How Inductive Bias in Machine Learning Aligns with Optimality in Economic Dynamics

Mahdi Ebrahimi Kahou[1] James Yu[2] Jesse Perla[2] Geoff Pleiss[3,4]

June 25, 2024

[1]Bowdoin College, Econ Dept

[2]University of British Columbia, Vancouver School of Economics

[3]University of British Columbia, Stats Dept

[4]Vector Institute

Motivation

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

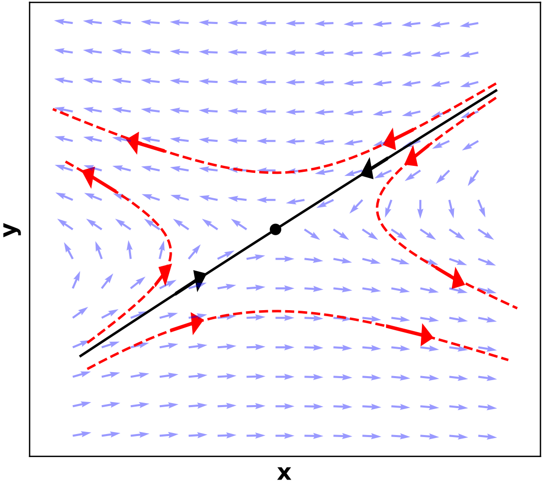
- Numerical solutions to dynamical systems are central to many quantitative fields in economics.
- Dynamical systems in economics are **boundary value** problems:
  1. Boundary is at **infinity**.
  2. The values at the boundary are potentially **unknown**. condition.
- Resulting from **forward looking** behavior of agents.
- Examples: Transversality and "no-bubble" condition.
- Without them the problems are ill-posed and have infinitely many solutions:
  - These forward-looking boundary conditions are the key limitation on increasing dimensionality.

Contribution

1. **Inductive bias alignment**:
   - The minimum-norm implicit bias of modern ML models automatically satisfies economic boundary conditions at infinity.
2. **Learning the right set of steady-states**:
   - Deep neural networks and kernel machines learn the boundary values, thereby extrapolating very accurately.
3. **Robustness and speed**:
   - Competitive in speed and more stable than traditional methods.
4. **Consistency of ML estimates**.

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 equations, coming from an economic optimization problem:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t), \mathbf{y}(t)) \tag{1}$$

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

$$\mathbf{0} = \mathbf{H}(\mathbf{x}(t), \mathbf{y}(t)) \tag{3}$$

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

$$\mathbf{0} = \lim_{t \to \infty} \mathbf{B}(t, \mathbf{x}(t), \mathbf{y}(t)) \tag{4}$$

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 $\hat{\mathbf{y}}(t)$

$$\min_{\hat{\mathbf{x}},\hat{\mathbf{y}}} \sum_{t_i \in \mathcal{D}} \left[\eta_1 \left\| \dot{\hat{\mathbf{x}}}(t_i) - \mathbf{F}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)(t_i)) \right\|_2^2 + \eta_2 \left\| \dot{\hat{\mathbf{y}}}(t_i) - \mathbf{G}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)) \right\|_2^2 \right.$$
$$\left. + \eta_3 \left\| \mathbf{H}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)) \right\|_2^2 + \eta_4 \left\| \hat{\mathbf{x}}(0) - \hat{\mathbf{x}}_0 \right\|_2^2 \right],$$

- 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

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

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

- $K(\cdot,\cdot)$: Matérn Kernel, with smoothness parameter ν and length scale ℓ.

We also solve the ridgeless kernel regression

$$\lim_{\lambda \to 0} \min_{\hat{\mathbf{x}},\hat{\mathbf{y}}} \sum_{t_i \in \mathcal{D}} \left[\eta_1 \left\| \dot{\hat{\mathbf{x}}}(t_i) - \mathbf{F}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)(t_i)) \right\|_2^2 + \eta_2 \left\| \dot{\hat{\mathbf{y}}}(t_i) - \mathbf{G}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)) \right\|_2^2 \right.$$
$$\left. + \eta_3 \left\| \mathbf{H}(\hat{\mathbf{x}}(t_i), \hat{\mathbf{y}}(t_i)) \right\|_2^2 \right] + \eta_4 \left\| \hat{\mathbf{x}}(0) - \hat{\mathbf{x}}_0 \right\|_2^2 + \lambda \left[\sum_{m=1}^{N_x} \|\dot{\hat{\mathbf{x}}}^{(m)}\|_{\mathcal{H}}^2 + \sum_{m=1}^{N_y} \|\dot{\hat{\mathbf{y}}}^{(m)}\|_{\mathcal{H}}^2 \right]$$

Applications

Linear asset pricing

$$\dot{\mathbf{x}}(t) = c + g\mathbf{x}(t) \tag{5}$$

$$\dot{\mathbf{y}}(t) = r\mathbf{y}(t) - \mathbf{x}(t) \tag{6}$$

$$0 = \lim_{t\to\infty} e^{-rt}\mathbf{y}(t) \tag{7}$$

- $\mathbf{x}(t) \in \mathbb{R}$: dividends, $\mathbf{y}(t) \in \mathbb{R}$: prices, and $\mathbf{x}_0$ given.
- Equation (5): how the dividends evolve in time.
- Equation (6): how the prices evolve in time.
- Equation (7): "no-bubble" condition, the boundary condition at infinity.

Why do we need the boundary condition?

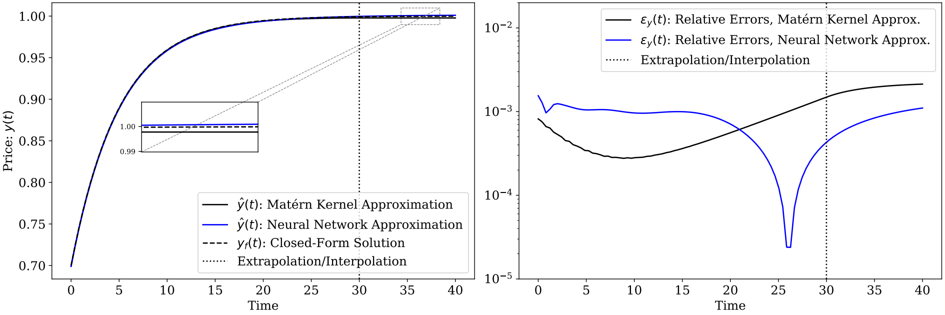
$$\dot{\mathbf{x}}(t) = c + g\mathbf{x}(t)$$
$$\dot{\mathbf{y}}(t) = r\mathbf{y}(t) - \mathbf{x}(t)$$

- 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: $\|\mathbf{y}_f\|_{\mathcal{H}} < \|\mathbf{y}\|_{\mathcal{H}}$.
- The price based on the fundamentals has the **lowest norm**.

Results



$$\mathcal{D} = \{0, 1, \cdots, 30\}$$

- The explosive solutions are ruled out without directly imposing the boundary condition.
- Provides a very accurate approximation, both in the short- and medium-run.
- Learns the steady-state.

Neoclassical growth model: the agent's problem

$$\max_{\mathbf{y}(t)} \int_0^\infty e^{-rt} \ln(\mathbf{y}(t)) dt$$
$$\text{s.t.} \quad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) - \mathbf{y}(t) - \delta \mathbf{x}(t)$$

for a given $\mathbf{x}_0$.

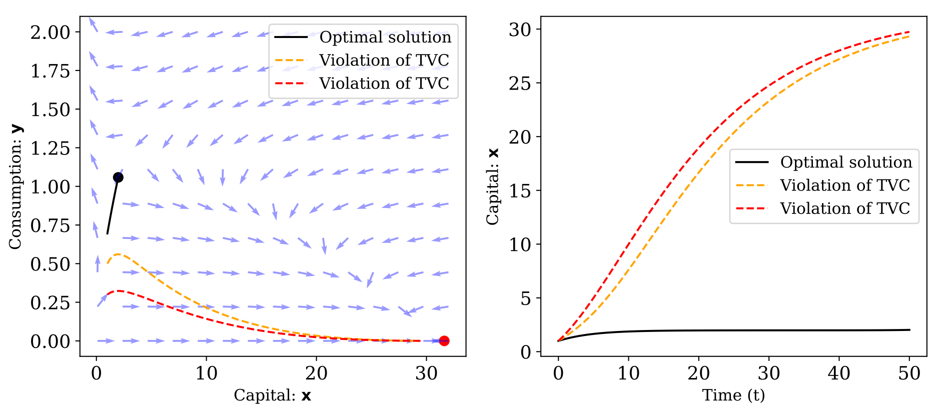
Constructing the Hamiltonian ...

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) - \mathbf{y}(t) - \delta \mathbf{x}(t) \tag{8}$$
$$\dot{\mathbf{y}}(t) = \mathbf{y}(t)\big[f'(\mathbf{x}(t)) - \delta - r\big] \tag{9}$$
$$0 = \lim_{t\to\infty} e^{-rt} \frac{\mathbf{x}(t)}{\mathbf{y}(t)} \tag{10}$$

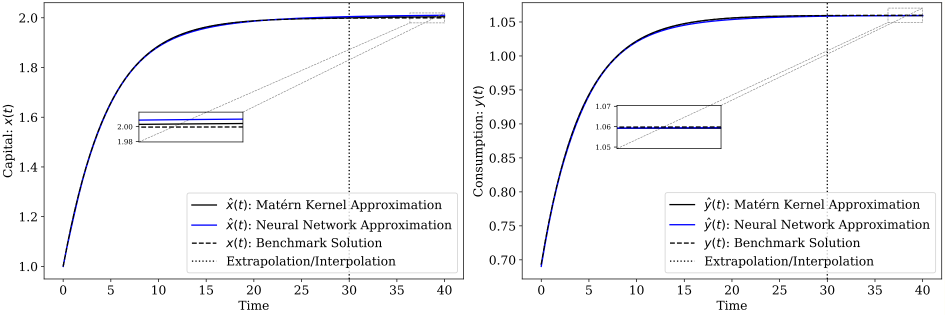
- $\mathbf{x}(t) \in \mathbb{R}$: capital and $\mathbf{y}(t) \in \mathbb{R}$: consumption.

Why do we need the boundary condition?

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) - \mathbf{y}(t) - \delta \mathbf{x}(t)$$
$$\dot{\mathbf{y}}(t) = \mathbf{y}(t)\left[f'(\mathbf{x}(t)) - \delta - r\right]$$



Results

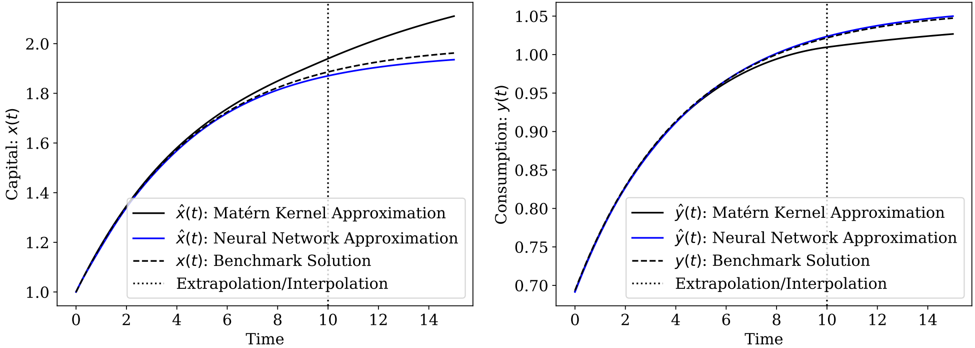


$$\mathcal{D} = \{0, 1, \cdots, 30\}$$

- The explosive solutions are ruled out without directly imposing the boundary condition.
- Provides a very accurate approximation, both in the short- and medium-run.
- Learns the **right steady-state**.

Relative errors

Short time planning: "In the long run we are all dead"



$$\mathcal{D} = \{0, 1, \cdots, 10\}$$

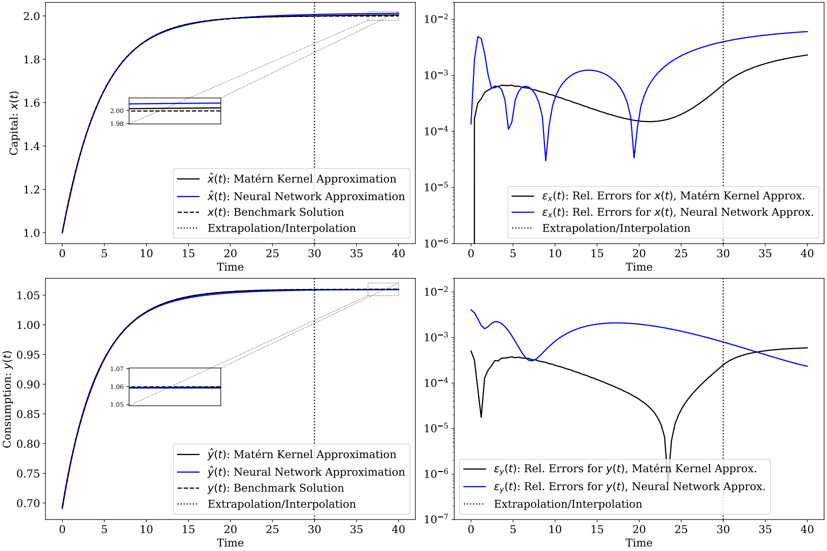
- The explosive solutions are ruled out without directly imposing the boundary condition.
- Provides a very accurate approximation in the short-run.

Neoclassical growth model: concave-convex production function

- In all examples so far we have had one **saddle-path** converging to a unique **saddle** fixed point.

Appendix

Neoclassical growth: relative errors



Back

contd

content...