

On mixture approximations for dynamical problems with distribution variables

Tobias K. S. Ritschel

Assistant Professor
tobk@dtu.dk, +45 4525 3315
Department of Applied Mathematics and Computer Science,
Technical University of Denmark

August 14th, 2025

Outline

1. What are mixture approximations?
 - ▶ Three different types of mixture approximations
 - ▶ Convergence
 - ▶ Practical implementation
 - ▶ What can you do with them?
2. Mixture approximations for delay differential equations (DDEs)
 - ▶ Examples of DDEs with distributed time delays
 - ▶ The linear chain trick (DDEs $\rightarrow$ ODEs)
 - ▶ Convergence
 - ▶ Simulation example (bifurcation analysis)
 - ▶ Kernel identification

Mixture approximations

Problem statement

Approximate the $L^1(\mathcal{D})$ function α with the following properties

1. Non-negative and bounded

$$0 \leq \alpha(t) \leq K, \quad t \in \mathcal{D}$$

2. Integrates to one

$$\int_{\mathcal{D}} \alpha(t) \, dt = 1$$

3. Continuous

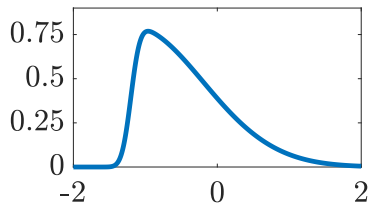
The domain may be

- ▶ the real numbers, $\mathcal{D} = (-\infty, \infty)$,
- ▶ the non-negative real numbers, $\mathcal{D} = [0, \infty)$, or
- ▶ an interval, $\mathcal{D} = [a, b]$

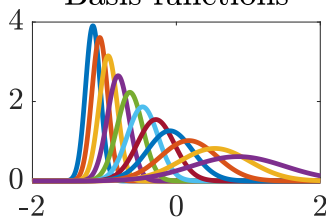
Example for $\mathcal{D} = (-\infty, \infty)$

Skew normal distribution

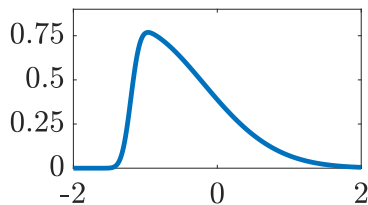
True density



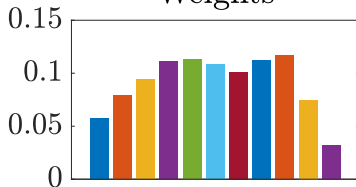
Basis functions



Approximate density



Weights



Mixture approximation

Approximate α by a sum of basis functions

$$\alpha(t) \approx \hat{\alpha}(t) = \sum_{m=0}^M c_m \ell_m(t)$$

Weight constraints

$$\sum_{m=0}^M c_m = 1, \quad 0 \leq c_m \leq 1, \quad m = 0, \dots, M$$

Non-negative basis functions

$$\ell_m(t) \geq 0, \quad m = 0, \dots, M$$

Basis functions integrate to one

$$\int_{\mathcal{D}} \ell_m(t) \, dt = 1, \quad m = 0, \dots, M$$

Gaussian, Erlang, and beta mixture approximations

Gaussian basis functions, $t \in \mathcal{D} = (-\infty, \infty)$

$$\ell_m(t) = b_m \exp \left(-\frac{1}{2} \left(\frac{t - \mu_m}{\sigma_m} \right)^2 \right), \quad b_m = \frac{1}{\sqrt{2\pi\sigma_m^2}}$$

Erlang basis functions, $t \in \mathcal{D} = [0, \infty)$

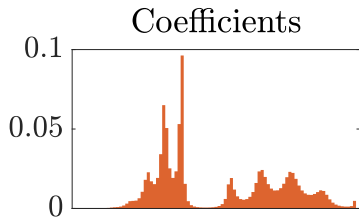
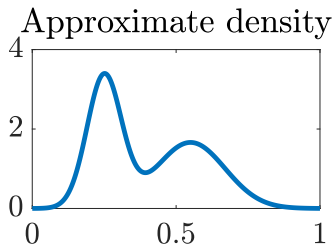
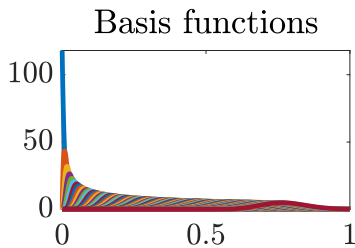
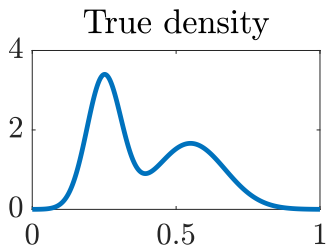
$$\ell_m(t) = b_m t^m e^{-at}, \quad b_m = \frac{a^{m+1}}{m!}$$

Beta basis functions, $t \in \mathcal{D} = [0, \Delta t]$

$$\ell_m(t) = b_m t^m (\Delta t - t)^{M-m}, \quad b_m = \frac{1}{\Delta t^{M+1}} \frac{(M+1)!}{m!(M-m)!}$$

Example: Erlang mixture approximation ($M = 90$)

Folded normal mixture distribution



Convergence of Erlang mixture approximations ($M = \infty$)

Coefficients

$$c_m = \int_{s_m}^{s_{m+1}} \alpha(s) \, ds, \quad s_m = m\Delta s, \quad \Delta s = 1/a \quad (1)$$

Weak convergence (Tijms [4, 5])

$$\int_0^t \hat{\alpha}(s) \, ds \rightarrow \int_0^t \alpha(s) \, ds \quad \text{as } a \rightarrow \infty$$

Pointwise convergence (based on delta families [6])

$$\hat{\alpha}(t) \rightarrow \alpha(t) \quad \text{as } a \rightarrow \infty$$

Vitali's convergence lemma

$$\int_0^\infty |\hat{\alpha}(s) - \alpha(s)| \, ds \rightarrow 0 \quad \text{as } a \rightarrow \infty$$

if

- ▶ $\hat{\alpha}$ converges pointwise ✓
- ▶ $\hat{\alpha}$ is tight ✓
- ▶ $\hat{\alpha}$ is uniformly integrable ✓

Convergence of beta and Gaussian mixture approximations

1. Beta mixture approximations, $\mathcal{D} = [0, 1]$

- ▶ ℓ_m is a Bernstein polynomial scaled by $M + 1$
- ▶ pointwise convergence is guaranteed by Weierstrass' approximation theorem [1] as $M \rightarrow \infty$

$$c_m = \frac{1}{M+1} \alpha(s_m), \quad s_m = m\Delta s, \quad \Delta s = 1/M,$$
$$c_m = \int_{s_m}^{s_{m+1}} \alpha(s) \, ds, \quad s_m = m\Delta s, \quad \Delta s = 1/(M+1)$$

2. Gaussian mixture approximations, $\mathcal{D} = (-\infty, \infty)$

- ▶ Error analysis by Maz'ya and Schmidt [2]
- ▶ Convergence proof by Sorenson and Alspach [3]

Optimal coefficients and rate parameter

Time domain ($k = 0, \dots, N - 1$)

$$\int_0^{t_h} \alpha(s) \, ds = 1 - \epsilon, \quad t_k = k/t_h$$

Least-squares optimization problem

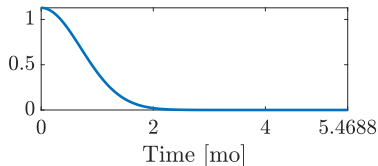
$$\begin{aligned} \min_{\{c_m\}_{m=0}^M, a} \quad & \phi = \frac{1}{2} \sum_{k=0}^{N-1} (\alpha(t_k) - \hat{\alpha}(t_k))^2 \Delta t, \\ \text{subject to} \quad & \sum_{m=0}^M c_m = 1, \\ & 0 \leq c_m \leq 1, \quad m = 0, \dots, M, \\ & a_{\min} \leq a \end{aligned}$$

- ▶ Use Matlab's `fmincon`
- ▶ Provide analytical first- and second-order derivatives

Example: Gaussian kernel

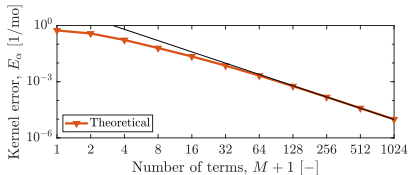
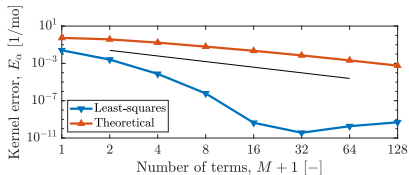
Kernel

$$\alpha(t) = \frac{2}{\sqrt{\pi}} e^{-t^2}$$



Error

$$E_{\alpha} = \sum_{k=0}^{N_{\alpha}-1} (\hat{\alpha}(t_k) - \alpha(t_k))^2 \Delta t$$



Outline

1. What are mixture approximations?
 - ▶ Three different types of mixture approximations
 - ▶ Convergence
 - ▶ Practical implementation
 - ▶ What can you do with them?
2. Mixture approximations for delay differential equations (DDEs)
 - ▶ Examples of DDEs with distributed time delays
 - ▶ The linear chain trick (DDEs $\rightarrow$ ODEs)
 - ▶ Convergence
 - ▶ Simulation example (bifurcation analysis)
 - ▶ Kernel identification

Differential equations with distributed time delays

Delay differential equations

Logistic equation

$$\dot{x}(t) = x(t) (1 - x(t))$$

General form (ODE)

$$\dot{x}(t) = f(x(t))$$

Logistic equation w. absolute time delay

$$\dot{x}(t) = x(t) (1 - x(t - \tau))$$

General form (DDE)

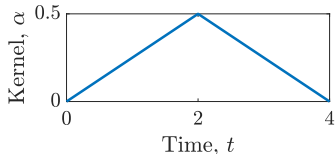
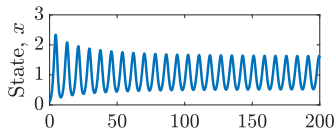
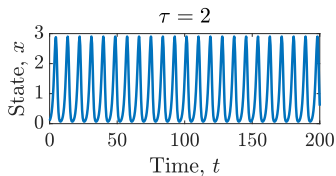
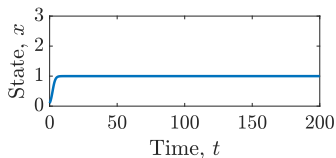
$$\dot{x}(t) = f(x(t), x(t - \tau))$$

Logistic equation w. distributed time delay

$$\dot{x}(t) = x(t) \left(1 - \int_{-\infty}^t \alpha(t-s)x(s) ds \right)$$

General form (DDDE)

$$\dot{x}(t) = f(x(t), \int_{-\infty}^t \alpha(t-s)x(s) ds)$$



Epidemiology

SIR model without delays

$$\dot{S}(t) = -\beta S(t)I(t),$$

$$\dot{I}(t) = \beta S(t)I(t) - \eta I(t),$$

$$\dot{R}(t) = \eta I(t)$$

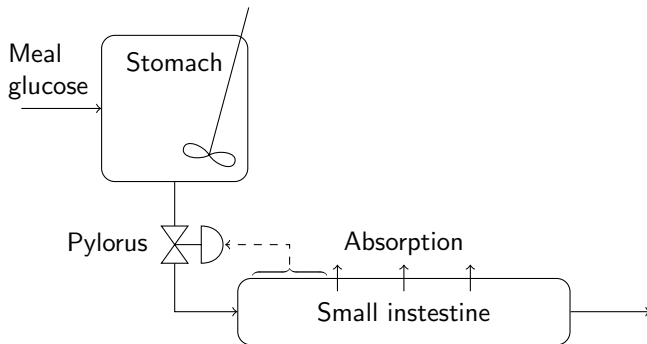
SIR model with distributed time delay

$$\dot{S}(t) = -\beta S(t) \int_{-\infty}^t \alpha(t-s)I(s) \, ds,$$

$$\dot{I}(t) = \beta S(t) \int_{-\infty}^t \alpha(t-s)I(s) \, ds - \eta I(t),$$

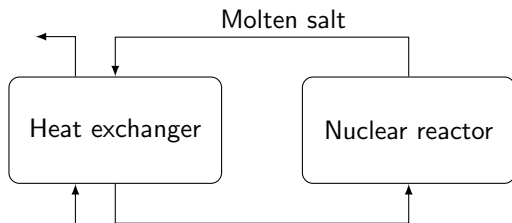
$$\dot{R}(t) = \eta I(t)$$

Diabetes

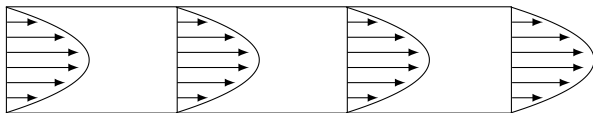


Ritschel, T.K.S., Reenberg, A.T., Carstensen, P.E., Bendsen, J., Jørgensen, J.B., 2023. Mathematical Meal Models for Simulation of Human Metabolism. arXiv: 2307.16444.

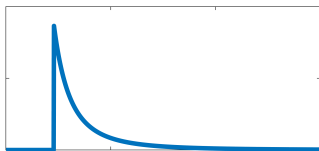
Molten salt nuclear reactor and nonuniform flow in pipes



Non-uniform velocity profile



Kernel for Hagen-Poiseuille flow (quadratic velocity profile)



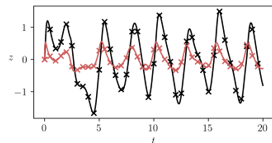
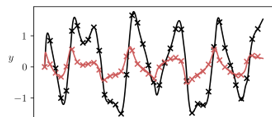
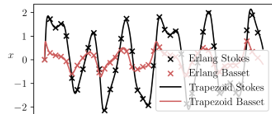
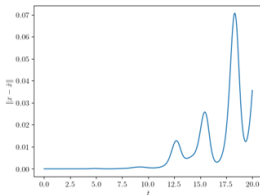
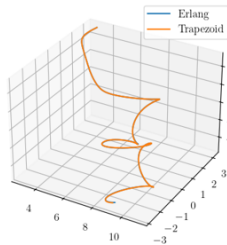
Ritschel, T.K.S., 2025. Numerical Optimal Control for Distributed Delay Differential Equations: A Simultaneous Approach based on Linearization of the Delayed Variables. In: Proceedings of the 2025 European Control Conference (ECC), June 24-27, Thessaloniki, Greece. arXiv: 2410.15083.

Particle flow in velocity field

Particle subject to Stoke's drag force and Basset history force

$$\dot{x}_p = u_p,$$

$$\dot{u}_p = \underbrace{\frac{1}{St} \mathcal{F}(a)(u_p - u_f)}_{\text{Nonlinear drag}} + \underbrace{C \int_0^t \frac{1}{\sqrt{t-s}} (\dot{u}_p - \dot{u}_f) ds}_{\text{Basset history force}}$$



Collaboration with PhD candidate Zejian You, Asst. Prof. Qi Wang, and Prof. Gustaaf Jacobs from San Diego State University.

Linear chain trick (DDE $\rightarrow$ ODE)

Linear chain trick

DDE with distributed time delay

$$\dot{x}(t) = f(x(t), z(t)), \quad z(t) = \int_{-\infty}^t \alpha(t-s)x(s) \, ds$$

Approximate kernel

$$\alpha(t) \approx \hat{\alpha}(t) = \sum_{m=0}^M c_m \ell_m(t)$$

Substitute into integral

$$z(t) \approx \hat{z}(t) = \sum_{m=0}^M c_m \int_{-\infty}^t \ell_m(t-s)x(s) \, ds = \sum_{m=0}^M c_m z_m(t)$$

Differentiate z_m ($\dot{\ell}_0 = -a\ell_0(t)$, $\dot{\ell}_m(t) = a(\ell_{m-1}(t) - \ell_m(t))$)

$$\begin{aligned} \dot{z}_m(t) &= \ell_m(0)x(t) + \int_{-\infty}^t \dot{\ell}_m(t-s)x(s) \, ds \\ &= \begin{cases} a(x(t) - z_0(t)), & m = 0, \\ a(z_{m-1}(t) - z_m(t)), & m > 0 \end{cases} \end{aligned}$$

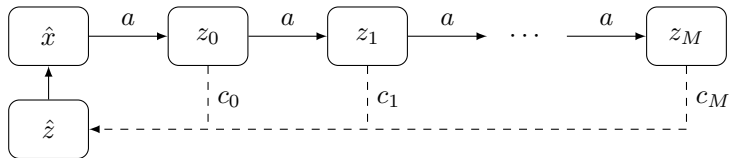
Approximate ODEs

System

$$\dot{\hat{x}}(t) = f(\hat{x}(t), \hat{z}(t)), \quad \hat{z}(t) = \sum_{m=0}^M c_m z_m(t)$$

Auxiliary memory states

$$\begin{aligned} \dot{z}_0(t) &= a(\hat{x}(t) - z_0(t)), \\ \dot{z}_m(t) &= a(z_{m-1}(t) - z_m(t)), \quad m = 1, \dots, M \end{aligned}$$



Approximate ODEs

Approximate ODEs

$$\begin{aligned}\dot{\hat{x}}(t) &= f(\hat{x}(t), \hat{z}(t)), \\ \dot{Z}(t) &= AZ(t) + B\hat{x}(t), \\ \hat{z}(t) &= CZ(t)\end{aligned}$$

Matrices

$$A = a \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}, \quad B = a \begin{bmatrix} 1 \\ \\ \\ \end{bmatrix},$$
$$C = [c_0 \quad c_1 \quad \cdots \quad c_M]$$

Does the approximate state converge ($M = \infty$)?

Bound on state error [7, Thm. 1] when f is Lipschitz and x is bounded

$$\int_0^\infty |\hat{\alpha}(t) - \alpha(t)| dt < \epsilon \quad \Rightarrow \quad |\hat{x}(t) - x(t)| \leq \epsilon w(t) \leq \epsilon w(t_f)$$

Auxiliary function

$$w(t) = 0, \quad t \in (-\infty, t_0],$$

$$\dot{w}(t) = L_z K_x + L_x w(t) + L_z \int_{-\infty}^t \hat{\alpha}(t-s) w(s) ds, \quad t \in [t_0, t_f]$$

- ▶ The steady state of the approximate ODEs and the DDEs is the same
- ▶ The eigenvalues converge to the roots for the DDEs (Hurwitz' convergence theorem)
- ▶ Error dynamics are locally stable if the DDEs and ODEs are locally stable

Ritschel, T.K.S., 2025. On Erlang Mixture Approximations for Differential Equations with Distributed Time Delays. arXiv: 2502.12984. In submission.

Numerical example: Simulation

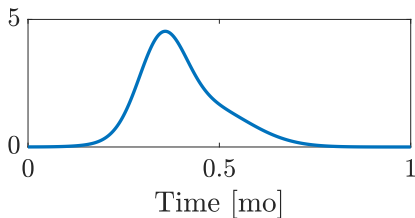
Numerical example

► System

$$\dot{N}(t) = \kappa N(t) \left(1 - \frac{1}{K} \int_{-\infty}^t \alpha(t-s) N(s) \, ds \right)$$

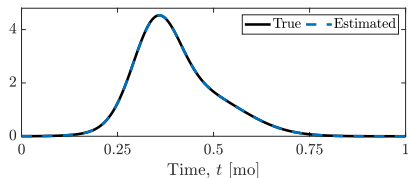
► Kernel

$$\alpha(t) = \gamma_1 F(t; \mu_1, \sigma_1) + \gamma_2 F(t; \mu_2, \sigma_2),$$
$$F(t; \mu, \sigma) = \frac{\exp\left(-\frac{1}{2} \left(\frac{t-\mu}{\sigma}\right)^2\right) + \exp\left(-\frac{1}{2} \left(\frac{t+\mu}{\sigma}\right)^2\right)}{\sqrt{2\pi}\sigma}$$

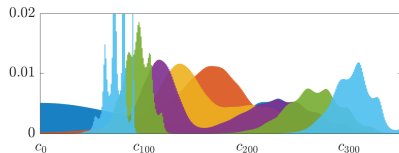
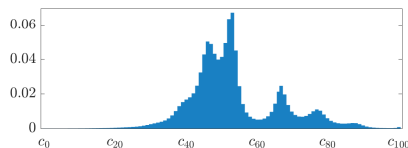
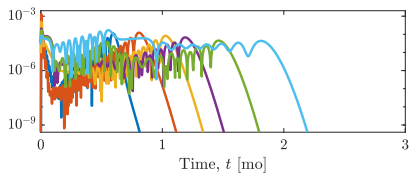
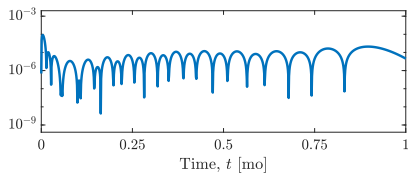
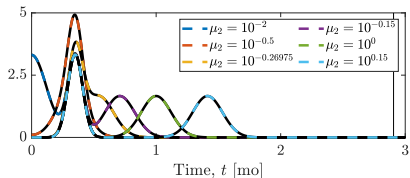


Numerical example: Bifurcation analysis

Model parameter

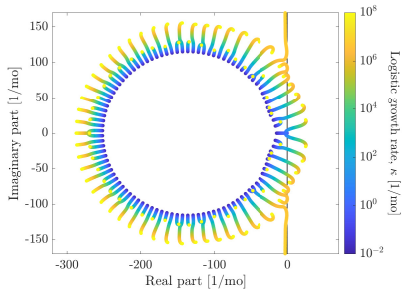


Kernel parameter

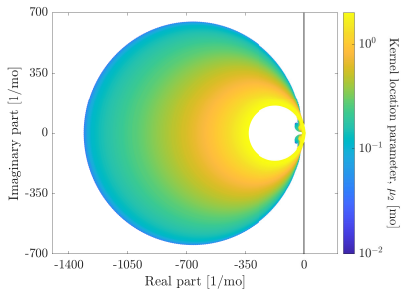


Numerical example: Bifurcation analysis

Model parameter

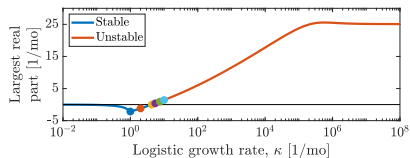


Kernel parameter

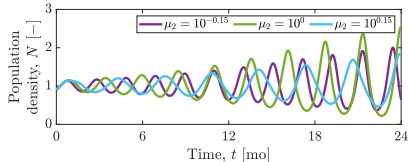
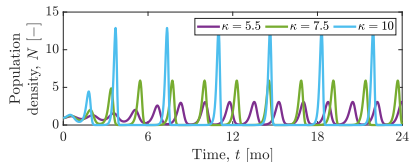
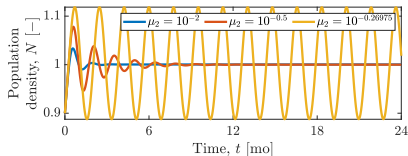
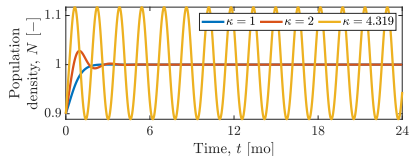
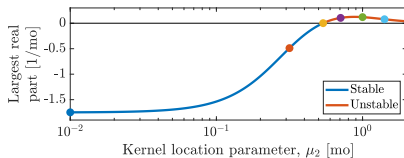


Numerical example: Bifurcation analysis

Model parameter



Kernel parameter

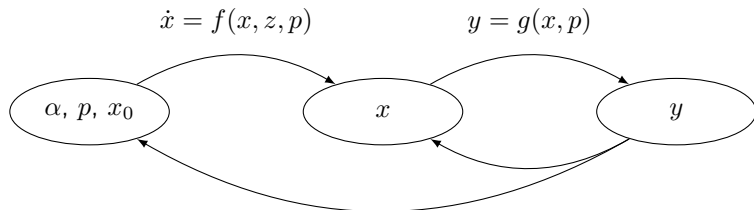


The numerical simulations are obtained with Euler's implicit method and a right rectangle rule for approximating the integral.

Kernel identification

Kernel identification (with John Wyller, NMBU, Norway)

Forward problem



Inverse problem

Ritschel, T.K.S., Wyller, J., 2025. An Algorithm for Distributed Time Delay Identification without A Priori Knowledge of the Kernel. *Automatica* 178, pp. 112382. DOI: 10.1016/j.automatica.2025.112382.

Dynamical least-squares problem

$$\min_{\{c_m\}_{m=0}^M, a, p, x_0} \frac{1}{2} \sum_{k=0}^N \|y_k - y(t_k)\|_2^2$$

subject to

$$x(t_0) = x_0,$$

$$Z(t_0) = Z_0(x_0, p),$$

$$\dot{x}(t) = f(x(t), z(t), p), \quad t \in [t_0, t_f],$$

$$\dot{Z}(t) = AZ(t) + Bx(t), \quad t \in [t_0, t_f],$$

$$z(t) = CZ(t), \quad t \in [t_0, t_f],$$

$$y(t_k) = g(x(t_k), p), \quad k = 0, \dots, N,$$

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

$$a_{\min} \leq a,$$

$$p_{\min} \leq p \leq p_{\max},$$

$$x_{\min} \leq x_0 \leq x_{\max}$$

Example

► System

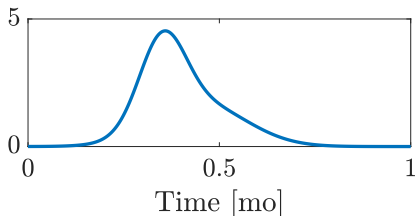
$$\dot{N}(t) = \kappa N(t) \left(1 - \frac{1}{K(t)} \int_{-\infty}^t \alpha(t-s) N(s) \, ds \right),$$

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

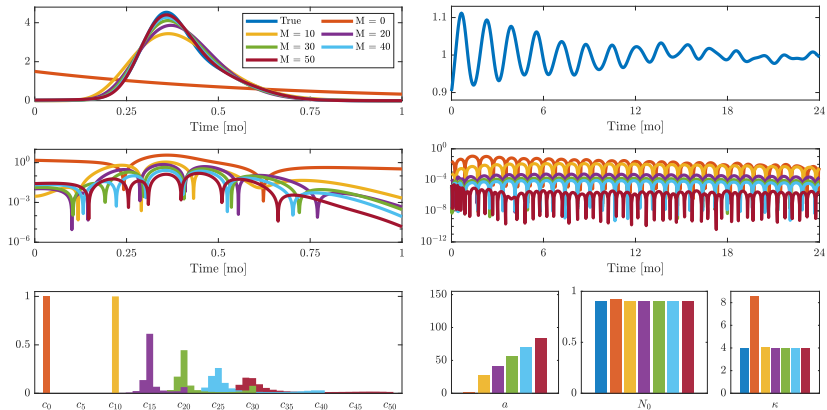
► Kernel

$$\alpha(t) = \gamma_1 F(t; \mu_1, \sigma_1) + \gamma_2 F(t; \mu_2, \sigma_2),$$

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



Example



Approximate controllers for DDEs

System

$$\dot{x}(t) = Fx(t) + Gu(t) + H \int_{-\infty}^t \alpha(t-s)x(s) \, ds$$

Control law

$$u(t_k) = -L_x x(t) - L_\gamma \int_{-\infty}^t \gamma(t-s)x(s) \, ds$$

- ▶ Optimal control
- ▶ State estimation
- ▶ Delay-adaptive optimal control

Controllers with memory for ODEs

PID controller with forgetting factor

$$u(t) = K_p e(t) + \int_0^t \alpha(t-s)e(s) \, ds + K_d \dot{e}(t)$$

Relation/similarity to fractional order PID controllers?

System

$$\dot{x}(t) = Fx(t) + Gu(t)$$

Control law

$$u(t_k) = -L_x x(t) - L_\gamma \int_{-\infty}^t \gamma(t-s)x(s) \, ds$$

Summary

1. Approximate $L^1(\mathcal{D})$ functions

$$\alpha(t) \approx \hat{\alpha}(t) = \sum_{m=0}^M c_m \ell_m(t), \quad t \in \mathcal{D}$$

2. Approximate solution to DDEs with distributed time delays

$$\dot{x}(t) = f(x(t), \int_{-\infty}^t \alpha(t-s)x(s) \, ds)$$

3. Many different applications/possible extensions

- ▶ Model reduction
- ▶ Stochastic differential equations
- ▶ Continuous thermodynamics (chemical phase/reaction equilibria)
- ▶ Optimal/feedback/adaptive control for DDEs
- ▶ Controllers with memory for ODEs (PID, LQR, MPC, etc.)

Model reduction

Thank you!

Bibliography I

- [1] G. G. Lorentz, *Bernstein polynomials*. New York: Chelsea Publishing Company, 2nd ed., 1986.
- [2] V. Maz'ya and G. Schmidt, "On approximate approximations using Gaussian kernels," *IMA Journal of Numerical Analysis*, vol. 16, pp. 13–29, 1996.
- [3] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, 1971.
- [4] H. C. Tijms, *Stochastic models: An algorithmic approach*. Wiley Series in Probability and Statistics, Wiley, 1995.
- [5] 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.
- [6] J. Korevaar, *Mathematical methods*, vol. 1. Academic Press, 1968.
- [7] N. Guglielmi and E. Hairer, "Applying stiff integrators for ordinary differential equations and delay differential equations to problems with distributed delays," *SIAM Journal on Scientific Computing*, vol. 47, no. 1, pp. A102–A123, 2025.

Bibliography II

- [8] T. K. S. Ritschel, “Numerical optimal control for distributed delay differential equations: A simultaneous approach based on linearization of the delayed variables.” arXiv:2410.15083, 2024. Preprint.
- [9] T. K. S. Ritschel, “On Erlang mixture approximations for differential equations with distributed time delays.” arXiv:2502.12984, 2025. In submission.
- [10] T. K. S. Ritschel and J. Wyller, “An algorithm for distributed time delay identification without a priori knowledge of the kernel,” *Automatica*, vol. 178, p. 112382, 2025.

Pointwise convergence based on delta families

Identity based on Dirac's delta, δ

$$\alpha(t) = \int_{\mathcal{D}} \delta(t-s)\alpha(s) \, ds$$

Express approximations in terms of delta family, δ_a

$$\hat{\alpha}(t) = \int_{\mathcal{D}} \delta_a(t-s)\alpha(s) \, ds$$

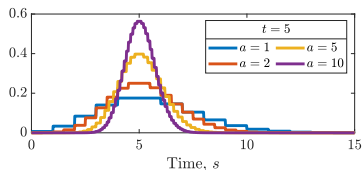
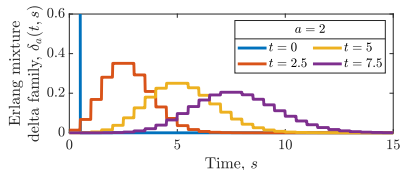
If δ_a “converges to” δ , then $\hat{\alpha}(t)$ converges to $\alpha(t)$

See, e.g., the book by Korevaar [6]

What should the delta family be?

Erlang mixture delta family

$$\delta_a(t, s) = \ell_m(t), \quad s \in [s_m, s_{m+1}), \quad s_m = m\Delta s, \quad \Delta s = 1/a,$$
$$\ell_m(t) = b_m t^m e^{-at}$$



Opportunities for model reduction?

Linear system

$$\begin{aligned}\dot{x}(t) &= Fx(t) + Ez(t) + Gu(t), \\ y(t) &= Hx(t)\end{aligned}$$

Approximation of integral term $z(t) = \int_{-\infty}^t \alpha(t-s)x(s) \, ds$

$$\begin{aligned}\dot{Z}(t) &= AZ(t) + Bx(t), \\ z(t) &= CZ(t)\end{aligned}$$

Can we reduce the combined system when Z is large?

Can we use model reduction to determine better basis functions, ℓ_m ?

Stochastic differential equations

Stochastic differential equation

$$dx(t) = f(x(t)) dt + g(x(t)) dw(t)$$

Solution (time-varying probability density function)

$$p(t, x) = ?$$

Fokker-Planck equation

$$\partial_t p(t, x) = -\partial_x (f(x)p(t, x)) + \frac{1}{2} \partial_{xx} (g^2(x)p(t, x))$$

Mixture approximation

$$p(t, x) \approx \hat{p}(t, x) = \sum_{m=0}^M c_m(t) \ell_m(x)$$

Substitute into Fokker-Planck equation

$$\begin{aligned} R(t, x) &= \sum_{m=0}^M \left(\ell_m(x) \dot{c}_m(t) + \left[\partial_x (f(x) \ell_m(x)) - \frac{1}{2} \partial_{xx} (g^2(x) \ell_m(x)) \right] c_m(t) \right) \\ &= \sum_{m=0}^M \left(\ell_m(x) \dot{c}_m(t) + A_m(x) c_m(t) \right) \end{aligned}$$

Minimize residual function

Least-squares approach

$$\min_{\{\dot{c}_m\}_{m=0}^M} \phi = \frac{1}{2} \int_{\mathcal{D}} R^2(t, x) w(x) dx$$

First-order optimality conditions

$$\begin{aligned} \frac{\partial \phi}{\partial \dot{c}_j} = \sum_{m=0}^M \left(\int_{\mathcal{D}} \ell_m(x) \ell_j(x) w(x) dx \dot{c}_m(t) \right. \\ \left. + \int_{\mathcal{D}} \ell_j(x) A_m(x) w(x) dx c_m(t) \right) = 0 \end{aligned}$$

If the basis functions are orthonormal wrt. the weight function w

$$\dot{c}_j(t) = - \sum_{m=0}^M \int_{\mathcal{D}} \ell_j(x) A_m(x) w(x) dx c_m(t) = \sum_{m=0}^M w_{jm} c_m(t)$$

How do we compute w_{jm} and ensure that $0 \leq c_m$ and $\sum_{m=0}^M c_m(t) = 1$?