WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

ODE Solvers as Bayesian State Estimation

Filip Tronarp

CHALMERS UNIVERSITY OF TECHNOLOGY · KTH VETENSKAP OCH KONST · LINKÖPINGS UNIVERSITET · LUND UNIVERSITY · UMEÅ UNIVERSITET

Differential equation:

$$\dot{y}(t) = v(y(t)), \quad y(0) = \zeta, \quad t \in [0, T]$$

Solution:

$$\begin{aligned} y(t) &= y(0) + \dot{y}(0)t + \ddot{y}(0)\frac{t^2}{2} + \dots, \\ &= \zeta + v(\zeta)t + \frac{dv}{dy}(\zeta)\frac{t^2}{2} + \dots \end{aligned}$$

Taylor series are impractical in general, what do?

Ordinary differential equations

Integral representation

Equivalent integral equation:

$$y(t) = \zeta + \int_0^t v(y(\tau))\,\mathrm{d}\tau = \underbrace{\zeta + \int_0^s v(y(\tau))\,\mathrm{d}\tau}_{=y(s)} + \int_s^t v(y(\tau))\,\mathrm{d}\tau$$

$$= y(s) + \int_s^t v(y(\tau))\,\mathrm{d}\tau = \varphi_{t,s}(y(s))$$

Solution on a grid:

$$y(t_m) = \varphi_{t_m,t_{m-1}}(y(t_{m-1})), \quad m = 1, \ldots, n.$$

Ordinary differential equations

Numerical solutions

Need approximation:

$$\begin{aligned}\hat{\varphi}_{t_m,t_{m-1}}(y(t_{m-1})) &\approx \varphi_{t_m,t_{m-1}}(y(t_{m-1})) \\ &= y(t_{m-1}) + \int_{t_{m-1}}^{t_m} v(y(\tau))\,\mathrm{d}\tau\end{aligned}$$

"Easy" to approximate if $\delta_m = t_m - t_{m-1}$ is small:

$$\int_{t_{m-1}}^{t_m} v(y(\tau))\,\mathrm{d}\tau \approx \delta_m v(y(t_{m-1})) + O(\delta_m^2)$$

Explicit Euler:

$$\hat{\varphi}_{t_m,t_{m-1}}(y) = y + \delta_m v(y).$$

A first attempt at probabilistic solvers

Problem formulation

- Differentiable prior:

$$x = \begin{pmatrix} y \\ \dot{y} \end{pmatrix} \sim \mathcal{GP}(\mu_0, k_0)$$

- Initial data:

$$x(0) = \begin{pmatrix} \zeta \\ v(\zeta) \end{pmatrix}$$

- Data:

$$0 = x_2(t_m) - v(x_1(t_m)), \quad m = 1, \dots, n$$

"Nonlinear GP regression without noise"

A first attempt at probabilistic solvers

Solution

Maximum a posterior problem:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} (\boldsymbol{x} - \boldsymbol{\mu}_0)^* \boldsymbol{K}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_0)$$

$$\text{s.t } \boldsymbol{x}_2 = v(\boldsymbol{x}_1)$$

Problem: inversion of Gram matrix is $O(n^3)$

Need: some model that mimics the behaviour:

$$y(t_{m+1}) = \varphi_{t_{m+1},0}(\zeta) = \varphi_{t_{m+1},t_m}(\varphi_{t_m,0}(\zeta))$$

A first attempt at probabilistic solvers

Illustration

Markov processes

Definition

Markov property:

$$\pi(t, x \mid t_{1:n}, x_{1:n}) = \pi(t, x \mid t_n, x_n) = f_{t,t_n}(x \mid x_n)$$

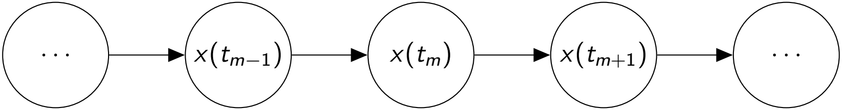
Joint distribution:

$$\begin{aligned}
\pi(t_{0:n}, x_{0:n}) &= \pi(t_n, x_n \mid t_{1:n-1}, x_{1:n-1})\pi(t_{1:n-1}, x_{1:n-1}) \\
&= f_{t_n,t_{n-1}}(x_n \mid x_{n-1})\pi(t_{1:n-1}, x_{1:n-1}) \\
&= \dots \\
&= \pi_0(x_0) \prod_{m=1}^{n} f_{t_m,t_{m-1}}(x_m \mid x_{m-1})
\end{aligned}$$

Compact representation: only need to specify π_0 and $f_{t,s}$.

Markov processes

Cutting time in half



Future is conditionally independent of the past:

$$
\begin{aligned}
&\pi(t_{m+1}, t_{m-1}, x_{m+1}, x_{m-1} \mid t_m, x_m) \\
&\quad = f_{t_{m+1},t_m}(x_{m+1} \mid x_m) \frac{f_{t_m,t_{m-1}}(x_m \mid x_{m-1})\pi(t_{m-1}, x_{m-1})}{\pi(t_m, x_m)} \\
&= \pi(t_{m+1}, x_{m+1} \mid t_m, x_m)\pi(t_{m-1}, x_{m-1} \mid t_m, x_m)
\end{aligned}
$$

Gauss–Markov process:

$$x(0) \sim \mathcal{N}(\mu_0, \Sigma_0), \tag{1a}$$
$$x(t) \mid x(s) \sim \mathcal{N}(\Phi(t,s)x(s), Q(t,s)) \tag{1b}$$

For instance solutions to linear SDEs:

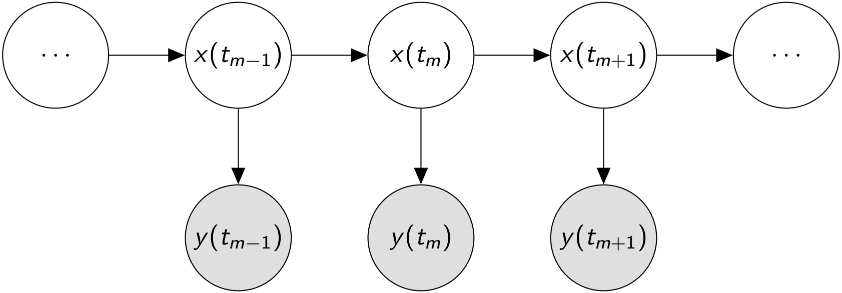
$$\mathrm{d}x(t) = Ax(t)\,\mathrm{d}t + B\,\mathrm{d}w(t)$$

Transition parameters:

$$\Phi(t,s) = e^{A(t-s)} \tag{2a}$$
$$Q(t,s) = \int_0^{t-s} e^{A\tau} BB^* e^{A^*\tau}\,\mathrm{d}\tau \tag{2b}$$

Bayesian state estimation

Partially observed Markov processes



Bayesian state estimation

Problem formulation

Partially observed Markov process:

$$x(0) \sim \pi_0(\cdot\,), \tag{3a}$$

$$x(t) \mid x(s) \sim f_{t|s}(\cdot \mid x(s)) \tag{3b}$$

$$y(t) \mid x(t) \sim g_t(\cdot \mid x(t)) \tag{3c}$$

Estimation problem

Given observations y on a grid $t_1, t_2, \ldots, t_n$ of a Markov process, estimate x.

Bayesian state estimation

Prior, Likelihood, and marginal likelihood

Prior:

$$\pi(t_{0:k}, x_{0:k}) = \pi(t_0, x_0) \prod_{m=1}^{k} f_{t_m, t_{m-1}}(x_m \mid x_{m-1})$$
$$= f_{t_k, t_{k-1}}(x_k \mid x_{k-1}) \pi(t_{0:k-1}, x_{0:k-1})$$

Likelihood:

$$L(t_{1:k}, x_{1:k}) = \prod_{m=1}^{k} g_{t_m}(y_m \mid x_m) = g_{t_k}(y_k \mid x_k) L(t_{1:k-1}, x_{1:k-1})$$

Marginal likelihood:

$$M(t_{1:k}, y_{1:k}) = \int L(t_{1:k}, x_{1:k}) \pi(t_{0:k}, x_{0:k}) \, \mathrm{d}x_{0:k}$$

Bayesian state estimation

The posterior recursion

The posterior:

$$\gamma(t_{1:k}, x_{1:k} \mid t_{1:k}, y_{1:k}) = \frac{L(t_{1:k}, y_{1:k})\pi(t_{0:k}, x_{0:k})}{M(t_{1:k}, y_{1:k})}$$

Rercursion:

$$\begin{aligned}
&\gamma(t_{1:k}, x_{1:k} \mid t_{1:k}, y_{1:k}) \\
&= \frac{g_{t_k}(y_k \mid x_k) f_{t_k, t_{k-1}}(x_k \mid x_{k-1})}{M(t_{1:k}, y_{1:k})} L(t_{1:k-1}, y_{1:k-1})\pi(t_{0:k-1}, x_{0:k-1}) \\
&= \frac{M(t_{1:k-1}, y_{1:k-1})}{M(t_{1:k}, y_{1:k})} g_{t_k}(y_k \mid x_k) f_{t_k, t_{k-1}}(x_k \mid x_{k-1}) \\
&\quad \times \gamma(t_{1:k-1}, x_{1:k-1} \mid t_{1:k-1}, y_{1:k-1})
\end{aligned}$$

N.B:

$$\frac{M(t_{1:k-1}, y_{1:k-1})}{M(t_{1:k}, y_{1:k})} = M^{-1}(t_k, y_k \mid t_{1:k-1}, y_{1:k-1})$$

Bayesian state estimation

The marginal likelihood recursion

Marginal likelihood recursion:

$$\frac{M(t_{1:k}, y_{1:k})}{M(t_{1:k-1}, y_{1:k-1})} = M^{-1}(t_{1:k-1}, y_{1:k-1}) \int L(t_{1:k}, x_{1:k})\pi(t_{0:k}, x_{0:k})\, \mathrm{d}x_{0:k}$$

$$= M^{-1}(t_{1:k-1}, y_{1:k-1})$$

$$\times \int g_{t_k}(y_k \mid x_k) f_{t_k, t_{k-1}}(x_k \mid x_{k-1}) L(t_{1:k-1}, x_{1:k-1})\pi(t_{0:k-1}, x_{0:k-1})\, \mathrm{d}x_{0:k}$$

$$= \int g_{t_k}(y_k \mid x_k) f_{t_k, t_{k-1}}(x_k \mid x_{k-1}) \gamma(t_{1:k-1}, x_{1:k-1} \mid t_{1:k-1}, y_{1:k-1})\, \mathrm{d}x_{0:k}$$

Bayesian state estimation

Filtering and prediction densities

- Filtering density at time $k-1$:

$$\gamma(t_{k-1}, x_{k-1} \mid t_{1:k-1}, y_{1:k-1}) \\ = \int \gamma(t_{1:k-1}, x_{1:k-1} \mid t_{1:k-1}, y_{1:k-1}) \, \mathrm{d}x_{0:k-2}$$

- Prediction density at time k:

$$\gamma(t_k, x_k \mid t_{1:k-1}, y_{1:k-1}) = \int \gamma(t_{1:k}, x_{1:k} \mid t_{1:k-1}, y_{1:k-1}) \, \mathrm{d}x_{0:k-1} \\ = \int f_{t_k, t_{k-1}}(x_k \mid x_{k-1}) \gamma(t_{1:k-1}, x_{1:k-1} \mid t_{1:k-1}, y_{1:k-1}) \, \mathrm{d}x_{0:k-1}$$

Bayesian state estimation
Back to the marginal likelihood

Marginal likelihood recursion again:

$$
\begin{aligned}
&\frac{M(t_{1:k}, y_{1:k})}{M(t_{1:k-1}, y_{1:k-1})} \\
&= \int g_{t_k}(y_k \mid x_k) f_{t_k, t_{k-1}}(x_k \mid x_{k-1}) \gamma(t_{1:k-1}, x_{1:k-1} \mid t_{1:k-1}, y_{1:k-1}) \, \mathrm{d}x_{0:k} \\
&= \int g_{t_k}(y_k \mid x_k) f_{t_k, t_{k-1}}(x_k \mid x_{k-1}) \gamma(t_{k-1}, x_{k-1} \mid t_{1:k-1}, y_{1:k-1}) \, \mathrm{d}x_{k-1:k} \\
&= \int g_{t_k}(y_k \mid x_k) \gamma(t_k, x_k \mid t_{1:k-1}, y_{1:k-1}) \, \mathrm{d}x_k \\
&= \frac{M(t_{1:k}, y_{1:k})}{M(t_{1:k-1}, y_{1:k-1})} \int g_{t_k}(y_k \mid x_k) \gamma(t_k, x_k \mid t_{1:k-1}, y_{1:k-1}) \, \mathrm{d}x_k
\end{aligned}
$$

Bayesian state estimation

Bayesian filtering

- Prediction:

$$\gamma(t_k, x_k \mid t_{1:k-1}, y_{1:k-1}) \\ = \int f_{t_k,t_{k-1}}(x_k \mid x_{k-1})\gamma(t_{k-1}, x_{k-1} \mid t_{1:k-1}, y_{1:k-1})\, \mathrm{d}x_{k-1}$$

- Marginal likelihood increment:

$$\frac{M(t_{1:k}, y_{1:k})}{M(t_{1:k-1}, y_{1:k-1})} = \int g_{t_k}(y_k \mid x_k)\gamma(t_k, x_k \mid t_{1:k-1}, y_{1:k-1})\, \mathrm{d}x_k$$

- Filter update:

$$\gamma(t_k, x_k \mid t_{1:k}, y_{1:k}) \propto g_{t_k}(y_k \mid x_k)\gamma(t_k, x_k \mid t_{1:k-1}, y_{1:k-1})$$

Bayesian state estimation

Woah?!

What we wanted:

- Recursive construction of the posterior

What we got:

- Recursive computation of the marginal likelihood
- Got filtering density (current best guess)

What is left:

- A *useful* representation of the posterior
- Posterior marginals:

$$\gamma(t_k, x_k \mid t_{1:n}, y_{1:n}), \quad t_k \leq t_n$$

Bayesian state estimation

Backward Markov representation I

Backward recursion for posterior marginals:

$$\gamma(t_k, x_k \mid t_{1:n}, y_{1:n}) = \int \gamma(t_k, x_k \mid t_{k+1:n}, x_{k+1:n}, t_{1:n}, y_{1:n}) \times \gamma(t_{k+1:n}, x_{k+1:n} \mid t_{1:n}, y_{1:n}) \, \mathrm{d}x_{k+1:n}$$

Markov property of prior implies Markov property of posterior:

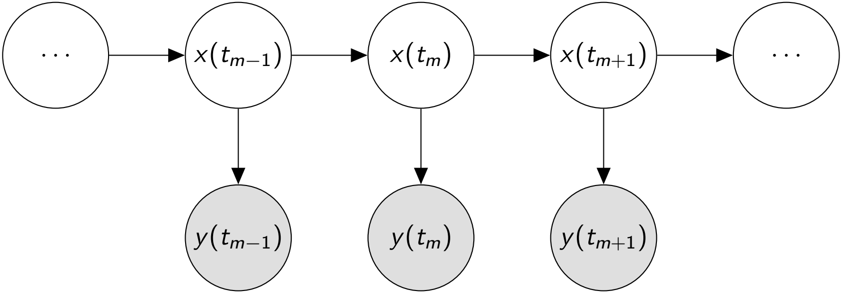
$$\gamma(t_k, x_k \mid t_{k+1:n}, x_{k+1:n}, t_{1:n}, y_{1:n}) = \gamma(t_k, x_k \mid t_{k+1}, x_{k+1}, t_{1:k}, y_{1:k})$$

Consequence:

$$\gamma(t_k, x_k \mid t_{1:n}, y_{1:n}) = \int \gamma(t_k, x_k \mid t_{k+1}, x_{k+1}, t_{1:k}, y_{1:k}) \times \gamma(t_{k+1}, x_{k+1} \mid t_{1:n}, y_{1:n}) \, \mathrm{d}x_{k+1}$$

Bayesian state estimation

Backward Markov representation II



Bayesian state estimation
The Backward kernel

The backward transition:

$$b_{t_k,t_{k+1}}(x_k \mid x_{k+1}) = \gamma(t_k, x_k \mid t_{k+1}, x_{k+1}, t_{1:k}, y_{1:k})$$

Bayes' rule:

$$\begin{aligned} b_{t_k,t_{k+1}}(x_k \mid x_{k+1}) &= \frac{\gamma(t_{k+1}, x_{k+1} \mid t_k, x_k, t_{1:k}, y_{1:k})\gamma(t_k, x_k \mid t_{1:k}, y_{1:k})}{\gamma(t_{k+1}, x_{k+1} \mid t_{1:k}, y_{1:k})} \\ &= \frac{f_{t_{k+1},t_k}(x_{k+1} \mid x_k)\gamma(t_k, x_k \mid t_{1:k}, y_{1:k})}{\gamma(t_{k+1}, x_{k+1} \mid t_{1:k}, y_{1:k})} \end{aligned}$$

Gauss–Markov regression

The beloved Gaussian case:

$$\begin{aligned} x(0) &\sim \mathcal{N}(\mu_0, \Sigma_0) \\ x(t) \mid x(s) &\sim \mathcal{N}(\Phi(t,s)x(s), Q(t,s)) \\ y(t) \mid x(t) &\sim \mathcal{N}(Cx(t), R) \end{aligned}$$

Gaussian models closed under marginalization/Bayes' rule – everything is Gaussian

Gaussian filtering

Filtering stuff:

$$\begin{aligned}\gamma(t_k, x_k \mid t_{1:k-1}, y_{1:k-1}) &= \mathcal{N}(x_k; \mu(t_k^-), \Sigma(t_k^-)) \\ \gamma(t_k, x_k \mid t_{1:k}, y_{1:k}) &= \mathcal{N}(x_k; \mu(t_k), \Sigma(t_k)) \\ \frac{M(t_{1:k}, y_{1:k})}{M(t_{1:k-1}, y_{1:k-1})} &= \mathcal{N}(y_k; \hat{y}(t_k), S(t_k))\end{aligned}$$

Gaussian filtering

The Kalman filter

- Prediction:

$$\mu(t_k^-) = \Phi(t_k, t_{k-1})\mu(t_{k-1})$$
$$\Sigma(t_k^-) = \Phi(t_k, t_{k-1})\Sigma(t_{k-1})\Phi^*(t_k, t_{k-1}) + Q(t_k, t_{k-1})$$

- Update:

$$\begin{aligned} S(t_k) &= C\Sigma(t_k^-)C^* + R \\ \hat{y}(t_k) &= C\mu(t_k^-) \\ K(t_k) &= \Sigma(t_k^-)C^*S^{-1}(t_k) \\ \mu(t_k) &= \mu(t_k^-) + K(t_k)\Big(y_k - \hat{y}(t_k)\Big) \\ \Sigma(t_k) &= \Sigma(t_k^-) - K(t_k)S(t_k)K^*(t_k) \end{aligned}$$

Gaussian smoothing

Smoothing stuff:

$$\begin{aligned} b_{t_k,t_{k+1}}(x_k \mid x_{k+1}) &= \mathcal{N}(x_k; a_{t_k,t_{k+1}}(x_{k+1}), P(t_k, t_{k+1})) \\ \gamma(t_k, x_k \mid t_{1:n}, y_{1:n}) &= \mathcal{N}(x_k; \xi(t_k), \Lambda(t_k)) \end{aligned}$$

Gaussian smoothing

The Rauch–Tung–Striebel Smoother

- Backward kernel parameters:

$$
\begin{aligned}
G(t_k, t_{k+1}) &= \Sigma(t_k)\Phi^*(t_{k+1}, t_k)\Sigma^{-1}(t_{k+1}^-) \\
a_{t_k, t_{k+1}}(x) &= \mu(t_k) + G(t_k, t_{k+1})(x - \mu(t_{k+1}^-)) \\
P(t_k, t_{k+1}) &= \Sigma(t_k) - G(t_k, t_{k+1})\Sigma(t_{k+1}^-)G^*(t_k, t_{k+1})
\end{aligned}
$$

- Backward recursion:

$$
\begin{aligned}
\xi(t_k) &= \mu(t_k) + G(t_k, t_{k+1})(\xi(t_{k+1}) - \mu(t_{k+1}^-)) \\
\Lambda(t_k) &= G(t_k, t_{k+1})\Lambda(t_{k+1})G^*(t_k, t_{k+1}) + P(t_k, t_{k+1})
\end{aligned}
$$

Bayesian ODE solvers

Vector field:

$$v \colon \mathbb{R}^d \to \mathbb{R}^d$$

Differential equation:

$$\dot{y}(t) = v(y(t)), \quad y(0) = y_0, \quad t \in [0, T].$$

Unknown quantity:

$$y^\dagger(t) = \varphi_{t,0}(y_0; v)$$

State-space realizable priors

State-space realizable prior:

$$\begin{aligned} \mathrm{d}x(t) &= Ax(t)\,\mathrm{d}t + \sqrt{\kappa}B\,\mathrm{d}w(t), \quad x(0) = x_0^\dagger, \\ \mathrm{D}^m y(t) &= E_m x(t), \quad m = 0, \dots, \nu. \end{aligned}$$

Assumptions:

- $E_m x_0^\dagger = \mathrm{D}^m y^\dagger(0)$ because autodiff
- (A, B) is controllable
- (A, E_0) is observable

State-space realizable priors

The Gauss–Markov property

Gauss–Markov property:

$$x(t) \mid x(s) \sim \mathcal{N}(\Phi(t,s)x(s), Q_\kappa(t,s)), \quad t > s$$

Parameters:

$$\Phi(t,s) = e^{A(t-s)},$$
$$Q_\kappa(t,s) = \kappa \int_0^{t-s} e^{A\tau} B B^* e^{A^*\tau} \, \mathrm{d}\tau.$$

What State-space model?

The integrated Wiener processes

ν-times integrated Wiener process:

$$\mathrm{d}y^{(\nu)}(t) = \sqrt{\kappa}\,\mathrm{d}w(t)$$

Corresponds to Taylor polynomial + stochastic remainder:

$$y(t) = \sum_{m=0}^{\nu} y^{(m)}(0)\frac{t^m}{m!} + \sqrt{\kappa}\int_0^t \frac{(t-\tau)^\nu}{\nu!}\,\mathrm{d}w(\tau)$$

What State-space model?

The integrated Ornstein–Uhlenbeck processes

Semi-linear problem:

$$\mathrm{D}y(t) = v(y(t)) = Ly(t) + N(y(t))$$

Differentiate ν times:

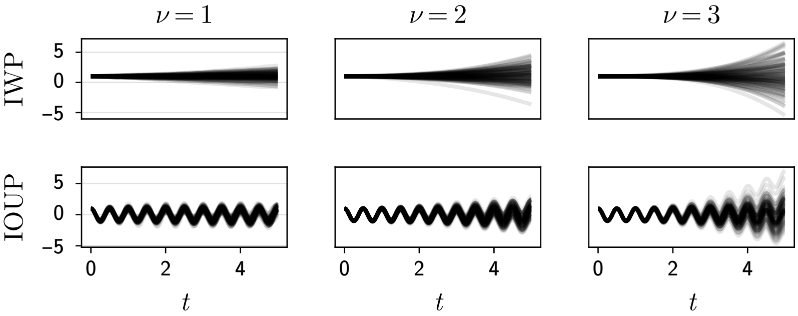
$$\mathrm{D}y^{(\nu+1)}(t) = Ly^{(\nu)}(t) + \mathrm{D}^{\nu} N(y(t))$$

Formally approximate $\mathrm{D}^{\nu} N(y(t))$ with white noise:

$$\mathrm{d}y^{(\nu)}(t) = Ly^{(\nu)}(t)\,\mathrm{d}t + \sqrt{\kappa}\,\mathrm{d}w(t)$$

What State-space model?

Illustration of priors



Solving as Bayesian state estimation

Inference problem:

$$
\begin{aligned}
x(t_{m+1}) \mid x(t_m) &\sim \mathcal{N}(\Phi(t_{m+1}, t_m)x(t_m), Q_\kappa(t_{m+1}, t_m)) \\
0 &= \dot{y}(t_m) - v(y(t_m)) = E_1 x(t_m) - v(E_0 x(t_m))
\end{aligned}
$$

Likelihood:

$$
g_{t_m}(y_m \mid x_m) = \mathcal{N}(0; E_1 x(t_m) - v(E_0 x(t_m)), 0)
$$

We *almost* know how to do this

Inference in nonlinear problems

(pretend its linear)

Zeroth order linearization:

$$v_0(t_m) = v(E_0\mu(t_m^-)), \quad C(t_m) = E_1$$

First order linearization:

$$v_0(t_m) = v(E_0\mu(t_m^-)), \quad C(t_m) = E_1 - J(E_0\mu(t_m^-))E_0$$

Maximum a posteriori objective:

$$V(x(t_{1:n})) = \sum_{m=1}^{n} \left\| Q_\kappa^{-1/2}(t_m, t_{m-1})(x(t_m) - \Phi(t_m, t_{m-1})x(t_{m-1})) \right\|^2,$$

$$0 = E_1 x(t_m) - v(E_0 x(t_m)), \quad m = 1, \dots, n.$$

Inference in linear problems

Calibration

Marginal log-likelihood:

$$\log M(\kappa) = \sum_{m=1}^{n} \log \mathcal{N}(v_0; C\mu(t_m^-), S_\kappa(t_m)) \tag{11}$$

Result:

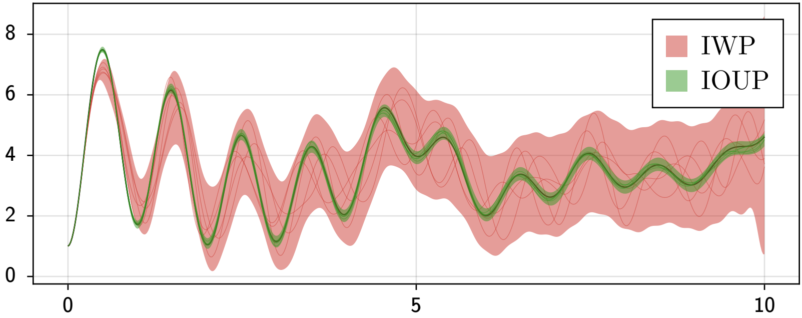
$$V_\kappa = \kappa V_1, \text{ for } V_\kappa \in \{Q_\kappa, \Sigma_\kappa, S_\kappa, P_\kappa, \Lambda_\kappa\} \tag{12}$$

Consequence:

$$\hat{\kappa} = \frac{1}{nd} \sum_{m=1}^{n} \|S_\kappa^{-1/2}(t_m)(v_0 - C\mu(t_m^-))\|^2 \tag{13}$$

Bayesian differential equation solvers

Some illustrations



Properties of solvers

- Stability?
- Convergence rates?

Stability of Bayesian solvers

A stability

Linear test equation:

$$\dot{y}(t) = Hy(t), \quad y(0) = y_0.$$

A-stability

A method is *A-stable* if, on a uniform grid,

$$y^\dagger(t) \to 0, \quad n \to \infty, \text{ implies } \hat{y}(t) \to 0, \quad n \to \infty.$$

Stability of Bayesian solvers

A stability of Bayesian solvers

Suppose (A, B) is controllable, then a Bayesian solver is A stable if and only if (Φ, C) is detectable.

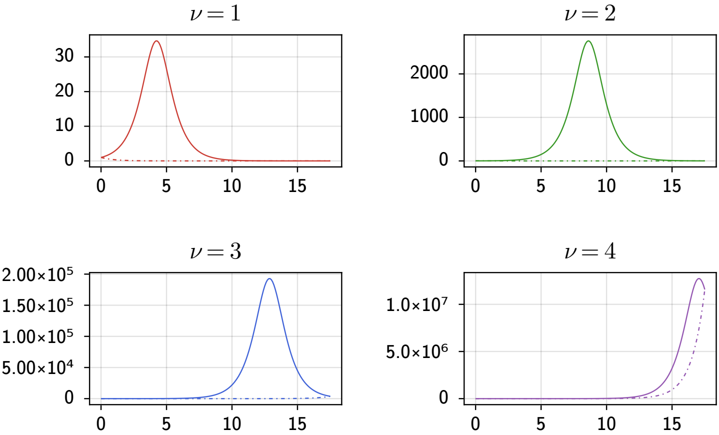
- (A, B) controllable implies $Q > 0$, hence stabilizability.
- Controllability of (A, B) is a property of the prior.
- Detectability of (Φ, C) is to some extent determined by method.

Note: (Φ, C) can be detectable even though H is not Hurwitz.

Stability of Bayesian solvers

Stable, all too stable?

$$\dot{y}(t) = y(t), \quad y(0) = 1.$$



Convergence of the MAP estimate

The setup

Solution estimate:

$$E_0\hat{x}(t) = \hat{y}(t) = y_0 + \int_0^t \dot{\hat{y}}(\tau)\,\mathrm{d}\tau$$
$$= y_0 + \int_0^t v(\hat{y}(\tau))\,\mathrm{d}\tau + \int_0^t Z[\hat{y}](\tau)\,\mathrm{d}\tau$$

Approximation of zero function:

$$Z[\hat{y}](t) = \dot{\hat{y}}(t) - v(\hat{y}(t))$$

Maximal interval length between zeros:

$$\delta = \max_{0 \le m \le n} |t_m - t_{m-1}|$$

Convergence of the MAP estimate
Proof ingredients

Sobolev interpolant of zero implies (Arcangéli et al. 2007):

$$\left|Z_i[\hat{y}]\right|_{\mathrm{H}_q^m} \le c_3 \delta^{\nu - m - (1/2 - 1/q)_+} \left|Z_i[\hat{y}]\right|_{\mathrm{H}_2^\nu}$$

Norm domination:

$$\|\hat{y}\|_{H^{\nu+1}} \le c_0 \|y^\dagger\|_{H^{\nu+1}}$$

Lipschitz property:

$$\begin{aligned}\left|Z_i[\hat{y}]\right|_{\mathrm{H}_2^\nu} &\le \|Z_i[\hat{y}]\|_{\mathrm{H}_2^\nu} \le \|Z_i[\hat{y}] - Z_i[y^\dagger]\|_{\mathrm{H}_2^\nu} \\ &\le c_2(y^\dagger, v_i) \|\hat{y} - y^\dagger\|_{H^{\nu+1}} \le c_3(y^\dagger, v_i) \|y^\dagger\|_{H^{\nu+1}}\end{aligned}$$

Convergence of the MAP estimate

Convergence rate

Let $v \in \mathscr{C}^{\nu+1}(\mathbb{R}^d, \mathbb{R}^d)$, and define e by

$$e[y](t) = \int_0^t Z[y](\tau)\, \mathrm{d}\tau.$$

Assume the unique solution, $y^\dagger$, exists up until $T^\dagger \geq T$, then there exists a positive constant $c_4(y^\dagger, \nu, f_i,)$

$$\begin{aligned} \left|e_i[\hat{y}]\right|_{\mathrm{H}_q^0} &\leq \delta^\nu T^{1/q} c_4(y^\dagger, \nu, f_i) \left\|y^\dagger\right\|_{\mathrm{H}_2^{\nu+1}} \\ \left|e_i[\hat{y}]\right|_{\mathrm{H}_q^m} &\leq \delta^{\nu+1-m-(1/2-1/q)_+} T^{1/q} c_4(y^\dagger, \nu, v_i) \left\|y^\dagger\right\|_{\mathrm{H}_2^{\nu+1}}, \end{aligned}$$

where $m = 1, \ldots, \nu$.

Convergence of the MAP estimate

What we need to do

Need to asusume:

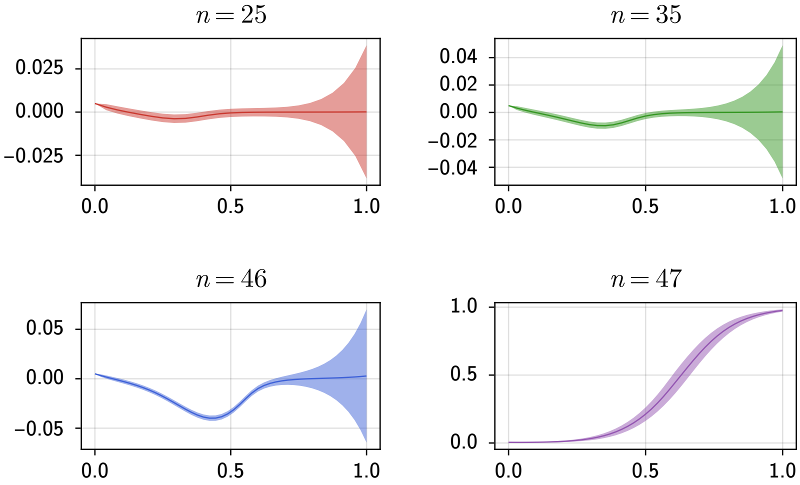
- $v \in \mathscr{C}^{\nu+1}(\mathbb{R}^d, \mathbb{R}^d)$
- $y^\dagger$ exists until $T^\dagger > T$

Need to establish:

- $y^\dagger \in \mathrm{H}_2^{\nu+1}$.
- $\hat{y} \in \mathrm{H}_2^{\nu+1}$ and that
- $\|\hat{y}\|_{H^{\nu+1}} \leq c_0 \|y^\dagger\|_{H^{\nu+1}}$
- $Z_i \colon \mathrm{H}^{\nu+1}([0, T], \mathbb{R}^d) \to \mathrm{H}^{\nu}([0, T], \mathbb{R})$ is locally Lipshitz (Valent 2013)

Convergence of the MAP estimate

Convergence only eventually



Some things to think about

- Which prior?
- How does the *actual* posterior behave?
- When is Gaussian approximation reasonable?
- How to implement Bayes' rule in practice? particle filters?