Solving Models of Economic Dynamics with Ridgeless Kernel Regressions

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

ASSA 2026

[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

- 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:
  - These forward-looking boundary conditions are a key limitation on increasing dimensionality.

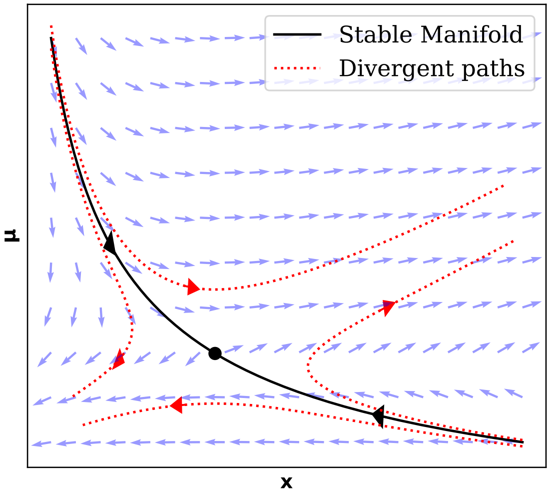
Contribution

Using kernel method to solve a broad class of infinite-horizon, deterministic, continuous-time model

1. **Minimum-norm alignment**:
   - The minimum-norm kernel method aligns with asymptotic boundary conditions.
2. **Learning the right set of steady-states**:
   - Kernel machines learn the boundary values, thereby extrapolating outside the training data.
3. **Robustness and speed**:
   - Competitive in speed and more stable than traditional methods.
4. **Consistency of ML estimates**.

Intuition

- **Violation of the boundary conditions**:
  - Sub-optimal solutions explode over time.
  - They have large derivatives.
  - This behavior is due to the **saddle-path** nature of the problem.
- **Minimum-norm solution** :
  - Penalizing large derivatives rules out explosive paths.
  - The remaining solution is the optimal solution.



The Problem

The class of problems

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

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

$$\dot{\boldsymbol{\mu}}(t) = r\boldsymbol{\mu}(t) - \boldsymbol{\mu}(t) \odot \mathbf{G}(\mathbf{x}(t), \boldsymbol{\mu}(t), \mathbf{y}(t)) \tag{2}$$

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

boundary conditions (at infinity)

$$\mathbf{0} = \lim_{t \to \infty} e^{-rt}\mathbf{x}(t) \odot \boldsymbol{\mu}(t), \tag{4}$$

initial value $\mathbf{x}(0) = \mathbf{x}_0$.

- $\mathbf{x} \in \mathbb{R}^M$: state variables.
- $\boldsymbol{\mu} \in \mathbb{R}^M$: co-state variables.
- $\mathbf{y} \in \mathbb{R}^P$: jump variables.

Challenges

Goal: Find an approximation to $\mathbf{x}(t)$, $\mu(t)$, and $\mathbf{y}(t)$.

What is the problem?

- Initial conditions $\mathbf{y}_0$ and μ_0 are unknown.
- The optimal solution follows a **saddle path**.
  - If T is small, solutions are inaccurate due to premature enforcement of the steady state.
  - If T is large, the algorithms become increasingly numerically unstable.

Example: Neoclassical Growth Model

$$\dot{x}(t) = f(x(t)) - \delta x(t) - y(t) := F(x(t), \mu(t), y(t))$$

$$\dot{\mu}(t) = r\mu(t) - \mu(t) \underbrace{\left(f'(x(t)) - \delta\right)}_{:=G(x(t),\mu(t),y(t))}$$

$$0 = \mu(t)y(t) - 1 := H(x(t), \mu(t), y(t))$$

$$x(0) = x_0, \qquad \lim_{t\to\infty} e^{-rt}\mu(t)x(t) = 0$$

capital $x(t)$, consumption $y(t)$, utility $\log(y)$, present-value co-state variable $\mu(t)$, discount rate $r > 0$, depreciation rate $0 < \delta < 1$, and production function $f(x)$.

Method

Method: approximation

- Pick a set of points $\mathcal{D} \equiv \{t_1, \cdots, t_N\}$ for some fixed interval $[0, T]$

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

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

- $\boldsymbol{\alpha}_j^x$, $\boldsymbol{\alpha}_j^\mu$, $\boldsymbol{\alpha}_j^y$, $\boldsymbol{\mu}_0$, and $\mathbf{y}_0$ are parameters to be found.
- $k(t, t_j)$ is a kernel that measures "similarity" between t and t_j.
- We use a Matérn kernel with smoothness ν and length ℓ.

  ▸▸ Matérn kernel

Method: Ridgeless kernel regression

We solve

$$\begin{aligned}
&\min_{\hat{\mathbf{x}},\hat{\boldsymbol{\mu}},\hat{\mathbf{y}}} \left(\sum_{m=1}^{M} \|\hat{\dot{\mathbf{x}}}^{(m)}\|_{\mathcal{H}}^2 + \sum_{m=1}^{M} \|\hat{\dot{\boldsymbol{\mu}}}^{(m)}\|_{\mathcal{H}}^2 \right) \\
&\text{s.t. } \hat{\dot{\mathbf{x}}}(t_i) = \mathbf{F}(\hat{\mathbf{x}}(t_i), \hat{\boldsymbol{\mu}}(t_i), \hat{\mathbf{y}}(t_i)), \quad \text{for all } t_i \in \mathcal{D} \\
&\qquad \hat{\dot{\boldsymbol{\mu}}}(t_i) = r\hat{\boldsymbol{\mu}}(t_i) - \hat{\boldsymbol{\mu}}(t_i) \odot \mathbf{G}(\hat{\mathbf{x}}(t_i), \hat{\boldsymbol{\mu}}(t_i), \hat{\mathbf{y}}(t_i)), \quad \text{for all } t_i \in \mathcal{D} \\
&\qquad \mathbf{0} = \mathbf{H}(\hat{\mathbf{x}}(t_i), \hat{\boldsymbol{\mu}}(t_i), \hat{\mathbf{y}}(t_i)), \quad \text{for all } t_i \in \mathcal{D}.
\end{aligned}$$

- $\|\hat{\dot{\mathbf{x}}}^{(m)}\|_{\mathcal{H}}^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i^{x^{(m)}} \alpha_j^{x^{(m)}} k(t_i, t_j)$ and $\|\hat{\dot{\boldsymbol{\mu}}}^{(m)}\|_{\mathcal{H}}^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i^{\mu^{(m)}} \alpha_j^{\mu^{(m)}} k(t_i, t_j)$
- TVC is **not imposed**.
- The penalization is used to control the smoothness of the approximating functions.
- For Matérn kernels, it also controls the smoothness of derivatives.

Applications

Neoclassical growth model

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

for a given x_0.

- $x(t) \in \mathbb{R}$: capital, $y(t) \in \mathbb{R}$: consumption, and a concave production function $f(x) = x^a$.

Constructing the Hamiltonian ...

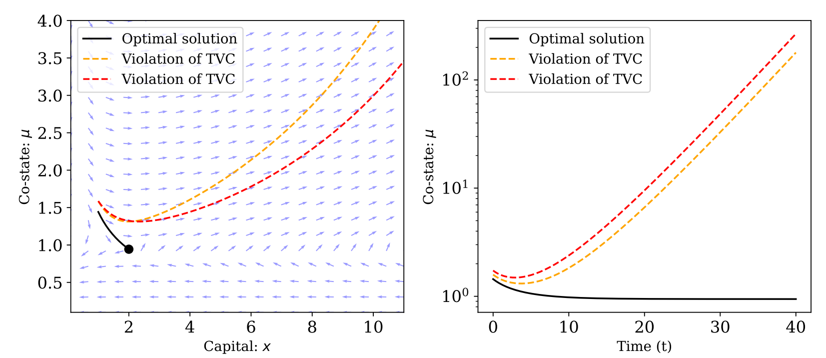
$$\begin{aligned}
\dot{x}(t) &= f(x(t)) - \delta x(t) - y(t) \\
\dot{\mu}(t) &= r\mu(t) - \mu(t) \\
0 &= \mu(t)y(t) - 1 \\
x(0) &= x_0, \qquad \lim_{t\to\infty} e^{-rt}\mu(t)x(t) = 0
\end{aligned}$$

- Last Equation : transversality condition (TVC)

Why do we need the boundary condition?

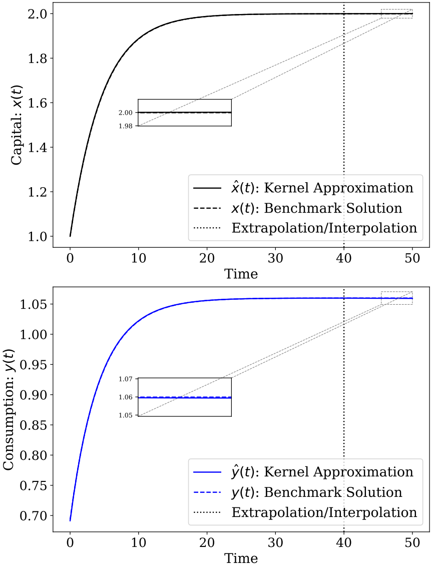
Ignoring the transversality condition:

$$\begin{aligned} \dot{x}(t) &= f(x(t)) - \delta x(t) - y(t) \\ \dot{\mu}(t) &= r\mu(t) - \mu(t) \\ 0 &= \mu(t)y(t) - 1 \\ x(0) &= x_0. \end{aligned}$$



Neoclassical growth model: results

- $\mathcal{D} = \{0, 1, \cdots, 40\}$.
- $f(x) = x^{\frac{1}{3}}$, $\delta = \frac{1}{3}$, and $r = 0.11$.
- The explosive solutions are ruled out without directly imposing the boundary condition.
- Very accurate approximations, both in the short- and medium-run.
- Learns the **right steady-state**. Relative errors



Neoclassical growth model: learning the steady state

content...

Extensions

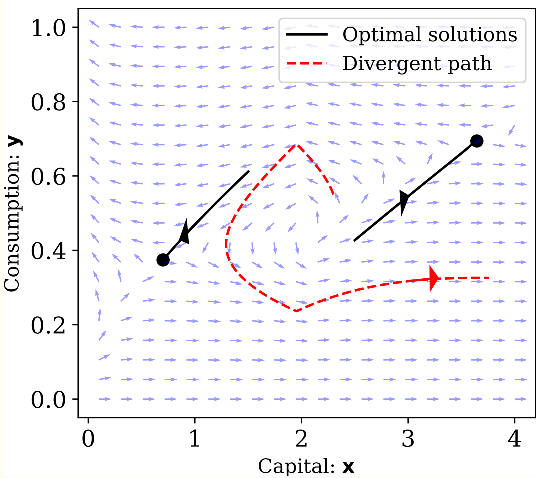
Neoclassical Growth Model: Non-Concave Production Function

- So far we have had a **unique** saddle-path converging to a unique **saddle** steady state.
- What if we have **two** saddle steady states, very close to each other (equilibrium multiplicity)?
- Neoclassical growth model with a non-concave production function (threshold externalities):

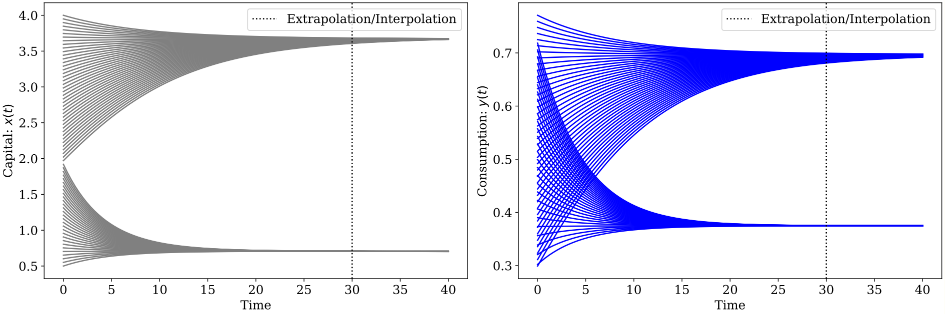
$$f(x) = A \max\{x^a, b_1 x^a - b_2\}$$

Non-concave production function: vector field

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) - \mathbf{y}(t) - \delta\mathbf{x}(t)$$
$$\dot{\mathbf{y}}(t) = \mathbf{y}(t)\left[f'(\mathbf{x}(t)) - \delta - r\right]$$
$$\mathbf{x}(0) = \mathbf{x}_0 \text{ given.}$$



Results



- The approximate solutions approach the right steady states.
- The transversality conditions are satisfied without being directly imposed.
- The steady states are learned. Full DAE

Linear asset pricing

Linear asset pricing model

$$\begin{aligned}
\dot{x}(t) &= c + gx(t) \\
\dot{\mu}(t) &= r\mu(t) - x(t) := r\mu(t) - \mu(t)\frac{x(t)}{\mu(t)} \\
0 &= \lim_{t\to\infty} e^{-rt}\mu(t)x(t).
\end{aligned}$$

- $x(t) \in \mathbb{R}$: flow payoffs from a claim to an asset.
- $\mu(t) \in \mathbb{R}$ be the price of a claim to that asset.
- x_0 given.

Why do we need the boundary condition?

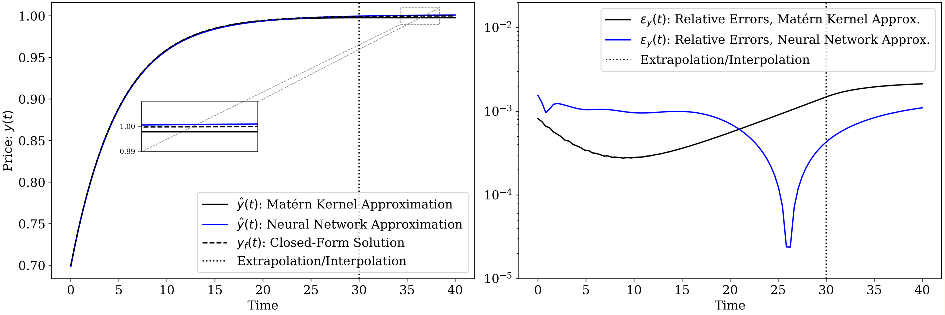
$$\dot{\mathbf{x}}(t) = c + g\mathbf{x}(t)$$
$$\dot{\mathbf{y}}(t) = r\mathbf{y}(t) - \mathbf{x}(t)$$

- 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**.

Results



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

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

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

Matérn kernel

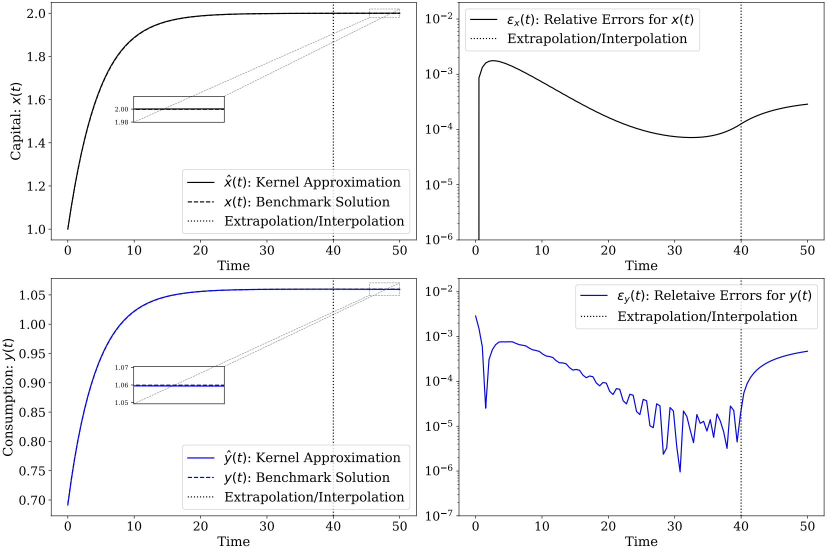
$$K(t, t_j) = C_{\frac{1}{2}}(t, t_j) = \sigma^2 \exp\left(-\frac{|t - t_j|}{\ell}\right),$$

$$K(t, t_j) = C_{\frac{3}{2}}(t, t_j) = \sigma^2 \left(1 + \frac{\sqrt{3}|t - t_j|}{\ell}\right) \exp\left(-\frac{\sqrt{3}|t - t_j|}{\ell}\right),$$

$$K(t, t_j) = C_{\frac{5}{2}}(t, t_j) = \sigma^2 \left(1 + \frac{\sqrt{5}|t - t_j|}{\ell} + \frac{5|t - t_j|^2}{3\ell^2}\right) \exp\left(-\frac{\sqrt{5}|t - t_j|}{\ell}\right).$$

Back

Neoclassical growth: relative errors



Back

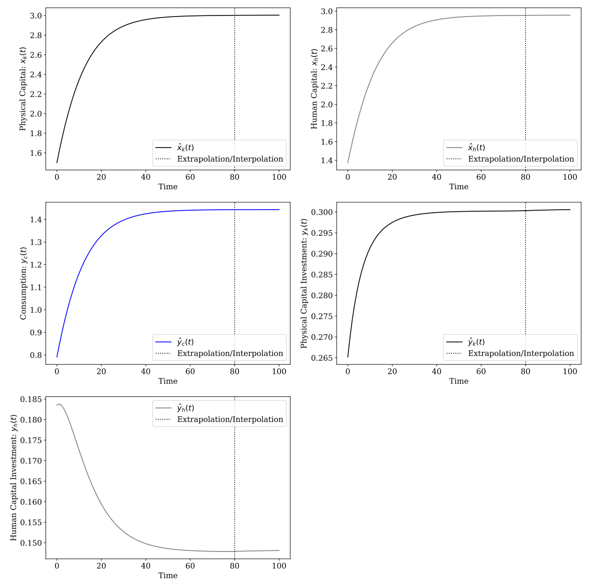
Human capital and growth

$$
\begin{aligned}
\dot{\mathbf{x}}_k(t) &= \mathbf{y}_k(t) - \delta_k \mathbf{x}_k(t), \\
\dot{\mathbf{x}}_h(t) &= \mathbf{y}_h(t) - \delta_k \mathbf{x}_h(t) \\
\dot{\mathbf{y}}_c(t) &= \mathbf{y}_c(t) \left[f_1 \left(\mathbf{x}_k(t), \mathbf{x}_h(t) \right) - \delta_k - r \right], \\
0 &= f \left(\mathbf{x}_k(t), \mathbf{x}_h(t) \right) - \mathbf{y}_c(t) - \mathbf{y}_k(t) - \mathbf{y}_h(t), \\
0 &= f_2 \left(\mathbf{x}_k(t), \mathbf{x}_h(t) \right) - f_1 \left(\mathbf{x}_k(t), \mathbf{x}_h(t) \right) + \delta_k - \delta_h. \\
0 &= \lim_{t \to \infty} e^{-rt} \frac{\mathbf{x}_k(t)}{\mathbf{y}_c(t)}, \; 0 = \lim_{t \to \infty} e^{-rt} \frac{\mathbf{x}_h(t)}{\mathbf{y}_c(t)}.
\end{aligned}
$$

- $\mathbf{x}_k$: physical capital, $\mathbf{x}_h$: human capital, $\mathbf{y}_c$: consumption, $\mathbf{y}_k$: investment in physical capital, $\mathbf{y}_h$: investment in human capital
- $f\left(x_k, x_h\right) = x_k^{a_k} x_h^{a_h}$

Back

Results


Physical Capital: $x_k(t)$
$\hat{x}_k(t)$
Extrapolation/Interpolation
Time
Human Capital: $x_h(t)$
$\hat{x}_h(t)$
Extrapolation/Interpolation
Time
Consumption: $y_c(t)$
$\hat{y}_c(t)$
Extrapolation/Interpolation
Time
Physical Capital Investment: $y_k(t)$
$\hat{y}_k(t)$
Extrapolation/Interpolation
Time
Human Capital Investment: $y_h(t)$
$\hat{y}_h(t)$
Extrapolation/Interpolation
Time