JAX-DIPS: Neuro-symbolic bootstrapping method for solving elliptic problems with discontinuities across irregular interfaces

Pouria A. Mistani* [†,a], Samira Pakravan[†,b], Rajesh Ilango[a], Frederic G. Gibou[b]

[a] *NVIDIA Corporation, Santa Clara, CA 95051, USA*
[b] *University of California, Santa Barbara, CA 93106-5070, USA*

Abstract

We present a scalable strategy for development of mesh-free hybrid neuro-symbolic partial differential equation solvers based on existing mesh-based numerical discretization methods. Particularly, this strategy can be used to efficiently train neural network surrogate models for the solution of partial differential equations while retaining the accuracy and convergence properties of the state-of-the-art numerical solvers. The presented neuro-symbolic bootstrapping method (hereby dubbed NSBM) is based on evaluation of the finite discretization residuals of the PDE system obtained on implicit Cartesian cells centered on a set of random collocation points with respect to trainable parameters of the neural network. We apply NSBM to the challenging class of elliptic problems with jump conditions across irregular sharp interfaces in three spatial dimensions. We show the method is convergent such that model accuracy improves by increasing number of collocation points in the domain. The algorithms presented here are implemented and released in a software package named `JAX-DIPS` (https://github.com/JAX-DIPS/JAX-DIPS), standing for differentiable interfacial PDE solver. `JAX-DIPS` is purely developed in `JAX`, offering end-to-end differentiability from mesh generation to the higher level discretization abstractions, geometric integrations, and interpolations, thus facilitating research into use of differentiable algorithms for developing hybrid PDE solvers.

Keywords: level-set method, free boundary problems, surrogate models, jump conditions, differentiable programming, neural networks

1. Introduction

1.1. Problem statement

Consider a closed irregular interface (Γ) that partitions the computational domain (Ω) into interior (Ω^-) and exterior (Ω^+) subdomains; *i.e.*, $\Omega = \Omega^- \cup \Gamma \cup \Omega^+$. We are interested in the solutions $\mathrm{u}^\pm \in \Omega^\pm$ to the following class of linear elliptic problems in $\mathbf{x} \in \Omega^\pm$:

$$\begin{aligned} k^\pm u^\pm - \nabla \cdot (\mu^\pm \nabla u^\pm) &= f^\pm, && \mathbf{x} \in \Omega^\pm \\ [u] &= \alpha, && \mathbf{x} \in \Gamma \\ [\mu \partial_{\mathbf{n}} u] &= \beta, && \mathbf{x} \in \Gamma \end{aligned}$$

Here $f^\pm = f(\mathbf{x} \in \Omega^\pm)$ is the spatially varying source term, $\mu^\pm = \mu(\mathbf{x} \in \Omega^\pm)$ are the diffusion coefficients, and $k^\pm$ are the reaction coefficients in the two domains. We consider Dirichlet boundary conditions in a cubic domain $\Omega = [-\mathrm{L}/2, \mathrm{L}/2]^3$.

This class of problems arise ubiquitously in describing diffusion dominated processes in physical systems and life sciences where sharp and irregular interfaces regulate transport across regions with different properties. Examples include Poisson-Boltzmann equation for describing electrostatic

*Corresponding author: p.a.mistani@gmail.com
†These authors contributed equally to this work

properties of membranes, colloids and solvated biomolecules with jump in dielectric permittivities [55, 45], electroporation of cell aggregates with nonlinear membrane jump conditions [47], epitaxial growth in fabrication of opto-electronic devices where atomic islands grow by surface diffusion of adatoms across freely moving interfaces [46], solidification of multicomponent alloys used for manufacturing processes with free interfaces separating different phases of matter [60, 10], directed self-assembly of diblock copolymers for next generation lithography [21, 50, 9], multiphase flows with and without phase change, and Poisson-Nernst-Planck equations for electrokinetics. Much of these processes are multiscale and the changes across interfaces must be mathematically modeled and numerically solved as sharp surfaces. Smoothing strategies introduce unphysical characteristics in the solution and lead to systemic errors.

1.2. Literature on relevant finite discretization methods

Several numerical methods have been proposed for accurate solution of this class of problems based on explicit or implicit representation of the interface. Finite element methods rely on explicit meshing of the surface that poses severe challenges [2, 12]. Implicit methods include the Immersed Interface Method (IIM) [35] and its variants [1, 39, 19, 23] that rely on Taylor expansions of the solution on both sides of the interface and modifying the local stencils to impose the jump conditions. The main challenge is evaluating high order jump conditions and surface derivatives along interface. Another method is the Ghost Fluid Method (GFM) [20] that was originally introduced to approximate two-phase compressible flows and later applied to the Poisson problem with jump conditions [40]. The basic idea is to define fictitious fluid regions across the discontinuities by adding jump conditions to the true fluid. While GFM captures the normal jump in solution accurately, the tangential jump is smeared. This was solved by the Voronoi Interface Method (VIM) [24] by applying the GFM treatment on a local Voronoi mesh by adapting a local Cartesian mesh which introduces numerical challenges. Several other approaches include the cut-cell method [16], discontinuous Galerkin and eXtended Finite Element Method (XFEM) [36, 48, 3] among others.

In this work we bootstrap the level-set based finite volume method on Cartesian grids proposed by Bochkov & Gibou (2020) [7]. This method is based on the idea of Taylor expansions in the normal direction and employing one-sided least-square interpolations for imposing jump conditions. In particular, this method offers second order accurate numerical solutions with first order accurate gradients in the L^∞-norm.

1.3. Literature on solving PDEs with neural networks

Since early 1990s, artificial neural networks have been used for solving partial differential equations by (i) mapping the algebraic operations of the discretized PDE systems onto specialized neural network architectures and minimizing the network energy, or (ii) treating the whole neural network as the basic approximation unit whose parameters are adjusted to minimize a specialized error function that includes the differential equation itself with its boundary/initial conditions.

In the first category, neurons output the discretized solution values over a set number of grid points and minimizing the network energy drives the neuronal values towards the solution of the linear system at the mesh points. In this case, the neural network energy is the residual of the finite discretization method summed over all neurons of the network [34]. Although the convergence properties of the finite discretization methods gaurantee and control quality of the obtained solutions, the computational costs grow by increasing resolution and dimensionality. Interestingly, due to regular and sparse structure of the finite discretizations, such locally connected neural network PDE solvers have been implemented on VLSI analog CMOS circuits [22, 15, 14].

The second strategy proposed by Lagaris *et al.* [33] relies on the function approximation capabilities of the neural networks. Encoding the solution everywhere in the domain within a neural network offers a mesh-free, compact, and memory efficient surrogate model for the solution function that can be utilized in subsequent inference tasks. This method has recently re-emerged as the physics-informed neural networks (PINNs) [53] and is widely used. Despite their advantages, these methods lack controllable accuracy and convergence properties of finite discretization methods.

Pursuit of hybrid solvers aims at leveraging the performance gains of neural network inference on modern accelerated hardware with the gauranteed accuracy of finite discretization methods. The hybridization efforts are algorithmic or architectural.

One important algorithmic method is the deep Galerkin method (DGM) [57] that is a neural network extension of the mesh-free Galerkin method where the solution is represented as a deep neural network rather than a linear combination of basis functions. The mesh-free nature of DGM, that stems from the underlying mesh-free Galerkin method, enables solving problems in higher dimensions by training the neural network model to satisfy the PDE operator and its initial and boundary conditions on a randomly sampled set of points rather than on an exponentially large grid. Although the number of points is huge in higher dimensions, the algorithm can process training on smaller batches of data points sequentially. Besides, second order derivatives in PDEs are calculated by a Monte Carlo method that retain scaling to higher dimensions. Another important algorithmic method is the deep Ritz method for solving variational problems [63] that implements a deep nerual network approximation of the trial function that is constrained by numerical quadrature rule for the variational functional, followed by stochastic gradient descent.

Architectural hybridization methods are based on differentiable numerical linear algebra. One emerging class involves implementing differentiable finite discretization solvers and embedding them in the neural network architectures that enable application of end-to-end differentiable gradient based optimization methods. Recently, differentiable solvers have been developed in JAX [11] for fluid dynamic problems, such as `Phi-Flow` [28], `JAX-CFD` [31], and `JAX-FLUIDS` [5]. These methods are suitable for inverse problems where an unknown field is modeled by the neural network, while the model influence is propagated by the differentiable solver into a measurable residual [51, 17, 42]. We also note the classic strategy for solving inverse problems is the adjoint method to obtain the gradient of the loss without differentiation across the solver [4]; however, deriving analytic expression for the adjoint equations is tedious, should be repeated after modification of the problem or its loss function, and can become impractical for multiphysics problems. Other important utilities of differentiable solvers are to model and correct for the solution errors of finite discretization methods [61], learning and controling PDE systems [18, 27].

Neural networks are not only universal approximators of continuous functions, but also of nonlinear operators [13]. Although this fact has been leveraged using data-driven strategies for learning differential operators by many authors [41, 6, 38, 37], current authors have demonstrated utility of differentiable solvers to effectively train nonlinear operators without any data in a completely physics-driven fashion, see section on learning the inverse transforms in [51].

1.4. Our contributions

In this work we propose a novel algorithm for solving PDEs based on deep neural networks by lifting any existing mesh-based finite discretization method off of its underlying grid and extend it into a mesh-free method that can be applied to high dimensional problems on unstructured random points in an embarrasingly parallel format. We call our method Neuro-Symbolic Bootstrapping Method (NSBM). Bootstrapping is a technique by which a simple computer program activates a more complicated system of programs.

JAX-DIPS Poisson problem solver provides the two sides of $Au = b$ which is obtained after discretization of the governing PDE problem over a uniform 3D grid. Two solution schemes for the forwards problem are possible: (1) using usual iterative methods having left-hand-side and right-hand-side of the PDE discretization, and (2) using autodifferentiation across loss function (difference between lhs and rhs in 2-norm) with respect to a given estimate for the solution vector on the underlying grid $\mathrm{u_{ijk}}$. Both approaches focus on minimizing the residual over the grid points

$$\min_{u_{ijk}} ||Au - b||_2^2$$

The first method attempts to span the residual space in an iterative fashion by *estimating* the gradient of the minimizing function, while the second method offered in JAX-DIPS is directly computing the exact gradient of the minimizing function with respect to the current estimate for solution. Therefore, the advantage of this method is fewer iterations and faster convergence specially

for irregular geometries where the condition number of the linear system leads to much difficulties that need to be resolved by complex preconditioning and increased number of iterations.

We designed `JAX-DIPS` as a modular open source library for the research community on the level-set methods for solving interfacial PDE problems. We invite the community to contribute to this library for further development of novel numerical algorithms for solving both forward and inverse interfacial PDE problems.

2. Dynamic Geometry Representation

A differentiable solver for free boundary problems requires a rich data structure capable to handle several geometric operations over a given mesh. We understand the mesh as a *topological* relationships between cells that discretize a domain, the geometric information of this discretization as well as the relative configuration of the interfaces with respect to the cells need to be computed.

In `JAX-DIPS` we have implemented a uniform grid that supports operations such as interpolations, interface advection, integrations over interfaces as well as in domains. We describe the numerical algorithms used in this section.

2.1. Interpolation methods

To support finite discretization methods for the free boundary problems (namely solver, level-set method, and geometric integrations) a widely needed building block is the ability to interpolate field values anywhere inside a grid cell given the values on the grid points. In `JAX-DIPS` we currently implement two types of interpolation schemes that have been used in the context of the level set method for achieving second-order accurate solutions by Min & Gibou (2007a)[44]: (i) trilinear interpolation, and (ii) quadratic non-oscillatory interpolation.

2.1.1. Trilinear interpolation

In a unit grid cell, rescaled to $\mathcal{C} \in [0,1]^3$, the trilinear interpolation at a point $(x,y,z) \in \mathcal{C}$ uses the grid values on the parent cell vertices according to equation 11 of [44],

$$\phi(x,y,z) = \sum_{i,j,k\in 0,1} \phi(i,j,k)(-1)^{i+j+k}(1-x-i)(1-y-j)(1-z-k)$$

Trilinear interpolation is based on polynomials of order 1 and offers accuracy of order 2 using 8 immediate vertices in a grid cell.

2.1.2. Quadratic non-oscillatory interpolation

Quadratic interpolation extends the trilinear interpolation by adding second order derivatives of the interpolating field. This is needed because trilinear interpolation is sensitive to presence of discontinuities and kinks which are abundant in the context of free boundary problems. The extension reads

$$\begin{aligned}\phi(x,y,z) = &\sum_{i,j,k\in 0,1} \phi(i,j,k)(-1)^{i+j+k}(1-x-i)(1-y-j)(1-z-k) \\ &- \phi_{xx}\frac{x(1-x)}{2} - \phi_{yy}\frac{y(1-y)}{2} - \phi_{zz}\frac{z(1-z)}{2}\end{aligned}$$

where second order derviatives are sampled as the minimum value on the parent cell vertices to enhance numerical stability of the interpolation

$$\begin{aligned}\phi_{xx} &= \min_{v\in \text{vertices}(\mathcal{C})} |D_{xx}\phi_v| \\ \phi_{yy} &= \min_{v\in \text{vertices}(\mathcal{C})} |D_{yy}\phi_v| \\ \phi_{zz} &= \min_{v\in \text{vertices}(\mathcal{C})} |D_{zz}\phi_v|\end{aligned}$$

The second order derivative operator is the familiar 5-point finite difference stencil.

2.2. Level-set method

The level-set method for solving free boundary problems was introduced by Osher & Sethian (1988) [49]. The free boundary is described as the zero contour of a signed-distance function, ϕ, whose evolution is given by the advection equation according to some velocity field dictated by the physics of the problem, $\mathbf{v}$, that is defined over the moving boundary

$$\frac{\partial \phi}{\partial t} + \mathbf{v} \cdot \nabla \phi = 0 \tag{1}$$

This *implicit* representation of the moving boundaries resolves the need for the challenging task of adapting the underlying grid to the yet-unknown discontinuities in the solution field. The computational simplicity of using Cartesian grids for solving free boundary problems with irregular geometries, as well as the ability to simulate freely moving discontinuities in a *sharp*-manner, that is required by the physics of these problems, are the two main offerings of the level-set method for this class of PDE problems.

The signed-distance property of the level-set function, *i.e.* $|\nabla \phi| = 1$, deteriorates with timestepping the discretized advection equation. This is resolved by solving the Sussman reinitialization equation [59] in fictitious time (τ) every few iterations of the physical time (t),

$$\frac{\partial \phi}{\partial \tau} + sgn(\phi_0)(|\nabla \phi| - 1) = 0 \tag{2}$$

Note the asymptotic solution of the Sussman equation is the signed-distance property of the level-set function. Here $sgn(\phi_0)$ is the sign function evaluated on the level-set function before reinitialization in order to preserve the sign of grid points during the fictitous updates.

Besides implicit representation of the free boundaries, the level-set function can be used to compute normal vectors to the interface

$$\mathbf{n} = \nabla \phi / |\nabla \phi|$$

as well as the curvature of the interface

$$\kappa = \nabla \cdot \mathbf{n}$$

Below we describe numerical methods implementd in `JAX-DIPS` to solve the level-set equations.

2.2.1. Second order accurate semi-Lagrangian advection scheme

Semi-Lagrangian methods are unconditionally stable for solving advection equations, therefore avoiding the restrictive CFL condition on the timestep size from time t^n to t^{n+1}. The general procedure to solve for $\phi^{n+1}(\mathbf{x}^{n+1})$ is to first evaluate for each grid point, $\mathbf{x}^{n+1}$, the departure point, $\mathbf{x}_d$, in the upwind direction along the characteristic curve; then use interpolation methods to recover value of the solution field at the departure point; and finally updating the solution field by setting $\phi^{n+1}(\mathbf{x}^{n+1}) = \phi^n(\mathbf{x}_d)$.

The task of evaluating departure points is treated by the second-order mid-point method [62]

$$\hat{\mathbf{x}} = \mathbf{x}^{n+1} - \frac{\Delta t}{2} \cdot \mathbf{v}^n(\mathbf{x}^{n+1})$$
$$\mathbf{x}_d = \mathbf{x}^{n+1} - \Delta t \cdot \mathbf{v}^{n+\frac{1}{2}}(\hat{\mathbf{x}})$$

The first step is trivial, but the second step demands velocity at the mid-timestep $t^{n+\frac{1}{2}}$ which can be evaluated from the velocity field at the previous two timesteps, t^n and t^{n-1}, according to

$$\mathbf{v}^{n+\frac{1}{2}} = \frac{3}{2}\mathbf{v}^n - \frac{1}{2}\mathbf{v}^{n-1}$$

this intermediate velocity field is then fed into a trilinear interpolation scheme to sample the velocities on the intermediate points $\mathbf{v}^{n+\frac{1}{2}}(\hat{\mathbf{x}})$. At this point departure points can be computed, and the advected solution is updated by sampling a nonoscillatory interpolant of the level-set function at timestep t^n over the computed departure points $\phi^n(\mathbf{x}_d)$.

2.2.2. Godunov Hamiltonian for reinitialization

The Sussman equation 2 is generically discretized over its spatial dimensions to obtain

$$\frac{d\phi}{d\tau} = -sgn(\phi^0)\left[H_G(D_x^+\phi, D_x^-\phi, D_y^+\phi, D_y^-\phi, D_z^+\phi, D_z^-\phi,) - 1\right] \tag{3}$$

where H_G is the so-called Gudonov Hamiltonian defined as

$$H_G(a,b,c,d,e,f) = \sqrt{\sum_{(i,j)\in(a,b),(c,d),(e,f)} \max\left(|\min(s\cdot i, 0)|^2, |\max(s\cdot j, 0)|^2\right)}$$
$$s = sgn(\phi^0)$$

where the one-sided derivatives are computed using second order accurate discretizations in the bulk far from interfaces

$$(D_x^+\phi)_{i,j,k} = \frac{\phi_{i+1,j,k} - \phi_{i,j,k}}{\Delta x} - \frac{\Delta x}{2}\text{minmod}((D_{xx}\phi)_{i,j,k}, (D_{xx}\phi)_{i+1,j,k})$$
$$(D_x^-\phi)_{i,j,k} = \frac{\phi_{i,j,k} - \phi_{i-1,j,k}}{\Delta x} + \frac{\Delta x}{2}\text{minmod}((D_{xx}\phi)_{i,j,k}, (D_{xx}\phi)_{i-1,j,k})$$

where we used $\text{minmod}(a,b) \triangleq \text{median}(a,0,b)$ slope-limiter [56]; and in the vicinity of interfaces these derivatives take into account distance to the interface using the level-set function. With similar results along y and z axes, the derivative along the positive x-axis is given by

$$(D_x^+\phi)_{i,j,k} = \frac{0 - \phi_{i,j,k}}{s_I} - \frac{s_I}{2}c_2^+$$
$$s_I = \frac{\Delta x}{2} + \begin{cases} -c_0^+/c_1^+ & \text{if } |c_2^+| < \epsilon \\ \left(-c_1^+ - sgn(\phi^0_{i,j,k})\sqrt{(c_1^+)^2 - 4c_2^+c_0^+}\right)/(2c_2^+) & \text{if } |c_2^+| \geq \epsilon \end{cases}$$

where

$$c_2^+ = \text{minmod}((D_{xx}\phi)_{i,j,k}, (D_{xx}\phi)_{i+1,j,k})$$
$$c_1^+ = \frac{\phi_{i+1,j,k} - \phi_{i,j,k}}{\Delta x}$$
$$c_0^+ = \frac{\phi_{i+1,j,k} + \phi_{i,j,k}}{2} - \frac{c_2^+(\Delta x)^2}{4}$$

And along the negative x-axis we obtain

$$(D_x^-\phi)_{i,j,k} = \frac{\phi_{i,j,k} - 0}{s_I} + \frac{s_I}{2}c_2^-$$
$$s_I = \frac{\Delta x}{2} + \begin{cases} c_0^-/c_1^- & \text{if } |c_2^-| < \epsilon \\ \left(c_1^- - sgn(\phi^0_{i,j,k})\sqrt{(c_1^-)^2 - 4c_2^-c_0^-}\right)/(2c_2^-) & \text{if } |c_2^-| \geq \epsilon \end{cases}$$

where

$$c_2^- = \text{minmod}((D_{xx}\phi)_{i,j,k}, (D_{xx}\phi)_{i-1,j,k})$$
$$c_1^- = \frac{\phi_{i,j,k} - \phi_{i-1,j,k}}{\Delta x}$$
$$c_0^- = \frac{\phi_{i-1,j,k} + \phi_{i,j,k}}{2} - \frac{c_2^-(\Delta x)^2}{4}$$

Having the right-hand-side, the TVD-RK2 method of Shu & Osher [56] is applied to the update the level-set function in fictitious time

$$\tilde{\phi}^{n+1} = \phi^n - \Delta\tau \cdot sgn(\phi^0)[H_G^n - 1]$$
$$\tilde{\phi}^{n+2} = \tilde{\phi}^{n+1} - \Delta\tau \cdot sgn(\phi^0)[\tilde{H}_G^{n+1} - 1]$$
$$\phi^{n+1} = \frac{\phi^n + \tilde{\phi}^{n+2}}{2}$$

with a timestep chosen by $\Delta\tau = \min(s_I, \Delta x, \Delta y, \Delta z)/3$.

2.3. Geometric operations

2.3.1. Integration over 3D surfaces and volumes

We use uniform Cartesian grids. For computational cells that are crossed by the interface, *i.e.* $\mathcal{V}_{i,j,k} \cap \Gamma \neq 0$, we use the geometric integrations proposed by Min & Gibou (2007b) [43]. In this scheme each grid cell, $\mathcal{C}$, is decomposed into five tetrahedra by the middle-cut triangulation [54] (each cell is rescaled to $[0,1]^3$) that are described below (also see figure 1 (right) of [43]):

$\mathrm{S}_1 \equiv \mathrm{conv}(\mathrm{P}_{000}; \mathrm{P}_{100}; \mathrm{P}_{010}; \mathrm{P}_{001})$	x = 0 face, y = 0 face, z = 0 face
$\mathrm{S}_2 \equiv \mathrm{conv}(\mathrm{P}_{110}; \mathrm{P}_{100}; \mathrm{P}_{010}; \mathrm{P}_{111})$	x = 1 face, y = 1 face, z = 0 face
$\mathrm{S}_3 \equiv \mathrm{conv}(\mathrm{P}_{101}; \mathrm{P}_{100}; \mathrm{P}_{111}; \mathrm{P}_{001})$	x = 1 face, y = 0 face, z = 1 face
$\mathrm{S}_4 \equiv \mathrm{conv}(\mathrm{P}_{011}; \mathrm{P}_{111}; \mathrm{P}_{010}; \mathrm{P}_{001})$	x = 0 face, y = 1 face, z = 1 face
$\mathrm{S}_5 \equiv \mathrm{conv}(\mathrm{P}_{111}; \mathrm{P}_{100}; \mathrm{P}_{010}; \mathrm{P}_{001})$	no face exposure

Hence each 3D grid cell is the union of 5 tetrahedra (simplices) $\mathcal{C} = \cup_{i=1}^{5} S_i$, where each simplex is identified by the pre-existing vertices of the grid cell (hence not creating new grid points). Given the values of the level set function sampled at these vertices one can compute coordinates of intersection points of the interface with each of the simplices $S_i \cap \Gamma$ as well as the negative domain $S_i \cap \Omega^-$. If $P_0, \cdots, P_3$ are the four vertices of a simplex S, then Γ crosses an edge P_iP_j if and only if $\phi(P_i)\phi(P_j) < 0$ and the intersection point across this edge is given by linear interpolation:

$$P_{ij} = P_j \frac{\phi(P_i)}{\phi(P_i) - \phi(P_j)} - P_i \frac{\phi(P_j)}{\phi(P_i) - \phi(P_j)}$$

Number of negative level-set values on the 4 (in 3D) tetrahedron vertices classifies the specific configuration for intersection between simplex S and the interface through a variable $\eta(\phi, S) = n(P_i | \phi(P_i) < 0)$. In 3D, possible values are $\eta = 0, 1, 2, 3, 4$ that correspond to the four configurations for the intersection cross section enumerated below:

- $S \cap \Gamma$, see table 2 and figure 2 of [43]:
  - $\eta = 0$: tetrahedron (S) is completely in positive domain with no intersection, $S \cap \Gamma = \emptyset$.
  - $\eta = 1$: with a single vertex in negative domain and remaining three in positive domain, the tetrahedron and interface have exactly 3 intersection points, the simplex $S \cap \Gamma$ has exactly 3 vertices; *cf.*, see figure 2 (center) of [43].
  - $\eta = 2$: with two vertices in negative domain and remaining two in positive domain, the cross section has four vertices that is splitted into two 3-vertex simplices; *cf.*, see figure 2 (right) of [43].
  - $\eta = 3$: with one vertex in positive domain and remaining three vertices in negative domain, the cross section has 3 vertices that is computed by inverting the sign of the level-set values on vertices and following the instruction for case $\eta = 1$.
  - $\eta = 4$: tetrahedron is completely in negative domain with no intersection, $S \cap \Gamma = \emptyset$.
- $S \cap \Omega^-$, see table 4 and figure 4 of [43]:

- $\eta = 0$: tetrahedron is completely in positive domain with no intersection, $S \cap \Omega^- = \emptyset$;
- $\eta = 1$: the intersection $S \cap \Omega^-$ is characterized by a single tetrahedron with 4 vertices according to figure 4 (left) of [43]; *i.e.*, one vertex is the negative level-set vertex of the parent tetrahedron and three others are interpolated points over the three edges pertaining to the negative vertex.
- $\eta = 2$: the intersection $S \cap \Omega^-$ is characterized by three tetrahedra with 12 vertices according to figure 4 (center) of [43]. Note that there
- $\eta = 3$: the intersection $S \cap \Omega^-$ is characterized by three tetrahedra with 12 vertices according to figure 4 (right) of [43].
- $\eta = 4$: tetrahedron is completely in negative domain and $S \cap \Omega^- = S$;

Note that although we only need to allocate memory for at most 4 vertices to uniquely identify $S \cap \Gamma$, in `JAX-DIPS` we choose to pre-allocate memory for two 3-vertex simplex data structure per S with a total of 6 vertices to separately store information for the cross section geometry. Similarly for $S \cap \Omega^-$ we pre-allocate memory for a three 4-vertex simplex data structure per S. Altogether, in the current implementation the geometric information of intersection points for each simplex S is expressed in terms of 5 simplicies (2 three-vertex simplices for surface area and 3 four-vertex simplicies for volume) using 18 points; this is an area for future optimization.

Having the intersection points, we compute surface and volume integrals of a given field over the interface Γ and in negative domain Ω^- as a summation of integrals over the identified simplices. For each simplex (with $n = 3$ or $n = 4$ vertices) surface and volume integrals can be numerically computed by having these vertices $P_0, \cdots, P_n$ and the values of the field f at these vertices according to

$$\int_S f dx = \text{vol}(S) \cdot \frac{f(P_0) + \cdots + f(P_n)}{n+1}$$

where

$$\text{vol}(S) = \frac{1}{n!} \left| \det \begin{pmatrix} (P_1 - P_0)\hat{e}_1 & \cdots & (P_n - P_0)\hat{e}_1 \\ \vdots & \ddots & \vdots \\ (P_1 - P_0)\hat{e}_n & \cdots & (P_n - P_0)\hat{e}_n \end{pmatrix} \right|$$

with $\hat{e}_i$ being the i^{th} Cartesian unit basis vector.

2.3.2. Cross sections of interface with grid cell faces

For the numerical discretizations considered in this work we also need the surface areas for simplices created at the intersection of Γ with each of the 6 faces of a grid cell $\mathcal{C}$. In `JAX-DIPS` for each face we reuse two corresponding simplices exposed to that face that were calculated in the geometric integrations module, explicitly:

- $x = 0$ face has contributions from (S_1, S_4)
- $x = 1$ face has contributions from (S_2, S_3)
- $y = 0$ face has contributions from (S_1, S_3)
- $y = 1$ face has contributions from (S_2, S_4)
- $z = 0$ face has contributions from (S_1, S_2)
- $z = 1$ face has contributions from (S_3, S_4)

For each face, we extract vertices from (S_i, S_j)-pair that lie on the considered face and sum the surface areas in (negative domain) contributed from each simplex on that face; therefore, the portion of the face surface area in the positive domain is simply the complementing value $\text{area}^+ = \text{area}_{\text{face}} - \text{area}^-$; *i.e.*, this ensures sum of areas adds up to the exact face area in downstream computations.

3. Differentiable Residual Minimization (DRM) Method

One of the main advantages of neural networks for solving partial differential equations is to provide surrogate models that can be readily evaluated on any point of their input space. For parameterized problems, a free parameter such as the diffusion coefficient could be variable for these problems, in this case having an oracle function that evaluates solutions in near-real-time is an invaluable tool for general engineering optimization workflows.

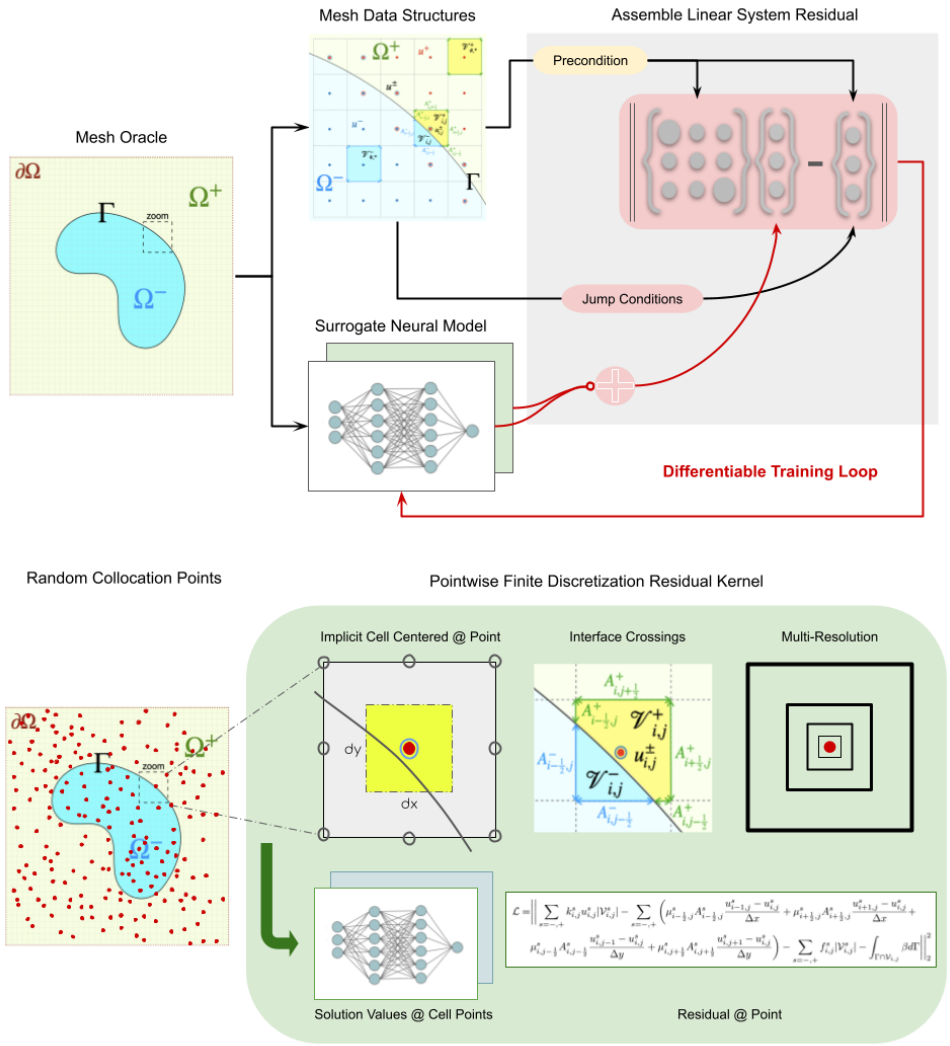

Figure 1: Fully classic numerical solver is used to train neural surrogate models. Training occurs in the finite dimensional space spanned by the finite discretization methods.

3.1. Neural network approximators for the solution

In 1987, Hecht and Nielson [25] applied an improved version of Kolmogorov's 1957 superposition theorem [32], due to Sprecher [58], to the field of neurocomputing and demonstrated that a 3-layer feedforward neural network (one input layer with n inputs, one hidden layer with $2n+1$ neurons, one output layer) are universal approximators for all *continuous* functions from the n-dimensional cube to a finite m-dimensional real vector space; *i.e.*, $f : [0,1]^n \to \mathbb{R}^m$. Recently, Ismailov (2022) [29]

demonstrated existence of neural networks implementing *discontinuous* functions, however efficient learning algorithms for such networks are not still available.

The solutions of interfacial PDE problems are discontinuous, with jumps appearing not only in the solution but also in the solution gradient. In light of above considerations, we define two separate neural networks to represent solution in Ω^- and Ω^+ regions:

$$u^+ = \mathcal{N}^+(\mathbf{x}) : \mathbb{R}^3 \cap \Omega^+ \to \mathbb{R} \qquad\qquad u^- = \mathcal{N}^-(\mathbf{x}) : \mathbb{R}^3 \cap \Omega^- \to \mathbb{R}$$

We use SIREN neural networks, where we implement fully connected feedforward architecture with `sin` activation function and the output layer is a single linear neuron. Note that piecewise differentiable nonlinearities such as the `ReLU` function are inappropriate choices for representing solutions to differential equations. Weights and biases are initialized from a truncated normal distribution with zero mean and unit variance.

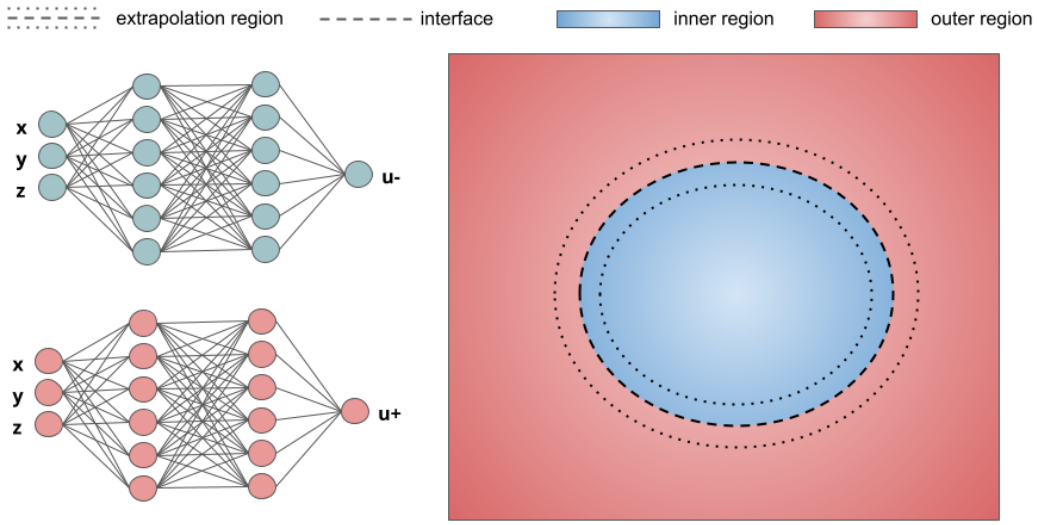

Figure 2: Two neural networks are defined for the two regions of the computational domain.

Solution networks are evaluated on grid points while the parameters of these networks are optimized using the loss function. We define the loss function by the mean-squared-error (MSE) of the residual of the discretized partial differential equation with jump conditions derived in section 3.2 that is evaluated on the grid points:

$$\mathcal{L}(u) = ||A\hat{u}(\mathbf{x}_{ijk} \in \Omega) - b||_2^2 + \text{vol}_{\mathcal{C}}||\hat{u}(\mathbf{x}_{ijk} \in \partial\Omega) - u(\mathbf{x}_{ijk} \in \partial\Omega)||_2^2$$

JAX-DIPS allows for exact computation of the gradient of the loss function using automatic differentiation, *i.e.* $\nabla_u \mathcal{L}(u)$. This is advantageous over existing approximate iterative methods such as GMRES, Conjugate Gradient, *etc.* Therefore, our strategy is to leverage this capability and use more sophisticated optimizers developed in the deep learning community (*e.g.* Adam [30], RMSProp, *etc*) to minimize the afforementioned loss function with the desired solution vector u^* of the PDE problem.

3.2. Approach I. Finite discretization method fused with regression extrapolation

For spatial discretizations at the presence of jump conditions we employ the numerical algorithm proposed by [8] on Cartesian grids. Method of [8] produces second-order accurate solutions and

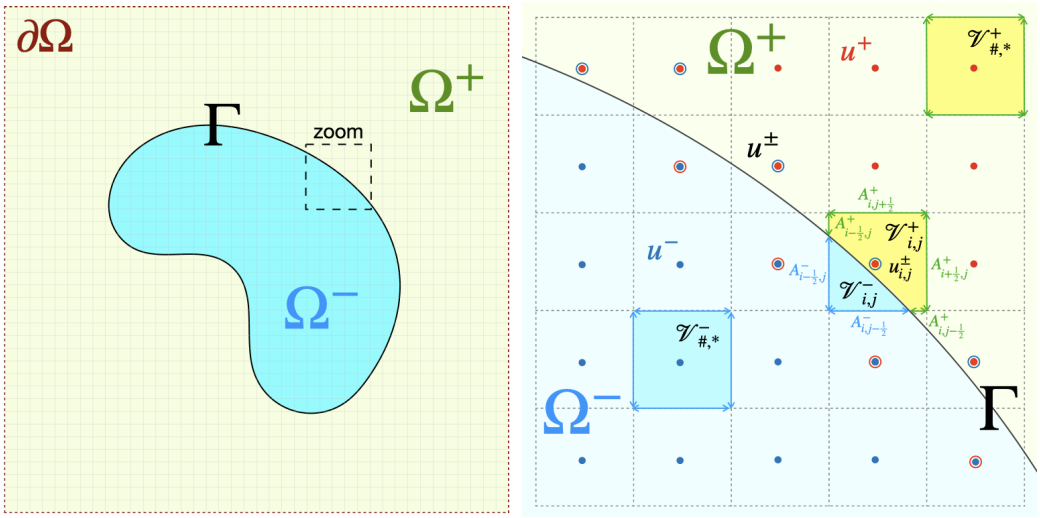

Figure 3: Notation used in this paper. Close to the interface where finite volumes are crossed by the interface, there are extra degrees of freedom (open circles) that are extrapolations of solutions from each domain to the opposite domain. Jump conditions are implicitly encoded in these extrapolated values.

first-order accuracte gradients in the L^∞-norm, while having a compact stencil that makes it a good candidate for parallelization. Moreover, treatment of the interface jump conditions do not introduce any augmented variables, this preserves the homogeneous structure of the linear system.

Here we use a finite volume discretization equation uniformly for all grid points. At grid points where the finite volumes are crossed by Γ we have

$$\sum_{s=-,+}\int_{\Omega^s\cap\mathcal{V}_{i,j}} k^s u^s d\Omega - \sum_{s=-,+}\int_{\Omega^s\cap\partial\mathcal{V}_{i,j}} \mu^s \partial_{\mathbf{n}^s} u^s d\Gamma = \sum_{s=-,+}\int_{\Omega^s\cap\mathcal{V}_{i,j}} f^s d\Omega + \int_{\Gamma\cap\mathcal{V}_{i,j}} [\mu\partial_{\mathbf{n}} u] d\Gamma$$

following standard treatment of volumetric integrals and using central differencing for derivatives we obtain in 2D (with trivial 3D extension)

$$\begin{aligned}
&\sum_{s=-,+} k^s_{i,j} u^s_{i,j} |\mathcal{V}^s_{i,j}| - \sum_{s=-,+}\left(\mu^s_{i-\frac{1}{2},j} A^s_{i-\frac{1}{2},j} \frac{u^s_{i-1,j}-u^s_{i,j}}{\Delta x} + \mu^s_{i+\frac{1}{2},j} A^s_{i+\frac{1}{2},j} \frac{u^s_{i+1,j}-u^s_{i,j}}{\Delta x} + \right.\\
&\left.\mu^s_{i,j-\frac{1}{2}} A^s_{i,j-\frac{1}{2}} \frac{u^s_{i,j-1}-u^s_{i,j}}{\Delta y} + \mu^s_{i,j+\frac{1}{2}} A^s_{i,j+\frac{1}{2}} \frac{u^s_{i,j+1}-u^s_{i,j}}{\Delta y}\right)\\
&= \sum_{s=-,+} f^s_{i,j} |\mathcal{V}^s_{i,j}| + \int_{\Gamma\cap\mathcal{V}_{i,j}} \beta d\Gamma + \mathcal{O}(\max(\Delta x,\Delta y)^{\mathcal{D}})
\end{aligned}$$

where $\mathcal{D}$ is the problem dimensionality.

Note that far from interface either $s=-$ (for $\mathbf{x}\in\Omega^-$) or $s=+$ (for $\mathbf{x}\in\Omega^+$) is retained. This is automatically considered through zero values for sub-volumes $|\mathcal{V}^+_{i,j}|$ and $|\mathcal{V}^-_{i,j}|$ as well as their face areas. Note that $\mu^-_{i-1/2,j}$ (or $\mu^+_{i-1/2,j}$) corresponds to the value of diffusion coefficient at the middle of segment $A^-_{i-1/2,j}$ (or $A^+_{i-1/2,j}$) respectively, same is true for other edges as well. However, there are extra degrees of freedom on grid points whose finite volumes are crossed by the interface; *i.e.*, see double circles in figure 3. [8] derived analytical expressions for the extra degrees of freedom (u^+ in Ω^- and u^- in Ω^+) in terms of the original degrees of freedom (u^- in Ω^- and u^+ in Ω^+) as well as the jump conditions, this preserves the original $\mathrm{N_x}\times\mathrm{N_y}$ system size.

In this scheme the basic idea is to extrapolate the jump at grid point from jump condition at the projected point onto the interface using a Taylor expansion: $u^+_{i,j} - u^-_{i,j} = [u]_{\mathbf{r}^{pr}_{i,j}} + \delta_{i,j}(\partial_{\mathbf{n}} u^+(\mathbf{r}^{pr}_{i,j}) -$

$\partial_{\mathbf{n}} u^-(\mathbf{r}^{pr}_{i,j})$). The unknown value ($u^-_{i,j}$ or $u^+_{i,j}$) is obtained based on approximation of the normal derivatives (*i.e.* $\partial_{\mathbf{n}} u^\pm(\mathbf{r}^{pr}_{i,j})$) which are computed using a least squares calculation on neighboring grid points that are in the fast-diffusion region (referred to as "Bias Fast") or in the slow diffusion region (referred to as "Bias Slow"). This makes two sets of ruls for unknown values $u^\pm_{i,j}$.

In two dimensions and on uniform grids, the gradient operator at the grid cell (i,j) that is crossed by an interface is estimated by a least squares solution given by

$$(\nabla u^\pm)_{i,j} = \mathbf{D}^\pm_{i,j} \begin{bmatrix} u_{i-1,j-1} - u^\pm_{i,j} \\ u_{i,j-1} - u^\pm_{i,j} \\ \vdots \\ u_{i+1,j+1} - u^\pm_{i,j} \end{bmatrix} \qquad \mathbf{D}^\pm_{i,j} = \left(X^T_{i,j} W^\pm_{i,j} X_{i,j}\right)^{-1} \left(W^\pm_{i,j} X_{i,j}\right)^T$$

and

$$W^\pm_{i,j} = \begin{bmatrix} \omega^\pm_{i,j}(-1,-1) & & & \\ & \omega^\pm_{i,j}(0,-1) & & \\ & & \ddots & \\ & & & \omega^\pm_{i,j}(1,1) \end{bmatrix} \qquad X_{i,j} = \begin{bmatrix} -h_x & -h_y \\ 0 & -h_y \\ h_x & -h_y \\ -h_x & 0 \\ 0 & 0 \\ h_x & 0 \\ -h_x & h_y \\ 0 & h_y \\ h_x & h_y \end{bmatrix}$$

and

$$\omega^\pm_{i,j}(p,q) = \begin{cases} 1 & (p,q) \in N^\pm_{i,j} \\ 0 & \text{else} \end{cases} \tag{4}$$

In this case, $D^\pm_{i,j}$ is a 2×9 matrix and we denote each of its 2×1 columns with $d^\pm_{i,j,p,q}$

$$\mathbf{D}^\pm_{i,j} = \begin{bmatrix} d^\pm_{i,j,-1,-1} & d^\pm_{i,j,0,-1} & d^\pm_{i,j,1,-1} & d^\pm_{i,j,-1,0} & d^\pm_{i,j,0,0} & d^\pm_{i,j,1,0} & d^\pm_{i,j,-1,1} & d^\pm_{i,j,0,1} & d^\pm_{i,j,1,1} \end{bmatrix}$$

The least square coefficients are then obtained by dot product of normal vector with these columns

$$c^\pm_{i,j,p,q} = \mathbf{n}^T_{i,j} d^\pm_{i,j,p,q}$$

and normal derivative can be computed (noting that $c^\pm_{i,j} = -\sum_{(p,q)\in N^\pm_{i,j}} c^\pm_{i,j,p,q}$)

$$\partial_n u^\pm(\mathbf{r}^{proj}_{i,j}) = c^\pm_{i,j} u^\pm_{i,j} + \sum_{(p,q)\in N^\pm_{i,j}} c^\pm_{i,j,p,q} u^\pm_{i+p,j+q} + \mathcal{O}(h)$$

At this point we can define a few intermediate variables at each grid point to simplify the presentation of the method,

$$\zeta^\pm_{i,j,p,q} := \delta_{i,j} \frac{[\mu]}{\mu^\mp} c^\pm_{i,j,p,q} \qquad \zeta^\pm_{i,j} := -\sum_{(p,q)\in N^\pm_{i,j}} \zeta^\pm_{i,j,p,q}$$

$$\gamma^\pm_{i,j,p,q} := \frac{\zeta^\pm_{i,j,p,q}}{1 \pm \zeta^\pm_{i,j}} \qquad \gamma^\pm_{i,j} := -\sum_{(p,q)\in N^\pm_{i,j}} \gamma^\pm_{i,j,p,q}$$

where the set of neighboring grid points are

$$N^\pm_{i,j} = \{(p,q): \quad p = -1,0,1, \quad q = -1,0,1, \quad (p,q) \neq (0,0), \quad \mathbf{x}_{i+p,j+q} \in \Omega^\pm\}$$

and $\delta_{i,j}$ is the signed distance from $\mathbf{x}_{i,j}$ that is computed from the level-set function $\phi(\mathbf{x})$

$$\delta_{i,j} = \frac{\phi(\mathbf{x}_{i,j})}{|\nabla\phi(\mathbf{x}_{i,j})|}$$

- Rules based on approximating $\partial_{\mathbf{n}} u^{+}(\mathbf{r}_{i,j}^{pr})$:

$$u_{i,j}^{-} = \begin{cases} u_{i,j} & \mathbf{x}_{i,j} \in \Omega^{-} \\ u_{i,j}(1-\gamma_{i,j}^{-}) - \sum_{(p,q)\in N_{i,j}^{-}} \gamma_{i,j,p,q}^{-} u_{i+p,j+q} - (\alpha + \frac{\delta_{i,j}\beta}{\mu^{+}})(1-\gamma_{i,j}^{-}) & \mathbf{x}_{i,j} \in \Omega^{+} \end{cases} \tag{5}$$

$$u_{i,j}^{+} = \begin{cases} u_{i,j}(1-\zeta_{i,j}^{-}) - \sum_{(p,q)\in N_{i,j}^{-}} \zeta_{i,j,p,q}^{-} u_{i+p,j+q} + \alpha + \delta_{i,j}\frac{\beta}{\mu^{+}} & \mathbf{x}_{i,j} \in \Omega^{-} \\ u_{i,j} & \mathbf{x}_{i,j} \in \Omega^{+} \end{cases} \tag{6}$$

It is useful to cast this in the form of matrix operations through defining intermediate tensors:

$$\mathbf{\Gamma}_{i,j} := \begin{bmatrix} \gamma_{i-1,j+1}^{-} & \gamma_{i,j+1}^{-} & \gamma_{i+1,j+1}^{-} \\ \gamma_{i-1,j}^{-} & \gamma_{i,j}^{-} & \gamma_{i+1,j}^{-} \\ \gamma_{i-1,j-1}^{-} & \gamma_{i,j-1}^{-} & \gamma_{i+1,j-1}^{-} \end{bmatrix}, \qquad \boldsymbol{\zeta}_{i,j} := \begin{bmatrix} \zeta_{i-1,j+1}^{-} & \zeta_{i,j+1}^{-} & \zeta_{i+1,j+1}^{-} \\ \zeta_{i-1,j}^{-} & \zeta_{i,j}^{-} & \zeta_{i+1,j}^{-} \\ \zeta_{i-1,j-1}^{-} & \zeta_{i,j-1}^{-} & \zeta_{i+1,j-1}^{-} \end{bmatrix}$$

$$\mathbf{U}_{i,j} := \begin{bmatrix} u_{i-1,j+1} & u_{i,j+1} & u_{i+1,j+1} \\ u_{i-1,j} & u_{i,j} & u_{i+1,j} \\ u_{i-1,j-1} & u_{i,j-1} & u_{i+1,j-1} \end{bmatrix}, \quad \mathbf{N}_{i,j}^{\pm} := \begin{bmatrix} \omega_{i,j}^{\pm}(-1,1) & \omega_{i,j}^{\pm}(0,1) & \omega_{i,j}^{\pm}(1,1) \\ \omega_{i,j}^{\pm}(-1,0) & 0 & \omega_{i,j}^{\pm}(1,0) \\ \omega_{i,j}^{\pm}(-1,-1) & \omega_{i,j}^{\pm}(0,-1) & \omega_{i,j}^{\pm}(1,-1) \end{bmatrix}$$

where $\mathbf{N}^{-}$ is a masking filter that passes the values in the negative neighborhood of node (i,j).

We also introduce the Hadamard product $\odot$ between two identical matrices that creates another identical matrix with each entry being elementwise products. Moreover, double contraction of two tensors A and B is defined by $A : B = \sum A \odot B$ which is a scalar value and equals the sum of all entries of the Hadamard product of the tensors; *i.e.*, note $A : A$ is square of Frobenius norm of A. Using these notations, the substitution rules read

$$u_{i,j}^{-} = \begin{cases} u_{i,j} & \mathbf{x}_{i,j} \in \Omega^{-} \\ \big(1 + \mathbf{\Gamma}_{i,j}^{-} : \mathbf{N}_{i,j}^{-}\big)u_{i,j} - \big(\mathbf{\Gamma}_{i,j}^{-} \odot \mathbf{N}_{i,j}^{-}\big) : \mathbf{U}_{i,j} - (\alpha + \delta_{i,j}\frac{\beta}{\mu^{+}})\big(1 + \mathbf{\Gamma}_{i,j}^{-} : \mathbf{N}_{i,j}^{-}\big) & \mathbf{x}_{i,j} \in \Omega^{+} \end{cases} \tag{7}$$

$$u_{i,j}^{+} = \begin{cases} \big(1 + \boldsymbol{\zeta}_{i,j}^{-} : \mathbf{N}_{i,j}^{-}\big)u_{i,j} - \big(\boldsymbol{\zeta}_{i,j}^{-} \odot \mathbf{N}_{i,j}^{-}\big) : \mathbf{U}_{i,j} + \alpha + \delta_{i,j}\frac{\beta}{\mu^{+}} & \mathbf{x}_{i,j} \in \Omega^{-} \\ u_{i,j} & \mathbf{x}_{i,j} \in \Omega^{+} \end{cases} \tag{8}$$

- Rules based on approximating $\partial_{\mathbf{n}} u^{-}(\mathbf{r}_{i,j}^{pr})$:

$$u_{i,j}^{-} = \begin{cases} u_{i,j} & \mathbf{x}_{i,j} \in \Omega^{-} \\ u_{i,j}(1-\zeta_{i,j}^{+}) - \sum_{(p,q)\in N_{i,j}^{+}} \zeta_{i,j,p,q}^{+} u_{i+p,j+q} - \alpha - \delta_{i,j}\frac{\beta}{\mu^{-}} & \mathbf{x}_{i,j} \in \Omega^{+} \end{cases} \tag{9}$$

$$u_{i,j}^{+} = \begin{cases} u_{i,j}(1-\gamma_{i,j}^{+}) - \sum_{(p,q)\in N_{i,j}^{+}} \gamma_{i,j,p,q}^{+} u_{i+p,j+q} + (\alpha + \delta_{i,j}\frac{\beta}{\mu^{-}})(1-\gamma_{i,j}^{+}) & \mathbf{x}_{i,j} \in \Omega^{-} \\ u_{i,j} & \mathbf{x}_{i,j} \in \Omega^{+} \end{cases} \tag{10}$$

in matrix notation we have

$$u_{i,j}^{-} = \begin{cases} u_{i,j} & \mathbf{x}_{i,j} \in \Omega^{-} \\ \big(1 + \boldsymbol{\zeta}_{i,j}^{+} : \mathbf{N}_{i,j}^{+}\big)u_{i,j} - \big(\boldsymbol{\zeta}_{i,j}^{+} \odot \mathbf{N}_{i,j}^{+}\big) : \mathbf{U}_{i,j} - \alpha - \delta_{i,j}\frac{\beta}{\mu^{-}} & \mathbf{x}_{i,j} \in \Omega^{+} \end{cases} \tag{11}$$

$$u_{i,j}^{+} = \begin{cases} \big(1 + \mathbf{\Gamma}_{i,j}^{+} : \mathbf{N}_{i,j}^{+}\big)u_{i,j} - \big(\mathbf{\Gamma}_{i,j}^{+} \odot \mathbf{N}_{i,j}^{+}\big) : \mathbf{U}_{i,j} + (\alpha + \delta_{i,j}\frac{\beta}{\mu^{-}})\big(1 + \mathbf{\Gamma}_{i,j}^{+} : \mathbf{N}_{i,j}^{+}\big) & \mathbf{x}_{i,j} \in \Omega^{-} \\ u_{i,j} & \mathbf{x}_{i,j} \in \Omega^{+} \end{cases} \tag{12}$$

procedure BIAS SLOW

if $\Gamma \cap \mathcal{C}_{i,j} = \emptyset$ **then**

$$B^{\pm}_{i,j} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \quad r^{\pm}_{i,j} = 0$$

else

if $\mu^{-}_{i,j} > \mu^{+}_{i,j}$ **then**

if $\phi_{i,j} \geq 0$ **then**

$$B^{+}_{i,j} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \quad r^{+}_{i,j} = 0$$

$$B^{-}_{i,j} = \begin{bmatrix} -\gamma^{-}_{i,j,-1,1} & -\gamma^{-}_{i,j,0,1} & -\gamma^{-}_{i,j,1,1} \\ -\gamma^{-}_{i,j,-1,0} & 1-\gamma^{-}_{i,j} & -\gamma^{-}_{i,j,1,0} \\ -\gamma^{-}_{i,j,-1,-1} & -\gamma^{-}_{i,j,0,-1} & -\gamma^{-}_{i,j,1,-1} \end{bmatrix}; \quad r^{-}_{i,j} = -(\alpha^{proj}_{i,j} + \delta_{i,j}\frac{\beta^{proj}_{i,j}}{\mu^{+}_{proj}})(1-\gamma^{-}_{i,j})$$

else

$$B^{+}_{i,j} = \begin{bmatrix} -\zeta^{-}_{i,j,-1,1} & -\zeta^{-}_{i,j,0,1} & -\zeta^{-}_{i,j,1,1} \\ -\zeta^{-}_{i,j,-1,0} & 1-\zeta^{-}_{i,j} & -\zeta^{-}_{i,j,1,0} \\ -\zeta^{-}_{i,j,-1,-1} & -\zeta^{-}_{i,j,0,-1} & -\zeta^{-}_{i,j,1,-1} \end{bmatrix}; \quad r^{+}_{i,j} = \alpha^{proj}_{i,j} + \delta_{i,j}\frac{\beta^{proj}_{i,j}}{\mu^{+}_{proj}}$$

$$B^{-}_{i,j} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \quad r^{-}_{i,j} = 0$$

else

if $\phi_{i,j} \geq 0$ **then**

$$B^{+}_{i,j} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \quad r^{+}_{i,j} = 0$$

$$B^{-}_{i,j} = \begin{bmatrix} -\zeta^{+}_{i,j,-1,1} & -\zeta^{+}_{i,j,0,1} & -\zeta^{+}_{i,j,1,1} \\ -\zeta^{+}_{i,j,-1,0} & 1-\zeta^{+}_{i,j} & -\zeta^{+}_{i,j,1,0} \\ -\zeta^{+}_{i,j,-1,-1} & -\zeta^{+}_{i,j,0,-1} & -\zeta^{+}_{i,j,1,-1} \end{bmatrix}; \quad r^{-}_{i,j} = \alpha^{proj}_{i,j} + \delta_{i,j}\frac{\beta^{proj}_{i,j}}{\mu^{-}_{proj}}$$

else

$$B^{+}_{i,j} = \begin{bmatrix} -\gamma^{+}_{i,j,-1,1} & -\gamma^{+}_{i,j,0,1} & -\gamma^{+}_{i,j,1,1} \\ -\gamma^{+}_{i,j,-1,0} & 1-\gamma^{+}_{i,j} & -\gamma^{+}_{i,j,1,0} \\ -\gamma^{+}_{i,j,-1,-1} & -\gamma^{+}_{i,j,0,-1} & -\gamma^{+}_{i,j,1,-1} \end{bmatrix}; \quad r^{+}_{i,j} = (\alpha^{proj}_{i,j} + \delta_{i,j}\frac{\beta^{proj}_{i,j}}{\mu^{-}_{proj}})(1-\gamma^{+}_{i,j})$$

$$B^{-}_{i,j} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \quad r^{-}_{i,j} = 0$$

Algorithm 1: Bias Slow approximation of the non-existing solution value on a grid point based on existing solution values in its neighborhood. The notation is used for $u^{\pm}_{i,j} = B^{\pm}_{i,j} : \mathbf{U}_{i,j} + r^{\pm}_{i,j}$.

3.3. Approach II. Finite discretization method fused with neural extrapolation

Our proposed hybrid PDE solver is based on the observation that the neural network models for negative and positive domains can be explicitly used as interpolation and extrapolation functions within the finite discretization schemes.

The basic idea of the finite discretization method of section 3.2 is to impose the jump conditions on the already-available grid points in the computational cells that are crossed by the interface, instead of aspiring to impose the jump conditions exactly on the interface itself. This is achieved by a Taylor expansion to project the interfacial jump conditions onto adjacent grid points. Using the neural network models for solutions, we are able to naturally evaluate extensions of the solution functions in a banded region around the interface as illustrated in figure 2.

Starting from the jump conditions, for points on the interface, $\mathbf{x} \in \Gamma$, we have

$$u^+ - u^- = \alpha$$
$$\mu^+ \partial_n u^+ - \mu^- \partial_n u^- = \beta$$

and after Taylor expansion in the normal direction we obtain on the adjacent grid points (i, j)

$$u^+_{i,j} - u^-_{i,j} = [u]_{\mathbf{r}^{pr}_{i,j}} + \delta_{i,j}(\partial_{\mathbf{n}} u^+(\mathbf{r}^{pr}_{i,j}) - \partial_{\mathbf{n}} u^-(\mathbf{r}^{pr}_{i,j})) \tag{13}$$

which explicitly incorporates the jump condition in the solutions. To incorporate the jump condition in fluxes we can rewrite either of the normal gradients in terms of the other

$$\partial_{\mathbf{n}} u^+(\mathbf{r}^{pr}_{i,j}) = \frac{\mu^-}{\mu^+} \partial_{\mathbf{n}} u^-(\mathbf{r}^{pr}_{i,j}) + \frac{\beta}{\mu^+}$$
$$\partial_{\mathbf{n}} u^-(\mathbf{r}^{pr}_{i,j}) = \frac{\mu^+}{\mu^-} \partial_{\mathbf{n}} u^+(\mathbf{r}^{pr}_{i,j}) - \frac{\beta}{\mu^-}$$

which leads to two relationships among predictions of the two neural networks at each grid point in the banded extrapolation region

$$u^+_{i,j} - u^-_{i,j} = \alpha(\mathbf{r}^{pr}_{i,j}) + \delta_{i,j}\left((\frac{\mu^-}{\mu^+} - 1)\partial_{\mathbf{n}} u^-(\mathbf{r}^{pr}_{i,j}) + \frac{\beta(\mathbf{r}^{pr}_{i,j})}{\mu^+} \right) \tag{14}$$

$$u^+_{i,j} - u^-_{i,j} = \alpha(\mathbf{r}^{pr}_{i,j}) + \delta_{i,j}\left((1 - \frac{\mu^+}{\mu^-})\partial_{\mathbf{n}} u^+(\mathbf{r}^{pr}_{i,j}) + \frac{\beta(\mathbf{r}^{pr}_{i,j})}{\mu^-} \right) \tag{15}$$

Note that we are representing solution functions, $\hat{u}^{\pm}(\mathbf{r})$, with neural networks where computing the normal derivatives is trivial using automatic differentiation of the network along the normal directions. In contrast to finite discretization methods, solutions at off-grid points is readily accessible by simply evaluating the neural network function at any desired points. Note that we can compute the projected location on the interface starting from each grid point (i, j) using the level-set function:

$$\mathbf{r}^{proj}_{ij} = \mathbf{r}_{ij} - \delta_{i,j}\mathbf{n}_{i,j}$$

In the second approach, the loss function remains as before, except the unknown $u^{\pm}$ values are derived using equations 14–15, instead of computing a regression-based extrapolation function based on the points in the neighborhood of interface cells:

$$\mathcal{L} = \Bigg|\Bigg| \sum_{s=-,+} k^s_{i,j} u^s_{i,j} |\mathcal{V}^s_{i,j}| - \sum_{s=-,+} \Bigg(\mu^s_{i-\frac{1}{2},j} A^s_{i-\frac{1}{2},j} \frac{u^s_{i-1,j} - u^s_{i,j}}{\Delta x} + \mu^s_{i+\frac{1}{2},j} A^s_{i+\frac{1}{2},j} \frac{u^s_{i+1,j} - u^s_{i,j}}{\Delta x} +$$
$$\mu^s_{i,j-\frac{1}{2}} A^s_{i,j-\frac{1}{2}} \frac{u^s_{i,j-1} - u^s_{i,j}}{\Delta y} + \mu^s_{i,j+\frac{1}{2}} A^s_{i,j+\frac{1}{2}} \frac{u^s_{i,j+1} - u^s_{i,j}}{\Delta y} \Bigg) - \sum_{s=-,+} f^s_{i,j} |\mathcal{V}^s_{i,j}| - \int_{\Gamma \cap \mathcal{V}_{i,j}} \beta d\Gamma \Bigg|\Bigg|^2_2$$

There is a major downside with this approach, that we need to compute second-order derivatives of the network during training. This slows down convergence dramatically, and the time-to-solution increases with increasing the number of neural layers. As a result we use 2 hidden layers for this approach with 10 neurons in each layer.

3.4. Approach III. Penalty minimization method

Physics-informed neural networks (PINNs) are based on minimizing a penalty function on evaluated on a point cloud, where the actual PDE constitutes the residual and automatic differentiation is used to compute the spatial derivaties in the PDE. Each of the two relations 14–15 offers an extra residual term that can be used to penalize the loss function in PINNs

$$\mathcal{L}_{interface} = \left|\left| u^+_{i,j} - u^-_{i,j} - \alpha(\mathbf{r}^{pr}_{i,j}) - \delta_{i,j}\left((\frac{\mu^-}{\mu^+} - 1)\partial_{\mathbf{n}} u^-(\mathbf{r}^{pr}_{i,j}) + \frac{\beta(\mathbf{r}^{pr}_{i,j})}{\mu^+} \right) \right|\right|^2_2 \tag{16}$$

or

$$\mathcal{L}_{interface} = \left|\left| u^+_{i,j} - u^-_{i,j} - \alpha(\mathbf{r}^{pr}_{i,j}) - \delta_{i,j}\left((1 - \frac{\mu^+}{\mu^-})\partial_{\mathbf{n}} u^+(\mathbf{r}^{pr}_{i,j}) + \frac{\beta(\mathbf{r}^{pr}_{i,j})}{\mu^-} \right) \right|\right|^2_2 \tag{17}$$

Far from interface, the usual procedure for physics-informed neural networks is applicable, namely,

$$\mathcal{L}_{bulk} = \left|\left| k^{\pm}u^{\pm} - \nabla\cdot(\mu^{\pm}\nabla u^{\pm}) - f^{\pm} \right|\right|^2_2$$
$$\mathcal{L}_{boundary} = \left|\left| u(\mathbf{r}_{bc}) - \hat{u}(\mathbf{r}_{bc}) \right|\right|^2_2$$

hence, the overall loss function for this class of problems shall be

$$\mathcal{L} = \mathcal{L}_{bulk} + \mathcal{L}_{boundary} + \mathcal{L}_{interface}$$

4. Optimization scheme

4.1. Preconditioners are ideal network regularizers

Finite discretization methods lead to solving a linear algebraic system with gaurantees on convergence and accuracies. The geometric irregularities and fine-grain details of the system often lead to bad condition number for the linear system, which can be remedied by advanced preconditioners. Preconditioners are a powerful technique to accelerate convergence of tradional numerical linear algebraic solvers. Given a poorly conditioned linear system $Ax = b$ one can obtain an equivalent system $\hat{A}\hat{x} = \hat{b}$ with accelerated convergence rate when using iterative gradient based methods. For the conjugate gradient method convergence iteration is proportional to $\sqrt{\kappa(A)}$ where $\kappa(A)$ is the condition number of matrix A. This is achieved by mapping the linear system with a nonsingular matrix M into a new space $M^{-1}Ax = M^{-1}b$ where $M^{-1}A$ has more regular spread of eigenvalues, hence a better condition number. The precondition matrix M should approximate A^{-1} such that $|I - M^{-1}A| < 1$. The simplest choice is the Jacobi prconditioner which amounts to using the diagonal part of A as the preconditioner, $M = diag(A)$.

Intuitively, condition number is caused by a separation of scales for geometric lengthscales or material properties that underly the solution patterns. One of the strengths of the presented approach is to readily enable usage of preconditioners for training neural network surrogate models. The workflow for training neural network surrogate models for PDEs with discontinuities is illustrated in figure 1.

In this work we use the Jacobi pre-conditioner. Basically, every element of the left-hand-side (Au) and right-hand-side (b) vectors are divided by the coefficient of the diagonal term of the matrix given by:

$$a_{ii} = \sum_{s=-,+}\left(k^s_{i,i}|\mathcal{V}^s_{i,i}| + (\mu^s_{i-\frac{1}{2},i}A^s_{i-\frac{1}{2},i} + \mu^s_{i+\frac{1}{2},i}A^s_{i+\frac{1}{2},i})/\Delta x + (\mu^s_{i,i-\frac{1}{2}}A^s_{i,i-\frac{1}{2}} + \mu^s_{i,i+\frac{1}{2}}A^s_{i,i+\frac{1}{2}})/\Delta y \right)$$

Note that we never explicitly compute the matrix, instead we compute the effect of matrix product of Au.

4.2. Learning rate scheduling

First order methods are slow but cheap; second order methods are fast but expensive. In `JAX-DIPS` we primarily utilize first order optimization methods such as Adam, RMSProp, etc. Second order methods such as Newton or BFGS are certainly faster convergence but require much more memory. Traditionally used GMRES or Conjugate Gradient methods for sparse linear systems are somewhere between first order and second order optimization methods that are based on build a basis vectors by computing gradients that are conjugate to each other $\mathbf{p}_j^T A \mathbf{p}_i = 0$ and will converge to the solution in at most n steps; *i.e.*, at most the solution vector is spanned in the full basis.

We found that starting from a zero guess for the solution it is important to start from a large learning rate and gradually decay the learning rate in a process of exponential annealing. For this purpose, we use the exponential decay scheduler provided by `Optax` [26] to control the learning rate in the Adam [30] optimizer:

$$r_k = r_0 \alpha^{k/T}$$

where r_k is the learning rate at step k of optimization, $\alpha < 1$ is the decay rate, and T is the decay count-scale. By default, we set $T = 100$, $\alpha = 0.975$, starting from an initial value of $r_0 = 10^{-2}$ and clip gradients by maximum global gradient norm (to a value 1) [52] before applying the Adam updates in each step. We note a larger decay rate, *e.g.* $\alpha = 0.98$, leads to small oscillations after 10000 steps and although similar levels of accuracy can be achieved at much less iterations, here we report results with the more robust decay rate.

4.3. Domain switching optimization scheme

The linear system suffers from worse condition number in the domain with more variability in diffusion coefficient, or where diffusion coefficient is larger; *i.e.,* the fast region. This leads to regionally unbalanced solution error where the overall error is systematically lopsided by the faster diffusion region.

We found this problem can be drastically improved by interleaving region-specific optimization epochs in the training pipeline, where only one of the networks is updated based on the loss computed in the its region. See algorithm 2 for details of the algorithm.

5. Numerical Examples

We consider solution to the problem

$$\begin{aligned} k^{\pm} u^{\pm} - \nabla \cdot (\mu^{\pm} \nabla u^{\pm}) &= f^{\pm}, && \mathbf{x} \in \Omega^{\pm} \\ [u] &= \alpha, && \mathbf{x} \in \Gamma \\ [\mu \partial_{\mathbf{n}} u] &= \beta, && \mathbf{x} \in \Gamma \end{aligned}$$

In both cases a pair of neural networks with architecture $3 \times 20 \times 20 \times 20 \times 20 \times 20 \times 1$ are used with `tanh` activation functions for hidden layers and linear output neuron. There are $3,562$ trainable parameters (*i.e.*, weights and biases) in the model.

5.1. Case I.

We consider a sphere $\phi(\mathbf{x}) = \sqrt{x^2 + y^2 + z^2} - 0.5$ centered in a box $\Omega : [-1, 1]^3$ with the exact solution

$$\begin{aligned} u^-(x, y, z) &= e^z, && \phi(\mathbf{x}) < 0 \\ u^+(x, y, z) &= \cos(x) \sin(y), && \phi(\mathbf{x}) \geq 0 \end{aligned}$$

and the diffusion coefficient

$$\begin{aligned} \mu^-(x, y, z) &= y^2 \ln(x + 2) + 4 && \phi(\mathbf{x}) < 0 \\ \mu^+(x, y, z) &= e^{-z} && \phi(\mathbf{x}) \geq 0 \end{aligned}$$

procedure DOMAIN SWITCHING OPTIMIZATION
  for *epoch* in $0 \cdots$ N **do**
    $region = Region(epoch)$
    if $region > 0$ **then**
      if $\mu^- > \mu^+$ **then**
        optimize u^-_{NN} in Ω^- given fixed u^+_{NN}
      else
        optimize u^+_{NN} in Ω^+ given fixed u^-_{NN}

    if $region == 0$ **then**
      optimize both networks in $\Omega^- \cup \Omega^+$

    if $region < 0$ **then**
      if $\mu^- < \mu^+$ **then**
        optimize u^-_{NN} in Ω^- given fixed u^+_{NN}
      else
        optimize u^+_{NN} in Ω^+ given fixed u^-_{NN}

procedure REGION(epoch)
  if *mode* == whole region $\rightarrow$ fast region **then**
    $region = epoch \ \% \ \tau$
  if *mode* == fast region $\rightarrow$ whole region $\rightarrow$ slow region **then**
    $region = \tau//2 - epoch \ \% \ \tau$

Algorithm 2: Domain switching method. Switching interval is τ.

that imply

$$f^-(x,y,z) = -[y^2 \ln(x+2) + 4]e^z \qquad \phi(\mathbf{x}) < 0$$
$$f^+(x,y,z) = 2\cos(x)\sin(y)e^{-z} \qquad \phi(\mathbf{x}) \geq 0$$

Table 8 of [24] reports convergence results for the solution and its gradient over the surface of the sphere in the L^∞-norm, here we report similar results for comparison with VIM. Order of convergence, denoted by p, is computed by doubling the number of grid points in every dimension and measuring the L^∞ error of solution and its gradient over all the grid points in the domain:

$$\frac{\text{err}(2h)}{\text{err}(h)} = 2^p \rightarrow p = \log_2\left(\frac{\text{err}(2h)}{\text{err}(h)}\right) \qquad h = \min(h_x, h_y, h_z)$$

After $10,000$ epochs on a grid of $16 \times 16 \times 16$ the L^∞-norm error of the solution is 1.1×10^{-1}, corresponding to a 10% relative error in the solution.

5.2. Case II.

We use a pair of fully connected feedforward neural networks, each composed of 1 hidden layer and 100 neurons with `sine` activation function, followed by an output layer with 1 linear neuron. There are a total of $1,002$ trainable parameters in the model.

We consider a star-shaped interface with inner and outer radii $r_i = 0.151$ and $r_e = 0.911$ that is immersed in a box $\Omega : [-1,1]^3$ described by the level-set function

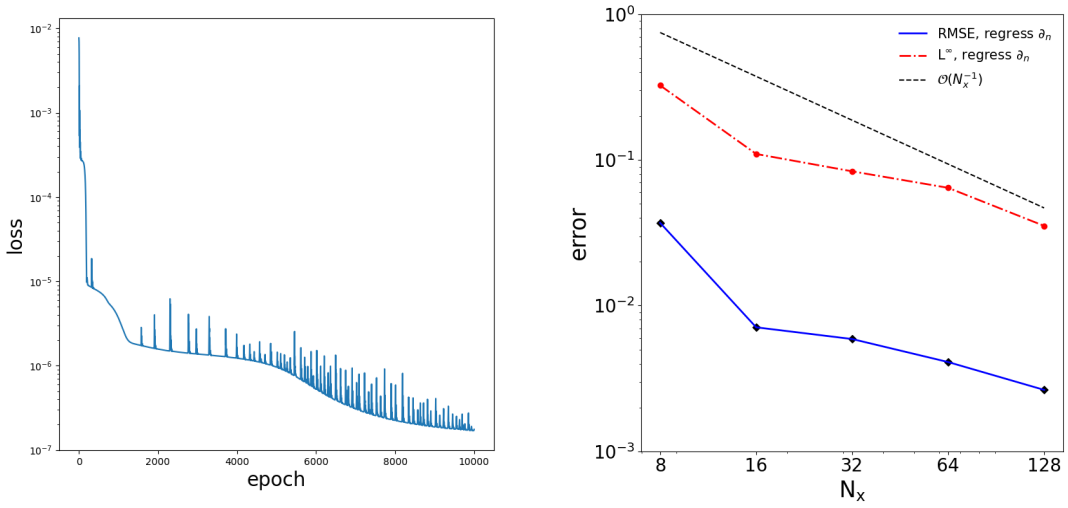

Figure 4: Loss evolution with epochs for the sphere of $16 \times 16 \times 16$ grid (left) and different accuracy measures at 5 resolutions (right).

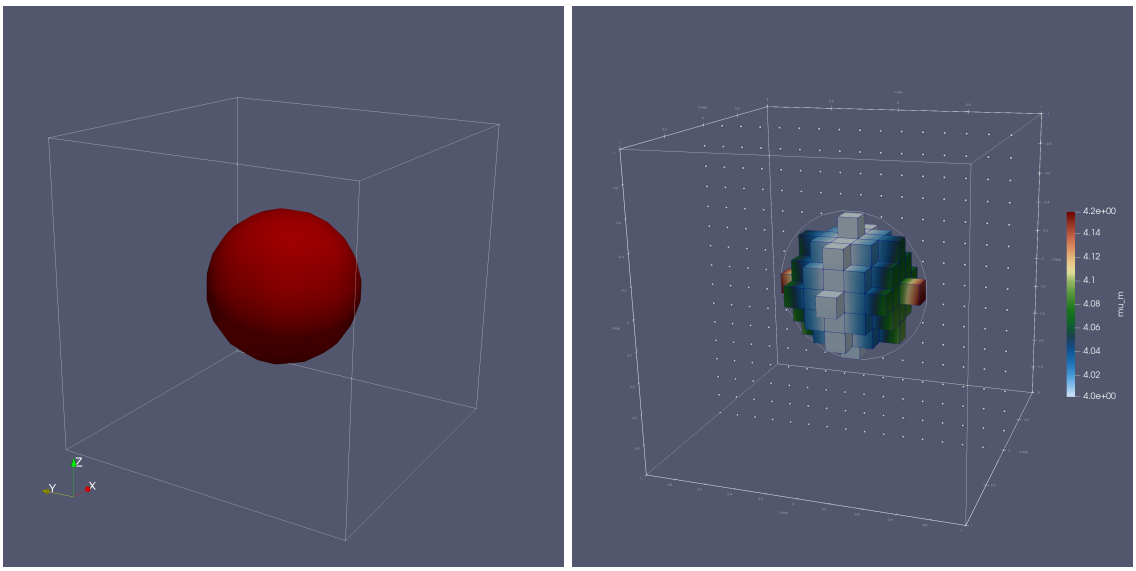

Figure 5: Illustration of three dimensional interface used (left), and $\mu^{\pm}$ on the $16 \times 16 \times 16$ grid (right).

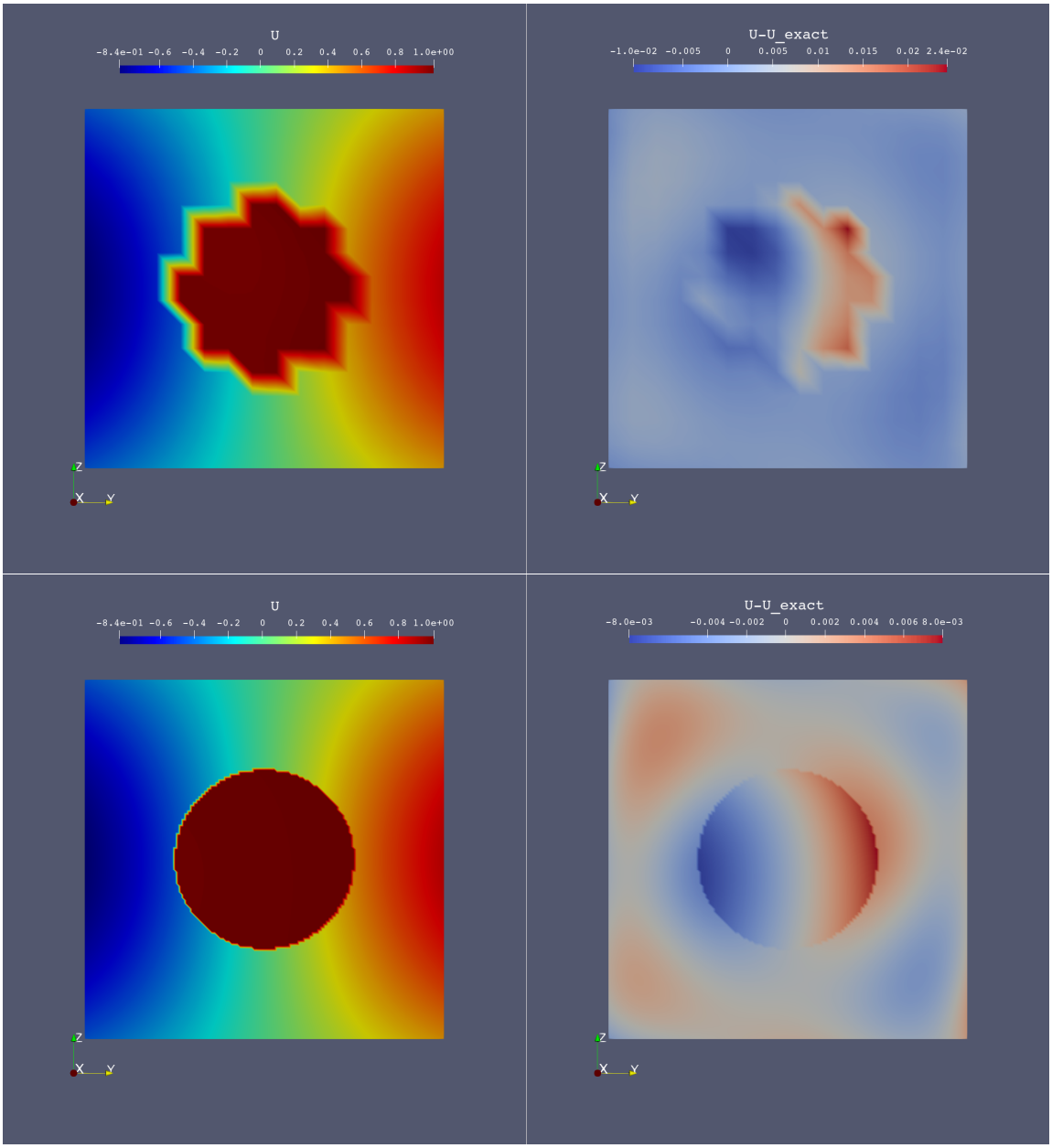

Figure 6: Illustration of numerical solution and absolute error using the regression based solver on a cross section of the domain.

regress ∂_n	RMSE		L^∞		GPU Statistics	
$N_{x,y,z}$	Solution	Order	Solution	Order	t (sec/epoch)	VRAM (GB)
2^3	3.7×10^{-2}	-	3.25×10^{-1}	-	0.0306	1.05
2^4	7.1×10^{-3}	2.38	1.1×10^{-1}	1.56	0.056	1.72
2^5	5.9×10^{-3}	0.27	8.36×10^{-2}	0.4	0.053	2.15
2^6	4.1×10^{-3}	0.53	6.44×10^{-2}	0.38	0.287	5.57
2^7	2.64×10^{-3}	0.64	3.53×10^{-2}	0.87	2.125	32.1

neural ∂_n	RMSE		L^∞		GPU Statistics	
$N_{x,y,z}$	Solution	Order	Solution	Order	t (sec/epoch)	VRAM (GB)
2^3		-		-		
2^4						
2^5						
2^6						
2^7						

Table 1: Convergence on the solution using the regression-based solver. We report L^∞-norm error as well as root-mean-squared-error (RMSE) of the solution field evaluated everywhere in the domain. Rightmost column reports the overall time to solution for `JAX-DIPS` which constitutes $10,000$ epochs in each case and the initial compilation time of jaxpressions. The neural network has 982 trainable parameters. In each case GPU compute occupancy is at 100% on a single NVIDIA RTX A6000 GPU. The neural method has 2 hidden layers with 10 neurons each, overall 322 trainable parameteres. The regression-based method has 5 hidden layers with 10 neurons each, overall 928 trainable parameters.

$$\phi(\mathbf{x}) = \sqrt{x^2+y^2+z^2} - r_0\Big(1 + \big(\frac{x^2+y^2}{x^2+y^2+z^2}\big)^2 \sum_{k=1}^{3} \beta_k \cos\big(n_k\big(\arctan\big(\frac{y}{x}\big) - \theta_k\big)\big)\Big)$$

with the parameters

$$r_0 = 0.483, \qquad \begin{pmatrix} n_1 \\ \beta_1 \\ \theta_1 \end{pmatrix} = \begin{pmatrix} 3 \\ 0.1 \\ 0.5 \end{pmatrix}, \begin{pmatrix} n_2 \\ \beta_2 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 4 \\ -0.1 \\ 1.8 \end{pmatrix}, \qquad \begin{pmatrix} n_3 \\ \beta_3 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 0.15 \\ 0 \end{pmatrix}$$

Considering an exact solution

$$u^-(x,y,z) = \sin(2x)\cos(2y)e^z, \qquad \phi(\mathbf{x}) < 0$$
$$u^+(x,y,z) = \Big[16\big(\frac{y-x}{3}\big)^5 - 20\big(\frac{y-x}{3}\big)^3 + 5\big(\frac{y-x}{3}\big)\Big]\ln(x+y+3)\cos(z), \qquad \phi(\mathbf{x}) \geq 0$$

and the diffusion coefficient

$$\mu^-(x,y,z) = 10\Big[1 + 0.2\cos(2\pi(x+y))\sin(2\pi(x-y))\cos(z)\Big] \qquad \phi(\mathbf{x}) < 0$$
$$\mu^+(x,y,z) = 1 \qquad \phi(\mathbf{x}) \geq 0$$

6. Conclusion & Future Directions

We developed a differentiable GPU-based framework for solving partial differential equations with jump conditions across irregular interfaces in three spatial dimensions. Solutions in each

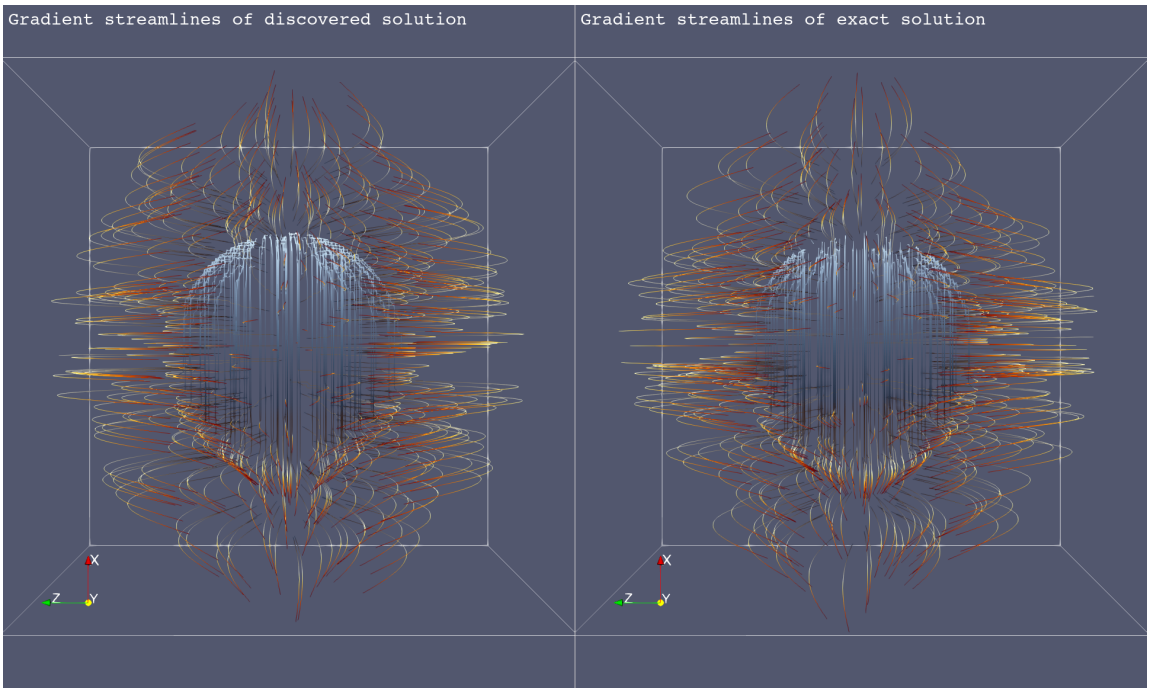

Figure 7: Streamlines of solution gradient for (left) the surrogate neural model colored by model solution value, (right) exact streamlines colored by exact solution values.

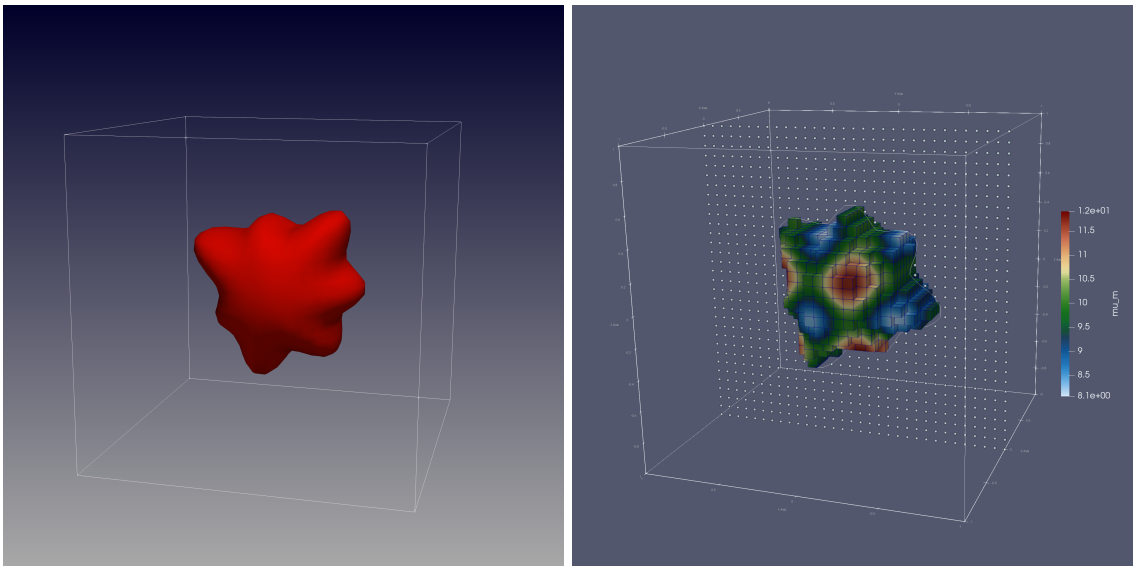

Figure 8: Illustration of three dimensional interface used (left), and $\mu^{\pm}$ on the $32 \times 32 \times 32$ grid (right).

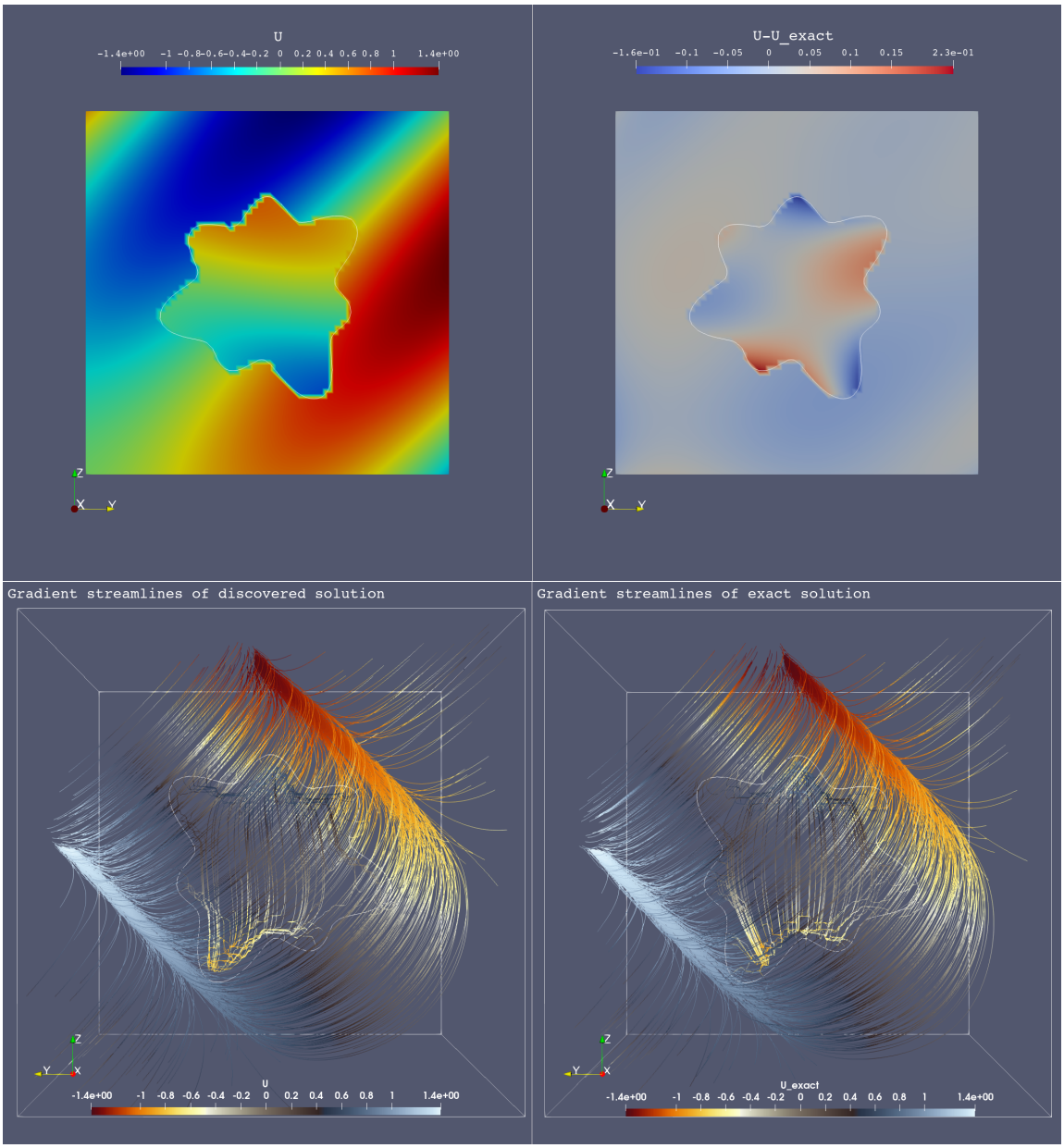

Figure 9: Illustration of exact and numerical solutions on a $64 \times 64 \times 64$ grid.

regress ∂_n	RMSE		L^∞		GPU Statistics	
$N_{x,y,z}$	Solution	Order	Solution	Order	t (sec/epoch)	VRAM (GB)
2^3	1.36×10^{-1}	-	1.27	-	0.019	0.98
2^4	7.98×10^{-2}		8.23×10^{-1}		0.022	1.01
2^5	4.36×10^{-2}		3.85×10^{-1}		0.032	1.30
2^6	2.43×10^{-2}		2.28×10^{-1}		0.200	3.7
2^7						21.4
neural ∂_n	RMSE		L^∞		GPU Statistics	
$N_{x,y,z}$	Solution	Order	Solution	Order	t (sec/epoch)	VRAM (GB)
2^3	2.17×10^{-1}	-	2.89	-	0.0259	0.93
2^4	1.34×10^{-1}		1.66		0.0408	1.19
2^5	5.68×10^{-2}		8.17×10^{-1}		0.0712	2.96
2^6	2.77×10^{-2}		3.94×10^{-1}		0.334	13.6
2^7	OOM	-	OOM	-	OOM	OOM

Table 2: Convergence on the solution using the regression-based solver. We report L^∞-norm error as well as root-mean-squared-error (RMSE) of the solution field evaluated everywhere in the domain. Rightmost column reports the overall time to solution for `JAX-DIPS` which constitutes $10,000$ epochs in each case and the initial compilation time of jaxpressions. The neural network has 982 trainable parameters. In each case GPU compute occupancy is at 100% on a single NVIDIA RTX A6000 GPU. The neural network pair have 1 hidden layer each with 100 neurons, overall $1,002$ trainable parameteres. We use a domain switching scheme with the whole region $\rightarrow$ fast region $\rightarrow$ fast region optimization sequence.

domain are represented by a simple multi-layer perceptron (MLP) and Cartesian grid points of the underlying numerical discretization scheme are treated as collocation points for optimizing the unknown parameters of the MLPs.

There are many improvements for `JAX-DIPS` that we will pursue for future development:

- More sophisticated neural architectures can be considered in `JAX-DIPS` by adding to the model class of the library. We only considered MLPs, however in recent years there have a plethora of deep neural network models that have shown great promise such as transformers, graph neural networks, *etc.*
- Multi-GPU training using domain decomposition and distribution using the `pmap` primitive in JAX is another very immediate improvement that will enable higher resolution simulations.
- Modification of the numerical discretization scheme in order to operate over unstructured point clouds rather than the uniform Cartesian grids.
- Extension to adaptive grids with enhanced resolutions closer to the interfaces while coarsening the grid cells in the bulk.

Acknowledgement

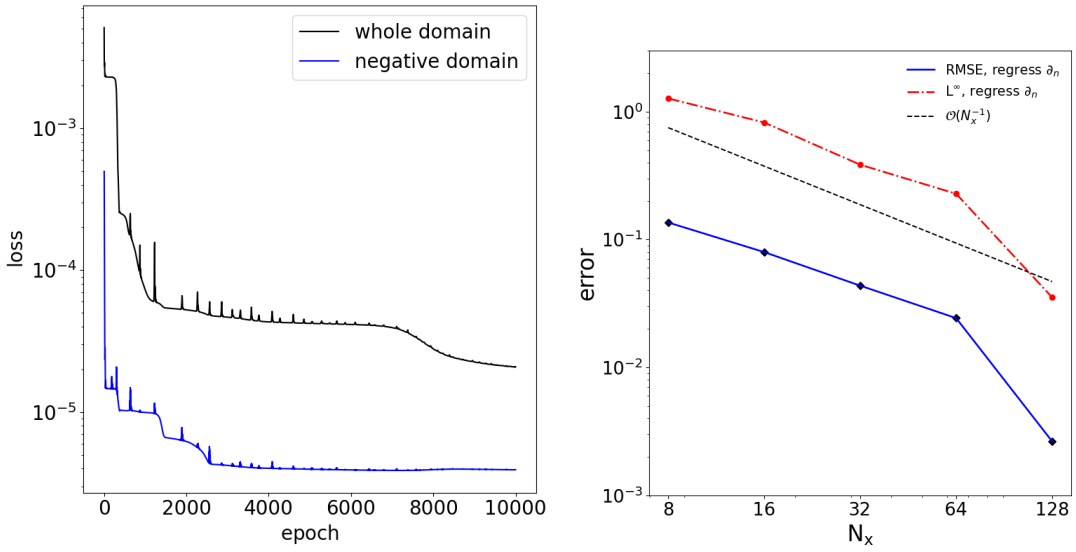

Figure 10: Loss evolution with epochs for the sphere of $64 \times 64 \times 64$ grid (left) and different accuracy measures at 5 resolutions (right).

References

[1] L. Adams and Z. Li. The immersed interface/multigrid methods for interface problems. *SIAM Journal on Scientific Computing*, 24(2):463–479, 2002.

[2] I. Babuška. The finite element method for elliptic equations with discontinuous coefficients. *Computing*, 5(3):207–213, 1970.

[3] T. Belytschko, N. Moës, S. Usui, and C. Parimi. Arbitrary discontinuities in finite elements. *International Journal for Numerical Methods in Engineering*, 50(4):993–1013, 2001.

[4] J. Berg and K. Nyström. Neural network augmented inverse problems for pdes. *arXiv preprint arXiv:1712.09685*, 2017.

[5] D. A. Bezgin, A. B. Buhendwa, and N. A. Adams. Jax-fluids: A fully-differentiable high-order computational fluid dynamics solver for compressible two-phase flows. *arXiv preprint arXiv:2203.13760*, 2022.

[6] K. Bhattacharya, B. Hosseini, N. B. Kovachki, and A. M. Stuart. Model reduction and neural networks for parametric pdes. *arXiv preprint arXiv:2005.03180*, 2020.

[7] D. Bochkov and F. Gibou. Solving elliptic interface problems with jump conditions on cartesian grids. *Journal of Computational Physics*, 407:109269, 2020.

[8] D. Bochkov and F. Gibou. Solving elliptic interface problems with jump conditions on cartesian grids. *Journal of Computational Physics*, 407:109269, 2020.

[9] D. Bochkov and F. Gibou. A non-parametric shape optimization approach for solving inverse problems in directed self-assembly of block copolymers. *arXiv preprint arXiv:2112.09615*, 2021.

[10] D. Bochkov, T. Pollock, and F. Gibou. Sharp-interface simulations of multicomponent alloy solidification. *arXiv preprint arXiv:2112.08650*, 2021.

[11] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang. JAX: composable transformations of Python+NumPy programs, 2018.

[12] J. H. Bramble and J. T. King. A finite element method for interface problems in domains with smooth boundaries and interfaces. *Advances in Computational Mathematics*, 6(1):109–138, 1996.

[13] T. Chen and H. Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995.

[14] L. O. Chua and L. Yang. Cellular neural networks: Applications. *IEEE Transactions on circuits and systems*, 35(10):1273–1290, 1988.

[15] L. O. Chua and L. Yang. Cellular neural networks: Theory. *IEEE Transactions on circuits and systems*, 35(10):1257–1272, 1988.

[16] R. Crockett, P. Colella, and D. T. Graves. A cartesian grid embedded boundary method for solving the poisson and heat equations with discontinuous coefficients in three dimensions. *Journal of Computational Physics*, 230(7):2451–2469, 2011.

[17] N. Dal Santo, S. Deparis, and L. Pegolotti. Data driven approximation of parametrized pdes by reduced basis and neural networks. *Journal of Computational Physics*, 416:109550, 2020.

[18] F. de Avila Belbute-Peres, K. Smith, K. Allen, J. Tenenbaum, and J. Z. Kolter. End-to-end differentiable physics for learning and control. *Advances in neural information processing systems*, 31, 2018.

[19] R. E. Ewing, Z. Li, T. Lin, and Y. Lin. The immersed finite volume element methods for the elliptic interface problems. *Mathematics and Computers in Simulation*, 50(1-4):63–76, 1999.

[20] R. P. Fedkiw, T. Aslam, B. Merriman, and S. Osher. A non-oscillatory eulerian approach to interfaces in multimaterial flows (the ghost fluid method). *Journal of computational physics*, 152(2):457–492, 1999.

[21] K. Galatsis, K. L. Wang, M. Ozkan, C. S. Ozkan, Y. Huang, J. P. Chang, H. G. Monbouquette, Y. Chen, P. Nealey, and Y. Botros. Patterning and templating for nanoelectronics. *Advanced Materials*, 22(6):769–778, 2010.

[22] D. Gobovic and M. Zaghloul. Design of locally connected cmos neural cells to solve the steady-state heat flow problem. In *Proceedings of 36th Midwest Symposium on Circuits and Systems*, pages 755–757. IEEE, 1993.

[23] Y. Gong, B. Li, and Z. Li. Immersed-interface finite-element methods for elliptic interface problems with nonhomogeneous jump conditions. *SIAM Journal on Numerical Analysis*, 46(1):472–495, 2008.

[24] A. Guittet, M. Lepilliez, S. Tanguy, and F. Gibou. Solving elliptic problems with discontinuities on irregular domains–the voronoi interface method. *Journal of Computational Physics*, 298:747–765, 2015.

[25] R. Hecht-Nielsen. Kolmogorov's mapping neural network existence theorem. In *Proceedings of the international conference on Neural Networks*, volume 3, pages 11–14. IEEE Press New York, NY, USA, 1987.

[26] M. Hessel, D. Budden, F. Viola, M. Rosca, E. Sezener, and T. Hennigan. Optax: composable gradient transformation and optimisation, in jax!, 2020.

[27] P. Holl, V. Koltun, and N. Thuerey. Learning to control pdes with differentiable physics. *arXiv preprint arXiv:2001.07457*, 2020.

[28] P. Holl, V. Koltun, K. Um, and N. Thuerey. phiflow: A differentiable pde solving framework for deep learning via physical simulations. In *NeurIPS Workshop*, volume 2, 2020.

[29] V. Ismailov. A three layer neural network can represent any multivariate function, 2020.

[30] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[31] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer. Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 118(21), 2021.

[32] A. N. Kolmogorov. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. In *Doklady Akademii Nauk*, volume 114, pages 953–956. Russian Academy of Sciences, 1957.

[33] I. E. Lagaris, A. Likas, and D. I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE transactions on neural networks*, 9(5):987–1000, 1998.

[34] H. Lee and I. S. Kang. Neural algorithm for solving differential equations. *Journal of Computational Physics*, 91(1):110–131, 1990.

[35] R. J. LeVeque and Z. Li. The immersed interface method for elliptic equations with discontinuous coefficients and singular sources. *SIAM Journal on Numerical Analysis*, 31(4):1019–1044, 1994.

[36] A. J. Lew and G. C. Buscaglia. A discontinuous-galerkin-based immersed boundary method. *International Journal for Numerical Methods in Engineering*, 76(4):427–454, 2008.

[37] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[38] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*, 2020.

[39] Z. Li, T. Lin, and X. Wu. New cartesian grid methods for interface problems using the finite element formulation. *Numerische Mathematik*, 96(1):61–98, 2003.

[40] X.-D. Liu, R. P. Fedkiw, and M. Kang. A boundary condition capturing method for poisson's equation on irregular domains. *Journal of computational Physics*, 160(1):151–178, 2000.

[41] L. Lu, P. Jin, and G. E. Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.

[42] P. Y. Lu, S. Kim, and M. Soljačić. Extracting interpretable physical parameters from spatiotemporal systems using unsupervised learning. *Physical Review X*, 10(3):031056, 2020.

[43] C. Min and F. Gibou. Geometric integration over irregular domains with application to level-set methods. *Journal of Computational Physics*, 226(2):1432–1443, 2007.

[44] C. Min and F. Gibou. A second order accurate level set method on non-graded adaptive cartesian grids. *Journal of Computational Physics*, 225(1):300–321, 2007.

[45] M. Mirzadeh, M. Theillard, and F. Gibou. A second-order discretization of the nonlinear Poisson-Boltzmann equation over irregular geometries using non-graded adaptive Cartesian grids. *Journal of Computational Physics*, 230(5):2125–2140, Mar. 2011.

[46] P. Mistani, A. Guittet, D. Bochkov, J. Schneider, D. Margetis, C. Ratsch, and F. Gibou. The island dynamics model on parallel quadtree grids. *Journal of Computational Physics*, 361:150–166, 2018.

[47] P. Mistani, A. Guittet, C. Poignard, and F. Gibou. A parallel voronoi-based approach for mesoscale simulations of cell aggregate electropermeabilization. *Journal of Computational Physics*, 380:48–64, 2019.

[48] N. Moës, J. Dolbow, and T. Belytschko. A finite element method for crack growth without remeshing. *International journal for numerical methods in engineering*, 46(1):131–150, 1999.

[49] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.

[50] G. Y. Ouaknin, N. Laachi, K. Delaney, G. H. Fredrickson, and F. Gibou. Level-set strategy for inverse dsa-lithography. *Journal of Computational Physics*, 375:1159–1178, 2018.

[51] S. Pakravan, P. A. Mistani, M. A. Aragon-Calvo, and F. Gibou. Solving inverse-pde problems with physics-aware neural networks. *Journal of Computational Physics*, 440:110414, 2021.

[52] R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks, 2012.

[53] M. Raissi, P. Perdikaris, and G. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[54] J. F. Sallee. The middle-cut triangulations of the n-cube. *SIAM Journal on Algebraic Discrete Methods*, 5(3):407–419, 1984.

[55] K. A. Sharp and B. Honig. Calculating total electrostatic energies with the nonlinear poisson-boltzmann equation. *Journal of Physical Chemistry*, 94(19):7684–7692, 1990.

[56] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of computational physics*, 77(2):439–471, 1988.

[57] J. Sirignano and K. Spiliopoulos. Dgm: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364, 2018.

[58] D. A. Sprecher. On the structure of continuous functions of several variables. *Transactions of the American Mathematical Society*, 115:340–355, 1965.

[59] M. Sussman, P. Smereka, and S. Osher. A level set approach for computing solutions to incompressible two-phase flow. *Journal of Computational Physics*, 114(1):146–159, 1994.

[60] M. Theillard, F. Gibou, and T. Pollock. A sharp computational method for the simulation of the solidification of binary alloys. *Journal of scientific computing*, 63(2):330–354, 2015.

[61] K. Um, R. Brand, Y. R. Fei, P. Holl, and N. Thuerey. Solver-in-the-loop: Learning from differentiable physics to interact with iterative pde-solvers. *Advances in Neural Information Processing Systems*, 33:6111–6122, 2020.

[62] D. Xiu and G. E. Karniadakis. A semi-lagrangian high-order method for navier–stokes equations. *Journal of Computational Physics*, 172(2):658–684, 2001.

[63] B. Yu et al. The deep ritz method: a deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, 2018.