Data Assimilation

Juan M. Restrepo
Department of Mathematics
Physics Department
University of Arizona

Three Estimation Problems:

Given a random time series $\{X(t): t < t_0\}$

$X(t)$ 2 R^N

- Prediction:

  Estimate $\{x(t): t > t_0\}$

- Filtering (Nudiction):

  Estimate $\{x(t_0)\}$

- Smoothing (Retrodiction):

  Estimate $\{x(t): t$ · $t_0\}$

Collaborators:

- Gregory Eyink, Johns Hopkins University
- Frank Alexander, LANL

Turning a model into a state estimation problem

Example:

$\partial_t u(z,t) = \nu\, \partial_{zz} u(z,t) + f(t)$

$u(z,0) = u_0(z)$

$u(0,t) = g(t) \quad u(1,t) = h(t)$

Discretizing:

$x(t)$ ´ $[g(t),u_1(t),u_2(t)\ldots u_N(t)]^T$

is the state variable, and it obeys

$x(t+\delta t) = A\, x(t) + B\, q(t)$

$x(t) \qquad = A\, x(t-\delta t) + B\, q(t-\delta t)$

....

Which leads to:

$L(x(0),\ldots,x(t-\delta t),x(t),x(t+\delta t),\ldots,x(t_f),\ldots,$

$\qquad Bq(t),\, Bq(t+\delta t),\ldots,t) = 0$

$x(t)$ 2 R^N B q 2 R^N

Statement of the Problem

MODEL (Langevin Problem):

$$
\begin{aligned}
dx(t) &= f(x(t),t)dt + [2D(x,t)]^{1/2} dW(t), \qquad t > t_0, \\
x(t_0) &= x_0. \\
x, f, W &\in \boldsymbol{R}^N,
\end{aligned}
$$

DATA:

$$
\begin{aligned}
y(t_m) &= h(x_m) + [2R(x_m,t)]^{1/2} \epsilon_m \\
&\quad \text{where } m = 1, 2, ..., M \\
h, \epsilon &: \boldsymbol{R}^N \to \boldsymbol{R}^{N_y}
\end{aligned}
$$

GOAL: estimate moments

(at least) find mean conditioned on data:

$x_S(t) = E[\, x(t)|\, y_1,\ldots, y_M]$

and

Covariance matrix (uncertainty)

$C_S(t) = E[(x(t)-x_S(t))(x(t) - x_S(t))^{>}|y_1,\ldots,y_M]$

The conditional mean $x_S(t)$ minimizes

tr $C_S(t) = E[|(x(t)-x_S(t))|^2|y_1,\ldots,y_M]$.

It is termed the smoother estimate.

A Nonlinear Example

Stochastic Dynamics (Langevin Problem):

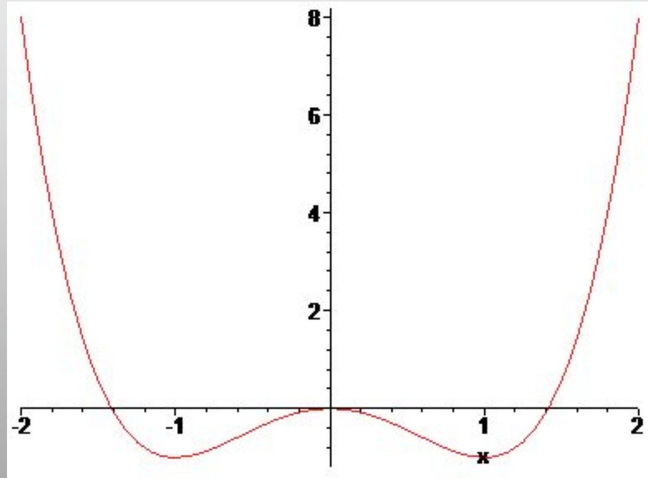
$dx(t) = f(x(t))\,dt + \kappa\,dW(t)$

with

$V(x) = -2x^2+x^4$

$f(x) = -V'(x)=4x(1-x^2)$

$\kappa = 0.5$



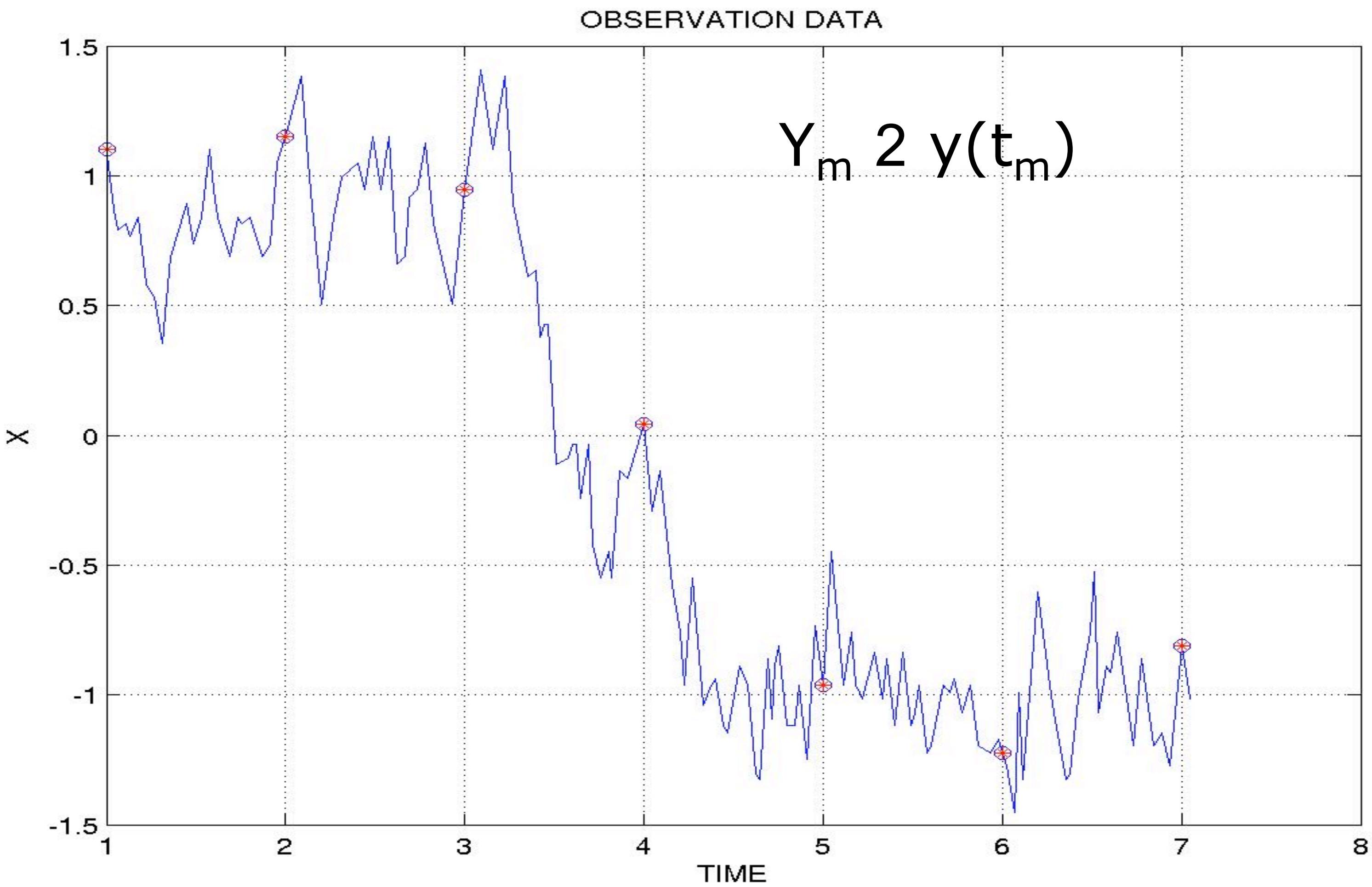
Measurements:

at times $m\,\Delta t$, m=1,…, M one observes

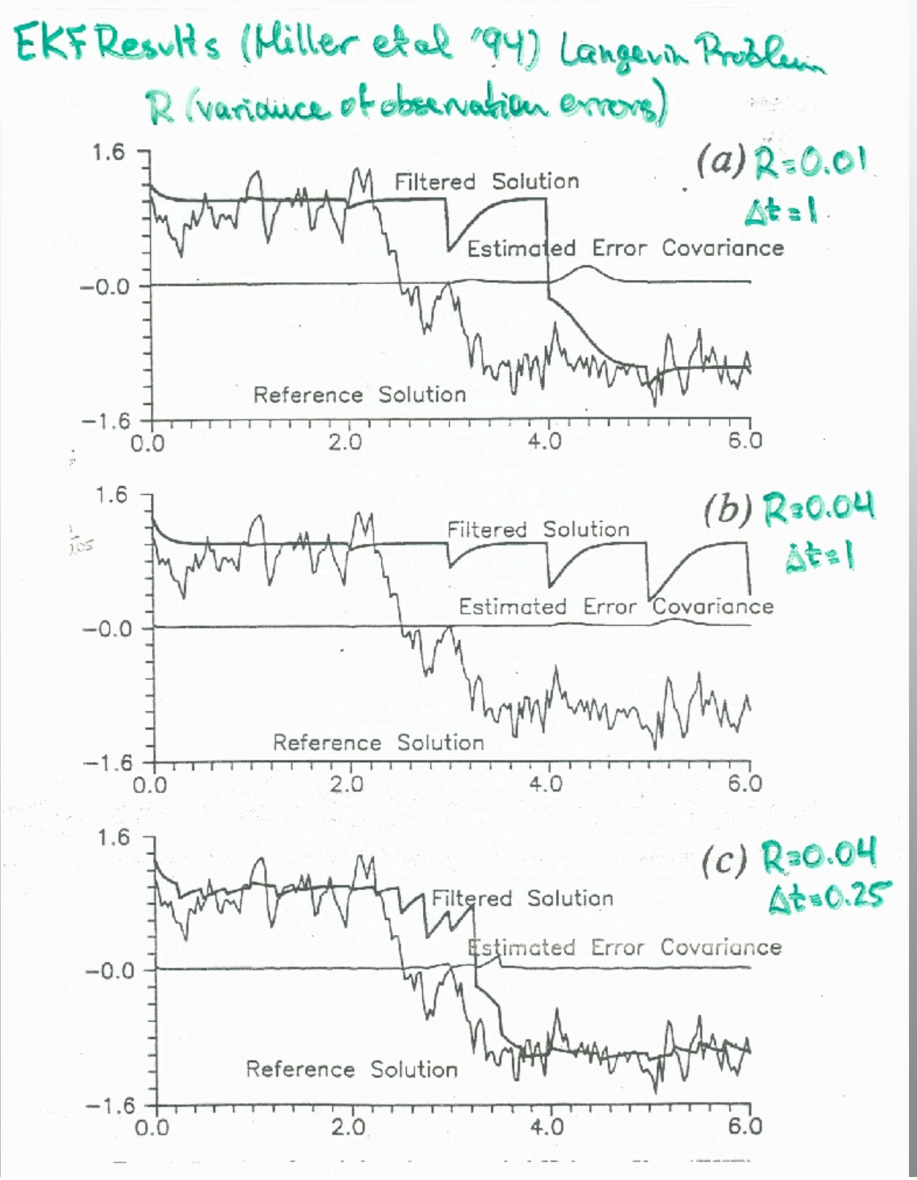
$y_m := X(t_m) + \rho\,N_m$

to have measured values Y_m, m=1,2,…,M

Observations



Extended Kalman Filter



Alternative Approaches

- KSP: optimal, but impractical
- ADJOINT/4D-VAR: optimal on linear/Gaussian

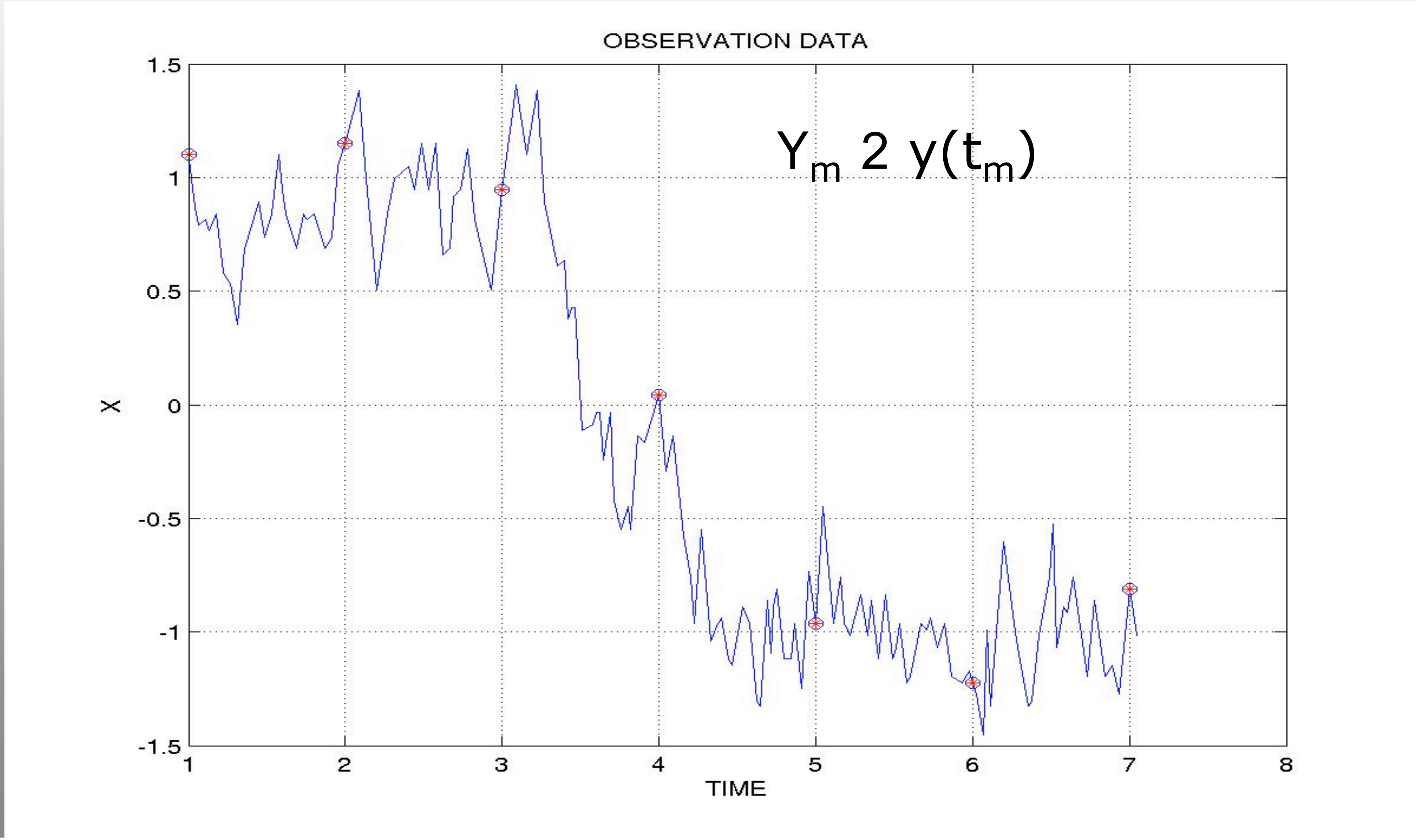
  (Restrepo, Leaf, Griewank, SIAM J. Sci Comp 1995)
- Mean Field Variational Method

  (Eyink, Restrepo, Alexander, Physica D, 2003)
- enKF (ensemble Kalman Filter)
- Particle Method

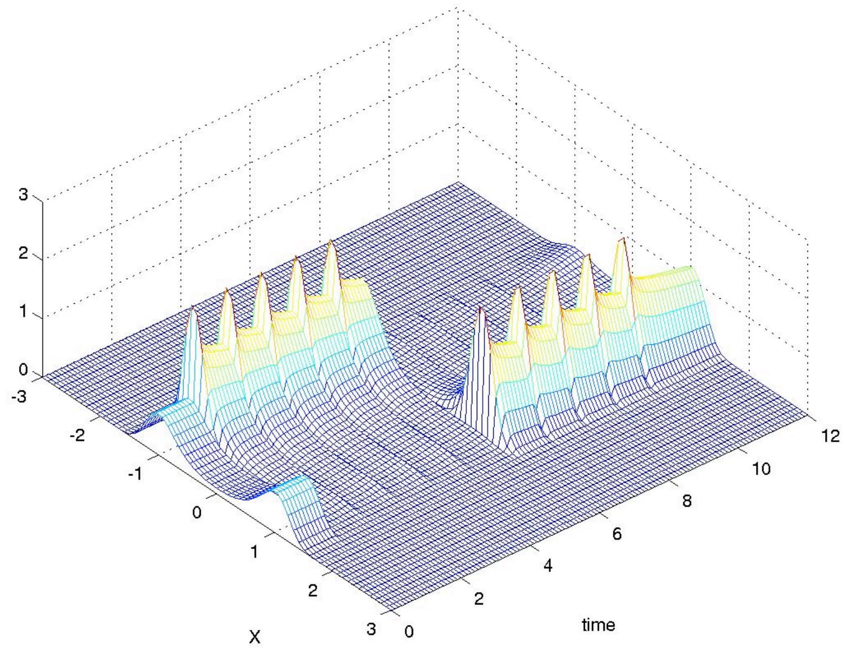
  (Kim Eyink Restrepo Alexander Johnson, Mon. Wea. Rev. 2002)
- Path Integral Method

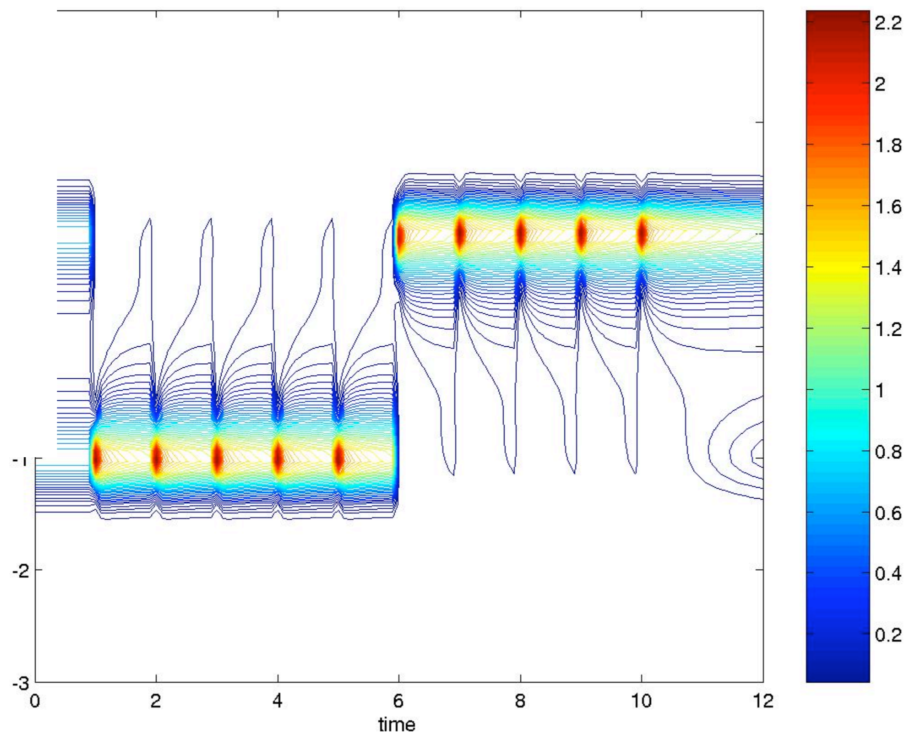
  (Alexander Eyink Restrepo, J. Stat. Phys. 2005)

Observations



KSP Filter Results





Why not KSP?

Impractical!

Very tiny example: a scalar PDE in 1-d space and time solved on a lattice of N points in space: Dim(x(t))=N. Each component of x(t) requires the solution of 2 stiff PDE’s

A Statistical-Mechanical Digression

- Configuration probability for N particles with (q_i, p_i) and fixed energy E_i

  Divide phase space into m cells thus

$$G = N!/(n_1!\, n_2! \ldots n_m!)$$

ways of distributing N particles in m energy cells.

THUS P / G is a probability.

Fact: log n! ¼ n log n - n

$$\text{Log } P = N \log N - \sum_j (n_j \log n_j - n_j) + C$$

$$P / \exp(-N \sum_j p_j \log p_j)$$

$$p_j = n_j/N$$

BAYESIAN STATEMENT

- P(X|D) / Prior £ Likelihood
- Use data for the likelihood
- Use model for the prior

$$P(X|D) = \exp(-A_{data}) \exp(-A_{model})$$

Path Integral Method

- Related to simulated annealing
- It could be developed as a black box
- Simple to implement
- Can handle nonlinear/non-Gaussian problems
- Calculates sample moments

PROBLEM: Relies on MC!!!

$$
\begin{aligned}
dx(t) &= f(x(t),t)dt + [2D(x,t)]^{1/2} dW(t), \qquad t > t_0, \\
x(t_0) &= x_0.
\end{aligned}
$$

Discretized using explicit Euler-Maruyama scheme

$$
\begin{aligned}
x_{k+1} &= x_k + f(x_k,t_k)\delta t + (2D)^{1/2}(x_k,t_k)(W(t_k+\delta t) - W(t_k)), \\
& \quad k = 0,1,2,\ldots \\
x_{k=0} &= x_0.
\end{aligned}
$$

Let η (t_k) = $W(t_k + \delta t)$-$W(t_k)$,
at times t_k, k=0,1,2,…,

Suppose $\eta(t_k)$ is Gaussian
Prob $\eta(t)$ »exp(-1/2 $\sum_k$ | η (t_k) $|^2$).

Hence exp(-A_{dyn}), for $t = t_0, t_1, \ldots t_T$

A_{dyn} ´ ¼ $\sum_{k = 0}^{T-1}$ [$[(x_{k+1} - x_k)/\delta t - f(x_k,t_k)]^{˃}$ $D^{-1}(x_k, t_k)$
$[(x_{k+1}-x_k)/\delta t - f(x_k,t_k)]$]

$$A_{dyn} \; ´ \; \sum_{k=0}^{T-1} \left[\left[(x_{k+1} - x_k)/\delta t - f(x_k, t_k) \right]^{>} D^{-1}(x_k, t_k) \left[(x_{k+1} - x_k)/\delta t - f(x_k, t_k) \right] \right] / 4$$

To include influence of observations

use Bayes' rule.

This modifies Action:

$$A_{obs} = \sum_{m=1}^{M} [h(x(t_m) - y(t_m)]^{>} R^{-1} [h(x(t_m)) - y(t_m)]$$

The Total Action:

$$A = A_{dyn} + A_{obs}$$

The Action is like the log-likelihood.

If it is known that A is convex and has a unique extremizer: use principle of least action:

get Euler Lagrange equations to solve

OR

use a minimization scheme

Otherwise use sampling:

- Hybrid Monte Carlo (HMC)
- Unigrid Monte Carlo (UMC)
- Generalized Monte Carlo (GHMC)
- Shadow Monte Carlo (SHMC)

Hybrid Monte Carlo

- molecular dynamics: used to propose a new system configuration
- Metropolis MC: accept/reject based on the energy

Configuration is specified by

degrees of freedom q_0, q_1, ... , q_T.

q_i 2 R^N

The HMC algorithm:

To each q_i, a conjugate generalized momemtum, p_i, is assigned.
The momenta p_i give rise to a kinetic energy $H_K = \sum_i p_i^2/2$.
The total Hamiltonian of the system $H = A + H_K$.
The dynamics are:
$dq_i/d\tau = p_i$
$dp_i/d\tau = F_i$ where
$F_i = -\partial A/\partial q_i$
is the force on the i^{th} degree of freedom.

What's going on?

Write Probability $P(x) = \exp(-E(x))/Z$

$E(x)$ and grad($E(x)$) are easily evaluated:

Gradient indicates which direction one should go to find states with higher probability!

Note: $H(q,p) = A(q) + H_K$

$P_H(q,p)=\exp(-H(q,p))/Z=\exp(-A(q))\exp(H_K(p))/Z$

separable

then marginal distribution $\exp(-A(q))/Z_q$

1) A chain of states is generated:
(q_i', p_i') $i=0,1,2,\ldots,T$, by J steps (ficticious time).

2) Detailed balance achieved if new configuration is accepted with probability min[1, exp ΔH], where $\Delta H = H(q',p') - H(q,p)$.

The Metropolis step corrects for time discretization errors.

3) p' refreshed after each acceptance/rejection taken from Gaussian distribution of independent variables $\exp(-H_K)$.

Unigrid Monte Carlo

Updates system by taking coherent moves on a number of length'' scales.

Decompose system into blocks of contiguous lattice points.
B=Block sizes 1, 2, 4, ..., 2^s.
B=1 the standard local Metropolis.

Update: to each site in B local value has a random (Gaussian) $\delta\phi$ added to it.

Metropolis accept/reject as before.

Generalized HMC

Dynamics replaced by:

$dq_i/dt = A\, p_i \quad dp_i/dt = [A]^T F_i$

A is an N£ N matrix.

when A =I we obtain HMC.

Discretize:

$q' = q + \delta\tau\, Ap + ½\, \delta\tau\, A\, A^T F([q])$

$p' = p + ½\, \delta\tau\, A^T(F[q]+F[q'])$

Challenge: find A that leads to a significant reduction of the correlation time.

Used the circulant matrix on example:

$A = \mathrm{circ}(1, \exp(-\alpha), \exp(-2\alpha), \ldots, \exp(-T\alpha))$

A Nonlinear Example

Stochastic Dynamics (Langevin Problem):

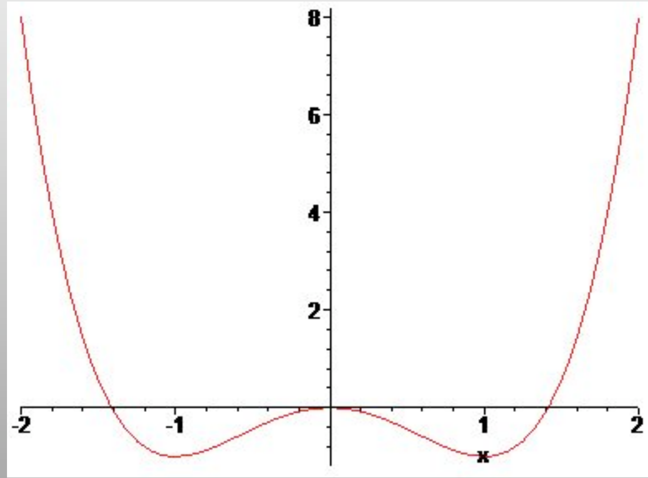
$dx(t) = f(x(t))\,dt + \kappa\,dW(t)$

with

$V(x) = -2x^2+x^4$

$f(x) = -V'(x)=4x(1-x^2)$

$\kappa = 0.5$



Measurements:

at times $m\,\Delta t$, $m=1,\ldots, M$ one observes

$y_m := x(t_m) + \rho\,N_m$

to have measured values Y_m, $m=1,2,\ldots,M$

PIMC Results

mean and variance consistent with the KSP results.

In general the issue is not whether the estimate is good. It’s whether the computational cost is justified.

RESULTS:
decorrelation time

$T+1$	HMC (J=1)	HMC (J=8)	UMC	GHMC (J=1)
8	900(125)	170(7)	800(40)	40(8) [0.20]
16	5300(1600)	560(20)	1040(60)	60(10) [0.10]
32	13300(8300)	2700 (140)	1430(100)	200(30) [0.05]
64	30000(7800)	2800(400)	1570(100)	420(70) [0.0245]

T+1 number of (time steps X dimension)
J number of Hamiltonian solve steps per sweep
(standard deviation)
[α] in A

Conclusions (Sampling)

- GHMC could affect a speedup
- No clear way to produce the matrix
- Cost of matrix/vector multiply can be significant
- Path Integral can be used as a black box data assimilator (if several sampling options are available). Faster samplers needed.

OVERALL CONCLUSIONS

•(BAYESIAN) DATA ASSIMILATION:
USES MODEL AND DATA AS A FORWARD PROBLEM
REQUIRES KNOWLEDGE OF NOISE STATISTICS

•PIMC IS SUITABLE IN HIGHLY NONLINEAR AND/OR NON-GAUSSIAN DISCRETIZED PROBLEMS
Computational cost is the issue:
Order is (D T MC) Dimension, Time steps, MC trials

•IF GAUSSIAN/LINEAR: *Least-squares is optimal (Kalman filter/smoother is sequential variant)*
•IF MILDLY NONLINEAR/NEAR-GAUSSIAN:
Extended KF, ensemble KF

Further Information:

http://www.physics.arizona.edu/~restrepo