Quantifying Uncertainty

Sai Ravela

M. I. T

2012

Summary

- Uncertainties can be aleatory or epistemic.
- Uncertainty Quantification is the process of converting epistemic uncertainties to aleatory ones.
- From a systems perspective, we typically employ models of real world processes. These models may be empirical or be realizations from theory.
- Typically, model parameters, model states and model structure are subject to epistemic uncertainties that we need to quantify
- Uncertainty Quantification requires the representation, propagation and estimation of probability density or mass functions associated with the states, parameters and structure estimates.
- We study methods with simple and accessible examples in this course.

Potentials and Probabilities

It is easy to define a "potential function" in many cases that expresses belief. For example:-

$$g(x;\sigma) = e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

$$g(0;\cdot) = 1 \quad (2)$$

Where, x is a scalar variable and σ^2 is the variance.

So, provided the sum exists, we commonly exchange beliefs with probabilities via the normalization:

$$P(X) = Pr(X = x) = f_X(x) = \frac{\phi(x)}{\sum_y \phi(y)} \quad (3)$$

Identical and Independent

- A discrete random variable (rv) X takes on a random value from a distribution.
- Sampling the distribution implies producing a sequence of rvs, each holding a sample. These rvs are **identical** by construction.
- Each rv is **independent** of the other in the sequence.
- The rvs are exchangeable; the order doesn't matter here.
- We refer to this as iid.

Sampling

Sampling: A **sample** of size n is a set of iid rvs $X_1 \ldots X_n$.

- Sometimes, we don't have an explicit distribution to start with, but a process real or numerical, that we sample, using sensors or observation equations.
- Variously denoted as a transfer function, sampler, model, sensor or other abstractions, these objects give us access to the distribution by repeated inquiry.

Estimator

- There is a distribution.
- There is an iid sample.
- There is a statistic calculated from the sample (e.g. mean, variance).
- The statistics estimate the parameters of the distribution.
- The estimate is uncertain and this uncertainty has its own distribution; the **sampling distribution** of the statistic.

Central Limit, a good start?

- The sampling distribution of the statistic can be estimated by repeatedly drawing n-length sample sequences from a distribution, calculating the statistic and then considering the resulting distribution.
- If the original distribution had a mean m and variance s, then in large n, the sampling distribution:
- Converges to a Gaussian.
- The sample mean converges $m_s \rightarrow m$ and has variance $v = \frac{s}{n}$
- The sample variance converges as $(n-1)v = \chi^2(n-1)$.
- Small sample problem!

Confidence and Error

- When communicating an estimate of a parameter of the distribution (e.g. mean) of a quantity of interest (e.g. model state or parameter), be sure to communicate the uncertainty in the estimate (the parameters of the sampling distribution).
- These include "standard error", "variance", "confidence interval", "the error bars" etc.
- Often this will require iid sampling from the available data. Various techniques such as jacknife and bootstrap are useful.

Random/Stochastic Process

- The future outcome is not deterministic. e.g. temperature over time at the intersection of Ring Road and Mehrauli road...
- Let's consider an rv sequence: $X_1 \ldots X_n$; causality is implied going right.
- We may think of the distributions $P(X_i)$ or more generally $P(X_1 \ldots X_n)$.

Stationarity: (of the distribution, not process!) implies
$P(X_{1+\tau} \ldots X_{n+\tau}) = P(X_1 \ldots X_n)$
Markovianity: (is that a word!) implies
$P(X_n|X_{n-1} \ldots X_1) = P(X_n|X_{n-1})$
Correlation: $(C_{1\tau}) = E[(X_n - E(X_n))(X_{n+\tau} - E(X_{n+\tau}))]$ and the relaxation time τ_r is when $C_{1\tau_r}$ is small.
Ergodicity: Implies one can exchange time averages with ensemble averages $\bar{x} =< x >$.

In a Linear Gaussian World

1. Two examples of linear models.
2. Assumption of Gaussian uncertainties.
3. Parameter and State Estimation
4. Uncertainty propagation.
5. We may skip the derivations!

AR Model–Some History

Did you know – the Yule Walker Equations for identifying an AR model are:

1. YULE (1927): AR(2) model for sunspots.
2. WALKER (1931): AR(4) model for darwin pressure/Southern Oscillation by observing Tahiti-Darwin correlation.

Methodology has had explosive impact in many, many areas.

Emperical models or Physical models?

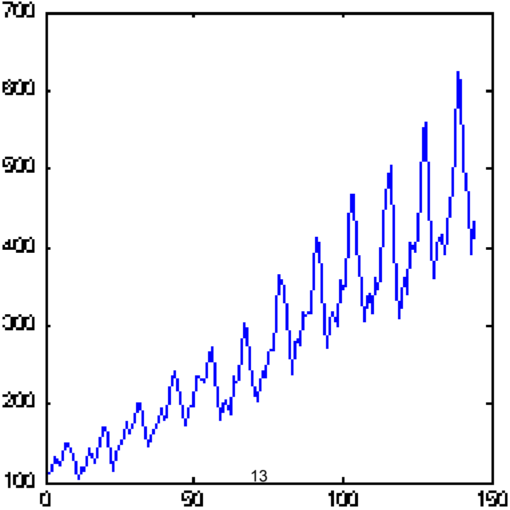
This is a dilemma for modeling many physical processes:

1. Physics-based models applicable to the full-range of dynamics, but difficult to implement and often with too-many degrees of freedom for the problem of interest.
2. Empirical ones can't generalize, limited predictability.

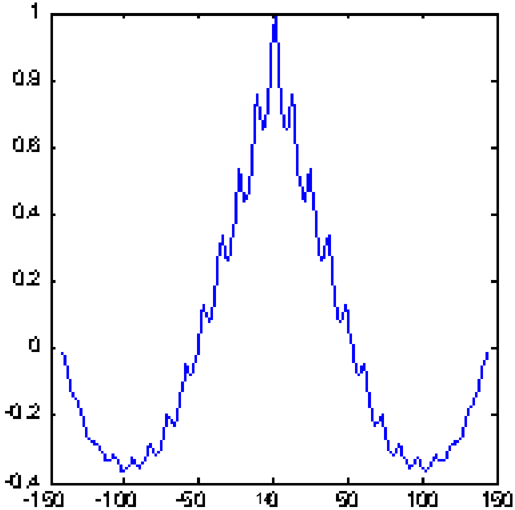
Often, it takes both skills to build a good model but the two don't speak the same language or communicate well.
As Walker says:
There is, today, always a risk that specialists in two subjects, using languages full of words that are unintelligible without study, will grow up not only without knowledge of each other's work, but also will ignore the problems which require mutual assistance.

Plot of Airline Passenger data

Source: Hyndman, R.J., Time Series Data Library,



Plot of Airline Passenger data



AR (n) Model

$$X_n = \alpha_1 X_{n-1} + \alpha_2 X_{n-2} + \ldots + \alpha_p X_{n-p} + \eta_n \qquad (4)$$

Where X_i is a zero-mean scalar measurement at discrete time iΔt and η_n is the random perturbation (noise, uncorrelated in time) at discrete time nΔt. Taking expectation up to a lag of p.

$$\begin{aligned} < X_n X_{n-1} > &= \alpha_1 < X_{n-1} X_{n-1} > + \alpha_2 < X_{n-1} X_{n-2} > + \\ &\quad \ldots + \alpha_p < X_{n-1} X_{n-p} > + < X_{n-1} \eta_n > \qquad (5) \\ \sigma_{01}^2 &= \alpha_1 \sigma_{00}^2 + \alpha_2 \sigma_{01}^2 + \ldots + \alpha_p \sigma_{0p-1}^2 \qquad (6) \\ \rho_1 &= \alpha_1 1 + \alpha_2 \rho_1 + \ldots + \alpha_p \rho_{p-1} \qquad (7) \end{aligned}$$

Where, ρ_k is the lag-k autocorrelation coefficient, $\rho_k = \frac{\sigma_{0k}^2}{\sigma_{00}^2}$ and σ_{0k}^2 the lag-k auto-covariance, by definition.

Expanding further, we get

$$
\begin{aligned}
\rho_1 &= \alpha_1 1 + \alpha_2\rho_1 + \ldots + \alpha_p\rho_{p-1} \\
\rho_2 &= \alpha_1\rho_1 + \alpha_2 1 + \ldots + \alpha_p\rho_{p-2} \\
\vdots &\quad \vdots \\
\rho_p &= \alpha_1\rho_{p-1} + \alpha_2\rho_{p-2} + \ldots + \alpha_p 1
\end{aligned}
$$

Providing the Yule-Walker model

$$\underline{\rho} = R\underline{\alpha}$$

We might also model directly;

$$
\begin{aligned}
X_p &= \alpha_1 X_{p-1} + \alpha_2 X_{p-2} + \ldots + \alpha_p X_0 + \eta_p \\
X_{p+1} &= \alpha_1 X_p + \alpha_2 X_{p-1} + \ldots + \alpha_p X_1 + \eta_{p+1} \\
X_{2p-1} &= \alpha_1 X_{2p-2} + \alpha_2 X_{2p-3} + \ldots + \alpha_p X_{p-1} + \eta_{2p-1}
\end{aligned}
$$

Rewritten, in vector form:

$$\underline{\mathrm{x}} = H\underline{\alpha} + \underline{\eta}$$

How to Solve?

Least Squares:

$$J(\underline{\alpha}) := ||\underline{x} - H\underline{\alpha}||$$

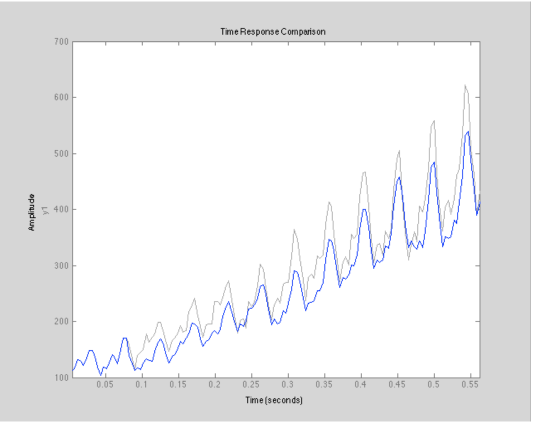
Understand the notation and terms.

$$\begin{aligned} dJ/d\underline{\alpha} &= 0 & (8)\\ &\Rightarrow H^T\underline{x} = H^T H\underline{\alpha} & (9)\\ &\Rightarrow \underline{\hat{\alpha}} = (H^T H)^{-1} H^T \underline{x} & (10) \end{aligned}$$

Least Squares Estimate using the Pseudo inverse. Stationary Point.

AR model identification

AR(10) model (YW): 1.0000 – 1.04140.03020.0952 – 0.0694 – 0.00770.1268 – 0.04630.03400.0185 – 0.1309
AR(10) model (LS): 1.0000 – 1.15230.4705 – 0.19670.3101 – 0.38860.2488 – 0.27410.4393 – 0.2832 – 0.1736



Parameter Estimators: What about Noise?

Least Square Estimate: $\underline{\hat{\alpha}} = (H^T H)^{-1} H^T \underline{x}$
There is a noise term $\underline{\eta}$. How to account for it?
Maximum Likelihood Estimate, when $\underline{\eta} \sim N(0, C_{XX})$:

$$\underline{\hat{\alpha}} = (H^T C_{xx}^{-1} H)^{-1} H^T C_{xx}^{-1} \underline{x} \tag{11}$$

With uncertainty estimate:

Bayesian Estimate

$$P(\underline{\alpha}|\underline{x}) \propto P(\underline{x}|\underline{\alpha})P(\underline{\alpha}) \tag{12}$$

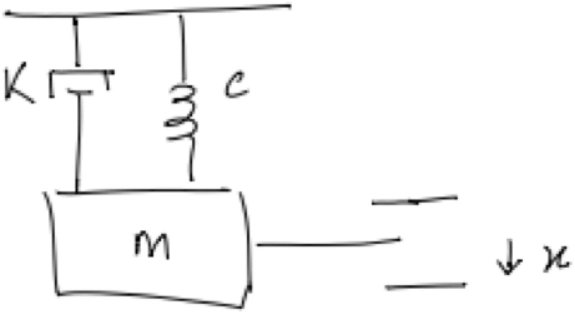
Maximum a posteriori (Bayes) estimate, when $\underline{\eta} \sim N(0, C_{XX})$ and $\underline{\alpha} \sim N(\underline{\bar{\alpha}}, C_{\alpha\alpha})$, of mean and covariance:

$$\begin{aligned}
\underline{\hat{\alpha}} &= \underline{\bar{\alpha}} + C_{\alpha\alpha}H^T(HC_{\alpha\alpha}H^T + C_{xx})^{-1}(\underline{x} - H\underline{\bar{\alpha}}) \\
\hat{C}_{\alpha\alpha} &= (H^T C_{xx}^{-1} H + C_{\alpha\alpha}^{-1})^{-1} \\
&= C_{\alpha\alpha} - C_{\alpha\alpha}H^T(HC_{\alpha\alpha}H^T + C_{xx})^{-1}HC_{\alpha\alpha}
\end{aligned}$$

Chalk talk: HOW IS THIS DERIVED?

A Physical Example

Exchange the positions of letters "K" ad "C" below!



Equations of Motion

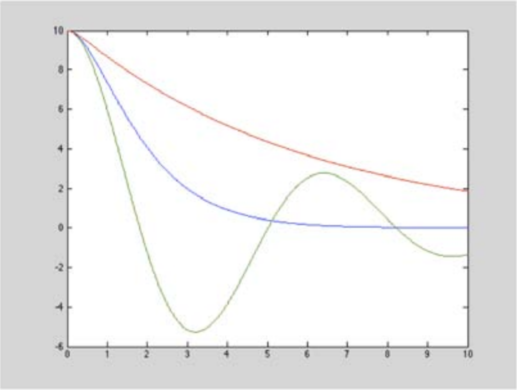
$$
\begin{aligned}
m\ddot{x} + c\dot{x} + kx &= F \qquad (13) \\
\ddot{x} + 2\eta\omega_0\dot{x} + \omega_0^2 x &= F \qquad (14)
\end{aligned}
$$

Where, the damping ratio $\eta = \frac{c}{2\sqrt{km}}$ and the natural frequency $\omega_0 = \sqrt{\frac{k}{m}}$ We write this in state-space form:

Spring Mass Damper System

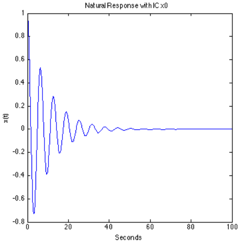
$$\dot{X} = AX \tag{15}$$

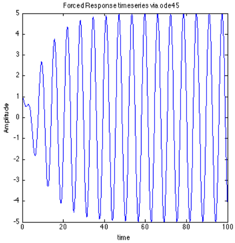
Chalk Talk: State-space Notation.



Spring Mass System

$$\dot{X} = AX + F \tag{16}$$





Discretization

We can solve

$$\dot{X} = AX + F \tag{17}$$

using numerical methods, e.g. Runge-Kutta methods (In matlab, ode45). Let's do demo.
But to understand, let's take an Euler discretization with zero-order hold of forcing. Then

$$\begin{aligned} A_d &= e^{A\Delta t} = \mathcal{L}^{-1}[(sI - A)^{-1}]_{t=\Delta t} && (18) \\ F_d &= A^{-1}(A_d - I)B && (19) \\ x_{t+\Delta t} &= A_d x_t + F_d && (20) \end{aligned}$$

We Just Say

$$x_{n+1} = M(x_n; \underline{\alpha}) \tag{21}$$

Recap

- Setup a Bayesian problem.
- Assumption of Linearity and Gaussianity.
- Specify the Objective.
- Minimize.
- Calculate Uncertainty.

There appeared to be some questions about Gaussian distributions.

Time Dependent Example

$$\underline{x}_{n+1} = M(\underline{x}_n; \underline{\alpha})$$
$$\underline{y}_n = H\underline{x}_n + \underline{\eta}$$

We have assumed the parameter vector is known constant, the model is deterministic, the observations are linearly related, but additively noisy and time-independent with $\underline{\eta} = N(\underline{0}, R)$. We are given a series of measurements $\underline{y}_0 \ldots \underline{y}_m$ and we are asked to estimate the initial condition $\underline{x}_0$. We may simply produce a least-squares function:

$$J(\underline{x}_0) := (\underline{x}_0 - \underline{x}_b)^T C_{00}^{-1} (\underline{x}_0 - \underline{x}_b) + \sum_{i=1}^{m} (\underline{y}_i - H\underline{x}_i)^T R^{-1} (\underline{y}_i - H\underline{x}_i)$$

Formulation

- There is some "background" state $\underline{x}_b$, from which the optimal initial conditions are a perturbation.
- There is a cost for departing from measurements and background, the former estimated from the known sensor noise.
- There's no Bayesian MAP formulation here.
- Even so, the objective isn't well-formulated. There aren't enough constraints to relate "future" measurements to the initial condition.
- But we do not know M, the model, provides such a constraint. Because this model is deterministic, we introduce the lagrange multiplier:

Objective

$$J(\underline{x}_0 := \frac{1}{2}(\underline{x}_0 - \underline{x}_b)^T C_{00}^{-1}(\underline{x}_0 - \underline{x}_b) +$$
$$\sum_{i=1}^{m} \left\{ \frac{1}{2}(\underline{y}_i - H\underline{x}_i)^T R^{-1}(\underline{y}_i - H\underline{x}_i) + \underline{\lambda}_i^T[\underline{x}_i - M(x_{i-1};\underline{\alpha})] \right\}$$

The solution must be a variation on the state space trajectory that the model constrains it to be in.

A Bayesian Perspective

$$P(\underline{x}_0|\underline{y}_1 \dots \underline{y}_n) \quad \propto \quad P(\underline{y}_n|\underline{x}_n)P(\underline{x}_n|\underline{x}_{n-1})\dots P(\underline{y}_1|\underline{x}_1)P(\underline{x}_1|\underline{x}_0) P(\underline{x}_0) \tag{22}$$

Assuming a perfect model has the effect of embedding the model directly. Thus we get

$$J(\underline{x}_0 := \frac{1}{2}(\underline{x}_0 - \underline{x}_b)^T C_{00}^{-1}(\underline{x}_0 - \underline{x}_b) + \sum_{i=1}^{m} \quad \frac{1}{2}(\underline{y}_i - HM(\underline{x}_{i-1}))^T R^{-1}(\underline{y}_i - HM(\underline{x}_{i-1}))$$

The solution must again be a variation on the state space trajectory that the model constrains it to be in.

Solution

$$\frac{dJ}{d\underline{x}_i} = \underline{\lambda}_i - \frac{\partial M^T}{\partial \underline{x}_i}\underline{\lambda}_{i+1} - H^T R^{-1}(\underline{y}_i - H\underline{x}_i);\ 0 < i < m$$

$$\frac{dJ}{d\underline{x}_m} = \underline{\lambda}_m - H^T R^{-1}(\underline{y}_m - H\underline{x}_m)$$

$$\frac{dJ}{d\underline{x}_0} = C_{00}^{-1}(\underline{x}_0 - \underline{x}_b) - \frac{\partial M^T}{\partial x_0}\underline{\lambda}_1$$

$$\frac{dJ}{d\underline{\lambda}_j} = \underline{x}_j - M(\underline{x}_{j-1}; \underline{\alpha});\ 0 < j \leq m$$

For a stationary point, we consider Euler-Lagrange equations.

"Forward Backward"

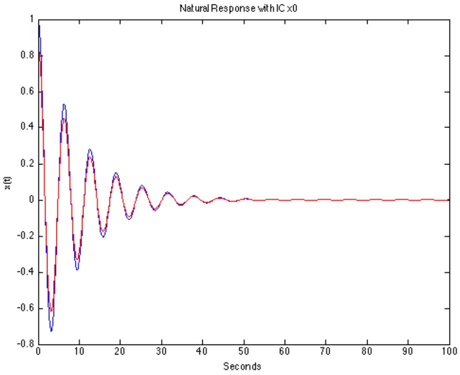
Forward(from $\underline{\lambda}$)

$$\begin{aligned} \underline{x}_i &= M(\underline{x}_{i-1}; \underline{\alpha}) \\ & 0 < i < M \end{aligned}$$

Backward(from $\underline{x}$)

$$\begin{aligned} \underline{\lambda}_m &= H^T R^{-1}(\underline{y}_m - H\underline{x}_m) \\ \underline{\lambda}_i &= \frac{\partial M^T}{\partial x_i}\underline{\lambda}_{i+1} + H^T R^{-1}(\underline{y}_i - H\underline{x}_i) \\ \hat{\underline{x}}_0 &= \underline{x}_b + C_{00}\frac{\partial M^T}{\partial x_0}\underline{\lambda}_1 \end{aligned}$$

On the Buggy


Natural Response with IC x0
x(t)
Seconds
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
10
20
30
40
50
60
70
80
90
100

Uncertainty?

Via Linearization

Forward(you'll need this in the end)

$$C_{ii} = \frac{\partial M}{\partial x_{i-1}} C_{i-1i-1} \frac{\partial M^T}{\partial x_{i-1}} \; 0 < i \leq m$$

What about backward? Convenient via information form:

$$\begin{aligned} \hat{I}_{mm} &= H^T R^{-1} H \\ \hat{I}_{ii} &= \frac{\partial M^T}{\partial x_i} I_{i+1i+1} \frac{\partial M}{\partial x_i} + H^T R^{-1} H \\ \hat{C}_{00} &= \left[C_{00}^{-1} + \frac{\partial M}{\partial x_0} \hat{I}_{11} \frac{\partial M^T}{\partial x_0} \right]^{-1} \end{aligned}$$

Here $\mathcal{L} = \frac{\partial M}{\partial x_i}$ is the Jacobian of M and $\mathcal{L}^T$ is its adjoint.

Example: Double Pendulum

C00 =

0.0012	0	0	0
0	0.0012	0	0
0	0	0.0012	0
0	0	0	0.0012

Ten seconds later:

C_{nn} =

0.0002	-0.0005	0.0001	0.0020
-0.0005	0.0130	0.0007	-0.0172
0.0001	0.0007	0.0003	-0.0000
0.0020	-0.0172	-0.0000	0.0409

Example: Double Pendulum

With a measurement every second with uncertainty same as C_{00}

$$\hat{C}_{00} = \text{1.0e-03 *} \begin{matrix} 0.0130 & -0.0033 & 0.0127 & -0.0009 \\ -0.0033 & 0.1259 & -0.0004 & -0.0148 \\ 0.0127 & -0.0004 & 0.0268 & -0.0061 \\ -0.0009 & -0.0148 & -0.0061 & 0.2176 \end{matrix}$$

Issues

Key components: Propagating forward, propagating backward, updating . . . to convergence.

1. There are huge dimensionality issues! - Monte-carlo, spectral and multiscale methods.
2. What about the jacobian and adjoint calculations? -Statistical approximations.
3. What about non-Gaussian uncertainties? -Bayesian inference
4. Can we not linearize? -Yes
5. What about model error? -Bayesian approaches.
6. How about a full fixed interval estimate? -Natural extension.

Unknown Parameter

We've assumed the model is perfect. How would we deal with stochastic forcing, parameterization, or uncertain parameters? Let's look at parameter case, it is now an unknown constant of the simulation:

$$
\begin{aligned}
J(\underline{x}_0, \underline{\alpha}) :=& \frac{1}{2}(\underline{x}_0 - \underline{x}_b)^T C_{00}^{-1} (\underline{x}_0 - \underline{x}_b) \\
& + \sum_{i=1}^{m} \left\{ \frac{1}{2}(\underline{y}_i - H\underline{x}_i)^T R^{-1} (\underline{y}_i - H\underline{x}_i) + \underline{\lambda}_i^T \left[\underline{x}_i - M(x_{i-1}; \underline{\alpha})\right] \right\} \\
& + (\underline{\alpha} - \underline{\alpha}_b)^T C_{\alpha\alpha}^{-1} (\underline{\alpha} - \underline{\alpha}_b)
\end{aligned}
$$

Gradients

$$\frac{dJ}{d\underline{x}_i} = \underline{\lambda}_i - \frac{\partial M^T}{\partial \underline{x}_i}\underline{\lambda}_{i+1} - H^T R^{-1}(\underline{y}_i - H\underline{x}_i);\ 0 < i < m$$

$$\frac{dJ}{d\underline{x}_m} = \underline{\lambda}_m - H^T R^{-1}(\underline{y}_m - H\underline{x}_m)$$

$$\frac{dJ}{d\underline{x}_0} = C_{00}^{-1}(\underline{x}_0 - \underline{x}_b) - \frac{\partial M^T}{\partial x_0}\underline{\lambda}_1$$

$$\frac{dJ}{d\underline{\lambda}_j} = \underline{x}_j - M(\underline{x}_{j-1}; \underline{\alpha});\ 0 < j \leq m$$

$$\frac{dJ}{d\underline{\alpha}} = C_{\alpha\alpha}^{-1}(\underline{\alpha} - \underline{\alpha}_b) - \sum_i \frac{\partial M^T}{\partial \alpha}\underline{\lambda}_i$$

Joint state-parameter estimate, is typically difficult, often approached by interleaving. When the parameters don't change, as here, then the solution may avoid overfitting issues. Sometimes parameter errors are referred to as model errors, but this is not strictly correct.

What about model error?

$$\underline{x}_{n+1} = M(\underline{x}_n; \underline{\alpha}) + \underline{w}$$
$$\underline{y}_n = H\underline{x}_n + \underline{\eta}$$

Under Gaussian conditions,$\underline{w} \sim N(0, C_{ww})$, say. Then, we can't impose a Lagrange multiplier as we've done, why?

Still, least-squares formulations can be constructed, but it is better to think about the problem in Bayesian terms, for the fixed point:

$$P(\underline{x}_0|\underline{y}_1) \propto P(Y_1 = \underline{y}_1|X_1 = \underline{x}_1) \sum_{X_1} P(X_1 = \underline{x}_1|X_0 = \underline{x}_0)P(X_0 = \underline{x}_0)$$

Missing Data

0.4414	0.3793	0.5958	0.2166	-0.1713
0.8154	1.0269	2.0754	-0.2729	-0.6826
-1.0004	-0.2160	-1.5771	-0.0468	-1.8362
-0.1735	1.5519	-0.8164	-0.7408	-0.0367
0.2328	0.5149	-0.0080	0.4749	0.0625
NaN	0.8682	-0.1971	0.4203	0.0107
NaN	-0.8680	-0.4592	0.1704	0.9506
NaN	0.8888	-0.8460	0.4277	0.2475
NaN	0.3925	-1.3717	1.4422	0.7604
NaN	-0.8056	0.5971	-0.8732	0.6066

Estimating a Model from Data

$$\underline{y_i} = M\underline{x_i} + \underline{n} \tag{23}$$

$$\underline{z_i} = [\underline{x_i}^T \underline{y_i}^T]^T \tag{24}$$

$$P(M|\underline{z_i}) \propto P(\underline{z_i}|M)P(M) \tag{25}$$

$$= P(\underline{y_i}|\underline{x_i}, M)P(\underline{x_i}|M)P(M) \tag{26}$$

$$= P(\underline{y_i}|\underline{x_i}, M)P(\underline{x_i}) \tag{27}$$

We assume an uninformative prior on the model and calculate the MLE.

The Objective

$$J(M) := \frac{1}{N}\sum_i (\underline{y_i} - M\underline{x_i})^T C_{yy}^{-1} (\underline{y_i} - M\underline{x_i}) + (\underline{x_i} - \underline{\bar{x}})^T C_{xx}^{-1} (\underline{x_i} - \underline{\bar{x}}) \quad (28)$$

The Stationary Point

$$dJ/dM = 0 \quad (29)$$

$$\frac{1}{N}\sum_i C_{yy}^{-1}(\underline{y_i} - M\underline{x_i})\underline{x_i}^T = 0 \quad (30)$$

The Linear Model

$$
\begin{aligned}
dJ/dM &= 0 && (31)\\
\frac{1}{N}\sum_i C_{yy}^{-1}(\underline{y_i} - M\underline{x_i})x_i^T &= 0 && (32)\\
< \underline{y_i}\underline{x_i}^T > &= M < \underline{x_i x_i}^T > && (33)\\
M &= C_{yx}C_{xx}^{-1} && (34)
\end{aligned}
$$

Where, x_i and y_i are zero-mean variables, wlog.

The Missing Data Estimate and Uncertainty

Estimate:

$$M = C_{yx} C_{xx}^{-1} \quad (35)$$

$$\underline{\hat{y}} = C_{yx} C_{xx}^{-1} \underline{\hat{x}} \quad (36)$$

$$(37)$$

Uncertainty follows from objective:

$$J(\underline{x}_i) := (\underline{y_i} - M\underline{x_i})^T C_{yy}^{-1} (\underline{y_i} - M\underline{x_i}) + (\underline{x_i} - \underline{\bar{x}_i})^T C_{xx}^{-1} (\underline{x_i} - \underline{\bar{x}_i}) \quad (38)$$

Fisher Information

Uncertainty Estimate:

$$d^2J/d\underline{x_i}^2 \quad = \quad M^T C_{yy}^{-1} M + C_{xx}^{-} 1 \tag{39}$$

$$\hat{C}_{yy} \quad = \quad (C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} C_{xx}^{-1} + C_{xx}^{-1})^{-1} \tag{40}$$

$$\tag{41}$$

Uncertainty follows from objective:

$$J(\underline{x}_i) \quad := \quad (\underline{y_i} - M\underline{x_i})^T C_{yy}^{-1} (\underline{y_i} - M\underline{x_i}) + (\underline{x_i} - \underline{\bar{x}_i})^T C_{xx}^{-1} (\underline{x_i} - \underline{\bar{x}_i}) \tag{42}$$

Application to Missing Information

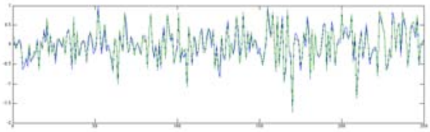
- Step 1: Estimate the full covariance C_{zz} from available data. Update the Linear Model from its components C_{yx} and C_{xx}.
- Step 2: Estimate the missing data from the most recent model.
- Repeat

This technique generalizes to Expectation-Maximization, and is a very interesting way of solving, often complex, Bayesian inference problems.
READING: Reg-EM.

Infilled

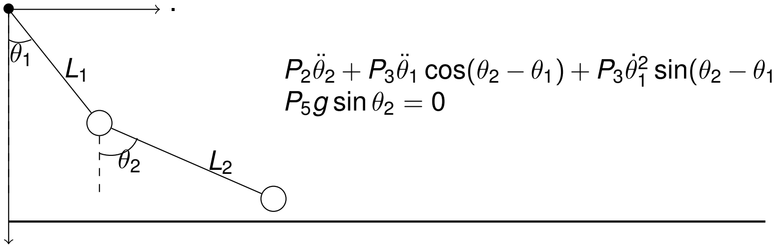
Noise iid, with std. 0.1

0.2081	0.8682	-0.1971	0.4203	0.0107
-0.1081	-0.8680	-0.4592	0.1704	0.9506
-0.0021	0.8888	-0.8460	0.4277	0.2475
0.0879	0.3925	-1.3717	1.4422	0.7604
-0.0376	-0.8056	0.5971	-0.8732	0.6066



Towards a Nonlinear World.

$$P_1\ddot{\theta}_1 + P_3\ddot{\theta}_2\cos(\theta_2-\theta_1) - P_3\dot{\theta}_2^2\sin(\theta_2-\theta_1) + P_4 g\sin\theta_1 = 0$$



$$P_2\ddot{\theta}_2 + P_3\ddot{\theta}_1\cos(\theta_2-\theta_1) + P_3\dot{\theta}_1^2\sin(\theta_2-\theta_1) + P_5 g\sin\theta_2 = 0$$

Let,

$$P_1 = (m_1+m_2)L_1^2 \qquad P_4 = (m_1+m_2)L_1$$
$$P_2 = m_2L_2^2 \qquad P_5 = m_2L_2$$
$$P_3 = m_2L_1L_2$$

12.S990 Quantifying Uncertainty
Fall 2012