Spooky Boundaries at a Distance: Inductive Bias, Dynamic Models, and Behavioral Macro

Mahdi Ebrahimi Kahou[1] Jesús Fernández-Villaverde[2] Sebastián Gómez-Cardona[3]
Jesse Perla[4] Jan Rosa[4]

Conference on Frontiers in Machine Learning and Economics, Federal Reserve Bank of Philadelphia

[1]Bowdoin College

[2]University of Pennsylvania

[3]Morningstar, Inc.

[4]University of British Columbia

Motivation, Question, and Contribution

Motivation

In the long run, we are all dead—*J.M. Keynes, A Tract on Monetary Reform (1923)*

- Numerical solutions to dynamical systems are central to many quantitative fields in economics.
- Dynamical systems in economics are **boundary value** problems:
  1. The boundary is at **infinity**.
  2. The values at the boundary are potentially **unknown**.
- Resulting from **forward looking** behavior of agents.
- Examples include the *transversality* and the *no-bubble* condition.
- Without them, the problems are ill-posed and have infinitely many solutions:
  - The problems are ill-posed in the Hadamard sense, meaning the solutions are not unique.
  - These forward-looking boundary conditions are a key limitation on increasing dimensionality.

Question

Question:

Can we (economists and agents) **ignore** *these long-run boundary conditions and still have accurate short/medium-run dynamics disciplined by these long-run conditions?*

Contribution

1. **Yes**, it is possible to meet long-run boundary conditions **without** strictly enforcing them as a constraint on the model's dynamics.
    - We show how using Machine Learning (ML) methods achieve this method.
    - This is due to the **inductive bias** of ML methods.
    - In this paper focusing on <u>deep neural networks</u>
2. We argue how inductive bias can serve as a micro-foundation for modeling forward-looking behavioral agents.
    - Easy to compute.
    - Provides short-run accuracy.
    - Satisfies the necessary long-run constraints.

Background: Economic Models, Deep learning and inductive bias

Economic Models: functional equations

Many theoretical models can be written as functional equations:

- Economic object of interest: f where $f : \mathcal{X} \to \mathcal{R} \subseteq \mathbb{R}^N$
  - e.g., asset price, investment choice, best-response, etc.
- Domain of f: $\mathcal{X}$
  - e.g. space of dividends, capital, opponents state or time in sequential models.
- The "model" error: $\ell(x, f) = \mathbf{0}$, for all $x \in \mathcal{X}$
  - e.g., Euler and Bellman residuals, equilibrium FOCs.

Then a **solution** is an $f^* \in \mathcal{F}$ where $\ell(x, f^*) = \mathbf{0}$ for all $x \in \mathcal{X}$.

Approximate solution: deep neural networks

1. Sample $\mathcal{X}$: $\mathcal{D} = \{x_1, \cdots, x_N\}$
2. Pick a deep neural network $f_\theta(\cdot) \in \mathcal{H}(\theta)$:
   - θ: parameters for optimization (i.e., weights and biases).
3. To find an approximation for f solve:

$$\min_\theta \frac{1}{N} \sum_{x \in \mathcal{D}} \|\ell(x, f_\theta)\|_2^2 \tag{1}$$

- Deep neural networks are highly over-parameterized.
- Formally, $|\theta| \gg N$

Over-parameterized interpolation

- Over-parameterized ($|\theta| \gg N$), the optimization problem can have many solutions.
- Since individual θ are irrelevant it is helpful to think of optimization directly within $\mathcal{H}$

$$\min_{f_\theta \in \mathcal{H}} \sum_{x \in \mathcal{D}} \|\ell(x, f_\theta)\|_2^2$$

- But which f_θ?
- **Mental model:** chooses min-norm interpolating solution for a (usually) unknown functional norm ψ

$$\begin{aligned} &\min_{\hat{f} \in \mathcal{H}} ||f_\theta||_\psi \\ &\text{s.t. } \ell(x, f_\theta) = 0, \quad \text{for all } x \in \mathcal{D} \end{aligned}$$

- That is what we mean by **inductive bias** (see Belkin, 2021 and Ma and Yang, 2021).
- Characterizing S (e.g., sobolev norms or semi-norms?) is an active research area in ML.

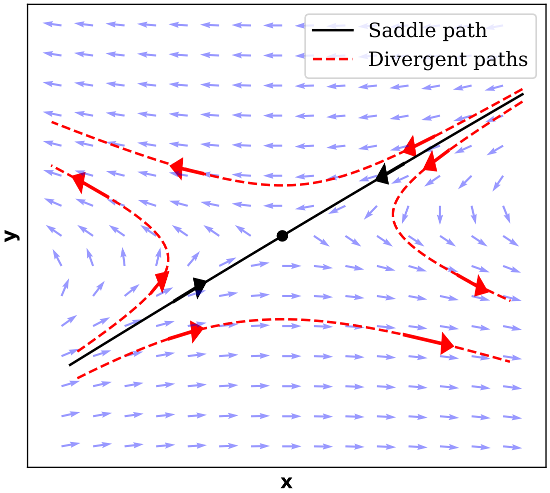
Smooth interpolation

- Intuition: biased toward solutions which are flattest and have smallest derivatives

Intuition

- **Minimum-norm implicit bias**:
  - Over-parameterized models (e.g., large neural networks) have more parameters than data points and potentially interpolate the data.
  - They are biased towards interpolating functions with smallest norm.

- **Violation of economic boundary conditions**:
  - Sub-optimal solutions diverge (explode) over time.
  - They have large or explosive norms.
  - This is due to the **saddle-path** nature of econ problems.



The Problem

The class of problems

A differential-algebraic system of equations, coming from an economic optimization problem:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t), \mathbf{y}(t)) \tag{2}$$

$$\dot{\mathbf{y}}(t) = \mathbf{G}(\mathbf{x}(t), \mathbf{y}(t)) \tag{3}$$

$$\mathbf{0} = \mathbf{H}(\mathbf{x}(t), \mathbf{y}(t)) \tag{4}$$

$\mathbf{x} \in \mathbb{R}^{N_x}$: state variables, $\mathbf{y} \in \mathbb{R}^{N_y}$: jump variables. Initial value $\mathbf{x}(0) = \mathbf{x}_0$ and boundary conditions (at infinity)

$$\mathbf{0} = \lim_{t \to \infty} \mathbf{B}(t, \mathbf{x}(t), \mathbf{y}(t)) \tag{5}$$

Goal: finding an approximation for $\mathbf{x}(t)$ and $\mathbf{y}(t)$.

What is the problem?

- $\mathbf{y}_0$ is unknown.
- The optimal solutions is a **saddle-path**: unstable nature

Method

Method

- Pick a set of points $\mathcal{D} \equiv \{t_1, \cdots, t_N\}$ for some fixed interval $[0, T]$
- Large machine learning models to learn $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{y}}(t)$

$$\min_{\hat{\mathbf{x}},\hat{\mathbf{y}}} \sum_{t_i \in \mathcal{D}} \Bigg[\eta_1 \underbrace{\left\| \dot{\hat{\mathbf{x}}}(t_i) - \mathbf{F}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)(t_i)) \right\|_2^2}_{\text{Residuals}^2\text{: state variables}} + \eta_2 \underbrace{\left\| \dot{\hat{\mathbf{y}}}(t_i) - \mathbf{G}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)) \right\|_2^2}_{\text{Residuals}^2\text{: jump variables}}$$

$$+ \eta_3 \underbrace{\left\| \mathbf{H}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)) \right\|_2^2}_{\text{Residuals}^2\text{: algebraic constraint}} \Bigg] + \eta_4 \underbrace{\left\| \hat{\mathbf{x}}(0) - \mathbf{x}_0 \right\|_2^2}_{\text{Residuals}^2\text{: initial conditions}} .$$

- This optimization **ignores** the boundary conditions.
- The implicit bias automatically satisfy the boundary conditions.
- Recent works suggest the implicit bias is toward smallest Sobolev semi-norms.

Ridgeless kernel regression

$$\hat{\dot{\mathbf{x}}}(t) = \sum_{j=1}^{N} \boldsymbol{\alpha}_j^x K(t, t_j), \quad \hat{\dot{\mathbf{y}}}(t) = \sum_{j=1}^{N} \boldsymbol{\alpha}_j^y K(t, t_j)$$

$$\hat{\mathbf{x}}(t) = \mathbf{x}_0 + \int_0^t \hat{\dot{\mathbf{x}}}(\tau) d\tau, \quad \hat{\mathbf{y}}(t) = \hat{\mathbf{y}}_0 + \int_0^t \hat{\dot{\mathbf{y}}}(\tau) d\tau$$

- $\mathbf{x}_0$ is given.
- $\hat{\mathbf{y}}_0$, $\boldsymbol{\alpha}_j^x$, and $\boldsymbol{\alpha}_j^y$ are learnable parameters.
- $K(\cdot, \cdot)$: Matérn Kernel, with smoothness parameter ν and length scale ℓ.

Ridgeless kernel regression: minimum Sobolev seminorm solutions

We also solve the ridgeless kernel regression

$$\lim_{\lambda \to 0} \min_{\hat{\mathbf{x}}, \hat{\mathbf{y}}} \sum_{t_i \in \mathcal{D}} \left[\eta_1 \left\| \dot{\hat{\mathbf{x}}}(t_i) - \mathbf{F}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)(t_i)) \right\|_2^2 + \eta_2 \left\| \dot{\hat{\mathbf{y}}}(t_i) - \mathbf{G}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)) \right\|_2^2 \right.$$

$$\left. + \eta_3 \left\| \mathbf{H}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)) \right\|_2^2 \right] + \eta_4 \left\| \hat{\mathbf{x}}(0) - \hat{\mathbf{x}}_0 \right\|_2^2 + \lambda \underbrace{\left[\sum_{m=1}^{N_x} \| \dot{\hat{\mathbf{x}}}^{(m)} \|_{\mathcal{H}}^2 + \sum_{m=1}^{N_y} \| \dot{\hat{\mathbf{y}}}^{(m)} \|_{\mathcal{H}}^2 \right]}_{\text{The Sobolev semi-norm}}$$

- Targeting Sobolev semi-norm.
- This choice is very natural: it solves the instability issues of the classical algorithm.

Applications

Linear asset pricing

$$\dot{\mathbf{x}}(t) = c + g\mathbf{x}(t) \tag{6}$$

$$\dot{\mathbf{y}}(t) = r\mathbf{y}(t) - \mathbf{x}(t) \tag{7}$$

$$0 = \lim_{t \to \infty} e^{-rt}\mathbf{y}(t) \tag{8}$$

- $\mathbf{x}(t) \in \mathbb{R}$: dividends, $\mathbf{y}(t) \in \mathbb{R}$: prices, and $\mathbf{x}_0$ given.
- Equation (6): how the dividends evolve in time.
- Equation (7): how the prices evolve in time.
- Equation (8): "no-bubble" condition, the boundary condition at infinity.

Why do we need the boundary condition?

$$\dot{\mathbf{x}}(t) = c + g\mathbf{x}(t)$$
$$\dot{\mathbf{y}}(t) = r\mathbf{y}(t) - \mathbf{x}(t)$$

- The solutions:

$$\mathbf{y}(t) = \mathbf{y}_f(t) + \zeta e^{rt}$$

- $\mathbf{y}_f(t) = \int_0^\infty e^{-r\tau}\mathbf{x}(t+s)ds$: price based on the fundamentals.
- ζe^{rt}: explosive bubble terms, it has to be **ruled out** by the boundary condition.
- Triangle inequality: $\|\mathbf{y}_f\| < \|\mathbf{y}\|$.
- The price based on the fundamentals has the **lowest norm**.

Conclusion

- Long-run (**global**) conditions can be replaced with appropriate regularization (**local**) to achieve the optimal solutions.
- The minimum-norm implicit bias of large ML models aligns with optimality in economic dynamic models.
- Both kernel and neural network approximations accurately learn the right steady state(s).
- Proceeding with **caution**: can regularization be thought of as an equilibrium selection device?

Appendix