

Linear Least Squares

Given a linear system $\mathbf{Ax} - \mathbf{b} = \mathbf{e}$,

$$\mathbf{a}_1 \bullet \mathbf{x} - b_1 = e_1$$

\vdots

$$\mathbf{a}_i \bullet \mathbf{x} - b_i = e_i$$

\vdots

$$\mathbf{a}_m \bullet \mathbf{x} - b_m = e_m$$

We want to minimize the sum of squares of the errors

$$\min_{\mathbf{x}} \sum_i e_i^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b})$$

Sometimes write this as $\mathbf{Ax} \cong \mathbf{b}$

601.455/655 Sep-21
Copyright R. H. Taylor

1

Linear Least Squares

- Many methods for $\mathbf{Ax} \approx \mathbf{b}$
- One simple one is to compute

$$\mathbf{Ax} \approx \mathbf{b}$$

$$\mathbf{A}^T \mathbf{Ax} \approx \mathbf{A}^T \mathbf{b}$$

$$\mathbf{x} \approx (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$$

- Better methods based on orthogonal transformations exist
- These methods are available in standard math libraries
- A short review follows

601.455/655 Sep-21
Copyright R. H. Taylor

2

Orthogonal Transformations

The key property is:

$$\mathbf{Q}^{-1} = \mathbf{Q}^T$$

Some implications of this are as follows

$$\text{if } \mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_n]$$

$$\text{then } \mathbf{q}_i \bullet \mathbf{q}_j = \begin{cases} 1, & \text{if } i=j \\ 0, & \text{if } i \neq j \end{cases}$$

$$\begin{aligned} \|\mathbf{Qx}\| &= \sqrt{(\mathbf{Qx})^T (\mathbf{Qx})} \\ &= \sqrt{\mathbf{x}^T \mathbf{Q}^T \mathbf{Q} \mathbf{x}} = \sqrt{\mathbf{x}^T \mathbf{x}} \\ &= \|\mathbf{x}\| \end{aligned}$$

601.455/655 Sep-21
Copyright R. H. Taylor

3

General Approach

The discussion below generally follows the development in
D. Lawson and R. Hanson, *Solving Least Squares Problems*,
Prentice-Hall, 1974

However, similar discussions may be found in many textbooks.

Given the problem

$$\min \|\mathbf{Ax} - \mathbf{b}\|$$

Observe that for any orthogonal matrix \mathbf{Q}

$$\|\mathbf{Ax} - \mathbf{b}\| = \|\mathbf{Q}(\mathbf{Ax} - \mathbf{b})\| = \|\mathbf{QAx} - \mathbf{Qb}\|$$

601.455/655 Sep-21
Copyright R. H. Taylor

4

Theorem (from Lawson & Hanson pp. 5-6)

Suppose \mathbf{A} is an $m \times n$ matrix with rank k and

$$\mathbf{A} = \mathbf{H}\mathbf{R}\mathbf{K}^T$$

where

$\mathbf{H} = m \times m$ orthogonal matrix

$\mathbf{K} = n \times n$ orthogonal matrix

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \text{ with } \text{rank}(\mathbf{R}_{11}) = k$$

This is called an orthogonal decomposition of \mathbf{A}

Define

$$\mathbf{g} = \mathbf{H}^T \mathbf{b} = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} \begin{matrix} \} k \\ \} n-k \end{matrix} \quad \mathbf{y} = \mathbf{K}^T \mathbf{x} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \begin{matrix} \} k \\ \} n-k \end{matrix}$$

and define $\tilde{\mathbf{y}}_1$ to be the unique solution of

$$\mathbf{R}_{11} \mathbf{y}_1 = \mathbf{g}_1$$

601.455/655 Sep-21
Copyright R. H. Taylor

5

Theorem (from Lawson & Hanson pp 5-6)

Then ...

- 1) All solutions to the problem of minimizing $\|\mathbf{Ax} - \mathbf{b}\|$ are of the form

$$\hat{\mathbf{x}} = \mathbf{K} \begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \mathbf{y}_2 \end{bmatrix} \text{ where } \mathbf{y}_2 \text{ is arbitrary}$$

- 2) Any such $\hat{\mathbf{x}}$ produces the same residual vector \mathbf{r} satisfying

$$\mathbf{r} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}} = \mathbf{H} \begin{bmatrix} \mathbf{0} \\ \mathbf{g}_2 \end{bmatrix}$$

- 3) The norm of \mathbf{r} satisfies

$$\|\mathbf{r}\| = \|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\| = \|\mathbf{g}_2\|$$

- 4) The unique solution of minimum length is

$$\tilde{\mathbf{x}} = \mathbf{K} \begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \mathbf{0} \end{bmatrix}$$

601.455/655 Sep-21
Copyright R. H. Taylor

6

Householder Decomposition

One method uses repeated Householder transformations to produce an upper triangular matrix \mathbf{R} .

$$\mathbf{H}^T \mathbf{A} \mathbf{K} = \mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} & 0 & \cdots & 0 \\ 0 & r_{22} & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & & \ddots & r_{k-1,k} & \vdots & & \vdots \\ 0 & \cdots & 0 & r_{kk} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

where $\mathbf{H}^T = \mathbf{H}_{k-1}^T \cdots \mathbf{H}_2^T \mathbf{H}_1^T$ is a product of Householder transformations and $\mathbf{K} = \mathbf{K}_1 \mathbf{K}_2 \cdots \mathbf{K}_p$ is a series of permutations, if needed, to avoid division by 0. Then, we solve the problem $\mathbf{A} \mathbf{x} \approx \mathbf{b}$ by solving $\mathbf{R}_{11} \tilde{\mathbf{y}}_1 = \mathbf{g}_1$ and

forming $\tilde{\mathbf{x}} = \mathbf{K} \begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \mathbf{0} \end{bmatrix}$ as outlined before.

601.455/655 Sep-21
Copyright R. H. Taylor

7

Singular Value Decomposition

- Developed by Golub, et al in late 1960's
- Commonly available in mathematical libraries
- E.g.,
 - MATLAB
 - IMSL
 - Numerical Recipes (Wm. Press, et. al., Cambridge Press)
 - CISST ERC Math Library

601.455/655 Sep-21
Copyright R. H. Taylor

8

Singular Value Decomposition

Given an arbitrary m by n matrix \mathbf{A} , there exist orthogonal matrices \mathbf{U} , \mathbf{V} and a diagonal matrix \mathbf{S} that:

$$\mathbf{A}_{m \times n} = \mathbf{U}_{m \times m} \begin{bmatrix} \mathbf{S}_{n \times n} \\ \mathbf{0}_{(m-n) \times n} \end{bmatrix} \mathbf{V}_{n \times n}^T \quad \text{for } m \geq n$$

or

$$\mathbf{A}_{m \times n} = \mathbf{U}_{m \times m} \begin{bmatrix} \mathbf{S}_{n \times n} & \mathbf{0}_{(m \times (n-m))} \end{bmatrix} \mathbf{V}_{n \times n}^T \quad \text{for } m \leq n$$

601.455/655 Sep-21
Copyright R. H. Taylor

9

SVD Least Squares

$$\mathbf{A}_{m \times n} \mathbf{x} \approx \mathbf{b}$$

$$\mathbf{U}_{m \times m} \begin{bmatrix} \mathbf{S}_{n \times n} \\ \mathbf{0}_{(m-n) \times n} \end{bmatrix} \mathbf{V}_{n \times n}^T \mathbf{x} = \mathbf{b}$$

$$\begin{bmatrix} \mathbf{S}_{n \times n} \\ \mathbf{0}_{(m-n) \times n} \end{bmatrix} \mathbf{y} = \mathbf{U}_{m \times m}^T \mathbf{b} \quad \text{where } \mathbf{y} = \mathbf{V}^T \mathbf{x}$$

Solve this for \mathbf{y} (trivial, since \mathbf{S} is diagonal), then compute

$$\mathbf{V} \mathbf{y} = \mathbf{V} \mathbf{V}^T \mathbf{x} = \mathbf{x}$$

601.455/655 Sep-21
Copyright R. H. Taylor

10

Least squares adjustment

Given a vector function $\bar{\mathbf{G}}(\bar{\mathbf{q}}; \bar{\mathbf{u}})$ of parameters $\bar{\mathbf{q}}$ and experimental variables $\bar{\mathbf{u}}$, together with a set of observations

$$\bar{\mathbf{v}}_k = \bar{\mathbf{G}}(\bar{\mathbf{q}}; \bar{\mathbf{u}}_k)$$

and an initial guess $\bar{\mathbf{q}}_0$ of the values of $\bar{\mathbf{q}}$, we wish to find a better estimate of $\bar{\mathbf{q}}$.

601.455/655 Sep-21
Copyright R. H. Taylor

11

Least Squares Adjustment

Step 0 $j \leftarrow 0$;

Step 1 Compute $\bar{\varepsilon}_k \leftarrow \bar{\mathbf{v}}_k - \bar{\mathbf{G}}(\bar{\mathbf{q}}; \bar{\mathbf{u}}_k)$ for $k=1 \dots N$; $\bar{\mathbf{E}}_j \leftarrow [\bar{\varepsilon}_1, \dots, \bar{\varepsilon}_N]^T$

Step 2 If $\|\bar{\mathbf{E}}_j\|$ is small or some other convergence criterion is met, then stop. Otherwise go on to Step 3.

Step 3 Solve the least squares problem

$$\begin{bmatrix} \vdots \\ \mathbf{J}_G(\bar{\mathbf{q}}_j, \bar{\mathbf{u}}_k) \\ \vdots \end{bmatrix} \bullet \Delta \bar{\mathbf{q}} \approx \begin{bmatrix} \vdots \\ -\bar{\varepsilon}_k \\ \vdots \end{bmatrix}$$

for $\Delta \bar{\mathbf{q}}$.

Step 4 Set $\bar{\mathbf{q}}_{j+1} \leftarrow \bar{\mathbf{q}}_j + \Delta \bar{\mathbf{q}}$; $j \leftarrow j + 1$; Go back to Step 1.

601.455/655 Sep-21
Copyright R. H. Taylor

12