Exercise 1. Subspace Sum

Suppose $U_1, U_2, \ldots, U_m$ are subspaces of a vector space V. The (subspace) sum of $U_1, U_2, \ldots, U_m$, denoted by $U_1 + U_2 + \ldots + U_m$ is defined to be the set of all possible sums from elements of $U_1, U_2, \ldots, U_m$:

$$U_1 + U_2 + \ldots + U_m = \{u_1 + u_2 + \ldots + u_m : u_1 \in U_1, \ldots, u_m \in U_m\}$$

(NB: This is not to be confused with the (subspace) direct sum $U_1 \oplus U_2 \oplus \cdots \oplus U_m = \{(u_1, \ldots, u_m) : u_1 \in U_1, \ldots, u_m \in U_m\}$.)

1.) Is $U_1 + U_2 + \ldots + U_m$ a subspace of V?

2.) Prove or give a counterexample: if U_1, U_2, W are subspaces of V such that $U_1 + W = U_2 + W$, then $U_1 = U_2$.

1.) Yes. It's clear that $U_1 + \ldots + U_m \subseteq V$, because U_i is subspace. And for any $u_i \in U_i$, we have $u_{sum} = u_1 + \ldots + u_m \in U_{sum} = U_1 + \ldots + U_m$. If U_{sum} is a subspace of V, the vector u_{sum} in it should sastisfy: ① $\forall \alpha, \beta \in U_{sum}$, $\alpha + \beta \in U_{sum}$ ② $\forall \alpha \in U_{sum}$, $\lambda \in F$, $\lambda\alpha \in U_{sum}$ ③ $0 \in U_{sum}$, according to linear algebra.

Since U_i is subspace, so let $u_i = 0$, we have $0_1 + \ldots + 0_m = 0 \in U_1 + \ldots + U_m$, which satisfies "③".

Since $u^\alpha_{sum} + u^\beta_{sum} = u'_\alpha + \ldots u^m_\alpha + u'_\beta + \cdots + u^m_\beta = (u'_\alpha + u'_\beta) + \cdots + (u^m_\alpha + u^m_\beta)$ and U_i is a subspace, so $u^i_\alpha + u^i_\beta \in U_i$ $(u^i_\alpha, u^i_\beta \in U_i)$, let $u^i_\alpha + u^i_\beta = u^i_{\alpha\beta}$, hence $u^\alpha_{sum} + u^\beta_{sum} = u'_{\alpha\beta} + \ldots + u^m_{\alpha\beta} \in U_1 + \ldots + U_m = U_{sum}$, i.e. for $\forall \alpha, \beta \in U_{sum}$, $\alpha + \beta \in U_{sum}$, which satisfies "①".

Since $\lambda(u_1 + \ldots + u_m) = \lambda u_1 + \ldots + \lambda u_m$ $(\lambda \in F)$ and $\lambda u_i \in U_i$, hence $\lambda(u_1 + \ldots + u_m) = \lambda u_1 + \ldots + \lambda u_m \in U_1 + \ldots + U_m = U_{sum}$, i.e. for $\forall \alpha \in U_{sum}$, $\lambda \in F$, we have $\lambda\alpha \in U_{sum}$, which satisfies "②".

So in conclusion, $U_1 + \ldots + U_2 + \ldots + U_m$ is a subspace of V.

2) Proof: As we all know, a vector space is a subspace of itself. Assume that U_1 and U_2 are different subspaces of W, which are also subspaces of V. Hence according to question 1), we have:

$$\begin{cases} U_1 \subseteq W \subseteq V \\ U_2 \subseteq W \subseteq V \\ U_1 \neq U_2 \end{cases} \Longrightarrow \begin{cases} U_1 + W = W \text{ is a subspace of } W \\ U_2 + W = W \text{ is a subspace of } W \end{cases}$$

Hence we have $U_1 + W = U_2 + W$, but $U_1 \neq U_2$, which means "2)" is not true if W has at least two diffrent subspaces, such as $W = \mathbb{R}^3$.

Exercise 2. NORMS

Let $x \in \mathbb{R}^n$ and $f : [0,1] \to \mathbb{R}$ be a continuous function. Prove that the following are norms:

1.) ($\ell^1(\mathbb{R}^n)$-Norm) $\|x\|_1 := \sum_{i=1}^n |x_i|$;

2.) ($L^\infty([0,1])$-Norm) $\|f\|_\infty = \sup_{t\in[0,1]} |f(t)|$.

Proof: A norm should satisfy: $\begin{cases} \|x\| \geq 0, \forall x \in V, \|x\| = 0 \Leftrightarrow x = 0 \\ \|\alpha x\| = |\alpha| \|x\|, \forall \alpha \in \mathbb{R}, \forall x \in V \\ \|x+y\| \leq \|x\| + \|y\|, \forall x, y \in V \end{cases}$

1.) ① It's clear that $\sum_i |x_i| \geq 0, \forall x \in V$ and "=" iff $x = 0$.

② $\|\alpha x\|_1 := \sum_{i=1}^n |\alpha x_i| = \sum_{i=1}^n |\alpha| |x_i| = |\alpha| \sum_{i=1}^n |x_i| = |\alpha| \|x\|_1, \forall \alpha \in \mathbb{R}$

③ $\|x+y\|_1 = \sum_{i=1}^n |x_i + y_i| \leq \sum_{i=1}^n (|x_i| + |y_i|) = \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i|$
$= \|x\|_1 + \|y\|_1$ $(\forall x, y \in \mathbb{R}^n)$

So in conclusion, $\|x\|_1$ is a norm.

2.) ① It's clear that $\sup_{t\in[0,1]} |f(t)| \geq 0$ and "=" iff $f(t) \equiv 0$.

② $\|\alpha f\|_\infty = \sup_{t\in[0,1]} |\alpha f(t)| = |\alpha| \sup_{t\in[0,1]} |f(t)| = |\alpha| \|f\|_\infty$

③ $\|f_1 + f_2\|_\infty = \sup_{t\in[0,1]} |f_1(t) + f_2(t)| \leq \sup_{t\in[0,1]} (|f_1(t)| + |f_2(t)|) =$
$\sup_{t\in[0,1]} |f_1(t)| + \sup_{t\in[0,1]} |f_2(t)| = \|f_1\|_\infty + \|f_2\|_\infty$

So in conclusion, $\|f\|_\infty$ is a norm.

Exercise 3. COMPARING NORMS

Prove the inequality on the norms of $\mathbb{R}^n$, $\|x\|_2 \leq \|x\|_1 \leq n^{\frac{1}{2}}\|x\|_2$.

Proof: $\|X\|_2 = \sqrt{XX^T} = \sqrt{\sum_{i=1}^{n} X_i^2} \leq \sum_{i=1}^{n} |X_i| = \|X\|_1$ ①

Consider the Cauchy Inequality: $\|\langle X, y\rangle\| \leq \|X\|\|y\|$, let $y = (\frac{1}{n}, \cdots, \frac{1}{n})$

$\|\langle X, y\rangle\| = \left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \leq \|X\|\|y\| = \sqrt{\sum_{i=1}^{n} X_i^2} \cdot \sqrt{n \cdot \frac{1}{n^2}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$

$\therefore \sum_{i=1}^{n} |X_i| \leq \sqrt{n\sum_{i=1}^{n} X_i^2} \Longrightarrow \|X\|_1 \leq n^{\frac{1}{2}}\|X\|_2$ ②

(① is clear because $(\sum_{i=1}^{n}|X_i|)^2 = \sum_{i=1}^{n} X_i^2 + R$ $(R \geq 0)$)

So in conclusion, $\|X\|_2 \leq \|X\|_1 \leq n^{\frac{1}{2}}\|X\|_2$.

Exercise 4. INNER PRODUCTS

1.) For any finite-dimensional, positive definite matrix $Q \in \mathbb{R}^{n,n}$, prove that $\langle x, y\rangle_Q = x^\top Q y$ defines an inner product on $\mathbb{R}^n$.

2.) If $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ defines an inner product, then prove that $\sqrt{\langle \cdot, \cdot \rangle}$ defines a norm on V.

Proof: $\mathbb{R}^n$ inner Product should satisfy:

$$\begin{cases} \langle , \rangle : V \times V \to \mathbb{R}^n \\ \langle X, y\rangle = \langle y, X\rangle \\ \langle aX + by, z\rangle = a\langle X, z\rangle + b\langle y, z\rangle \\ \langle X, X\rangle \geq 0, \text{"="} \text{ iff } X = 0. \\ (\forall X, y, z \in V, a, b \in \mathbb{R}) \end{cases}$$

1.) $X^TQy = q_{11}X_1y_1 + q_{12}X_1y_2 + \cdots + q_{1n}X_1y_n$
$+ q_{21}X_2y_1 + q_{22}X_2y_2 + \cdots + q_{2n}X_2y_n$
$+ \cdots$
$+ q_{n1}X_ny_1 + q_{n2}X_ny_2 + \cdots + q_{nn}X_ny_n$

So it's easy to know that if we exchange X and y, the value of X^TQy won't change, i.e. $X^TQy = y^TQX$. Hence, $\langle X, y\rangle_Q = \langle y, X\rangle_Q$.

And $\langle aX + by, z\rangle_Q = (aX + by)^TQz = z^TQ(aX + by) = z^TQaX + z^TQby$
$= az^TQX + bz^TQy = a\langle z, X\rangle_Q + b\langle z, y\rangle_Q = a\langle X, z\rangle_Q + b\langle y, z\rangle$

And $\langle X, X\rangle_Q = X^TQX$ is a quadratic form of X, since Q is a positive define matrix, hence $x^TQx \geq 0$, "=" iff $x=0$. By the way, $x,y \in R^n$.

So in conclusion, $\langle x,y\rangle_Q$ defines a inner product on R^n.

2.) let "·" to be x, $f(x)=\sqrt{\langle x,x\rangle}$;

① since $\langle x,x\rangle \geq 0$, "=" iff $x=0$, hence $f(x)\geq 0$ and "=" iff $x=0$.

② $f(\alpha x)=\sqrt{\langle \alpha x,\alpha x\rangle}=\sqrt{\alpha\langle x,\alpha x\rangle}=\sqrt{\bar{\alpha}\langle \alpha x,x\rangle}=\sqrt{\alpha^2\langle x,x\rangle}=|\alpha|\sqrt{\langle x,x\rangle}=|\alpha| f(x)$

③ Since $(x+\lambda y, x+\lambda y)=(x,x)+\bar{\lambda}(x,y)+\lambda(y,x)+|\lambda|^2(y,y)$

$\geq 0 \quad (\forall x,y\in V, \forall \lambda\in C)$

if $y\neq 0$, let $\lambda=-\frac{(x,y)}{(y,y)}$, we have: $(x,x)-\frac{(y,x)}{(y,y)}(x,y)$

$-\frac{(x,y)}{(y,y)}(y,x)+\left|\frac{(x,y)}{(y,y)}\right|^2(y,y)=(x,x)-\frac{2(x,y)(y,x)}{(y,y)}$

$+\frac{|(x,y)|^2}{(y,y)}\geq 0$

hence $(x,x)-\frac{2|(x,y)|^2}{(y,y)}+\frac{|(x,y)|^2}{(y,y)}\geq 0$

$\Rightarrow (x,x)(y,y)\geq |(x,y)|^2$.

Since $f(x+y)=\sqrt{\langle x+y,x+y\rangle}=\sqrt{\langle x,x\rangle+\langle y,y\rangle+\langle x,y\rangle+\langle y,x\rangle}$

$f(x)+f(y)=\sqrt{\langle x,x\rangle}+\sqrt{\langle y,y\rangle}$

hence $[f(x+y)]^2=\langle x,x\rangle+\langle y,y\rangle+\langle x,y\rangle+\langle y,x\rangle$

$[f(x)+f(y)]^2=\langle x,x\rangle+\langle y,y\rangle+2\sqrt{\langle x,x\rangle\langle y,y\rangle}$

$\geq \langle x,x\rangle+\langle y,y\rangle+2|\langle x,y\rangle|$

let $\langle x,y\rangle=a+bi$, then $\langle y,x\rangle=\overline{\langle x,y\rangle}=a-bi$

Since it's obvious $\langle x,y\rangle+\langle y,x\rangle=\langle x,y\rangle+\overline{\langle x,y\rangle}=2a\leq$
$2|(x,y)|^2=2\sqrt{a^2+b^2}\geq 2|a|$
Hence $[f(x+y)]^2\leq[f(x)+f(y)]^2 \Rightarrow f(x+y)\leq f(x)+f(y)$
And if $y=0$, it's obvious that $f(x+y)=f(x)+f(y)$.
So in conclusion, $f(x)$ satisfies all the conditions of a norm, which means $\sqrt{\langle\cdot,\cdot\rangle}$ defines a norm on V.

Exercise 5. LINEAR TRANSFORMATIONS I

Given $A, B, C \in \mathbb{C}^{n\times n}$, determine if the following maps (involving matrix multiplication) from $A, B, C \in \mathbb{C}^{n\times n}$ are linear:

1.) $X \mapsto AX + XB$;

2.) $X \mapsto AX + BXC$; and,

3.) $X \mapsto AX + XBX$.

Solution: A linear map should satisfy: $f(\alpha v+\beta u)=\alpha f(v)+\beta f(u)$
$(\forall \alpha,\beta,v,u)$

☆ let X to be replaced by $\alpha v+\beta u$, then we have:

1.) $f(\alpha v+\beta u)=A(\alpha v+\beta u)+(\alpha v+\beta u)B$
$=\alpha Av+\beta Au+\alpha vB+\beta uB=\alpha(Av+vB)+\beta(Au+uB)$
$=\alpha f(v)+\beta f(u)$, hence this is a linear map.

2.) $f(\alpha v+\beta u)=A(\alpha v+\beta u)+B(\alpha v+\beta u)C$
$=\alpha Av+\beta Au+(\alpha Bv+\beta Bu)C$
$=\alpha Av+\beta Au+\alpha BvC+\beta BuC$
$=\alpha(Av+BvC)+\beta(Au+BuC)=\alpha f(v)+\beta f(u)$,

hence this is also a linear map.

3.) $f(\alpha v+\beta u)=A(\alpha v+\beta u)+(\alpha v+\beta u)B(\alpha v+\beta u)$

$=\alpha Av+\beta Au+(\alpha v+\beta u)(\alpha Bv+\beta Bu)$

$=\alpha Av+\beta Au+\alpha v\alpha Bv+\beta u\alpha Bv+\alpha v\beta Bu+\beta u\beta Bu$

$=\alpha Av+\beta Au+\alpha^2 vBv+\alpha\beta uBv+\alpha\beta vBu+\beta^2 uBu$

$=\alpha(Av+\alpha vBv)+\beta(Au+\beta uBu)+\alpha\beta uBv+\alpha\beta vBu$

$\neq \alpha f(v)+\beta f(u)=\alpha(Av+vBv)+\beta(Au+uBu)$

So this is not a linear map.

Exercise 6. Range and Null spaces

Given finite dimensional spaces V and W, and a linear operator $A\colon V\to W$. Let $\mathsf{im}(A)$ and $\mathsf{ker}(A)$ respectively denote the range space and null space (or kernel) of A. Prove the following:

1.) The spaces $\mathsf{im}(A)$ and $\mathsf{ker}(A)$ are subspaces;

2.) [Rank-Nullity Theorem] $\dim V=\dim \mathsf{im}(A)+\dim \mathsf{ker}(A)$.

Proof:

1.) $im(A)=\{w\in W: \text{there exists } v\in V \text{ s.t. } Av=w\}$, since $im(A)\subseteq W$, hence $im(A)$ is a subspace of W iff $\begin{cases} ① \ \forall \alpha,\beta\in im(A), (\alpha+\beta)\in im(A) \\ ② \ \forall \lambda\in F, \forall\in im(A), \lambda\alpha\in im(A) \\ ③ \ 0\in im(A) \end{cases}$

Since $0\in V$ and $0\in W$, hence $A0=0\in im(A)$.

Since $\lambda v\in V$, $\lambda w\in W$, hence $A\lambda v=\lambda w\in im(A)$. (Assume $Av=w$)

Since $(v_\alpha+v_\beta)\in V$, $(w_\alpha+w_\beta)\in W$, hence $A(v_\alpha+v_\beta)=(w_\alpha+w_\beta)\in im(A)$ $\left(\text{Assume } \begin{matrix} Av_\alpha=w_\alpha \\ Av_\beta=w_\beta \end{matrix}\right)$

So in conclusion, $im(A)$ is a subspace of W.

$ker(A)=\{v\in V: Av=0\}$. Since $A0=0$, hence $0\in ker(A)$. Since $A\alpha v=\alpha Av=\alpha 0=0$, hence $\alpha v\in ker(A)$. Since $A\alpha=0$, $A\beta=0\Rightarrow A(\alpha+\beta)=0$, hence $\forall\alpha,\beta\in ker(A)$, $(\alpha+\beta)\in ker(A)$. So in conclusion, $ker(A)$ is a subspace.

2.)

Assume $\dim(V)=n$, $\dim(W)=m$, so that $A=A_{n\times m}$. We know that $\ker(A)=N(A)$ (nullspace) and $\mathrm{im}(A)=C(A)$ (column space). Hence in order to prove $\dim V=\dim \mathrm{im}(A)+\dim\ker(A)$, we should prove that:

$n=\dim C(A)+\dim N(A)$. Since $\dim C(A)=\mathrm{rank}(A)$ is the number of pivot columns and $\dim N(A)$ is the number of free variables, i.e. the number of non-pivot columns. Hence it obvious that $\dim C(A)+\dim N(A)=n$ = pivot + non-pivot columns. So in conclusion, we prove $\dim(V)=\dim\mathrm{im}(A)+\dim\ker(A)$.

Exercise 7. Matrix Norm

Let $A\in\mathbb{R}^{m\times n}$. Prove that the induced matrix norm $\|A\|_1=\max_{1\le j\le n}\sum_{i=1}^m|a_{ij}|$ and $\|A\|_\infty=\max_{1\le i\le m}\sum_{j=1}^n|a_{ij}|$

Proof: 1) $\|AX\|_1=\sum_{i=1}^m\left|\sum_{j=1}^n a_{ij}x_j\right|\le\sum_{i=1}^m\sum_{j=1}^n|x_j|\max_{1\le j\le n}|a_{ij}|=\|X\|_1\max_{1\le j\le n}\sum_{i=1}^m|a_{ij}|$

$\therefore \|A\|_1=\sup_{X\ne0}\frac{\|AX\|_1}{\|X\|_1}\le\max_{1\le j\le n}\sum_{i=1}^m|a_{ij}|$. Now we should prove "=" can be satisfied: assume that when $j=p$, $\frac{\|AX\|_1}{\|X\|_1}$ get the maximum value and let $x_j=\begin{cases}1, & j=p\\ 0, & j\ne p\end{cases}$ so that $\|X\|_1=1$ and $\frac{\|AX\|_1}{\|X\|_1}=\sum_{i=1}^m\left|\sum_{j=1}^n a_{ij}x_j\right|$

$=\sum_{i=1}^m|a_{ip}|$

so that there is at least one X can satisfy $\|A\|\le\max_{1\le j\le n}\sum_{i=1}^m|a_{ij}|$, which means $\|A\|_1=\max_{1\le j\le n}\sum_{i=1}^m|a_{ij}|$.

$=\sum_{i=1}^m|a_{ij}|\ (j=p)$

2.) $\|Ax\|_\infty=\max_{1\le i\le m}\left|\sum_{j=1}^n a_{ij}x_j\right|\le\max_{1\le i\le m}\sum_{j=1}^n|a_{ij}x_j|\le\max_{1\le j\le n}|x_j|\max_{1\le i\le m}\sum_{j=1}^n|a_{ij}|$

$=\|x\|_\infty\max_{1\le i\le m}\sum_{j=1}^n|a_{ij}|$

Hence $\|A\|_\infty=\sup_{x\ne0}\frac{\|Ax\|_\infty}{\|x\|_\infty}\le\max_{1\le i\le m}\sum_{j=1}^n|a_{ij}|$. Assume $\sum_{j=1}^n|a_{pj}|=\max_{1\le i\le m}\sum_{j=1}^n|a_{ij}|=R$ and let $y=[y_1,y_2,\dots,y_n]^T$, where:

$y_i = \begin{cases} 1, & a_{pj} \geq 0 \\ -1, & a_{pj} < 0 \end{cases}$, so that $\|x\|_\infty = 1$ and $\|A\|_\infty = \max_{\|x\|_\infty = 1} \|Ax\|_\infty \geq \|Ay\|_\infty$

$= \max_{1 \leq i \leq m} \left|\sum_{j=1}^n a_{ij} y_j\right| = \sum_{j=1}^n |a_{pj}|$

$= \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$.

So in conclusion, $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$.

Exercise 8. PSEUDO-INVERSE

Suppose a matrix $A \in \mathbb{R}^{m \times n}$ is decomposed in the form

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^*$$

where U and V are unitary matrices and Σ is an invertible $r \times r$ matrix (the singular value decomposition could be used to produce such a decomposition). Then the "Moore-Penrose inverse", or pseudo-inverse of A, denoted by A^+, can be defined as the $n \times m$ matrix

$$A^+ = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^*.$$

(You can invoke it in Matlab with pinv(A))

1.) Show that A^+A and AA^+ are symmetric, and that $AA^+A = A$ and $A^+AA^+ = A^+$. (These four conditions actually constitute an alternative definition of the pseudo-inverse.)

2.) Show that when A has full column rank then $A^+ = (A^*A)^{-1}A^*$ and that when A has full row rank then $A^+ = A^*(AA^*)^{-1}$.

3.) Show that of all x that minimize $\|y - Ax\|_2$ (and there will be many if A does not have full column rank), the one with smallest length $\|x\|_2$ is given by $\hat{x} = A^+y$.

Proof:

1.) $(A^+A)^* = \left(V\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix}U^*U\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}V^*\right)^*$, since U, V are unitary matrices, i.e. $U^* = U^{-1}$, so that $(A^+A)^* = \left(V\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}V^*\right)^*$, where I is identity matrix, hence $(A^+A)^* = (V^*)^*\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}V^* = V\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}V^* = A^+A$. And for AA^+, is obvious that we can just exchange V, U and Σ^{-1}, Σ, and hence: $(AA^+)^* = \left(U\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}U^*\right)^* = U\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}U^* = AA^+$. So in conclusion, AA^+ and A^+A are symmetric.

And $AA^+A = U\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}U^* \cdot U\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}V = U\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}V = A$

$A^+AA^+ = V\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}V^*V\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix}U^* = V\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix}U^* = A^+$

2) Since $\text{rank}(A) \le \min(m,n)$, so if A has full column rank, we have $n \le m$ and $A = U\begin{pmatrix}\Sigma\\0\end{pmatrix}V^*$ $(r=n)$, $A^+ = V\begin{pmatrix}\Sigma^{-1}\\0\end{pmatrix}U^*$.

$$(A^*A)^{-1}A^* = \left(V\begin{pmatrix}\Sigma^*\\0\end{pmatrix}U^*U\begin{pmatrix}\Sigma\\0\end{pmatrix}V^*\right)^{-1}\left(V\begin{pmatrix}\Sigma^*\\0\end{pmatrix}U^*\right)$$

$$= \left(V\begin{pmatrix}\Sigma^{-1}\\0\end{pmatrix}U^*U\begin{pmatrix}(\Sigma^*)^{-1}\\0\end{pmatrix}V^*\right)\left(V\begin{pmatrix}\Sigma^*\\0\end{pmatrix}U^*\right) = V\begin{pmatrix}\Sigma^{-1}\\0\end{pmatrix}U^* = A^+$$

similarly, whe A has full ~~row~~ column rank, we have:

$$A^*(AA^*)^{-1} = V(\Sigma^*\ 0)U^*(U(\Sigma\ 0)V^*V(\Sigma^*\ 0)U^*)^{-1}$$

$$= V(\Sigma^*0)U^*(U((\Sigma^*)^{-1}0)(\Sigma^{-1}0)U^*) = V(\Sigma^{-1}0)U^* = A^+$$

3) $\underset{x}{\text{minimize}}\|y-Ax\|_2 \iff \underset{x}{\text{minimize}}\|y-Ax\|_2^2 \iff (y-A\hat{x}) \perp Ax$

$\iff \langle y-\hat{x}, \tilde{x}\rangle = 0, \forall \tilde{x} \in X \iff (y-A\hat{x})^*A = 0 \iff A^*(y-A\hat{x}) = 0$

$\iff A^*A\hat{x} = A^*y$ (, when A has full column rank, $\hat{x} = (A^*A)^{-1}A^*y$
$= A^+y$

And when A has not full column rank, Let $\hat{x} = A^+y = A^*(AA^*)^{-1}y$, we have $A^*A\hat{x} = A^*AA^*(AA^*)^{-1}y = A^*y$, which is also a solution. So now we should just prove $\|\hat{x}\|_2 \le \|x^+\|_2$, x^+ is any solution.

$$\begin{cases}A^*A\hat{x} = A^*y\\ A^*Ax^+ = A^*y\end{cases} \Rightarrow A^*A(\hat{x}-x^+) = 0 \Rightarrow \hat{x}-x^+ \in N(A^*A) = N(A),$$

if $\hat{x} \perp (\hat{x}-x^+)$, we have $\|x^+\|_2^2 = \|\hat{x}\|_2^2 + \|\hat{x}-x^+\|_2^2 \ge \|\hat{x}\|_2^2$. And since from exercise 6) we know that $n = \dim N(A) + \dim C(A)$. Since A and A^TA has the same n, and it's obvious $A^TAx=0$ and $Ax=0$ have the same solution, so $\dim C(A)$ and $\dim C(A^TA)$ are the same. Since $\dim C(A) = \dim C(A^T)$, hence $C(A^TA) = C(A^T)$. Since $\hat{x} = A^+y = A^*(AA^*)^{-1}y$ or $= (A^*A)^{-1}A^*$, so it's obvious that $\hat{x} \in C(A)$. Since $N(A) \perp Row(A) = C(A)$, so that $\hat{x} \perp (\hat{x}-x^+)$ is true, hence $\|\hat{x}\|_2^2 \le \|x^+\|_2^2$. So in conclusion, $\hat{x} = A^+y$ is the solution that has the smallest lenth.