1$, $\nu=\frac{1}{2}$ (left) and $a=2$, $\nu=\frac{3}{2}$ (right). In both cases the correlation length is $\kappa^{-1}=0.1$ and the number of gridcells is $N=256$. The samples are scaled by σ^{-1} to make their variance ~ 1.

generate a sample $\alpha^{(h)k}$ from the target distribution $\pi^{(h)}(\cdot|\nu,\kappa)$ by drawing $\xi^{(h)k}\sim\pi_0^{(h)}$ and solving the tridiagonal system $Q^{(h)}\alpha^{(h)k}=\xi^{(h)k}$ for $\alpha^{(h)k}$. For $a=1$, $\nu=\frac{1}{2}$ we Cholesky factorise $Q^{(h)}=L^{(h)}(L^{(h)})^\top$ and solve the bi-diagonal system $(L^{(h)})^\top\alpha^{(h)k}=\xi^{(h)k}$. Observe that the approach is readily extended to arbitrary integer $a>1$. Fig. 1 showns some random samples generates in this way for both $a=1$, $\nu=\frac{1}{2}$ and $a=2$, $\nu=\frac{3}{2}$. The corresponding covariance between a and $y=\frac{3}{4}$ is visualised in Fig. 2 where we show both an estimator for the covariance on the interval $\Omega=[0,1]$ and the Matern covariance function. The two functins differ since von Neumann boundary conditions have been imposed to generate the samples.

B.1 Two-dimensional case

In $d=2$ dimensions we only consider the case $a=2$ which, since $a=\nu+\frac{d}{2}$, corresponds to $\nu=1$. The Matern function is

$$C_{\nu=1,\kappa}(\kappa||x-y||)=\sigma^2\kappa||x-y||K_1(\kappa||x-y||) \tag{55}$$

with the marginal covariance

$$\sigma^2=\sigma^2(\kappa)=\frac{1}{4\pi\kappa^2}. \tag{56}$$

To draw samples from the distribution, we start by generating a vector $\xi^{(h)}=(\xi_0,\dots,\xi_{(n+1)^2-1})\in\mathbb{R}^{(n+1)^2}$ where $\xi_j\sim\mathcal{N}(0,h^{-1})$. Next, we solve the system $Q^{(h)}\alpha^{(h)}=\xi^{(h)}$ where $Q^{(h)}$ is a finite difference discretisation of the shifted Laplace operator $-\Delta+\kappa^2$ with homogeneous von Neumann boundary conditions. The resulting $\alpha^{(h)}$ is a sample from the desired distribution. Fig. 3 shows a selection of random samples generated with this method. As shown in Fig. 4 we also verified that the empirical covariance closely matches the Matern covariance function for small distances $||x-x_0||$.

References

[1] Raphaël Pestourie, Youssef Mroueh, Chris Rackauckas, Payel Das, and Steven G Johnson. Physics-enhanced deep surrogates for partial differential equations. *Nature Machine Intelligence*, 5(12):1458–1465, 2023.

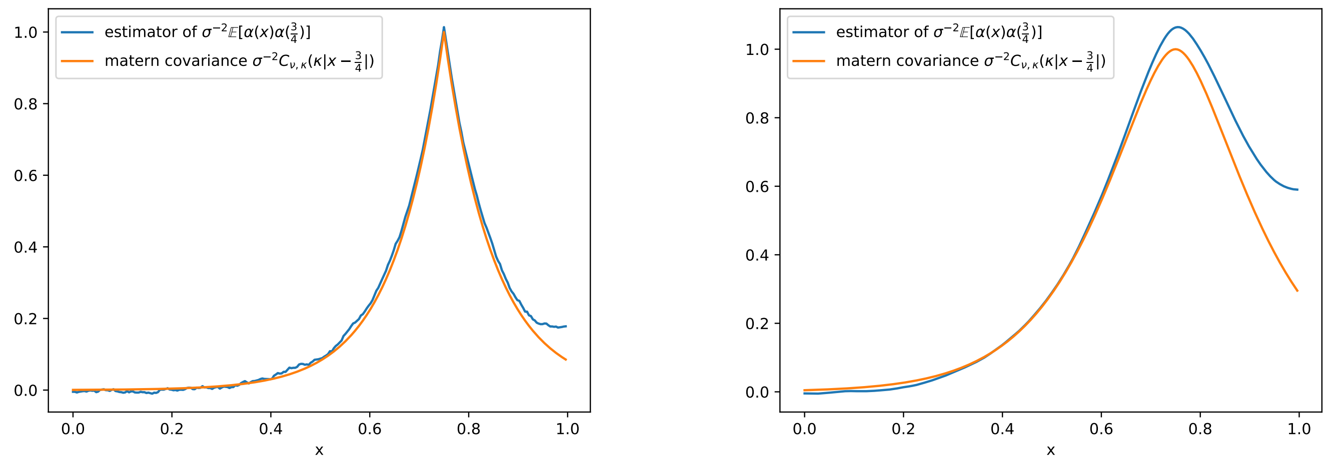

Figure 2: Empirical covariance between the points x and $y = \frac{3}{4}$ and Matern covariance function for one-dimensional sampling. Results are shown for $a = 1$, $\nu = \frac{1}{2}$ (left) and $a = 2$, $\nu = \frac{3}{2}$ (right). In both cases the correlation length is $\kappa^{-1} = 0.1$ and the number of gridcells is $N = 256$. The samples are scaled by σ^{-1} to make their variance ~ 1.

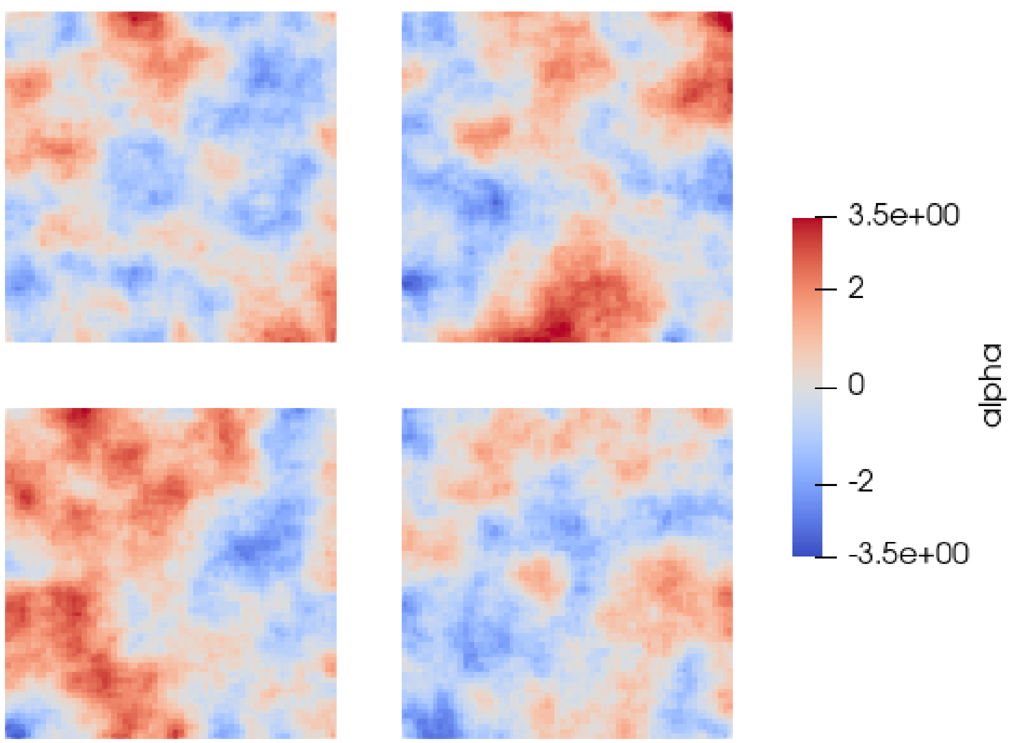

Figure 3: Visualisation of four random samples $\alpha^{(h)}$ drawn from the 2d distribution for $\Lambda = 0.1$ and a 64×64 grid. Samples are rescaled by σ^{-1} to make their variance ~ 1.

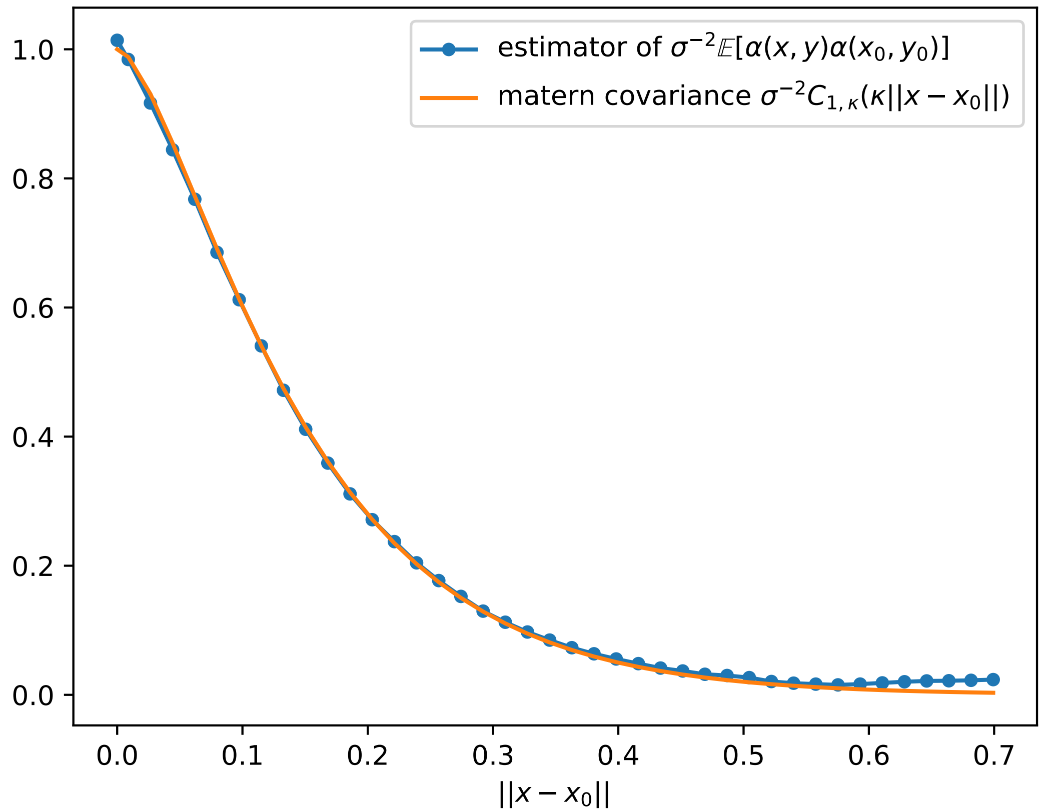

Figure 4: Empirical covariance function between the point $x_0 = (\frac{1}{2}, \frac{1}{2})$ and all other points x of the domain. Results are shown for $\Lambda = 0.1$ and a 64×64 grid and 10000 samples were used for the estimator. Samples are rescaled by σ^{-1} to make their variance ~ 1.

[2] Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):423–498, 2011.