Range, Null Space and the Least Squares solution

Definition 10: Let $\mathbf{A}$ be an arbitrary $m \times n$ $(m \geq n)$ matrix. The function $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is then said to be a **linear mapping** from $\mathbb{R}^n \to \mathbb{R}^m$. The **range** of the linear mapping is the set $B(\mathbf{A}) \subseteq \mathbb{R}^m$, that meets the condition that for any $\mathbf{y} \in B(\mathbf{A})$ exists an $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{A}\mathbf{x} = \mathbf{y}$. **The null space** of the linear mapping is the set $N(\mathbf{A}) \subseteq \mathbb{R}^n$, that meets the condition that for any $\mathbf{x} \in N(\mathbf{A})$ it applies that $\mathbf{A}\mathbf{x} = 0$.

Theorem 5 Let $\mathbf{u}_1 \ldots, \mathbf{u}_K$ be an arbitrary orthonormal basis for $B(\mathbf{A})$. Then the least squares solution $\mathbf{x}$ that minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ satisfies

$$\mathbf{A}\mathbf{x} = \sum_{k=1}^{K} (\mathbf{b} \cdot \mathbf{u}_k)\mathbf{u}_k \equiv \mathbf{b}_{LS}$$

Singular Value Decomposition

Theorem 6 Consider an arbitrary $m \times n$ matrix $\mathbf{A}$. Then we can write $\mathbf{A}$ as $\mathbf{A} = \mathbf{UWV}^T$, where $\mathbf{U}$ is an $m \times n$ column othonormal matrix, $\mathbf{V}$ is an $n \times n$ orthonormal matrix and $\mathbf{W}$ is an $n \times n$ diagonal matrix having non-negative diagonal elements $w_1, \ldots w_n$ ordered such that $w_1 \geq w_2 \geq \ldots \geq w_n$. This is said to be a **Singular Value Decomposition** (SVD) of $\mathbf{A}$.

$$\begin{pmatrix} \\ \\ & \mathbf{A} & \\ \\ \\ \end{pmatrix} = \begin{pmatrix} \\ \\ & \mathbf{U} & \\ \\ \\ \end{pmatrix} \cdot \begin{pmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{pmatrix} \cdot \begin{pmatrix} & \mathbf{V}^T & \end{pmatrix}$$

$\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, where $\mathbf{U}$ is an $m \times n$ column othonormal matrix, $\mathbf{V}$ is an $n \times n$ orthonormal matrix and $\mathbf{W}$ is an $n \times n$ diagonal matrix having non-negative diagonal elements $w_1, \ldots w_n$ ordered such that $w_1 \geq w_2 \geq \ldots \geq w_n$.

Theorem 7 Consider an arbitrary $m \times n$ matrix $\mathbf{A}$ and assume that for $\mathbf{W}$ it applies that $w_1, \ldots, w_K$ are positive and $w_{K+1}, \ldots, w_n$ are equal to zero. Then it applies that

i) $N(\mathbf{A})$ has dimension $n - K$ and the last $n - K$ columns of $\mathbf{V}$ form an orthonormal basis for $N(\mathbf{A})$.

ii) $B(\mathbf{A})$ has dimension K and the first K columns of $\mathbf{U}$ form an orthonormal basis for $B(\mathbf{A})$.

iii) The SVD solution $\mathbf{x} = \mathbf{V}\tilde{\mathbf{W}}^{-1}\mathbf{U}^T\mathbf{b}$, where $[\tilde{\mathbf{W}}^{-1}]_{jj} = 0$ if $\mathbf{W}_{jj} = 0$, otherwise $[\tilde{\mathbf{W}}^{-1}]_{jj} = 1/\mathbf{W}_{jj}$, is the least squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$. (Notice that it then follows that if all of the $\mathbf{W}_{jj}$'s are positive, i.e. $\mathbf{A}$ has full rank it applies that $\mathbf{x} = \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\mathbf{b}$ is the least squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$).

Error analysis for systems of linear equations. Residual errors

The residual error should be computed as a relative error, namely

$$\epsilon_{residual} = \frac{\|\mathbf{Ax} - \mathbf{b}\|}{\|\mathbf{b}\|}$$

If $m = n$, the residual error should be very close to zero unless the matrix is near singular. For $m > n$, the linear equations are typically from some sort of fitting problem such as a Least Squares Problem. The value $\epsilon_{residual}$ indicates how good the fitting model is. It is easy to see that a random fitting model would produce $\epsilon_{residual} \simeq \sqrt{\frac{m-n}{m}}$. If your result is not much better than that, you should consider the quality of your model.

Error analysis for systems of linear equations. Errors on solution

Even though solving a set of linear equations seem very deterministic, it is relevant to consider the error $\delta\mathbf{x}$ on the result $\mathbf{x}$. In typical applications, there are two very different sources to this error.

The first source is the error on the right hand side $\delta\mathbf{b}$. The error $\delta\mathbf{b}$ is typically is some kind of measurement error and therefore may be quite large.

The second source is the error on the matrix $\delta\mathbf{A}$ which is typically due to the real number precision. Hence, $||\delta\mathbf{A}||$ is mostly of the order $||\delta\mathbf{A}|| \simeq 10^{-18}$.

Error on solution. Impact from error on right hand side

$$\begin{aligned} \mathbf{A}_{ij} &:= \mathbf{A}_{ij}/\sigma_i \quad i = 1, \ldots m, \quad j = 1, \ldots n \\ \mathbf{b}_i &:= \mathbf{b}_i/\sigma_i \quad i = 1, \ldots m \end{aligned}$$

where σ_i is the inaccuracy on $\mathbf{b}_i$.

The error estimate $\delta\mathbf{x}$ is then purely given by the SVD matrices using Eq.15.4.19

$$[\delta\mathbf{x}]_j \simeq \sqrt{\sum_{i=1}^{n} \left(\frac{V_{ji}}{w_i}\right)^2} \quad j = 1, \ldots, n \tag{4}$$

Error on solution. Impact from error on matrix

$$(\mathbf{A} + \delta\mathbf{A})(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b}$$

For simplicity, we restrict ourselves to inaccuracies $\delta\mathbf{A}$ that are in the range of $\mathbf{A}$. Hence, $\delta\mathbf{A} = \mathbf{U}\mathbf{U}^T\delta\mathbf{A}$. Formulating using SVD, we get

$$\mathbf{U}\mathbf{W}\mathbf{V}^T(\mathbf{I} + \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\delta\mathbf{A})(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b}$$

$$\begin{aligned}(\mathbf{x} + \delta\mathbf{x}) &= (\mathbf{I} + \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\delta\mathbf{A})^{-1}\mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\mathbf{b} \\ &= (\mathbf{I} + \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\delta\mathbf{A})^{-1}\mathbf{x}\end{aligned}$$

and hence

$$\begin{aligned}(\mathbf{x} + \delta\mathbf{x}) &= (\mathbf{I} + \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\delta\mathbf{A})^{-1}\mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\mathbf{b} \\ &= (\mathbf{I} + \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\delta\mathbf{A})^{-1}\mathbf{x}\end{aligned}$$

which yields

$$\delta\mathbf{x} = [(\mathbf{I} + \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\delta\mathbf{A})^{-1} - \mathbf{I}]\mathbf{x}$$

Using geometric series and approximations, we get (see notes):

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|\mathbf{W}^{-1}\|\|\delta\mathbf{A}\|}{1 - \|\mathbf{W}^{-1}\|\|\delta\mathbf{A}\|} \tag{5}$$

If we compute $\mathbf{A}^T\mathbf{A}$ using SVD, we get

$$\mathbf{A}^T\mathbf{A} = (\mathbf{U}\mathbf{W}\mathbf{V}^T)^T(\mathbf{U}\mathbf{W}\mathbf{V}^T) = (\mathbf{V}\mathbf{W}\mathbf{U}^T)(\mathbf{U}\mathbf{W}\mathbf{V}^T) = \mathbf{V}\mathbf{W}^2\mathbf{V}^T$$

which is itself the SVD of $\mathbf{A}^T\mathbf{A}$.

We then get the SVD solution to $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ as

$$\begin{aligned}
(\mathbf{x}+\delta\mathbf{x}) &= (\mathbf{I}+\mathbf{V}\mathbf{W}^{-2}\mathbf{V}^T\delta\mathbf{A})^{-1}\mathbf{V}\mathbf{W}^{-2}\mathbf{V}^T(\mathbf{V}\mathbf{W}\mathbf{U}^T\mathbf{b}) \\
&= (\mathbf{I}+\mathbf{V}\mathbf{W}^{-2}\mathbf{V}^T\delta\mathbf{A})^{-1}\mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\mathbf{b} \\
&= (\mathbf{I}+\mathbf{V}\mathbf{W}^{-2}\mathbf{V}^T\delta\mathbf{A})^{-1}\mathbf{x}
\end{aligned}$$

Performing the derivation completely equivalent to the above, we get

$$\frac{||\delta\mathbf{x}||}{||\mathbf{x}||} \leq \frac{||\mathbf{W}^{-2}||\,||\delta\mathbf{A}||}{1-||\mathbf{W}^{-2}||\,||\delta\mathbf{A}||} \tag{6}$$

Here we see the problem with the Normal Equations. If for example $w_n = 10^{-9}$ for $\mathbf{A}$, we get $||\mathbf{W}^{-2}|| = 10^{-18}$ and hence $||\mathbf{W}^{-2}||\,||\delta\mathbf{A}||$ becomes around one for a double precision number representation. This is exactly the problem we see in Filip with the Normal Equations.