

# Normed & Inner Product Vector Spaces

ECE 275A – Statistical Parameter Estimation

Ken Kreutz-Delgado  
ECE Department, UC San Diego

# Normed Linear Vector Space

In a vector space **it is useful to have a meaningful measure of size, distance, and neighborhood.** The existence of a **norm** allows these concepts to be well-defined.

A norm  $\|\cdot\|$  on a vector space  $\mathcal{X}$  is a mapping from  $\mathcal{X}$  to the the nonnegative real numbers which obeys the following three properties:

- 1  $\|\cdot\|$  is **homogeneous**,  $\|\alpha x\| = |\alpha|\|x\|$  for all  $\alpha \in \mathcal{F}$  and  $x \in \mathcal{X}$ ,
- 2  $\|\cdot\|$  is **positive-definite**,  $\|x\| \geq 0$  for all  $x \in \mathcal{X}$  and  $\|x\| = 0$  iff  $x = 0$ , and
- 3  $\|\cdot\|$  satisfies the **triangle-inequality**,  $\|x + y\| \leq \|x\| + \|y\|$  for all  $x, y \in \mathcal{X}$ .

A norm provides a **measure of size** of a vector  $x$ ,  $\text{size}(x) = \|x\|$

A norm provides a **measure of distance** between two vectors,  $d(x, y) = \|x - y\|$

A norm provides a well-defined  $\epsilon$ -ball or  **$\epsilon$ -neighborhood of a vector  $x$ ,**

$$N_{\epsilon}(x) = \{y \mid \|y - x\| \leq \epsilon\} = \text{closed } \epsilon \text{ neighborhood}$$

$$\overset{\circ}{N}_{\epsilon}(x) = \{y \mid \|y - x\| < \epsilon\} = \text{open } \epsilon \text{ neighborhood}$$

# Normed Linear Vector Space – Cont.

**There are innumerable norms** that one can define on a given vector space.

Assuming a canonical representation  $x = (x[1], \dots, x[n])^T \in \mathcal{F}^n$ ,  $\mathcal{F} = \mathbb{C}$  or  $\mathbb{R}$ , for a vector  $x$ , the most commonly used norms are

**The 1-norm:** 
$$\|x\|_1 = \sum_{i=1}^n |x[i]|,$$

the **2-norm:** 
$$\|x\|_2 = \sqrt{\sum_{i=1}^n |x[i]|^2},$$

and the  **$\infty$ -norm, or sup-norm:** 
$$\|x\|_\infty = \max_i |x[i]|$$

These norms are all special cases of the family of **p-norms**

$$\|x\|_p = \left( \sum_{i=1}^n |x[i]|^p \right)^{\frac{1}{p}}$$

In this course we focuss on the **weighted 2-norm**,  $\|x\| = \sqrt{x^H \Omega x}$ , where the **weighting matrix, aka metric matrix**,  $\Omega$  is hermitian and positive-definite.

# Banach Space

- A **Banach Space** is a **complete** normed linear vector space.
- Completeness is a technical condition which is the requirement that every so-called **Cauchy convergent sequence** is a convergent sequence.
- This condition is necessary (but not sufficient) for iterative numerical algorithms to have well-behaved and testable convergence behavior.
- As this condition is automatically guaranteed to be satisfied for every **finite-dimensional** normed linear vector space, it is not discussed in courses on Linear Algebra.
- Suffice it to say that the finite dimensional spaces normed-vector spaces, or subspace, considered in this course are perforce Banach Spaces.

# Minimum Error Norm Soln to Linear Inverse Problem

- An important theme of this course is that one can learn unknown parameterized models by minimizing the discrepancy between model behavior and observed real-world behavior.
- If  $y$  is the observed behavior of the world, which is assumed (modeled) to behave as  $y \approx \hat{y} = Ax$  for known  $A$  and unknown parameters  $x$ , one can attempt to learn  $x$  by minimizing a **model behavior discrepancy measure**  $D(y, \hat{y})$  wrt  $x$ .
  - In this way we can rationally deal with an inconsistent inverse problem. Although no solution may exist, we try to find an approximate solution which is “good enough” by minimizing the discrepancy  $D(y, \hat{y})$  wrt  $x$ .
- Perhaps the simplest procedure is to work with a discrepancy measure,  $D(e)$ , that depends directly upon the **prediction error**  $e \triangleq y - \hat{y}$ .
- A logical choice of a discrepancy measure when  $e$  is a member of a normed vector space with norm  $\| \cdot \|$  is

$$D(e) = \|e\| = \|y - Ax\|$$

- Below, we will see how this procedure is facilitated when  $y$  belongs to a Hilbert space.

# Inner Product Space and Hilbert Space

Given a vector space  $\mathcal{X}$  over the field of scalars  $\mathcal{F} = \mathbb{C}$  or  $\mathbb{R}$ , an **inner product** is an  $\mathcal{F}$ -valued binary operator on  $\mathcal{X} \times \mathcal{X}$ ,

$$\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{F}; \quad \{x, y\} \mapsto \langle x, y \rangle \in \mathcal{F}, \quad \forall x, y \in \mathcal{X}.$$

The inner product has the following three properties:

- 1 **Linearity in the second argument**
- 2 **Real positive-definiteness of  $\langle x, x \rangle$**  for all  $x \in \mathcal{X}$ .

$$0 \leq \langle x, x \rangle \in \mathbb{R} \text{ for any vector } x, \text{ and } 0 = \langle x, x \rangle \text{ iff } x = 0.$$

- 3 **Conjugate-symmetry**,  $\langle x, y \rangle = \overline{\langle y, x \rangle}$  for all  $x, y \in \mathcal{X}$ .

Given an inner product, one can construct the associated **induced norm**,

$$\|x\| = \sqrt{\langle x, x \rangle},$$

as the right-hand side of the above can be shown to satisfy all the properties demanded of a norm. **It is this norm that is used in an inner product space.**

If the resulting normed vector space is a Banach space, one calls the inner product space a **Hilbert Space**. All finite-dimensional inner product spaces are Hilbert spaces.

# The Weighted Inner Product

- On a finite  $n$ -dimensional Hilbert space, a general inner product is given by the **weighted inner product**,

$$\langle x_1, x_2 \rangle = x_1^H \Omega x_2,$$

where the **Weighting or Metric Matrix**  $\Omega$  is hermitian and positive-definite.

- The corresponding induced norm is the weighted 2-norm mentioned above

$$\|x\| = \sqrt{x^H \Omega x}$$

- When the metric matrix takes the value  $\Omega = I$  we call the resulting inner product and induced norm the **standard or Cartesian inner-product** and the **standard or Cartesian 2-norm** respectively. The Cartesian inner-product on real vector spaces is what is discussed in most undergraduate courses on linear algebra.

# Orthogonality Between Vectors

- The existence of an inner product enables us to define and exploit the concepts of orthogonality and angle between vectors and vectors; vectors and subspaces; and subspaces and subspaces.
- Given an **arbitrary** (not necessarily Cartesian) inner product, we **define orthogonality** (with respect to that inner product) of two vectors  $x$  and  $y$ , which we denote as  $x \perp y$ , by

$$x \perp y \quad \text{iff} \quad \langle x, y \rangle = 0$$

- If  $x \perp y$ , then

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\ &= \langle x, x \rangle + 0 + 0 + \langle y, y \rangle = \|x\|^2 + \|y\|^2 \end{aligned}$$

yielding the (generalized) **Pythagorean Theorem**

$$x \perp y \implies \|x + y\|^2 = \|x\|^2 + \|y\|^2$$



# C-S Inequality and the Angle Between Two Vectors

- An important relationship that exists between an inner product  $\langle x, y \rangle$  and its corresponding induced norm  $\|x\| = \sqrt{\langle x, x \rangle}$  is given by the

## Cauchy–Schwarz (C-S) Inequality

$$|\langle x, y \rangle| \leq \|x\| \|y\| \quad \text{for all } x, y \in \mathcal{X}$$

with equality iff and only if  $y = \alpha x$  for some scalar  $\alpha$ .

- One can meaningfully define **the angle  $\theta$  between two vectors in a Hilbert space** by

$$\cos \theta \triangleq \frac{|\langle x, y \rangle|}{\|x\| \|y\|}$$

since as a consequence of the C-S inequality we must have

$$0 \leq \cos \theta \leq 1$$

# Subspace Orthogonality and Orthogonal Complements

- Two Hilbert subspaces are said to be **orthogonal subspaces**,  $\mathcal{V} \perp \mathcal{W}$  if and only if every vector in  $\mathcal{V}$  is orthogonal to every vector in  $\mathcal{W}$ .
  - If  $\mathcal{V} \perp \mathcal{W}$  it must be the case that  $\mathcal{V}$  and  $\mathcal{W}$  are disjoint  $\mathcal{W}$ ,  $\mathcal{V} \cap \mathcal{W} = \{0\}$ .
- Given a subspace  $\mathcal{V}$  of  $\mathcal{X}$ , one defines the **orthogonal complement**  $\mathcal{V}^\perp$  of  $\mathcal{V}$  to be the set  $\mathcal{V}^\perp$  of all vectors in  $\mathcal{X}$  which are perpendicular to  $\mathcal{V}$ .
  - The orthogonal complement (in the finite dimensional case assumed here) obeys the property  $\mathcal{V}^{\perp\perp} = \mathcal{V}$ .
  - The orthogonal complement  $\mathcal{V}^\perp$  is **unique and a subspace** in its own right for which
$$\mathcal{X} = \mathcal{V} \oplus \mathcal{V}^\perp .$$
  - Thus  $\mathcal{V}$  and  $\mathcal{V}^\perp$  are complementary subspaces.
  - Thus  $\mathcal{V}^\perp$  is *more* than a complementary subspace to  $\mathcal{V}$ ,  
 **$\mathcal{V}^\perp$  is the orthogonally complementary subspace to  $\mathcal{V}$ .**
  - Note that it must be the case that

$$\dim \mathcal{X} = \dim \mathcal{V} + \dim \mathcal{V}^\perp .$$

# Orthogonal Projectors

- In a Hilbert space the projection onto a subspace  $\mathcal{V}$  along its (unique) orthogonal complement  $\mathcal{V}^\perp$  is an **orthogonal projection operator**, denoted by

$$P_{\mathcal{V}} \triangleq P_{\mathcal{V}|\mathcal{V}^\perp}$$

- Note that for an orthogonal projection operator the complementary subspace does not have to be explicitly denoted.
- Furthermore if the subspace  $\mathcal{V}$  is understood to be the case, one usually denotes the orthogonal projection operator simply by

$$P \triangleq P_{\mathcal{V}}$$

- Of course, as is the case for all projection operators, an orthogonal projection operator is idempotent

$$P^2 = P$$

# Four Fundamental Subspaces of a Linear Operator

Consider a linear operator  $A : \mathcal{X} \rightarrow \mathcal{Y}$  between two finite-dimensional Hilbert spaces  $\mathcal{X}$  and  $\mathcal{Y}$ .

We must have that

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

- If  $\dim(\mathcal{X}) = n$  and  $\dim(\mathcal{Y}) = m$ , we must have

$$\dim(\mathcal{R}(A)) = r \quad \dim(\mathcal{R}(A)^\perp) = m - r \quad \dim(\mathcal{N}(A)) = \nu \quad \dim(\mathcal{N}(A)^\perp) = n - \nu$$

where  $r$  is the rank, and  $\nu$  the nullity, of  $A$ .

- The unique subspaces  $\mathcal{R}(A)$ ,  $\mathcal{R}(A)^\perp$ ,  $\mathcal{N}(A)$ , and  $\mathcal{N}(A)^\perp$  are called

## The Four Fundamental Subspaces of the linear operator $A$ .

- Understanding these four subspaces yields great insight into solving ill-posed linear inverse problems  $y = Ax$ .

# Projection Theorem & Orthogonality Principle

Given a vector  $x$  in a Hilbert space  $\mathcal{X}$ , what is the best approximation,  $v$ , to  $x$  in a subspace  $\mathcal{V}$  in the sense that the norm of the error  $D(e) = e = x - v$ ,  $\|e\| = \|x - v\|$ , is to be minimized over all possible vectors  $v \in \mathcal{V}$ ?

- We call the resulting optimal vector  $v$  the **least-squares estimate** of  $x$  in  $\mathcal{V}$ , because **in a Hilbert space minimizing the (induced norm) of the error is equivalent to minimizing the “squared-error”**  $\|e\|^2 = \langle e, e \rangle$ .

- Let  $v_0 = P_{\mathcal{V}} x$  be the orthogonal projection of  $x$  onto  $\mathcal{V}$ .

- Note that

$$P_{\mathcal{V}^\perp} x = (I - P_{\mathcal{V}})x = x - P_{\mathcal{V}} x = x - v_0$$

must be orthogonal to  $\mathcal{V}$ .

- For any vector  $v \in \mathcal{V}$  we have

$$\|e\|^2 = \|x - v\|^2 = \|(x - v_0) + (v_0 - v)\|^2 = \|x - v_0\|^2 + \|v_0 - v\|^2 \geq \|x - v_0\|^2,$$

as an easy consequence of the Pythagorean theorem. (Note that the vector  $v - v_0$  must be in the subspace  $\mathcal{V}$ .)

- **Thus the error is minimized when  $v = v_0$ .**

# Projection Theorem & Orthogonality Principle – Cont.

- Because  $v_0$  is the orthogonal projection of  $x$  onto  $\mathcal{V}$ , the least-squares optimality of  $v_0$  is known as the

**Projection Theorem:**  $v_0 = P_{\mathcal{V}} x$

- Alternatively, recognizing that the optimal error must be orthogonal to  $\mathcal{V}$ ,  $(x - v_0) \perp \mathcal{V}$ , this result is also equivalently known as the

**Orthogonality Principle:**  $\langle x - v_0, v \rangle = 0$  for all  $v \in \mathcal{V}$ .

# Least-Squares Soln to Ill-Posed Linear Inverse Problem

- Consider a linear operator  $A : \mathcal{X} \rightarrow \mathcal{Y}$  between two finite-dim Hilbert spaces and the associated inverse problem  $y = Ax$  for a specified measurement vector  $y$ .
- In the prediction-error discrepancy minimization approach to solving inverse problems discussed above, it is now natural to use the inner product induced norm as the model discrepancy measure

$$D^2(e) = \|e\|^2 = \langle e, e \rangle = \|y - \hat{y}\|^2 = \|y - Ax\|^2$$

- With  $\mathcal{R}(A)$  a subspace of the Hilbert space  $\mathcal{Y}$ , we see that we are looking for the best approximation  $\hat{y} = Ax$  to  $y$  in the subspace  $\mathcal{R}(A)$ ,

$$\min_{\hat{y} \in \mathcal{R}(A)} \|y - \hat{y}\|^2 = \min_{x \in \mathcal{X}} \|y - Ax\|^2$$

- From the Projection Theorem, we know that the solution to this problem is given by the following geometric condition

**Geometric Condition for a Least-Squares Solution:  $e = y - Ax \perp \mathcal{R}(A)$**

which must hold for any  $x$  which produces a least-squares solution  $\hat{y}$ .

# Generalized Solns to Ill-Posed Linear Inverse Problems

- Taking the linear operator  $A$  to be a mapping between Hilbert spaces, we can obtain a **generalized least-squares solution** to an ill-posed linear inverse problem  $Ax = y$  by looking for the unique solution to the

Regularized Least-squares Problem: 
$$\min_x \|y - Ax\|^2 + \beta \|x\|^2$$

where the indicated norms are the inner product induced norms on the domain and codomain and  $\beta > 0$ .

- The solution to this problem,  $\hat{x}_\beta$ , is a function of the **regularization parameter**  $\beta$ . The choice of the precise value of the regularization parameter  $\beta$  is often a nontrivial problem.
- The unique limiting solution

$$\hat{x} \triangleq \lim_{\beta \rightarrow 0} \hat{x}_\beta,$$

is a **minimum norm least-squares solution**, aka **pseudoinverse solution**

- The operator  $A^+$  which maps  $y$  to the solution,  $\hat{x} = A^+y$  is called the **pseudoinverse of  $A$** . The pseudoinverse  $A^+$  is a linear operator.
- In the special case when  $A$  is square and full-rank, it must be the case that  $A^+ = A^{-1}$  showing that the pseudoinverse is a **generalized inverse**.



# The Pseudoinverse Solution

- The pseudoinverse solution,  $\hat{x}$ , is the unique least-squares solution to the linear inverse problem having **minimum norm among all least-squares solutions** to the least squares problem of minimizing  $\|e\|^2 = \|y - Ax\|^2$ ,

$$\hat{x} = \arg \min_{x'} \left\{ \|x'\| \mid x' \in \arg \min_x \|y - Ax\|^2 \right\}$$

Thus the pseudoinverse solution is a least-squares solution.

- Because  $Ax \in \mathcal{R}(A)$ ,  $\forall x$ , **any** particular least-squares solution,  $x'$ , to the inverse problem  $y = Ax$  yields a value  $\hat{y} = Ax'$  which is the unique least-squares approximation of  $y$  in the subspace  $\mathcal{R}(A) \subset \mathcal{Y}$ ,

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

- As discussed above, the orthogonality condition determines the least-squares approximation  $\hat{y} = Ax'$  from the

**Geometric Condition for a Least-Squares Solution:  $e = y - Ax' \in \mathcal{R}(A)^\perp$**

The pseudoinverse solution has the smallest norm,  $\|x'\|$ , among all vectors  $x'$  that satisfy the orthogonality condition.

# The Pseudoinverse Solution – Cont.

- Because  $\mathcal{X} = \mathcal{N}(A)^\perp \oplus \mathcal{N}(A)$  we can write any *particular* least-squares solution,  $x' \in \arg \min_x \|y - Ax\|^2$  as

$$x' = (P_{\mathcal{N}(A)^\perp} + P_{\mathcal{N}(A)}) x' = P_{\mathcal{N}(A)^\perp} x' + P_{\mathcal{N}(A)} x' = \hat{x} + x'_{\text{null}},$$

- Note that

$$\hat{y} = Ax' = A(\hat{x} + x'_{\text{null}}) = A\hat{x} + Ax'_{\text{null}} = A\hat{x}.$$

- $\hat{x} \in \mathcal{N}(A)^\perp$  is **unique**. I.e., independent of the particular choice of  $x'$ .
- The least squares solution  **$\hat{x}$  is the unique minimum norm least-squares solution**.
- This is true because the Pythagorean theorem yields

$$\|x'\|^2 = \|\hat{x}\|^2 + \|x'_{\text{null}}\|^2 \geq \|\hat{x}\|^2,$$

showing that  $\hat{x}$  is indeed the minimum norm least-squares solution.

- Thus the geometric condition that a least-squares solution  $x'$  is also a minimum norm solution is that  $x' \perp \mathcal{N}(A)$ :

**Geometric Condition for a Minimum Norm LS Solution:  $x' \in \mathcal{N}(A)^\perp$**

# Geometric Conditions for a Pseudoinverse Solution

1. **Geometric Condition for a Least-Squares Solution:**  $e = y - Ax' \in \mathcal{R}(A)^\perp$

2. **Geometric Condition for a Minimum Norm LS Solution:**  $x' \in \mathcal{N}(A)^\perp$

- The primary condition (1) ensures that  $x'$  is a least-squares solution to the inverse problem  $y = Ax$ .
- The secondary condition (2) then ensures that  $x'$  is the unique minimum norm least squares solution,  $x' = \hat{x} = A^+y$ .
- We want to move from the insightful geometric conditions to equivalent algebraic conditions that will allow us to solve for the pseudoinverse solution  $\hat{x}$ .
  - To do this we introduce the concept of the **Adjoint Operator**,  $A^*$ , of a linear operator  $A$ .

# The Adjoint Operator - Motivation

Given a linear operator  $A : \mathcal{X} \rightarrow \mathcal{Y}$  mapping between two Hilbert spaces  $\mathcal{X}$  and  $\mathcal{Y}$ , **suppose** that we can find a companion linear operator  $M$  that maps in the reverse direction  $M : \mathcal{Y} \rightarrow \mathcal{X}$  such that

$$(\star\star) \quad \mathcal{N}(M) = \mathcal{R}(A)^\perp \quad \text{and} \quad \mathcal{R}(M) = \mathcal{N}(A)^\perp$$

Then the two geometric conditions for a least-squares solution become

$$M(y - Ax) = 0 \quad \Rightarrow \quad MAx = My$$

and

$$x = M\lambda \quad \Rightarrow \quad x - M\lambda = 0 \quad \text{for some } \lambda \in \mathcal{Y}$$

which we can write as

$$\begin{pmatrix} MA & 0 \\ I & -M \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} My \\ 0 \end{pmatrix},$$

which can be jointly solved for the pseudoinverse solution  $x = \hat{x}$  and a “nuisance parameter”  $\lambda$ .

- The companion linear operator  $M$  having the properties **( $\star\star$ )** shown above: (1) exists; (2) is unique; and (3) allows for the determination of  $\hat{x}$  and  $\lambda$ . It is known as the **Adjoint Operator,  $A^*$ , of  $A$ .**

# Existence of the Adjoint Operator

Given  $A : \mathcal{X} \rightarrow \mathcal{Y}$  define  $A^* \triangleq M$  by

$$\langle My, x \rangle = \langle y, Ax \rangle \quad \forall x \in \mathcal{X} \quad \forall y \in \mathcal{Y}$$

**Uniqueness:** Suppose  $M$  and  $M'$  both satisfy the above condition. Then

$$\begin{aligned}\langle My, x \rangle &= \langle M'y, x \rangle = \quad \forall x, \quad \forall y \\ \langle My - M'y, x \rangle &= 0 \quad \forall x, \quad \forall y \\ \langle (M - M')y, x \rangle &= 0 \quad \forall x, \quad \forall y \\ \langle (M - M')y, (M - M')y \rangle &= 0 \quad \forall y \\ \|(M - M')y\|^2 &= 0 \quad \forall y \\ (M - M')y &= 0 \quad \forall y \\ M - M' &= \mathbf{0} \\ M &= M'\end{aligned}$$

# Existence of the Adjoint Operator – Cont.

**Linearity:** For any  $\alpha_1, \alpha_2, y_1, y_2$ , and for all  $x$ , we have

$$\begin{aligned}\langle M(\alpha_1 y_1 + \alpha_2 y_2), x \rangle &= \langle \alpha_1 y_1 + \alpha_2 y_2, Ax \rangle \\ &= \bar{\alpha}_1 \langle y_1, Ax \rangle + \bar{\alpha}_2 \langle y_2, Ax \rangle \\ &= \bar{\alpha}_1 \langle My_1, x \rangle + \bar{\alpha}_2 \langle My_2, x \rangle \\ &= \langle \alpha_1 My_1 + \alpha_2 My_2, x \rangle \\ \Rightarrow \langle M(\alpha_1 y_1 + \alpha_2 y_2) - (\alpha_1 My_1 + \alpha_2 My_2), x \rangle &= 0\end{aligned}$$

**Existence:** Typically shown by construction for a given problem.

For example, assuming the standard column-vector representation for finite-dimensional vectors, take  $\mathcal{X} = \mathbb{C}^n$  with inner product  $\langle x_1, x_2 \rangle = x_1^H \Omega x_2$ , for hermitian, positive-definite  $\Omega$ , and  $\mathcal{Y} = \mathbb{C}^m$  with inner product  $\langle y_1, y_2 \rangle = y_1^H W y_2$ , for hermitian, positive-definite  $W$ . Then

$$\langle My, x \rangle = y^H M^H \Omega x \quad \text{and} \quad \langle y, Ax \rangle = y^H W A x = y^H \underbrace{(W A \Omega^{-1})}_{M^H} \Omega x$$
$$M = (W A \Omega^{-1})^H \quad \iff \quad M = \Omega^{-1} A^H W$$

# Existence of the Adjoint Operator – Cont.

**Proof that  $\mathcal{N}(M) = \mathcal{R}(A)^\perp$ .**

Recall that two sets are equal iff they contain the same elements.

We have

$$\begin{aligned}y \in \mathcal{R}(A)^\perp &\iff \langle y, Ax \rangle = 0, \quad \forall x \\ &\iff \langle My, x \rangle = 0, \quad \forall x \\ &\iff My = 0 \quad (\text{prove this last step}) \\ &\iff y \in \mathcal{N}(M)\end{aligned}$$

showing that  $\mathcal{R}(A)^\perp = \mathcal{N}(M)$ .

# Existence of the Adjoint Operator – Cont.

**Proof that  $\mathcal{R}(M) = \mathcal{N}(A)^\perp$ .**

Note that  $\mathcal{R}(M) = \mathcal{N}(A)^\perp$  iff  $\mathcal{R}(M)^\perp = \mathcal{N}(A)$

We have

$$\begin{aligned}x \in \mathcal{R}(M)^\perp &\iff \langle My, x \rangle = 0, \quad \forall y \\ &\iff \langle y, Ax \rangle = 0, \quad \forall y \\ &\iff Ax = 0 \quad (\text{prove this last step}) \\ &\iff x \in \mathcal{N}(A)\end{aligned}$$

showing that  $\mathcal{R}(M)^\perp = \mathcal{N}(A)$ .



# Algebraic Conditions for the Pseudoinverse Solution

We have transformed the geometric conditions (1) and (2) for obtaining the minimum norm least squares solution to the linear inverse problem  $y = Ax$  into the corresponding algebraic conditions

1. Algebraic Condition for an LS Solution – The Normal Equation:  $A^*Ax = A^*y$

2. Algebraic Condition for a Minimum Norm LS Solution:  $x = A^*\lambda$

where  $A^* : \mathcal{Y} \rightarrow \mathcal{X}$ , the adjoint of  $A : \mathcal{X} \rightarrow \mathcal{Y}$ , is given by

Definition of the Adjoint Operator,  $A^*$ :  $\langle A^*y, x \rangle = \langle y, Ax \rangle \quad \forall x, \forall y$

When the (finite dimensional) domain has a metric matrix  $\Omega$  and the (finite dimensional) codomain has metric matrix  $W$ , then (assuming the standard column vector coordinate representation) adjoint operator is given by

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

which is a type of “generalized transpose” of  $A$ .

# Solving for the Pseudoinverse Solution

Note that the normal equation is always consistent by construction

$$A^* e = A^*(y - \hat{y}) = A^*(y - Ax) = 0 \text{ (consistent)} \Rightarrow A^* Ax = A^* y$$

Similarly, one can always enforce the minimum norm condition

$$x = A^* \lambda$$

by an appropriate projection of  $n$  onto  $\mathcal{R}(A^*) = \mathcal{N}(A)^\perp$ . In principle, the combination of these two equations

$$\begin{pmatrix} A^* A & 0 \\ I & -A^* \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} A^* y \\ 0 \end{pmatrix}$$

can always be solved for  $\hat{x}$  and  $\lambda$ . In practice, this is typically only done for the case when  $A$  is full rank.

Recall that the process of solving for  $\hat{x}$  given a measurement  $y$  is described as an action, or operation, on  $y$  by the so-called pseudoinverse operator  $A^+$ ,  $\hat{x} = A^+ y$ . Because  $A^*$  is a linear operator (as was shown above) and the product of any two linear operators is also a linear operator (as can be easily proved), the above system of equations is linear. Thus the solution of  $\hat{x}$  depends linearly on  $y$  and therefore:

**The pseudoinverse  $A^+$  is a linear operator.**

## Solving for the Pseudoinverse Solution - Cont.

In general, solving for the pseudoinverse operator  $A^+$  or solution  $\hat{x}$  requires nontrivial numerical machinery (such as the utilization of the Singular Value Decomposition (SVD)).

However, there are two special cases for which the pseudoinverse equations given above have a straightforward solution.

These correspond to the two possible ways that  $A$  can have full rank:

- When  $A$  is onto. (Corresponding to a matrix  $A$  having full row rank.)
- When  $A$  is one-to-one. (Corresponding to a matrix  $A$  with full column rank.)