REDUCED-ORDER MODELING, MORI-ZWANZIG, DISCRETE MODELING, RENORMALIZATION

Alexandre Chorin, Ole Hald, Fei Lu, Matthias Morzfeld, Xuemin Tu

Problem: you have a system of differential equations of the form:

$$\frac{d}{dt}\phi = R(\phi), \quad \phi(0) = x, \tag{1}$$

where $\phi = (\widehat{\phi}, \widetilde{\phi})$, $dim(\widehat{\phi}) = m$, the full system is too complex or uncertain to solve, but you are interested only in $\widehat{\phi}$ and you have measurements of $\widehat{\phi}$ available. Definite initial data $\widehat{x}$ are available only for $\widehat{\phi}$. For the $\widetilde{\phi}$ you have to sample data from a pdf, with values conditioned by $\widehat{\boldsymbol{x}}$.

First, what kind of reduced equation can one expect? Answer from the Mori-Zwanzig formalism.

First, we convert the problem into a linear problem.

Step 1: imbed the problem in a family of problems, $d\phi/dt = R(\phi(t))$, $\phi(0) = x$ for any x, making the solution $\phi = \phi(x,t)$. (this is the key).

Step 2: if one can solve the nonlinear system of ODE's, one can use it solve the linear partial differential equation

$$\frac{\partial u}{\partial t} + \sum_i R_i(x) \frac{\partial u}{\partial x_i} = 0,\ u(x,0) = x$$

by the method of characteristics, and vice versa.

Step 3. The solution of the PDE $u(x,t) \neq \phi(x,t)$; in particular because it is not a function of a given x, t (to find the x at which you have a solution you have to follow a characteristic). However, if you reverse the direction of time, you can choose your x in advance. So reverse the direction of time. This gives the Liouville equation:

$$\frac{\partial u}{\partial t} = \sum_i R_i(x) \frac{\partial u}{\partial x_i} = Lu, \; u(x,0) = x.$$

If the initial data for this equation are $\phi(x,0) = g(x)$, then $u(x,t) = g(\phi(x,t))$. If $g(x) = x_i$, then $u_i(x,t) = \phi_i(x,t)$. If you can solve the linear PDE you can solve the ODE and vice versa.

Notation: We denote the solution of the equation $u_t = Lu,\ u(x,0) = g(x)$, by $u = e^{tL}g$ The equation can be rewritten as

$$\frac{\partial}{\partial t}e^{tL}x = Le^{tL}x.$$

Introduce a projection operator P that projects the space of solutions of this equation onto the span of $\widehat{x}$, for example, onto a space spanned by the components of $\widehat{\phi}$ or a space defined by conditional expectations, i.e., $Pf(\phi) = E[f(x)|\widehat{x}]$.

In the new notations, the linear PDE becomes:

$$\frac{\partial}{\partial t} e^{tL} x = e^{tL} PLx + e^{tL} QLx.$$

The Dyson (or Duhamel) formula says:

$$e^{t(A+B)} = e^{tA} + \int_0^t e^{(t-s))(A+B)} B e^{sA} ds.$$

Setting $A = PL,\ B = QL$, substituting into the previous formula, and applying to the the initial components of $\widehat{\phi}$, you find:

$$\frac{\partial}{\partial t} e^{tL} \widehat{x} = e^{tL} PL\widehat{x} + e^{tQL} QL\widehat{x} + \int_0^t e^{(t-s)L} PL e^{sQL} QL\widehat{x} ds.$$

Multiplying by P the noise term disappears, and we have an evolution equation for the projection of $\widehat{\phi}$ onto the range of P.

The RHS is the sum of a "secular" term, a noise term, and a dissipation/memory term, as follows:

$e^{tL}\widehat{x}$ is $\widehat{\phi}$,, what we want to calculate.

$L = \Sigma R_j \frac{\partial}{\partial x_j}$, so that $L\widehat{x} = \hat{R}(x)$, and $PL\widehat{x}$ is its projection on the resolved subspace (=range of P). The first term on the right is just a function of $\widehat{x}$.

To understand the second term, define $w = e^{tQL}QL\widehat{x}$. By definition, this is the solution of the equation $\frac{\partial}{\partial t}w = QLw$ with initial data $w(0) = QL\widehat{x}$. this is a "noise" term because it lives in the unresolved subspace (the range of Q); it starts from the unresolved part of the data and propagates it orthogonally to the resolved space. This noise is generally neither white nor even Gaussian.

The last term looks at the noise at every instant s between 0 and t, and propagates its effect on $\widehat{\phi}$ up to time t. It embodies a memory.

This is only a road map for how to solve the problem- one has to evaluate e^{tL}, e^{sQL}, P, Q, etc. If the system is chaotic, the various steps are ill-conditioned. This is useful mainly as a starting point for approximation, and a guide as to what to expect. In particular, it shows that a reduced system requires a memory; unresolved degrees of freedom couple past and future. Data are used at many points in the calculation.

Another approach: create a parametric model and fit the data into it ("stochastic parametrization") (work with Fei Lu).

Setup: write the problem in the form:

$$d\widehat{\phi}/dt = \widehat{R}(\phi), \;\; d\widetilde{\phi}/dt = \tilde{R}(\phi).$$

Let $R_0(\widehat{\phi})$ be a low-dimensional approximation of $\widehat{R}(\phi)$, and assume that measurements b_i of $\widehat{\phi}$ are available at a long enough sequence of points in time $t_1, t_2, \ldots, t_n$. Try to use the b_i to determine a remainder function $z = z(\widehat{\phi}, t)$ such that

$$d\widehat{\phi}/dt = R_0(\widehat{\phi}) + z$$

gives a good approximation of $\widehat{\phi}$. In general z must be random.

The equation for z is:

$$db/dt - R_0(b) = z.$$

Problems: data may be not differentiable; there may be too few of them to difference them; this is in general a non-Markovian SDE, which is hard to solve accurately.

Solution: make the problem discrete.

Short detour: The NARMAX parametric representaion of time series.

A time series is a statistically stationary recursion $\ldots u_{i-1}, u_i, u_{i+1} \ldots$. It may have several different representations.

Representation as a moving average (MA):

$$u_i = c_i\xi_i + c_{i-1}\xi_{i-1} + c_{i-2}\xi_{i-2} + \cdots + c_{i-n}\xi_{i-n} = A_n(\xi),$$

where the c_i are constants and the ξ_i are IID Gaussians.

representation as an autoregression (AR):

$u_i - b_1 u_{i-1} - \cdots - b_m u_{i-m} = \xi$, or $(I - B_m)u = \xi$.

A given time series may have few terms in one representation, and a long one in the other. An effective representation should be of the form (ARMA):

$$(I - B_m)u = A_n(\xi).$$

To take into account nonlinear, non-Gaussian effects and external memory, add

$$\Sigma_{k,l>0} b_{k,l} P_k(x^{n+1-l}),$$

where the x_i are known external inputs.

This is the NARMAX (Nonlinear AutoRegression Moving Average with eXternal input) representation (differs a little from what is in the CS literature because x and y are not independent).

The functions P provide "structure", i.e., they define the subspace in which the solution is to live (analogous to the projection P in MZ). Methods for picking the P functions have been discussed in detail in various publications.

If the Ps are known, there are standard ways for estimating the orders and the length of the memory. The coefficients are obtained by maximum likelihood estimation.

Criteria for picking structure, memory length, and coefficients: (i) the reduced system should reproduce selected long-term statistics of the data, such as marginals of the stationary pdf), (ii) it should make reliable short-time predictions, and (iii) it should be stable.

Steps to construct a discrete model:

(1) Discretize the equation $\frac{d}{dt}\hat{\phi} = R_0(p\hat{h}i)$:

$$\hat{\phi}^{n+1} = \hat{\phi}^n + \delta R_\delta(\hat{\phi}^n).$$

(2)Introduce a discrete remainder z:

$$\hat{\phi}^{n+1} = \hat{\phi}^n + \delta R_\delta(\hat{\phi}^n) + \delta z^{n+1},$$

so that unambiguously

$$z^{n+1} = (b^{n+1} - b^n)/\delta - R_\delta(b^n).$$

(3) Find a NARMA representation for the time series z.

Note: (i) Model compared directly with the data; (ii) No stochastic ODE to solve; (iii) No need to worry about a continuum representation of the memory.

Example: The Lorenz96 model.

$$\frac{d}{dt}x_k = x_{k-1}\left(x_{k+1} - x_{k-2}\right) - x_k + F + z_k,$$
$$\frac{d}{dt}y_{j,k} = \frac{1}{\varepsilon}[y_{j+1,k}(y_{j-1,k} - y_{j+2,k}) - y_{j,k} + h_y x_k]$$

with $z_k = \frac{h_x}{J}\sum_j y_{j,k}$, and $k = 1, \ldots, K$, $j = 1, \ldots, J$.

In these equations, $x_k = x_{k+K}$, $y_{j,k} = y_{j,k+K}$ and $y_{j+J,k} = y_{j,k+1}$. We set $\varepsilon = 0.5, K = 18, J = 20, F = 10, h_x = -1$ and $h_y = 1$. R_0 is chosen by setting all the y variables to zero. We observe $b_k = \widehat{\phi}(t_k)$ and assume there is no observation error.

Summary of the model:

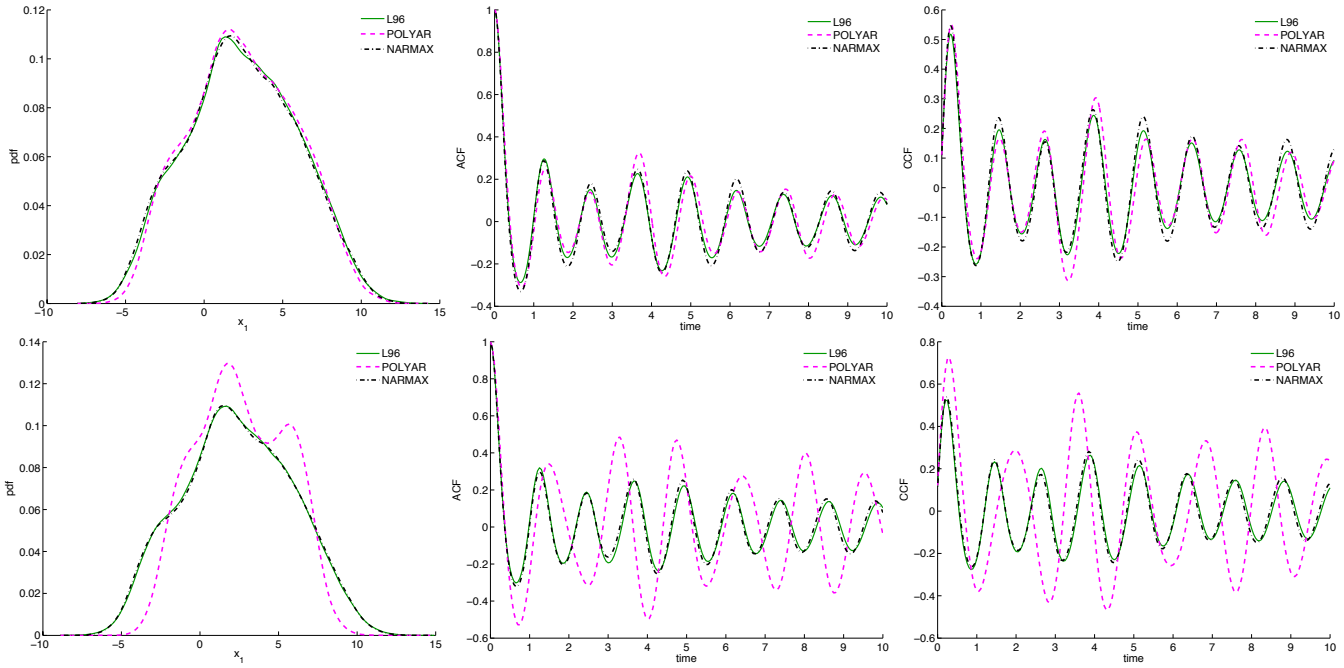
$$x^{n+1} = x^n + \delta R_\delta(x^n) + \delta z^{n+1}, \;\; z^{n+1} = \Phi^{n+1} + \xi^{n+1}, \qquad (2)$$

for $n = 1, 2, \ldots$, where the ξ^{n+1} are independent Gaussian random variables with mean zero and variance σ^2, and Φ^n is the sum:

$$\Phi^n = \mu + \sum_{j=1}^{p} a_j z^{n-j} + \sum_{j=1}^{r} \sum_{i=1}^{s} b_{i,j} P_i(x^{n-j}) + \sum_{j=1}^{q} c_j \xi^{n-j}, \qquad (3)$$

μ, σ^2 and $\left\{a_j, b_{i,j}, c_j\right\}$ are parameters to be inferred.

pdfs,autocorrelations, crosscorrelations, produced by the full Lorenz system, the POLYAR (Wilks) model, and our NARMAX, with data spacing $h = .01$ (top) and $h = 0.05$ bottom. ($h = 0.05$ corresponds to meteorological time approx. 6 hours).



Questions:

I. Compare NARMA and MZ: How come the much less detailed parametric NARMA provides such good results?

II. where should we find data? Experimental data are typically tainted by observation error, and using a large calculationn to get data for a small useful calculation often requires too large a "large" calculation.

Answers: the renormalization group (RNG).

The renormalization group: the idea is that the global statistics of a system that consists of many interacting variables depends only on a few of the small-scale features. The RNG is a way to jettison the irrelevant information step by step.

Example: the central limit theorem. Given a large number of independent random variables $\eta_1, \eta_2, \eta_3, \ldots$ with variance 1, the scaled sum $n^{-1/2}\Sigma\eta_i$ tends to a Gaussian independently of the pdfs of the individual variables.

We now derive this result by a general method that can be applied numerically in complexsituations.

Construct the following sequences of variables:

$$T_{0,1} = \eta_1, \;\; T_{0,2} = \eta_2, \;\; T_{0,3} = \eta_3, \ldots \tag{4}$$

$$T_{1,1} = \frac{1}{\sqrt{2}}(\eta_1 + \eta_2), T_{1,2} = \frac{1}{\sqrt{2}}(\eta_3 + \eta_4), T_{1,3} = \frac{1}{\sqrt{2}}(\eta_5 + \eta_6), \ldots \tag{5}$$

and

$$T_{n+1,1} = \frac{1}{\sqrt{2}}(T_{n,1} + T_{n,2}), \;\; T_{n+1,2} = \frac{1}{\sqrt{2}}(T_{n,3} + T_{n,4}), \ldots \tag{6}$$

for $n \geq 1$.

Let f_n be the pdf of the $T_{n,i}$, with f_0 the original pdf. Then

$$f_{n+1}(x) = \sqrt{2} \int_{-\infty}^{+\infty} f_n(t) f_n(\sqrt{2}x - t) dt. \tag{7}$$

If the f_n converge to a limit f_∞, this equation becomes

$$f_\infty(x) = \sqrt{2} \int_{-\infty}^{+\infty} f_\infty(t) f_\infty(\sqrt{2}\, x - t) dt. \tag{8}$$

This equation has a solution, which is a $N(0, 1)$ Gaussian. The iteration (2.11) converges to that solution, and the limit is independent of the starting pdf f_0, which is forgotten.

If we did not know how to find the fixed point analytically, we could have found it computationally: Form the "block" variables $T_{n,i}$, sample them, make a histogram of the sample values, represent the histogram in some basis, and observe its components. As n increases, the components change; this is a "parameter flow". At each step, information relating to the previous f_n is jettisoned.

Note: we knew to use the coefficient $1/\sqrt{2}$ to preserve the variance of the variables, otherwise one gets a fixed point which is either 0 ir ∞, both of no interest.

Another example: the 2D Ising model.

At each node of a 2D regular lattice with nodes (i, j), $1 \leq i, j \leq N$ lives a variables $s_{i,j}$ taking on the values ± 1. The probability distribution $e^{-H/T}$, where T is a temperature and H is a Hamiltonian $H = -J\Sigma s_i s_{i'}$, and i' denotes the near neighbors of i. In the limit $N- > \infty$ the system undergoes a phase transition from an ordered phase at $T < T_c$ and a disordered phase at $T > T_c$. To estimate these exponents naively one need a gigantic calculation, but one can use the RNG to get around it.

One can start with a finite lattice of modest size, say 8 by 8. One can divide this lattice into 2 by 2 blocks, and define in each block a "super spin" $S_{I,J}$ taking on the values ± 1; the pdf of of these new spins depends on the probabilities of the initial spins.

One can look for the pdf by positing a renormalized Hamiltonian of the form

$$\Sigma c_i P_i(S)$$

and estimating the coefficients by a marginalization. This can be repeated until a fixed point is reached. The parameter flow can be analyzed and exhibits the information loss. There is one interesting fixed point that corresponds to the phase transition; the properties of that transition can be read from the parameter flow in its neighborhood.

The variables S at each step are analogous to the x variables in the L96 problem or the $\widehat{\phi}$ variables in the MZ calculation. The s variables are the y. One can calculate the marginals by stochastic parametrization- one can consider the "Glauber dynamics" of the spins, i.e., the Markov chain Monte Carlo dynamics that sample the pdf at each level, and then derive the MCMC dynamics of the reduced chain by stochastic parametrization of the original chain.

One then identifies fixed points, and one can observe that the parameter space is shrinking, indicating that irrelevant information is being jettisoned.

It would be desirable to factor the transition between a detailed description of a complex system (such as a turbulent flow) and a useful description (e.g. a LES model) into a sequence of RNG steps based on stochastic parametrization (see my talk at the forthcoming meeting in Tel Aviv).