MAESTRO

Ludovic Goudenège[*]   Pierre Monmarché[†]   Benjamin Rotenberg[‡]

Hadrien Vroylandt[§]   Rodolphe Vuilleumier[¶]   Fabio Pietrucci[‖]

Abstract

MAESTRO

[*]CNRS, FR 3487, Fédération de Mathématiques de CentraleSupélec, CentraleSupélec, 91190 Gif-sur-Yvette, Université Paris-Saclay, France, goudenege@math.cnrs.fr

[†]Sorbonne Université, France, pierre.monmarche@upmc.fr

[‡]Sorbonne Université, France, benjamin.rotenberg@sorbonne-universite.fr

[§]Sorbonne Université, France, hadrien@hvroylandt.eu

[¶]Sorbonne Université, France, rodolphe.vuilleumier@ens.fr

[‖]Sorbonne Université, France, fabio.pietrucci@sorbonne-universite.fr

Introduction and main results

1 Derivation of the generalized Langevin equation

Let's consider a first order differential equation on the state $\boldsymbol{X}(t)$

$$\mathrm{d}\boldsymbol{X}(t) = \mathcal{L}\boldsymbol{X}(t)\mathrm{d}t \tag{1.1}$$

where $\mathcal{L}$ denote the evolution operator, that can be a Liouville operator for Hamiltonian dynamics, or a Markov operator for stochastic dynamics. We assume that we have only a partial view of the system trough a set of observables.

Most common choices of observables are either a subset of coordinates (like the coordinates of a particular particle) or coarse-grained function of the coordinates (like the center of mass of some of the particles).

1.1 Projection operator

To define projection, we need to introduce a scalar product, we have using equilibrium probability ρ_{eq}

$$\langle f, g \rangle = \int \mathrm{d}\boldsymbol{X} \rho_{eq}(\boldsymbol{X}) f^{\mathrm{T}}(\boldsymbol{X}) g(\boldsymbol{X}) \tag{1.2}$$

Orthogonal projection For orthogonal projection onto a set of m basis function $\{b\}$ (we don't really restrict m to be finite or event the family to be countable), we introduce the matrix $B \in \mathbb{R}^{m \times \cdot}$ such that each column of B is a basis function, then the orthogonal projection onto $\{b\}$ applied on a function $F(\boldsymbol{X})$ is

$$\mathcal{P}_{\{b\}}F = B \cdot \langle B, B \rangle^{-1} \cdot \langle B, F \rangle \tag{1.3}$$

Oblique projection We can also define an oblique projection[*] still using the same matrix B that is a basis for the range of the projector, but we also define the orthogonal complement D to the null space of the projector

$$\mathcal{P}^o_{\{b\}}F = B \cdot \langle D, B \rangle^{-1} \cdot \langle D, F \rangle \tag{1.4}$$

Examples of family basis

- $\{b\} = \{\mathcal{O}(\boldsymbol{X})\}$; we project onto the observable, that is the Mori projection.

- $\{b\} = \{\delta(\alpha - \mathcal{O}(\boldsymbol{X}))\}_{\alpha \in \mathbb{R}} = \mathrm{Span}(\delta(\alpha - \mathcal{O}(\boldsymbol{X})))$; we project onto the set of function of the observable, that is the Zwanzig projection[†]

- $\{b\} = \{b_k(\mathcal{O}(\boldsymbol{X}))\}_{k=1}^m$; we project onto a set of basis function b_k of the observable (Hermite polynomial, Legendre polynomial,...)

- $\{b\} = \{\{b_k(\mathcal{O}(\boldsymbol{X}))\}_{k=1}^m, \{h_j(\boldsymbol{X})\}\}$; we project onto a set of basis function of the observable and linear projection onto a set of hidden variable.

1.2 Mori-Zwanzig formalism

Let's denote the projection operator onto the set of basis family as $\mathcal{P}_{\{b\}}$, this operator have several properties

The evolution equation of our observables are given by

$$\frac{\partial \mathcal{O}(\boldsymbol{X}, t)}{\partial t} = \mathcal{L}\mathcal{O}(\boldsymbol{X}, t) = e^{t\mathcal{L}}\mathcal{L}\mathcal{O}(\boldsymbol{X}, 0) = e^{t\mathcal{L}}\mathcal{P}_{\{b\}}\mathcal{L}\mathcal{O}(\boldsymbol{X}, 0) + e^{t\mathcal{L}}(1 - \mathcal{P}_{\{b\}})\mathcal{L}\mathcal{O}(\boldsymbol{X}, 0) \tag{1.5}$$

---

[*]that is more or less what [?] are doing

[†]In this case, the family is not countable and the product part of the projector is $\int \mathrm{d}\alpha$

We have the operator identity (Duhamel-Dyson)

$$e^{t\mathcal{L}} = e^{t(1-\mathcal{P}_{\{b\}})\mathcal{L}} + \int \mathrm{d}s e^{(t-s)\mathcal{L}} \mathcal{P}_{\{b\}} \mathcal{L} e^{s(1-\mathcal{P}_{\{b\}})\mathcal{L}} \tag{1.6}$$

Both equations can be combined to get

$$\frac{\partial \mathcal{O}(\boldsymbol{X},t)}{\partial t} = e^{t\mathcal{L}} \mathcal{P}_{\{b\}} \mathcal{L} \mathcal{O}(\boldsymbol{X},0) + \int \mathrm{d}s e^{(t-s)\mathcal{L}} \mathcal{P}_{\{b\}} \mathcal{L} e^{s(1-\mathcal{P}_{\{b\}})\mathcal{L}} (1 - \mathcal{P}_{\{b\}}) \mathcal{L} \mathcal{O}(\boldsymbol{X},0) \tag{1.7}$$
$$+ e^{t(1-\mathcal{P}_{\{b\}})\mathcal{L}} (1 - \mathcal{P}_{\{b\}}) \mathcal{L} \mathcal{O}(\boldsymbol{X},0)$$

We define

$$F(\boldsymbol{X},t) = e^{t(1-\mathcal{P}_{\{b\}})\mathcal{L}} (1 - \mathcal{P}_{\{b\}}) \mathcal{L} \mathcal{O}(\boldsymbol{X},0) \tag{1.8}$$

that allow to simplify a bit the Eq. (1.7)

$$\frac{\partial \mathcal{O}(\boldsymbol{X},t)}{\partial t} = e^{t\mathcal{L}} \mathcal{P}_{\{b\}} \mathcal{L} \mathcal{O}(\boldsymbol{X},0) + \int \mathrm{d}s e^{(t-s)\mathcal{L}} \mathcal{P}_{\{b\}} \mathcal{L} F(\boldsymbol{X},s) + F(\boldsymbol{X},t) \tag{1.9}$$

1.3 Hamiltonian case

The global evolution of the system is given by the Hamilton's equations

$$\begin{cases} \dot{\boldsymbol{q}} = \boldsymbol{M}^{-1} \boldsymbol{p} \\ \dot{\boldsymbol{p}} = -\boldsymbol{\nabla} U(\boldsymbol{q}) \end{cases} \tag{1.10}$$

We choose as observable a subset of positions and velocities $\mathcal{O}(\boldsymbol{X},t) = (\delta(x - \boldsymbol{X}), \delta(p - \boldsymbol{X}))$.[‡]

$$\begin{cases} \dot{x} = M^{-1} \boldsymbol{p} \\ \dot{\boldsymbol{p}} = F_{\mathrm{eff}}(x) + \int_0^t K(\boldsymbol{p}(\tau), t-\tau) \mathrm{d}\tau + R(t) \end{cases} \tag{1.11}$$

1.4 Langevin case

The main difference with Hamiltonian case is the presence of an extra heat bath on observed variable.

1.5 Second Fluctuation-Dissipation Theorem

2 Short memory and auxiliary variable approach

$$\begin{cases} \dot{q} &= M^{-1} p \\ \dot{p} &= F(q) - \boldsymbol{A}_{ph} \boldsymbol{h} - A_{pp} p + \Sigma_p \xi(t) \\ \dot{\boldsymbol{h}} &= -\boldsymbol{A}_{hh} \boldsymbol{h} - \boldsymbol{A}_{hp} p + \boldsymbol{\Sigma}_h \boldsymbol{W}(t) \end{cases} \tag{2.1}$$

2.1 Fluctuation-Dissipation Theorem

2.2 Equivalence of GLE and extended Markov system

2.3 Alternative derivation via Markovian approximation

Can we use an extended basis directly from the beginning

2.4 Path probability for extended system

For the extended system, the transition probability follow from the gaussian nature of the noise

---

[‡]The x part of the projection is exact is the function p is included into the basis

Transition probability First we discretize evolution of (3.4) following a ABOBA scheme, *i.e.* position-Verlet instead of velocity-Verlet that would be OBABO.

A-step Generator for the A-step is $\mathcal{L}_A = v \, \nabla_x$ leading to propagator

$$x_{t+\Delta t} = x_t + \Delta t v_t \tag{2.2}$$

B-step Generator for the B-step is $\mathcal{L}_B = F(x) \cdot \nabla_v$ leading to propagator

$$v_{t+\Delta t} = v_t + \Delta t F(x_t) \tag{2.3}$$

O-step Generator for the O-step is $\mathcal{L}_O = -A \, \nabla_{v,z} + A \Delta_{v,z}$ leading to the (exact) propagator

$$\begin{pmatrix} v_{t+\Delta t} \\ z_{t+\Delta t} \end{pmatrix} = \exp\left[-\Delta t \boldsymbol{A}\right] \begin{pmatrix} v_t \\ z_t \end{pmatrix} + \boldsymbol{S} \cdot \boldsymbol{G} \tag{2.4}$$

where $\boldsymbol{G}$ is a vector of gaussian number of mean 0 and variance 1 and the matrix $\boldsymbol{S}$ is obtained trough

$$\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} = k_B T \left(\mathrm{Id} - \exp\left[-\Delta t \boldsymbol{A}\right] \exp\left[-\Delta t \boldsymbol{A}^{\mathrm{T}}\right] \right) \tag{2.5}$$

$$= k_B T \left(\begin{pmatrix} \mathrm{Id}_{vv} - e^{-\Delta t \boldsymbol{A}}{}_{vv} e^{-\Delta t \boldsymbol{A}^{\mathrm{T}}}{}_{vv} - e^{-\Delta t \boldsymbol{A}}{}_{vz} e^{-\Delta t \boldsymbol{A}^{\mathrm{T}}}{}_{vz} & -e^{-\Delta t \boldsymbol{A}}{}_{vv} e^{-\Delta t \boldsymbol{A}^{\mathrm{T}}}{}_{zv} - e^{-\Delta t \boldsymbol{A}}{}_{vz} e^{-\Delta t \boldsymbol{A}^{\mathrm{T}}}{}_{zz} \\ -e^{-\Delta t \boldsymbol{A}}{}_{zv} e^{-\Delta t \boldsymbol{A}^{\mathrm{T}}}{}_{vv} - e^{-\Delta t \boldsymbol{A}}{}_{zz} e^{-\Delta t \boldsymbol{A}^{\mathrm{T}}}{}_{vz} & \mathrm{Id}_{zz} - e^{-\Delta t \boldsymbol{A}}{}_{zz} e^{-\Delta t \boldsymbol{A}^{\mathrm{T}}}{}_{zz} - e^{-\Delta t \boldsymbol{A}}{}_{zv} e^{-\Delta t \boldsymbol{A}^{\mathrm{T}}}{}_{zv} \end{pmatrix} \right) \tag{2.6}$$

In case of a velocity dependent, we could take instead an Euler-Maruyama step with

$$\begin{pmatrix} v_{t+\Delta t} \\ z_{t+\Delta t} \end{pmatrix} = \begin{pmatrix} v_t \\ z_t \end{pmatrix} - \Delta t \boldsymbol{A} \begin{pmatrix} v_t \\ z_t \end{pmatrix} + \boldsymbol{S}_{\mathrm{Euler}} \cdot \boldsymbol{G} \tag{2.7}$$

with matrix $\boldsymbol{S}_{\mathrm{Euler}}$ now being

$$\boldsymbol{S}_{\mathrm{Euler}} \boldsymbol{S}_{\mathrm{Euler}}^{\mathrm{T}} = k_B T \Delta t \boldsymbol{A} \boldsymbol{A}^{\mathrm{T}} \tag{2.8}$$

ABOBA scheme This lead to the following scheme for one timestep evolution

$$x_{1/2} = x_0 + v_0 \Delta t/2 \qquad\qquad (A) \tag{2.9}$$
$$v_{1/2} = v_0 + F(x_{1/2}) \Delta t/2 \qquad\qquad (B) \tag{2.10}$$
$$v'_{1/2} = \exp\left[-\Delta t \boldsymbol{A}\right]_{vv} v_{1/2} + \exp\left[-\Delta t \boldsymbol{A}\right]_{vz} z_0 + [\boldsymbol{S} \cdot \boldsymbol{G}]_v \qquad\qquad (O) \tag{2.11}$$
$$\boldsymbol{z}_1 = \exp\left[-\Delta t \boldsymbol{A}\right]_{zv} v_{1/2} + \exp\left[-\Delta t \boldsymbol{A}\right]_{zz} z_0 + [\boldsymbol{S} \cdot \boldsymbol{G}]_z \qquad\qquad (O) \tag{2.12}$$
$$v_1 = v'_{1/2} + F(x_{1/2}) \Delta t/2 \qquad\qquad (B) \tag{2.13}$$
$$x_1 = x_{1/2} + v_1 \Delta t/2 \qquad\qquad (A) \tag{2.14}$$

That lead to the final transition as

$$\begin{pmatrix} x_1 \\ v_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{\Delta t}{2}(1 + \exp\left[-\Delta t \boldsymbol{A}\right]_{vv}) & \frac{\Delta t}{2} \exp\left[-\Delta t \boldsymbol{A}\right]_{vz} \\ 0 & \exp\left[-\Delta t \boldsymbol{A}\right]_{vv} & \exp\left[-\Delta t \boldsymbol{A}\right]_{vz} \\ 0 & \exp\left[-\Delta t \boldsymbol{A}\right]_{zv} & \exp\left[-\Delta t \boldsymbol{A}\right]_{zz} \end{pmatrix} \begin{pmatrix} x_0 \\ v_0 \\ z_0 \end{pmatrix}$$
$$+ \begin{pmatrix} (1 + \exp\left[-\Delta t \boldsymbol{A}\right]_{vv}) \Delta t^2/4 \\ (1 + \exp\left[-\Delta t \boldsymbol{A}\right]_{vv}) \Delta t/2 \\ \exp\left[-\Delta t \boldsymbol{A}\right]_{zv} \Delta t/2 \end{pmatrix} F\left(x_0 + v_0 \frac{\Delta t}{2}\right) \tag{2.15}$$
$$+ \begin{pmatrix} 0 & \frac{\Delta t}{2} \boldsymbol{S}_{vv} & \frac{\Delta t}{2} \boldsymbol{S}_{vz} \\ 0 & \boldsymbol{S}_{vv} & \boldsymbol{S}_{vz} \\ 0 & \boldsymbol{S}_{zv} & \boldsymbol{S}_{zz} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{G}_x \\ \boldsymbol{G}_v \\ \boldsymbol{G}_z \end{pmatrix}.$$

Then we have the gaussian probability transition from (x_t, v_t, z_t) to $(x_{t+\Delta t}, v_{t+\Delta t}, z_{t+\Delta t})$

$$\mathbb{P}\left((x_{t+\Delta t}, v_{t+\Delta t}, z_{t+\Delta t}|(x_t, v_t, z_t))\right) = \sqrt{(2\pi)^M \det \mathbf{\Sigma}_\Theta} \tag{2.16}$$

$$e^{-\frac{1}{2}\left[\begin{pmatrix} x_{t+\Delta t} \\ v_{t+\Delta t} \\ z_{t+\Delta t} \end{pmatrix} - M_\Theta \cdot \begin{pmatrix} x_t \\ v_t \\ z_t \end{pmatrix} - F_\Theta \begin{pmatrix} x_t \\ v_t \\ z_t \end{pmatrix} \mathbf{\Gamma}_\Theta\right]^{\mathrm{T}} \mathbf{\Sigma}_\Theta^{-1} \left[\begin{pmatrix} x_{t+\Delta t} \\ v_{t+\Delta t} \\ z_{t+\Delta t} \end{pmatrix} - M_\Theta \cdot \begin{pmatrix} x_t \\ v_t \\ z_t \end{pmatrix} - F_\Theta \begin{pmatrix} x_t \\ v_t \\ z_t \end{pmatrix} \mathbf{\Gamma}_\Theta\right]}$$

where we define

$$\mathbf{\Sigma}_\Theta = \begin{pmatrix} 0 & \dfrac{\Delta t}{2} \mathbf{S}_{vv} & \dfrac{\Delta t}{2} \mathbf{S}_{vz} \\ 0 & \mathbf{S}_{vv} & \mathbf{S}_{vz} \\ 0 & \mathbf{S}_{zv} & \mathbf{S}_{zz} \end{pmatrix} \cdot \begin{pmatrix} 0 & \dfrac{\Delta t}{2} \mathbf{S}_{vv} & \dfrac{\Delta t}{2} \mathbf{S}_{vz} \\ 0 & \mathbf{S}_{vv} & \mathbf{S}_{vz} \\ 0 & \mathbf{S}_{zv} & \mathbf{S}_{zz} \end{pmatrix}^{\mathrm{T}} \tag{2.17}$$

Initial probability We assume that the hidden degree of freedom are equilibrated, therefore the initial probability is

$$\mathbb{P}\left((x_0, v_0, z_0)\right) = \tag{2.18}$$

Complete path probability Given the Markovian nature of the model, the path probability to observe trajectory $\{(x_t, v_t, z_t)\}_0^T$ is given by

$$\mathbb{P}\left(\{(x_t, v_t, z_t)\}_0^T\right) = \tag{2.19}$$

that can be written under the form of an exponential family as

$$\mathbb{P}\left(\{(x_t, v_t, z_t)\}_0^T\right) = \tag{2.20}$$

Therefore the log-likelihood of the model is

$$-\log \mathcal{L} \tag{2.21}$$

3 Data-driven parametrization of GLE

3.1 Computation of memory kernel trough correlation function

For reference, see [**?**]

We start from the GLE

$$\dot{x}(t) = \boldsymbol{v}(t) M\dot{v}(t) = F_{\text{eff}}(x(t)) - \int_0^t K(t-\tau)v(\tau)\mathrm{d}\tau + R(t) \tag{3.1}$$

Since, the correlation between the initial velocity and the noise term is zero, $\langle v(0)R(t)\rangle = 0$, we get the Volterra equation between velocity correlation and force correlation

$$\langle v(0)(M\dot{v}(t) - F_{\text{eff}}(x(t)))\rangle = -\int_0^t K(t-\tau)\langle v(0)v(\tau)\rangle \, \mathrm{d}\tau \tag{3.2}$$

Iterative solution To solve the Volterra equation (3.2), we can discretize the integral using trapezoidal rule. Denoting $C_i^{vv} = \langle v(0)v(i\Delta t)\rangle$, and $C_i^{vf} = \langle v(0)[M\dot{v}(i\Delta t) - F_{\text{eff}}(x(i\Delta t))]\rangle$, and $K_i = K(i\Delta t)$, we get

$$K_i = -\frac{2}{\Delta t C_0^{vv}}\left(C_i^{vf} + \frac{\Delta t}{2}K_0 C_i^{vv} + \sum_{j=1}^{i-1}\Delta t K_j C_{i-j}^{vv}\right) \tag{3.3}$$

Laplace/Fourier transform The Volterra equation (3.2) can also be solved through Laplace or Fourier transform.

Other approaches   We can also transform the equation into a second order Volterra equation.
There exist more involved method to solve the Volterra integral equation, see collocation method

3.2   Hidden Markov approach

As the extended system is Markov, it is amendable to maximum likelihood approach. Due to the hidden nature of the auxillary variable, we turn to EM-type algorithm.

Parametrization of the system   We consider the following system of SDE.

$$\begin{cases} \dot{x} = M^{-1}p \\ \dot{p} = F(x) + A_{ph}\boldsymbol{h} - \gamma p + \Sigma_p \xi(t) \\ \dot{\boldsymbol{h}} = \boldsymbol{A}_{hh}\boldsymbol{h} + \boldsymbol{A}_{hp}p + \boldsymbol{\Sigma}_h \boldsymbol{W}(t) \end{cases} \tag{3.4}$$

We are given time series of corrected force $\{\dot{p} - F(q)\}_0^T$.

The parameters to compute are therefore only the dissipative part of the dynamics and the stochastic part is retrieved through the FDT

- A parametrization of the mean force over a chosen functionnal basis.

- A_{ph} and γ

- $\boldsymbol{A}_{hh}$ and $\boldsymbol{A}_{hp}$

Sufficient statistics   As our complete probabilistic model (2.21) is an exponential family, there is a complete set of sufficient statistic $T_1(\{(x_t, v_t, z_t)\}_0^T)$ and $T_2(\{(x_t, v_t, z_t)\}_0^T)$ (Pitman–Koopman–Darmois theorem) given by

$$T_1(\{(x_t, v_t, z_t)\}_0^T) = \mathbf{E}\left[\begin{pmatrix} x_{t+\Delta t} \\ v_{t+\Delta t} \\ z_{t+\Delta t} \end{pmatrix} - \begin{pmatrix} x_t \\ v_t \\ z_t \end{pmatrix} \middle| (x, v, z) \right] \tag{3.5}$$

$$T_2(\{(x_t, v_t, z_t)\}_0^T) = \mathbf{E}\left[\left(\begin{pmatrix} x_{t+\Delta t} \\ v_{t+\Delta t} \\ z_{t+\Delta t} \end{pmatrix} - \begin{pmatrix} x_t \\ v_t \\ z_t \end{pmatrix} \right) \left(\begin{pmatrix} x_{t+\Delta t} \\ v_{t+\Delta t} \\ z_{t+\Delta t} \end{pmatrix} - \begin{pmatrix} x_t \\ v_t \\ z_t \end{pmatrix} \right)^{\mathrm{T}} \right] \tag{3.6}$$

They are associated with natural parameters

$$\boldsymbol{\eta}_1(\boldsymbol{\Theta}, (x, v, z)) = \boldsymbol{\Sigma}_\Theta^{-1}\left((\boldsymbol{M}_\Theta - \mathrm{Id}) \cdot \begin{pmatrix} x \\ v \\ z \end{pmatrix} + F_\Theta \begin{pmatrix} x \\ v \\ z \end{pmatrix} \boldsymbol{\Gamma}_\Theta \right) \tag{3.7}$$

$$\boldsymbol{\eta}_2(\boldsymbol{\Theta}) = -\frac{1}{2}\boldsymbol{\Sigma}_\Theta^{-1} \tag{3.8}$$

and the log-partition function is

$$f(\Theta, (x, v, z)) = -\frac{1}{4}\int \mathrm{d}(x, v, z)\boldsymbol{\eta}_1(\boldsymbol{\Theta}, (x, v, z))^{\mathrm{T}}\boldsymbol{\eta}_2(\boldsymbol{\Theta})^{-1}\boldsymbol{\eta}_1(\boldsymbol{\Theta}, (x, v, z)) - \frac{1}{2}\log(-2\det \boldsymbol{\eta}_2(\boldsymbol{\Theta})) \tag{3.9}$$

Expectation step   The first step is computing the probability of the hidden variables given the parameters and the observables. That means we need $\mathbb{P}\left(\boldsymbol{h}_t | \{y_\tau\}_0^T, \boldsymbol{\Theta}_{k-1}\right)$[§]

Once we have $\mathbb{P}\left(\boldsymbol{h}_t | \{y_\tau\}_0^T, \boldsymbol{\Theta}_{k-1}\right)$, we can compute the sufficient statistics averaged over this probability distribution

Foward-Backward algorithm

[§]For a multivariate gaussian $p(x_A, x_B) = \mathcal{N}((\mu_A, \mu_B), \Sigma)$, the conditionnal gaussian is $p(x_A | x_B) = \mathcal{N}(\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(x_B - \mu_B), \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA})$

$$T_1^h(\{(x_t, v_t, z_t)\}_0^T) = \mathbf{E}\left[T_1(\{(x_t, v_t, z_t)\}_0^T)\right]_{\boldsymbol{h}_t | \{y_\tau\}_0^T, \boldsymbol{\Theta}_{k-1}} \tag{3.10}$$

$$T_2^h(\{(x_t, v_t, z_t)\}_0^T) = \mathbf{E}\left[T_2(\{(x_t, v_t, z_t)\}_0^T)\right]_{\boldsymbol{h}_t | \{y_\tau\}_0^T, \boldsymbol{\Theta}_{k-1}} \tag{3.11}$$

Maximisation step The derivative of the MLE are given by

$$TEST \tag{3.12}$$

The maximum likelihood estimators from transition probabilities (2.15) write

Stochastic approximation Given that we are unable to fully obtain the probability distribution of the hidden variable given the observed variables, we can instead sample the hidden variables and use Robbins–Monro algorithm or Polyak–Ruppert averaging. That is going to lead to the SAEM algorithm.

References

[1] Fokker, Planck and Langevin. A new equation. Nature, 1 (2020), 03K91L.