ECE 174

Linear Vector Space Concepts
Need for Vector Space Theory

- What are *Vectors* and *Linear Vector Spaces*?
- What are *Norms* and *Normed Linear Vector Spaces*? (Theory of *Banach Spaces*.)
- What are *Inner Products* and *Inner Product Vector Spaces*? (Theory of *Hilbert Spaces*.)
- What are *Linear Operators* and the *Geometry Induced by Linear Operators*. (The ‘*Four Fundamental Subspaces*’ associated with a linear operator.)
- What is a *Linear Inverse Problem*? (*Well-Posed* and *Ill-Posed* Inverse Problems.)
- How does one solve a linear inverse problem?
  - *Minimum Norm Solution* and *Weighted Least Squares Solution*.
  - *Projection Theorem* in Hilbert Spaces. (*Orthogonality Principle*.)
  - *Generalized Inverses*. (*Pseudo-Inverse, QR-factorization, SVD.*)
Heuristic Concept of a Vector

Many important physical, engineering, biological, sociological, economic, scientific quantities, which we call *vectors*, have the following conceptual properties.

- They have a natural ‘zero point’ or “origin”, the *zero vector*, 0, or admit the use of a conventional zero point.

- They can be *added* in a symmetric *commutative* and *associative* manner to produce new vectors

  \[ z = x + y = y + x , \quad x,y,z \text{ are vectors} \quad \text{(commutativity)} \]

  \[ x + y + z \equiv x + (y + z) = (x + y) + z , \quad x,y,z \text{ are vectors} \quad \text{(associativity)} \]

- They can be symmetrically scaled by the multiplication of scalars (*scalar multiplication*) to produce new vectors

  \[ z = \alpha x = x\alpha , \quad x,z \text{ are vectors, } \alpha \text{ is a scalar} \quad \text{(scalar multiplication of } x \text{ by } \alpha) \]

- Scalars can be members of any fixed *field* (such as the field of rational polynomials). We will work only with the fields of real and complex numbers.

- They have an *additive inverse*, \(-x = (-1)x\)

  \[ x - x \equiv x + (-1)x = x + (-x) = 0 \]
**Linear Vector Spaces**

- A **Vector Space**, \( \mathcal{X} \), is a collection of **vectors**, \( x \in \mathcal{X} \), over a **field**, \( \mathcal{F} \), of **scalars**.
  - If the scalars are the field of real numbers, then we have a **Real** Vector Space.
  - If the scalars are the field of complex numbers, then we have a **Complex** Vector Space.

- Any vector \( x \in \mathcal{X} \) can be multiplied by an arbitrary scalar \( \alpha \) to form \( \alpha x = x \alpha \in \mathcal{X} \). This is called **scalar multiplication**.
  - Note that we must have **closure of scalar multiplication**. I.e, we demand that the new vector formed via scalar multiplication must also be in \( \mathcal{X} \).

- Any two vectors \( x, y \in \mathcal{X} \) can be added to form \( x + y \in \mathcal{X} \) where the operation “+” of **vector addition** is associative and commutative.
  - Note that we must have **closure of vector addition**.

- The vector space \( \mathcal{X} \) must contain an **additive identity** (the **zero vector** \( 0 \)) and, for every vector \( x \), an **additive inverse** \( -x \).

- In this course we primarily **finite dimensional** vector spaces \( \dim \mathcal{X} = n < \infty \) and mostly give results appropriate for this restriction.
Linear Vector Spaces – Cont.

• Any vector $x$ in an $n$-dimensional vector space can be represented (with respect to an appropriate basis—see below) as an $n$-tuple ($n \times 1$ column vector) over the field of scalars,

$$x = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \in \mathcal{X} = \mathbb{F}^n = \mathbb{C}^n \text{ or } \mathbb{R}^n.$$ 

• We refer to this as a **canonical representation** of a finite-dimensional vector. We often (but not always) assume that vectors in an $n$-dimensional vector space are canonically represented by $n \times 1$ column vectors.
Any linear combination of arbitrarily selected vectors \( x_1, \cdots, x_r \) drawn from \( \mathcal{X} \)

\[
\alpha_1 x_1 + \cdots + \alpha_r x_r
\]

for arbitrary \( r \), and scalars \( \alpha_i, i = 1, \cdots, r \), must also be a vector in \( \mathcal{X} \).

- This is easily shown via induction using the properties of closure under pairwise vector addition, closure under scalar multiplication, and associativity of vector addition.

- This global ‘closure of linear combinations property of \( \mathcal{X} \)’ (i.e., the property holds everywhere on \( \mathcal{X} \)) is why we often refer to \( \mathcal{X} \) as a (globally) Linear Vector Space.

- This is in contradistinction to locally linear spaces, such as differentiable manifolds, of which the surface of a ball is the classic example of a space which is locally linear (flat) but globally curved.

- Some physical phenomenon of interest cannot be modeled by linear vector spaces, the classic example being rotations of a rigid body in three dimensional space. (Finite rotations do not commute.)
Examples of Vectors

Voltages, Currents, Power, Energy, Forces, Displacements, Velocities, Accelerations, Temperature, Torques, Angular Velocities, Income, Profits, ...., can all be modeled as vectors.

Example: Set of all $m \times n$ matrices. Define matrix addition by component-wise addition and scalar multiplication by component-wise multiplication of the matrix component by the scalar. This is easily shown to be a vector space.

- We can place the elements of this vector space into canonical form by stacking the columns of an $m \times n$ matrix $A$ to form an $mn \times 1$ column vector denoted by $\text{vec}(A)$ (sometimes also denoted by $\text{stack}(A)$).

Example: Take

$$\mathcal{X} = \{ f(t) = x_1 \cos(\omega_1 t) + x_2 \cos(\omega_2 t) \text{ for } -\infty < t < \infty; x_1, x_2 \in \mathbb{R}; \omega_1 \neq \Omega_2 \}$$

and define vector addition and scalar multiplication component wise. Note that any vector $f \in \mathcal{X}$ has a canonical representation $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$. Thus $\mathcal{X} \cong \mathcal{X}' \triangleq \mathbb{R}^2$, and without loss of generality (wlog) we usually work with $\mathcal{X}'$ in lieu of $\mathcal{X}$.
Important Example: Set of all Functions forms a Vector Space

- Consider functions (say of time $t$) $f$ and $g$, which we sometimes also denote as $f(\cdot)$ and $g(\cdot)$.
  - $f(t)$ is the value of the function $f$ at time $t$. (Think of $f(t)$ as sample of $f$ at time $t$.) Strictly speaking, then, $f(t)$ is not the function $f$ itself.
- Functions are single-valued by definition. Therefore

$$f(t) = g(t), \quad \forall t \iff f = g$$

I.e., functions are uniquely defined once we know their output values for all possible input values $t$

- We can define vector addition to create a new function $h = f + g$ by specifying the value of $h(t)$ for all $t$, which we do as follows:

$$h(t) = (f + g)(t) \triangleq f(t) + g(t), \quad \forall t$$

- We define scalar multiplication of the function $f$ by the scalar $\alpha$ to create a new function $g = (\alpha f)$ via

$$(\alpha f)(t) = \alpha \cdot f(t), \quad \forall t$$

- Finally we define the zero function $o$ as the function that maps to the scalar value 0 for all $t$, $o(t) = 0, \quad \forall t$.  

Examples of Vectors – Cont.
**Vector Subspaces**

- A subset $\mathcal{V} \subset \mathcal{X}$ is a **subspace** of a vector space $\mathcal{X}$ if it is a vector space in its own right.

- If $\mathcal{V}$ is a subspace of a vector space $\mathcal{X}$, we call $\mathcal{X}$ the **parent space** or **ambient space** of $\mathcal{V}$.

- *It is understood that a subspace $\mathcal{V}$ “inherits” the vector addition and scalar multiplication operations from the ambient space $\mathcal{X}.* To be a subspace, $\mathcal{V}$ must also inherit the zero vector element.

- Given this fact, to determine if a subset $\mathcal{V}$ is also a subspace one needs to check that every linear combination of vectors in $\mathcal{V}$ yields a vector in $\mathcal{V}$.

- This latter property is called the property of **closure of the subspace $\mathcal{V}$ under linear combinations of vectors in $\mathcal{V}$**.
  
  - Therefore if closure fails to hold for a subset $\mathcal{V}$, then $\mathcal{V}$ is **not** a vector subspace.
  
  - Note that testing for closure includes as a special case testing whether the zero vector belongs to $\mathcal{V}$. 
Vector Subspaces - Cont.

Consider the complex vector space \( \mathcal{X} = \text{complex } n \times n \text{ matrices}, \ n > 1 \), with matrix addition and scalar multiplication defined component-wise. Are the following subsets of \( \mathcal{X} \) vector subspaces?

- \( \mathcal{V} = \) upper triangular matrices. This is a subspace as it is closed under the operations of scalar multiplication and vector addition inherited from \( \mathcal{X} \).

- \( \mathcal{V} = \) positive definite matrices. This is not a subspace as it is not closed under scalar multiplication. (Or, even simpler, it does not contain the zero element.)

- \( \mathcal{V} = \) symmetric matrices, \( A = A^T \). This is a subspace as it is closed under the operators inherited from \( \mathcal{X} \).

- \( \mathcal{V} = \) hermitian matrices, \( A = A^H \) (complex symmetric matrices, \( A = (A^T) = (\bar{A})^T \)). This is not a subspace as it is not closed under scalar multiplication (check this!). It does include the zero element.
Subspace Sums

Given two subsets \( \mathcal{V} \) and \( \mathcal{W} \) of vectors, we define their **set sum** by

\[
\mathcal{V} + \mathcal{W} = \{ v + w | v \in \mathcal{V} \text{ and } w \in \mathcal{W} \}.
\]

Let the sets \( \mathcal{V} \) and \( \mathcal{W} \) in addition both be **subspaces** of \( \mathcal{X} \). In this case we call \( \mathcal{V} + \mathcal{W} \) a **subspace sum** and we have

- \( \mathcal{V} \cap \mathcal{W} \) and \( \mathcal{V} + \mathcal{W} \) are also subspaces of \( \mathcal{X} \)
- \( \mathcal{V} \cup \mathcal{W} \subseteq \mathcal{V} + \mathcal{W} \) where in general \( \mathcal{V} \cup \mathcal{W} \) is **not** a subspace.

In general, we have the subspace ordering

\[
0 \triangleq \{0\} \subset \mathcal{V} \cap \mathcal{W} \subset \mathcal{V} + \mathcal{W} \subset \mathcal{X},
\]

where \( \{0\} \) is the **trivial subspace** consisting only of the zero vector (additive identity) of \( \mathcal{X} \). The trivial subspace has dimension zero.
Linear Independence

• By definition, $r$ vectors $x_1, \cdots, x_r \in \mathcal{X}$ are **linearly independent** when,

\[
\alpha_1 x_1 + \cdots + \alpha_r x_r = 0 \quad \text{if and only if} \quad \alpha_1 = \cdots = \alpha_r = 0
\]

• Suppose this condition is violated because (say) $\alpha_1 \neq 0$, then we have

\[
x_1 = -\frac{1}{\alpha_1} (\alpha_2 x_2 + \cdots + \alpha_r x_r)
\]

• A collection of vectors are **linearly dependent** if they are **not** linearly independent.
Linear Independence - Cont.

- The definition of linear independence can be written in matrix-vector form as

\[
X \alpha = \begin{pmatrix} x_1 & \cdots & x_r \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_r \end{pmatrix} = 0 \iff \alpha \triangleq \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_r \end{pmatrix} = 0
\]

- Thus \( x_1, \ldots, x_r \) are linearly independent iff the associated matrix

\[
X \triangleq \begin{pmatrix} x_1 & \cdots & x_r \end{pmatrix}
\]

has full column rank (equivalently, iff the null space of \( X \) is trivial).

- Assuming that \( x_i \in \mathcal{F}^n \), the resulting matrix \( X = (x_1 \cdots x_r) \) is \( n \times r \).
**Span of a Set of Vectors**

- The *span* of the collection $x_1, \cdots, x_r \in \mathcal{X}$ is the set of all linear combinations of the vectors,

$$\text{Span}\{x_1, \cdots, x_r\} = \{y \mid y = \alpha_1 x_1 + \cdots + \alpha_r x_r = X \alpha, \ \forall \alpha \in F^r\} \subset \mathcal{X}$$

- The subset $\mathcal{V} = \text{Span}\{x_1, \cdots, x_r\}$ is a vector subspace of $\mathcal{X}$.

- If, in addition, the spanning vectors $x_1, \cdots, x_r$ are linearly independent we say that the collection is a *linearly independent spanning set* or a *basis* for the subspace $\mathcal{V}$.

- We denote a basis for a subspace $\mathcal{V}$ by

$$B_{\mathcal{V}} = \{x_1, \cdots, x_r\}$$
Basis and Dimension

- Given a basis for a vector space or subspace, the number of basis vectors in the basis is unique.

- For a given space or subspace, there are many different bases, but they must all have the same number of vectors.

- This number, then, is an intrinsic property of the space itself and is called the dimension \( d = \dim \mathcal{V} \) of the space or subspace \( \mathcal{V} \).

If the number of elements, \( d \), in a basis is finite, we say that the space is finite dimensional, otherwise we say that the space is infinite dimensional.

- Linear algebra is the study of linear mappings between finite dimensional vector spaces. The study of linear mappings between infinite dimensional vector spaces is known as Linear Functional Analysis or Linear Operator Theory.
Basis and Dimension – Cont.

- The dimension of the trivial subspace is zero, \( 0 = \dim \{0\} \).
- If \( \mathcal{V} \) is a subspace of \( \mathcal{X} \), \( \mathcal{V} \subset \mathcal{X} \), we have \( \dim \mathcal{V} \leq \dim \mathcal{X} \).
- In general for two arbitrary subspaces \( \mathcal{V} \) and \( \mathcal{W} \) of \( \mathcal{X} \) we have,

\[
\dim (\mathcal{V} + \mathcal{W}) = \dim \mathcal{V} + \dim \mathcal{W} - \dim (\mathcal{V} \cap \mathcal{W}) ,
\]

and

\[
0 \leq \dim (\mathcal{V} \cap \mathcal{W}) \leq \dim (\mathcal{V} + \mathcal{W}) \leq \dim \mathcal{X} .
\]

- Furthermore, if \( \mathcal{X} = \mathcal{V} + \mathcal{W} \) then,

\[
\dim \mathcal{X} \leq \dim \mathcal{V} + \dim \mathcal{W} ,
\]

with equality if and only if \( \mathcal{V} \cap \mathcal{W} = \{0\} \).
Two subspaces, $V$ and $W$, of a vector space $X$ are *independent* or *disjoint* when $V \cap W = \{0\}$. In this case we have

$$\dim (V + W) = \dim V + \dim W.$$ 

If $X = V + W$ for two *independent* subspaces $V$ and $W$ we say that $V$ and $W$ are *companion subspaces* and we write,

$$X = V \oplus W.$$ 

In this case $\dim X = \dim V + \dim W$.

Given two companion subspaces $V$ and $W$ any vector $x \in X$ can be written *uniquely* as

$$x = v + w, \quad v \in V \text{ and } w \in W.$$ 

The unique component $v$ is called *the projection of $x$ onto $V$ along its companion space $W$*.

The unique component $w$ is called *the projection of $x$ onto $W$ along its companion space $V$*. 
Independent Subspaces and Projections – Cont.

\[ \mathbf{x} = \mathbf{v} \oplus \mathbf{w} \]

Projection of \( \mathbf{x} \) onto \( \mathbf{v} \) along \( \mathbf{w} \)

Projection of \( \mathbf{x} \) onto \( \mathbf{w} \) along \( \mathbf{v} \)
**Projection Operators**

- Given the unique decomposition of a vector $x$ along two companion subspaces $V$ and $W$, $x = v + w$, we define the **companion projection operators** $P_{V|W}$ and $P_{W|V}$ by,

  $$ P_{V|W} x \triangleq v \quad \text{and} \quad P_{W|V} x = w $$

- Obviously $P_{V|W} + P_{W|V} = I$. i.e., $P_{V|W} = I - P_{W|V}$.

- It is straightforward to show that $P_{V|W}$ and $P_{W|V}$ are both **idempotent**, 

  $$ P^2_{V|W} = P_{V|W} \quad \text{and} \quad P^2_{W|V} = P_{W|V} $$

  where $P^2_{V|W} = (P_{V|W}) \ (P_{V|W})$. For example

  $$ P^2_{V|W} x = P_{V|W} (P_{V|W} x) = P_{V|W} v = v = P_{V|W} x $$

  and since this is true for all $x \in X$ it must be the case that $P^2_{V|W} = P_{V|W}$.

- It can also be shown that the projection operators $P_{V|W}$ and $P_{W|V}$ are **linear operators**.
Linear Operators and Matrices

Consider a function $A$ which maps between two vector spaces $\mathcal{X}$ and $\mathcal{Y}$, $A : \mathcal{X} \rightarrow \mathcal{Y}$.

- $\mathcal{X}$ is called the **input space** or the **source space** or the **domain**.
- $\mathcal{Y}$ is called the **output space** or the **target space** or the **codomain**.
- The mapping or operator $A$ is said to be **linear** if
  \[
  A (\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 A x_1 + \alpha_2 A x_2 \quad \forall x_1, x_2 \in \mathcal{X}, \forall \alpha_1, \alpha_2 \in \mathcal{F}.
  \]

- Note that in order for this definition to be well-posed the vector spaces $\mathcal{X}$ and $\mathcal{Y}$ both must have the same field of scalars $\mathcal{F}$.
- For example, $\mathcal{X}$ and $\mathcal{Y}$ must be both real vectors spaces, or must be both complex vector spaces.
It is well-known that any linear operator between finite dimensional vectors spaces has a matrix representation.

In particular if $n = \dim \mathcal{X} < \infty$ and $m = \dim \mathcal{Y} < \infty$ for two vector spaces over the field $\mathcal{F}$, then a linear operator $A$ which maps between these two spaces has an $m \times n$ matrix representation over the field $\mathcal{F}$.

Note that projection operators on finite-dimensional vector spaces must have matrix representations.

Often, for convenience, we assume that any such linear mapping $A$ is an $m \times n$ matrix and we write $A \in \mathcal{F}^{m \times n}$.

**Example:** Differentiation as a linear mapping between 2nd order polynomials

$$b + 2cx = \frac{d}{dx} (a + bx + cx^2) \iff \begin{pmatrix} b \\ 2c \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

using the simple polynomial basis functions $1$, $x$, and $x^2$. If a different set of polynomial basis functions are used, then we would have a different vector-matrix representation of the differentiation. **Again we note: representations of vectors and operators are basis dependent.**
Two Linear Operator Induced Subspaces

• Every linear operator has two natural vector subspaces associated with it.

The **Range Space**, 

\[ \mathcal{R}(A) \triangleq A(\mathcal{X}) \triangleq \{ y \mid y = Ax, \; x \in \mathcal{X} \} \subset \mathcal{Y}, \]

The **Nullspace**, 

\[ \mathcal{N}(A) = \{ x \mid Ax = 0 \} \in \mathcal{X}. \]

• Note that the nullspace is a subspace of the source space (domain), while the range space is a subspace of the target space (the codomain).

• It is straightforward to show that \( \mathcal{N}(A) \) and \( \mathcal{R}(A) \) are linear subspaces using the fact that \( A \) is a *linear* operator.

• When attempting to solve a linear problem \( y = Ax \), a solution exists if and only if \( y \in \mathcal{R}(A) \).

• If \( y \in \mathcal{R}(A) \) we say that the problem is *consistent*. Otherwise the problem is *inconsistent*. 
Two Linear Operator Induced Subspaces – Cont.

- The dimension of the range space of a linear operator $A$ is called the *rank* of $A$,

$$r(A) = \text{rank}(A) = \dim \mathcal{R}(A),$$

- the dimension of the nullspace of a linear operator $A$ is called the *nullity* of $A$,

$$\nu(A) = \text{nullity}(A) = \dim \mathcal{N}(A),$$

- The rank and nullity of a linear operator $A$ have unique values which are independent of the specific matrix representation of $A$. They are *intrinsic* properties of the linear operator $A$ and *invariant* with respect to all changes of representation. Note that, as dimensions, the rank and nullity must take on nonnegative integer or zero values.

- Given a matrix representation for $A \in \mathbb{F}^{m \times n}$ standard undergraduate courses in linear algebra explain how to determine the rank and nullity via LU factorization (aka Gaussian elimination) to place a matrix into upper echelon form. The rank, $r = r(A)$ is then given by the number of nonzero pivots while the nullity, $\nu = \nu(A)$, is given by $\nu = n - r$. 
**Linear Forward and Inverse Problem**

- Given a linear mapping between two vector spaces $A : \mathcal{X} \rightarrow \mathcal{Y}$ the problem of computing an “output” $y$ in the codomain given an “input” vector $x$ in the domain,

\[ Ax \rightarrow y \]

is called the **forward problem**.

- **the forward problem is always well-posed** in that knowing $A$ and given $x$ one can construct $y$ by (say) a straightforward matrix-vector multiplication.

- Given a vector $y$ in the codomain, the problem of determining an $x$ in the domain for which

\[ y \rightarrow Ax \]

is known as an **inverse problem**.

- Solving the linear inverse problem is much harder than solving the forward problem, even when the problem is well-posed.

Furthermore the **inverse problem is often ill-posed** compounding the problem difficulty
Well-Posed and Ill-Posed Linear Inverse Problems

Given an $m$-dimensional vector $y$ in the codomain, the inverse problem of determining an $n$-dimensional vector $x$ in the domain for which $Ax = y$ is said to be well-posed if and only if the following three conditions are true for the linear mapping $A$:

1. $y \in \mathcal{R}(A)$ for all $y \in \mathcal{Y}$ so that a solution exists for all $y$. I.e., we demand that $A$ be onto, $\mathcal{R}(A) = \mathcal{Y}$ or, equivalently, that $r(A) = m$. It is not enough to merely require consistency for a given $y$ because even the tiniest error or misspecification in $y$ can render the problem inconsistent.

2. If a solution exists, we demand that it be unique. I.e., we demand that $A$ be one-to-one, $\mathcal{N}(A) = \{0\}$. Equivalently, $\nu(A) = 0$.

3. The solution $x$ does not depend sensitively on the value of $y$. I.e., we demand that $A$ be numerically well-conditioned.

If any of these three conditions is violated we say that the inverse problem is ill-posed.

Condition three is studied in great depth in courses on Numerical Linear Algebra. In this course, we ignore the numerical conditioning problem and focus on the first two conditions only.

In particular, we will generalize the concept of solution by looking for a minimum-norm least-squares solution which will exist even when the first two conditions are violated.
**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 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 $x = (x[1], \cdots, x[n])^T \in \mathcal{F}^n$, $\mathcal{F} = \mathbb{C}$ 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 focus 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.
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

$$x \perp y \iff \langle x, y \rangle = 0$$

- If $x \perp y$, then

$$||x + y||^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle = \langle x, x \rangle + 0 + 0 + \langle y, y \rangle = ||x||^2 + ||y||^2$$

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

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

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**, \( V \perp W \) if and only if every vector in \( V \) is orthogonal to every vector in \( W \).

- If \( V \perp W \) it must be the case that \( V \) are disjoint \( W \), \( V \cap W = \{0\} \).

- Given a subspace \( V \) of \( X \), one defines the **orthogonal complement** \( V^\perp \) of \( V \) to be the set \( V^\perp \) of all vectors in \( X \) which are perpendicular to \( V \).

- The orthogonal complement (in the finite dimensional case assumed here) obeys the property \( V^\perp \perp = V \).

- The orthogonal complement \( V^\perp \) is **unique and a subspace** in its own right for which

\[
X = V \oplus V^\perp.
\]

- Thus \( V \) and \( V^\perp \) are complementary subspaces.

- Thus \( V^\perp \) is **more** than a complementary subspace to \( V \),

\[
V^\perp \text{ is the orthogonally complementary subspace to } V.
\]

- Note that it must be the case that

\[
\dim X = \dim V + \dim V^\perp.
\]
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 $A : \mathcal{X} \to \mathcal{Y}$ between two finite-dim Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$.

We must have that

$$\mathcal{Y} = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$$

and

$$\mathcal{X} = \mathcal{N}(A)^\perp \oplus \mathcal{N}(A).$$

- 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

  \textit{Projection Theorem:} \quad 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

  \textit{Orthogonality Principle:} \quad \langle x - v_0, v \rangle = 0 \text{ for all } v \in \mathcal{V}.
Consider a linear operator \( A : \mathcal{X} \rightarrow \mathcal{Y} \) between two finite-dim Hilbert spaces and the associated inverse problem \( y = Ax \) for 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

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

which must hold for any \( x \) which produces a least-squares solution \( \hat{y} \).
Taking the linear operator $A$ to be a mapping between Hilbert spaces, we can obtain a \textit{generalized least-squares solution} to an ill-posed linear inverse problem $Ax = y$ by looking for the unique solution to the \textit{regularized least-squares problem}:

$$\min_x \|y - Ax\|^2 + \beta \|x\|^2, \quad \beta > 0$$

where the indicated norms are the inner product induced norms on the domain and codomain.

The solution to this problem, $\hat{x}_\beta$, is a function of the \textit{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 \textit{minimum norm least-squares solution}, aka \textit{pseudoinverse solution}.

The operator $A^+$ which maps $y$ to the solution, $\hat{x} = A^+ y$ is called the \textit{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 \textit{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

**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 \( x' = N(A) \bot \oplus N(A) \) we can write any particular least-squares solution, \( x' \in \arg\min_x \| y - Ax \|^2 \) as

\[
x' = \left( P_{N(A)} \bot + P_{N(A)} \right) x' = P_{N(A)} \bot x' + P_{N(A)} x' = \hat{x} + x'_{\text{null}},
\]

• Note that

\[
\hat{y} = Ax' = A \left( \hat{x} + x'_{\text{null}} \right) = A\hat{x} + Ax'_{\text{null}} = A\hat{x}.
\]

• \( \hat{x} \in N(A) \bot \) 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 \( x' \perp N(A) \), or equivalently that

\[
\text{Geometric Condition for a Minimum Norm LS Solution: } \ x' \in N(A) \bot
\]
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

1. \( \mathcal{N}(M) = \mathcal{R}(A)^\perp \)
2. \( \mathcal{R}(M) = \mathcal{N}(A)^\perp \)

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, if the coefficient matrix is full rank, can be jointly solved for the pseudoinverse solution \( x = \hat{x} \) and the “nuisance parameter” \( \lambda \).

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

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

$$\langle My, x \rangle = \langle M'y, x \rangle = \forall x, \forall y$$

$$\langle My - M'y, x \rangle = 0 \quad \forall x, \forall y$$

$$\langle (M - M')y, x \rangle = 0 \quad \forall x, \forall y$$

$$\langle (M - M')y, (M - M')y \rangle = 0 \quad \forall y$$

$$\|(M - M')y\|^2 = 0 \quad \forall y$$

$$(M - M')y = 0 \quad \forall y$$

$$M - M' = 0$$

$$M = M'$$
Existence of the Adjoint Operator – Cont.

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

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

Existence: Typically shown by construction.

For example 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 $W$. Assume the standard column-vector representation for all vectors. Then

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

showing that

$$
M = (W A \Omega^{-1})^H \\
\Rightarrow \\
M = \Omega^{-1} A^H W
$$
Proof of Property (1): $\mathcal{N}(M) = \mathcal{R}(A)\perp$.

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

We have

\[
y \in \mathcal{R}(A)\perp \iff \langle y, Ax \rangle = 0, \quad \forall x
\]
\[
\iff \langle My, x \rangle = 0, \quad \forall x
\]
\[
\iff My = 0 \quad \text{(prove this last step)}
\]
\[
\iff y \in \mathcal{N}(M)
\]

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

Proof of Property (2): $\mathcal{R}(M) = \mathcal{N}(A) \perp$.

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

We have

\begin{align*}
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{align*}

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


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

where \( A^* : \mathcal{Y} \to \mathcal{X} \), the adjoint of \( A : \mathcal{X} \to \mathcal{Y} \), is given by

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

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

\[ A^* e = A^* (y - \hat{y}) = A^* (y - Ax) = 0 \text{ (consistent)} \implies A^* Ax = A^* y \]

Similarly, one can always enforce the minimum norm condition

\[ x = A^* \lambda \]

by an appropriate projection of \( n \) onto \( \mathcal{R}(A^*) = \mathcal{N}(A) \perp \).

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 uniquely for \( \hat{x} \) and \( \lambda \), as we shall come to understand.

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

\textit{the pseudoinverse} \( A^+ \) \textit{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 are

- When $A$ is onto
- When $A$ is one-to-one
The Four Fundamental Subspaces of a Linear Operator

For a linear operator \( A : \mathcal{X} \rightarrow \mathcal{Y} \),

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

Defining the adjoint \( A^* : \mathcal{Y} \rightarrow \mathcal{X} \) by

\[
\langle y, Ax \rangle = \langle A^* y, x \rangle
\]

we obtain (as was shown last lecture)

\[
\mathcal{R}(A^*) = \mathcal{N}(A)^\perp \quad \text{and} \quad \mathcal{N}(A^*) = \mathcal{R}(A)^\perp
\]

yielding

\[
\mathcal{X} = \mathcal{R}(A^*) \oplus \mathcal{N}(A) \quad \text{and} \quad \mathcal{Y} = \mathcal{R}(A) \oplus \mathcal{N}(A^*)
\]

- 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 \( \mathcal{R}(A^*) \).
Four Fundamental Subspaces – Cont.

Because

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

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

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

Thus the relationships between \( A \) and \( A^* \) and their four fundamental subspaces are entirely symmetrical,

\[ A : \mathcal{X} \rightarrow \mathcal{Y} , \quad A^* : \mathcal{Y} \rightarrow \mathcal{X} \]

\[ \langle y, Ax \rangle = \langle A^* y, x \rangle , \quad \langle x, A^* y \rangle = \langle Ax, y \rangle \]

\[ \mathcal{Y} = \mathcal{R}(A) \oplus \mathcal{N}(A^*) , \quad \mathcal{X} = \mathcal{R}(A^*) \oplus \mathcal{N}(A) \]

\[ \mathcal{R}(A) = \mathcal{N}(A^*)^{\perp} , \quad \mathcal{R}(A^*) = \mathcal{N}(A)^{\perp} \]
Because of this symmetry, sometimes we can get twice the number of results per derivation. For example, here we prove that

\[ \mathcal{N}(A) = \mathcal{N}(A^* A) \quad \text{and} \quad \mathcal{R}(A^*) = \mathcal{R}(A^* A) \]

from which it directly follows that

\[ \mathcal{N}(A^*) = \mathcal{N}(AA^*) \quad \text{and} \quad \mathcal{R}(A) = \mathcal{R}(AA^*) \]

Proofs of the first statements:

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

\[
x \in \mathcal{N}(A^* A) \iff A^* A x = 0
\]
\[
\quad \iff \langle \xi, A^* A x \rangle = 0, \quad \forall \xi
\]
\[
\quad \iff \langle A \xi, A x \rangle = 0, \quad \forall \xi
\]
\[
\quad \Rightarrow \langle A x, A x \rangle = \|A x\|^2 = 0
\]
\[
\quad \Rightarrow x \in \mathcal{N}(A)
\]
Proofs continued:

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

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

$$\iff x \perp \mathcal{N}(A)$$

$$\iff x \in \mathcal{R}(A^*)$$

Note that we have also established that

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

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

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

Note that $\dim \mathcal{R}(A) = \dim \mathcal{R}(A^*)$. 
### Adjoint-Based Conditions for a P-Inv Solution

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

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

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

- The geometric conditions conditions easily lead to the algebraic conditions

1. **Algebraic Cond. for an LS Solution – The Normal Equation:**
   \[ A^* Ax = A^* y \]

2. **Algebraic Cond. for a Minimum Norm LS Solution:**
   \[ x = A^* \lambda \]

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

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

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

- 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 On to

• 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

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

• Directly solving the equation \( AA^*\lambda = y \) for \( \lambda \) 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

\[
\langle Px_1, (I - P)x_2 \rangle = 0 \iff \langle (I - P)^* Px_1, x_2 \rangle = 0 \iff (I - P)^* P = 0 \iff P = P^* P
\]

which yields the property

Orthogonal Projection Operators are Self-Adjoint: \( P = P^* \)

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

**The restricted mapping \( A : \mathcal{R}(A^*) \rightarrow \mathcal{R}(A) \) 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 \( A : \mathcal{R}(A^*) \rightarrow \mathcal{R}(A) \) 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).
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,

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

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

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

\[
\hat{x} \xrightarrow{A} \hat{y} \xleftarrow{A^+}
\]

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 \textit{vice versa}.

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

\[
A : \mathcal{R}(A^*) \rightarrow \mathcal{R}(A) \quad \text{and} \quad A^+ : \mathcal{R}(A) \rightarrow \mathcal{R}(A^*)
\]

\textit{they are inverses of each other.}
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

$$\hat{y} = P_{\mathcal{R}(A)} y$$  \hspace{1cm} (\hat{y} \text{ is the least-squares estimate of } y)

$$= A\hat{x}$$  \hspace{1cm} (\hat{x} \text{ is a least-squares solution })

$$= AA^+ y$$

or

$$P_{\mathcal{R}(A)} y = AA^+ y, \hspace{0.5cm} \forall y \in \mathcal{Y}$$

Therefore

$$P_{\mathcal{R}(A)} = AA^+ \quad \text{and} \quad P_{\mathcal{R}^\perp(A)} = I - AA^+$$
Pseudoinverse & Orthogonal Projections – Cont.

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

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

Then

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

Now let

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

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

$$0 = A(x - \hat{x})$$

$$\Rightarrow x - \hat{x} \in \mathcal{N}(A) = \mathcal{R}(A^*)^\perp$$

$$\Rightarrow x - \hat{x} \perp \mathcal{R}(A^*)$$

$$\Rightarrow \hat{x} = P_{\mathcal{R}(A^*)} x \quad \text{(by the orthogonality principle)}$$

$$\Rightarrow A^+ A x = P_{\mathcal{R}(A^*)} x$$

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

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

Having shown that $AA^+ = P_R(A)$ and $A^+A = P_R(A^*)$, 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,

\[
\text{i. } (AA^+)^* = AA^+ \quad \text{ii. } (A^+A)^* = A^+A
\]

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

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

\[
AA^+AA^+ = AA^+ \quad \text{and} \quad A^+AA^+A = A^+A
\]

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

\[
\text{iii. } AA^+A = A \quad \text{iv. } A^+AA^+ = A^+
\]
Four M-P P-Inv Conditions

M-P THEOREM: (To be proven in the next lecture.)

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 $\alpha$

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

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

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

as $C^+ B^+ A^+$ in general does not satisfy the M-P conditions to be a p-inv of $ABC$. 
**Example 3:** In example 2, suppose the spaces have the standard Cartesian metric.

Furthermore, assume that $A$ and $C$ are unitary, $A^* = A^{-1}$ and $C^* = C^{-1}$. For complex spaces this means that $A^{-1} = A^* = A^H$, whereas for real spaces this means that $A$ and $B$ are orthogonal, $A^{-1} = A^* = A^T$.

Then we have that

$$(ABC)^+ = C^* B^+ A^* = C^{-1} B^+ A^{-1} = C^+ B^+ A^+$$

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

Thus, with the above *additional assumptions* on $A$ and $C$. we now *can* claim that $(ABC)^+ = C^+ B^+ A^+$.

**Example 4:** Suppose that $\Sigma$ 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

$$\Sigma^+ = \begin{pmatrix} S^+ & 0^T_2 \\ 0^T_1 & 0^T \end{pmatrix} = \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix}$$
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 $m \times m$, $\Sigma$ 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, $\sigma_i$, $i = 1, \ldots, r$, 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

$$S^+ = S^{-1} = \text{diag}^{-1}(\sigma_1 \cdots \sigma_r) = \text{diag}(\frac{1}{\sigma_1} \cdots \frac{1}{\sigma_r})$$
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 $\sigma_i$ 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 $\sigma_i$ 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

\[
A^+ = V \Sigma^+ U^H \quad \text{(from example 3)}
\]

\[
= \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^H & U_2^H \end{pmatrix} \quad \text{(from example 4)}
\]

\[
= V_1 S^{-1} U_1^H \quad \text{(from invertibility of } S)\]

\[
= \begin{pmatrix} v_1 & \cdots & v_r \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1} \\ \cdot \\ \cdot \\ \frac{1}{\sigma_r} \end{pmatrix} \begin{pmatrix} u_1^H \\ \vdots \\ u_r^H \end{pmatrix}
\]

\[
= \frac{1}{\sigma_1} v_1 u_1^H + \cdots + \frac{1}{\sigma_r} v_r u_r^H
\]

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.
**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,

$$A_{m \times n} : \mathcal{X} = \mathbb{C}^n \rightarrow \mathcal{Y} = \mathbb{C}^m$$

Note that $A$ in general is non-square and therefore does not have a spectral representation (because eigenvalues and eigenvectors are 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 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 $(\lambda_i, x_i)$, $i = 1, \cdots, n$, $A$ 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

\[
AA^H \quad \text{and} \quad A^HA
\]

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 \( \langle x, Mx \rangle = x^H M x \) is real and positive-semidefinite

\[
\langle x, Mx \rangle = x^H M x \geq 0 \quad \forall x
\]

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

\[
\langle x, Mx \rangle = x^H M x = 0 \quad \text{if and only if} \quad x = 0
\]
Singular Value Decomposition (SVD) – Cont.

In Lecture 9 we will show that 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 of Lecture 7. As noted in that example, knowledge of the SVD enables us to compute the pseudoinverse of $A$ in the rank deficient case.

The SVD will also allow us 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 will also provide a simple 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 $\mathcal{R}(A)$ in addition to the fact that $A$ maps $\mathcal{N}(A)$ to 0.
Singular Value Decomposition – Cont.

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

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

Therefore the number of nonzero (and hence strictly positive) eigenvalues of $AA^H$ and $A^HA$ must be equal to $r = r(A)$. 
**Singular Value Decomposition – Cont.**

Example 5 – Cont.

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 $m \times m$, $\Sigma$ 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, $\sigma_i, i = 1, \cdots, r$, 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

$$S^+ = S^{-1} = \text{diag}^{-1}(\sigma_1 \cdots \sigma_r) = \text{diag}(\frac{1}{\sigma_1} \cdots \frac{1}{\sigma_r})$$
Four M-P P-Inv Conditions – Cont.

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 $\sigma_i$ 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 $\sigma_i$ for $i = 1, \cdots, r$.

We have

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


**Four M-P P-Inv Conditions – Cont.**

**Example 5 – Cont.**

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

\[
A^+ = V \Sigma^+ U^H \quad \text{(from example 3)}
\]

\[
= \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^H & U_2^H \end{pmatrix} \quad \text{(from example 4)}
\]

\[
= V_1 S^{-1} U_1^H \quad \text{(from invertibility of } S)\]

\[
= \begin{pmatrix} v_1 & \cdots & v_r \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1} \\ \vdots \\ \frac{1}{\sigma_r} \end{pmatrix} \begin{pmatrix} u_1^H \\ \vdots \\ u_r^H \end{pmatrix}
\]

\[
= \frac{1}{\sigma_1} v_1 u_1^H + \cdots + \frac{1}{\sigma_r} v_r u_r^H
\]

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.
**Eigenstructure of $A^H A$**

Let $A : \mathcal{X} = \mathbb{C}^n \rightarrow \mathcal{Y} = \mathbb{C}^m$ 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 $AA^H : \mathbb{C}^m \rightarrow \mathbb{C}^m$ and $A^H A : \mathbb{C}^n \rightarrow \mathbb{C}^n$ must both be equal to $r = r(A)$.

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

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

with corresponding $n$-dimensional orthonormal eigenvectors

$$\begin{align*}
\text{span of } \mathcal{R}(A^H) & = \mathcal{N}(A)^\perp \\
\text{span } \mathcal{N}(A) & = \mathcal{N}(A^H A)
\end{align*}$$
Thus we have

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

and

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

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 \(\sigma_i^2\), \(1 \leq i \leq r\), is unique up to sign \(v_i \mapsto \pm v_i\).

Eigenvectors \(v_i\) associated with the same nondistinct nonzero eigenvalue \(\sigma_i^2\) 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

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

where

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

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

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

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

$$I_{r \times r} = V_1^H V_1, \quad I_{\nu \times \nu} = V_2^H V_2$$

It is straightforward to show that $V_1 V_1^H$ and $V_2 V_2^H$ are idempotent and self-adjoint.
Eigenstructure of $A^H A$ – Cont.

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

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

$$S_{r \times r} \triangleq \text{diag}(\sigma_1 \cdots \sigma_r)$$

Then

$$A^H A v_i = \sigma_i^2 v_i \quad 1 \leq i \leq r$$

can be written as

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

which yields

$$I_{r \times r} = S^{-1} V_1^H A^H A V_1 S^{-1} \quad (1)$$

We also note that

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

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

$$0_{m \times \nu} = AV_2 \quad (2)$$
**Eigenstructure of $AA^H$**

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

Above we have shown that

$$(A^HA)v_i = \sigma_i^2 v_i \quad i = 1, \ldots, r$$

where $\sigma_i^2$, $1 \leq i \leq r$, are nonzero. If we multiply both sides of this equation by $A$ we get

(recall that $r = r(AA^H) \leq m$)

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

Showing that $Av_i$ and $\sigma_i^2$ are eigenvector-eigenvalue pairs.

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

$$u_i \triangleq \frac{1}{\sigma_i} Av_i \quad 1 \leq i \leq r$$

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

$$S_{r \times r} = U_1^H AV_1$$ (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 $V_1$. 
Because $u_i$ correspond to the nonzero eigenvalues of $AA^H$ they must span $\mathcal{R}(AA^H) = \mathcal{R}(A)$. Therefore

$$\mathcal{R}(U_1) = \mathcal{R}(A) \quad \text{and} \quad P_{\mathcal{R}(A)} = U_1U_1^H$$

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 by Gram-Schmidt orthonormalization.) Let

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

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

$$0_{n \times \mu} = A^H U_2$$

and

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

(4)
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

\[ I_{m \times m} = U^H U = UU^H = U_1 U_1^H + U_2 U_2^H = P_R(A) + P_N(A^H) \]
Derivation of the SVD

\[ \Sigma_{m \times n} \triangleq U^H A V = \begin{pmatrix} U_1^H \\ U_2^H \end{pmatrix} A \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} U_1^H A V_1 & U_1 A V_2 \\ U_2^H A V_1 & U_2^H A V_2 \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \]

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 \( A \)

**SVD:** \( A = U \Sigma V^H = U_1 S V_1^H \)

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 $\sigma_i$ 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 $\sigma_i$ for $i = 1, \cdots, r$.

We have

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

Each term in this dyadic expansion is unique (i.e., does not depend on how the ambiguities mentioned above are resolved).
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 $\mathcal{N}(A)$,

$$Av_i = 0 \quad i = r + 1, \ldots, n$$

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

$$v_i \xrightarrow{A} \sigma_i u_i \quad i = 1, \ldots, r$$
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

\[
1 = \det I = \det UU^H = \det U \det U = |\det U|^2
\]

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 \( U \), \( \det U = e^{i\phi} \) 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 \( A \).

\[
|\det A| = \det USV^H = |\det U| \cdot |\det S| \cdot |\det V| = \det S = \sigma_1 \sigma_1 \cdots \sigma_n
\]
Exploiting identities provable from the M-P Theorem (see Lecture 7 and Homework 3) we have

\[ A^+ = V \Sigma^+ U^H \]

\[ = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} S^+ & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^H & U_2^H \end{pmatrix} \]

\[ = V_1 S^{-1} U_1^H \]

\[ = \begin{pmatrix} v_1 \cdots v_r \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1} \\ \vdots \\ \frac{1}{\sigma_r} \end{pmatrix} \begin{pmatrix} u_1^H \\ \vdots \\ u_r^H \end{pmatrix} \]

\[ = \frac{1}{\sigma_1} v_1 u_1^H + \cdots + \frac{1}{\sigma_r} v_r u_r^H \]

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

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

Obviously having access to the SVD is tremendously useful. With the background we have now covered, one can now 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 ambiguities 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 third 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 carefully designed so that you do not have to perform eigendecompositions to obtain the SVD’s. Rather, you can easily force the matrices into SVD factored form via a series of simple steps based on understanding the geometry underlying the SVD.

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

\[
\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{5}}{2} \\ \frac{\sqrt{5}}{2} \end{pmatrix} \cdot \sqrt{5} \cdot 1 = \begin{pmatrix} \frac{\sqrt{5}}{2} \\ \frac{\sqrt{5}}{2} \end{pmatrix} \cdot X \cdot \begin{pmatrix} \frac{\sqrt{5}}{2} \\ \frac{\sqrt{5}}{2} \end{pmatrix} \cdot 1 = \begin{pmatrix} \frac{\sqrt{5}}{2} \\ \frac{\sqrt{5}}{2} \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ 0 \end{pmatrix} \cdot 1
\]

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

\[
\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \cdot 2 \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & X \\ \frac{1}{\sqrt{2}} & X \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}
\]

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

\[
\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}
\]

Note the \( \pm \) sign ambiguity in the choice of \( U_2 \) and \( V_2 \).
**Weighted Least Squares and the Linear Gaussian Model**

**Linear Gaussian Model**

\[ y = Ax + v, \quad v \sim \mathcal{N}(0, C) \]

Equivalent to

**Parametric Probability Model**

\[ y \sim \mathcal{N}(Ax, C), \quad p_x(Y) = \frac{1}{\sqrt{(2\pi)^m \det C}} \exp \left\{ -\frac{1}{2} ||y - Ax||^2_{C^{-1}} \right\} \]

where

\[ ||y - Ax||^2_{C^{-1}} \triangleq (y - Ax)^T C^{-1} (y - Ax) \]

defines a **weighted 2-norm** with weighting matrix \( W = C^{-1} \).
The Likelihood Function

**Likelihood** of $x$ given $y$: $\ell_y(x) \triangleq p_x(y)$

In the figure below note that it is rational to prefer probability model $p_{x_2}(\cdot)$ over model $p_{x_1}(\cdot)$ given the observed value of $y$.

\[
\ell_y(x_2) = p_{x_2}(y) > p_{x_1}(y) = \ell_y(x_1)
\]
Model Fitting by Likelihood Maximization

The **Maximum Likelihood Estimate** of \( x \) given \( y \) is determined as

\[
\hat{x} = \arg \max_x \ell_y(x) = \arg \max_x p_x(y)
\]

For the Linear Gaussian model this is equivalent to solving the

**weighted least-squares problem**

\[
\hat{x} = \arg \min_x \| y - Ax \|_C^{-1}^2
\]

Which corresponds to solving a **Linear Inverse Problem**

\[
y \approx Ax
\]

in an appropriate **Minimum Norm** sense, where \( y \) and \( x \) are **Vectors**, \( A \) is a (matrix representation of) a **Linear Operator**, and \( \| \cdot \|_C^{-1} \) is a **weighted 2-Norm**.