AReS & MaRS - Adversarial and MMD-Minimizing Regression for SDEs

Gabriele Abbati*[1], Philippe Wenk*[23], Michael A Osborne[1], Andreas Krause[2], Bernhard Schölkopf[4], Stefan Bauer[4]

[1]University of Oxford, [2]ETH Zürich, [3]Max Planck ETH Center for Learning Systems, [4]Max Planck Institute for Intelligent Systems

Abstract

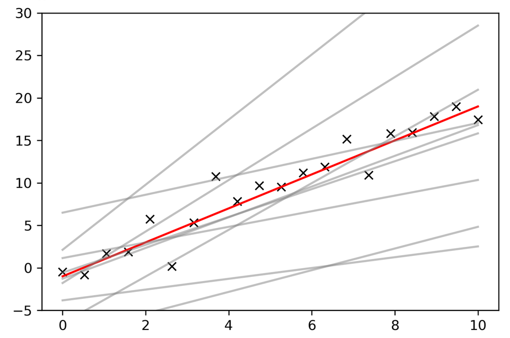
Stochastic differential equations (SDEs) are an important modeling class in many disciplines.
We propose a **novel, probabilistic model** for estimating the drift and diffusion given noisy observations of the underlying stochastic system.
Using state-of-the-art **adversarial** and **moment matching** inference techniques, we avoid the classical expensive discretization schemes.
We achieve significant **improvements** in parameter inference **accuracy** and robustness given random initial guesses.

Gradient Matching

ODE $\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \\ \mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon} \quad \text{with } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_{\mathbf{y}}) \end{cases}$

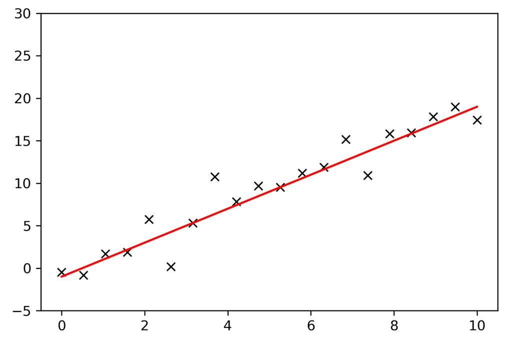
Given f and **y**, **infer x** and $\boldsymbol{\theta}$

Integration-based methods



parameters → trajectory

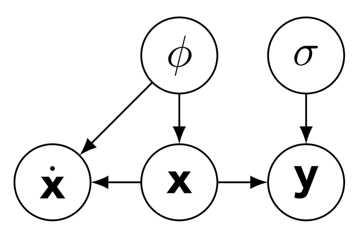
Integration-free methods



trajectory → parameters

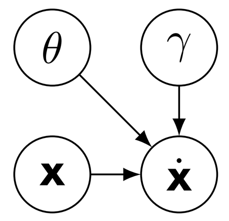
Models

(1) Gaussian Process prior on states



$$p(\mathbf{x} \mid \boldsymbol{\phi}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_{\mathbf{y}}, \mathbf{C}_{\boldsymbol{\phi}})$$
$$p(\dot{\mathbf{x}} \mid \mathbf{x}, \boldsymbol{\phi}) = \mathcal{N}(\dot{\mathbf{x}} \mid \mathbf{D}\mathbf{x}, \mathbf{A})$$

(2) ODE Model



$$p(\dot{\mathbf{x}} \mid \mathbf{x}, \boldsymbol{\theta}, \gamma) = \mathcal{N}(\dot{\mathbf{x}} \mid \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}), \gamma \mathbf{I})$$

(3) Observational Error $p(\mathbf{y} \mid \mathbf{x}, \sigma) = \mathcal{N}(\mathbf{y} \mid \mathbf{x}, \sigma^2 \mathbf{I})$

Inference

- Calderhead *et al.*[1], Dondelinger *et al.*[2]: **Product of Experts**
  $$p(\dot{\mathbf{x}}) \propto p_{\text{data}}(\dot{\mathbf{x}}) p_{\text{ODE}}(\dot{\mathbf{x}})$$
- Wenk *et al.*[3]: **Forced equality**
  $$p(\dot{\mathbf{x}}) \propto p_{\text{data}}(\dot{\mathbf{x}}_{\text{data}}) p_{ODE}(\dot{\mathbf{x}}_{\text{ODE}}) \delta(\dot{\mathbf{x}}_{\text{data}} - \dot{\mathbf{x}}) \delta(\dot{\mathbf{x}}_{\text{ODE}} - \dot{\mathbf{x}})$$
- Wenk, Abbati *et al.*[4]: **ODEs as constraints**

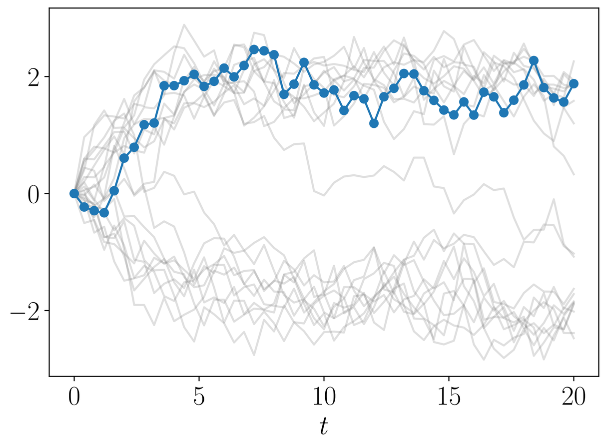
Stochastic Gradient Matching

General SDE Problem $\begin{cases} d\mathbf{x} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) dt + \mathbf{G} d\mathbf{W} \\ \mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon} \quad \text{with } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_{\mathbf{y}}) \end{cases}$

Given f and **y**, **infer x**, **G** and $\boldsymbol{\theta}$

Problems:

- Both **observation** and **process** noise
- **Stochastic** sample paths
- Paths are **not differentiable**



Doss-Sussmann Transformation

Definition (Ornstein-Uhlenbeck) *The Ornstein-Uhlenbeck process is the stochastic process* $\mathbf{o}$ *defined by the equation:* $d\mathbf{o} = -\mathbf{o}dt + \mathbf{G}d\mathbf{W}$.
We introduce the latent variable

$$\mathbf{z} = \mathbf{x} - \mathbf{o}$$

to get the **stochastic gradients**

$$d\mathbf{z}(t) = \{\mathbf{f}(\mathbf{z}(t) + \mathbf{o}(t), \boldsymbol{\theta}) + \mathbf{o}(t)\} \, dt$$

A Novel Noise Model

Previous Error Model $\mathbf{Y} = \mathbf{X} + \mathbf{E}$ → **New Error Model** $\mathbf{Y} = \mathbf{Z} + \mathbf{O} + \mathbf{E}$

Resulting observation marginal distribution:

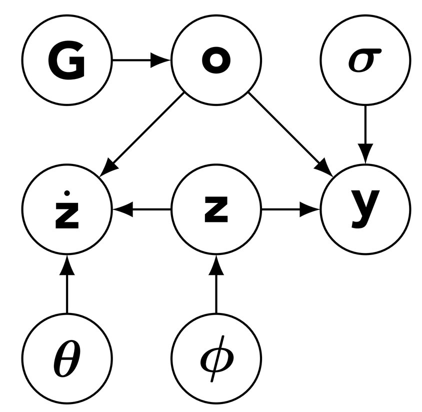
$$p(\tilde{\mathbf{y}} \mid \boldsymbol{\phi}, \mathbf{G}, \boldsymbol{\sigma}) = \mathcal{N}(\mathbf{0}, \mathbf{C}_{\boldsymbol{\phi}} + \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^T + \mathbf{T})$$

Gaussian prior $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\boldsymbol{\phi}})$

OU process $\mathbf{o} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^\top\right)$

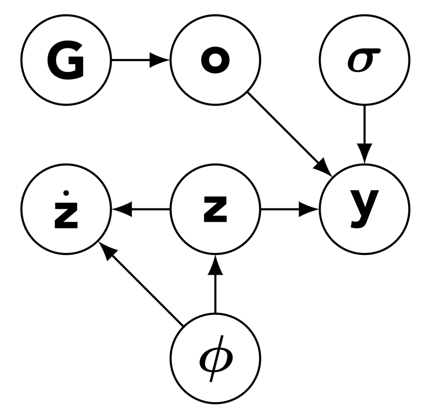
Obs. noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$

SDE-based model



$$p(\dot{\mathbf{z}} \mid \mathbf{o}, \mathbf{z}, \boldsymbol{\theta}) = \delta(\dot{\mathbf{z}} - \mathbf{f}(\mathbf{z} + \mathbf{o}, \boldsymbol{\theta}) - \mathbf{o})$$

Data-based model



$$p(\dot{\mathbf{z}} \mid \mathbf{z}, \boldsymbol{\phi}) = \mathcal{N}(\dot{\mathbf{z}} \mid \mathbf{D}\mathbf{z}, \mathbf{A})$$

AReS & MaRS: Inference and Results

GP fit
(ZOE noise model)

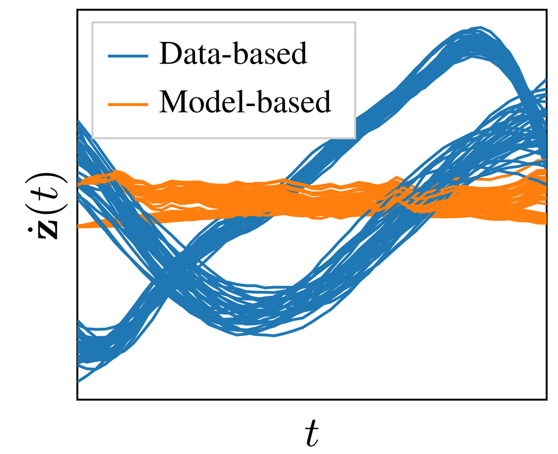
Samples from p_{SDE}
$\dot{\mathbf{z}}_{\text{SDE}} \sim p(\dot{\mathbf{z}} \mid \mathbf{o}, \mathbf{z}, \boldsymbol{\theta})$

Samples from p_{data}
$\dot{\mathbf{z}}_{\text{data}} \sim p(\dot{\mathbf{z}} \mid \mathbf{z}, \boldsymbol{\phi})$

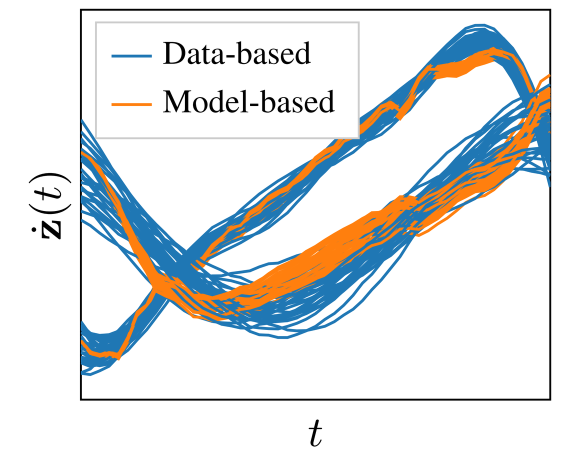
Iterative
Gradient-based
optimization

(1) AReS (WGAN[5]) $\boldsymbol{\theta} \leftarrow -\nabla_{\boldsymbol{\theta}} \left[\frac{1}{M} \sum_{i=1}^{M} f_\omega(\dot{\mathbf{z}}^{(i)}_{\text{SDE}})\right]$

(2) MaRS (MMD[6]) $\boldsymbol{\theta} \leftarrow \nabla_{\boldsymbol{\theta}} MMD_u^2 \left[\dot{\mathbf{z}}_{\text{SDE}}, \dot{\mathbf{z}}_{\text{data}}\right]$



Samples before training



Samples after training

DW, GT	NPSDE[7]	VGPA[8]	ESGF[9]	AReS	MaRS
$\theta_0 = 0.1$	0.09 ± 7.00	0.05 ± 0.04	0.01 ± 0.03	0.09 ± 0.04	**0.10 ± 0.05**
$\theta_1 = 4$	3.36 ± 248.82	1.11 ± 0.66	0.11 ± 0.16	3.68 ± 1.34	**3.85 ± 1.10**
$H = 0.25$	0.00 ± 0.02	/	0.20 ± 0.05	**0.21 ± 0.09**	

LV, GT	NPSDE[7]	ESGF[9]	AReS	MaRS
$\theta_0 = 2$	1.58 ± 0.71	2.04 ± 0.09	2.36 ± 0.18	**2.00 ± 0.09**
$\theta_1 = 1$	0.74 ± 0.31	1.02 ± 0.05	1.18 ± 0.9	**1.00 ± 0.04**
$\theta_2 = 4$	2.26 ± 1.51	3.87 ± 0.59	**3.97 ± 0.63**	3.70 ± 0.51
$\theta_3 = 1$	0.49 ± 0.35	0.96 ± 0.14	**0.98 ± 0.18**	0.91 ± 0.14
$\mathbf{H}_{1,1} = 0.05$	/	0.01 ± 0.03	**0.03 ± 0.004**	
$\mathbf{H}_{1,2} = 0.03$	/	0.01 ± 0.01	**0.02 ± 0.01**	
$\mathbf{H}_{2,1} = 0.03$	/	0.01 ± 0.01	**0.02 ± 0.01**	
$\mathbf{H}_{2,2} = 0.09$	/	0.03 ± 0.02	**0.09 ± 0.03**	

References

[1] Calderhead et al., *Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes*, NIPS, 2009.
[2] Dondelinger et al., *ODE parameter inference using adaptive gradient matching with Gaussian processes*, AISTATS, 2013.
[3] Wenk et al., *Fast Gaussian Process Based Gradient Matching for Parameter Identification in Systems of Nonlinear ODEs*, AISTATS, 2019.
[4] Wenk*, Abbati* et al., *Odin: ODE-informed regression for parameter and state inference in time-continuous dynamical systems*, arXiv preprint 1902.06278, 2019.
[5] Arjovsky and Chintala, *Wasserstein gan*, arXiv preprint 1701.07875, 2017.
[6] Gretton et al., *A kernel two-sample test*, JMLR, 2012.
[7] Yildiz et al., *Learning stochastic differential equations with gaussian processes without gradient matching*, IEEE 28th International Workshop on Machine Learning for Signal Processing, 2012.
[8] Vrettas et al., *Variational mean-field algorithm for efficient inference in large systems of stochastic differential equations*, Physical Review E, 2015.
[9] Särkkä et al., *Posterior inference on parameters of stochastic differential equations via non-linear Gaussian filtering and adaptive MCMC*, Statistics and Computing, 2015.

Contacts
Gabriele Abbati
gabb@robots.ox.ac.uk
Philippe Wenk
wenkph@ethz.ch