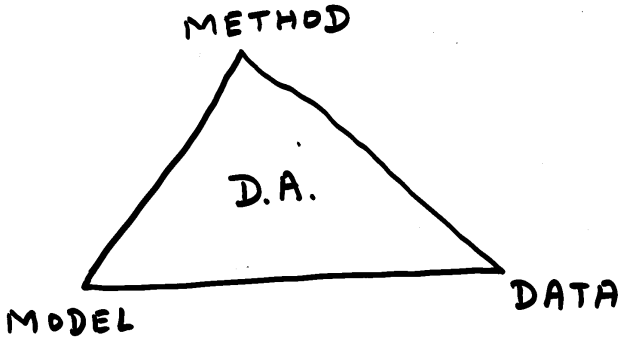
DATA ASSIMILATION: AN OVERVIEW

S. LAKSHMIVARAHAN

ASSIMILATE - TO BE OBSORBED AND INCORPORATED

ASSIMILATION - ACT OF BRINGING OR COMING TO A RESEMBLANCE



- FITTING MODELS TO DATA
- WHY DO THIS? - TO PREDICT THE BEHAVIOR OF SYSTEM.

- GAUSS (1777-1855):
  - LEAST SQUARE METHOD
  - FITTED A MODEL TO THE TRAJECTORY
  - USED IT IN PREDICTING FUTURE POSITION VERY ACCURATELY

- A WELL FITTED MODEL $\not\Rightarrow$ GOOD PREDICTION

- NOT EVERY SYSTEM IS PREDICTABLE TO THE SAME DEGREE

<u>NOTE</u>: AN ASTRONOMER CAN PREDICT VERY PRECISELY THE POSITION OF JUPITER AND ITS MOON AT 11:30 PM TONIGHT BUT HE WILL HAVE NO CLUE WHERE HIS TEEN AGE DAUGHTER WILL BE AT THAT TIME.

- MODEL
  - STATIC
  - DYNAMIC
- MODEL
  - EMPIRICAL
  - UNDERLYING SCIENCE

EMPIRICAL

- REGRESSION
- TIME SERIES
- MEDICINE
- FINANCE
- METEOROLOGY
  - EL NINO/LANINA USING SOI

UNDERLYING SCIENCE

- METEOROLOGY
  - SHALLOW WATER
  - MIXED LAYER
  - PRIMITIVE EQN
- INVERSE PROBLEM IN GEO PHYSICS

⇒ KNOWN WITH UNKNOWN PARAMETERS

- IN DYNAMIC CASE
  INITIAL CONDITION
  BOUNDARY CONDITION

<u>NOTE</u>:- USE OF DETERMINISTIC MODELS IN WEATHER PREDICTION.

INITIAL/BOUNDARY VALUE PROBLEM

DATA

- TIME SERIES
  - TEMPERATURE IN OKC AIRPORT
- SPATIAL DATA
  - RAIN in OK ON MAY 3rd
- SPACE-TIME
  - RAIN FALL IN OK FOR EACH DAY IN OCTOBER

- DATA **DISTRIBUTION** - GAPS
  - VARIATION OF DENSITY

- **ERRORS** IN DATA / OBSERVATION
  - BALOONS
  - RADAR
  - SHIPS / PLANES
  - GROUND STATIONS
  - SATELLITE
  - OIL PLATFORMS

- OVERALL DESIGN - WHAT TO / HOW TO MEASURE?

- RARE EVENTS

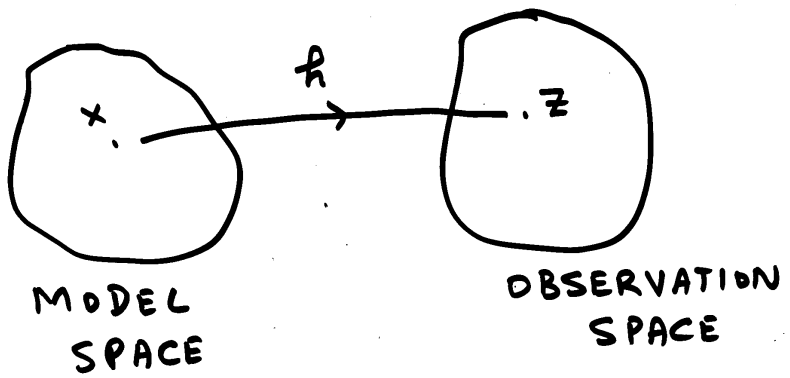
ASSIMILATION - STATIC MODELS

- $z = (z_1, z_2, \cdots z_m)^T \in R^m$

  z_i - VELOCITY, POSITION, RAIN
  TEMP., ENERGY, VOLTAGE,
  REFLECTIVITY, INTEREST RATE
  EXCHANGE RATE

- $x = (x_1, x_2, \cdots x_n)^T \in R^n$ - MODEL STATE

MODEL: $h : R^n \to R^m$



$$z = h(x)$$

- GIVEN $h(\cdot)$ AND z, FIND x
- INVERSE PROBLEM
- $h(\cdot)$ - NONLINEAR, NOT ONE-TO-ONE
  $h^{-1}(\cdot)$ - MAY NOT EXIST

EXAMPLES

x	z	$h(\cdot)$
· RAIN	· REFLECTIVITY	· RADAR METEOROLOGY
· TEMP	· RADIATED ENERGY IN A GIVE SLICE OF THE SPECTRUM	· RADIATION PHYSICS STEFAN'S LAW
· SPEED	VOLTAGE	· FARADAY'S LAW

IN PRACTICE:

$$Z = h(x) + v$$

MEASUREMENT ERROR

- LINEAR
- NONLINEAR
- MAY NOT HAVE AN INVERSE

$\Downarrow$

"GENERALIZED INVERSE"

$E(v) = 0$

$E(vv^T) = \Sigma_v$

- DIAGONAL
- NON DIAGONAL
- KNOWN

FORMULATION

$Z = h(x) + V$

Z - OBSERVABLE

V - IS NOT OBSERVABLE

$Z \in R^m \quad x \in R^n$

$m : n$

$m > n$

- OVER DETERMINED

$\Downarrow$

- INCONSISTENT

$\Rightarrow$ NO SOLUTION

$\Downarrow$

- RESIDUAL $r(x) = z - h(x)$
- MINIMIZE THE "<u>SIZE</u>" OF $r(x)$

$m = n$

$m < n$

- UNDER DETERMINED

$\Downarrow$

- INFINITE SOLUTIONS

$\Downarrow$

$r(x) = 0$

- INDUCE UNIQUENESS BY SOME REGULARITY CONDITION
- MIN. $\|x\|$

CRITERION: RESIDUAL + REGULARITY

$$f(x) = f_0(x) + f_R(x)$$

$$f(x) = f_o(x) + f_R(x)$$

$$f_o(x) = \frac{1}{2}\left[z - h(x)\right]^T \Sigma_V^{-1} \left[z - h(x)\right]$$

$$f_R(x) = \frac{\alpha}{2} \|x\|^2 = \frac{\alpha}{2} x^T x$$

$$= \frac{1}{2}(x - x_b)^T B^{-1} (x - x_b)$$

$$= \frac{\alpha}{2} \|\phi(x)\|^2 \qquad \phi(x) = 0$$

CONSTRAINT

- PENALTY FUNCTION
- CONSTRAINTS OFTEN IMPLY BALANCE CONDITIONS

<u>PROBLEM</u>: GIVEN z, Σ_V, $h(\cdot)$

$\{\alpha, x_b, B, \phi(\cdot)\}$

FIND x THAT <u>MINIMIZES</u> $f(x)$

$$h(\cdot) \begin{cases} \text{LINEAR} \\ \text{NONLINEAR} \end{cases} \qquad \begin{matrix} m \approx 10^5 \\ n \approx 10^7 \end{matrix}$$

DYNAMIC CASE

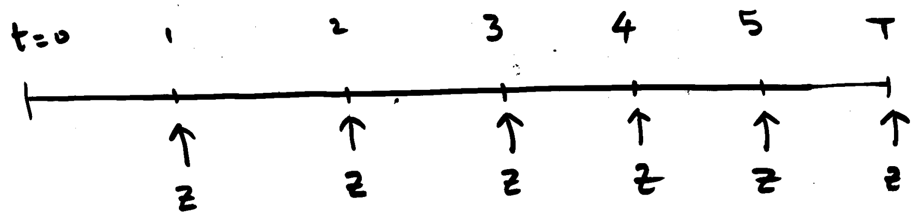
$X = (x_1, x_2, x_3)^T \quad t \geq 0$

$$\eta(x,t) = \begin{bmatrix} \eta_1(x,t) \\ \eta_2(x,t) \\ \vdots \\ \eta_K(x,t) \end{bmatrix} \begin{array}{l} \to u\text{ - VELOCITY} \\ \to v\text{ - VELOCITY} \\ \to p\text{ - pressure} \\ \to T\text{ - TEMP} \\ \to \text{MOISTURE} \end{array}$$

$$\frac{\partial \eta}{\partial t} = F[\eta, \alpha, t]$$

$\alpha \to$ parameters

- $x \in D$, DOMAIN - FINITE/INFINITE
- $\eta(x,0)$ - INITIAL CONDITION
- $\eta[x,t]$ = GIVEN FOR $x \in$ BOUNDARY OF D
- GIVEN I.C./B.C., WE CAN IN PRINCIPLE INTEGRATE THE MODEL NUMERICALLY.
- I.C./B.C. ARE TO BE <u>ESTIMATED</u> USING DATA.



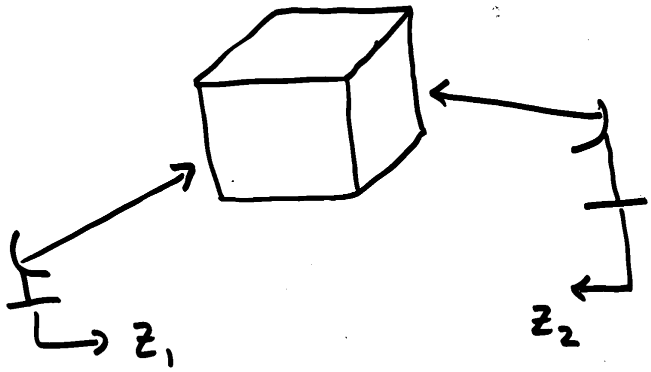
$$\frac{1}{2}\sum_{i=1}^{T}\sum_{x_i \in D}\left[z_i - \eta(x_i, i)\right]^T\left[z_i - \eta(x_i, i)\right]$$

$= f(\eta(x,0))$ - MEASURE OF MIS-FIT BETWEEN MODEL AND DATA

- MODEL SOLUTION DEPENDS ON THE I.C.

<u>IDEA:-</u>
- START WITH $\eta(x,0)$
- COMPUTE SOLUTION $\eta(x,t)$
- COMPUTE $f(\eta(x,0))$
- COMPUTE $\nabla f(\eta(x,0))$
- USE IN A MINIMIZATION SCHEME TO FIND THE BEST $\eta(x,0)$

MULTI SENSOR PROBLEM



x : $h(x)$ - MODEL COUNTER PART

$z_1 \simeq h(x)$ $z_2 \simeq h(x)$

QUESTION : FIND x : IT FITS BOTH SET OF OBSERVATIONS

- METR 5803 DYNAMIC DATA ASSIMILATION

SPRING 2003

Book: DYNAMIC DATA ASSIMILATION: A LEAST SQUARE APPROACH by J. LEWIS, S. LAKSHMIVARAHAN, S.K. DHALL