Multi-scale Hydrological Data Assimilation in Layered Media

Juan M. Restrepo

Department of Mathematics

Physics Department

University of Arizona

Collaborators:

- Daniel Tartakovsky, UCSD
- Michael Holst, UCSD

TIME SERIES
Estimation Problems:

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

$z(t)$ 2 R^N

- Prediction:

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

- Filtering (Nudiction):

Estimate $\{z(t_0)\}$

- Smoothing (Retrodiction):

Estimate $\{z(t): t \cdot t_0\}$

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:

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

is the state variable, obeying

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

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

....

Leads to LINEAR PROBLEM:

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

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

$x(t)$ 2 R^N $\quad$ $B\, q$ 2 R^N

Statement of the Problem

MODEL (Langevin Problem):

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

$$x, f, dW \in \boldsymbol{R}^N,$$

DATA:

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

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
$\mathrm{tr}\, C_S(t) = E[|(x(t)-x_S(t))|^2 | y_1, \ldots, y_M]$.
It is termed the smoother estimate.

Optimal Estimate of Discretized Linear Model with Gaussian Noise

Let $z_i = u(x_i)$ where x_i 2 Ω

$$B\, z + n_m = F$$

$$D\, z + n_d = Y$$

OR

$$M\, z + N = T$$

$$\min_z J = \langle N^T N \rangle$$

Least Squares, SVD (Kalman)

A Nonlinear Example

Stochastic Dynamics (Langevin Problem):

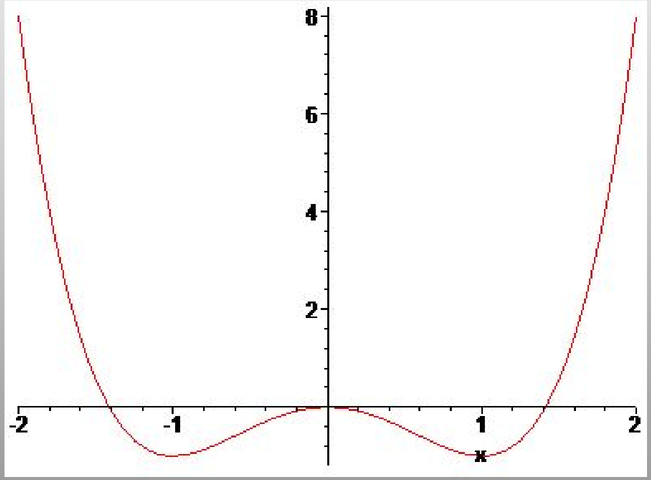
$dx(t) = f(x(t))\, dt + \kappa\, dW(t)$

with

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

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

$\kappa = 0.5$



Measurements:

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

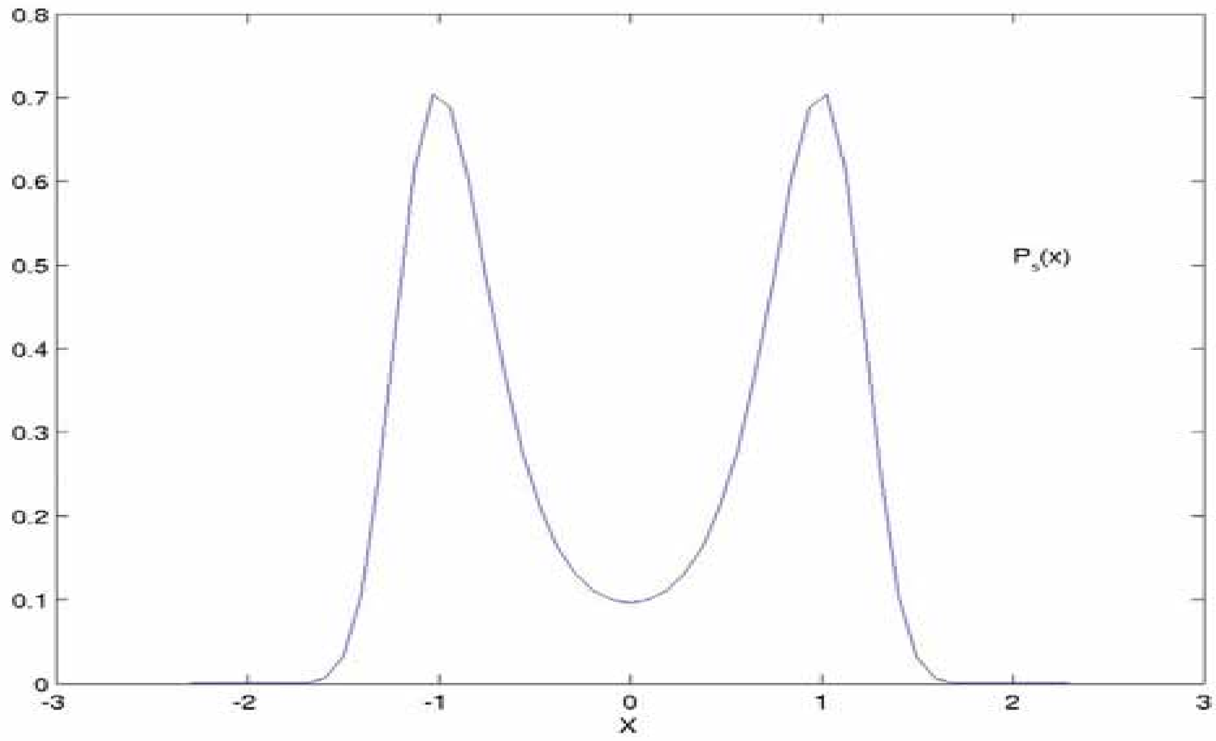
$y_m := X(t_m) + \rho\, N_m$

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

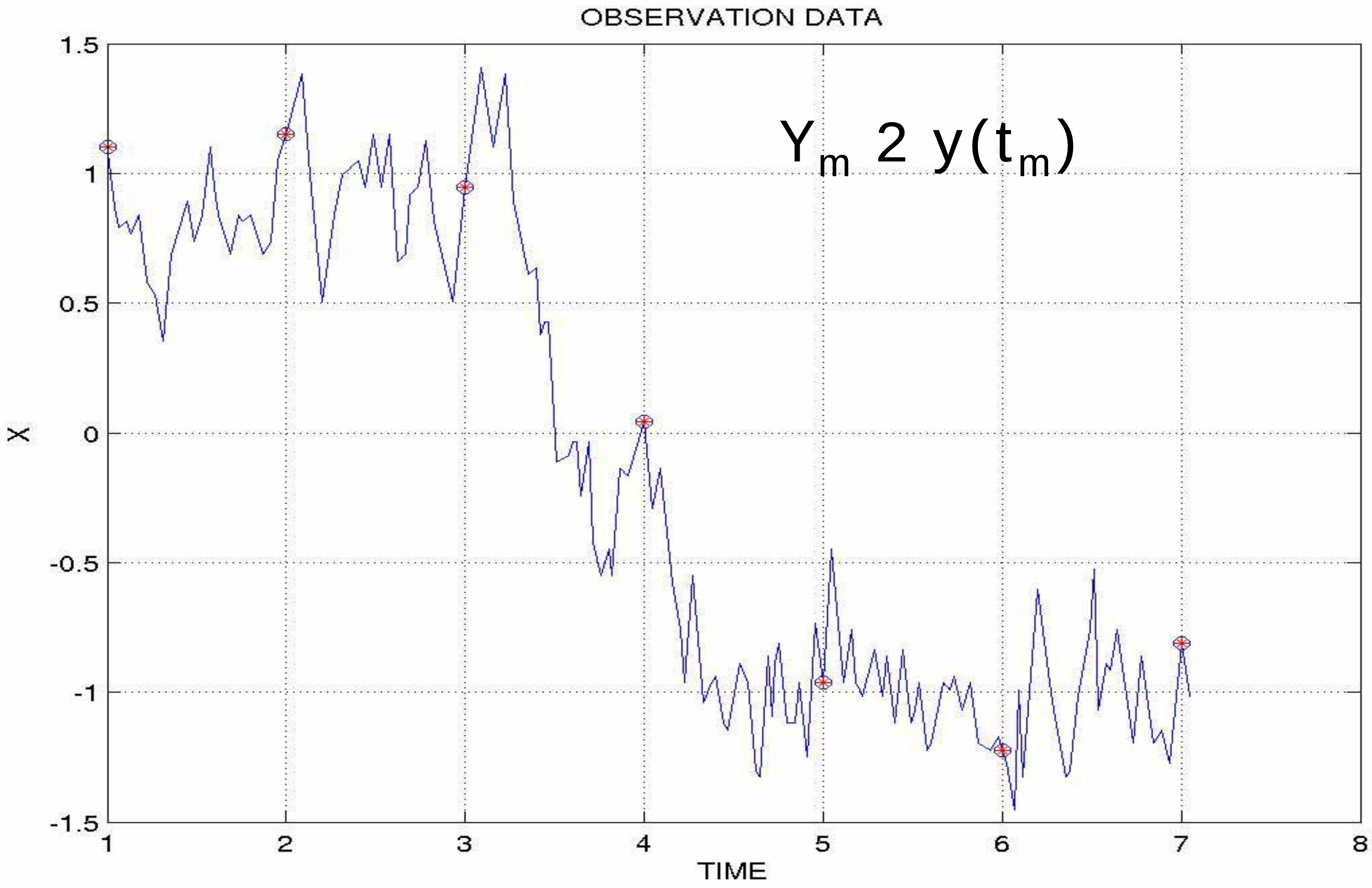
Kolmogorov Equation

$$\partial_t P = -\partial_x [f(x) P] + \kappa^2 \partial_{xx} P/2$$

$$P(x,t)_{t\ !\ 1} = P_s(x)$$



Observations



Y_m 2 $y(t_m)$

BAYESIAN STATEMENT

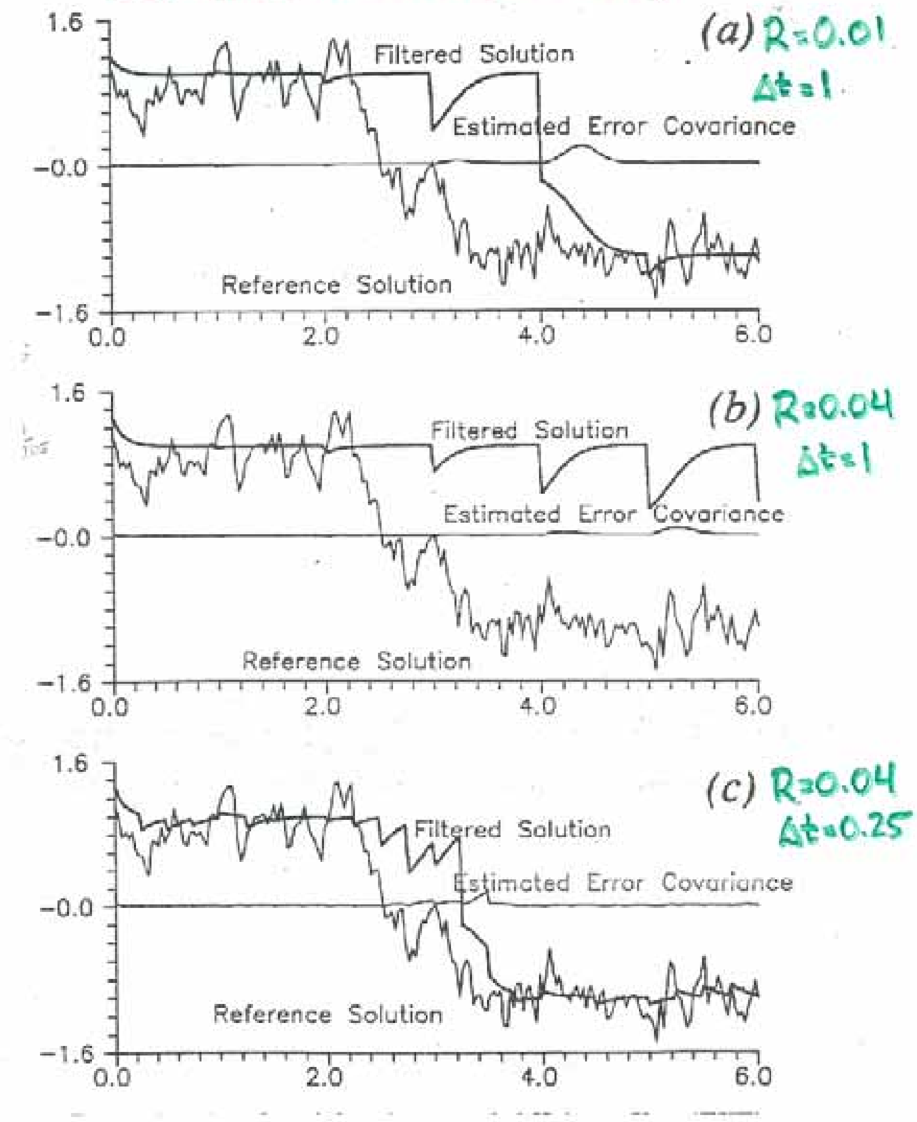
- P(X|D) / Prior £ Likelihood
- Use data for the likelihood
- Use model for the prior

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

Extended Kalman Filter

EKF Results (Miller et al '94) Langevin Problem

R (variance of observation errors)



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 and Restrepo Physica D, 2007)

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 $\eta\ (t_k) = W(t_k + \delta t) - W(t_k)$,
at times t_k, $k=0,1,2,\ldots,$

Suppose $\eta(t_k)$ is Gaussian

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

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

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

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

To include influence of observations

use Bayes' rule.

This modifies Action:

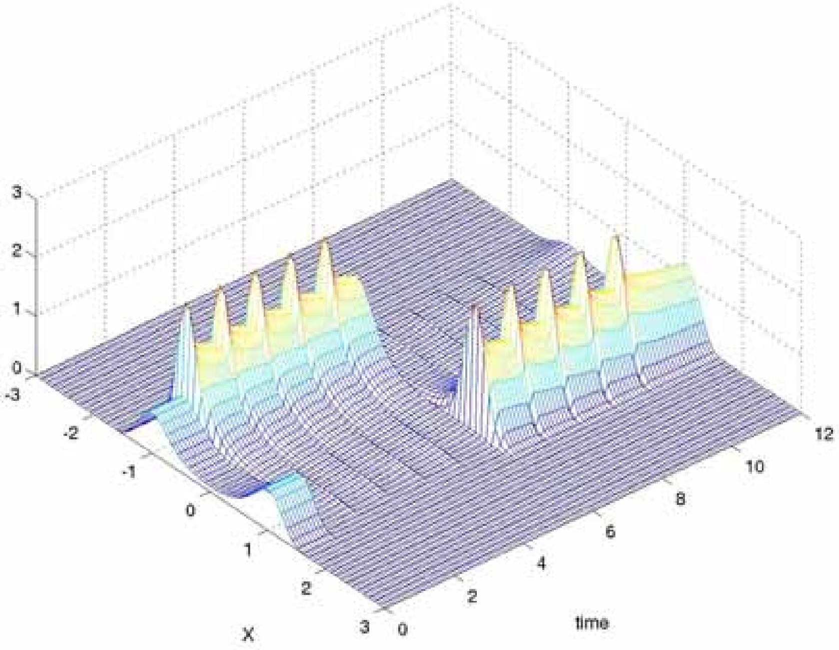
$$A_{obs} = \sum_{m=1}^{M} [h(x(t_m) - y(t_m)]^{>} R^{-1} [h(x(t_m)) - y(t_m)]$$

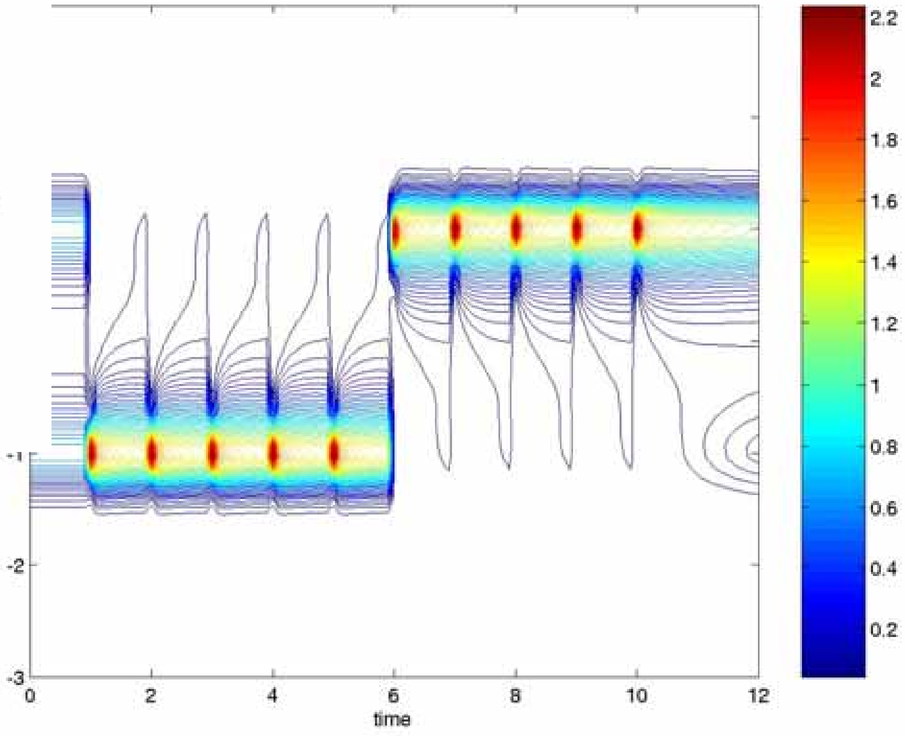
The Total Action:

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

The Action is like the log-likelihood.

PIMC Filter Results





Estimation Applied To Steady State Hydrology

- Estimate hydraulic head in domain
- Estimate material properties in domain
- Estimate “best” boundary values
- Estimate all of the above

Simplest Boundary Value Problem

MODEL:

$$\begin{array}{rcll} -\nabla\cdot[K(x)\nabla u]-f(x) & = & n(x) & x\in\Omega \\ u|_{\partial\Omega_i} & & \text{continuous} & \\ \hat{n}\cdot K(x)\nabla u|_{\partial\Omega_i} & & \text{continuous} & \\ u|_{\partial\Omega} & = & U+\theta(x) & \end{array}$$

DATA:

$$Y(x)=T(u,K)+\rho$$

$n(x)$, $\theta(x)$, $\rho(x)$ are known statistical quantities

OUR APPROACH

- USE DATA-DRIVEN CLASSIFICATION: estimates partitioning into homogeneous layers.

  Support Vector Machines

- DISCRETIZE Variational formulation for the model plus constraint (via Lagrange multiplier): constrained minimum satisfies E-L. Coupling of each subproblem is automatically satisfied.

  Weak form (using Dirichlet energy)

- SOLVE nonlinear system in each subdomain:

  Newton

Data-Driven Classification

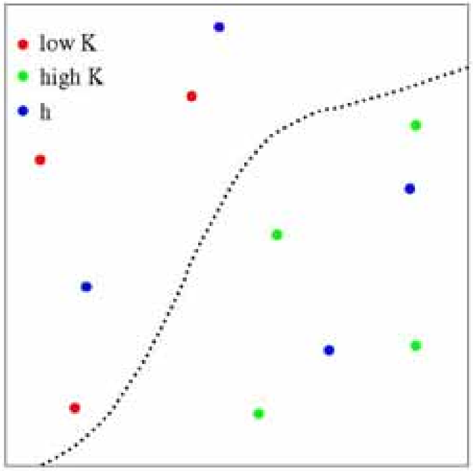
Estimate the boundaries between heterogeneous geologic facies

- Data

  $K_i = K(\mathbf{x}_i)$, e.g., conductivity
  $h_{jk} = h(\mathbf{x}_j, t_k)$, e.g., head

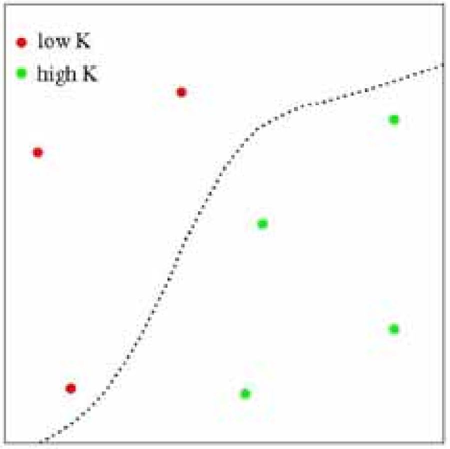
- Data are sparse

- Measurements are well differentiated



Measurements of system parameters (K) $\Longrightarrow$ forward FD problem

Measurements of system states (h) $\Longrightarrow$ inverse FD problem



- Assign indicators to data,

  $$I(\mathbf{x}_i) = 1(0) \quad \text{if} \quad \mathbf{x}_i \in M_1(M_2)$$

- $\mathcal{I}(\mathbf{x}, \boldsymbol{\alpha}) \equiv$ an estimate of $I(\mathbf{x})$
- $\min R = \int ||I - \mathcal{I}|| \mathrm{d}P(I, \{\mathbf{x}\}_{i=1}^N)$

- Geostatistics (Kriging)
  1. the L^2 norm
  2. the indicator function $I(\mathbf{x})$ is a random field, and
  3. the choice of sampling locations $\{\mathbf{x}_i\}_{i=1}^N$ as deterministic. $\Longrightarrow$
  4. Variance: $\sigma_I^2 = \int (I - \mathcal{I})^2 \mathrm{d}P(I)$

- SVMs
  1. the L^1 norm
  2. the indicator function $I(\mathbf{x})$ as deterministic, and
  3. the choice of sampling locations $\{\mathbf{x}_i\}_{i=1}^N$ as random. $\Longrightarrow$
  4. Expected risk: $\min R_{\mathrm{exp}} = \int |I - \mathcal{I}| \mathrm{d}P(\{\mathbf{x}\}_{i=1}^N)$

Support Vector Machines

- Alternative to Kriging
- Very good alternative when sample densities are too low for Kriging
- Highly automated
- Can be incorportated in the solver problem

Heterogeneous Sub-Surface

In each subdomain i = 1, 2, .., M

$$\begin{aligned} K(x,\omega) &= \exp\left[\sum_{j=1}^{\infty} \kappa_j(\omega)\phi_j(x)\right] \\ u(x,\omega) &= \sum_{j=1}^{\infty} \mu_j(\omega)\phi_j(x) \end{aligned}$$

$$-\nabla \cdot (K\nabla u) - \overline{f} = n(x,\omega)$$

$$\begin{aligned} E(n) &= 0 \\ E(n(x)n(y)) &= g(|x-y|) \end{aligned}$$

(Weak) Variational Formulation

- Let P:=[u,K]
- Use standard machinery to solve nonlinear problem but use weighted norms (locally in each subdomain).
- Use Newton solver but decide whether to do global estimate of partial estimates (increasing or decreasing the uncertainty in each subdomain).
- Use Galerkin discretization of Newton Systems.

Weak Formulation (no noise)

$$\phi(P) \quad = \quad \frac{1}{2}||T(P) - y||^2 + S(P - P_0)$$

Dirichlet-like Energy $\quad S(P) \quad = \quad \sum_{i=1}^{M}\left[\int_{\Omega_i}(\frac{1}{2}|\nabla P|^2 + kP^2)\,dx\right]$

$$G(P) = -\nabla \cdot (K\nabla P) - \overline{f} = 0$$

$$\Phi(P,\Lambda) = \phi(P) - \langle \Lambda, G(P) \rangle$$

LEADS TO: Find [P,Λ]^T such that

$$\begin{aligned} \langle \Phi'(P), v \rangle - \langle \Lambda, G'(P)v \rangle \quad &= \quad 0, \quad \forall\, v \in X \\ \langle \nu, G(P) \rangle \quad &= \quad 0, \quad \forall\, \nu \in Y^* \end{aligned}$$

X, Y* Banach spaces

Newton Solution

Find corrections $[\pi,\lambda]^T$ to $[P,\Lambda]^T$

$$\begin{bmatrix} \phi''(P) - [G''(P)^T \Lambda] & -G'(P)^T \\ G'(P) & 0 \end{bmatrix} \begin{bmatrix} \pi \\ \lambda \end{bmatrix} = \begin{bmatrix} -\phi'(P) + G'(P)^T \Lambda \\ -G(P) \end{bmatrix}$$

To find Hessians and Jacobians, use ADIFOR/C

Final Comments

- Model error formulation vs. closure?
- Already existing nonlinear solvers.
- Weak formulation automatically takes care of boundary conditions at the layer interfaces.
- Can give a-priori estimates of error.
- Unlike Inverse Method (Tikhonov, e.g.) problem is greatly more numerically stable.
- Use PIMC (see Restrepo, 2007) to benchmark results.
- Constrain number of SVM subdomains to the Newton solve.

Further Information:

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