1 Pseudoinverse (4.3)

We will solve the least squares with the SVD with full rank matrix A. Using SVD also works when the rank of A is not full. Recall that

$$A \in \mathbb{R}^{n\times m} \qquad n \geq m \qquad \min_{\mathbf{x}\in\mathbb{R}^m} ||\mathbf{b} - A\mathbf{x}||_2^2.$$

1.1 Brief Review

Recall that, for a full rank A, i.e., $\text{rank}(A) = m$, we can use full QR decomposition ($A = QR$) or reduced QR decomposition ($A = \hat{Q}\hat{R}$). In particular, for reduced QR, $A = \hat{Q}\hat{R}$ where $\hat{Q} \in \mathbb{R}^{n\times m}$ and $\hat{R} \in \mathbb{R}^{m\times m}$. Recall that $A\mathbf{x} = \mathbf{b}$ as well, so

$$\begin{aligned}
&A\mathbf{x} = \mathbf{b} \\
&\implies \hat{Q}\hat{R}\mathbf{x} = \mathbf{b} \\
&\implies \hat{Q}^T\hat{Q}\hat{R}\mathbf{x} = \hat{Q}^T\mathbf{b} \\
&\implies \hat{R}\mathbf{x} = \hat{Q}^T\mathbf{b} \qquad \text{Since } Q^TQ = I.
\end{aligned}$$

Remember that, because $\hat{Q}$ is orthogonal, $\hat{Q}\hat{Q}^T = I_{n\times n}$ and $\hat{Q}^T\hat{Q} = I_{m\times m}$. Now, remember that if $\hat{R}$ has full rank, then it'll look like

$$\hat{R} = \begin{bmatrix} * & * & * & * & \dots & * \\ 0 & * & * & * & \dots & * \\ 0 & 0 & * & * & \dots & * \\ 0 & 0 & 0 & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & * \end{bmatrix}.$$

Then, the associated equation will have a unique solution since it has m equations and m unknowns. **Now**, what if $\text{rank}(R) = r < m$? Then, it is not full rank and it'll look like

$$\hat{R} = \begin{bmatrix} * & * & * & * & \dots & * \\ 0 & * & * & * & \dots & * \\ 0 & 0 & * & * & \dots & * \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

In particular, some of the diagonal entries are 0's. So, $\hat{R}\mathbf{x} = \hat{Q}^T\mathbf{b}$. We have m unknowns, but we only have r independent equations. Those equations are not helpful since we can't use them to solve for the unknowns – we'll end up with infinitely many solutions. So, we'll have to choose one of these infinitely many solutions based on $||\mathbf{x}||_2$.

1.2 A New Problem

We now have a new problem to solve.

- Find $\min ||\mathbf{b} - A\mathbf{x}||_2^2$ (infinitely many solutions.)
- Pick $\mathbf{x}$ with minimal $||\mathbf{x}||_2$.

With the above two minimizers, there *will* be a unique solution $\mathbf{x}$. We will see that the unique $\mathbf{x}$ can be written as $\mathbf{x} = A^+\mathbf{b}$, where A^+ is known as the **pseudoinverse**[1].

[1] Note that if A is invertible, then $\mathbf{x} = A^{-1}\mathbf{b}$, so in some sense A^+ is mimicking A^{-1}.

With this in mind, how do we find the least squares solution of the minimal 2-norm? With $A = U\Sigma V^T$, we have

$$\begin{aligned}
||\mathbf{b} - A\mathbf{x}||_2^2 &= ||\mathbf{b} - U\Sigma V^T\mathbf{x}||_2^2 \\
&= ||U(U^T\mathbf{b}) - U(\Sigma V^T\mathbf{x})||_2^2 && \text{Recall that } UU^T = U^TU = I \\
&= ||U(U^T\mathbf{b} - \Sigma V^T\mathbf{x})||_2^2 \\
&= ||U^T\mathbf{b} - \Sigma V^T\mathbf{x}||_2^2 \\
&= \left|\left| \underbrace{\begin{bmatrix} \hat{c} \\ d \end{bmatrix}}_{U^T\mathbf{b} \in \mathbb{R}^{n\times 1}} - \left[\begin{array}{c|c} \hat{\Sigma} & 0 \\ \hline 0 & 0 \end{array}\right] \underbrace{\begin{bmatrix} \hat{y} \\ z \end{bmatrix}}_{V^T\mathbf{x} \in \mathbb{R}^{m\times 1}} \right|\right|_2^2 && \text{See remark.} \\
&= \left|\left| \begin{bmatrix} \hat{c} \\ d \end{bmatrix} - \begin{bmatrix} \hat{\Sigma}\hat{y} \\ \mathbf{0} \end{bmatrix} \right|\right|_2^2 \\
&= \left|\left| \begin{bmatrix} \hat{c} - \hat{\Sigma}\hat{y} \\ d \end{bmatrix} \right|\right|_2^2 \\
&= ||\hat{c} - \hat{\Sigma}\hat{y}||_2^2 + ||d||_2^2
\end{aligned}$$

Remarks:

- Note that $\hat{c} \in \mathbb{R}^r$ and $d \in \mathbb{R}^{n-r}$ and $\hat{y} \in \mathbb{R}^r$ and $z \in \mathbb{R}^{m-r}$.
- Additionally,
$$\Sigma = \left[\begin{array}{c|c} \hat{\Sigma} & 0 \\ \hline 0 & 0 \end{array}\right]$$
where $\hat{\Sigma}$ is an $r \times r$ matrix.
- d is independent of $\mathbf{x}$ when minimizing over $\mathbf{x}$.

In any case, we want to minimize $||\hat{c} - \hat{\Sigma}\hat{y}||_2^2$. This can be done by solving

$$\hat{c} = \hat{\Sigma}\hat{y}.$$

This yields

$$\begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \sigma_3 & & \\ & & & \ddots & \\ & & & & \sigma_r \end{bmatrix} \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_r \end{bmatrix} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \vdots \\ \hat{c}_r \end{bmatrix}.$$

And this gives us

$$\hat{y}_i = \frac{\hat{c}_i}{\sigma_i}, i = 1, \dots, r.$$

To find $\mathbf{x}$, we do

$$V^T\mathbf{x} = \begin{bmatrix} \hat{y} \\ z \end{bmatrix} \implies \mathbf{x} = V \begin{bmatrix} \hat{y} \\ z \end{bmatrix}.$$

No matter the value of z, it won't affect the value of $\hat{y}$. We can define many $\mathbf{x}$ values such that the least square problems has the minimum. However, we want to define the unique solution $\mathbf{x}$, so how do we choose z so we can find the minimized $||\mathbf{x}||_2$?

$$||\mathbf{x}||_2^2 = \left|\left| V \begin{bmatrix} \hat{y} \\ z \end{bmatrix} \right|\right|_2^2 = \left|\left| \begin{bmatrix} \hat{y} \\ z \end{bmatrix} \right|\right|_2^2 = ||\hat{y}||_2^2 + ||z||_2^2,$$

where recall that V is orthogonal. So, $||\mathbf{x}||_2^2$ is minimized when $z = \mathbf{0}$. So, in summary, $\boxed{\mathbf{x} = V \begin{bmatrix} \hat{y} \\ \mathbf{0} \end{bmatrix}}$ is the unique solution.