Linear Least Squares Derivation

Written by Jeff Wood in LaTeX environment
Updated 2016/01/31

Solution of $\vec{\mathbf{x}}$ to $_{\min}||\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}}||$ is obtained by solving $(\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}})^T \cdot \mathbf{A} = \vec{\mathbf{0}}$ for $\vec{\mathbf{x}}$, where

- $\mathbf{A}$ is $m \times n$
- $\vec{\mathbf{y}}$ is $m \times 1$
- $\vec{\mathbf{x}}$ is $n \times 1$

and

$$\mathbf{A} = \begin{bmatrix} \vec{\mathbf{a}}_1 & \vec{\mathbf{a}}_2 & \dots & \vec{\mathbf{a}}_n \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix}, \vec{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \vec{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \tag{1}$$

Minimizing $f(\vec{\mathbf{x}}) = ||\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}}||$ is the same as minimizing $(\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}})^T (\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}})$

$$_{\min}||f(\vec{\mathbf{x}})|| =_{\min} ||\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}}|| =_{\min} \left[(\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}})^T (\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}}) \right] \tag{2}$$

Multiplying out by the FOIL method gives:

$$\begin{aligned} f(\vec{\mathbf{x}}) &= (\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}})^T (\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}}) \\ &= \left(\vec{\mathbf{y}}^T - (\mathbf{A} \cdot \vec{\mathbf{x}})^T \right) (\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}}) \\ &= \left(\vec{\mathbf{y}}^T - \vec{\mathbf{x}}^T \cdot \mathbf{A}^T \right) (\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}}) \\ &= \vec{\mathbf{y}}^T \vec{\mathbf{y}} - \vec{\mathbf{x}}^T \mathbf{A}^T \vec{\mathbf{y}} - \vec{\mathbf{y}}^T \mathbf{A} \vec{\mathbf{x}} + \vec{\mathbf{x}}^T \mathbf{A}^T \mathbf{A} \vec{\mathbf{x}} \end{aligned} \tag{3}$$

This gives a summation of matrix products

$$f(\vec{\mathbf{x}}) = \vec{\mathbf{y}}^T \vec{\mathbf{y}} - \vec{\mathbf{x}}^T \mathbf{A}^T \vec{\mathbf{y}} - \vec{\mathbf{y}}^T \mathbf{A} \vec{\mathbf{x}} + \vec{\mathbf{x}}^T \mathbf{A}^T \mathbf{A} \vec{\mathbf{x}} \tag{4}$$

Which can further be reduced by writing as a partition of vectors and inner products:

$$= ||\vec{\mathbf{y}}|| - \vec{\mathbf{x}}^T \begin{bmatrix} \vec{\mathbf{a}}_1^T \vec{\mathbf{y}} & \vec{\mathbf{a}}_2^T \vec{\mathbf{y}} & \dots & \vec{\mathbf{a}}_n^T \vec{\mathbf{y}} \end{bmatrix}^T - \begin{bmatrix} \vec{\mathbf{y}}^T \vec{\mathbf{a}}_1 & \vec{\mathbf{y}}^T \vec{\mathbf{a}}_2 & \dots & \vec{\mathbf{y}}^T \vec{\mathbf{a}}_n \end{bmatrix} \vec{\mathbf{x}} + \vec{\mathbf{x}}^T \begin{bmatrix} \vec{\mathbf{a}}_1^T \vec{\mathbf{a}}_1 & \vec{\mathbf{a}}_1^T \vec{\mathbf{a}}_2 & \dots & \vec{\mathbf{a}}_1^T \vec{\mathbf{a}}_n \\ \vec{\mathbf{a}}_2^T \vec{\mathbf{a}}_1 & \vec{\mathbf{a}}_2^T \vec{\mathbf{a}}_2 & \dots & \vec{\mathbf{a}}_2^T \vec{\mathbf{a}}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{\mathbf{a}}_n^T \vec{\mathbf{a}}_1 & \vec{\mathbf{a}}_n^T \vec{\mathbf{a}}_2 & \dots & \vec{\mathbf{a}}_n^T \vec{\mathbf{a}}_n \end{bmatrix} \vec{\mathbf{x}} \tag{5}$$

And then further reduced by writing sums in terms of x-components of the vector $\vec{\mathbf{x}}$:

$$\begin{aligned} &= ||\vec{\mathbf{y}}|| - 2 \cdot \begin{bmatrix} \vec{\mathbf{y}}^T \vec{\mathbf{a}}_1 & \vec{\mathbf{y}}^T \vec{\mathbf{a}}_2 & \dots & \vec{\mathbf{y}}^T \vec{\mathbf{a}}_n \end{bmatrix} \vec{\mathbf{x}} + \begin{pmatrix} \vec{\mathbf{a}}_1^T \vec{\mathbf{a}}_1 x_1 x_1 + \vec{\mathbf{a}}_1^T \vec{\mathbf{a}}_2 x_2 x_1 + \ldots + \vec{\mathbf{a}}_1^T \vec{\mathbf{a}}_n x_n x_1 + \\ \vec{\mathbf{a}}_2^T \vec{\mathbf{a}}_1 x_1 x_2 + \vec{\mathbf{a}}_2^T \vec{\mathbf{a}}_2 x_2 x_2 + \ldots + \vec{\mathbf{a}}_2^T \vec{\mathbf{a}}_n x_n x_2 + \\ \vdots \\ \vec{\mathbf{a}}_n^T \vec{\mathbf{a}}_1 x_1 x_n + \vec{\mathbf{a}}_n^T \vec{\mathbf{a}}_2 x_2 x_n + \ldots + \vec{\mathbf{a}}_n^T \vec{\mathbf{a}}_n x_n x_n \end{pmatrix} \\ &= \sum_{j=1}^{m} y_j^2 - 2 \cdot \left[\vec{\mathbf{y}}^T \vec{\mathbf{a}}_1 x_1 + \vec{\mathbf{y}}^T \vec{\mathbf{a}}_2 x_2 + \ldots + \vec{\mathbf{y}}^T \vec{\mathbf{a}}_n x_n \right] + \begin{pmatrix} \vec{\mathbf{a}}_1^T \vec{\mathbf{a}}_1 x_1 x_1 + \vec{\mathbf{a}}_1^T \vec{\mathbf{a}}_2 x_2 x_1 + \ldots + \vec{\mathbf{a}}_1^T \vec{\mathbf{a}}_n x_n x_1 + \\ \vec{\mathbf{a}}_2^T \vec{\mathbf{a}}_1 x_1 x_2 + \vec{\mathbf{a}}_2^T \vec{\mathbf{a}}_2 x_2 x_2 + \ldots + \vec{\mathbf{a}}_2^T \vec{\mathbf{a}}_n x_n x_2 + \\ \vdots \\ \vec{\mathbf{a}}_n^T \vec{\mathbf{a}}_1 x_1 x_n + \vec{\mathbf{a}}_n^T \vec{\mathbf{a}}_2 x_2 x_n + \ldots + \vec{\mathbf{a}}_n^T \vec{\mathbf{a}}_n x_n x_n \end{pmatrix} \end{aligned} \tag{6}$$

This gives us a total *square*-distance of $||\vec{\mathbf{y}} - \mathbf{A}\vec{\mathbf{x}}||$ which we can attempt to minimize by varying each x-component. The value of each x_j that gives the minimum distance is calculated by setting the partial derivative of (6) equal to zero.

Setting $\partial f(\vec{\mathbf{x}})/\partial x_j = 0$ gives:

$$\begin{aligned}
\partial f(\vec{\mathbf{x}})/\partial x_j = 0 &= -2\vec{\mathbf{y}}^T\vec{\mathbf{a}}_j + 2\cdot\sum_{i=1}^{n}\vec{\mathbf{a}}_j^T\vec{\mathbf{a}}_i x_i \\
&= -2\vec{\mathbf{y}}^T\vec{\mathbf{a}}_j + 2\cdot\vec{\mathbf{a}}_j^T\sum_{i=1}^{n}\vec{\mathbf{a}}_i x_i \\
&= -2\vec{\mathbf{y}}^T\vec{\mathbf{a}}_j + 2\cdot\vec{\mathbf{a}}_j^T\mathbf{A}\vec{\mathbf{x}}
\end{aligned} \tag{7}$$

Rearranging the sum to eliminate the 0 term, then simplifying gives:

$$\begin{aligned}
2\vec{\mathbf{y}}^T\vec{\mathbf{a}}_j &= 2\cdot\vec{\mathbf{a}}_j^T\mathbf{A}\vec{\mathbf{x}} \\
2\left(\vec{\mathbf{y}}^T\vec{\mathbf{a}}_j\right) &= 2\cdot\vec{\mathbf{a}}_j^T\mathbf{A}\vec{\mathbf{x}} \\
2\left(\vec{\mathbf{a}}_j^T\vec{\mathbf{y}}\right) &= 2\cdot\vec{\mathbf{a}}_j^T\mathbf{A}\vec{\mathbf{x}} \\
\vec{\mathbf{a}}_j^T\vec{\mathbf{y}} &= \vec{\mathbf{a}}_j^T\mathbf{A}\vec{\mathbf{x}}
\end{aligned} \tag{8}$$

Since only gives the value for x_j, finding the vector $\vec{\mathbf{x}}$ requires solving (8) for all n cases.

This can be accomplished by creating a n-length vector from the last line of (8) as:

$$\begin{gathered}
\begin{bmatrix} \vec{\mathbf{a}}_1^T \vec{\mathbf{y}} \\ \vec{\mathbf{a}}_2^T \vec{\mathbf{y}} \\ \vdots \\ \vec{\mathbf{a}}_n^T \vec{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} \vec{\mathbf{a}}_1^T \mathbf{A} \vec{\mathbf{x}} \\ \vec{\mathbf{a}}_2^T \mathbf{A} \vec{\mathbf{x}} \\ \vdots \\ \vec{\mathbf{a}}_n^T \mathbf{A} \vec{\mathbf{x}} \end{bmatrix} = \\
\begin{bmatrix} \vec{\mathbf{a}}_1^T \\ \vec{\mathbf{a}}_2^T \\ \vdots \\ \vec{\mathbf{a}}_n^T \end{bmatrix} \vec{\mathbf{y}} = \begin{bmatrix} \vec{\mathbf{a}}_1^T \\ \vec{\mathbf{a}}_2^T \\ \vdots \\ \vec{\mathbf{a}}_n^T \end{bmatrix} \mathbf{A} \vec{\mathbf{x}} = \\
\mathbf{A}^T \vec{\mathbf{y}} = \mathbf{A}^T \mathbf{A} \vec{\mathbf{x}}
\end{gathered} \tag{9}$$

The last step involves solving for $\vec{\mathbf{x}}$ which is:

$$\left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T \vec{\mathbf{y}} = \vec{\mathbf{x}} \tag{10}$$

$$(\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}})^T \cdot \mathbf{A} = \vec{\mathbf{0}} \tag{11}$$

$$\begin{aligned} \mathbf{A}^T \cdot (\vec{\mathbf{y}} - \mathbf{A} \cdot \vec{\mathbf{x}}) &= \vec{\mathbf{0}} \\ \mathbf{A}^T \cdot \vec{\mathbf{y}} - \mathbf{A}^T \cdot \mathbf{A} \cdot \vec{\mathbf{x}} &= \vec{\mathbf{0}} \\ \mathbf{A}^T \cdot \vec{\mathbf{y}} &= \mathbf{A}^T \cdot \mathbf{A} \cdot \vec{\mathbf{x}} \\ \left(\mathbf{A}^T \cdot \mathbf{A}\right)^{-1} \mathbf{A}^T \cdot \vec{\mathbf{y}} &= \vec{\mathbf{x}} \end{aligned} \tag{12}$$