Stochastic Force Inference: User-Oriented Guide

Anna Frishman and Pierre Ronceray

Principle of the method

Input. A time series $\{\mathbf{x}^i\}_{i=1\dots N_{\text{steps}}}$ at times $t_i = 0\dots\tau$ of d-dimensional data. We define $\Delta\mathbf{x}^i = \mathbf{x}^{i+1} - \mathbf{x}^i$, $\Delta t_i = t_{i+1} - t_i$, and the mid-interval positions $\mathbf{x}^{i+\frac{1}{2}} = (\mathbf{x}^{i+1} + \mathbf{x}^i)/2$. We write x_μ^i the μ-th component of $\mathbf{x}^i$.

Model. The input data is assumed to obey Brownian dynamics,

$$\dot{x}_\mu = F_\mu(\mathbf{x}) + \sqrt{2D(\mathbf{x})}_{\mu\nu}\xi_\nu + \partial_\nu D_{\mu\nu}(\mathbf{x})$$

The gray term reflects the presence of multiplicative noise, written in Itô [2]. The force field $F_\mu(\mathbf{x})$ and the diffusion tensor field $D_{\mu\nu}(\mathbf{x})$ are assumed to be time-independent. The mobility matrix is absorbed in this definition of the force.

Output. Stochastic Force Inference returns the inferred diffusion tensor field $\hat{D}_{\mu\nu}(\mathbf{x})$, force field $\hat{F}_\mu(\mathbf{x})$ and out-of-equilibrium velocity field $\hat{v}_\mu(\mathbf{x})$. These are obtained by linear regression of local estimators with a set of n_b fitting functions $\{b_\alpha(\mathbf{x})\}_{\alpha=1\dots n_b}$ (the "projection basis"). These estimates are robust to localization error on the $\mathbf{x}^i$.

Projection basis

Choice of the basis. A good projection basis should be smooth, adapted to the symmetries of the problem, and with a number of functions in agreement with the information available in the trajectory (see "Error estimate" sections). If an educated guess on the functional form of the field to infer is available, use it, otherwise use a generic basis:

- Polynomial basis, $b = \{1, x_\mu, x_\mu x_\nu, \dots\}$, up to order n.
  Pros: parameter-free, complete. Order 1 captures circulation.
  Cons: unbounded; at high order, weight concentrates on large $\mathbf{x}$.
- Fourier basis, $b = \{1, \cos\left(\sum_\mu \frac{2\pi k_\mu x_\mu}{L_\mu}\right), \sin\left(\sum_\mu \frac{2\pi k_\mu x_\mu}{L_\mu}\right)\}$, for integer vectors k_μ with $\sum_\mu k_\mu \le n$.
  Pros: complete, regular, higher orders capture fine features.
  Cons: width parameters L_μ must be specified.

Normalization. The inference methods use normalized fitting functions $\hat{c}_\alpha(\mathbf{x})$, obtained from the unnormalized functions $b_\alpha(\mathbf{x})$ through:

$$\hat{c}_\alpha(\mathbf{x}) = \hat{B}_{\alpha\beta}^{-1/2} b_\beta(\mathbf{x}) \qquad \text{with} \qquad \hat{B}_{\alpha\beta} = \sum_i \frac{\Delta t_i}{\tau} b_\alpha(\mathbf{x}^i)$$

Diffusion inference

- Pick and normalize a basis $\hat{c}_\alpha(\mathbf{x})$ (see "Projection basis"). To infer a space-independent diffusion tensor choose $b = \{1\}$.
- Compute the diffusion projection coefficients:

$$\hat{D}_{\mu\nu\alpha} = \frac{1}{\tau}\sum_i \hat{c}_\alpha(\mathbf{x}^i)\left[\Delta x_\mu^i \Delta x_\nu^i + \Delta x_\mu^i \Delta x_\nu^{i+1} + \Delta x_\mu^{i+1}\Delta x_\nu^i\right]$$

  Note: the gray terms correct for biases induced by localization error [3]. For short, clean trajectories, omitting reduces inference error by a factor 4.
- The reconstructed diffusion tensor field is:

$$\hat{D}_{\mu\nu}(\mathbf{x}) = \hat{D}_{\mu\nu\alpha} c_\alpha(\mathbf{x})$$

Caveat: The reconstructed diffusion field $\hat{D}_{\mu\nu}(\mathbf{x})$ is not guaranteed to be positive definite at all points in space. If the spatial variations are strong, further regularization may be needed to ensure it does not get negative (this is important to infer the information and entropy production, which require inverting $\mathbf{D}$).

Error estimate. The relative error on the projection coefficients is

$$\frac{\delta\hat{D}^2}{\hat{D}^2} \sim \frac{4dn_b}{N_{\text{steps}}}$$

with n_b the number of basis functions and N_{steps} the number of time steps.

References

All these results, with additional details and proofs, are presented in Ref. [1]. A Python package implementing this method is available at xxx.

[1] Anna Frishman and Pierre Ronceray.
Learning force fields from stochastic trajectories.
arXiv:1809.09650, September 2018.

[2] A. W. C. Lau and T. C. Lubensky.
State-dependent diffusion: Thermodynamic consistency and its path integral formulation.
Phys. Rev. E, 76(1):011123, July 2007.

[3] Christian L. Vestergaard, Paul C. Blainey, and Henrik Flyvbjerg.
Optimal estimation of diffusion coefficients from single-particle trajectories.
Phys. Rev. E, 89(2):022726, February 2014.

Force inference

- Infer the diffusion field $\hat{D}_{\mu\nu}(\mathbf{x})$ (possibly constant).
- Pick and normalize a basis $\hat{c}_\alpha(\mathbf{x})$ (see "Projection basis"; it can be different from the one used for diffusion).
- Compute the out-of-equilibrium velocity projection:

$$\hat{v}_{\mu\alpha} = \frac{1}{\tau}\sum_i \Delta x_\mu^i \hat{c}_\alpha(\mathbf{x}^{i+\frac{1}{2}})$$

- The force projection coefficients are:

$$\hat{F}_{\mu\alpha} = \hat{v}_{\mu\alpha} - \sum_i \frac{\Delta t_i}{\tau}\partial_\nu[\hat{D}_{\mu\nu}\hat{c}_\alpha](\mathbf{x}^{i+\frac{1}{2}})$$

  Notes: the partial derivative acts on both the projection functions and the diffusion tensor (if it is spatially variable).
- The reconstructed force and velocity fields are:

$$\hat{F}_\mu(\mathbf{x}) = \hat{F}_{\mu\alpha} c_\alpha(\mathbf{x}) \qquad \hat{v}_\mu(\mathbf{x}) = \hat{v}_{\mu\alpha} c_\alpha(\mathbf{x})$$

Error estimate and overfitting criterion. The error on the projection coefficients is

$$\frac{\delta\hat{F}^2}{\hat{F}^2} \sim \frac{dn_b}{2\hat{I}_b} \quad \text{with} \quad \hat{I}_b = \frac{1}{4}\sum_i \hat{F}_\mu(\mathbf{x}^i)\hat{D}_{\mu\nu}^{-1}(\mathbf{x}^i)\hat{F}_\mu(\mathbf{x}^i)$$

Overfitting occurs if the number of functions $n_b > 2\hat{I}_b/d$.

This error estimate is rough in the case of spatially variable diffusion, and does not factor in the error in inferring $\mathbf{D}(\mathbf{x})$. It does not take localization error into account.

Entropy production

The estimator for the entropy produced along the trajectory is:

$$\Delta\hat{S}_b = \sum_i \hat{v}_\mu(\mathbf{x}^i)\hat{D}_{\mu\nu}^{-1}(\mathbf{x}^i)\hat{v}_\nu(\mathbf{x}^i)$$

with a absolute error of order $(2\Delta\hat{S}_b)^{1/2} + 2dn_b$. The error is uncontrolled in the case of multiplicative noise.