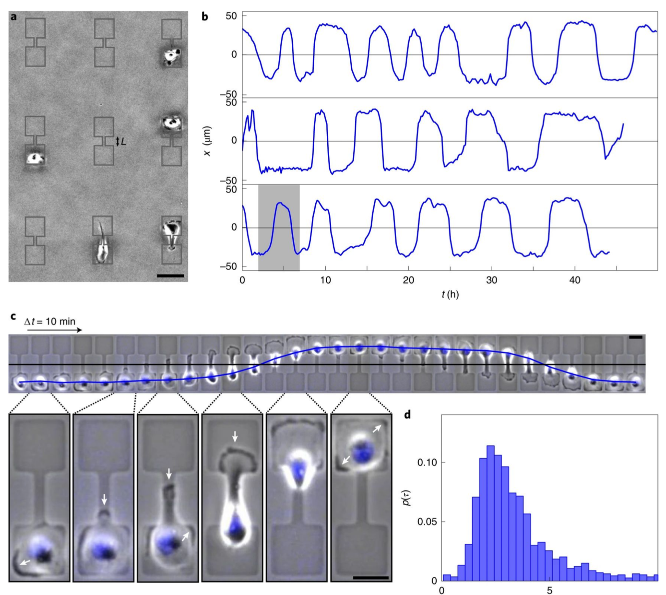
A Rigorous Approach to Stochastic Dynamics Inference from Tracked Data

Complex biological system
↓ Digital recording
Large data set
↓ Inference of mathematical model — Inverse problem
Detailed mathematical model
↓ Low-dimensional approximation — Forward problem
Analytically and/or numerically tractable model

Inference of the dynamical processes from the tracked data

Motivation

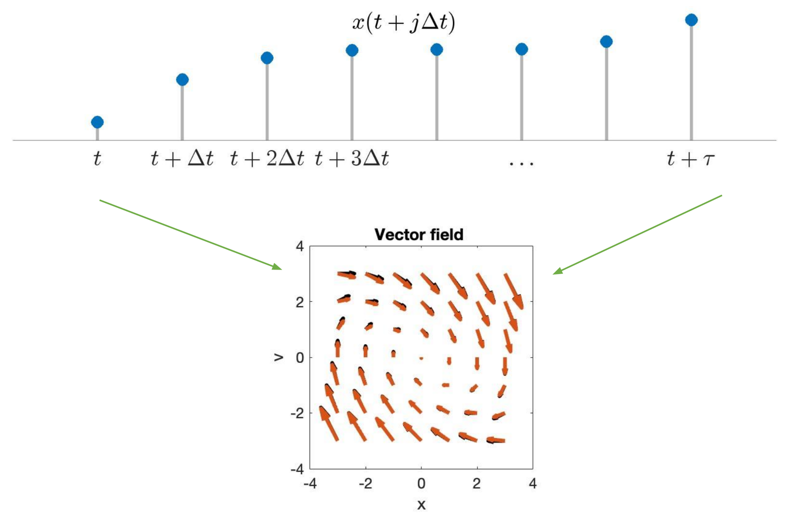


[*Stochastic nonlinear dynamics of confined cell migration in two-state systems*, D Brückner, et al. Nature Phys. 2019]

Formulation of the problem

$$dx = v dt$$
$$dv = F(x,v)dt + \sigma(x,v)d\xi(t)$$

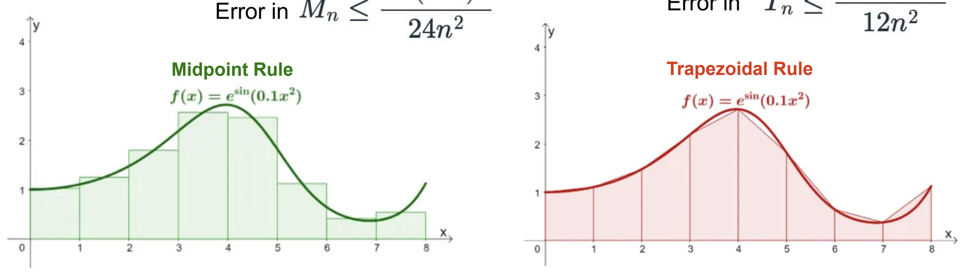
Infer F(x,v) and σ(x,v) from n points sampled at intervals Δt



Simple analogy – numerical quadrature

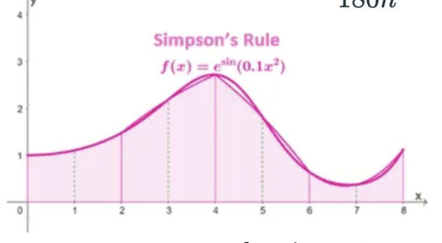
Approximate $\int_a^b f(x)\,dx$ from discrete data:

Error in $M_n \le \frac{M(b-a)^3}{24n^2}$ Error in $T_n \le \frac{M(b-a)^3}{12n^2}$



Error in $S_n \le \frac{M(b-a)^5}{180n^4}$

- same data
- all linear combinations
- different precision



www.storyofmathematics.com

ULI method (Underdamped Langevin Inference)

Project force & noise functions to a finite basis: $b = \{b_\alpha(x,v)\},\ 1 \le \alpha \le n_b$

Orthonormalization of basis functions:

$$c_\alpha(x,v) = \sum_{\beta=1}^{n_b} B_{\alpha\beta}^{-1/2} b_\beta(x,v), \quad \text{where} \quad B_{\alpha,\beta} = \langle b_\alpha(x,v)b_\beta(x,v)\rangle = \int b_\alpha(x,v)b_\beta(x,v)P(x,v)\,dxdv$$

↑ stationary probabilistic measure

Coefficients of the basis functions (low-dimensional projections):

$$F(x,v) \approx \sum_{\alpha=1}^{n_b} F_\alpha c_\alpha(x,v), \quad \sigma^2(x,v) \approx \sum_{\alpha=1}^{n_b} \sigma_\alpha^2 c_\alpha(x,v),$$
$$F_\alpha = \langle F(x,v)c_\alpha(x,v)\rangle = \int F(x,v)c_\alpha(x,v)P(x,v)\,dxdv,$$
$$\sigma_\alpha^2 = \langle \sigma^2(x,v)c_\alpha(x,v)\rangle = \int \sigma^2(x,v)c_\alpha(x,v)P(x,v)\,dxdv.$$

Challenges:

- x – observed, v – unobserved, loss of Markovian property, $P(x,v)$ – unobserved
- Typically a single trajectory of a limited length - small data , sparse sampling times
- Additional noise from measurement error distorting data

Discretization of integrals: Integrals through P replaced by time averages

$$\hat{c}_\alpha(x,v) = \sum_{\beta=1}^{n_b} \hat{B}_{\alpha\beta}^{-1/2} b_\beta(x,v), \quad \text{where} \quad \hat{B}_{\alpha\beta} = \frac{\Delta t}{\tau}\sum_t b_\alpha(x(t),v(t))b_\beta(x(t),v(t)).$$

Discretization of derivatives:

$$\Delta x^{(n)}(t) = x(t+n\Delta t) - x(t)$$
$$\hat{v}(t) = \frac{1}{\Delta t}\left[\beta\Delta x^{(2)} + \gamma\Delta x^{(1)}\right]$$
$$\hat{a}(t) = \frac{1}{\Delta t^2}\left[B\Delta x^{(2)} + C\Delta x^{(1)}\right]$$

- Brückner et al. considered: $\beta = \lambda/2$, $\gamma = 1-\lambda$, $B = 1$, $C = -1$
- What is the optimal λ?

How to obtain such method? Use estimator of the type: $\hat{F}_\alpha = \langle \hat{a}c_\alpha(x,\hat{v})\rangle$

- Requires multiple levels of stochastic expansions

$$\langle \hat{a}c_\alpha(x,\hat{v})\rangle = \langle F(x,v)c_\alpha(x,v)\rangle + \frac{1+2\lambda}{6}\left\langle \frac{\partial c_\alpha(x,v)}{\partial v}\sigma^2(x,v)\right\rangle + O(\Delta t)$$

Projection coefficient computed from data | Exact value of the projection coefficient | Corrections that may depend on stochastic terms $\lambda = -1/2$ → $\hat{v} = \frac{1}{\Delta t}\left[-\frac{1}{4}x(t+2\Delta t) + \frac{3}{2}x(t+\Delta t) - \frac{5}{4}x(t)\right]$

Inference method of Brückner et al.:

$$\hat{F}_\alpha = \langle \hat{a}c_\alpha(x,\hat{v})\rangle$$
$$\hat{v} = \frac{1}{\Delta t}\left[-\frac{1}{4}\Delta x^{(2)} + \frac{3}{2}\Delta x^{(1)}\right]$$
$$\hat{a} = \frac{1}{\Delta t^2}\left[\Delta x^{(2)} - \Delta x^{(1)}\right]$$

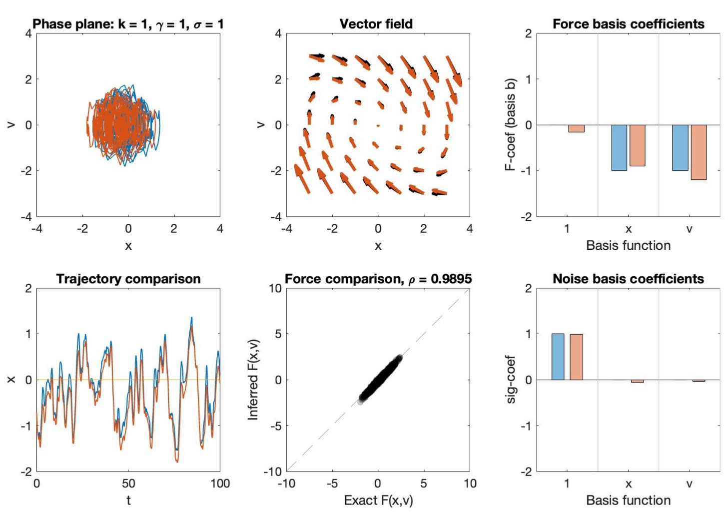
\+ measurement noise (magnitude Λ) →

$$\hat{F}_\alpha = \langle \hat{a}c_\alpha(x,\hat{v})\rangle - \frac{1}{2}\left\langle \frac{\partial c_\alpha(x,v)}{\partial v}\Big|_{v=\hat{v}}\hat{\sigma}^2(x,\hat{v})\right\rangle$$
$$\hat{v} = \frac{1}{2\Delta t}\Delta x^{(2)}$$
$$\hat{a} = \frac{1}{\Delta t^2}\left[\Delta x^{(2)} - \Delta x^{(1)}\right]$$

[*Inferring the dynamics of underdamped stochastic systems*, D Brückner, P Ronceray, CP Broedersz, PRL 2020]

Example: harmonic oscillator

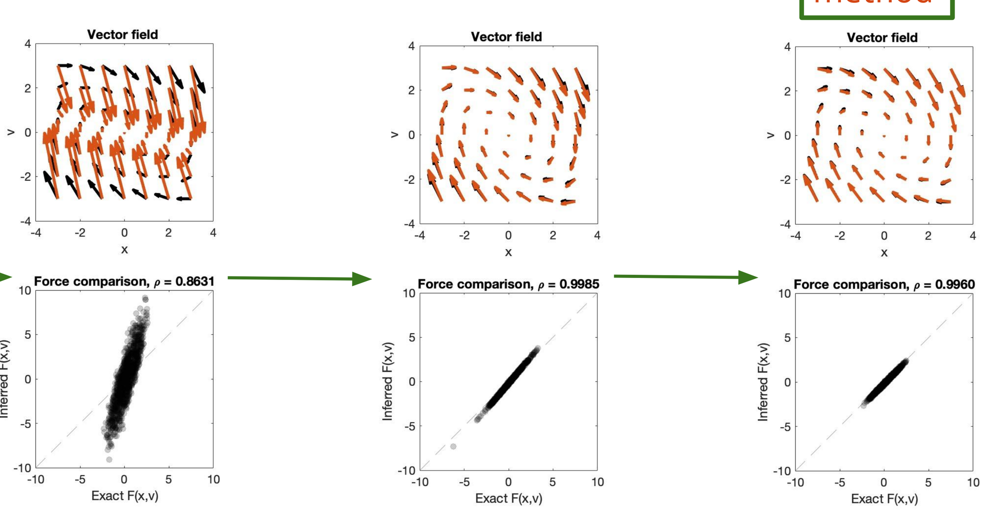
No measurement noise: $\lambda = -1/2$, $\Lambda = 0$
$\Delta t = 0.1$, $\tau = n\Delta t = 100$, $n = 1000$, $\gamma = \kappa = \sigma = 1$



\+ measurement noise ($\Lambda = 0.0004$)

$$x^\eta(t) = x(t) + \eta(t)$$
$$\langle \eta(t)\eta(t')\rangle = \Lambda\delta(t-t')$$

$$dx = v dt$$
$$dv = [-\gamma v - kx]dt + \sigma d\xi$$



Brückner et al. $\lambda = -1/2$	Brückner et al. $\lambda = 1$	Our method
Failure to converge for data with measurement noise	Ad hoc estimator offset Incompatibility w/ & w/o noise Force estimator depends on noise estimator	Systematic expansion No incompatibility Works for data with measurement noise

Main result: Improved method based on rigorous analysis

Extending stencil from 3 to 4 nodes

$$\hat{v}(t) = \frac{1}{\Delta t}\left[\alpha\Delta x^{(3)} + \beta\Delta x^{(2)} + \gamma\Delta x^{(1)}\right] \qquad \hat{a}(t) = \frac{1}{\Delta t^2}\left[A\Delta x^{(3)} + B\Delta x^{(2)} + C\Delta x^{(1)}\right]$$

Trajectory smoothing for inherent noise estimation

$$\bar{x}^\eta = ax^\eta(t+3\Delta t) + bx^\eta(t+2\Delta t) + cx^\eta(t+\Delta t) + dx^\eta(t)$$

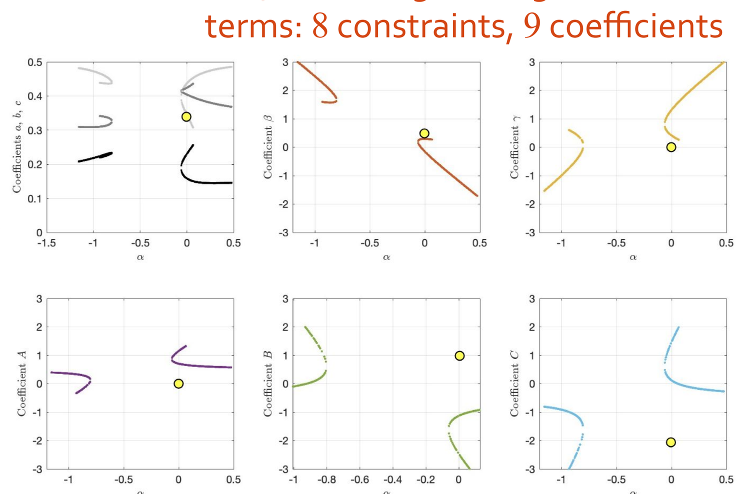
Selecting coefficients: cancellation of the error terms in presence/absence of measurement error

For Bruckner et al. requires λ=1 (inconsistency) | For Bruckner et al. a=0, b=c=d=1/3 (equal weights) | Does not cancel in Bruckner et al. (not enough parameters)

$$\hat{A}_\alpha^\eta = \underbrace{\langle \hat{a}c_\alpha(\bar{x},\hat{v})\rangle}_{\hat{A}_\alpha} + \frac{1}{\Delta t^3}\frac{\partial c_\alpha}{\partial v}\left\langle f^{(a)}f^{(v)}\right\rangle + \frac{1}{\Delta t^2}\frac{\partial c_\alpha}{\partial x}\left\langle f^{(a)}f^{(x)}\right\rangle + \frac{\hat{a}}{\Delta t}\frac{\partial^2 c_\alpha}{\partial x\partial v}\left\langle f^{(x)}f^{(v)}\right\rangle$$
$$+ \frac{1}{2\Delta t}\hat{a}\frac{\partial^2 c_\alpha}{\partial v^2}\left\langle {f^{(v)}}^2\right\rangle + \frac{1}{2}\hat{a}\frac{\partial^2 c_\alpha}{\partial x^2}\left\langle {f^{(x)}}^2\right\rangle + \cdots$$

additional conditions
more constraints

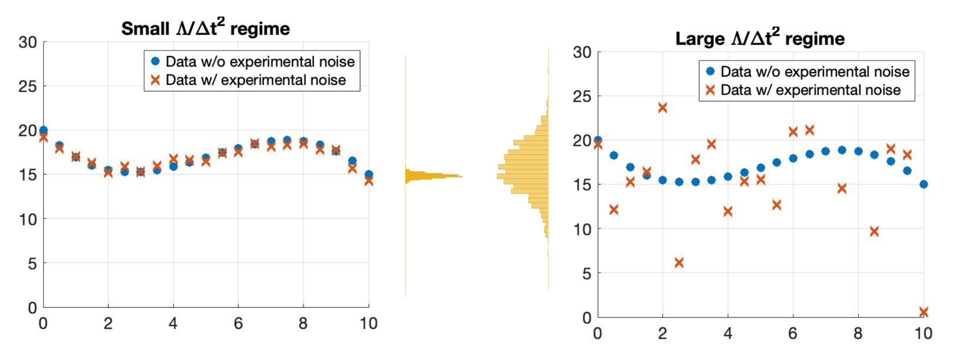
Coefficients, cancelling leading order error terms: 8 constraints, 9 coefficients



Small parameters: $C_x\Lambda \ll \frac{C_v\Lambda}{\Delta t} \ll \frac{C_a\Lambda}{\Delta t^2} < 1$

Tradeoff: magnitude of the measurement noise and step size

- Smaller Δt when no measurement = better approximation
- But Δt cannot be too low when measurement noise is present
- Magnitude of the measurement noise cannot be too large



FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
Univerzita Komenského v Bratislave

MATFYZ CONNECTIONS

Katarína Boďová, Richard Kollár
Comenius University Bratislava, Slovakia

Acknowledgements:
D. Brückner[1], G. Tkačik[2]
[1] University of Basel, Basel, Switzerland, [2] ISTA, Austria

This work was funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04- 00037