<u>More algebra relevant to optimization</u>

Consider decomposed $A \in \mathbb{R}^{m\times n}$, as:

(9.1) $$A = P \begin{bmatrix} \hat{A}_{r\times r} & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

invertible

$\perp$ $\perp$

maybe with some 0-blocks missing

- useful for finding A^+
- useful for finding $\mathcal{R}[A]$, $\mathcal{N}[A] \triangleq \{x : Ax = 0\}$

useful later for constrained optimization

Aside: the so-called singular value decomposition yields (9.1) with diagonal $\hat{A}$

(9.2) Th For $A \in \mathbb{R}^{m\times n}$ of (9.1):

(i) $\mathcal{R}[A] = \mathcal{L}[\underbrace{p_{*1}, \dots, p_{*r}}_{\text{colums of } P}]$

form an orthonormal basis for $\mathcal{R}[A]$

(ii) $\mathcal{N}[A] = \mathcal{L}[\underbrace{q_{*(r+1)}, \dots, q_{*n}}_{\text{columns of } Q}]$

$= \mathcal{R}[H]$ where $H = [q_{*(r+1)}, \dots, q_{*n}]$

(9.3) all x's in $\mathcal{N}[A]$ are generated by $H\theta$ as θ varies over $\mathbb{R}^{n-r}$

If some of the 0-blocks in (9.1) are missing, the effects on Th. (9.2) are easy to figure out.

Proof of (i) uses:

(9.4) <u>Lemma</u> If $B \in \mathbb{R}^{p \times p}$ is non-singular then $\mathcal{R}[B] = \mathbb{R}^p$

Aside: To show that:
set X = set Y
we have to show that
$X \subset Y$ and $Y \subset X$
To show that $X \subset Y$:
show $(x \in X) \Rightarrow (x \in Y)$

<u>Proof of Lemma (9.4):</u>

Proof that $\mathcal{R}[B] \subset \mathbb{R}^p$:

$$\mathcal{R}[B] = \{ \underbrace{Bx}_{\in \mathbb{R}^p} : x \in \mathbb{R}^p \} \subset \mathbb{R}^p$$

Proof that $\mathbb{R}^p \subset \mathcal{R}[B]$:

$$y \in \mathbb{R}^p \Rightarrow y = B\underbrace{B^{-1}y}_{x} = Bx \text{ for some } x$$

$$\Rightarrow y \in \{ Bx : x \in \mathbb{R}^p \} = \mathcal{R}[B]$$

QED

Proof of (i):

$$\mathcal{R}[A] = \{ Ax : x \in \mathbb{R}^n \}$$

$$= \{ P \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} \underbrace{Q^T x}_{y} : x \in \mathbb{R}^n \}$$

$\boxed{\text{$y$ varies over $\mathcal{R}[Q^T] = \mathbb{R}^n$ because Q^T = invertible}}$

$$= \{ P \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} y : y \in \mathbb{R}^n \}, \quad y = \begin{pmatrix} \tilde{y} \\ \bar{y} \end{pmatrix}$$

$$= \{ P \begin{bmatrix} \hat{A}\tilde{y} \\ 0 \end{bmatrix} : \begin{pmatrix} \tilde{y} \\ \bar{y} \end{pmatrix} \in \mathbb{R}^n \}, \quad z = \hat{A}\tilde{y}$$

$\boxed{\text{$z$ varies over $\mathcal{R}[\hat{A}] = \mathbb{R}^r$ - because $\hat{A}$ = invertible}}$

$$= \{ P \begin{pmatrix} z \\ 0 \end{pmatrix} : z \in \mathbb{R}^r \} = \{ \sum_{i=1}^{r} P_{*i} z_i : z \in \mathbb{R}^r \}$$

$$= \mathcal{L}[P_{*1}, \dots, P_{*r}]$$

QED

Orthogonal complement of a linear subspace

For a l.s.s. L of $\mathbb{R}^n$:

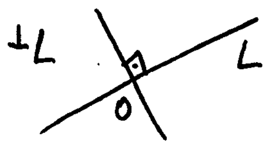
$\boxed{\text{linear sub-space}}$

$\boxed{\text{the orthogonal complement of } L}$

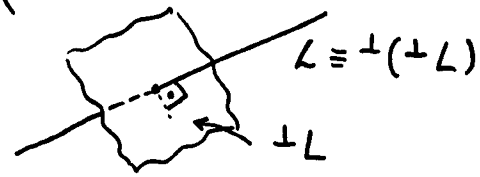
$^{\perp}L \triangleq$ the set of vectors which are orthogonal to every vector in L

$= \{ z \in \mathbb{R}^n : z^T x = 0, \ \forall x \in L \}$

in $\mathbb{R}^2$:



in $\mathbb{R}^3$:



(9.5) $\underline{\underline{\text{Th}}}$ For a l.s.s. L of $\mathbb{R}^n$:

(i) $^{\perp}L$ is a l.s.s. of $\mathbb{R}^n$

(ii) $^{\perp}(^{\perp}L) = L$

Orthogonal projection

(9.6) $\underline{\underline{\text{Th}}}$ Suppose L = a l.s.s. of $\mathbb{R}^n$

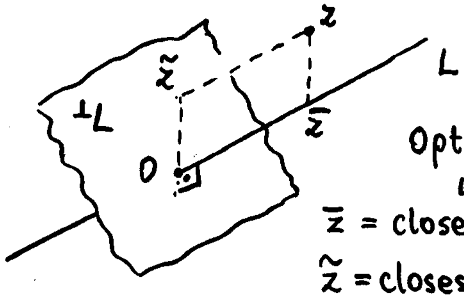
Then : every $z \in \mathbb{R}^n$ can be written as :

$$z = \underset{L}{\underset{\cap}{\bar{z}}} + \underset{^{\perp}L}{\underset{\cap}{\tilde{z}}}$$

$\boxed{\text{the orthogonal projection of } z \text{ on } L}$ $\boxed{\text{the orthogonal projection of } z \text{ on } ^{\perp}L}$

for unique $\bar{z}, \tilde{z}$

in $\mathbb{R}^3$:



Optimization again !

$\bar{z}$ = closest pt. in L to z

$\tilde{z}$ = closest pt. in $^{\perp}L$ to z

(9.7) $\underline{\underline{\text{Th}}}$ If $L = \mathcal{R}[A]$ for some $A \in \mathbb{R}^{m \times n}$

then : $\bar{z} = AA^+ z$

$\tilde{z} = (I - AA^+) z$

(£)

A useful fact about $A \in \mathbb{R}^{m\times n}$

(9.8) Th $\quad {}^{\perp}\mathcal{R}[A] = \mathcal{N}[A^T]$

Proof that ${}^{\perp}\mathcal{R}[A] \subset \mathcal{N}[A^T]$:

$$z \in {}^{\perp}\mathcal{R}[A]$$

$\Rightarrow z^T y = 0 \,,\ \forall y \in \mathcal{R}[A]$

$\Rightarrow z^T A x = 0 \,,\ \forall x \in \mathbb{R}^n$

$\Rightarrow (A^T z)^T x = 0 \,,\ \forall x \in \mathbb{R}^n$

$\boxed{\text{choose } x = A^T z}$

$\Rightarrow (A^T z)^T A^T z = 0 \ \Rightarrow \|A^T z\|^2 = 0$

$\Rightarrow A^T z = 0$

$\Rightarrow z \in \mathcal{N}[A^T]$

Hence $\ {}^{\perp}\mathcal{R}[A] \subset \mathcal{N}[A^T]$

(£) closest point in $\mathcal{R}[A] = L$ to z:

$$\underbrace{\hat{x}}_{A^+ z} = \arg\min_{x \in \mathbb{R}^n} \|Ax - z\|$$

$\Rightarrow \bar{z} = A\hat{x} = AA^+ z$

Proof that $\mathcal{N}[A^T] \subset {}^{\perp}\mathcal{R}[A]$:

$$z \in \mathcal{N}[A^T]$$

$\Rightarrow A^T z = 0$

$\Rightarrow x^T (\underbrace{A^T z}_{y}) = 0 \,,\ \forall x \in \mathbb{R}^n$

$\boxed{x^T y = \sum x_i y_i = \sum y_i x_i = y^T x}$

$\Rightarrow (A^T z)^T x = 0 \,,\ \forall x \in \mathbb{R}^n$

$\Rightarrow z^T \underbrace{A x}_{y} = 0 \,,\ \forall x \in \mathbb{R}^n$

varies all over $\mathcal{R}[A]$ as x varies over $\mathbb{R}^n$

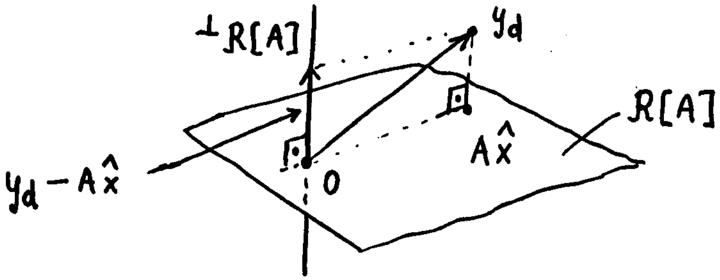
$\Rightarrow z^T y = 0 \,,\ \forall y \in \mathcal{R}[A]$

$\Rightarrow z \in {}^{\perp}\mathcal{R}[A]$

Hence $\mathcal{N}[A^T] \subset {}^{\perp}\mathcal{R}[A]$

QED

Application of Th. (9.8) to give increased insight for $\min_{x \in \mathbb{R}^n} \| y_d - Ax \|^2$



It is obvious geometrically that:

$y_d - A\hat{x} \perp$ to every vector in $\mathcal{R}[A]$

$$\Rightarrow \quad y_d - A\hat{x} \in {}^\perp\mathcal{R}[A] \underset{\boxed{\text{Th (9.8)}}}{=} \mathcal{N}[A^T]$$

$$\Rightarrow \quad A^T(y_d - A\hat{x}) = 0$$

$$\Rightarrow \quad A^T y_d = A^T A \hat{x}$$

assuming that $A^T A$ = invertible

$$\Rightarrow \quad \underbrace{\underset{\boxed{\text{unique}}}{\hat{x}} = (A^T A)^{-1} A^T y_d}$$

computable formula for $\hat{x}$!

So: if $A^T A$ = invertible

then the unique global minimizer is:

$$\hat{x} = (A^T A)^{-1} A^T y_d$$

Since there is only one global minimizer, it must be the global minimizer of least norm.

Hence:

$$A^+ y_d = (A^T A)^{-1} A^T y_d \quad , \ \forall y_d$$

$$\Rightarrow \quad A^+ \underset{\boxed{\text{only if } (A^T A)^{-1} \text{ exists}}}{=} (A^T A)^{-1} A^T$$

- the above is a new derivation of a formula for A^+ for a special case, <u>yielding new insight</u>.

BUT: forming $A^T A$ undesirable numerically

$\Rightarrow$ formula of § 7 for A^+ is better as it works even if $A^T A \neq$ invertible and avoids calculation of $A^T A$.