Linear Least Squares

Given a linear system **Ax -b = 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

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

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

Linear Least Squares

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

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

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

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

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

Orthogonal Transformations

The key property is:

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

Some implications of this are as follows

$$\text{if} \qquad \mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$$

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

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

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

$$\min \left\| \mathbf{A}x - \mathbf{b} \right\|$$

Observe than for any orthogonal matrix **Q**

$$\left\| \mathbf{A}x - \mathbf{b} \right\| = \left\| \mathbf{Q}\left(\mathbf{A}x - \mathbf{b} \right) \right\| = \left\| \mathbf{QA}x - \mathbf{Qb} \right\|$$

Theorem (from Lawson & Hanson pp 5-6)

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

$$\mathbf{A} = \mathbf{HRK}^T$$

where

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

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

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

This is called an orthogonal decomposition of **A**

Define

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

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

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

Theorem (from Lawson & Hanson pp 5-6)

Then …

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

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

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

Householder Decomposition

One method uses repeated Householder transformations to produce an upper triangular matrix **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.

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

Singular Value Decomposition

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

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

or

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

SVD Least Squares

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

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

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

Solve this for **y** (trivial, since **S** is diagonal), then compute

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

Least squares adjustment

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

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

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

Least Squares Adjustment

Step 0 $j \leftarrow 0;$

Step 1 Compute $\vec{\varepsilon}_{\mathrm{k}} \leftarrow \vec{\mathbf{v}}_k - \vec{\mathbf{G}}(\vec{\mathbf{q}};\vec{\mathbf{u}}_k)$ for k=1$\cdots$N; $\vec{\mathrm{E}}_{\mathrm{j}} \leftarrow \left[\vec{\varepsilon}_1,\cdots,\vec{\varepsilon}_N\right]^T$

Step 2 If $\left\|\vec{\mathrm{E}}_{\mathrm{j}}\right\|$ is small or some other convergence criterion is met, then stop. Otherwise go on to Step 3.

Step 3 Solve the least squares problem

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

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

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