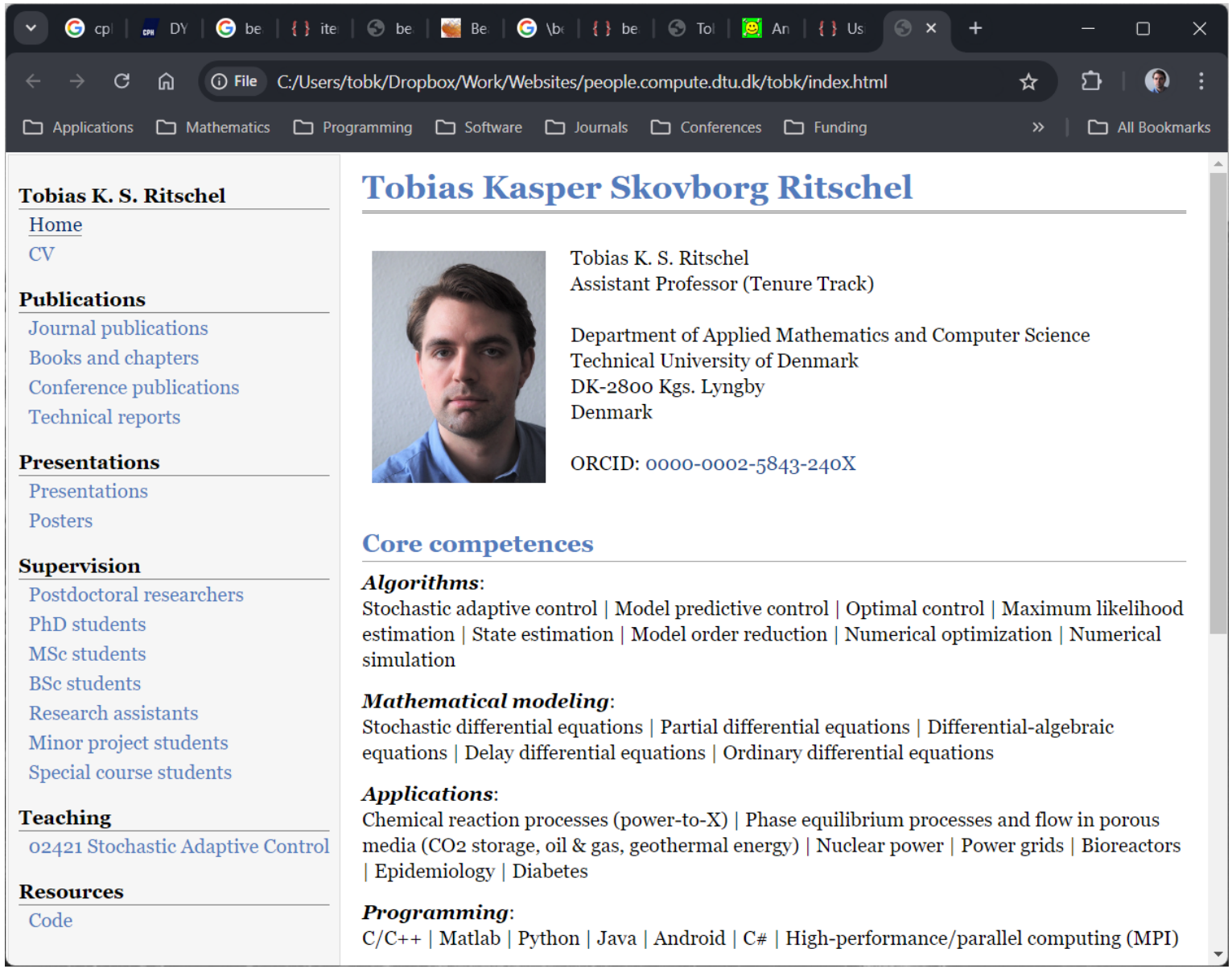
Computational methods for distributed delay differential equations based on kernel approximations

Tobias K. S. Ritschel

Department of Applied Mathematics and Computer Science
Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

August 29th, 2024

Briefly about myself . . .



Introduction

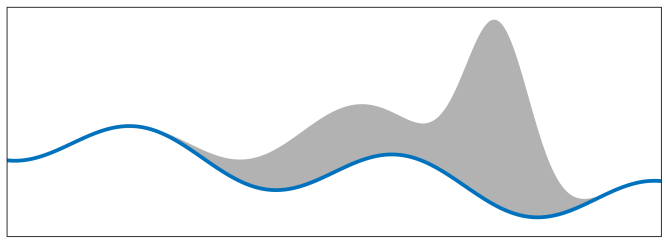
- Delay differential equation with distributed delay

$$\dot{x}(t) = f(x(t), z(t), u(t), d(t), p), \tag{1a}$$

$$z(t) = \int_{-\infty}^{t} \alpha(t-s) r(s)\, \mathrm{d}s, \tag{1b}$$

$$r(t) = h(x(t), p) \tag{1c}$$

- Sketch



- Measurement equation

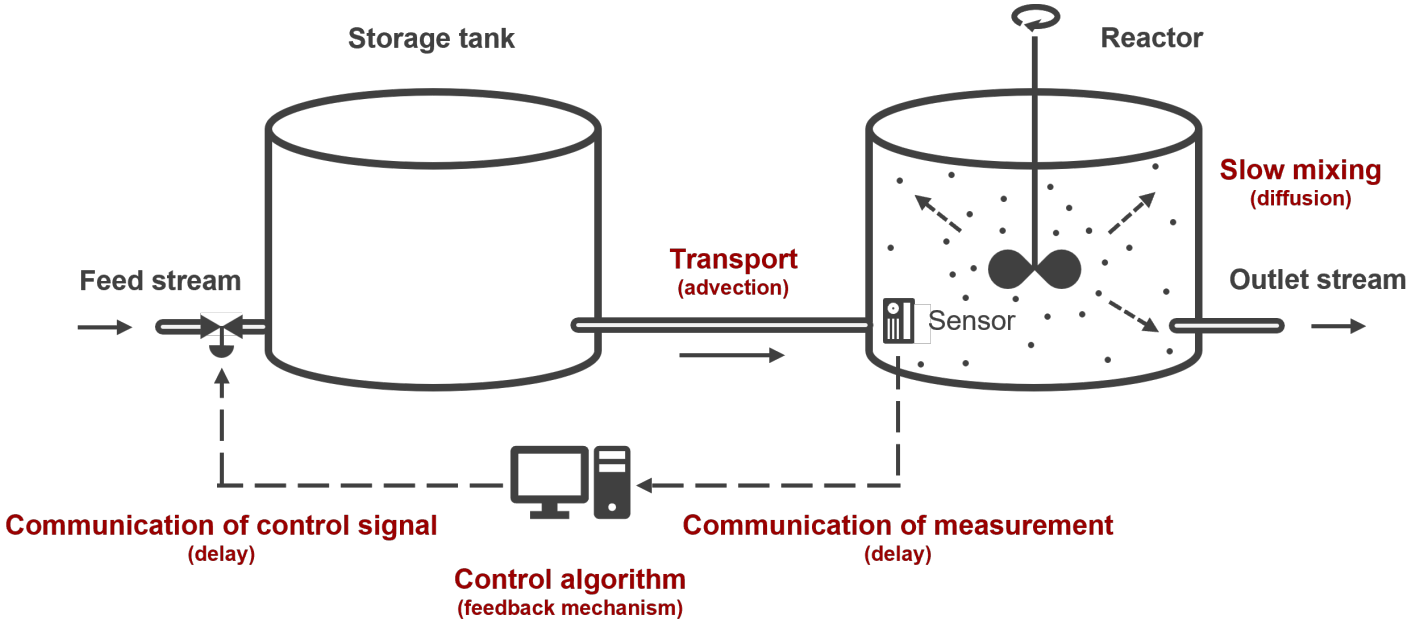
$$y(t_k) = g(x(t_k), p) \tag{2}$$

- Feedback control

$$u(t) = \mu(y(t_k)), \qquad t \in [t_k, t_{k+1}[\tag{3}$$

- Ecology
- Medicine (PK/PD)
- Physiology
- Economics
- Mechanics

Large, complex, and interconnected systems



Time-Varying Distributed Delays and Their Use in Aggregative Models of Large Systems

THOMAS J. MANETSCH, MEMBER, IEEE

***Abstract*—The distributed delay has found widespread application in the modeling of aggregative processes involving the flow of entities with random transit times through a given process. In most of these applications it is assumed that the mean entity transit time is fixed. There are, however, a number of important applications in aggregative economic,** the context of a larger system model. We conclude the paper with discussions of applications of the time-varying distributed delay to an aggregative biological process and an aggregative economic process.

Reference: [1]

Key questions and outline

1. Given a kernel, α, what is the state, x? (Simulation)
2. Given a set of measurements, $\{y(t_k)\}_{k=0}^{N}$, what is the kernel, α? (Identification)
3. Given an estimated kernel, $\hat{\alpha}$, what are the underlying dynamics? (Interpretation)

Main part

- Kernel approximation • ($\hat{\alpha} \approx \alpha$ or $\hat{z} \approx z$)
- Simulation/analysis based on approximate kernel •
- Kernel identification •
- Kernel for advective flow • (interpretation)

Current work and potential future topics

- Linearization •
- How much does a PID controller remember? •
- How much *should* a PID controller remember •

Kernel approximation

Mixed Erlang kernel approximation

- Approximate the kernel by the probability density function (pdf) of a mixed Erlang distribution

$$\hat{\alpha}^{(M)}(t) = \sum_{m=0}^{M} c_m \hat{\alpha}_m(t) \tag{4}$$

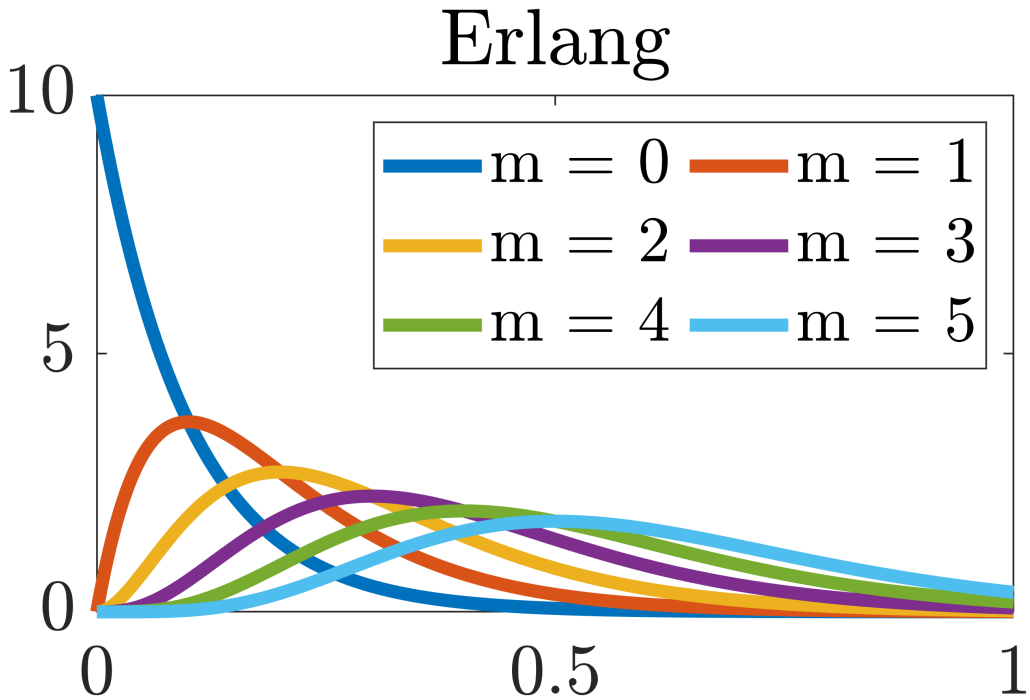
- pdf and cumulative distribution function (cdf) of $m+1$'th order Erlang distribution

$$\hat{\alpha}_m(t) = b_m t^m e^{-at}, \quad b_m = \frac{a^{m+1}}{m!}, \quad \hat{\beta}_m(t) = 1 - \frac{1}{a}\sum_{n=0}^{m} \hat{\alpha}_n(t) \tag{5}$$

- Conditions

$$c_m \in [0,1], \qquad a \in (0,\infty), \qquad \sum_{m=0}^{M} c_m = 1 \tag{6}$$

pdf of Erlang distributions



Linear chain trick •

- Substitute kernel approximation

$$\hat{z}(t) = \int_{-\infty}^{t} \hat{\alpha}^{(M)}(t-s)\hat{r}(s)\,\mathrm{d}s = \sum_{m=0}^{M} c_m \hat{z}_m(t), \tag{7a}$$

$$\hat{z}_m(t) = \int_{-\infty}^{t} \hat{\alpha}_m(t-s)\hat{r}(s)\,\mathrm{d}s \tag{7b}$$

- Differentiate $\hat{z}_m$

$$\begin{aligned}\dot{\hat{z}}_m(t) &= \hat{\alpha}_m(0)\hat{r}(t) + \int_{-\infty}^{t} \dot{\hat{\alpha}}_m(t-s)\hat{r}(s)\,\mathrm{d}s \\ &= \begin{cases} a(\hat{r}(t) - \hat{z}_0(t)), & m = 0, \\ a(\hat{z}_{m-1}(t) - \hat{z}_m), & m \geq 1 \end{cases}\end{aligned} \tag{8}$$

Linear chain trick

- From previous slide

$$\hat{z}(t) = \sum_{m=0}^{M} c_m \hat{z}_m(t), \tag{9a}$$

$$\dot{\hat{z}}_m(t) = \begin{cases} a(\hat{r}(t) - \hat{z}_0(t)), & m = 0, \\ a(\hat{z}_{m-1}(t) - \hat{z}_m), & m \geq 1 \end{cases} \tag{9b}$$

- Approximate system

$$\dot{\hat{x}}(t) = f(\hat{x}(t), \hat{z}(t), u(t), d(t), p), \tag{10a}$$

$$\dot{\hat{Z}}(t) = A\hat{Z}(t) + B\hat{r}(t), \tag{10b}$$

$$\hat{z}(t) = C\hat{Z}(t), \tag{10c}$$

$$\hat{r}(t) = h(\hat{x}(t), p) \tag{10d}$$

- System matrices

$$A = a \begin{bmatrix} -I & & & \\ I & -I & & \\ & \ddots & \ddots & \\ & & I & -I \end{bmatrix}, \qquad B = a \begin{bmatrix} I \\ \\ \\ \end{bmatrix}, \tag{11}$$

$$C = \begin{bmatrix} c_0 I & c_1 I & \cdots & c_M I \end{bmatrix} \tag{12}$$

Is it a good approximation?

According to Tijms [2] (see also [3]), the cdf of any non-negative random variable can be approximated by the cdf of an infinite mixed Erlang distribution.

- Cumulative distribution functions

$$\beta(t) = \int_0^t \alpha(s)\,\mathrm{d}s, \qquad \hat{\beta}^{(M)}(t) = \int_0^t \hat{\alpha}^{(M)}(s)\,\mathrm{d}s = \sum_{m=0}^{M} c_m \hat{\beta}_m(t) \tag{13}$$

- Coefficients and rate parameter

$$c_m = \beta(t_{m+1}) - \beta(t_m), \qquad a = 1/\Delta t, \qquad t_m = m\Delta t \tag{14}$$

- Pointwise convergence

$$\lim_{\Delta t \to 0} \hat{\beta}^{(\infty)}(t) = \beta(t) \tag{15}$$

for any $t \in [0, \infty)$ where β is continuous

What does Tijms' theorem tell us?

Can we approximate α?

- Scheffé [4]: Convergence of $\hat{\alpha}^{(\infty)}$ implies convergence of $\hat{\beta}^{(\infty)}$
- Boos [5]: Convergence of $\hat{\beta}^{(\infty)}$ implies convergence of $\hat{\alpha}^{(\infty)}$ if $\hat{\alpha}^{(\infty)}$ is bounded and equicontinuous (see also [6])
- If $\hat{\alpha}^{(\infty)}$ converges uniformly, it should also converge to α [7]
- I have not been able to prove convergence. Maybe start with simple examples of α and β.

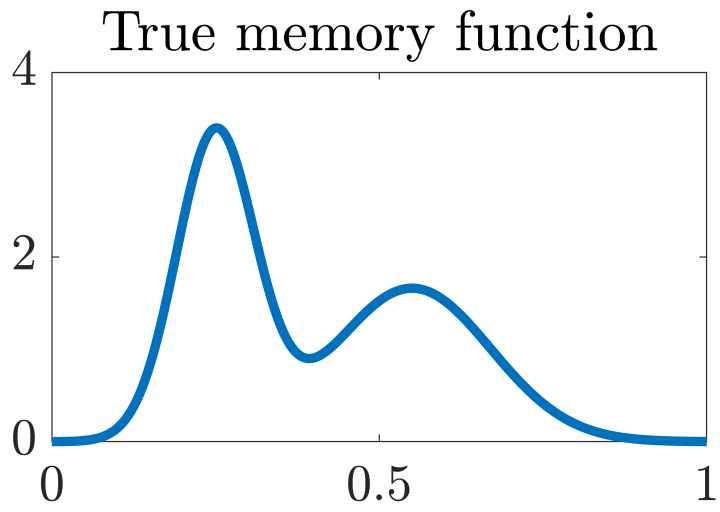
Can we approximate z?

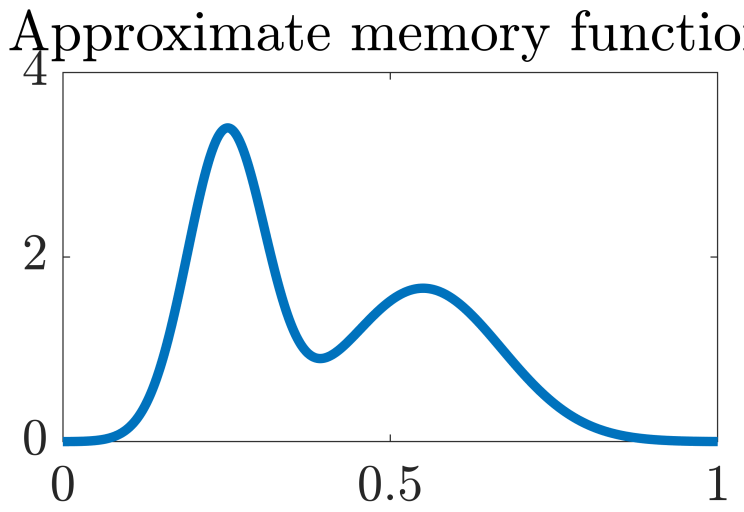
- Does $\hat{\beta}^{(\infty)} \to \beta$ imply that

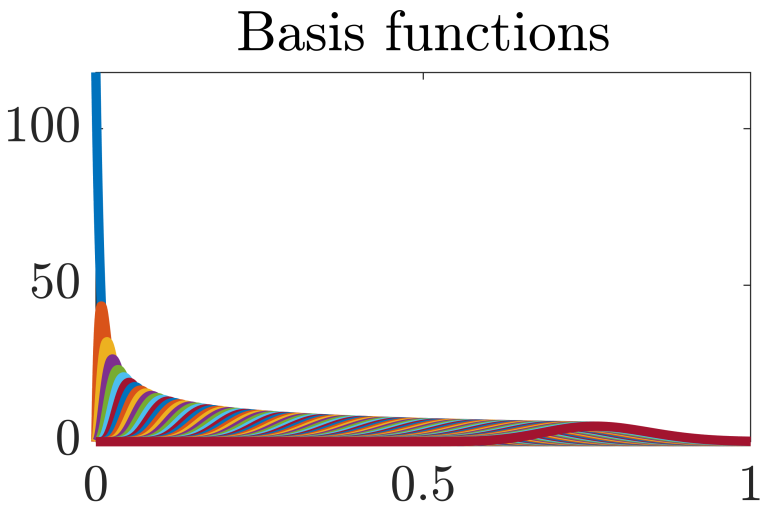
$$\hat{z}(t) = \int_{-\infty}^{t} \hat{\alpha}^{(\infty)}(t-s)r(s)\,\mathrm{d}s \to \int_{-\infty}^{t} \alpha(t-s)r(s)\,\mathrm{d}s = z(t)? \qquad (16)$$

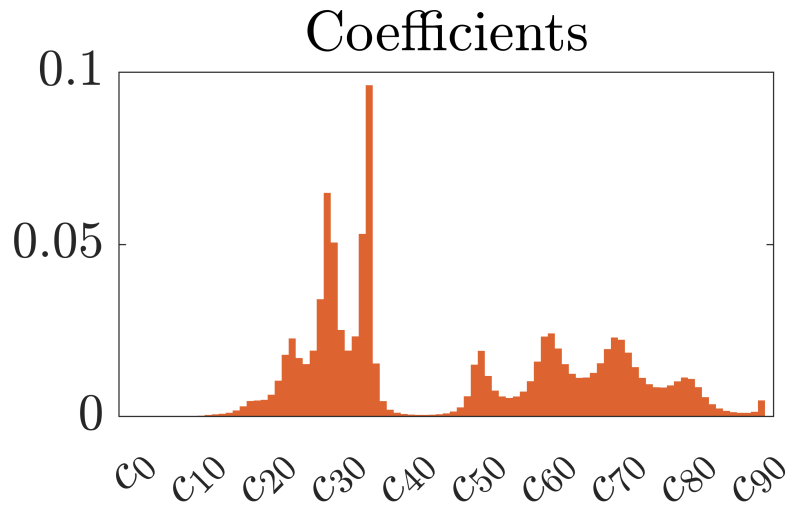
  We have a sketch of a proof (see appendix)

Example

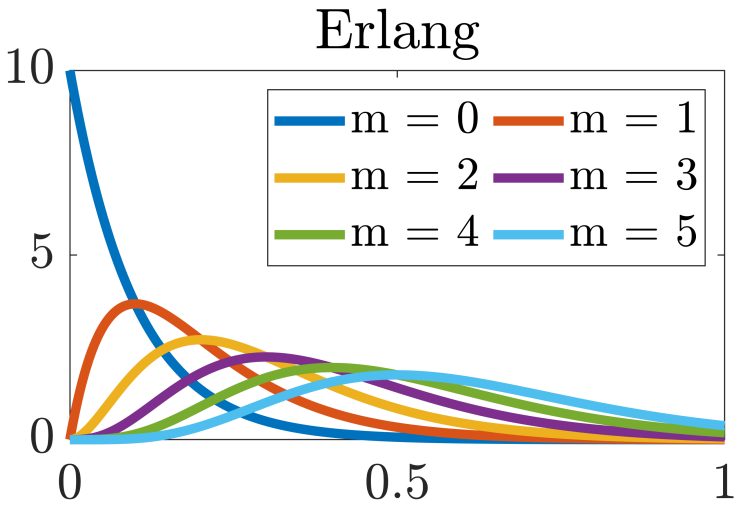

True memory function
4
2
0
0
0.5
1

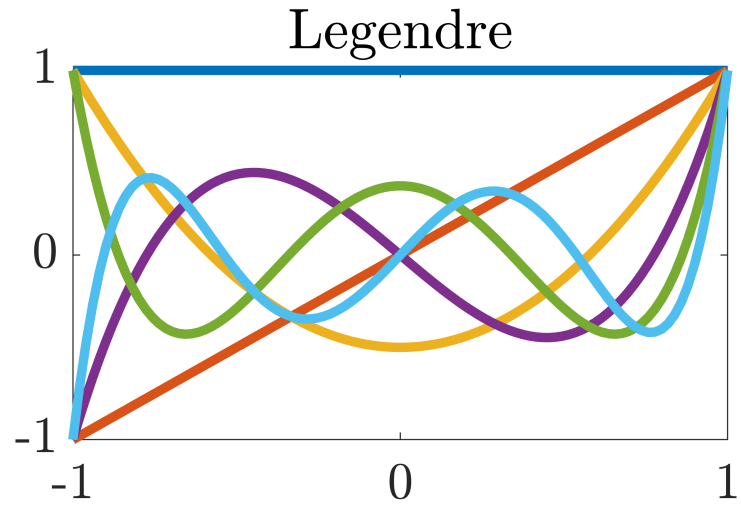

Approximate memory function
4
2
0
0
0.5
1

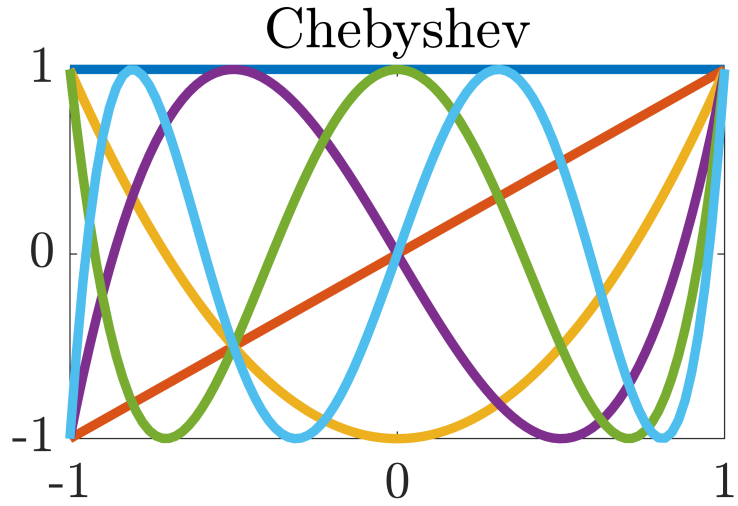

Basis functions
100
50
0
0
0.5
1

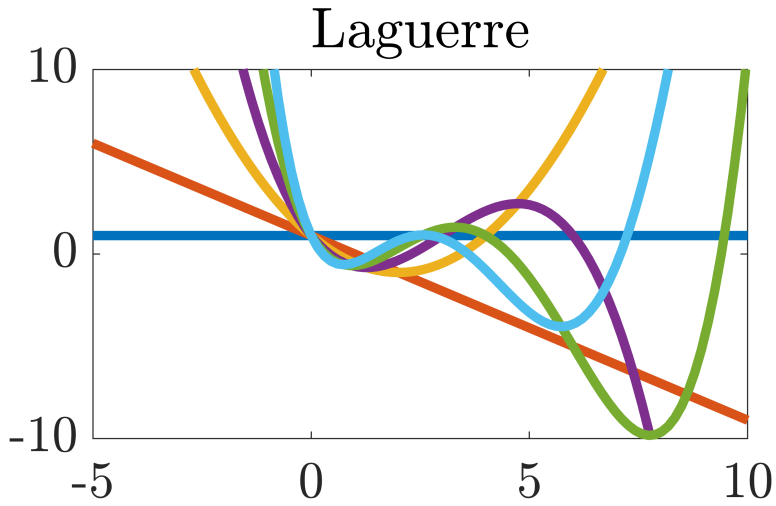

Coefficients
0.1
0.05
0
c0
c10
c20
c30
c40
c50
c60
c70
c80
c90

Alternative approximations









- Hermite polynomials
- Mixed phase-type distributions
- Splines

Simulation

Simulation/analysis – key steps •

- Approximate the kernel, $\hat{\alpha}^{(M)} \approx \alpha$
- Form the approximate system (of ordinary differential equations)
- Use any method for ordinary differential equations

Least-squares approximation

- Objective: For given M, identify the coefficients $\{c_m\}_{m=0}^{M}$ and the rate parameter a that minimize

$$\int_0^\infty (\alpha(t) - \hat{\alpha}^{(M)}(t))^2 \, \mathrm{d}t \tag{17}$$

- Approximations:
  1. Choose t_h such that $1 - \beta(t_h) \leq \epsilon$ (see appendix)
  2. Use a left rectangle rule with N rectangles to approximate the integral
- Least-squares optimization problem (solved using Matlab's fmincon)

$$\min_{\{c_m\}_{m=0}^{M}, a} \quad \sum_{k=0}^{N-1} \left(\alpha(t_k) - \hat{\alpha}^{(M)}(t_k)\right)^2 \Delta t \tag{18a}$$

subject to

$$\sum_{m=0}^{M} c_m = 1, \tag{18b}$$

$$0 \leq c_m \leq 1, \tag{18c}$$

$$a_{\min} \leq a \tag{18d}$$

Test of approximate simulation •

- Test system

$$\dot{N}(t) = \kappa N(t)\left(1 - \frac{\tilde{N}(t)}{K}\right) + Q(t), \qquad \alpha(t) = \frac{2}{\sqrt{\pi}} e^{-t^2}, \tag{19a}$$

$$\tilde{N}(t) = \int_{-\infty}^{t} \alpha(t-s) N(s)\,\mathrm{d}s, \qquad \beta(t) = \operatorname{erf}(t) \tag{19b}$$

- Solution (method of manufactured solutions)

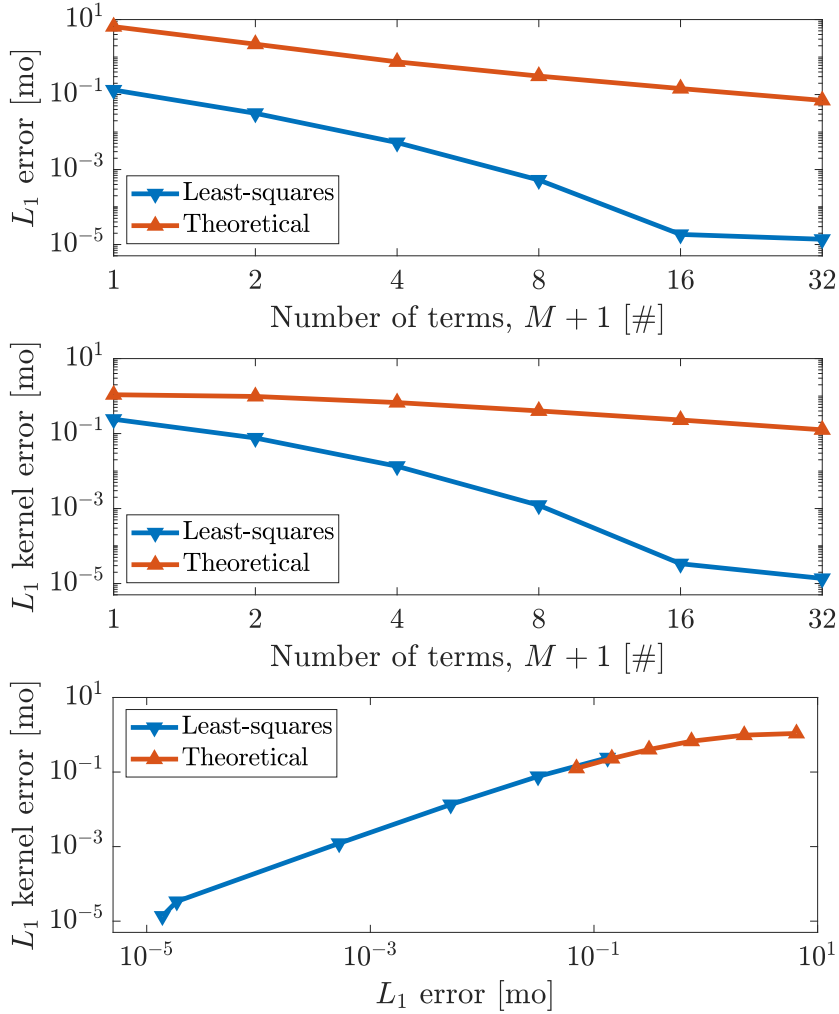
$$N^*(t) = 1 + e^{-\left(\frac{t}{\gamma}\right)^2}, \tag{20a}$$

$$\tilde{N}^*(t) = 1 + \frac{\gamma}{\sqrt{\gamma^2+1}} e^{-\frac{t^2}{\gamma^2+1}} \left(1 + \operatorname{erf}\left(\frac{t}{\gamma\sqrt{\gamma^2+1}}\right)\right) \tag{20b}$$

- Choose forcing term accordingly

$$Q(t) = \dot{N}^*(t) - \kappa N^*(t)\left(1 - \frac{\tilde{N}^*(t)}{K}\right) \tag{21}$$

Test results



Approximate L_1 error of the states (top) and kernel (middle) obtained with a mixed Erlang kernel approximation for different values of M. The bottom plot shows the results in the top and middle plot together.

Numerical example •

- System

$$\dot{N}(t) = \kappa N(t)\left(1 - \frac{\tilde{N}(t)}{K}\right), \tag{22a}$$

$$\tilde{N}(t) = \int_{-\infty}^{t} \alpha(t-s)N(s)\,\mathrm{d}s \tag{22b}$$

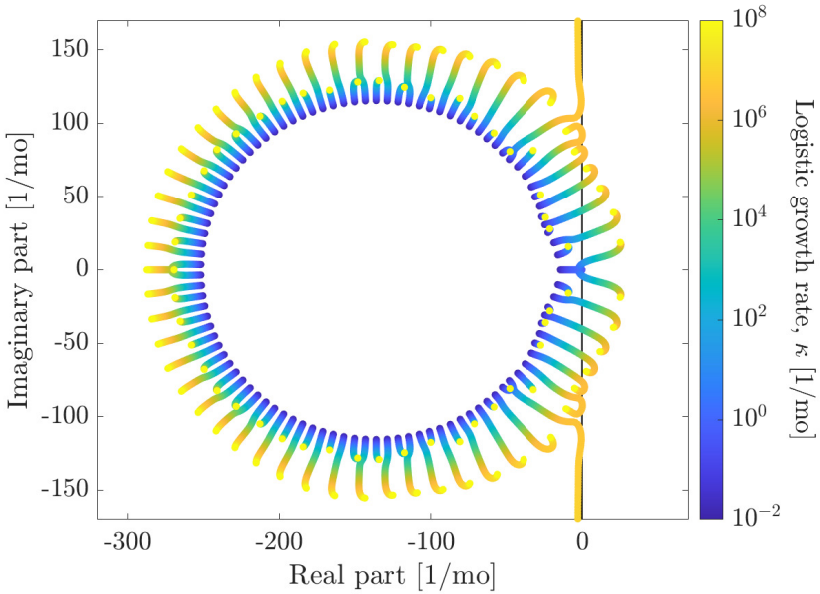
- Kernel

$$\alpha(t) = \gamma_1 F(t;\mu_1,\sigma_1) + \gamma_2 F(t;\mu_2,\sigma_2), \tag{23a}$$

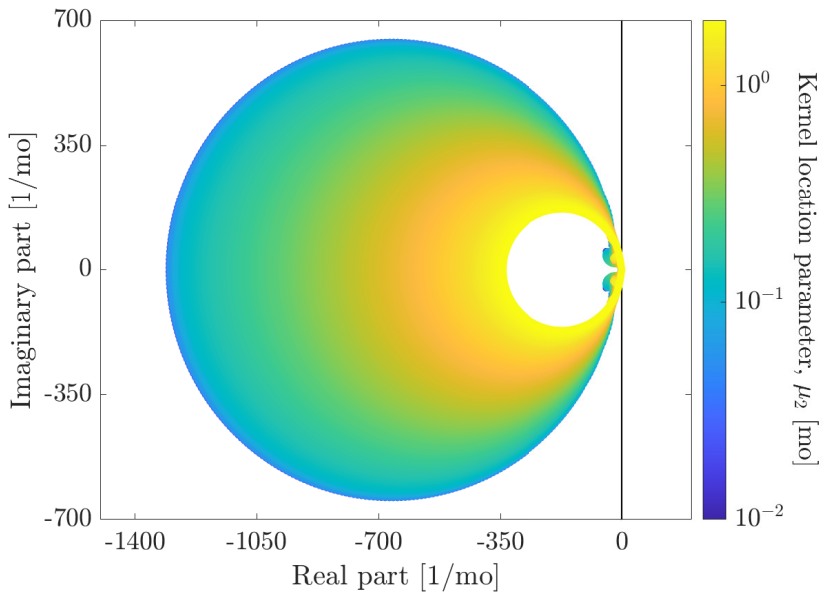
$$F(t;\mu,\sigma) = \frac{\exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) + \exp\left(-\frac{1}{2}\left(\frac{t+\mu}{\sigma}\right)^2\right)}{\sqrt{2\pi}\sigma} \tag{23b}$$

Numerical example: Bifurcation analysis

Model parameter

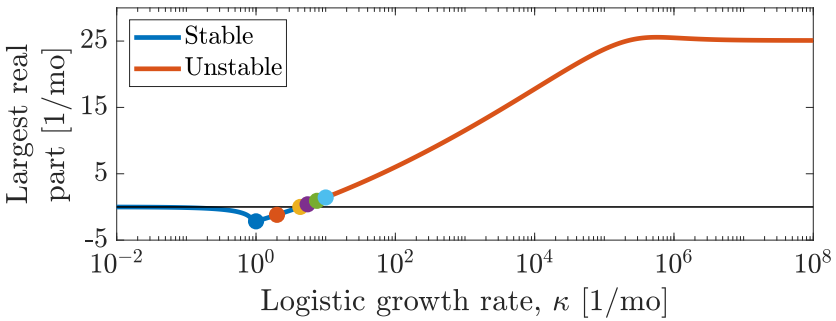


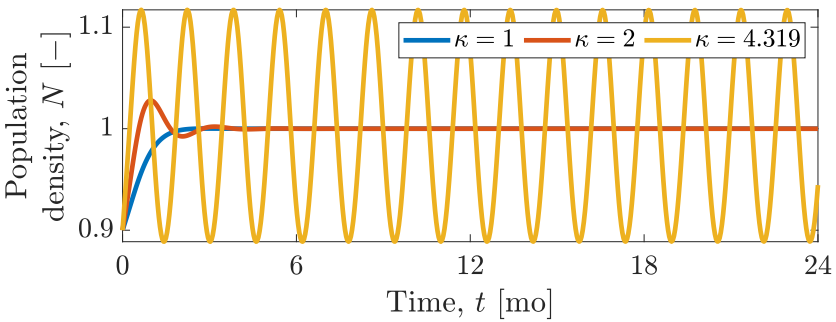
Kernel parameter

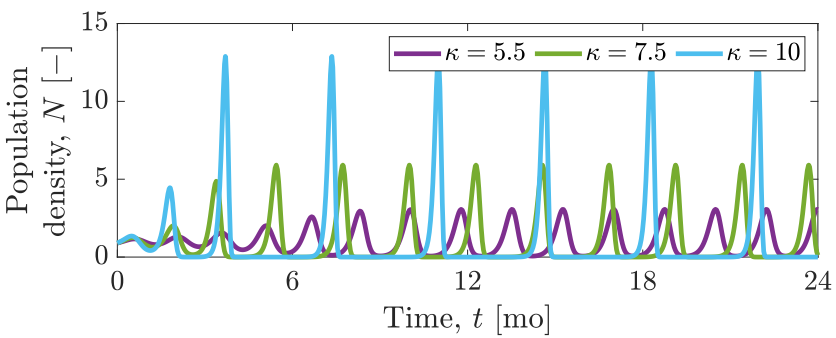


Numerical example: Bifurcation analysis

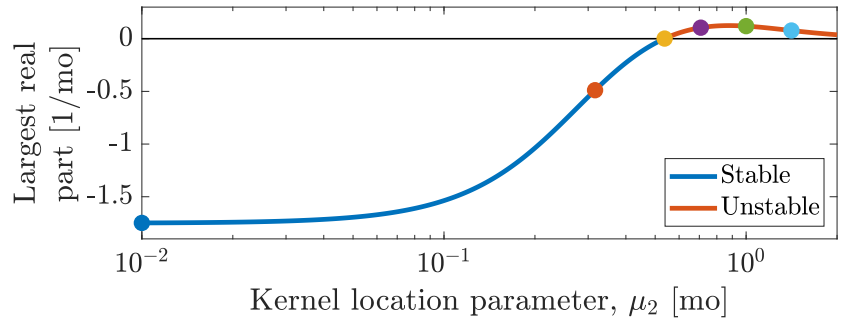
Model parameter

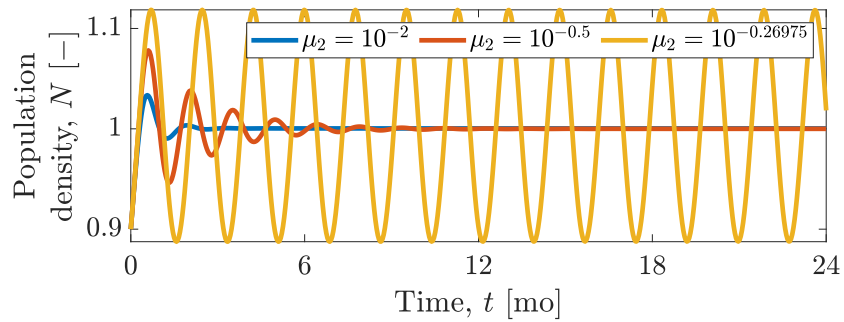


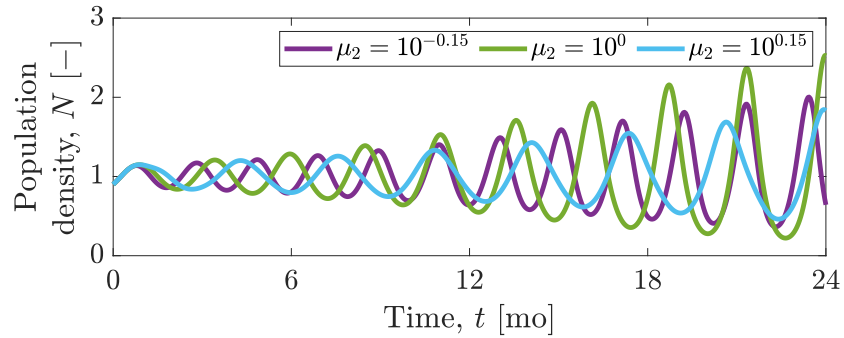




Kernel parameter

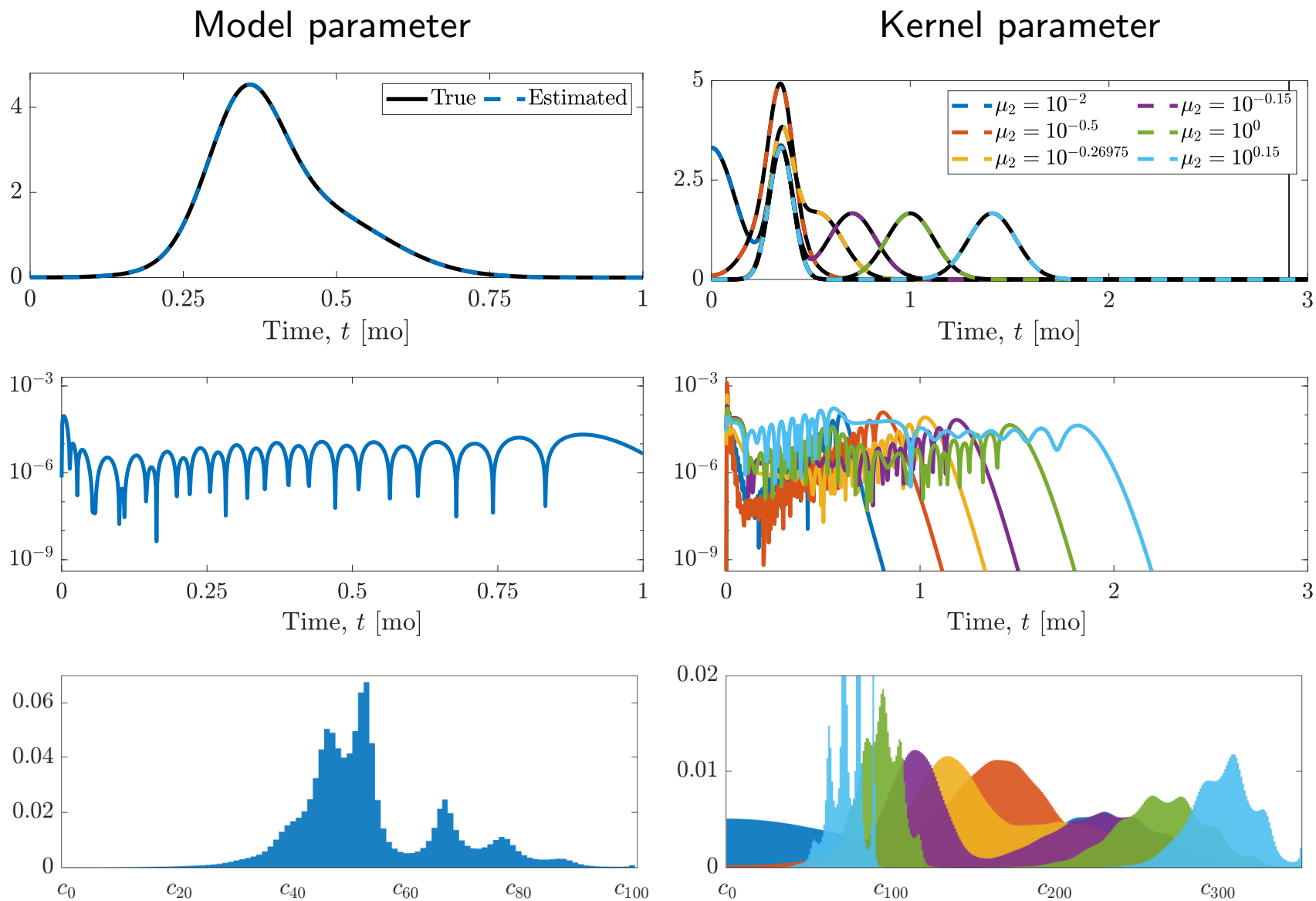






The numerical simulations are obtained with Euler's implicit method and a right rectangle rule for approximating the integral.

Numerical example: Bifurcation analysis



Identification

Identification (collaboration w. John Wyller)

- Key objective: Given a set of N measurements, $\{y(t_k)\}_{k=0}^{N}$, estimate the kernel, α, model parameters, p, and the initial state, x_0
- Choose M and approximate kernel, $\hat{\alpha}^{(M)} \approx \alpha$
- Assume that the system is initially in steady state
  (makes it easier to evaluate the initial memory states)
- Solve least-squares (or maximum likelihood) estimation problem

Dynamical least-squares problem

$\hat{q}$ are the coefficients, $\{c_m\}_{m=0}^{M}$, and the rate parameter, a, in the approximate kernel, $\hat{\alpha}^{(M)}$

$$\min_{\hat{p},\hat{q},\hat{x}_0} \quad \frac{1}{2}\sum_{k=0}^{N} \|y(t_k) - \hat{y}(t_k)\|_2^2 \tag{24a}$$

subject to

$$\hat{x}(t_0) = \hat{x}_0, \tag{24b}$$

$$\hat{Z}(t_0) = \hat{Z}_0(\hat{x}_0, \hat{p}), \tag{24c}$$

$$\dot{\hat{x}}(t) = f(\hat{x}(t), \hat{z}(t), u(t), d(t), p), \qquad t \in [t_0, t_f], \tag{24d}$$

$$\dot{\hat{Z}}(t) = A(\hat{q})\hat{Z}(t) + B(\hat{q})\hat{r}(t), \qquad t \in [t_0, t_f], \tag{24e}$$

$$\hat{z}(t) = C(\hat{q})\hat{Z}(t), \qquad t \in [t_0, t_f], \tag{24f}$$

$$\hat{r}(t) = h(\hat{x}(t), \hat{p}), \qquad t \in [t_0, t_f], \tag{24g}$$

$$\hat{y}(t_k) = g(\hat{x}(t_k), \hat{p}), \qquad k = 0, \ldots, N, \tag{24h}$$

$$w^T\hat{q} = 1, \tag{24i}$$

$$p_{\min} \leq \hat{p} \leq p_{\max}, \tag{24j}$$

$$q_{\min} \leq \hat{q} \leq q_{\max}, \tag{24k}$$

$$x_{\min} \leq \hat{p} \leq x_{\max} \tag{24l}$$

Example •

- System

$$\dot{N}(t) = \kappa N(t)\left(1 - \frac{\tilde{N}(t)}{K(t)}\right), \tag{25a}$$

$$\tilde{N}(t) = \int_{-\infty}^{t} \alpha(t-s)N(s)\,\mathrm{d}s, \tag{25b}$$

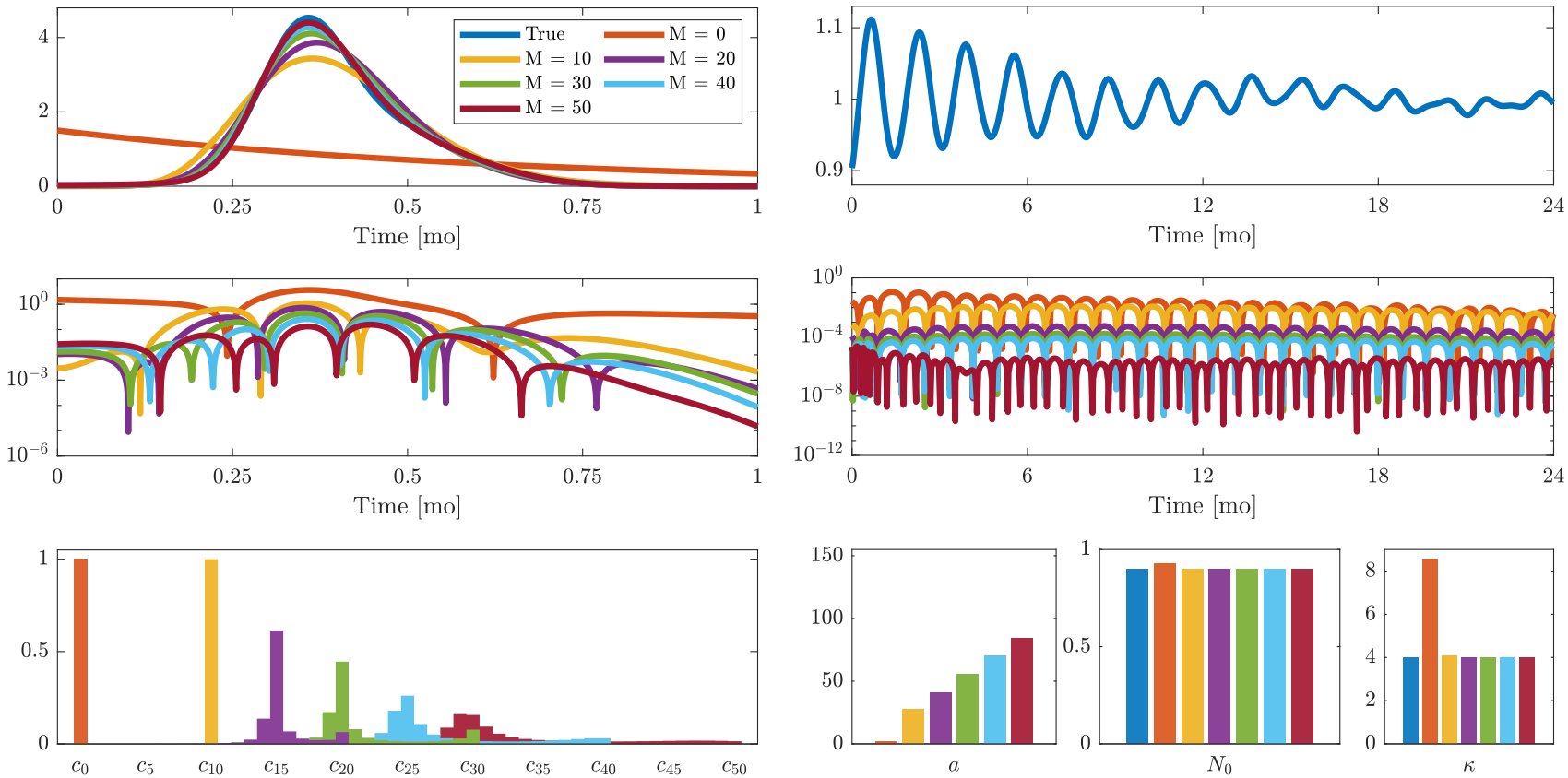
$$K(t) = (1 + A_1 \sin(2\pi\omega_1 t) + A_2 \sin(2\pi\omega_2 t))\bar{K} \tag{25c}$$

- Kernel

$$\alpha(t) = \gamma_1 F(t;\mu_1,\sigma_1) + \gamma_2 F(t;\mu_2,\sigma_2), \tag{26a}$$

$$F(t;\mu,\sigma) = \frac{\exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) + \exp\left(-\frac{1}{2}\left(\frac{t+\mu}{\sigma}\right)^2\right)}{\sqrt{2\pi}\sigma} \tag{26b}$$

Example

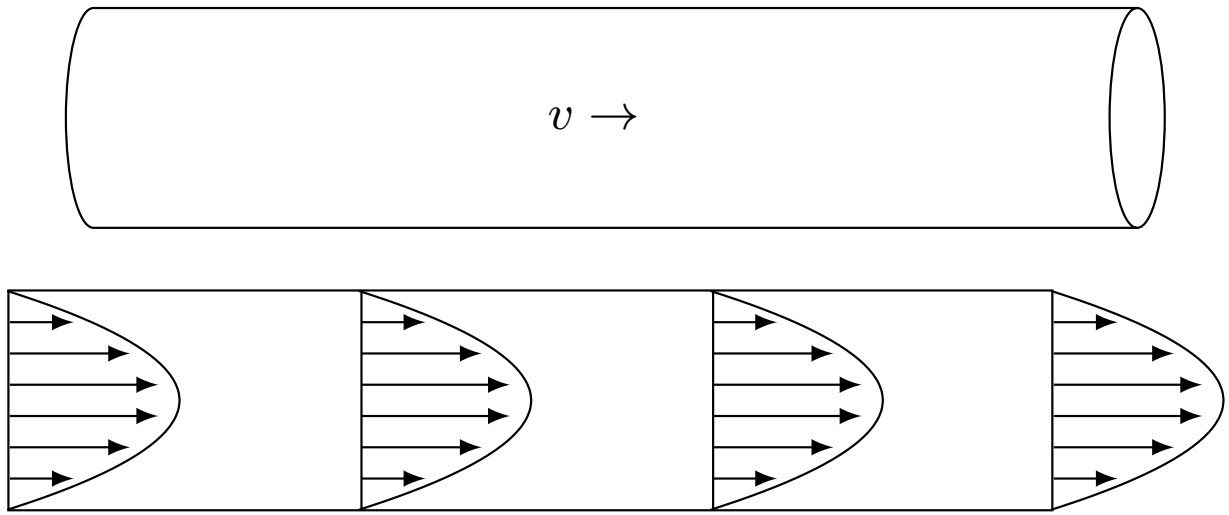


Left column: Kernel approximations (top), kernel approximation error (middle), and coefficients (bottom). Right column: True population density (top), which is measured very frequently, the error in the population density (middle), and the initial population density, N_0, the model parameter, κ, and the rate parameter, a (bottom). The colors are consistent across the plots.

Kernel for advective flow process

Kernel corresponding to advective flow

- Advective flow in a pipe with varying flow velocity



- What's coming out of the pipe?

$$f(t) = 2\pi \int_0^R C(t-\tau)v(r)r \, \mathrm{d}r \tag{27}$$

- Assumptions

$$v(R) = 0, \qquad \frac{\partial v}{\partial r} < 0 \tag{28}$$

Kernel corresponding to advective flow

- What's coming out of the pipe?

$$f(t) = 2\pi \int_0^R C(t-\tau) v(r) r \, \mathrm{d}r \tag{29}$$

  Objective: Change variable from r to τ

- Radius is an implicit function of the travel time: $r = r(\tau)$

$$v(r)\tau = L \tag{30}$$

- Increment in radius

$$\mathrm{d}r = \frac{\partial r}{\partial \tau} \mathrm{d}\tau \tag{31}$$

- Derivative of radius

$$\frac{\partial v}{\partial r}\frac{\partial r}{\partial \tau}\tau + v(r) = 0, \quad \frac{\partial r}{\partial \tau} = -\frac{v(r)}{\tau}\left(\frac{\partial v}{\partial r}\right)^{-1} = -\frac{L}{\tau^2}\left(\frac{\partial v}{\partial r}\right)^{-1} \tag{32}$$

- Change of variables (note the switch of integration bounds)

$$f(t) = \int_{\tau_0}^{\infty} \bar{\alpha}(\tau) C(t-\tau) \, \mathrm{d}\tau, \tag{33a}$$

$$\bar{\alpha}(\tau) = \begin{cases} -2\pi \frac{L^2}{\tau^3} \left(\frac{\partial v}{\partial r}\right)^{-1} r(\tau), & \tau \geq \tau_0, \\ 0, & \tau < \tau_0, \end{cases} \qquad \tau_0 = L/v(0) \tag{33b}$$

Kernel corresponding to advective flow

- Change of variables

$$f(t) = \int_{\tau_0}^{\infty} \bar{\alpha}(\tau) C(t-\tau)\,\mathrm{d}\tau, \tag{34a}$$

$$\bar{\alpha}(\tau) = \begin{cases} -2\pi \frac{L^2}{\tau^3} \left(\frac{\partial v}{\partial r}\right)^{-1} r(\tau), & \tau \geq \tau_0, \\ 0, & \tau < \tau_0, \end{cases} \qquad \tau_0 = L/v(0) \tag{34b}$$

- The normalization factor is the volumetric flow rate, F

$$\int_{\tau_0}^{\infty} \bar{\alpha}(\tau)\,\mathrm{d}\tau = \int_0^R v(r) r\,\mathrm{d}r = F \tag{35}$$

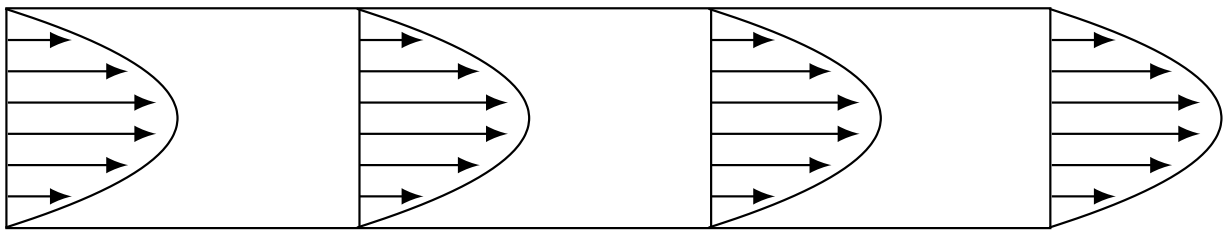
- Mass flow rate

$$f(t) = F \int_{-\infty}^{t} \alpha(t-s) C(s)\,\mathrm{d}s, \qquad \alpha(\tau) = \bar{\alpha}(\tau)/F \tag{36}$$

Example: Hagen-Poiseuille flow

- Quadratic velocity profile [8, Section 6.4.2]

$$v(r) = a(R^2 - r^2), \qquad a = \frac{\Delta P}{4\mu L} \tag{37}$$



- Minimum travel time and volumetric flow rate

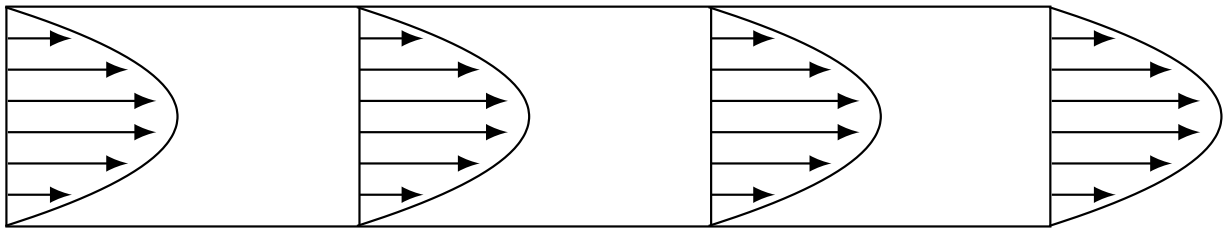
$$\tau_0 = \frac{L}{aR^2}, \qquad F = \frac{\pi}{2} a R^4 \tag{38}$$

- Kernel (and "mean")

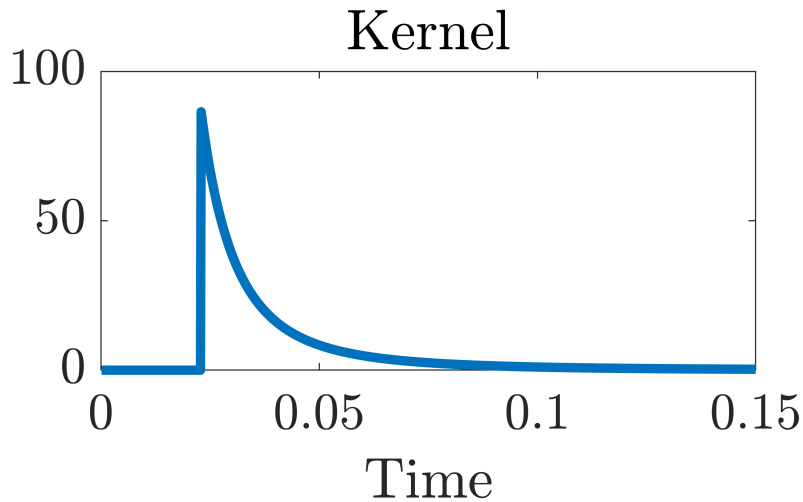
$$\alpha(\tau) = \begin{cases} 2\frac{\tau_0^2}{\tau^3}, & \tau \geq \tau_0, \\ 0, & \tau < \tau_0, \end{cases} \qquad \int_0^\infty \alpha(\tau)\tau \,\mathrm{d}\tau = 2\tau_0 \tag{39}$$

Example: Hagen-Poiseuille flow

- Quadratic velocity profile



- Kernel



Current and future topics

Linearization of the delayed quantity

- Memory state

$$z(t) = \int_{-\infty}^{t} \alpha(t-s) r(s) \, \mathrm{d}s \tag{40}$$

- Linearize the delayed quantity

$$r(s) \approx r(t) + \dot{r}(t)(s-t) \tag{41}$$

- Rewrite memory state

$$\begin{aligned} z(t) &= \int_{-\infty}^{t} \alpha(t-s) \left(r(t) + \dot{r}(t)(s-t) \right) \mathrm{d}s \\ &= r(t) \int_{-\infty}^{t} \alpha(t-s) \, \mathrm{d}s - \dot{r}(t) \int_{-\infty}^{t} \alpha(t-s)(t-s) \, \mathrm{d}s \\ &= r(t) - \dot{r}(t) \underbrace{\int_{0}^{\infty} \alpha(\tau) \tau \, \mathrm{d}\tau}_{\text{“Mean” of the density}} \end{aligned} \tag{42}$$

- Key observation: The result is a set of *implicit* differential equations
- Current work: Use the linearized model to solve optimal control problems
- Warning: The linearization might not capture the stability of steady states

How much does a PID controller remember?

- Classical example of a system with memory

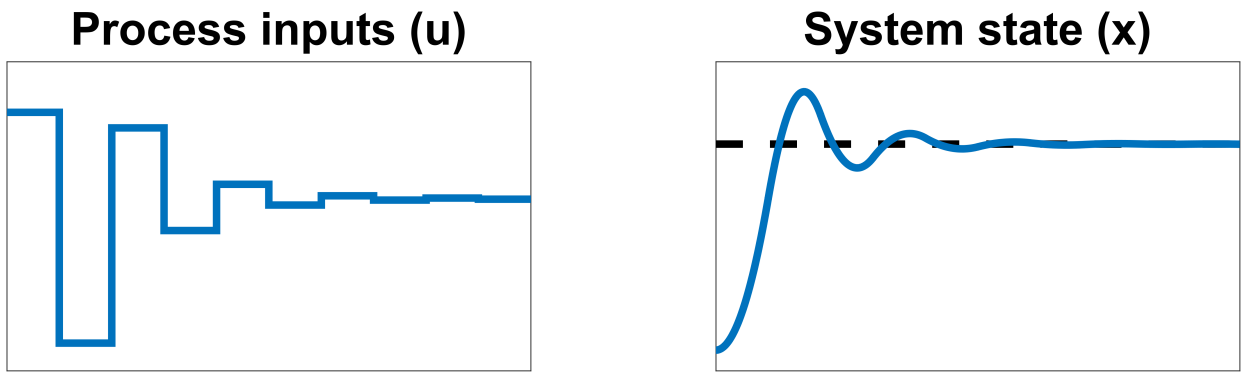
$$\dot{x}(t) = f(x(t), u(t)), \tag{43a}$$

$$u(t) = K_p e(t) + K_i \int_{t_0}^{t} e(s)\,\mathrm{d}s + K_p \dot{e}(t) \tag{43b}$$

- Tracking error

$$e(t) = x(t) - \bar{x} \tag{44}$$

- Example



- How much does the integral term "remember"?

$$u(t) = \cdots + K_i \int_{-\infty}^{t} \alpha(t-s) e(s)\,\mathrm{d}s + \cdots \tag{45}$$

How much *should* a PID controller remember? •

- Controlled system

$$\dot{x}(t) = f(x(t), u(t)), \tag{46}$$

$$u(t) = \cdots + K_i \int_{-\infty}^{t} \alpha(t-s)e(s)\,\mathrm{d}s + \cdots \tag{47}$$

- Tracking error

$$e(t) = x(t) - \bar{x} \tag{48}$$

- Question: How should we choose α, e.g., to stabilize the system?
- Motivation: Faults can cause *controller wind-up*

Summary

Summary

Key questions:

1. Given a kernel, α, what is the state, x?
   (Simulation)
2. Given a set of measurements, $\{y(t_k)\}_{k=0}^{N}$, what is the kernel, α?
   (Identification)
3. Given an estimated kernel, $\hat{\alpha}$, what are the underlying dynamics?
   (Interpretation)

Key innovation: Mixed Erlang kernel approximation

$$\alpha(t) \approx \sum_{m=0}^{M} c_m \hat{\alpha}_m(t), \qquad \hat{\alpha}_m(t) = b_m t^m e^{-at} \tag{49}$$

- Simulate
- Estimate kernel
- Derive analytical kernel (advective flow)
- Future topics (linearization and PID controllers)

Appendix 1: Sketch of proof

Sketch of proof – current status

- $\hat{\beta}^{(\infty)}(t) \to \beta(t)$ implies that $\hat{\beta}^{(\infty)}(t+\delta) - \hat{\beta}^{(\infty)}(t) \to \beta(t+\delta) - \beta(t)$
- Approximation error

$$
\begin{aligned}
|\hat{z}_i(t) - z_i(t)| &= \left| \int_{-\infty}^{t} \hat{\alpha}^{(\infty)}(t-s) r_i(s)\,\mathrm{d}s - \int_{-\infty}^{t} \alpha(t-s) r_i(s)\,\mathrm{d}s \right| \\
&= \left| \int_{-\infty}^{t} \left(\hat{\alpha}^{(\infty)}(t-s) - \alpha(t-s) \right) r_i(s)\,\mathrm{d}s \right| \\
&\leq \int_{-\infty}^{t} \left| \hat{\alpha}^{(\infty)}(t-s) - \alpha(t-s) \right| |r_i(s)|\,\mathrm{d}s \\
&\leq \int_{-\infty}^{t} \left| \hat{\alpha}^{(\infty)}(t-s) - \alpha(t-s) \right| \mathrm{d}s \sup_{s \in [0,\infty)} |r_i(s)| \qquad (50)
\end{aligned}
$$

- Pointwise error

$$
e(t) = \alpha^{(\infty)}(t) - \alpha(t) \qquad (51)
$$

- Error

$$
\int_{-\infty}^{t} \left| \hat{\alpha}^{(\infty)}(t-s) - \alpha(t-s) \right| \mathrm{d}s = \int_{0}^{t} |e(s)|\,\mathrm{d}s \qquad (52)
$$

Sketch of proof – current status

- Convergence of integral of error

$$\int_t^{t+\delta} \hat{\alpha}^{(\infty)}(s)\,\mathrm{d}s \to \int_t^{t+\delta} \alpha(s)\,\mathrm{d}s \tag{53}$$

- For every $\epsilon > 0$, there exists a $\Delta t > 0$ such that

$$\left| \int_t^{t+\delta} e(s)\,\mathrm{d}s \right| < \epsilon \tag{54}$$

- If there is a *finite* number of crossings (K) with the horizontal axis,

$$\int_0^{\infty} |e(s)|\,\mathrm{d}s = \sum_{k=1}^{K} \left| \int_{\tau_k}^{\tau_{k+1}} e(s)\,\mathrm{d}s \right| \tag{55}$$

- For any ϵ, choose Δt such that

$$\left| \int_{\tau_k}^{\tau_{k+1}} e(s)\,\mathrm{d}s \right| < \frac{\epsilon}{K \sup_{s\in[0,\infty)} |r_i(s)|} \tag{56}$$

  Then,

$$|\hat{z}_i(t) - z_i(t)| < \epsilon \tag{57}$$

- What should we require of α, r, etc.?

Appendix 2: Kernel domain

Identify domain for approximating kernel: Bisection

1. Choose t_h such that $1 - \beta(t_h) \leq \epsilon_h$
2. Given an initial interval, $[t_h^\ell, t_h^u]$, if $1 - \beta(t_h^u) < \epsilon_h$, shift the interval by $\Delta t = t_h^u - t_h^\ell$ to the right and repeat
3. Otherwise, a solution has been found if $|1 - \beta(t_h) - \epsilon_h|$ is sufficiently small (i.e., within some prespecified tolerance) for the midpoint $t_h^m = \frac{1}{2}(t_h^\ell + t_h^u)$
4. If a solution has not been found, repeat with the interval $[t_h^\ell, t_h^m]$ if $1 - \beta(t_h^m) < \epsilon_h$ and with the interval $[t_h^m, t_h^u]$ if $1 - \beta(t_h^m) > \epsilon_h$
5. That is, if t_h^m is too small, the next iteration is twice its value. If it is too large, the next iteration is halfway between the current and the last iteration. The iterations are stopped when $|1 - \beta(t_h) - \epsilon_h|$ is sufficiently small

Identify domain for approximating kernel: Newton's method

We parametrize $t_h = e^\lambda$ and solve the residual equation

$$F(t_h) = \ln(1 - \beta(t_h)) - \ln \epsilon_h = 0 \tag{58}$$

for λ using MATLAB's `fsolve`.
We supply the Jacobian of F to Matlab's fsolve. It is given by

$$\frac{\partial F}{\partial \lambda} = \frac{\partial F}{\partial t_h} \frac{\partial t_h}{\partial \lambda}, \tag{59}$$

where

$$\frac{\partial t_h}{\partial \lambda} = e^\lambda = t_h, \qquad \frac{\partial F}{\partial t_h} = \frac{-\alpha(t_h)}{1 - \beta(t_h)} \tag{60}$$

Identification revisited

The approximate system is **linear** in the coefficients and rate parameter •

- We did not exploit the linearity – can we do that?
- Rewrite approximate system

$$\dot{X}(t) = M(X(t), u(t), d(t), p)\hat{q} \tag{61}$$

- Step 1: Approximate the state time derivative, $\dot{X}(t_k)$
- Step 2: Solve the quadratic program

$$\min_{\hat{q}} \phi = \frac{1}{2} \sum_{k=0}^{N} \|\dot{X}(t_k) - M_k \hat{q}\|_2^2 \tag{62a}$$

subject to

$$w^T \hat{q} = 1, \tag{62b}$$

$$q_{\min} \leq \hat{q} \leq q_{\max} \tag{62c}$$

where $M_k = M(X(t_k), u(t_k), d(t_k), p)\hat{q}$

[1] T. J. Manetsch, "Time-varying distributed delays and their use in aggregative models of large systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-6, no. 8, pp. 547–553, 1976.

[2] H. C. Tijms, *Stochastic models: An algorithmic approach*. Wiley Series in Probability and Statistics, Wiley, 1995.

[3] H. Cossette, D. Landriault, E. Marceau, and K. Moutanabbir, "Moment-based approximation with mixed Erlang distributions," *Variance*, vol. 10, no. 1, pp. 166–182, 2016.

[4] H. Scheffé, "A useful convergence theorem for probability distributions," *The Annals of Mathematical Statistics*, vol. 18, no. 3, pp. 434–438, 1947.

[5] D. D. Boos, "A converse to Scheffé's theorem," *The Annals of Statistics*, vol. 13, no. 1, pp. 423–427, 1985.

[6] T. J. Sweeting, "On a converse to Scheffé's theorem," *The Annals of Statistics*, vol. 14, no. 3, pp. 1252–1256, 1986.

[7] W. Rudin, *Principles of mathematical analysis*. International Series in Pure and Applied Mathematics, McGraw-Hill, 3rd ed., 1976.

[8] D. P. Kessler and R. A. Greenkorn, *Momentum, heat, and mass transfer fundamentals*. Marcel Dekker, 1999.