Pseudoinverse & Orthogonal Projection Operators

ECE 174 – Linear & Nonlinear Optimization

Ken Kreutz-Delgado
ECE Department, UC San Diego

The Four Fundamental Subspaces of a Linear Operator

For a linear operator $A : \mathcal{X} \to \mathcal{Y}$,

$$\mathcal{X} = \mathcal{N}(A)^{\perp} \oplus \mathcal{N}(A) \quad \text{and} \quad \mathcal{Y} = \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$$

Defining the adjoint $A^* : \mathcal{Y} \to \mathcal{X}$ by

$$\langle y, Ax \rangle = \langle A^* y, x \rangle$$

we obtain (as was shown last lecture)

$$\boxed{\boldsymbol{\mathcal{R}(A^*) = \mathcal{N}(A)^{\perp} \quad \text{and} \quad \mathcal{N}(A^*) = \mathcal{R}(A)^{\perp}}}$$

yielding

$$\boxed{\boldsymbol{\mathcal{X} = \mathcal{R}(A^*) \oplus \mathcal{N}(A) \quad \text{and} \quad \mathcal{Y} = \mathcal{R}(A) \oplus \mathcal{N}(A^*)}}$$

- The four fundamental subspaces of A are the orthogonally complementary subspaces $\mathcal{R}(A)$ and $\mathcal{N}(A^*)$ and the orthogonally complementary subspaces, $\mathcal{N}(A)$ and $\mathcal{R}(A^*)$.

Because

$$\langle y, Ax\rangle = \langle A^* y, x\rangle \;\Leftrightarrow\; \overline{\langle Ax, y\rangle} = \overline{\langle x, A^* y\rangle} \;\Leftrightarrow\; \langle x, A^* y\rangle = \langle Ax, y\rangle$$

the adjoint of A^* is (merely by definition!)

$$\boxed{\boldsymbol{A^{**} \triangleq (A^*)^* = A}}$$

Thus the relationships between A and A^* and their four fundamental subspaces are <u>entirely symmetrical</u>:

$$\begin{array}{rcl} \boldsymbol{A : \mathcal{X} \to \mathcal{Y}} & \qquad & \boldsymbol{A^* : \mathcal{Y} \to \mathcal{X}} \\ \boldsymbol{\langle y, Ax\rangle = \langle A^* y, x\rangle} & & \boldsymbol{\langle x, A^* y\rangle = \langle Ax, y\rangle} \\ \boldsymbol{\mathcal{Y} = \mathcal{R}(A) \oplus \mathcal{N}(A^*)} & & \boldsymbol{\mathcal{X} = \mathcal{R}(A^*) \oplus \mathcal{N}(A)} \\ \boldsymbol{\mathcal{R}(A) = \mathcal{N}(A^*)^\perp} & & \boldsymbol{\mathcal{R}(A^*) = \mathcal{N}(A)^\perp} \end{array}$$

The Pseudoinverse Solution – Cont.

Because of this symmetry, sometimes we can get twice the number of results per derivation. For example, here we prove that

$$\boxed{\mathcal{N}(A) = \mathcal{N}(A^*A) \quad \textbf{and} \quad \mathcal{R}(A^*) = \mathcal{R}(A^*A)}$$

from which it directly follows that

$$\boxed{\mathcal{N}(A^*) = \mathcal{N}(AA^*) \quad \textbf{and} \quad \mathcal{R}(A) = \mathcal{R}(AA^*)}$$

Proofs of the first statements:

- If $x \in \mathcal{N}(A)$, then obviously $x \in \mathcal{N}(A^*A)$. On the other hand

$$\begin{aligned} x \in \mathcal{N}(A^*A) \quad &\iff \quad A^*Ax = 0 \\ &\iff \quad \langle \xi, A^*Ax \rangle = 0, \quad \forall \xi \\ &\iff \quad \langle A\xi, Ax \rangle = 0, \quad \forall \xi \\ &\implies \quad \langle Ax, Ax \rangle = \|Ax\|^2 = 0 \\ &\implies \quad x \in \mathcal{N}(A) \end{aligned}$$

Four Fundamental Subspaces – Cont.

Proofs continued:

- Having established that $\mathcal{N}(A) = \mathcal{N}(A^*A)$, we have

$$\begin{aligned} x \in \mathcal{R}(A^*A) \quad &\iff \quad x \perp \mathcal{N}(A^*A) \qquad \text{(using the fact that } (A^*A)^* = A^*A) \\ &\iff \quad x \perp \mathcal{N}(A) \\ &\iff \quad x \in \mathcal{R}(A^*) \end{aligned}$$

Note that we have also established that

$$\boxed{\boldsymbol{\nu(A) = \nu(A^*A), \quad r(A) = r(AA^*), \quad \nu(A^*) = \nu(AA^*), \quad \textbf{and} \quad r(A^*) = r(A^*A)}}$$

Furthermore, with A a mapping between two finite dimensional spaces one can show

$$\boxed{\boldsymbol{r(A^*A) = r(AA^*) = r(A) = r(A^*)}}$$

Note that $\boldsymbol{\dim \mathcal{R}(A) = \dim \mathcal{R}(A^*)}$.

Adjoint-Based Conditions for a P-Inv Solution

- Having defined the adjoint we obtain the geometric conditions for a pseudoinverse solution,

$$\boxed{\textbf{1. Geometric Cond. for a Least-Squares Solution:} \quad e = y - Ax' \in \mathcal{N}(A^*)}$$

$$\boxed{\textbf{2. Geometric Cond. for a Minimum Norm LS Solution:} \quad x' \in \mathcal{R}(A^*)}$$

- The geometric conditions conditions easily lead to the algebraic conditions

$$\boxed{\textbf{1. Algebraic Cond. for an LS Solution – The Normal Equation:} \quad A^*Ax = A^*y}$$

$$\boxed{\textbf{2. Algebraic Cond. for a Minimum Norm LS Solution:} \quad x = A^*\lambda}$$

- When the domain has a metric matrix Ω and the codomain has metric matrix W, then (assuming the standard canonical-basis representations of vectors and linear operators) the adjoint operator is

$$\boxed{A^* = \Omega^{-1}A^H W}$$

Solving for the P-Inv Solution – I: A One-to-One

- If a linear mapping A between finite dimensional spaces is either onto or one-to-one, we say that A is **full-rank.** Otherwise A is **rank deficient.**
  - If A is a matrix which is onto, we say that it is **full row rank.**
  - If A is a matrix which is one-to-one we say that it is **full column rank.**
- If A is one-to-one, then the least-squares solution to the inverse problem $y = Ax$ is unique. Thus the second algebraic condition, which serves to resolve non-uniqueness when it exists, is not needed.
  - A one-to-one yields A^*A one-to-one and onto, and hence invertible. Thus from the first algebraic condition (the normal equation), we have

$$A^*Ax = A^*y \Rightarrow \hat{x} = (A^*A)^{-1}A^*y = A^+y$$

    showing that the pseudoinverse operator that maps measurement y to the least-squares solution $\hat{x}$ is given by

$$\boxed{\boldsymbol{A \text{ one-to-one} \Rightarrow A^+ = (A^*A)^{-1}A^*}}$$

  - Directly solving the normal equation $A^*Ax = A^*y$ is a numerical superior way to obtain $\hat{x}$. The expressions $A^+ = (A^*A)^{-1}A^*$ and $\hat{x} = A^+y$ are usually preferred for mathematical analysis purposes.

Solving for the P-Inv Solution – II: A Onto

- If A is onto (has a full row rank matrix representation), then there is always a solution to the inverse problem $y = Ax$. Thus the first algebraic condition (the normal equation), which serves to obtain an approximate solution and stands in for $y = Ax$ when it is inconsistent, is not needed for our analysis purposes. (It may have numerical utility however).
  - A onto yields A^*A onto and one-to-one, and hence invertible. Thus from the second algebraic condition and the (consistent) equation $y = Ax$ we have

    $$x = A^*\lambda \Rightarrow y = Ax = AA^*\lambda \Rightarrow \lambda = (AA^*)^{-1}y \Rightarrow x = A^*(AA^*)^{-1}y = A^+y$$

    showing that the pseudoinverse operator that maps measurement y to the least-squares solution $\hat{x}$ is given by

    $$\boxed{A \text{ onto} \Rightarrow A^+ = A^*(AA^*)^{-1}}$$

  - Directly solving the equation $AA^*\lambda = y$ for λ and then computing $\hat{x} = A^*\lambda$ is a numerical superior way to obtain $\hat{x}$. The expressions $A^+ = A^*(AA^*)^{-1}$ and $\hat{x} = A^+y$ are usually preferred for mathematical analysis purposes.
- What if A is neither one-to-one nor onto? How to compute the p-inv then?

Orthogonal Projection Operators

Suppose that $P = P^2$ is an orthogonal projection operator onto a subspace $\mathcal{V}$ along its orthogonal complement $\mathcal{V}^\perp$. Then $I - P$ is the orthogonal projection operator onto $\mathcal{V}^\perp$ along $\mathcal{V}$. For all vectors x_1 and x_2, we have

$$\langle Px_1, (I-P)x_2 \rangle = 0 \Leftrightarrow \langle (I-P)^* Px_1, x_2 \rangle = 0 \Leftrightarrow (I-P)^* P = 0 \Leftrightarrow P = P^* P$$

which yields the property

$$\boxed{\textbf{Orthogonal Projection Operators are Self-Adjoint:} \quad \boldsymbol{P = P^*}}$$

Thus, if $P = P^2$, P is a projection operator. If in addition $P = P^*$, then P is an orthogonal projection operator.

- $A^+ A : \mathcal{X} \to \mathcal{X}$ and $AA^+ : \mathcal{Y} \to \mathcal{Y}$ are both orthogonal projection operators. The first onto $\mathcal{R}(A^*) \subset \mathcal{X}$, the second onto $\mathcal{R}(A) \subset \mathcal{Y}$.

$\mathcal{R}(A)$ and $\mathcal{R}(A^*)$ are Linearly Isomorphic

Consider the linear mapping $A : \mathcal{X} \to \mathcal{Y}$ restricted to be a mapping from $\mathcal{R}(A^*)$ to $\mathcal{R}(A)$, $A : \mathcal{R}(A^*) \to \mathcal{R}(A)$.

The restricted mapping $A : \mathcal{R}(A^*) \to \mathcal{R}(A)$ is onto.

For all $\hat{y} \in \mathcal{R}(A)$ there exists $\hat{x} = A^+ y \in \mathcal{R}(A^*)$ such that $A\hat{x} = \hat{y}$.

The restricted mapping $A : \mathcal{R}(A^*) \to \mathcal{R}(A)$ is one-to-one.

Let $\hat{x} \in \mathcal{R}(A^*)$ and $\hat{x}' \in \mathcal{R}(A^*)$ both map to $\hat{y} \in \mathcal{R}(A)$, $\hat{y} = A\hat{x} = A\hat{x}'$.

Note that $\hat{x} - \hat{x}' \in \mathcal{R}(A^*)$ while at the same time $A(\hat{x} - \hat{x}') = 0$.

Therefore $\hat{x} - \hat{x}' \in \mathcal{R}(A^*) \cap \mathcal{N}(A)$, yielding $\hat{x} - \hat{x}' = 0$. Thus $\hat{x} = \hat{x}'$.

Since all of the elements of $\mathcal{R}(A^*)$ and $\mathcal{R}(A)$ are in one-to-one correspondence, these subspaces must be isomorphic as sets (and therefore have the same cardinality).

$\mathcal{R}(A)$ and $\mathcal{R}(A^*)$ are Linearly Isomorphic – Cont.

The restricted mapping $A : \mathcal{R}(A^*) \to \mathcal{R}(A)$ is a linear isomorphism.

Note that the restricted mapping $A : \mathcal{R}(A^*) \to \mathcal{R}(A)$ is linear, and therefore it preserves linear combinations in the sense that

$$A(\alpha_1 \hat{x}_1 + \cdots + \alpha_\ell \hat{x}_\ell) = \alpha_1 A\hat{x}_1 + \cdots + \alpha_\ell A\hat{x}_\ell \in \mathcal{R}(A)$$

Furthermore it can be shown that A isomorphically maps bases in $\mathcal{R}(A^*)$ to bases in $\mathcal{R}(A)$. Thus the dimension (number of basis vectors in) $\mathcal{R}(A^*)$ must be the same as the dimension (number of basis vectors) in $\mathcal{R}(A)$. Since the restricted mapping A is an isomorphism that preserves the vector space properties of its domain, span, linear independence, and dimension, we say that it is a **linear isomorphism.**

Summarizing:

$\mathcal{R}(A^*)$ and $\mathcal{R}(A)$ are Linearly Isomorphic,

$$\boxed{\mathcal{R}(A^*) \cong \mathcal{R}(A)}$$

$$\boxed{r(A^*) = \dim(\mathcal{R}(A^*)) = \dim(\mathcal{R}(A)) = r(A)}$$

$\mathcal{R}(A)$ and $\mathcal{R}(A^*)$ are Linearly Isomorphic – Cont.

The relationship between $\hat{y} = A\hat{x} \in \mathcal{R}(A)$ and $\hat{x} = A^+\hat{y} \in \mathcal{R}(A^*)$

$$\hat{x} \underset{A^+}{\overset{A}{\rightleftarrows}} \hat{y}$$

is one-to-one in both mapping directions. I.e., every $\hat{x} = A\hat{y} \in \mathcal{R}(A^*)$ maps to the unique element $\hat{y} = A\hat{x} \in \mathcal{R}(A)$, and *vice versa.*

Therefore when A and A^+ are **restricted** to be mappings between the subspaces $\mathcal{R}(A^*)$ and $\mathcal{R}(A)$,

$$A : \mathcal{R}(A^*) \to \mathcal{R}(A) \quad \text{and} \quad A^+ : \mathcal{R}(A) \to \mathcal{R}(A^*)\,,$$

then **they are inverses of each other:**

$$\boxed{A^+\big|_{\mathcal{R}(A)} = \left(A\big|_{\mathcal{R}(A^*)}\right)^{-1}}$$

Pseudoinverse & Orthogonal Projections

Let $A : \mathcal{X} \to \mathcal{Y}$, with $\dim(\mathcal{X}) = n$ and $\dim(\mathcal{Y}) = m$.

For any $y \in \mathcal{Y}$, compute

$$\hat{x} = A^{+} y$$

We have

$$\begin{aligned} \hat{y} &= P_{\mathcal{R}(A)} y && (\hat{y} \text{ is the least-squares estimate of } y) \\ &= A\,\hat{x} && (\hat{x} \text{ is a least-squares solution }) \\ &= AA^{+} y \end{aligned}$$

or

$$P_{\mathcal{R}(A)} y = AA^{+} y, \quad \forall y \in \mathcal{Y}$$

Therefore

$$\boxed{P_{\mathcal{R}(A)} = AA^{+} \quad \text{and} \quad P_{\mathcal{R}^{\perp}(A)} = I - AA^{+}}$$

Pseudoinverse & Orthogonal Projections – Cont.

For any $x \in \mathcal{X}$, compute

$$\hat{y} = Ax \qquad \text{(note that } \hat{y} \in \mathcal{R}(A) \text{ by construction)}$$

Then

$$\hat{y} = P_{\mathcal{R}(A)}\hat{y} = Ax$$

Now let

$$\hat{x} = A^{+}\hat{y} = A^{+}Ax \in \mathcal{R}(A^{*})\,.$$

Then, since $\hat{y} = A\hat{x}$,

$$\begin{aligned}
0 &= A(x - \hat{x}) \\
&\Rightarrow x - \hat{x} \in \mathcal{N}(A) = \mathcal{R}(A^{*})^{\perp} \\
&\Rightarrow x - \hat{x} \perp \mathcal{R}(A^{*}) \\
&\Rightarrow \hat{x} = P_{\mathcal{R}(A^{*})}x \qquad \text{(by orthogonality principle \& } \hat{x} \in \mathcal{R}(A^{*}) \text{)} \\
&\Rightarrow A^{+}Ax = P_{\mathcal{R}(A^{*})}x
\end{aligned}$$

Since this is true for all $x \in \mathcal{X}$, we have

$$\boxed{P_{\mathcal{R}(A^{*})} = A^{+}A \quad \text{and} \quad P_{\mathcal{R}^{\perp}(A^{*})} = I - A^{+}A}$$

Properties of AA^+ and A^+A

Having shown that $AA^+ = P_{\mathcal{R}(A)}$ and $A^+A = P_{\mathcal{R}(A^*)}$, we now know that AA^+ and A^+A must satisfy the properties of being orthogonal projection operators.

In particular AA^+ and A^+A must be self-adjoint,

$$\textbf{I.} \quad (AA^+)^* = AA^+ \qquad \textbf{II.} \quad (A^+A)^* = A^+A$$

These are the first two of the four **Moore-Penrose (M-P) Pseudoinverse Conditions.**

AA^+ and A^+A must also be idempotent, yielding

$$AA^+AA^+ = AA^+ \quad \text{and} \quad A^+AA^+A = A^+A$$

Note that both of these conditions are consequences of either of the remaining two M-P conditions,

$$\textbf{III.} \quad AA^+A = A \qquad \textbf{IV.} \quad A^+AA^+ = A^+$$

Four M-P P-Inv Conditions

M-P THEOREM: (Proved below.)

Consider a linear operator $A : \mathcal{X} \to \mathcal{Y}$. A linear operator $M : \mathcal{Y} \to \mathcal{X}$ is the unique pseudoinverse of A, $M = A^+$, if and only if it satisfies the

Four M-P Conditions:

I. $(AM)^* = AM$ **II.** $(MA)^* = MA$ **III.** $AMA = A$ **IV.** $MAM = M$

■

Thus one can test any possible candidate p-inv using the M-P conditions.

Example 1: Pseudoinverse of a scalar α

$$\alpha^+ = \begin{cases} \frac{1}{\alpha} & \alpha \neq 0 \\ 0 & \alpha = 0 \end{cases}$$

Example 2: For *general* linear operators A, B, and C for which the composite mapping ABC is well-defined we have

$$(ABC)^+ \neq C^+B^+A^+$$

because $C^+B^+A^+$ *in general* does **not** satisfy the M-P conditions to be a p-inv of ABC.

Example 3: Specializing Example 2, now assume that A and C are both *unitary:* $A^* = A^{-1}$, $C^* = C^{-1}$. (If we further assume that our metric is Cartesian, then For complex spaces this means that $A^{-1} = A^* = A^H$, whereas for real spaces this means that A and B are orthogonal, $A^{-1} = A^* = A^T$.)

Assuming that A and C are unitary, we have

$$(ABC)^+ = C^* B^+ A^*$$

as can be verified by showing that $C^* B^+ A^*$ satisfies the M-P conditions to be a p-inv for ABC.

Note that with the *additional unitary assumptions* on A and C. we now *can* claim that $(ABC)^+ = C^+ B^+ A^+$ since $C^* B^+ A^* = C^{-1} B^+ A^{-1} = C^+ B^+ A^+$.

Example 4: Suppose that Σ is a block matrix $m \times n$ matrix with block entries

$$\Sigma = \begin{pmatrix} S & 0_1 \\ 0_2 & 0 \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}$$

where S is square and the remaining block matrices have only entries with value 0. Then

$$\Sigma^+ = \begin{pmatrix} S^+ & 0_2^T \\ 0_1^T & 0^T \end{pmatrix} = \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix}$$

Four M-P P-Inv Conditions – Cont.

Example 5: Let A be a complex $m \times n$ matrix mapping between $\mathcal{X} = \mathbb{C}^n$ and $\mathcal{Y} = \mathbb{C}^m$ where both spaces have the standard Cartesian inner product. We shall see that A can be factored as

$$A = U\Sigma V^H = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} = U_1 S V_1^H$$

This factorization is known as the **Singular Value Decomposition (SVD).** The matrices have the following dimensions: U is $m \times m$, Σ is $m \times n$, V is $n \times n$.

The matrix S is a *real* diagonal matrix of dimension $r \times r$ where $r = r(A)$. Its diagonal entries, σ_i, $i = 1, \cdots, r$, are all real, greater than zero, and are called the **singular values** of A. The singular values are usually ordered in descending value

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{r-1} \geq \sigma_r > 0$$

Note that

$$S^+ = S^{-1} = \operatorname{diag}^{-1}(\sigma_1 \ \cdots \ \sigma_r) = \operatorname{diag}\left(\frac{1}{\sigma_1} \ \cdots \ \frac{1}{\sigma_r}\right)$$

Example 5 – Cont.

The matrix U is unitary, $U^{-1} = U^H$, and its columns form an orthonormal basis for the codomain $\mathcal{Y} = \mathbb{C}^m$ wrt to the standard inner product. (Its rows are also orthonormal.) If we denote the columns of U, known as the **left singular vectors**, by u_i, $i = 1, \cdots, m$, the first r lsv's comprise the columns of the $m \times r$ matrix U_1, while the remaining lsv's comprise the columns of the $m \times \mu$ matrix U_2, where $\mu = m - r$ is the dimension of the nullspace of A^*. The lsv u_i is in one-to-one correspondence with the singular value σ_i for $i = 1, \cdots, r$.

The matrix V is unitary, $V^{-1} = V^H$, and its columns form an orthonormal basis for the domain $\mathcal{X} = \mathbb{C}^n$ wrt to the standard inner product. (Its rows are also orthonormal.) If we denote the columns of V, known as the **right singular vectors**, by v_i, $i = 1, \cdots, m$, the first r rsv's comprise the columns of the $n \times r$ matrix V_1, while the remaining rsv's comprise the columns of the $n \times \nu$ matrix V_2, where $\nu = n - r$ is the nullity (dimension of the nullspace) of A. The rsv v_i is in one-to-one correspondence with the singular value σ_i for $i = 1, \cdots, r$.

We have

$$A = U_1 S V_1^H = \begin{pmatrix} u_1 & \cdots & u_r \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} v_1^H \\ \vdots \\ v_r^H \end{pmatrix} = \sigma_1 u_1 v_1^H + \cdots + \sigma_r u_r v_r^H$$

Example 5 – Cont.

Using the results derived in Examples 3 and 4, we have

$$\begin{aligned}
A^+ &= V\Sigma^+ U^H \qquad \text{(from example 3)} \\
&= \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^H & U_2^H \end{pmatrix} \qquad \text{(from example 4)} \\
&= V_1 S^{-1} U_1^H \qquad \text{(from invertibility of } S\text{)} \\
&= \begin{pmatrix} v_1 & \cdots & v_r \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_r} \end{pmatrix} \begin{pmatrix} u_1^H \\ \vdots \\ u_r^H \end{pmatrix} \\
&= \frac{1}{\sigma_1} v_1 u_1^H + \cdots + \frac{1}{\sigma_r} v_r u_r^H
\end{aligned}$$

Note that this construction works regardless of the value of the rank r. This shows that knowledge of the SVD of a matrix A allows us to determine its p-inv even in the rank-deficient case.

Four Moore-Penrose Pseudoinverse Conditions

MOORE-PENROSE THEOREM

Consider a linear operator $A : \mathcal{X} \to \mathcal{Y}$.

A linear operator $M : \mathcal{Y} \to \mathcal{X}$ is the unique pseudoinverse of A, $M = A^+$, if and only if it satisfies the

Four Moore-Penrose Conditions:

I. $\boldsymbol{(AM)^* = AM}$ **II.** $\boldsymbol{(MA)^* = MA}$ **III.** $\boldsymbol{AMA = A}$ **IV.** $\boldsymbol{MAM = M}$

■

More simply we usually say that A^+ is the unique p-inv of A iff

I. $(AA^+)^* = AA^+$ II. $(A^+A)^* = A^+A$ III. $AA^+A = A$ IV. $A^+AA^+ = A^+$

- The theorem statement provides greater clarity because there we distinguish between a *candidate* p-inv M and the *true* p-inv A^+.
- If and only if the candidate p-inv M satisfies the four M-P conditions can we claim that indeed $A^+ = M$.

Proof of the M-P Theorem

First we reprise some basic facts that are consequences of the definitional properties of the pseudoinverse.

FACT 1: $\mathcal{N}(A^+) = \mathcal{N}(A^*)$

FACT 2: $\mathcal{R}(A^+) = \mathcal{R}(A^*)$

FACT 3: $P_{\mathcal{R}(A)} = A\,A^+$

FACT 4: $P_{\mathcal{R}(A^*)} = A^+A$

We now proceed to prove two auxiliary theorems (Theorems A and B).

THEOREM A

Let $C : \mathcal{X} \to \mathcal{Y}$ and $B : \mathcal{Y} \to \mathcal{Z}$ be linear mappings. It is readily shown that the composite mapping $BC : \mathcal{X} \to \mathcal{Z}$ is a linear mapping where BC is defined by

$$(BC)x \triangleq B(Cx) \quad \forall x \in \mathcal{X}$$

Then

$$\boldsymbol{\mathcal{N}(B) \cap \mathcal{R}(C) = \{0\} \quad \Rightarrow \quad BC = 0 \ \text{ iff } \ C = 0}$$

■

Proof

$$\begin{aligned}
BC = 0 &\Leftrightarrow (BC)x = 0 \quad \forall x && \text{(definition of zero operator)} \\
&\Leftrightarrow B(Cx) = 0 \quad \forall x && \text{(definition of composition)} \\
&\Leftrightarrow Cx = 0 \quad \forall x && \text{(because } Cx \in \mathcal{R}(C) \cap \mathcal{N}(B) = \{0\}, \ \ \forall x) \\
&\Leftrightarrow C = 0 && \text{(definition of zero operator)}
\end{aligned}$$

QED

Theorem B covers the uniqueness part of the M-P Theorem.

THEOREM B. The pseudoinverse of A is unique. ■

Proof. Suppose that A^+ and M are *both* p-inv's of A. Then Fact 3 gives $P_{\mathcal{R}(A)} = AA^+ = AM$ or

$$A(A^+ - M) = 0$$

From Fact 2, $\mathcal{R}(A^*) = \mathcal{R}(A^+) = \mathcal{R}(M)$ and as a consequence

$$\mathcal{R}(A^+ - M) \subset \mathcal{R}(A^*)$$

But $\mathcal{R}(A^*) \perp \mathcal{N}(A)$ and therefore

$$\mathcal{R}(A^+ - M) \subset \mathcal{R}(A^*) = \mathcal{N}(A)^\perp$$

so that

$$\mathcal{N}(A) \cap \mathcal{R}(A^+ - M) = \{0\}$$

Therefore from Theorem A,

$$A^+ - M = 0 \Rightarrow A^+ = M$$

QED

Proof of the M-P Theorem – Cont.

Necessity ('only if' part) of the M-P Conditions. Assume that $M = A^+$.

Necessity of M-P Conditions I & II. Easy consequences of Facts 3 and 4.

Necessity of M-P Condition III. Note that Fact 4 and indempotency of a projection operator implies $(A^+A)(A^+A) = A^+A$, or

$$A^+ \underbrace{(AA^+A - A)}_{\triangleq C = A(A^+A - I)} = AC = 0$$

We have $\mathcal{N}(A^+) = \mathcal{N}(A^*)$ (Fact 1) and $\mathcal{R}(C) \subset \mathcal{R}(A) = \mathcal{N}(A^*)^\perp$. Therefore $\mathcal{N}(A^+) \cap \mathcal{R}(C) = \mathcal{N}(A^*) \cap \mathcal{R}(C) = \{0\}$ so that by Theorem A, $C = 0$.

Necessity of M-P Condition IV. Note that Fact 3 and indempotency of a projection operator implies $(AA^+)(AA^+) = AA^+$, or

$$A \underbrace{(A^+AA^+ - A^+)}_{\triangleq C = A^+(AA^+ - I)} = AC = 0$$

With $\mathcal{R}(A^+) = \mathcal{R}(A^*)$ (Fact 2) we have $\mathcal{R}(C) \subset \mathcal{R}(A^+) = \mathcal{R}(A^*) = \mathcal{N}(A)^\perp$. Therefore $\mathcal{N}(A) \cap \mathcal{R}(C) = \{0\}$ so that by Theorem A, $C = 0$.

Proof of the M-P Theorem – Cont.

Sufficiency('if' part) of the M-P Conditions.

Here we assume that M satisfies all four of the M-P conditions and then show as a consequence that $M = A^+$.

We do this using the following steps.

(1) First prove that $P_{\mathcal{R}(A)} = AM$ (proving that $AM = AA^+$ via uniqueness of projection operators).

(2) Prove that $P_{\mathcal{R}(A^*)} = MA$ (proving that $MA = A^+A$).

(3) Finally, prove that as a consequence of (1) and (2), $M = A^+$.

Proof of the M-P Theorem – Cont.

Sufficiency – Cont.

Step (1):

From M-P conditions 1 & 3, $(AM)^* = AM$ and $AM = (AMA)M = (AM)(AM)$, showing that AM is an orthogonal projection operator. But onto what? Obviously onto a subspace of $\mathcal{R}(A)$ as $\mathcal{R}(AM) \subset \mathcal{R}(A)$. However

$$\mathcal{R}(A) = A(\mathcal{X}) = AMA(\mathcal{X}) = AM\,(A(\mathcal{X})) = AM\,(\mathcal{R}(A)) \subset AM(\mathcal{Y}) = \mathcal{R}(AM) \subset \mathcal{R}(A)$$

yields the stronger statement that $\mathcal{R}(AM) = \mathcal{R}(A)$. Thus AM is the orthogonal projector onto the range of A, $P_{\mathcal{R}(A)} = AM = AA^+$.

Step (2):

From M-P conditions 2 & 3, $(MA)^* = MA$ and $MA = M(AMA) = (MA)(MA)$, showing that MA is an orthogonal projection operator. Note that M-P conditions 3 and 2 imply $A^* = (AMA)^* = A^*M^*A^*$ and $MA = (MA)^* = A^*M^*$. We have

$$\mathcal{R}(A^*) = A^*(\mathcal{Y}) = (A^*M^*A^*)(\mathcal{Y}) = (A^*M^*)(\mathcal{R}(A^*)) \subset \underbrace{(A^*M^*)(\mathcal{X})}_{=\mathcal{R}(A^*M^*)=\mathcal{R}(MA)} \subset \mathcal{R}(A^*)$$

showing that $\mathcal{R}(MA) = \mathcal{R}(A^*)$. Thus $P_{\mathcal{R}(A^*)} = MA = A^+A$.

Proof of the M-P Theorem – Cont.

Sufficiency – Cont.

Step (3):

Note that we have yet to use M-P condition 4, $MAM = M$. From M-P condition 4 and the result of Step (2) we have

$$MAM = P_{\mathcal{R}(A^*)}M = M$$

Obviously, then $\mathcal{R}(M) \subset \mathcal{R}(A^*)$, as can be rigorously shown via the subspace chain

$$\mathcal{R}(M) = M(\mathcal{Y}) = P_{\mathcal{R}(A^*)}M(\mathcal{Y}) = P_{\mathcal{R}(A^*)}(\mathcal{R}(M)) \subset P_{\mathcal{R}(A^*)}(\mathcal{X}) = \mathcal{R}(A^*)$$

Recalling that $\mathcal{R}(A^+) = \mathcal{R}(A^*)$ (Fact 2), it therefore must be the case that

$$\mathcal{R}(M - A^+) \subset \mathcal{R}(A^*) = \mathcal{N}(A)^\perp$$

Using the result of Step (1), $P_{\mathcal{R}(A)} = AM = AA^+$, we have

$$A(M - A^+) = 0$$

with $\mathcal{N}(A) \cap \mathcal{R}(M - A^+) = \{0\}$. Therefore Theorem A yields $M - A^+ = 0$.

QED

Note the similarity of the latter developments in Step 3 to the proof of Theorem B. In fact, some thought should convince yourself that the latter part of Step 3 provides justification for the claim that the pseudoinverse is unique, so that Theorem B can be viewed as redundant to the proof of the M-P Theorem.

Theorem B was stated to introduce the student to the use of Theorem A (which played a key role in the proof of the M-P Theorem) and to present the uniqueness of the pseudoinverse as a key result in its own right.

Singular Value Decomposition (SVD)

Henceforth, let us consider only Cartesian Hilbert spaces (i.e., spaces with identity metric matrices) and consider all finite dimensional operators to be represented as complex $m \times n$ matrices,

$$\underset{m\times n}{A} : \mathcal{X} = \mathbb{C}^n \to \mathcal{Y} = \mathbb{C}^m$$

Note, in general, that the matrix A may be non-square and therefore not have a spectral representation (because eigenvalues and eigenvectors are then not defined).

Even if A is square, it will in general have complex valued eigenvalues and non-orthogonal eigenvectors. Even worse, a general $n \times n$ matrix can be *defective* and not have a full set of n eigenvectors, in which case A is not diagonalizable. In the latter case, one must one use generalized eigenvectors to understand the spectral properties of the matrix (which is equivalent to placing the matrix in *Jordan Canonical Form*).

It is well know that if a square, $n \times n$ complex matrix is self-adjoint (Hermitian), $A = A^H$, then its eigenvalues are all real and it has a full complement of n eigenvectors that can all be chosen to orthonormal. In this case for eigenpairs (λ_i, x_i), $i = 1, \cdots, n$, A has a simple spectral representation given by an orthogonal transformation,

$$A = \lambda_1 x_1 x_1^H + \cdots + \lambda_n x_n x_n^H = X \Lambda X^H$$

with $\Lambda = \text{diag}(\lambda_1 \; \cdots \; \lambda_n)$, and X is unitary, $X^H X = X X^H = I$, where the columns of X are comprised of the orthonormal eigenvectors x_i. If in addition, a hermitian matrix A is positive-semidefinite, denoted as $A \geq 0$, then the eigenvalues are all non-negative, and all strictly positive if the matrix A is invertible (positive-definite, $A > 0$).

Singular Value Decomposition (SVD) – Cont.

Given an arbitrary (nonsquare) complex matrix operator $A \in \mathbb{C}^{m \times n}$ we can 'regularized' its structural properties by 'squaring' it to produce a hermitian, positive-semidefinite matrix, and thereby exploit the very nice properties of hermitian, positive-semidefinite matrices mentioned above.

Because matrix multiplication is noncommutative, there are two ways to 'square' A to form a hermitian, positive-semidefinite matrix, viz

$$AA^H \qquad \text{and} \qquad A^H A$$

It is an easy exercise to proved that both of these forms are hermitian, positive-semidefinite, recalling that a matrix M is defined to be positive-semidefinite, $M \geq 0$, if and only if the associated quadratic form $\langle x, Mx \rangle = x^H Mx$ is real and positive-semidefinite

$$\langle x, Mx \rangle = x^H Mx \geq 0 \quad \forall x$$

Note that a sufficient condition for the quadratic form to be real is that M be hermitian, $M = M^H$. For the future, recall that a positive-semidefinite matrix M is positive-definite, $M > 0$, if in addition to the non-negativity property of the associated quadratic form we also have

$$\langle x, Mx \rangle = x^H Mx = 0 \quad \text{if and only if} \quad x = 0$$

Singular Value Decomposition (SVD) – Cont.

The eigenstructures of the well-behaved hermitian, positive-semidefinite 'squares' $A^H A$ and AA^H are captured in the Singular value Decomposition (SVD) introduced in Example 5 given earlier. As noted in that example, knowledge of the SVD enables one to compute the pseudoinverse of A in the rank deficient case.

The SVD also allow one to compute a variety of important quantities, including the rank of A, orthonormal bases for all four fundamental subspaces of A, orthogonal projection operators onto all four fundamental subspaces of the matrix operator A, the spectral norm of A, the Frobenius norm of A, and the condition number of A.

The SVD also provides a geometrically intuitive understanding of the nature of A as an operator based on the action of A as mapping hyperspheres in $\mathcal{R}(A^*)$ to hyperellipsoids in $\mathcal{R}(A)$.

Eigenstructure of $A^H A$

Let $A : \mathcal{X} = \mathbb{C}^n \to \mathcal{Y} = \mathbb{C}^m$ be an $m \times n$ matrix operator mapping between two Cartesian complex Hilbert spaces.

Recall that (with $A^H = A^*$ for A a mapping between Cartesian spaces)

$$r(AA^H) = r(A) = r(A^H) = r(A^H A)$$

Therefore the number of nonzero (and hence strictly positive) eigenvalues of $AA^H : \mathbb{C}^m \to \mathbb{C}^m$ and $A^H A : \mathbb{C}^n \to \mathbb{C}^n$ must both be equal to $r = r(A)$.

Let the nonnegative eigenvalues of $A^H A$ be denoted and ordered as

$$\sigma_1^2 \quad \geq \quad \cdots \quad \geq \quad \sigma_r^2 \quad > \quad \underbrace{\sigma_{r+1}^2 = \cdots = \sigma_n^2 = 0}_{\text{eigenvalue 0 has multiplicity } \nu = n - r}$$

with corresponding n-dimensional orthonormal eigenvectors

$$\underbrace{v_1 \quad \cdots \quad v_r}_{\text{span of } \mathcal{R}(A^H) = \mathcal{N}(A)^\perp} \quad \underbrace{v_{r+1} \quad \cdots \quad v_n}_{\text{spans } \mathcal{N}(A) = \mathcal{N}(A^H A)}$$

Eigenstructure of $A^H A$ – Cont.

Thus we have

$$(A^H A)v_i = \sigma_i^2 v_i \quad \text{with} \quad \sigma_i^2 > 0 \quad \text{for} \quad i = 1, \cdots, r$$

and

$$(A^H A)v_i = 0 \quad \text{for} \quad i = r+1, \cdots, n$$

The eigenvectors $v_{r+1} \cdots v_n$ can be chosen to be *any* orthonormal set spanning $\mathcal{N}(A)$.

An eigenvectors v_i associated with a distinct nonzero eigenvalues σ_i^2, $1 \le i \le r$, is unique up to sign $v_i \mapsto \pm v_i$.

Eigenvectors v_i associated with the same nondistinct nonzero eigenvalue σ_i^2 with multiplicity p can be chosen to be any orthonormal set that spans the p-dimensional eigenspace associated with that eigenvalue.

Thus we see that there is a lack of uniqueness in the eigen-decomposition of $A^H A$. This lack of uniqueness (as we shall see) will carry over to a related lack of uniqueness in the SVD.

What is unique are the values of the nonzero eigenvalues, the eigenspaces associated with those eigenvalues, and any projection operators we construct from the eigenvectors (uniqueness of projection operators).

Eigenstructure of $A^H A$ – Cont.

In particular, we uniquely have

$$P_{\mathcal{R}(A^H)} = V_1\, V_1^H \qquad \text{and} \qquad P_{\mathcal{N}(A)} = V_2\, V_2^H$$

where

$$V_1 \triangleq \begin{pmatrix} v_1 & \cdots & v_r \end{pmatrix} \in \mathbb{C}^{n\times r} \qquad V_2 \triangleq \begin{pmatrix} v_{r+1} & \cdots & v_n \end{pmatrix} \in \mathbb{C}^{n\times \nu}$$

and

$$V \triangleq \begin{pmatrix} V_1 & V_2 \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_r & v_{r+1} & \cdots & v_n \end{pmatrix} \in \mathbb{C}^{n\times n}$$

Note that

$$\mathcal{R}(V_1) = \mathcal{R}(A^H), \qquad \mathcal{R}(V_2) = \mathcal{N}(A), \qquad \mathcal{R}(V) = \mathcal{X} = \mathcal{C}^n$$

$$\underset{n\times n}{I} = V^H V = V V^H = V_1 V_1^H + V_2 V_2^H = P_{\mathcal{R}(A^H)} + P_{\mathcal{N}(A)}$$

$$\underset{r\times r}{I} = V_1^H V_1, \qquad \underset{\nu\times \nu}{I} = V_2^H V_2$$

It is straightforward to show that $V_1\, V_1^H$ and $V_2\, V_2^H$ are idempotent and self-adjoint.

We now prove two identities that will prove useful when deriving the SVD.

Taking $\sigma_i = \sqrt{\sigma_i^2}$ define

$$\underset{r\times r}{S} \triangleq \mathrm{diag}(\sigma_1 \ \cdots \ \sigma_r)$$

Then

$$A^H A v_i = \sigma_i^2 v_i \qquad 1 \le i \le r$$

can be written as

$$A^H A V_1 = V_1 S^2$$

which yields

$$\underset{r\times r}{\boldsymbol{I}} = \boldsymbol{S^{-1} V_1^H A^H A V_1 S^{-1}} \tag{1}$$

We also note that

$$A^H A v_i = 0 \Leftrightarrow A v_i \in \mathcal{R}(A) \cap \mathcal{N}(A^H) = \{0\}$$

so that $A^H A v_i = 0, \ i = r+1, \cdots, n$ yields

$$\underset{m\times\nu}{\boldsymbol{0}} = \boldsymbol{A V_2} \tag{2}$$

Eigenstructure of AA^H

The eigenstructure of $A^H A$ determined above places constraints on the eigenstructure of AA^H.

Above we have shown that

$$(A^H A)v_i = \sigma_i^2 v_i \quad i = 1, \cdots, r$$

where σ_i^2, $1 \le i \le r$, are nonzero. If we multiply both sides of this equation by A we get (recall that $r = r(AA^H) \le m$

$$(AA^H)(Av_i) = \sigma_i^2 (Av_i) \quad i = 1, \cdots, r$$

Showing that Av_i and σ_i^2 are eigenvector-eigenvalue pairs.

Since AA^H is hermitian, the vectors Av_i *must* be orthogonal. In fact, the vectors

$$u_i \triangleq \frac{1}{\sigma_i} Av_i \qquad 1 \le i \le r$$

are *orthonormal.*

Eigenstructure of AA^H – Cont.

This follows from defining

$$U_1 = \begin{pmatrix} u_1 & \cdots & u_r \end{pmatrix} \in \mathbb{C}^{m \times r}$$

which is equivalent to

$$U_1 = AV_1 S^{-1}$$

and noting that Equation (1) yields orthogonality of the columns of U_1

$$U_1^H U_1 = S^{-1} V_1^H A^H A V_1 S^{-1} = I$$

Note from the above that

$$\underset{r \times r}{\boldsymbol{S}} = \boldsymbol{U_1^H A V_1} \tag{3}$$

Also note that a determination of V_1 (based on a resolution of the ambiguities described above) completely specifies $U_1 = AV_1 S^{-1}$. Contrawise, it can be shown that a specification of U_1 provides a unique determination of V_1.

Eigenstructure of AA^H - Cont.

Because u_i correspond to the nonzero eigenvalues of AA^H they must span $\mathcal{R}(AA^H) = \mathcal{R}(A)$. Therefore

$$\mathcal{R}(U_1) = \mathcal{R}(A) \qquad \text{and} \qquad P_{\mathcal{R}(A)} = U_1 U_1^H$$

Complete the set u_i, $i = 1, \cdots, r$, to include a set of orthonormal vectors, u_i, $i = r+1, \cdots m$, orthogonal to $\mathcal{R}(U_1)$ (this can be done via random selection of new vectors in $\mathbb{C}^m$ followed b Gram-Schmidt orthonormalization.) Let

$$\underset{m\times\mu}{U_2} = \begin{pmatrix} u_{r+1} & \cdots & u_m \end{pmatrix}$$

with $\mu = m - r$.

By construction

$$\mathcal{R}(U_2) = \mathcal{R}(U_1)^\perp = \mathcal{R}(A)^\perp = \mathcal{N}(A^H)$$

and therefore

$$\underset{n\times\mu}{\mathbf{0}} = A^H U_2 \tag{4}$$

and

$$P_{\mathcal{N}(A^H)} = U_2 U_2^H$$

Eigenstructure of AA^H– Cont.

Setting

$$U = \begin{pmatrix} u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \end{pmatrix} = \begin{pmatrix} U_1 & U_2 \end{pmatrix}$$

we have

$$\underset{m \times m}{I} = U^H U = UU^H = U_1 U_1^H + U_2 U_2^H = P_{\mathcal{R}(A)} + P_{\mathcal{N}(A^H)}$$

$$\underset{m\times n}{\Sigma} \triangleq U^H A V = \begin{pmatrix} U_1^H \\ U_2^H \end{pmatrix} A \begin{pmatrix} V_1 & V_2 \end{pmatrix} = \begin{pmatrix} U_1^H A V_1 & U_1 A V_2 \\ U_2^H A V_1 & U_2^H A V_2 \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}$$

or

$$A = U \Sigma V^H$$

Note that

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} = U_1 S V_1^H$$

This yields the **Singular Value Decomposition (SVD)** factorization of A

$$\boxed{\textbf{SVD:} \quad \boldsymbol{A = U\Sigma V^H = U_1 S V_1^H}}$$

Note that when A is square and full rank, we have $U = U_1$, $V = V_1$, $\Sigma = S$, and

$$A = U S V^H$$

SVD Properties

The matrix U is unitary, $U^{-1} = U^H$, and its columns form an orthonormal basis for the codomain $\mathcal{Y} = \mathbb{C}^m$ wrt to the standard inner product. (Its rows are also orthonormal.) If we denote the columns of U, known as the **left singular vectors**, by u_i, $i = 1, \cdots, m$, the first r lsv's comprise the columns of the $m \times r$ matrix U_1, while the remaining lsv's comprise the columns of the $m \times \mu$ matrix U_2, where $\mu = m - r$ is the dimension of the nullspace of A^*. The lsv u_i is in one-to-one correspondence with the singular value σ_i for $i = 1, \cdots, r$.

The matrix V is unitary, $V^{-1} = U^H$, and its columns form an orthonormal basis for the domain $\mathcal{X} = \mathbb{C}^n$ wrt to the standard inner product. (Its rows are also orthonormal.) If we denote the columns of V, known as the **right singular vectors**, by v_i, $i = 1, \cdots, m$, the first r rsv's comprise the columns of the $n \times r$ matrix V_1, while the remaining rsv's comprise the columns of the $n \times \nu$ matrix V_2, where $\nu = n - r$ is the nullity (dimension of the nullspace) of A. The rsv v_i is in one-to-one correspondence with the singular value σ_i for $i = 1, \cdots, r$.

We have

$$A = U_1 S V_1^H = \begin{pmatrix} u_1 & \cdots & u_r \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} v_1^H \\ \vdots \\ v_r^H \end{pmatrix} = \sigma_1 u_1 v_1^H + \cdots + \sigma_r u_r v_r^H$$

Each term in this dyadic expansion is unique (i.e., does not depend on how the ambiguities mentioned above are resolved).

SVD Properties – Cont.

We can use the SVD to gain geometric intuition of the action of the matrix operator $A : \mathcal{X} = \mathbb{C}^n \to \mathcal{Y} = \mathbb{C}^m$ on the space $\mathcal{X} = \mathcal{R}(A^H) + \mathcal{N}(A)$.

The action of A on $\mathcal{N}(A) = \mathcal{R}(V_2)$ is trivial to understand from its action on the right singular vectors which form a basis for $\mathcal{N}(A)$,

$$Av_i = 0 \qquad i = r+1, \cdots, n$$

In class we discussed the geometric interpretation of the action of the operator A on $\mathcal{R}(A^H)$ based on the dyadic expansion

$$A = \sigma_1 u_1 v_1^H + \cdots + \sigma_r u_r v_r^H$$

as a mapping of a hypersphere in $\mathcal{R}(A^H)$ to an associated hyperellipsoid in $\mathcal{R}(A)$ induced by the basis vector mappings

$$v_i \xrightarrow{A} \sigma_i u_i \qquad i = 1, \cdots, r$$

SVD Properties – Cont.

When A is square and presumably full rank, $r = n$, this allows us to measure the numerical conditioning of A via the quantity (the **condition number** of A)

$$\text{cond}(A) = \frac{\sigma_1}{\sigma_n}$$

This measures the degree of 'flattening' (distortion) of the hypersphere induced by the mapping A. A perfectly conditioned matrix A has $\text{cond}(A) = 1$, and an infinitely ill-conditioned matrix has $\text{cond}(A) = +\infty$.

Using the fact that for square matrices $\det A = \det A^T$ and $\det AB = \det A \det B$, we note that

$$1 = \det I = \det UU^H = \det U \, \overline{\det U} = |\det U|^2$$

or

$$|\det U| = 1$$

and similarly for the unitary matrices U^H, V, and V^H. (Note BTW that this implies for a unitary matrix U, $\det U = e^{j\phi}$ for some $\phi \in \mathbb{R}$. When U is real and orthogonal, $U^{-1} = U^T$, this reduces to $\det U = \pm 1$.) Thus for a square matrix A.

$$|\det A| = \det USU^H = |\det U| \cdot |\det S| \cdot |\det U| = \det S = \sigma_1 \sigma_1 \cdots \sigma_n$$

SVD Properties – Cont.

Exploiting identities provable from the M-P Theorem (see Homework 3) we have

$$
\begin{aligned}
A^+ &= V\Sigma^+ U^H \\
&= \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^H & U_2^H \end{pmatrix} \\
&= V_1 S^{-1} U_1^H \\
&= \begin{pmatrix} v_1 & \cdots & v_r \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_r} \end{pmatrix} \begin{pmatrix} u_1^H \\ \vdots \\ u_r^H \end{pmatrix} \\
&= \frac{1}{\sigma_1} v_1 u_1^H + \cdots + \frac{1}{\sigma_r} v_r u_r^H
\end{aligned}
$$

Note that this construction works regardless of the value of the rank r. This shows that knowledge of the SVD of a matrix A allows us to determine its p-inv even in the rank-deficient case. Also note that the pseudoinverse is *unique*, regardless of the particular SVD variant (i.e., it does not depend on how the ambiguities mentioned above are resolved).

SVD Properties – Cont.

Note that having an SVD factorization of A at hand provides us with an orthonormal basis for $\mathcal{X} = \mathbb{C}^n$ (the columns of V), an orthonormal basis for $\mathcal{R}(A^H)$ (the columns of V_1), an orthonormal basis for $\mathcal{N}(A)$ (the columns of V_2), an orthonormal basis for $\mathcal{Y} = \mathbb{C}^m$ (the columns of U), an orthonormal basis for $\mathcal{R}(A)$ (the columns of U_1), and an orthonormal basis for $\mathcal{N}(A^H)$ (the columns of U_2).

Although the SVD factorization, the bases mentioned above, are not uniquely defined, it is the case that the orthogonal projectors constructed from the basis are unique (from uniqueness of projection operators). Thus *we can construct the unique orthogonal projection operators via*

$$P_{\mathcal{R}(A)} = U_1 U_1^H \qquad P_{\mathcal{N}(A^H)} = U_2 U_2^H \qquad P_{\mathcal{R}(A^H)} = V_1 V_1^H \qquad P_{\mathcal{N}(A)} = V_2 V_2^H$$

Obviously having access to the SVD is tremendously useful. With the background we have now covered, one can now can greatly appreciate the utility of the Matlab command svd(A) which returns the singular values, left singular vectors, and right singular vectors of A, from which one can construct all of the entities described above. (Note that the singular vectors returned by Matlab will not necessarily *all* agree with the ones you construct by other means because of the ambiguites mentioned above. However, *the singular values will be the same,* and the left and right singular vector associated with the same, *distinct* singular value should only differ from yours by a sign at most.) Another useful Matlab command is pinv(A) which returns the pseudoinverse of A regardless of the value of the rank of A.

Two Simple SVD Examples

In the 3rd homework assignment you are asked to produce the SVD for some simple matrices by hand and then construct the four projection operators for each matrix as well as the pseudoinverse. The problems in Homework 3 have been *very carefully designed* so that you do not have to perform eigendecompositions to obtain the SVD's. Rather, you can easily force the hand-crafted matrices into SVD factored form via a series of simple steps based on understanding the geometry underlying the SVD. Two examples of this (nongeneral) solution procedure are given here.

Example 1. $A = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. First note that $m = 2$, $n = 1$, $r = r(A) = 1$ (obviously), $\nu = n - r = 0$, and $\mu = m - r = 1$. This immediately tells us that $V = V_1 = v_1 = 1$. We have

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}}_{U_1} \cdot \underbrace{\sqrt{5}}_{S} \cdot \underbrace{1}_{V^H} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{5}} & \mathrm{X} \\ \frac{2}{\sqrt{5}} & \mathrm{X} \end{pmatrix}}_{U_1 \quad U_2} \cdot \underbrace{\begin{pmatrix} \sqrt{5} \\ 0 \end{pmatrix}}_{\Sigma} \cdot \underbrace{1}_{V^H} = \underbrace{\underbrace{\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}}_{U_1 \quad U_2}}_{U} \cdot \underbrace{\begin{pmatrix} \sqrt{5} \\ 0 \end{pmatrix}}_{\Sigma} \cdot \underbrace{1}_{V^H}$$

Note that we exploit the fact that we know the dimensions of the various matrices we have to compute. Here we first filled out Σ before determining the unknown values of $U_2 = u_2$, which was later done using the fact that $u_2 \perp u_1 = U_1$.

Two Simple SVD Examples – Cont.

Example 2. $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Note that $m = 2$, $n = 2$, $r = r(A) = 1$ (obviously), $\nu = n - r = 1$, and $\mu = m - r = 1$. Unlike the previous example, here we have a nontrivial nullspace.

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}}_{U_1} \cdot \underbrace{2}_{S} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{V_1^T} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & X \\ \frac{1}{\sqrt{2}} & X \end{pmatrix}}_{U_1 \quad U_2} \cdot \underbrace{\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}}_{\Sigma} \cdot \underbrace{\overbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ X & X \end{pmatrix}}^{V_1^T}}_{V_2^T}$$

Exploiting the facts that $U_1 = u_1 \perp u_2 = U_2$ and $V_1 = v_1 \perp v_2 = V_2$ we easily determine that

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \underbrace{\underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}}_{U_1 \quad U_2}}_{U} \cdot \underbrace{\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}}_{\Sigma} \cdot \underbrace{\underbrace{\overbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}}^{V_1^T}}_{V_2^T}}_{V^T}$$

Note the $\pm$ sign ambiguity in the choice of U_2 and V_2.