A machine learning least-squares method with weighted norm

Paulina Sepúlveda [1] and Ignacio Brevis[1] and Ignacio Muga[1]

Pontificia Universidad Católica de Valparaíso[1]

Minimum Residual Least-Squares Finite Element Methods,
Oct 5-7, 2022

Outline

Table of Contents

Introduction

Neural Networks + Finite element Methods

Recent Literature

[1] C. F. Higham et. al. **Deep learning: An introduction for applied mathematicians**, SIAM Rev., 61 (2019), pp. 860–891.

[2] W. E, **The Deep Ritz Method: A deep learning-based numerical algorithm for solving variational problems**, Commun. Math. Sci., 6 (2018), pp. 1–12.

[3] B. Khara, et. al. **NeuFENet: Neural finite element solutions with theoretical bounds for parametric pdes.** arXiv:2110.01601, 2021.

References for this talk

- S. Mishra, **A machine learning framework for data driven acceleration of computations of differential equations,** Mathematics in Engineering, 1 (2018), pp. 118–146.
- I. Brevis, I. Muga, and K. G. van der Zee, **A machine-learning minimal-residual (ML-MRes) framework for goal-oriented finite element discretizations.** Comput. Math. Appl., Vol. 95, pp. 186–199, 2021.
- I. Brevis, I. Muga, and K. G. van der Zee, **Neural Control of Discrete Weak Formulations: Galerkin, Least-Squares and Minimal-Residual Methods with Quasi-Optimal Weights** , arXiv:2206.07475, 2022.

Our interest

- Given U, V Hilbert spaces, U^*, V^* the dual spaces, a parameter k,

$$B_k : U \to V^*.$$

> Given a k, f_k, and B.C. , find $u_k \in U$ **satisfying a QoI** s.t.
>
> $$B_k u_k = \ell_k, \quad \text{in} \quad V^*,$$
>
> where B_k is a differential operator depending on k, ℓ_k functional in V^* depending on k.

- We consider its **Variational form**

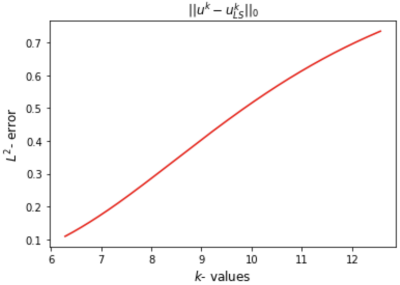
> Find $u_k \in U$ such that
>
> $$b_k(u_k, v) = \ell_k(v), \qquad \forall v \in V.$$

Motivation problem

1D time-harmonic wave propagation problem:

$$\begin{aligned} i\mathbf{k}u + u' &= 0, \qquad \text{in} \quad (0,1) \\ u(0) &= u_0. \end{aligned}$$

- Fixed mesh: 10 elements. (5 nodes per wavelength)
- LS method, lowest order F.E. , and $\mathbf{k}$ grows,
- phase error, and pollution effect.



The problem

- Can we use an ANN to approach a quantity of interest **QoI** for a family of parameters k ? How?
- Could we approach the solution in the whole domain?

Table of Contents

Accelerated method [I.Brevis, Muga, van der Zee]

Assume that $(V, (\cdot,\cdot)_V)$ can be endowed with weigthed inner products $\{(\cdot,\cdot)_{V,\omega}\}$, such that their endowed norms are equivalent, *i.e.*

$$C_1\|v\|_{V,\omega} \leq \|v\|_V \leq C_2\|v\|_{V,\omega}$$

Consider a coarse finite element $U_h \subset U$, where we want to solve the variational form, and $V_h \subset V$ such that $\dim(V_h) > \dim(U_h)$.

The discrete method

Find $(e_{h,k,\omega}, u_{h,k,\omega}) \in V_h \times U_h$ s.t.

$$\begin{aligned} (e_{h,k,\omega}, v_h)_{V,\omega} + b_k(u_{h,k,\omega}, v_h) &= \ell_k(v), \quad \forall\, v_h \in V_h, \\ b_k(w_h, e_{h,k,\omega}) &= 0, \quad \forall\, w_h \in U_h. \end{aligned}$$

- Training off-line to get the optimal weigths ω with an ANN. In-line, solve with the trained weigths.
- Choose the right cost function.

Table of Contents

Formulations

Discrete Least-Squares Formulation, $B_k : U \to L^2(\Omega)$

Find $u_{h,k} \in U_h$ such that

$$(B_k u_{h,k}, B_k v) = (\ell_k, B_k v), \quad \forall v \in U_h.$$

Weighted Least-Squares Formulation

Find $u_h \in U_h$ such that

$$(Bu_h, \bar{\omega} B w_h)_{L^2(\Omega)} = (\ell, \bar{\omega} B w_h), \qquad \forall w_h \in U_h.$$

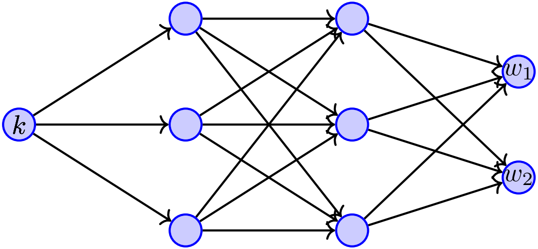
I. Brevis, I. Muga, and K. G. van der Zee, Neural Control of Discrete Weak Formulations: Galerkin, Least-Squares and Minimal-Residual Methods with Quasi-Optimal Weights, arXiv:2206.07475. (Non-parametric, continuous and positive weigths)

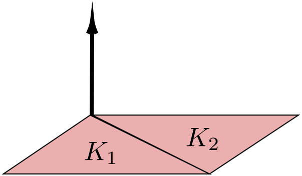
How to include the ANN

- The ANN will give us output **positive** constant weights, as many as elements of the mesh. $\omega^*(\cdot) = \sigma_L(\mathrm{ANN}(\cdot, \theta^*)$, where σ_L is a **positive activation function** then, the ANN returns positive values, as many as we want.

$$(r, w)_w = \sum_{K \in \Omega_h} (w_i r, w)_K$$

- The input of the ANN is one parametro k.





Steps

- We train (**off-line**) with different parameters, and define a family of positive weight-functions to be used in the weighted inner products $\{(\cdot,\cdot)_{V,\omega}, \omega \in W\}$.
- Next, given the discrete U_h, we construct the map $W \times \{k_i\}_{i=1}^{i=n} \ni (\omega, k_i) \to u_{h,k_i\omega} \in U_h$, and proceed to train the ANN by computing:

$$\theta^* = \text{argmin}_\theta \frac{1}{2} \| \sqrt{\omega(\cdot)}(\ell_k - B_k u_{h,k}) \|^2_{L^2(\Omega)},$$

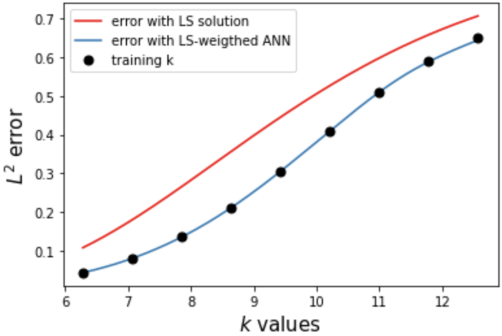
- Finally, we build the matrices of the linear system needed for the **online step**. Having θ^*, we solve $\omega^*(\cdot) = \sigma_L(ANN(\cdot, \theta^*))$, we construct:

$$B_{ij} = (B_k(u_j), w^* B_k(u_i)), \qquad (L_k)_i = (\ell_k, w^* B_k u_i)$$

Numerical results with LS

Find $u_h \in U_h$ such that

$$(B_k u_{h,k}, \omega^* B_k w_h) = (\ell_k, \omega^* B_k w_h), \qquad \forall w_h \in U_h.$$



mesh= 10 elements, lowest order, [Brevis, Muga, Van der Zee *et. al.* 2022.]

Alternative- method

Weighted Least-Squares Formulation

Find $u_h \in U_h$ such that

$$(Bu_h, \bar{\omega} Bw_h)_{L^2(\Omega)} = (\ell, \bar{\omega} Bw_h), \qquad \forall w_h \in U_h.$$

We could think in a trial mesh for U_h, and a finer test mesh V_h, with $\dim(U_h) < \dim V_h$ such that

Find $(e_{h,k,\omega}, u_{h,k,\omega}) \in V_h \times U_h$ s.t.

$$\begin{aligned} (\omega e_{h,k}, v_h)_{V,\omega} + (B_k u_{h,k,\omega}, v_h) &= \ell_k(v), \quad \forall\, v_h \in V_h, \\ (B_k w_h, e_{h,k,\omega}) &= 0, \qquad \forall\, w_h \in U_h. \end{aligned}$$

[I. Brevis, I. Muga, and K. G. van der Zee, 2022]

How to train

- We generate two meshes, one for the trial, and one for the test.

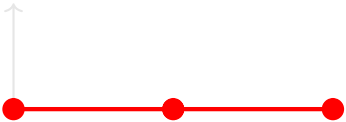
How to train

- We generate two meshes, one for the trial, and one for the test. Trial mesh

1 Element trial mesh

How to train

- We generate two meshes, one for the trial, and one for the test.
  Test mesh



2 Elements in test mesh

How to train

- We generate two meshes, one for the trial, and one for the test.
- Construct the variational form considering Ω_h and Ω_{2h}

How to train

- We generate two meshes, one for the trial, and one for the test.
- Construct the variational form considering Ω_h and Ω_{2h}
- Write down the quantity of interest we want to ritch

How to train

- We generate two meshes, one for the trial, and one for the test.
- Construct the variational form considering Ω_h and Ω_{2h}
- Write down the quantity of interest we want to ritch
- Train with weigths defined on the elements of the Test mesh.

How to train

- We generate two meshes, one for the trial, and one for the test.
- Construct the variational form considering Ω_h and Ω_{2h}
- Write down the quantity of interest we want to ritch
- Train with weigths defined on the elements of the Test mesh.
- Minimize the cost functional

How to train

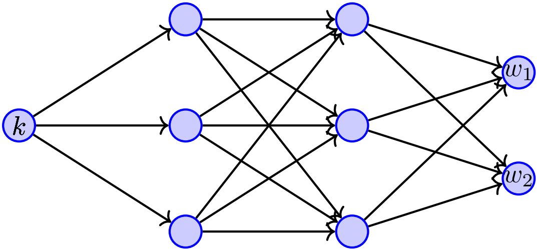
- We generate two meshes, one for the trial, and one for the test.
- Construct the variational form considering Ω_h and Ω_{2h}
- Write down the quantity of interest we want to ritch
- Train with weigths defined on the elements of the Test mesh.
- Minimize the cost functional
- Online, solve using the ω^* weigth.

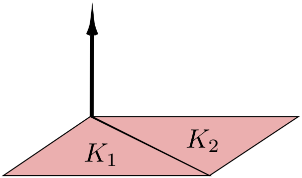
How to include the ANN

- The ANN will give us constant weights for the K elements of the net. $\text{ANN}_\theta = \sigma_L(\text{ANN})$, where σ_L is a positive activation function, then, the ANN returns positive values, as many as we want.

$$(r, w)_w = \sum_{K \in \Omega_h} (w_i r, w)_K$$

- The input of the ANN is one parametro k.





Advantage of the discontinuous V test space

$$\begin{bmatrix} G_\omega & B_k \\ B_k^T & 0 \end{bmatrix} = \begin{bmatrix} L_k \\ 0 \end{bmatrix}$$

The G matrix has a block structure, by element.

- Solve the Schur complement.

$$B_k^T G_\omega^{-1} B_k \mathbf{u} = B_k^T G_\omega^{-1} L_k.$$

Numerical results, parametric equation

1D time-harmonic wave propagation problem:

$$\begin{aligned} iku + u' &= 0, && \text{in } (0,1) \\ u(0) &= u_0 \end{aligned}$$

- 10 elements for the trial mesh.
- 20 elements for the test space.

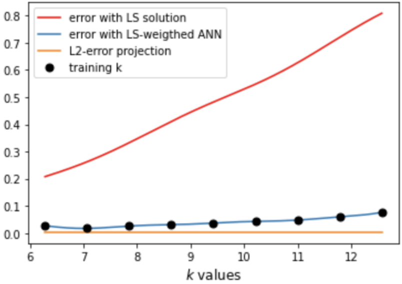


Table of Contents

A DPG-approach

Let's consider finding u_k in the advection-reacction parametric equation, such as:

$$u' + ku = 0$$
$$u(0) = u_0$$

Consider $\Omega_h = (0,1)$ trial mesh, y Ω_{2h} test mesh. Find $u \in L^2(\Omega_{2h})$, and $\hat{u} \in H^{1/2}(\Omega_h)$, such that, for all $w \in H^1(\Omega_h)$.
For smooth functions,

$$\underbrace{\sum_{i=1}^{m} \int_{x_{i-1}}^{x_i} -uv' + kuvdx + (\hat{u}_i v)\Big|_{x_{i-1}}^{x_i}}_{b((u,\hat{u}),v)} = \underbrace{\int_0^1 f_k v}_{\ell(v)}$$

- $U_h := P_0^d(\Omega_h^{\text{trial}}) \times P_1^c(\partial\Omega_h^{\text{test}})$, $\dim(U_h) = 3$. (Trial mesh 1 element)
- $V_h := P_1(\partial\Omega_h^{\text{test}})$. $\dim(V_h) = 4$. (Test mesh 2 elements)

The practical DPG method

Find $(e_{h,k,\omega}, (u_{h,k,\omega}, \hat{u}_{h,k,\omega})) \in V_h \times U_h$ s.t.

$$\begin{aligned} (e_{h,k,\omega}, v_h)_{V,\omega} + b_k((u_{h,k,\omega}, \hat{u}_{h,k,\omega}), v_h) &= \ell_k(v), \quad && \forall\, v_h \in V_h, \\ b_k((w_{h,k,\omega}, \hat{w}_{h,k,\omega}), e_{h,k,\omega}) &= 0, && \forall\, (w_{h,k,\omega}, \hat{w}_{h,k,\omega}) \in U_h. \end{aligned}$$

The matrix form:

$$\begin{bmatrix} G_\omega & B_k \\ B_k^T & 0 \end{bmatrix} = \begin{bmatrix} L_k \\ 0 \end{bmatrix}$$

The G matrix has a block structure, by element.

- Solve the Schur complement.

$$B_k^T G_\omega^{-1} B_k \mathbf{u} = B_k^T G_\omega^{-1} L_k.$$

Steps

- We train (**off-line**) with different parameters, and define a family of positive weight-functions to be used in the weighted inner products $\{(\cdot,\cdot)_{V,\omega}, \omega \in W\}$.
- Next, given the discrete U_h, we construct the map $W \times \{k_i\}_{i=1}^{i=n} \ni (\omega, k_i) \to \mathrm{u}_{h,k_i\omega} \in U_h$, and proceed to train the ANN by computing:

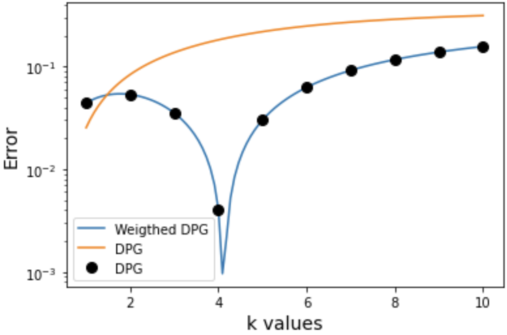
$$\theta^* = \operatorname{argmin}_\theta \frac{1}{2} \sum_{i=1}^{n} |q(u_{h,k_i,w}) - q(u_{k_i})|^2,$$

- Finally, we build the matrices of the linear system needed for the **online step**. We solve $\omega^*(\cdot) = \sigma_L(ANN(\cdot, \theta^*))$, we construct the matrices and solve.

Numerical results

How to approach a quantity of interest for a set of k- parameters ?

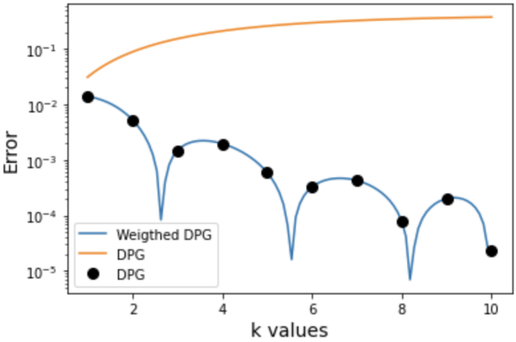
$$\mathbf{qoI}(u_k) := u_k(0.7)$$



Numerical results

How to approach a quantity of interest for a set of k- parameters ?

$$\mathbf{qoI}(u_k) := u_k(0.3)$$



On going work

- Different Cost functions to fix the phase gap.
- Instead of using positive piece-wise constants weigths, use positive functions.
- Study the phase gap of 1D time-harmonic wave propagation problem with DPG.

- S. Mishra, **A machine learning framework for data driven acceleration of computations of differential equations,** Mathematics in Engineering, 1 (2018), pp. 118–146.
- I. Brevis, I. Muga, and K. G. van der Zee, **A machine-learning minimal-residual (ML-MRes) framework for goal-oriented finite element discretizations.** Comput. Math. Appl., Vol. 95, pp. 186–199, 2021.
- I. Brevis, I. Muga, and K. G. van der Zee, **Neural Control of Discrete Weak Formulations: Galerkin, Least-Squares and Minimal-Residual Methods with Quasi-Optimal Weights** , arXiv:2206.07475, 2022.

Thank you!

E-mail: paulina.sepulveda@pucv.cl