AONN: An adjoint-oriented neural network method for all-at-once solutions of parametric optimal control problems

Changsha Institute for Computing and Digital Economy, Peking University

Kejun Tang
tangkejun@icode.pku.edu.cn

May 20, 2023

Joint work with Pengfei Yin, Guangqiang Xiao and Chao Yang

Outline

Background

- Aeronautics
- Microelectronics
- Reservoir simulations
- ...

Background

- Mathematical (physical) model: PDEs or ODEs
- Data-driven model (e.g., deep neural networks): no proper physical model but massive available data
- Numerical methods
  Both of them need numerical methods

Problem setup

OCP($\boldsymbol{\mu}$) Parametric optimal control problem: for any $\boldsymbol{\mu}$, find the solution to

$$\begin{aligned} &\min_{(y(\mathbf{x},\boldsymbol{\mu}),u(\mathbf{x},\boldsymbol{\mu}))\in Y\times U} J(y(\mathbf{x},\boldsymbol{\mu}),u(\mathbf{x},\boldsymbol{\mu});\boldsymbol{\mu}), \\ &\text{s.t.}\quad \mathbf{F}(y(\mathbf{x},\boldsymbol{\mu}),u(\mathbf{x},\boldsymbol{\mu});\boldsymbol{\mu})=0 \ \text{ in } \Omega(\boldsymbol{\mu}),\ \text{and}\ \ u(\mathbf{x},\boldsymbol{\mu})\in U_{ad}(\boldsymbol{\mu}), \end{aligned}.$$

- $\boldsymbol{\mu}\in\mathcal{P}\subset\mathbb{R}^D$: a vector that collects a finite number of parameters
- $\Omega(\boldsymbol{\mu})\subset\mathbb{R}^d$: a spatial domain depending on $\boldsymbol{\mu}$
- $\mathbf{x}\in\Omega(\boldsymbol{\mu})$: a spatial variable
- $J: Y\times U\times\mathcal{P}\mapsto\mathbb{R}$: a parameter-dependent objective functional. Y and U are two proper function spaces defined on $\Omega(\boldsymbol{\mu})$
- $\mathbf{F}$: the governing equation, parameter-dependent PDEs
- $U_{ad}(\boldsymbol{\mu})$: a parameter-dependent bounded closed convex subset of U

Problem setup

OCP($\boldsymbol{\mu}$) Parametric optimal control problem: for any $\boldsymbol{\mu}$, find the solution to

$$\begin{aligned} &\min_{(y(\mathbf{x},\boldsymbol{\mu}),u(\mathbf{x},\boldsymbol{\mu}))\in Y\times U} J(y(\mathbf{x},\boldsymbol{\mu}),u(\mathbf{x},\boldsymbol{\mu});\boldsymbol{\mu}), \\ &\text{s.t.} \quad \mathbf{F}(y(\mathbf{x},\boldsymbol{\mu}),u(\mathbf{x},\boldsymbol{\mu});\boldsymbol{\mu}) = 0 \ \text{ in } \Omega(\boldsymbol{\mu}), \ \text{ and } \ u(\mathbf{x},\boldsymbol{\mu}) \in U_{ad}(\boldsymbol{\mu}), \end{aligned} .$$

- The presence of parameters introduces extra prominent complexity
- Obtaining all-at-once solutions is challenge
- Additional constraints (e.g. box constraints) make NN-based methods hard to train

Problem setup

The corresponding KKT system

$$\begin{cases} J_y(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) - \mathbf{F}_y^*(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) p^*(\boldsymbol{\mu}) = 0, \\ \mathbf{F}(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) = 0, \\ (\mathrm{d}_u J(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}), v(\boldsymbol{\mu}) - u^*(\boldsymbol{\mu})) \geq 0, \ \forall v(\boldsymbol{\mu}) \in U_{ad}(\boldsymbol{\mu}). \end{cases}$$

- $(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}))$: the minimizer
- $p^*(\boldsymbol{\mu})$: the adjoint function which is also known as the Lagrange multiplier
- $\mathbf{F}_y^*(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu})$: the adjoint operator of $\mathbf{F}_y(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu})$
- $\mathrm{d}_u J(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) =$
  $J_u(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) - \mathbf{F}_u^*(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) p^*(\boldsymbol{\mu})$.

Main idea

The KKT system

$$\begin{cases} J_y(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) - \mathbf{F}_y^*(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) p^*(\boldsymbol{\mu}) = 0, \\ \mathbf{F}(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) = 0, \\ (\mathrm{d}_u J(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}), v(\boldsymbol{\mu}) - u^*(\boldsymbol{\mu})) \geq 0, \ \forall v(\boldsymbol{\mu}) \in U_{ad}(\boldsymbol{\mu}). \end{cases}$$

Solving this KKT system to get the optimal solution

- three neural networks to approximate $y^*(\boldsymbol{\mu})$, $u^*(\boldsymbol{\mu})$ and $p^*(\boldsymbol{\mu})$ separately
- deal with the parameters

goal: obtain the optimal solution for any parameters

Main idea

The KKT system

$$\begin{cases} J_y(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) - \mathbf{F}_y^*(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) p^*(\boldsymbol{\mu}) = 0, \\ \mathbf{F}(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}) = 0, \\ (\mathrm{d}_u J(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu}), v(\boldsymbol{\mu}) - u^*(\boldsymbol{\mu})) \geq 0, \ \forall v(\boldsymbol{\mu}) \in U_{ad}(\boldsymbol{\mu}). \end{cases}$$

Solving this KKT system to get the solution

- $\hat{y}(\mathbf{x}(\boldsymbol{\mu}); \boldsymbol{\theta}_y)$, $\hat{u}(\mathbf{x}(\boldsymbol{\mu}); \boldsymbol{\theta}_u)$, and $\hat{p}(\mathbf{x}(\boldsymbol{\mu}); \boldsymbol{\theta}_p)$: three independent deep neural networks
- $\mathbf{x}(\boldsymbol{\mu}) = \left[\ x_1, \ \ldots, \ x_d, \ \mu_1, \ \ldots, \ \mu_D \ \right].$

key point: construct a proper loss function

Main idea

$$\mathcal{L}_s(\boldsymbol{\theta}_y, \boldsymbol{\theta}_u) = \left(\frac{1}{N} \sum_{i=1}^{N} \left| r_s\left(\hat{y}(\mathbf{x}(\boldsymbol{\mu})_i; \theta_y), \hat{u}(\mathbf{x}(\boldsymbol{\mu})_i; \theta_u); \boldsymbol{\mu}_i\right) \right|^2 \right)^{\frac{1}{2}}, \tag{1a}$$

$$\mathcal{L}_a(\boldsymbol{\theta}_y, \boldsymbol{\theta}_u, \boldsymbol{\theta}_p) = \left(\frac{1}{N} \sum_{i=1}^{N} \left| r_a\left(\hat{y}(\mathbf{x}(\boldsymbol{\mu})_i; \theta_y), \hat{u}(\mathbf{x}(\boldsymbol{\mu})_i; \theta_u), \hat{p}(\mathbf{x}(\boldsymbol{\mu})_i; \theta_p); \boldsymbol{\mu}_i\right) \right|^2 \right)^{\frac{1}{2}}, \tag{1b}$$

$$\mathcal{L}_u(\boldsymbol{\theta}_u, u_{\text{step}}) = \left(\frac{1}{N} \sum_{i=1}^{N} \left| \hat{u}(\mathbf{x}(\boldsymbol{\mu})_i; \theta_u) - u_{\text{step}}(\mathbf{x}(\boldsymbol{\mu})_i) \right|^2 \right)^{\frac{1}{2}}. \tag{1c}$$

$$r_s(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu}) \triangleq \mathbf{F}(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu}), \tag{2a}$$

$$r_a(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}), p(\boldsymbol{\mu}); \boldsymbol{\mu}) \triangleq J_y(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu}) - \mathbf{F}_y^*(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu}) p(\boldsymbol{\mu}), \tag{2b}$$

Main idea

- $r_s(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu})$: residual of the state equation
- $r_a(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}), p(\boldsymbol{\mu}); \boldsymbol{\mu})$: residual of the adjoint equation
- $u_{\text{step}}(\mathbf{x}(\boldsymbol{\mu}))$: an intermediate variable for the third inequality in the KKT system

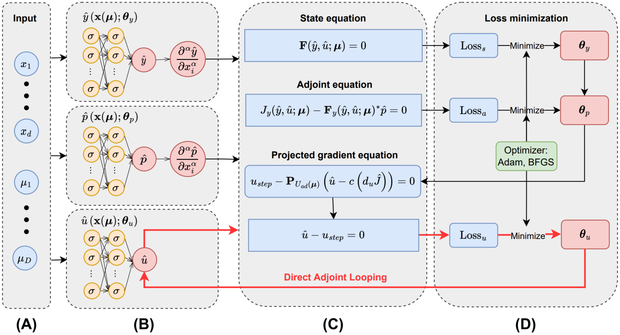


Figure: (A) Inputs (B) AONN: three separate neural networks $\hat{y}, \hat{p}, \hat{u}$ (C) The corresponding loss functions. (D) $\hat{y}, \hat{p}, \hat{u}$ are trained sequentially.

Some key ingredients

- the state equation and the adjoint equation: solving two parametric PDEs in $\Omega_{\mathcal{P}} = \{\mathbf{x}(\boldsymbol{\mu}) : \mathbf{x} \in \Omega(\boldsymbol{\mu})\}$
- projection gradient descent for inequality constraints in the KKT system

$$\mathbf{P}_{U_{ad}(\boldsymbol{\mu})}(u(\boldsymbol{\mu})) = \arg \min_{v(\boldsymbol{\mu}) \in U_{ad}(\boldsymbol{\mu})} \|u(\boldsymbol{\mu}) - v(\boldsymbol{\mu})\|_2,$$

$$u_{\text{step}}(\boldsymbol{\mu}) = \mathbf{P}_{U_{ad}(\boldsymbol{\mu})}\left(u(\boldsymbol{\mu}) - c\mathrm{d}_u J(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu})\right).$$

Because the optimal control function $u^*(\boldsymbol{\mu})$ satisfies

$$u^*(\boldsymbol{\mu}) - \mathbf{P}_{U_{ad}(\boldsymbol{\mu})}\left(u^*(\boldsymbol{\mu}) - c\mathrm{d}_u J(y^*(\boldsymbol{\mu}), u^*(\boldsymbol{\mu}); \boldsymbol{\mu})\right) = 0, \quad \forall c \geq 0.$$

The residual for the control function

$$r_v(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}), p(\boldsymbol{\mu})) \triangleq u(\boldsymbol{\mu}) - \mathbf{P}_{U_{ad}(\boldsymbol{\mu})}\left(u(\boldsymbol{\mu}) - c\mathrm{d}_u J(y(\boldsymbol{\mu}), u(\boldsymbol{\mu}); \boldsymbol{\mu})\right).$$

AONN algorithm

- training $\hat{y}(\mathbf{x}(\boldsymbol{\mu}); \boldsymbol{\theta}_y)$ for the state function

$$\boldsymbol{\theta}_y^k = \arg\min_{\boldsymbol{\theta}_y} \mathcal{L}_s \left(\boldsymbol{\theta}_y, \boldsymbol{\theta}_u^{k-1}\right).$$

- updating $\hat{p}(\mathbf{x}(\boldsymbol{\mu}); \boldsymbol{\theta}_p)$ for the adjoint function

$$\boldsymbol{\theta}_p^k = \arg\min_{\boldsymbol{\theta}_p} \mathcal{L}_a \left(\boldsymbol{\theta}_y^k, \boldsymbol{\theta}_u^{k-1}, \boldsymbol{\theta}_p\right).$$

- refining $\hat{u}(\mathbf{x}(\boldsymbol{\mu}); \boldsymbol{\theta}_u)$ for the control function

$$\boldsymbol{\theta}_u^k = \arg\min_{\boldsymbol{\theta}_u} \mathcal{L}_u \left(\boldsymbol{\theta}_u, u_{\text{step}}^{k-1}\right).$$

Comparison with other methods

- A straightforward way:

$$\mathrm{OCP}: \quad \begin{cases} \min_{(y,u)\in Y\times U} J(y,u), \\ \text{s.t. } \mathbf{F}(y,u)=0 \text{ in } \Omega, \quad \text{and } u \in U_{ad}. \end{cases}$$

cannot handle parametric optimal control efficiently

- NN-based methods

$$\min_{(y,u)\in Y\times U} J(y,u) + \beta_1 \mathbf{F}(y,u)^2 + \beta_2 \|u - \mathbf{P}_{U_{ad}}(u)\|_U + \beta_3 \ldots$$

too many penalty terms lead to failure and not suitable for nonsmooth problems

Numerical results

We start with the following nonparametric optimal control problem:

$$\begin{cases} \min\limits_{y,u} J(y,u) := \dfrac{1}{2}\|y-y_d\|^2_{L_2(\Omega)} + \dfrac{\alpha}{2}\|u\|^2_{L_2(\Omega)}, \\ \text{subject to } \begin{cases} -\Delta y + y^3 = u + f & \text{in } \Omega \\ y = 0 & \text{on } \partial\Omega, \end{cases} \\ \text{and} \quad u_a \le u \le u_b \quad \text{a.e. in } \Omega. \end{cases}$$

The corresponding adjoint equation

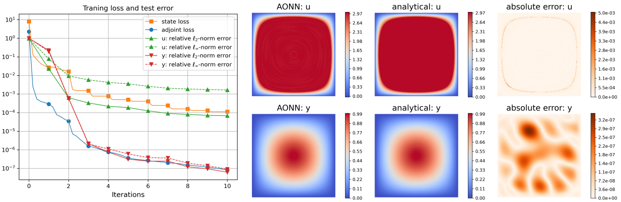
$$\begin{cases} -\Delta p + 3py^2 = y - y_d & \text{in } \Omega, \\ p = 0 & \text{on } \partial\Omega. \end{cases}$$

where $\Omega = (0,1)^2$, $\alpha = 0.01$, $u_a = 0$, and $u_b = 3$.

Numerical results

The analytical optimal solution is given by

$$
\begin{aligned}
y^* &= \sin(\pi x_1)\sin(\pi x_2), \\
u^* &= \mathbf{P}_{[u_a, u_b]}(2\pi^2 y^*), \text{pointwise projection operator onto } [u_a, u_b] \\
p^* &= -2\alpha\pi^2 y^*,
\end{aligned}
$$



(a) (b)

Figure: Test 1: training loss and test error. Test error is evaluated at 256×256 uniform grid points. (a) Loss behaviour test errors in both ℓ_2-norm and ℓ_∞-norm during training process. (b) Solution and error

Numerical results

The parametric version

$$\begin{cases} \min_{y(\boldsymbol{\mu}),u(\boldsymbol{\mu})} J(y(\boldsymbol{\mu}),u(\boldsymbol{\mu})) := \frac{1}{2}\|y(\boldsymbol{\mu})-y_d(\boldsymbol{\mu})\|^2_{L_2(\Omega)} + \frac{\alpha}{2}\|u(\boldsymbol{\mu})\|^2_{L_2(\Omega)}, \\ \text{subject to } \begin{cases} -\Delta y(\boldsymbol{\mu}) + y(\boldsymbol{\mu})^3 = u(\boldsymbol{\mu}) + f(\boldsymbol{\mu}) & \text{in } \Omega \\ y(\boldsymbol{\mu}) = 0 & \text{on } \partial\Omega, \end{cases} \\ \text{and} \quad u_a \le u(\boldsymbol{\mu}) \le \boldsymbol{\mu} \quad \text{a.e. in } \Omega. \end{cases}$$

where u_b is set to be a continuous variable $\boldsymbol{\mu}$ ranging from 3 to 20.

Numerical results

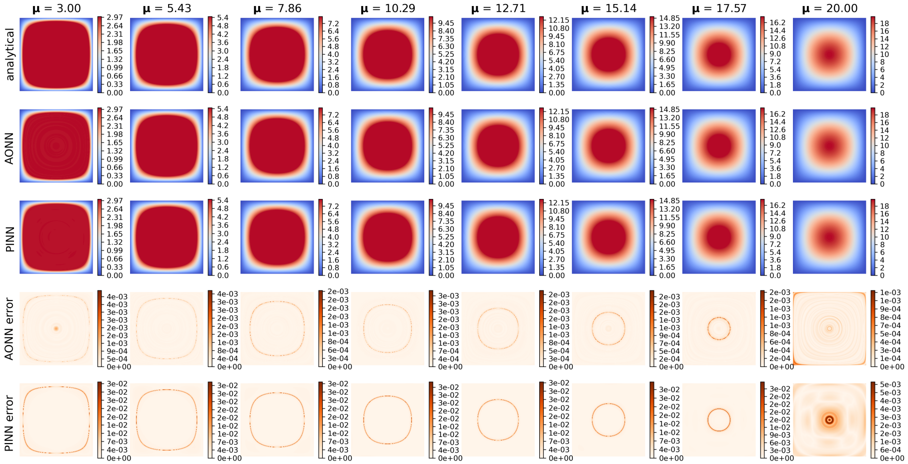


Figure: Test 2: the control solutions $u(\boldsymbol{\mu})$ of AONN and PINN with eight realizations of $\mu \in [3, 20]$, and their absolute errors.

Numerical results

Optimal control for the Navier-Stokes equations with physical parametrization

$$\min_{y(\boldsymbol{\mu}),u(\boldsymbol{\mu})} J(y(\boldsymbol{\mu}),u(\boldsymbol{\mu})) = \frac{1}{2}\|y(\boldsymbol{\mu}) - y_d(\boldsymbol{\mu})\|^2_{L_2(\Omega)} + \frac{1}{2}\|u(\boldsymbol{\mu})\|^2_{L_2(\Omega)},$$

$$\begin{cases} -\boldsymbol{\mu}\Delta y(\boldsymbol{\mu}) + (y(\boldsymbol{\mu})\cdot\nabla)y(\boldsymbol{\mu}) + \nabla p(\boldsymbol{\mu}) = u(\boldsymbol{\mu}) + f(\boldsymbol{\mu}) & \text{in } \Omega, \\ \operatorname{div} y(\boldsymbol{\mu}) = 0 & \text{in } \Omega, \\ y(\boldsymbol{\mu}) = 0 & \text{on } \partial\Omega, \end{cases}$$

where $\Omega = (0,1)^2$ with a parameter $\boldsymbol{\mu} \in [0.1, 100]$ representing the reciprocal of the Reynolds number, and a constraint for u
$u_1(\boldsymbol{\mu})^2 + u_2(\boldsymbol{\mu})^2 \leq r^2$ with $r = 0.2$

Numerical results

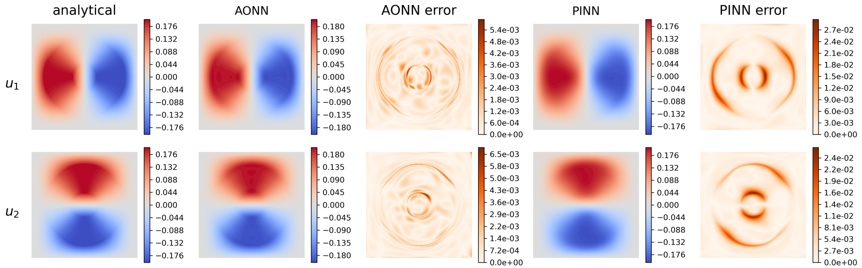


Figure: Test 3: optimal solutions of the control function $u = (u_1, u_2)$ obtained by AONN and PINN, and their absolute errors for a given parameter $\mu = 10$.

Numerical results

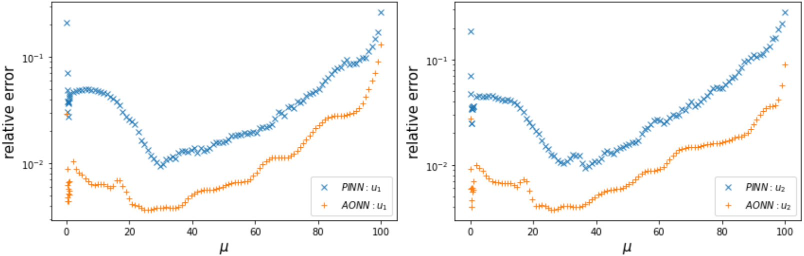


Figure: Test 3: the relative errors (in the ℓ_2-norm sense) of AONN and PINN for the two components of $u(\boldsymbol{\mu}) = (u_1(\boldsymbol{\mu}), u_2(\boldsymbol{\mu}))$. The relative errors are computed on the 256×256 meshgrid for each fixed parameter $\boldsymbol{\mu}$.

Numerical results

$$\begin{cases} \min\limits_{y(\boldsymbol{\mu}),u(\boldsymbol{\mu})} J(y(\boldsymbol{\mu}),u(\boldsymbol{\mu})) = \frac{1}{2}\|y(\boldsymbol{\mu})-y_d(\boldsymbol{\mu})\|^2_{L_2(\Omega(\boldsymbol{\mu}))} + \frac{\alpha}{2}\|u(\boldsymbol{\mu})\|^2_{L_2(\Omega(\boldsymbol{\mu}))}, \\ \text{subject to } \begin{cases} -\Delta y(\boldsymbol{\mu}) = u(\boldsymbol{\mu}) & \text{in } \Omega(\boldsymbol{\mu}), \\ y(\boldsymbol{\mu}) = 1 & \text{on } \partial\Omega(\boldsymbol{\mu}), \end{cases} \\ \text{and} \quad u_a \le u(\boldsymbol{\mu}) \le u_b \quad \text{a.e. in } \Omega(\boldsymbol{\mu}), \end{cases}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ is the paramter.
$\Omega(\boldsymbol{\mu}) = ([0,2]\times[0,1])\backslash B((1.5,0.5),\mu_1)$ and the desired state is given by

$$y_d(\boldsymbol{\mu}) = \begin{cases} 1 & \text{in } \Omega_1 = [0,1]\times[0,1], \\ \mu_2 & \text{in } \Omega_2(\boldsymbol{\mu}) = ([1,2]\times[0,1])\backslash B((1.5,0.5),\mu_1), \end{cases}$$

where $B((1.5,0.5),\mu_1)$ is a ball of radius μ_1 with center $(1.5,0.5)$, $\alpha = 0.001$ and $\boldsymbol{\mu} \in \mathcal{P} = [0.05,0.45]\times[0.5,2.5]$.

Numerical results

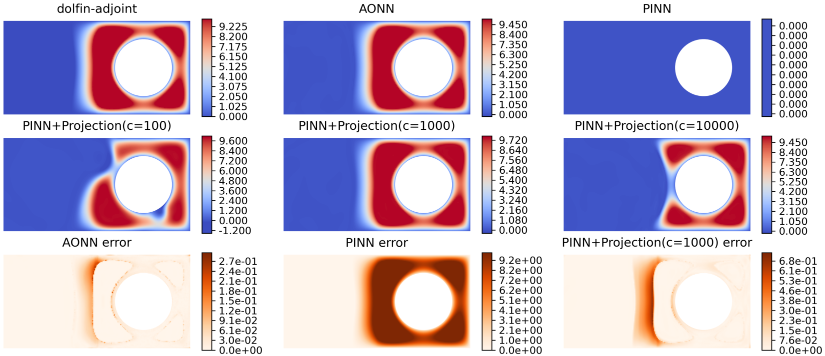
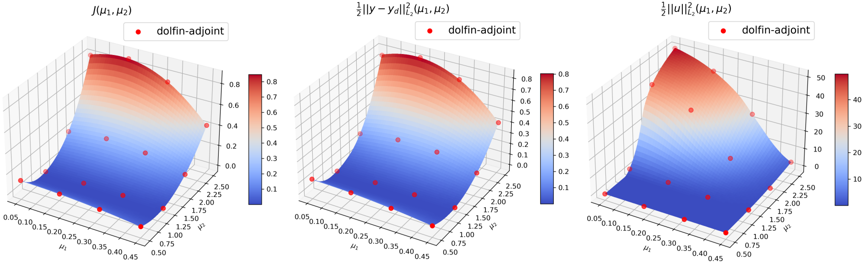


Figure: Test 4: the solution obtained by the dolfin-adjoint solver for a fixed parameter $\boldsymbol{\mu} = (0.3, 2.5)$, the approximate solutions of u obtained by AONN, PINN, PINN+Projection (with different $c = 100, 1000, 10000$), and the absolute errors of the AONN solution and the PINN+Projection solution with $c = \frac{1}{\alpha} = 1000$.

Numerical results



(a) (b) (c)

Figure: Test 4: several quantities as functions with respect to parameter $\boldsymbol{\mu} = (\mu_1, \mu_2)$ obtained by AONN. Each red dot denotes the quantity corresponding to a specific $\boldsymbol{\mu}$ computed from the dolfin-adjoint solver. (a) Objective value: J (b) Attainability of the desired state: $\frac{1}{2}\|y - y_d\|^2_{L_2}$. (c) L_2-norm of control function: $\frac{1}{2}\|u\|^2_{L_2}$.

Numerical results

$$
\begin{aligned}
&\min_{y(\boldsymbol{\mu}),u(\boldsymbol{\mu})} J := \frac{1}{2}\|y(\boldsymbol{\mu}) - y_d\|^2_{L_2(\Omega)} + \frac{\alpha}{2}\|u(\boldsymbol{\mu})\|^2_{L_2(\Omega)} + \boldsymbol{\mu}\|u(\boldsymbol{\mu})\|_{L_1(\Omega)}, \\
&\text{subject to } \begin{cases} -\Delta y(\boldsymbol{\mu}) + y(\boldsymbol{\mu})^3 = u(\boldsymbol{\mu}) & \text{in } \Omega, \\ y(\boldsymbol{\mu}) = 0 & \text{on } \partial\Omega, \end{cases} \\
&\text{and} \quad u_a \le u(\boldsymbol{\mu}) \le u_b \quad \text{a.e. in } \Omega.
\end{aligned}
$$

$$
\begin{aligned}
&\Omega = B(0,1), \\
&\alpha = 0.002, u_a = -12, u_b = 12, \\
&y_d = 4\sin(2\pi x_1)\sin(\pi x_2)\exp(x_1),
\end{aligned}
$$

The range of parameter is set to $\boldsymbol{\mu} \in [0, 0.128]$.

Numerical results

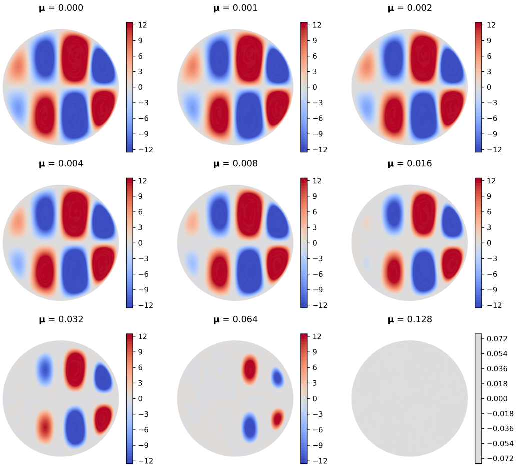


Figure: Test 5: the AONN solutions $u(\boldsymbol{\mu})$ of representative values for $\boldsymbol{\mu} = 2^i \times 10^{-3}, i = 0, 1, \ldots, 8$.

Numerical results

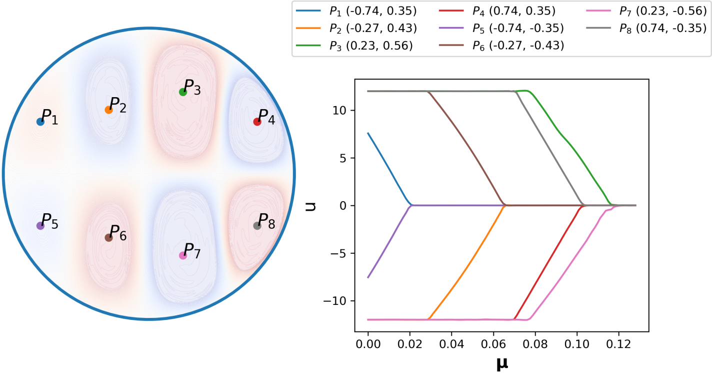


Figure: Test 5: the AONN solution $u(\boldsymbol{\mu})$ of eight fixed peaks $P_1 \sim P_8$ as a function respect to $\boldsymbol{\mu}$. The legend on the right is the coordinates of the eight points.

Sumary and outlook

summary

- develop AONN, an adjoint-oriented neural network method, for computing all-at-once solutions to parametric optimal control problems.
- integrate the idea of the direct-adjoint looping (DAL) approach in neural network approximation.
- meshless, without penalty-based loss function of the complex Karush–Kuhn–Tucker (KKT) system, thereby reducing the training difficulty of neural networks and improving the accuracy of solutions

outlook

- analysis
- adaptive sampling
- large scale problems and realistic applications

Q & A

Thank you for your attention