Ex 1

1) By definition, $\mathcal{R}(A)$ is the set of vectors y of the form $y = Ax$ for some x. Hence the only way for there to exist an x such that $y = Ax$ is to have $y \in \mathcal{R}(A)$.

2) This means that <u>every</u> y must be able to be written as $y = Ax$ for some x. In this case, the columns of A must span whatever space y lives in, which is $\mathbb{F}^M$ for $A \in \mathbb{F}^{M \times N}$.

Ex 2

We want $w \in \mathbb{R}^D$ such that $y_i \approx w^T x_i$. Stacking these, we get

$$\begin{aligned} y_1 &= w^T x_1 \\ y_2 &= w^T x_2 \\ &\vdots \\ y_N &= w^T x_N \end{aligned} \iff \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{y \in \mathbb{R}^N} = \underbrace{\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}}_{X \in \mathbb{R}^{N \times D}} w$$

$\uparrow$ data matrix

where we used the fact that $w^T x = x^T w$.

Ex 3

We have

$$f(x) = \|Ax - y\|^2 = (Ax - y)^T (Ax - y)$$

$$= x^T A^T A x + y^T y - 2 y^T A x$$

First let $v = A^T y$ and consider $\nabla_x v^T x$. For arbitrary x_i, we have

$$\frac{\partial}{\partial x_i} v^T x = \frac{\partial}{\partial x_i} \sum_{i=1}^{N} v_i x_i = v_i .$$

Stacking all these gives

$$\nabla_x v^T x = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = v$$

and substitution gives

$$\nabla_x y^T A x = \nabla_x (A^T y)^T x = A^T y .$$

Next consider the $x^T A^T A x$ term. From HW1, P7, we know that

$$x^T B x = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i B_{ij} x_j$$

substitute $B = A^T A$

Further, consider the case where B is symmetric, since $A^T A$ is.

Differentiating with respect to x_i, we need to account for x_i showing up in both sums, which gives

$$\frac{\partial}{\partial x_i} x^T B x = \sum_{j=1}^{n} B_{ij} x_j + \sum_{j=1}^{n} x_j B_{ji}$$

$$= 2 \sum_{j=1}^{n} B_{ij} x_j \qquad \text{by symmetry of } B,\ B_{ij} = B_{ji}$$

$$= 2 B_{i,:} x$$

Stacking all these, we get

$$\nabla_x x^T B x = \begin{bmatrix} B_{1,:} x \\ B_{2,:} x \\ \vdots \\ B_{n,:} x \end{bmatrix} = Bx$$

note the "inner product" view of matrix-vector multiplication from pg. 1.24 of the notes

$$\Rightarrow \nabla_x x^T A^T A x = A^T A x \quad \text{by substituting } A^T A = B.$$

Putting this all together, we get

$$\nabla_x \; x^T A^T A x + y^T y - 2 y^T A x = 2 A^T A x - 2 A^T y.$$

Ex 4

For A^TA to be invertible, it must be full rank. Using the SVD, we have

$$\begin{aligned} A^TA &= V\Sigma^T U^T U \Sigma V^T \\ &= U\Sigma^T\Sigma V^T \end{aligned}$$

where U is orthogonal, so $\operatorname{rank}(A^TA) = \operatorname{rank}(\Sigma^T\Sigma)$. Note that $\Sigma \in \mathbb{R}^{M\times N}$, so $\Sigma^T\Sigma \in \mathbb{R}^{N\times N}$ is the diagonal matrix of the form

$$\begin{bmatrix} \sigma_1^2 & & & & & & \\ & \sigma_2^2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r^2 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

Thus A^TA is full rank if $r = N$ above, or when $\operatorname{rank}(A) = N$. In other words, A^TA is invertible when the columns of A are linearly independent.

Ex 5

Full. Note that we use the fact that $UU^T = I$.

Ex 6

$$\underbrace{\Sigma_r^{-1} U_r^T y}_{v \in \mathbb{R}^r} = \begin{bmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_r \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_r \end{bmatrix} = \begin{bmatrix} v_1/\sigma_1 \\ v_2/\sigma_2 \\ \vdots \\ v_r/\sigma_r \end{bmatrix}$$

Now inspect $v = U_r^T y$

$$U_r^T y = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_r^T \end{bmatrix} y = \begin{bmatrix} u_1^T y \\ u_2^T y \\ \vdots \\ u_r^T y \end{bmatrix}$$

where u_i is the i^{th} column of U. Putting these together

$$\Sigma_r^{-1} U_r^T y = \begin{bmatrix} u_1^T y / \sigma_1 \\ u_2^T y / \sigma_2 \\ \vdots \\ u_r^T y / \sigma_r \end{bmatrix}$$

Ex 7

Recall we defined $z = V^T \hat{x}$ and V is orthogonal, so $V^{-1} = V^T$. This gives

$$\hat{x} = V V^T \hat{x} = V z$$

Ex 8

The number of LI columns in A is $\dim(\mathcal{R}(A))$. Note that $\mathcal{R}(A) \in \mathbb{R}^M$ which has maximum dimension M. If $N>M$ columns are LI, there would exist $N>M$ columns in $\mathbb{R}^M$ that are LI, contradicting the definition of dimension of a vector space.

Ex 9

Let $A \in \mathbb{R}^{m\times n}$ with $m \geq n$. Recall the SVD sizes are

$$A = \underset{m\times m}{U} \quad \underset{m\times n}{\Sigma} \quad \underset{n\times n}{V^T}$$

$$= \begin{bmatrix} \underset{m\times r}{U_r} & \underset{m\times(m-r)}{U_0} \end{bmatrix} \begin{bmatrix} \Sigma_r & \overset{r\times(n-r)}{0} \\ \underset{(m-r)\times r}{0} & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_0^T \end{bmatrix}$$

$V_r : n\times r$

$V_0 : n-r \times r$

If $r=n$, we get V_0 of size $n-r = n-n = 0$.

Ex 10

$$A^+ = (U\Sigma V^T)^+ = (U_r \Sigma_r V_r^T)^+$$

$$= (U_r (\Sigma_r V_r^T))^+$$

$$= (\Sigma_r V_r^T)^+ U_r^T$$ property (P1) since U_r has orthonormal columns

$$= V_r \Sigma_r^+ U_r^T$$ property (P2) since V_r^T has orthonormal rows

$$= V_r \Sigma_r^{-1} U_r^T$$ since Σ_r is invertible

Ex 11

$$\Sigma^+ = \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

1) $$\Sigma\Sigma^+\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}\Sigma = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} = \Sigma$$

2) $$\Sigma^+\Sigma\Sigma^+ = \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}\Sigma_r^+ = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \Sigma^+$$

3) $\Sigma^+\Sigma = \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ which is symmetric

4) $\Sigma\Sigma^+ = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ which is symmetric

Ex 12

By definition, $N(A) = \{x \in \mathbb{R}^n : Ax = 0\}$.

Ex 13

Among all vectors x of the form $x = A^+y + x_N$ where $x_N \in N(A)$, we want to find the one with minimum norm.

Ex 14

$A^+y = V_r\Sigma_r^{-1}U_r^Ty$. Let $v = \Sigma_r^{-1}U_r^Ty$. Then $A^+y = V_r y$, which satisfies the definition of a vector lying in $\mathcal{R}(V_r)$.

Ex 15

Using the normal equations, the solution satisfies $A^TAx = A^Tb$, but $a, b \in \mathbb{R}^m$ so $A^TA = a^Ta \in \mathbb{R}$ and $A^Tb = a^Tb \in \mathbb{R}$ as well. The solution is then

scalar

$$\hat{x} = \frac{a^Tb}{a^Ta} = \frac{a^Tb}{\|a\|^2} = \frac{\|a\|\,\|b\|\cos\theta}{\|a\|^2} = \frac{\|b\|}{\|a\|}\cos\theta$$

Now the LS cost becomes

$$\|a\hat{x}-b\|^2 = \left\| \frac{\|b\|}{\|a\|}\cos\theta\, a - b \right\|^2$$

$$= \left(\frac{\|b\|}{\|a\|}\cos\theta\, a - b\right)^T\left(\frac{\|b\|}{\|a\|}\cos\theta\, a - b\right)$$

$\|a\|\,\|b\|\cos\theta$

$$= \frac{\|b\|^2}{\|a\|^2}\cos^2\theta\, a^Ta + b^Tb - 2\frac{\|b\|}{\|a\|}\cos\theta\, a^Tb$$

$$= \|b\|^2\cos^2\theta + \|b\|^2 - 2\|b\|^2\cos^2\theta$$

$$= \|b\|^2\left(1-\cos^2\theta\right)$$

$$= \|b\|^2\sin^2\theta$$

$\sin^2\theta + \cos^2\theta = 1$