Module – 3.1

STATIC, DETERMINISTIC LINEAR INVERSE (WELL – POSED) PROBLEM

S. Lakshmivarahan

School of Computer Science

University of Oklahoma

Norman, Ok – 73069, USA

varahan@ou.edu

PROBLEM STATEMENT – A ST.LINE PROBLEM

- A particle is moving in a st. line:
  - Constant velocity, V – Not known
  - Initial position, Z_0 – Not known
- Observations of position Z_i at time t_i for $1 \le i \le m$ are available

TIME	t_1	t_2	… t_i…	T_m
POSITION	Z_1	Z_2	… Z_i …	Z_m

- Problem: Given the pair (t_i, Z_i), $1 \le i \le m$, estimate the unknowns Z_0, V

BUILD A LINEAR MODEL

- To enable estimation of the unknowns, we need to build a relation – called the model between the known and unknowns
- From basic Physics relating time and motion:

$$Z_i = Z_0 + Vt_i \qquad (1)$$

must hold for each $1 \leq i \leq m$

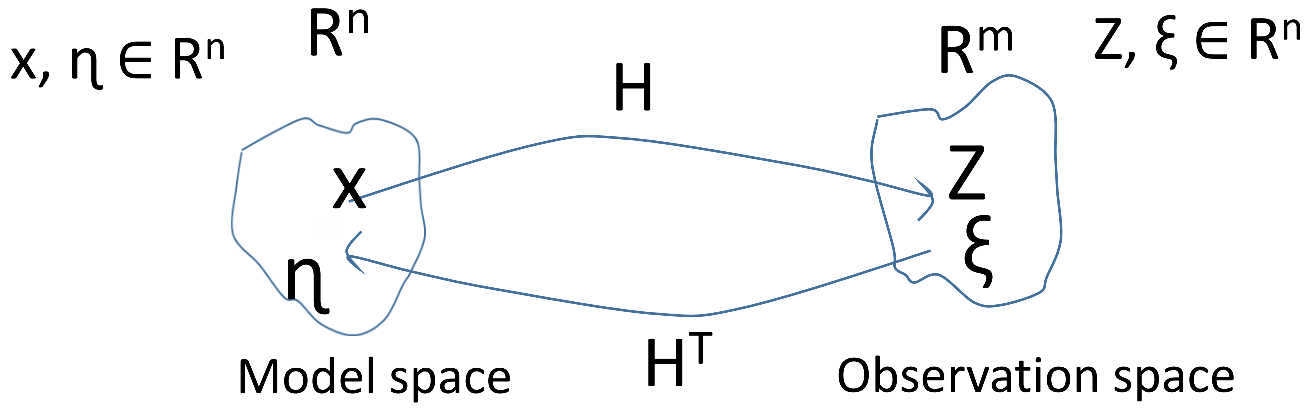
- In matrix-vector notation (6.1) becomes

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_i \\ \vdots \\ Z_m \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_i \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} Z_0 \\ V \end{bmatrix} = Hx \qquad (2)$$

- Or $Z = Hx \qquad Z \in R^m,\ H \in R^{mx2},\ x \in R^2 \qquad (3)$
- Equation (3) is a linear model
- Given (Z, H), find x, is the linear inverse problem

A GENERALIZATION – LINEAR MODEL

- Let $Z \in R^m$ be the <u>observation vector</u>
- R^m is called the <u>observation space</u>
- Let $x \in R^n$ be the <u>unknown vector</u>
- R^n is called the <u>model space</u>
- $H \in R^{mxn}$ is the relation between the model space and observation space



$$Z = Hx$$
$$\eta = H^T\xi$$

ON SOLVING Z = Hx

- When m = n and H is non-singular, then

$$x = H^{-1}Z \qquad (4)$$

- When $m \neq n$, H is a rectangular matrix and the standard notion of non-singularity does not apply
- Two cases arise:

  $m > n$ – overdetermined case – Inconsistent case

  $m < n$ – underdetermined case – Infinity many solution

OVERDETERMINED CASE: m > n

- m = 3 and n = 2, H = $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$
- Columns are H are linearly independent
- SPAN(H) = 2-D plane defined by these two columns which is a subset of R^3
- Let Z = $\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$, since Z = (-1)$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ + 1$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, Z ∈ SPAN(H)
- Z = Hx has a solution x = $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

INCONSISTENT CASE: m > n

- Recall that columns of H are defined by the mathematical model but the column Z of observation that come from the real world measurement
- Generally, observations have noise embedded in them and models are only approximations to reality
- Hence, move often than not, Z does not belong to the SPAN(H)
- In such cases Z = HX has no solution in the sense that there is no vector x that will satisfy equation Z = Hx

ANOTHER LOOK AT INCONSISTENT CASE: m > n

- m = 3, n = 2, H = $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$
- Z = (2, 3.5, 4.2)^T
- Z = Hx => $x_1 + x_2 = 2$, $x_1 + 2x_2 = 3.5$, $x_1 + 3x_2 = 4.2$
- Verify that $x_1 = \frac{1}{2}$ and $x_2 = \frac{3}{2}$ is the solution of the first two, but this does not satisfy the third
- Verify that solution of any two out of these three equations, does not satisfy the remaining equation
- In this sense there is no solution to Z = Hx when m > n

UNDERDETERMINED CASE: m < n

- m = 2, n = 3, H = $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \end{bmatrix}$
- Z = Hx becomes

$$Z_1 = x_1 + 2x_2 + 3x_3$$

$$Z_2 = x_1 + 4x_2 + 5x_3$$

- Rewrite:

$$x_1 + 2x_2 = Z_1 - 3x_3$$

$$x_1 + 4x_2 = Z_2 - 5x_3$$

- For each $x_3 \in R$, there is a pair $(x_1(x_3), x_2(x_3))^T$ that is the solution of this pair
- Z = Hx has infinite solution $(x_1(x_3), x_2(x_3), x_3)^T$
- Hence, there is no uniqueness in this case when m < n

SUMMARY – LINEAR INVERSE PROBLEM

- $Z = Hx$ and H is of full rank

$m : n$

$m > n$

- Overdetermined
- $Rank(H) = n$
- Inconsistent system
- No solution

$m = n$

- $Rank(H) = n$
- Unique solution $X = H^{-1}z$

$m < n$

- Underdetemined
- $Rank(H) = m$
- Infinitely many solution
- No uniqueness

- Thus, we need to generalize the concept of solution for the two extreme cases when $m > n$ and $m < n$
- This generalized solution in called the <u>lease square solution</u>

UNWEIGHTED LEAST SQUARES SOLUTION: m > n

- Define $\Lambda(x) = Z - Hx \in R^m$ – residual vector <- $\mathrm{r}(x)$
- Recall when m > n, there is no $x \in R^n$ for which $\mathrm{r}(x) = 0$
- As a compromise, we seek $x \in R^n$ for which the vector $\mathrm{r}(x)$ will have a minimum length
- To this end, define $f(x) = \|\mathrm{r}(x)\|_2^2 = \mathrm{r}^T(x)\,\mathrm{r}(x) = \sum_{i=1}^{n} \mathrm{r}_i^2(x)$ which is the square of the norm of the residual
- $\mathrm{r}_i(x) = Z_i - H_{i*}x$ where H_{i*} is the i^{th} row of H

  = i^{th} component of the residual vector
- Hence, f(x) = sum of the squares of the components of the residual vector
- The vector $x \in R^n$ that minimizes f(x) is called the least squares solution

LEAST SQUQARES METHOD: m > n

- $f(x) = r^T(x)r(x) = (Z - Hx)^T(Z - Hx)$

  $= (Z^T - (Hx)^T)(Z - Hx)$

  $= (Z^T - x^T H^T)(Z - Hx)$

  $= Z^T Z - Z^T Hx - x^T H^T Z + x^T(H^T H)x \qquad (5)$
- $Z^T Hx$ being a scalar: $Z^T Hx = (Z^T Hx)^T$

  $= x^T H^T Z \qquad (6)$
- Therefore, $f(x) = Z^T Z - 2Z^T Hx + x^T(H^T H)x \qquad (7)$
- Find x that minimizes f(x) in (7)

H^TH SPD WHEN H IS OF FULL RANK

- Since $H^TH = (H^TH)^T$, H^TH is symmetric
- Consider $x^T(H^TH)x = (x^TH^T)(Hx) = (Hx)^T(Hx)$

$$= ||Hx||_2^2 \qquad (8)$$

- Since m > n, Rank(H) = n and the columns of H are linearly independent
- That is, Hx = 0 exactly when x = 0

  ≠ 0 otherwise

- Hence $x^T(H^TH)x > 0$ for $x \neq 0$

  = 0 only when x = 0 } -> (9)

- (H^TH) is positive definite

GRADIENT AND HESSIAN OF f(x)

- Refer to f(x) in (7)
- $\nabla_x(Z^T Z) = 0$, $\nabla_x^2(Z^T Z) = 0$
- $\nabla_x(2Z^T Hx) = 2\nabla_x(a^T x)$ with $a = H^T Z$

  $= 2a = 2H^T Z$
- $\nabla_x^2(2Z^T Hx) = 0$
- $\nabla_x(x^T(H^T H)x) = 2(H^T H)x$
- $\nabla_x^2(x^T(H^T H)x) = 2H^T H$ – SPD
- Combining
- Gradient of f = $\nabla_x f(x) = -2H^T Z + 2(H^T H)x$ -> (10)
- Hessian of f = $\nabla_x^2 f(x) = 2(H^T H)$ -> (11)

UNCONSTRAINED MINIMIZATION OF f(x) – NORMAL EQUATION

- Setting $\nabla_x f(x) = -2H^T Z + 2(H^T H)x = 0$

- Least square solution is the solution of the Normal equations which is linear symmetric, positive definite system: $(H^T H)x = H^T Z$ -> (12)

- Or $X_{ls} = (H^T H)^{-1} H^T Z = H^+ Z$ -> (13)

  $H^+ = (H^T H)^{-1} H^T$ – Generalized inverse of H -> (14)

- Since the Hessian $\nabla_x^2 f(x) = 2(H^T H)$ is SPD, f(x) is a convex function and hence the minimum is unique

MINIMUM RESIDUAL

- The minimum residual $r(x_{LS}) = Z - Hx_{LS}$
- By (14): $r(x_{LS}) = [I - H(H^TH)^{-1}H^T]Z \neq 0$ -> (15)
- Here in lies the difference between the classical solution where $r(x) = 0$ and the least squares solution where $r(x_{ls}) \neq 0$ for the overdetermined case
- Verify $f(x_{ls}) = \|r(x)\|_2^2 = Z^T[I - H(H^TH)^{-1}H^T]Z$ -> (16)

  which is the minimum value of sum of square errors (SSE)

AN ILLUSTRATION – ST.LINE PROBLEM

- H = $\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}$

- $H^T H = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum_{i=1}^{m} t_i \\ \sum_{i=1}^{m} t_i & \sum_{i=1}^{m} t_i^2 \end{bmatrix}$

- $H^T Z = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} Z_i \\ \sum_{i=1}^{m} Z_i t_i \end{bmatrix}$

ILLUSTRATION CONTINUED

- Normal equations: $(H^T H)x = H^T Z$

$$\begin{bmatrix} m & \sum_{i=1}^{m} t_i \\ \sum_{i=1}^{m} t_i & \sum_{i=1}^{m} t_i^2 \end{bmatrix} \begin{bmatrix} Z_0 \\ V \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} Z_i \\ \sum_{i=1}^{m} Z_i t_i \end{bmatrix}$$

- Dividing by n => $\begin{bmatrix} 1 & \overline{t} \\ \overline{t} & \overline{t^2} \end{bmatrix} \begin{bmatrix} Z_0 \\ V \end{bmatrix} = \begin{bmatrix} \overline{Z} \\ \overline{Zt} \end{bmatrix}$

$$\overline{t} = \frac{1}{m} \sum_{i=1}^{m} t_i \ , \ \overline{t^2} = \frac{1}{m} \sum_{i=1}^{m} t_i^2, \ \overline{Z} = \frac{1}{m} \sum_{i=1}^{m} Z_i, \ \overline{Zt} = \frac{1}{m} \sum_{i=1}^{m} Z_i t_i$$

- <u>Solution:</u> $V^* = \frac{\overline{Zt} - \overline{t}\,\overline{Z}}{\overline{t^2} - (\overline{t})^2}$

  $Z^* = \overline{Z} - \overline{t} V^*$

- SSE = $f(Z_0^*, V^*) = \sum_{i=1}^{m} [Z_i - (Z_0^* + V^* t_i)]^2$ is the minimum value of the sum of squared errors

- RMS error = $[\frac{SSE}{m}]^{\frac{1}{2}} = [\frac{f(Z_0^*, V^*)}{m}]^{1/m}$ is a measure of the linear fit

NUMERICAL EXAMPLE – ALGEBRAIC METHOD

- m = 4, n = 2, H = $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$, Z = $\begin{bmatrix} 1.0 \\ 3.0 \\ 2.0 \\ 3.0 \end{bmatrix}$
- $\bar{t}$ = 1.5, $\overline{t^2}$ = 3.5, $\bar{Z}$ = 2.25, $\overline{Zt}$ = 4
- Normal equation: $\begin{bmatrix} 1 & 1.5 \\ 1.5 & 3.5 \end{bmatrix} \begin{bmatrix} Z_0 \\ V \end{bmatrix} = \begin{bmatrix} 2.25 \\ 4 \end{bmatrix}$
- Solution: V* = 0.5, Z_0* = 1.5
- Filted/assimilated model: $Z_i = 1.5 + 0.5t_i$
- SSE = 1.5, RMS error = 0.6124

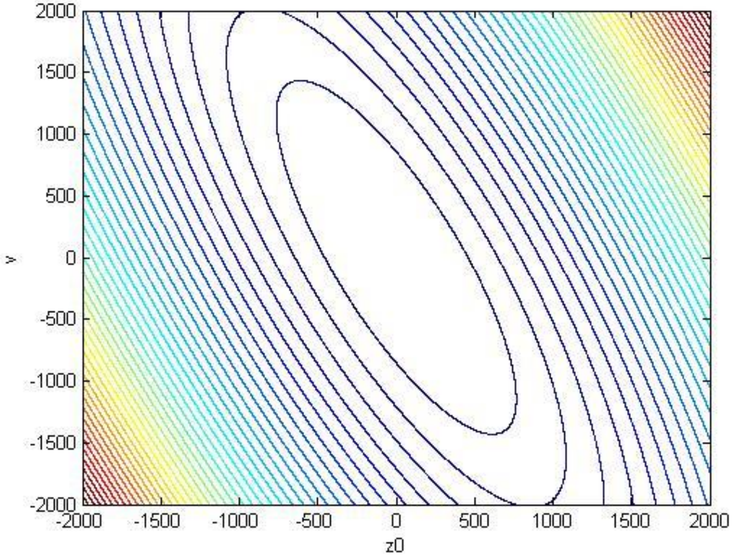
CONTOURS OF f(x) – GRAPHICAL METHOD

- Using the data in slide (19) we can get

$$f(Z_0, V) = Z^TZ - 2Z^THx + x^T(H^TH)x$$

$$= Z_0^2 + 3Z_0V + 3.5V^2 - 9Z_0 - 25V + 23$$

- The contours of $f(Z_0, V)$ using MATLAB is given below
- The minimum is $Z_0^* = 1.5$, $V^* = 0.5$



WEIGHTED LEAST SQUARES: m > n

- Let $W \in R^{mxm}$ be a SPD matrix
- The weighted sum of squared errors:

$$f_w(x) = (Z - Hx)^T W(Z - Hx)$$

- W – could be a diagonal matrix with different weights along the diagonal or a general SPD
- Verify that the normal equations in this case is

$$(H^T WH)x = H^T WZ$$

- The weighted least square solution is:

$$X_{ls} = (H^T WH)^{-1} H^T WZ \qquad \text{-> (17)}$$

UNDERDETERMINED CASE: $m < n$

- Recall: There are infinitely many solutions
- $r(x) = 0$ for infinitely many $x \in R^n$
- Unlike when $m > n$, in this case $f(x) = \|r(x)\|_2^2 = 0$
- Need a new approach
- To get an unique solution, formulate it as a constrained minimization problem using the standard Lagrangian multiplier methods for equality constrained problem (Module 5)

LAGRANGIAN FORMULATION: $m < n$

- Problem statement: Find $x \in R^n$ such that $||x||^2$ is a minimum when Z satisfies $Z = Hx$
- Let $\lambda \in R^m$ and define the Lagrangian

$$L(x, \lambda) = ||x||^2 + \lambda^T(Z - Hx) \qquad \text{-> (18)}$$

- Now the above constrained minimization is solved by minimizing $L(x, \lambda)$ with respect to $x \in R^n$ and $\lambda \in R^m$ as an unconstrained problem

LAGRANGIAN METHOD: m < n

- A necessary conditions for the minimum are:

$$\nabla_x L(x, \lambda) = 0$$

$$\nabla_\lambda L(x, \lambda) = 0$$

- By solving these two equations in the two unknowns x, λ, we get the optimal x and λ
- For L in (18)

$$\left.\begin{aligned} \nabla_x L(x, \lambda) &= 2x - H^T\lambda = 0 \\ \nabla_\lambda L(x, \lambda) &= Z - Hx = 0 \end{aligned}\right\} \quad \text{-> (19)}$$

LEAST SQUARES SOLUTION: m < n

- Solving (19): $x = \frac{1}{2}H^T\lambda$ -> (20)

  $Z = Hx = \frac{1}{2}HH^T\lambda$ -> (21)

- From (21): $\lambda = 2(HH^T)^{-1}Z$ -> (22)
- Using (22) in (19)

  $X_{Ls} = H^T(HH^T)^{-1}Z$ -> (23)

- If H is of full rank, Rank(H) = m then it can be verified (HH^T) is SPD
- X_{LS} is computed in two steps:
  - Solve normal equations: $(HH^T)y = Z$ and find $y = (HH^T)^{-1}Z$
  - $X_{Ls} = H^Ty$

RESIDUAL AT X_{LS}

- $r(x_{LS}) = Z - Hx_{LS}$

  $= Z - HH^T(HH^T)^{-1}Z$

  $= Z - Z = 0$

- This is to be expected since we start with the infinitely many solutions for which $r(x) = 0$

EXERCISES

6.1) Let $x_1 + x_2 = 1$, $x_1 + 2x_2 = 3.5$, $x_1 + 3x_2 = 4.2$

Solve any two and verify that this solution is not consistent with the third equation

6.2) Solve $\begin{bmatrix} 1 & \bar{t} \\ \bar{t} & \overline{t^2} \end{bmatrix} \begin{bmatrix} Z_0 \\ V \end{bmatrix} = \begin{bmatrix} \bar{Z} \\ \overline{Zt} \end{bmatrix}$

and verify that the solution is given: $V^* = \frac{\overline{Zt} - \bar{t}\,\bar{Z}}{\overline{t^2} - (\bar{t})^2}$, $Z^* = \bar{Z} - \bar{t}V^*$

6.3) Using MATLAB, plot the contours of

$f(Z_0, V) = Z_0^2 + 3Z_0V + 3.5V^2 - 9Z_0 - 25V + 23$

Find the minimizer (Z^*, V^*) graphically

EXERCISES

6.4) Find the minimizer of

$$f_w(x) = (Z - Hx)^T W (Z - Hx)$$

and verify that

$$x_{LS} = (H^T W H)^{-1} H^T W Z$$

6.5) The generalized inverse of H is

$$H^+ = (H^T H)^{-1} H^T \text{ if } m > n$$

$$= H^T (H H^T)^{-1} \text{ if } m < n$$

when H is of full rank

Verify that H^+ satisfies the Moore-Penrose Condition: (Module – 3)

a) $HH^+H = H$
b) $H^+HH^+ = H^+$
c) $(HH^+)^T = HH^+$
d) $(H^+H)^T = H^+H$

REFERENCES

- J. Lewis, S. Lakshmivarahan, S. Dhall (2006), Dynamic Data Assimilation: a least squares approach, *Cambridge University Press – Chapter 5*