Overlapping Schwarz Domain Decomposition in Continuous-Time

Hongli Zhao — University of Chicago

Sen Na — Georgia Institute of Technology

Mihai Anitescu — University of Chicago

Main Contributions

- Generalization for exponential decay of sensitivity (EDS) of nonlinear control to infinite-dimension setting
- Gradient-based optimization methods inspired by *optimize-then-discretize* method in optimal control
- Extensions to deep learning (e.g. image, PINN)

Main Takeaway: A nonlinear optimization problem defined on a large domain / time-horizon can be divided into smaller subproblems that can be solved independently

Background

- Many problems (e.g. power planning, trajectory tracking, reinforcement learning) involve optimization of an objective constrained by dynamics
- Domain decomposition has long been applied for solving large-scale linear and nonlinear elliptic PDEs
- In constrained optimization, Schwarz methods have been shown to converge to full problem optimality
- But optimization with discretized variables does not allow adaptive time-stepping or higher-order solvers
- Generalizing the result to infinite-dimensional spaces allow for the design of flexible numerical methods

Problem Formulation

- Constrained optimization

$$\min_{\{u_k\},\{x_k\}} \sum_{k=0}^{N-1} L(t_k, x_k, u_k)\cdot \Delta t + \Phi(x_N)$$
$$\text{s.t.}\quad x_{k+1} = x_k + \Delta t \cdot f(t_k, x_k, u_k),\quad k = 0,\dots,N-1$$
$$x_0 = x_{\text{init}} \in \mathbb{R}^{n_x},\quad u_k \in \mathbb{R}^{n_u}$$

↓

- Continuous-time nonlinear control

$$\min_{u(\cdot),\,x(\cdot)} \int_0^T L(t, x(t), u(t))\,dt + \Phi(x(T))$$
$$\text{s.t.}\quad \dot{x}(t) = f(t, x(t), u(t)),\quad t\in(0,T],\quad u(t)\in\mathbb{R}^{n_u}$$
$$x(0) = x_0 \in \mathbb{R}^{n_x}$$

Overlapping Schwarz Decomposition

- For $j = 0, 1, \dots, m$, define subproblem:

$$\min_{u_j(\cdot),x_j(\cdot)} \int_{\tau_j^{(0)}}^{\tau_j^{(1)}} L(t, x_j(t), u_j(t))\,dt + \tilde{\Phi}_j\big(x_j(\tau_j^{(1)});\,\bar{w}_j\big)$$
$$\text{s.t.}\quad \dot{x}_j(t) = f(t, x_j(t), u_j(t)),\quad t\in(\tau_j^{(0)}, \tau_j^{(1)}]$$
$$x_j(\tau_j^{(0)}) = \bar{x}(\tau_j^{(0)})$$

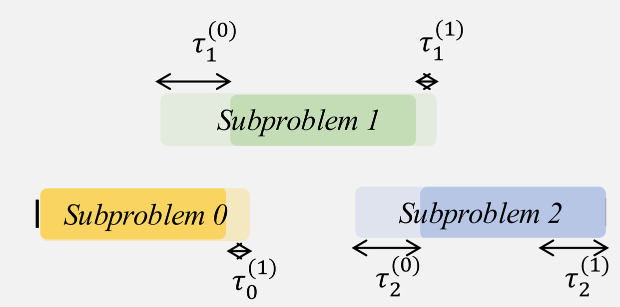


Figure: Dividing the domain into m = 3 sub-domains with overlaps

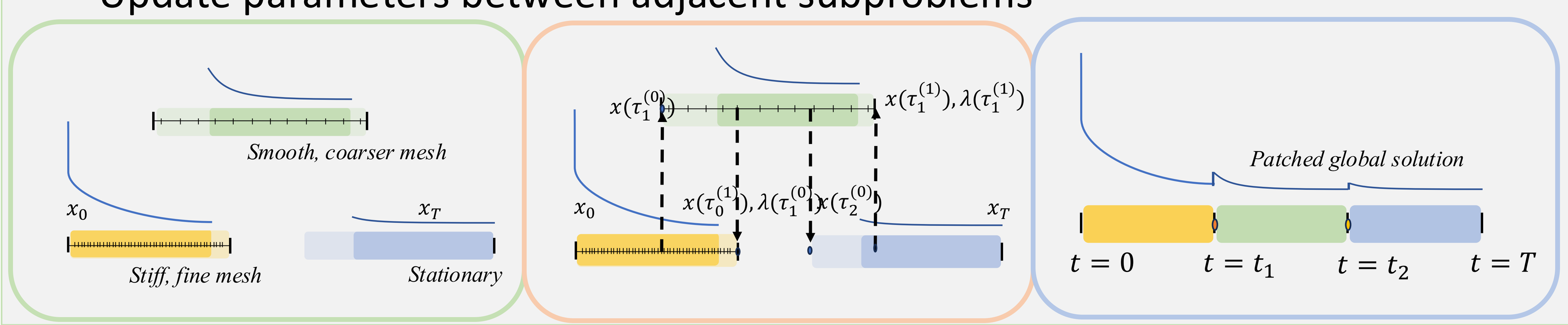
where: $\bar{x}, \bar{u}, \bar{\lambda}$ are external parameters, and the modified terminal cost is defined:

$$\tilde{\Phi}_j(x;\bar{w}_j) = \begin{cases} L(x, \bar{u}(\tau_j^{(1)})) - \bar{\lambda}_j^\top(\tau_j^{(1)}) f(t, x, \bar{u}(\tau_j^{(1)})) + \frac{\mu}{2}\left\| x - \bar{x}(\tau_j^{(1)})\right\|^2, & \text{if } j = 0,\dots,m-1 \\ \Phi(x), & \text{if } j = m \end{cases}$$

- Find the optimality of each subproblem by solving the coupled system:

$$\dot{x}_j(t) = f(t, x_j(t), u_j(t)),\qquad x_j(\tau_j^{(0)}) = \bar{x}(\tau_j^{(0)})$$
$$\dot{\lambda}_j(t) = -\nabla_x L(t, x_j(t), u_j(t)) - \nabla_x f(t, x_j(t), u_j(t))^\top \lambda_j(t),\qquad \lambda_j(\tau_j^{(1)}) = \nabla\tilde{\Phi}_j(x_j(\tau_j^{(1)});\bar{w}_j)$$
$$0 = \nabla_u L(t, x_j(t), u_j(t)) + \nabla_u f(t, x_j(t), u_j(t))^\top \lambda_j(t)$$

- Update parameters between adjacent subproblems



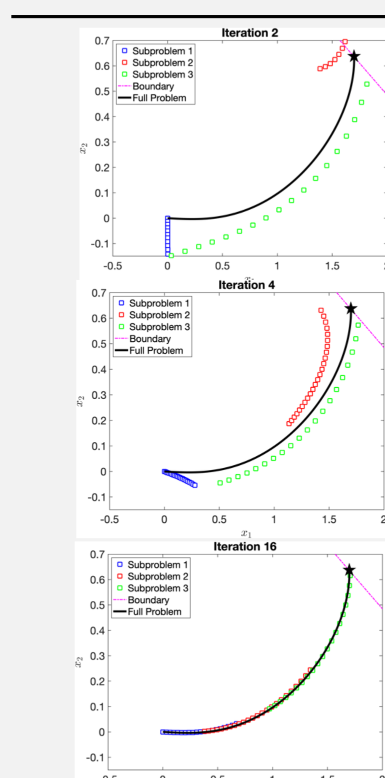
Main Proof: Exponential Decay of Sensitivity

- Sensitivity of optimal solutions (x^*, u^*, λ^*) to perturbations locally satisfies linear-quadratic control [1]
- **Special case**: linear-quadratic control with external data d

$$\min_{u(\cdot),x(\cdot)} \frac{1}{2}\int_0^T \begin{bmatrix} x(t)\\ u(t)\\ d(t)\end{bmatrix}^\top \begin{bmatrix} Q(t) & H^\top(t) & G^\top(t)\\ H(t) & R(t) & W^\top(t)\\ G(t) & W(t) & 0\end{bmatrix} \begin{bmatrix} x(t)\\ u(t)\\ d(t)\end{bmatrix} dt + \frac{1}{2}\begin{bmatrix} x(T)\\ d(T)\end{bmatrix}^\top \begin{bmatrix} Q_T & G_T^\top\\ G_T & 0\end{bmatrix}\begin{bmatrix} x(T)\\ d(T)\end{bmatrix},$$
$$\text{s.t.}\quad \dot{x}(t) = A(t)x(t) + B(t)u(t) + C(t)d(t),\quad t\in(0,T],$$
$$x(0) = d(0).$$

- Under SOSC and uniform complete controllability conditions, Pontryagin's minimization principle implies exponentially convergent linear evolution operator
- As a result, perturbations are exponentially damped as one moves into a domain
- Exists a choice of overlap size τ that yields a contractive mapping

Numerical Simulations



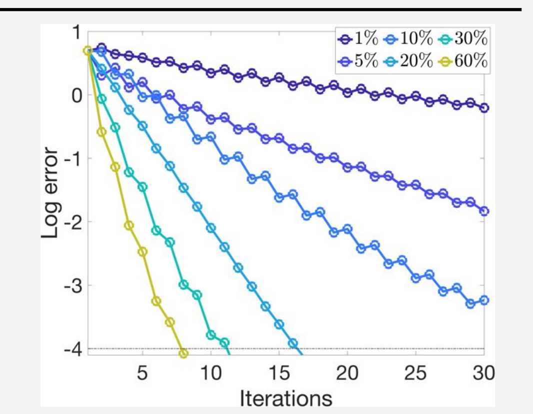
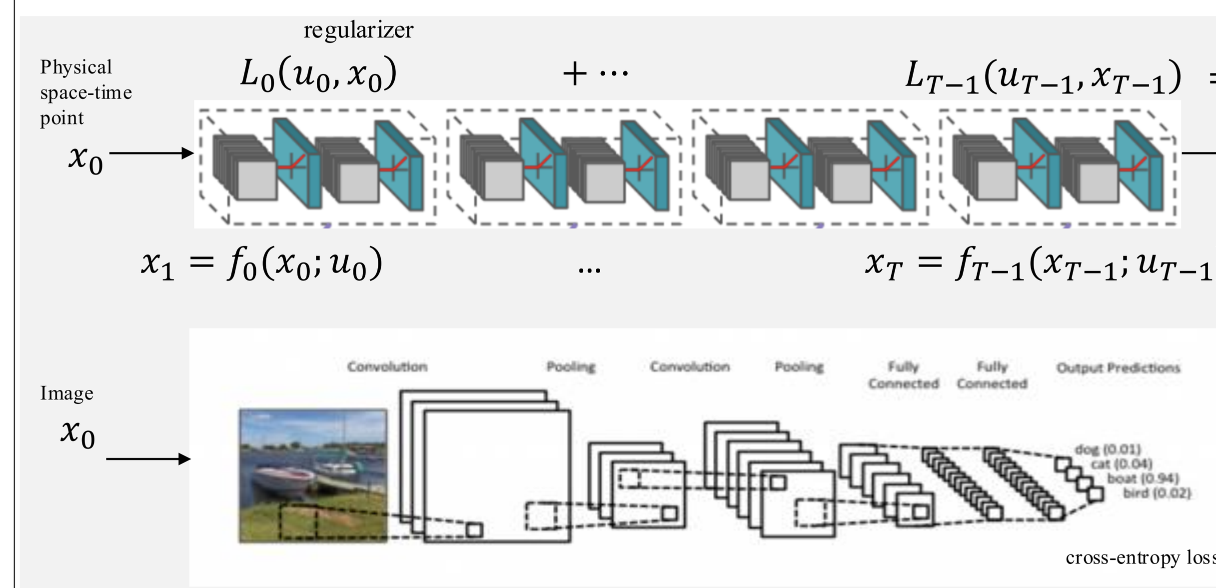


Figure: (Left) Overlapping Schwarz state solution visualization. (Right) Relative L^∞ error convergence to global optimal solution, with varying overlap sizes.

[1] H. Joseph et al. *Hyper-differential sensitivity analysis with respect to model discrepancy: Optimal solution updating (2024)*

Deep Learning

Insight: *ResNet is a discretized dynamical system* [3]



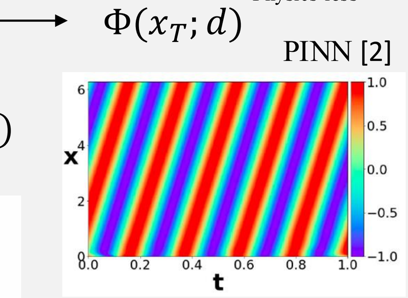


Figure: Illustrations of ResNets (top: ResNet for PINN applications; bottom:ResNet-CNN for image classification). The terminal costs can be modified to suit different learning tasks.

Empirical Study

Task 1: MNIST classification

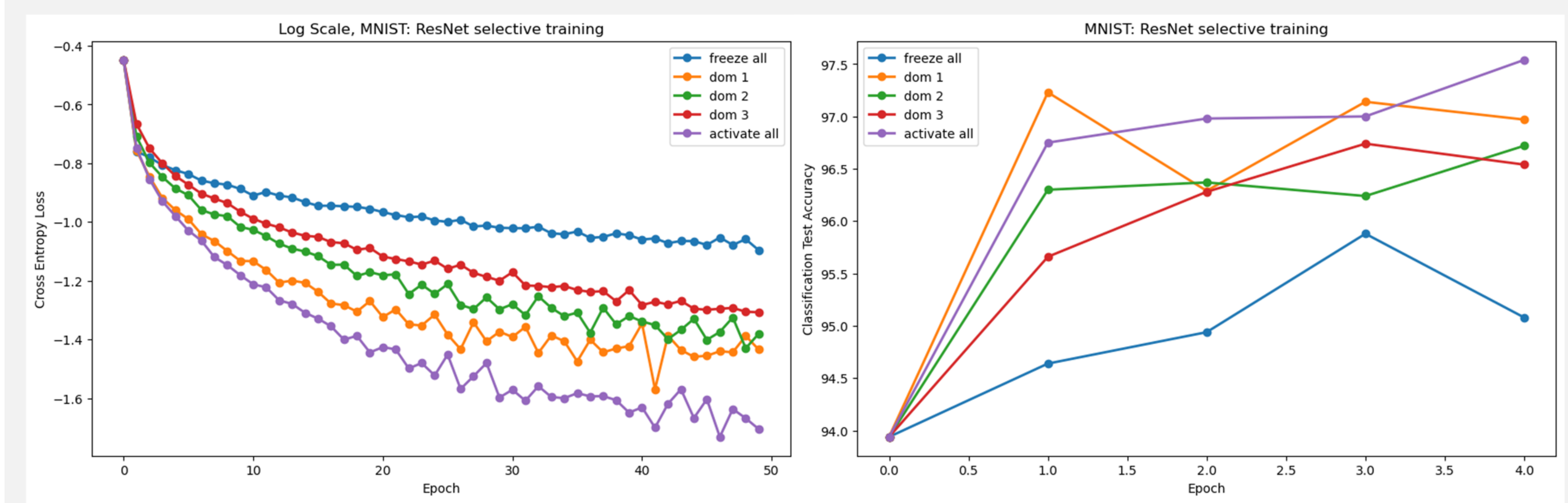
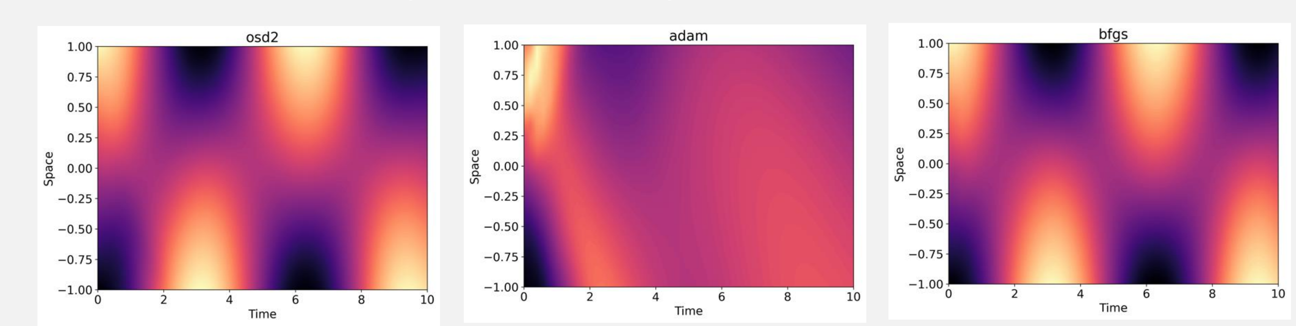


Figure 1: MNIST classification (Left: loss convergence; Right: classification accuracy) with overlapping Schwarz decomposition of 1-3 overlapped layers; the convergence rate is improved as overlap size increases.

Task 2: Klein-Gordon equation, comparison with Adam and LBFGS



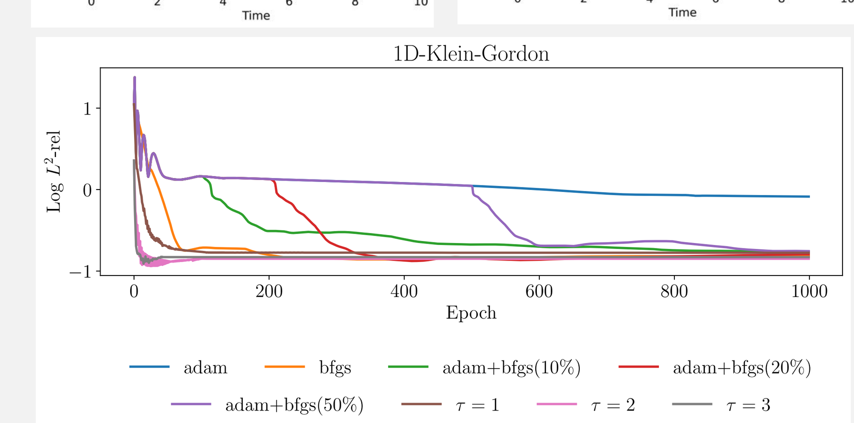


Figure 2: Comparison of PINN solutions of the Klein-Gordon equation optimized using Adam and LBFGS. The loss landscape shows that overlapping Schwarz decomposition converges much more rapidly.

[2] M. Mahoney et al. *Continuous-in-depth neural networks (2020)*

[3] Q. Li et al. *An Optimal Control Approach to Deep Learning and Applications to Discrete-Weight Neural Networks (2018)*

honglizhaobob@uchicago.edu