Least Squares and SVD

Sources:

`Matrix Methods for Computational Modeling and Data Analytics` author: Mark Embree, Virginia Tech

A solution of the linear system

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$$

can be found if $\mathbf{b}$ is in the column space of $\mathbf{A}$. Or expressed otherwise:

$$\mathbf{b} \in R(\mathbf{A})$$

For the more general case $\mathbf{b} \notin R(\mathbf{A})$ we are interested in a solution which minimises $||\mathbf{b} - \mathbf{A}\mathbf{x}||$.

The vector space $\mathbb{R}^m$ of which $\mathbf{b}$ is a vector is the sum of column space $R(\mathbf{A})$ and left null space $N(\mathbf{A}^T)$.

$$\mathbb{R}^m = R(\mathbf{A}) \oplus N(\mathbf{A}^T)$$

This allows us to decompose vector $\mathbf{b}$ into a part $\mathbf{b}_R$ in $R(\mathbf{A})$ and a orthogonal part $\mathbf{b}_N$ in $N(\mathbf{A}^T)$.

$$\mathbf{b} = \mathbf{b}_R + \mathbf{b}_N$$

With these notation the linear system can be formulated in terms of these vectors:

$$\mathbf{b} - \mathbf{A}\mathbf{x} = \mathbf{b}_R + \mathbf{b}_N - \mathbf{A}\mathbf{x} = (\mathbf{b}_R - \mathbf{A}\mathbf{x}) + \mathbf{b}_N$$

The quadratic norm is computed:

$$\begin{aligned}
||\mathbf{b} - \mathbf{A}\mathbf{x}||^2 &= ((\mathbf{b}_R - \mathbf{A}\mathbf{x}) + \mathbf{b}_N)^T \cdot ((\mathbf{b}_R - \mathbf{A}\mathbf{x}) + \mathbf{b}_N) &(1)\\
&= ||\mathbf{b}_R - \mathbf{A}\mathbf{x}||^2 + ||\mathbf{b}_N||^2 - 2\underbrace{(\mathbf{b}_R - \mathbf{A}\mathbf{x})^T \cdot \mathbf{b}_N}_{orthogornality} &(2)\\
&= ||\mathbf{b}_R - \mathbf{A}\mathbf{x}||^2 + ||\mathbf{b}_N||^2 &(3)
\end{aligned}$$

Minimising $||\mathbf{b} - \mathbf{A}\mathbf{x}||$ is then equivalent to minimise $||\mathbf{b}_R - \mathbf{A}\mathbf{x}||$. And this is equivalent to find the solution of

$$\mathbf{A}\mathbf{x} = \mathbf{b}_R$$

For the general case of a rectangular matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ we multiply $\mathbf{b} - \mathbf{A}\mathbf{x}$ by $\mathbf{A}^T$:

$$\mathbf{A}^T \cdot (\mathbf{b} - \mathbf{A}\mathbf{x}) = \mathbf{A}^T \cdot \mathbf{b} - \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T \cdot \mathbf{b}_R + \underbrace{\mathbf{A}^T \cdot \mathbf{b}_N}_{0} - \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T \cdot \mathbf{b}_R - \mathbf{A}^T\mathbf{A}\mathbf{x}$$

So we solve

$$\mathbf{A}^T\mathbf{A}\cdot\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

so there is no need to obtain $\mathbf{b}_R$ explicitely. If the inverse $\left(\mathbf{A}^T\mathbf{A}\right)^{-1}$ exists the solution vector $\mathbf{x}$ is just:

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

In many text on linear algebra the matrix

$$\mathbf{A}^+ = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\cdot\mathbf{A}^T$$

is referred to an `pseudoinverse` of $\mathbf{A}$.

With the reduced `SVD` an alternate expression for the pseudoinverse is computed.

$$\mathbf{A} = \mathbf{U}\cdot\mathbf{\Sigma}\cdot\mathbf{V}^T = \sum_{j=1}^{r}\sigma_j\cdot\mathbf{u}_j\cdot\mathbf{v}_j^T$$

$$\begin{align}
\mathbf{A}^+ &= \left(\left(\mathbf{U}\cdot\mathbf{\Sigma}\cdot\mathbf{V}^T\right)^T\mathbf{U}\cdot\mathbf{\Sigma}\cdot\mathbf{V}^T\right)^{-1}\cdot\left(\mathbf{U}\cdot\mathbf{\Sigma}\cdot\mathbf{V}^T\right)^T \tag{4}\\
&= \left(\mathbf{V}\cdot\mathbf{\Sigma}\cdot\mathbf{U}^T\cdot\mathbf{U}\cdot\mathbf{\Sigma}\cdot\mathbf{V}^T\right)^{-1}\cdot\mathbf{V}\cdot\mathbf{\Sigma}\cdot\mathbf{U}^T \tag{5}\\
&= \left(\mathbf{V}\cdot\mathbf{\Sigma}\cdot\mathbf{\Sigma}\cdot\mathbf{V}^T\right)^{-1}\cdot\mathbf{V}\cdot\mathbf{\Sigma}\cdot\mathbf{U}^T \tag{6}\\
&= \mathbf{V}\cdot\mathbf{\Sigma}^{-1}\cdot\mathbf{\Sigma}^{-1}\cdot\mathbf{V}^{-1}\cdot\mathbf{V}\cdot\mathbf{\Sigma}\cdot\mathbf{U}^T \tag{7}\\
&= \mathbf{V}\cdot\mathbf{\Sigma}^{-1}\cdot\mathbf{\Sigma}^{-1}\cdot\mathbf{\Sigma}\cdot\mathbf{U}^T \tag{8}\\
&= \mathbf{V}\cdot\mathbf{\Sigma}^{-1}\cdot\mathbf{U}^T = \sum_{j=1}^{r}\frac{1}{\sigma_j}\cdot\mathbf{v}\cdot\mathbf{u}_j^T \tag{9}
\end{align}$$

Note that deriving this expression we have utilised several properties of matrices $\mathbf{U}$ and $\mathbf{V}$:

1. Since $\mathbf{U}$ is orthogonal $\mathbf{U}^T\mathbf{U} = \mathbf{I}$
2. Since $\mathbf{V}$ is orthogonal and square it has an inverse matrix $\mathbf{V}^{-1} = \mathbf{V}^T$

In [ ]: