The Mori-Zwanzig Formalism and Stochastic Systems with Memory

Eric Darve

Department of Mechanical Engineering
Institute for Computational and Mathematical Engineering
Stanford University

April 11 2012
1 Motivation and goals
2 Mori-Zwanzig: expansion into means and fluctuation
3 Illustrative examples
4 Markov state models for computing reaction rates
Prediction with incomplete information

How accurately can we predict the evolution over time of some quantity with incomplete information?

This typically corresponds to a situation where we only have partial knowledge of the system, for example we only track certain variables, while the effect of the other variables is modeled or somehow approximated.
The Mori-Zwanzig approach can be derived in a general fashion without reference to a specific application.

We will focus on the general setting. From there, many different types of approximation methods can be derived for different applications.

We will present the discrete version, which is simpler to discuss.
Assume we have some evolution operator $M$. This operator may correspond to advancing the system forward in time.

For example we may have a dynamical system:

$$\frac{dx}{dt} = F(x(t))$$

Then:

$$Mx_0 = x(\tau|x_0, 0)$$

$$Mf(x_0) = f(x(\tau|x_0, 0))$$

Discrete dynamics (defined by $M$) is the natural setting if one uses a numerical integrator with time step $\tau$. 
The incomplete information is represented by a projection operator \( \mathcal{P} \) such that \( \mathcal{P}^2 = \mathcal{P} \).

This operator may correspond to an averaging operation, where the average is taken over the degrees of freedom that we are not observing.

In the optimal prediction interpretation, the averaging operation corresponds to a best guess, that is the one that does not produce any statistical bias given the information available.
Averaging operator over missing information

Assume now that we have some coarse grained variable $\xi(x)$ and we have some a priori knowledge about the equilibrium probability density function $\rho(x)$.

Take some field $f(x)$, we may define $\mathcal{P}$ as:

$$
\mathcal{P} f(x) = \frac{\int f(x') \delta(\xi(x') - \xi(x)) \rho(x') \, dx'}{\int \delta(\xi(x') - \xi(x)) \rho(x') \, dx'}
$$

We indeed have $\mathcal{P}^2 = \mathcal{P}$.

Moreover, $\mathcal{P} f$ is constant over the set $\{x' | \xi(x') = \xi(x)\}$, therefore we can think of $\mathcal{P} f$ as being a function of $\xi$ rather than $x$. 
Let us consider that we are tracking the evolution of $\xi$ in time or the evolution of a function of $\xi$.

Is it possible to predict $\xi(t + \Delta t)$ from the knowledge of $\xi(t)$ without knowing the full $x(t)$?

An exact prediction is not possible but perhaps we can find an equation that produces a trajectory $\xi(t)$ that has properties similar to the true trajectory.
In the Mori-Zwanzig approach, one strives to express the “unknown information” in terms of a quantity with zero mean.

If we only know $\xi(x)$, then we can estimate a function $f(x)$ using $P f(\xi)$.

The remainder is $(I - P) f$.

$(I - P) f$ has zero mean, that is $P (I - P) f = 0$. 


Expanding $\mathcal{M}^k$ using functions of $\xi$

Mori-Zwanzig works by simply decomposing the operators into their mean component and fluctuating component.

The mean is known and the fluctuations can be modeled, or even eliminated if one looks at averages.

The exact answer to the evolution problem is given by $\mathcal{M}^k$. This represents the state of the system at time $k\tau$.

Can we express $\mathcal{M}^k$ using an expansion in terms of functions of $\xi$ and a fluctuating term?
At each step, we decompose $M^k$ into mean and fluctuations.

Define:

$$L_0 = \mathcal{P}, \quad F_0 = \mathcal{I} - \mathcal{P}$$

The operator $L_0$ is a function of $\xi$.

Consider $D_1 = M - M\mathcal{P}$ and take its fluctuations.

Step 1:

$$F_1 = D_1 - \mathcal{P}D_1 = D_1 - L_1$$

$L_1$ is a function of $\xi$.

Then:

$$M = M\mathcal{P} + L_1 + F_1 = ML_0 + L_1 + F_1$$
Step 2

Advance by one step again:

\[ D_2 = M^2 - M^2L_0 - ML_1 \]

This is used to form an approximation of \( M^2 \):

\[ F_2 = D_2 - PD_2 = D_2 - L_2 \]

Again:

\[ M^2 = M^2L_0 + ML_1 + L_2 + F_2 \]
This process leads to the following recurrences:

\[ M^k = \sum_{l=0}^{k} M^{k-l} L_l + F_k \]

\[ D_k = M^k - \sum_{l=0}^{k-1} M^{k-l} L_l = M F_{k-1} \]

\[ L_k = \mathcal{P} D_k \quad F_k = D_k - L_k \]

By construction:

\[ \mathcal{P} F_k = 0 \]
Properties of this recurrence

\[ \mathcal{M}^k = \sum_{l=0}^{k} \mathcal{M}^{k-l} \mathcal{L}_l + \mathcal{F}_k \]

\[ e^{t\mathcal{L}} = e^{t\mathcal{L}} \mathcal{P} + \int_0^t e^{(t-s)\mathcal{L}} \mathcal{P} \mathcal{L} e^{sQ}(I - \mathcal{P}) ds + e^{tQ}(I - \mathcal{P}) \]

1. \( \mathcal{L}_l \) is a function of \( \xi \), the resolved variable.
2. \( \mathcal{F}_k \) is a fluctuating term: \( \mathcal{P} \mathcal{F}_k = 0 \).

If the system is sufficiently “chaotic,” it might be the case that at some point \( \mathcal{L}_k = \mathcal{P} \mathcal{D}_k \) is negligible. In that case:

\[ \mathcal{M}^k = \sum_{l=0}^{k_0} \mathcal{M}^{k-l} \mathcal{L}_l + \mathcal{F}_k \]
Example of application

This formula can be applied in many different ways. One needs to choose:

- What are the variables that are resolved? → choose $\mathcal{P}$
- What quantity are we trying to estimate? → choose function to which operators are applied.

For example:

$$\dot{\xi}(k\tau) = \sum_{l=0}^{k_0} \mathcal{L}_1(\xi((k - l)\tau)) + \mathcal{F}_k^\xi$$
Decay of $\mathcal{L}_k$

A more explicit expression can be obtained instead of the recurrences.

From the definition:

$$D_k = \mathcal{M}F_{k-1}$$

But then:

$$F_k = \mathcal{M}F_{k-1} - \mathcal{P}\mathcal{M}F_{k-1} = (I - \mathcal{P})\mathcal{M}F_{k-1} = [(I - \mathcal{P})\mathcal{M}]^k(I - \mathcal{P})$$

So:

$$\mathcal{L}_k = \mathcal{P}D_k = \mathcal{P}\mathcal{M}F_{k-1} = \mathcal{P}\mathcal{M}((I - \mathcal{P})\mathcal{M})^{k-1}(I - \mathcal{P})$$

$$= \mathcal{P}(\mathcal{M}(I - \mathcal{P}))^k$$

$$= \mathcal{P} [\mathcal{P}, \mathcal{M}] (\mathcal{M}(I - \mathcal{P}))^{k-1}$$

$$= \mathcal{P} [\mathcal{P}, [\mathcal{P}, [... , [\mathcal{P}, \mathcal{M}] \mathcal{M}] ... \mathcal{M}] \mathcal{M}]$$

$$= \mathcal{P} [\mathcal{P}, [\mathcal{P}, [... , \mathcal{M} \mathcal{M}] ... \mathcal{M}] \mathcal{M}]$$

$$= \mathcal{P} [\mathcal{P}, [\mathcal{P}, [... , \mathcal{M} \mathcal{M}]... \mathcal{M}] \mathcal{M}]$$
Under-resolved linear system

A simple example to illustrate these formulas. Consider:

\[ \xi_{k+1} = A \xi_k + B x_k \]
\[ x_{k+1} = a \xi_k + b x_k \]

Can we model the evolution of \( \xi_k \) in terms of the sequence \( \xi_l \) only?

Apply the previous expansion:

\[ \mathcal{F}_k = \begin{pmatrix} 0 & 0 \\ 0 & b^k \end{pmatrix} \quad \mathcal{L}_k = \begin{pmatrix} 0 & 0 \\ b^{k-1}a & 0 \end{pmatrix} \]

\[ \xi_{k+1} = A \xi_k + B \sum_{l=1}^{k} b^{l-1} a \xi_{k-l} + B b^k x_0 \]
Say that the linear operator $\mathcal{M}$ is constructed from a mass-spring system $X = (q, p)$:

$$\frac{dX}{dt} = KX$$

The eigenvalues of $K$ are all on the imaginary axis in the complex plane.

Split the system into resolved and unresolved variables where the resolved variables are associated with position and momentum of some particles.
With a finite step $\tau$, we need to consider $\exp(\tau K)$ as the operator. Although the eigenvalues of $\exp(\tau K)$ are on the unit circle, those of $b$ in general are not.

In fact it is possible to exhibit cases where the eigenvalues of $b$ are all strictly inside the unit circle so that $b^k \to 0$ when $k \to \infty$.

This is somewhat unexpected since in the time continuous case, the memory never decays. This is basically a consequence of the fact that all the eigenvalues of $K$ are on the imaginary axis. But this holds only in the time continuous case.

The system appears to “forget” about the initial conditions if we only observe it every step $\tau$. 
Eigenvalues of $\exp(\tau K)$

- $\ln(|\text{e-values}|)$; min = 0.123

 Eigenvalues of $b(\tau)$

Decay of $Bb^k x_0$
Eigenvalues of $\exp(\tau K)$

$-\ln(|\text{e-values}|); \text{min} = 0.023$

Eigenvalues of $b(\tau)$

Decay of $B b^k \chi_0$
Galerkin basis

- Numerically, we need a discrete approximation of functions of $\xi$.
- This can be obtained in different ways. A Galerkin approximation is possible.
- Consider the vector space spanned by $\phi_i(\xi)$. The projector $\mathcal{P}$ can be defined as:

$$
\mathcal{P} f(x) = \sum_i f_i \phi_i(\xi)
$$

$$
\sum_j \mathbb{E}(\phi_i \phi_j) f_j = \mathbb{E}(\phi_i f)
$$

- In many molecular systems, $\rho(C_i) \ll \rho(C_j)$. Therefore the most convenient choice is often simply to choose a basis $\phi_i$ that is piecewise constant over some cells $C_i$. 

Macro-state model

- Assume some appropriate cells $C_i$ have been defined.
- $\mathcal{P}$: macrostate projection operator

$$[\mathcal{P} f](x) = E(f \mid c(x)) = \frac{1}{\rho(c(x))} \int_{C_{c(x)}} f(y) \rho(y) dy,$$

$c(x)$: macrostate index at $x$

- $\mathcal{P} f$ is a function of the macrostate and $\mathcal{P}^2 = \mathcal{P}$. 
Master equation over cells

- $\phi_i$: characteristic function of $C_i$.
- $\mathcal{M}^k = \sum_{l=0}^{k} \mathcal{M}^{k-l} \mathcal{L}_l + \mathcal{F}_k$.

1. Apply operator $\mathcal{P}$:

$$\mathcal{P} \mathcal{M}^k = \sum_{l=0}^{k} \mathcal{P} \mathcal{M}^{k-l} \mathcal{L}_l$$

2. Apply this operator to the function $\mathcal{M} \phi_i$. 
Define the transition matrices $T_{ij}^{(k)} = \mathcal{P}_j \mathcal{M}^k \phi_i$. Then, Mori-Zwanzig becomes:

$$T^{(k+1)} = \sum_{l=0}^{k} Q_{k-l} T^{(l)}$$

where $[Q_k]_{ij} = [\mathcal{L}_k \mathcal{M} \phi_i]_j$.

The matrices $Q_k$ can be obtained using the following recurrence

$$Q_k = T^{(k+1)} - \sum_{l=1}^{k} Q_{k-l} T^{(l)}$$

Markov approximation: $T^{(k+1)} = Q_0 T^{(k)} = T^{(1)} T^{(k)}$. 
The Markov approximation corresponds to $T^{(2)} = (T^{(1)})^2$.

If $P[P, M] = 0$ then the Markov approximation is exact.

More generally, this condition is sufficient to truncate the recurrence:


\[ k \quad \text{and} \quad k \]
**Commutator**

- $\mathcal{P}\mathcal{M} = \mathcal{P}\mathcal{M}\mathcal{P}$: the fluctuations of a function in a macrostate are averaged out on the time scale $\Delta t$.
- Example: two bins separated by a barrier. Function $f$ is a ramp in the right bin, and 0 in the left bin.

\[ \Delta t = 4, \frac{\mathcal{P}Mf}{\mathcal{P}\mathcal{M}\mathcal{P}f} = 0.2 \]

\[ \Delta t = 16, \frac{\mathcal{P}Mf}{\mathcal{P}\mathcal{M}\mathcal{P}f} = 0.7 \]
Computing rates

- Assume a Markov model, and that matrix $T^{(1)}$ has a full set of eigenvectors.
- Denote: $\phi(k\Delta t)$ probability of $c(x)$ at step $k$.

$$\phi(k\Delta t) = \sum_{n=1}^{N} \phi_0 n (\lambda_n)^k v_n$$

- $\lambda_0 = 1$: equilibrium probability
  - $\lambda_1 = e^{-\kappa\Delta t}$ \Rightarrow relaxation rate $\kappa = - (\ln \lambda_1)/\Delta t$
- The rate can also be computed using

$$\phi(m\Delta t) = \sum_{l=0}^{m-1} Q_{m-1-l} \phi(l\Delta t)$$
One dimensional random walk

- Is there a difference between the Markov approximation with a lag time $m\tau$ and an $m$-term recurrence relation with step $\tau$?
- The computational cost to obtain both approximations is the same.
- 1D system, 16 fine states, macrostate = two adjacent states, Metropolis dynamics, 50% chance of attempting a move.
Convergence of rate

The Markov model converges very slowly.

Relative error in rate

The Markov model converges very slowly.
The statistical error when computing the eigenvalue $\lambda_1$ depends on $\tau$.

The variance of $\lambda_1$ can be obtained analytically from the left and right 2nd eigenvectors of $\mathbf{T}^{(1)}$. 
Long lag time increase of error

- At “long” lag times statistical errors increase rapidly.
- Plateau corresponds to $\tau \gg \lambda_3^{-1}$ = “direct” rate calculation
  = sampling trajectories going from one basin to the other. Benefits of Markov model are lost.
- We need to find a trade-off between two difficulties:
  1. At short lag times, we have a systematic error
  2. At long lag times, statistical errors are large
Memory significant at short time lags. Statistical error increases at longer time lags.
Error due to memory in Markov models

- Although in principle the Markov approximation is equivalent to
  \[ T^{(2)} = (T^{(1)})^2 \]
  this condition can be relaxed when one is interested only in the eigenvalue \( \lambda_1 \).
- Effect of memory is reduced when estimating \( \lambda_1 \) if the cells \( C_i \) are such that the 2nd eigenvector is nearly constant in each cell:
  \[ C_i = \{ x \mid i \varepsilon \leq \psi_2(x) < (i + 1)\varepsilon \}, \quad i \text{ integer} \]
Energy function $U(x, y)$. Cells are vertical slabs rotated by an angle $\theta$. Allows tuning memory or non-Markovity.
Convergence: $\theta = 0, 20, 40$
Corrections on Markov model for rates

More accurate models are available that involve not just local sampling but a global convergence of probabilities.

These models are exact even in the presence of long memory:

1. Weighted ensemble sampling; Kim, Huber, Zuckerman, Darve
2. Cell sampling with correct re-entry statistics; Venturoli, Vanden-Eijnden
3. Non-equilibrium umbrella sampling; Warmflash, Dinner