Finite-Dimensional Vector Spaces

Kenneth Kreutz–Delgado
Electrical and Computer Engineering
Jacobs School of Engineering
University of California, San Diego

Linear Vector Space. A Vector Space, $\mathcal{X}$, is a collection of vectors, $x \in \mathcal{X}$, over a field, $\mathcal{F}$, of scalars. 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. The vector space $\mathcal{X}$ must contain an additive identity (the zero vector $0$) and, for every vector $x$, an additive inverse $-x$. The required properties of vector addition are given on page 160 of the textbook by Meyer.  

The properties of scalar multiplication, $\alpha x = \alpha \cdot x$, of a vector $x \in \mathcal{X}$ by a scalar $\alpha \in \mathcal{F}$ are also given on page 160 of Meyer. The scalars used in this course are either the field of real numbers, $\mathcal{F} = \mathbb{R}$, or the field of complex numbers, $\mathcal{F} = \mathbb{C}$.  

In this course, we consider only finite dimensional vector spaces dim $\mathcal{X} = n < \infty$. Any vector $x$ in an $n$–dimensional vector space can be represented as an $n$–tuple $(n \times 1$ column matrix.

---

2 Other fields commonly encountered are the field of rational numbers $\mathcal{F} = \mathbb{Q}$, which is used in numerical analysis and approximation theory, and the Galois field $\mathcal{F} = \{0, 1\}$, which is used in the theory of convolutional codes.
3 More general results applicable to the infinite dimensional case can be found in a variety of good texts, including Linear Operator Theory in Engineering and Science, A.W. Naylor and G.R. Sell, Springer–Verlag, 1982, which is available in a relatively inexpensive paperback edition.
vector) over the field of scalars,
\[
x = \begin{pmatrix} x(1) \\ \vdots \\ x(n) \end{pmatrix} \in \mathcal{X} = \mathcal{F}^n = \mathbb{C}^n \text{ or } \mathbb{R}^n.
\]

We henceforth assume that all vectors in an \(n\)-dimensional vector space are represented by \(n \times 1\) column vectors.

The complex conjugate of a scalar \(\alpha\) is denoted by \(\bar{\alpha}\). The complex conjugate of a matrix \(A\) (vector \(x\)) is denoted by \(\bar{A}\) (\(\bar{x}\)) and is defined to be the matrix (vector) of the complex conjugates of the elements of the matrix \(A\) (of the vector \(x\)). The hermitian conjugate of a matrix \(A\) (vector \(x\)), denoted by \(A^H\) (\(x^H\)), is defined to be the transpose of the complex conjugate, or equivalently the complex conjugate of the transpose, of the matrix (vector) \(A^H = (\bar{A})^T = (\bar{A})^T (x^H = (\bar{x})^T = (\bar{x})^T)\). A matrix is said to be hermitian if \(A^H = A\).

**Linear Vector Subspace.** A subset \(\mathcal{V} \subset \mathcal{X}\) is a subspace of \(\mathcal{X}\) if it is a vector space in its own right. It is understood that a subspace \(\mathcal{V}\) “inherits” the vector addition and scalar multiplication operations from the ambient space \(\mathcal{X}\). 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 under linear combinations of vectors in \(\mathcal{V}\). If \(\mathcal{V}\) is a subspace of a vector space \(\mathcal{X}\), we call \(\mathcal{X}\) the parent space or ambient space of \(\mathcal{V}\).

Given two sets \(\mathcal{V}\) and \(\mathcal{W}\) of vectors, we define their sum by
\[
\mathcal{V} + \mathcal{W} = \{v + w \mid v \in \mathcal{V} \text{ and } w \in \mathcal{W}\}.
\]

Note that \(\mathcal{V} \cup \mathcal{W} \subset \mathcal{V} + \mathcal{W}\). If, in addition, the sets \(\mathcal{V}\) and \(\mathcal{W}\) are subspaces of \(\mathcal{X}\), then \(\mathcal{V} \cap \mathcal{W}\) and \(\mathcal{V} + \mathcal{W}\) are also subspaces of \(\mathcal{X}\). In general,
\[
\{0\} \subset \mathcal{V} \cap \mathcal{W} \subset \mathcal{V} + \mathcal{W} \subset \mathcal{X},
\]
where \(\{0\}\) is the trivial subspace consisting only of the zero vector (additive identity) of \(\mathcal{X}\).

**Linear Independence and Dimension.** By definition \(r\) vectors \(x_1, \ldots, x_r \in \mathcal{X}\) are linearly independent when,
\[
\alpha_1 x_1 + \cdots + \alpha_r x_r = 0 \text{ if and only if } \alpha_1 = \cdots = \alpha_r = 0.
\]

---

4 Note that a real hermitian matrix is symmetric, \(A^T = A\).

5 Note the difference between sum and set union. In order for the sum to exist, the elements of the two sets must have a well-defined operation of addition, whereas the union exists whether or not the elements can be added.

6 In general \(\mathcal{V} \cup \mathcal{W}\) is not a subspace.
Note that this definition can be written in matrix–vector form as,

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

Thus \( x_1, \cdots, x_r \) are linearly independent iff the associated matrix \( X = (x_1 \cdots x_r) \) has full column rank (equivalently, iff the null space of \( X \) is trivial).

If \( x_1, \cdots, x_r \) are not linearly independent, then at least one of them can be written as a linear combination of the remaining vectors. In this case we say that the collection of vectors is linearly dependent.

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

\[ \text{Span} \{x_1, \cdots, x_r\} = \{y = \alpha_1 x_1 \cdots \alpha_r x_r = X\alpha \mid \forall \alpha \in \mathbb{F}^r \} \subset \mathcal{X}. \]

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 \( \mathcal{V} \) by \( B_{\mathcal{V}} = \{x_1, \cdots, x_r\} \).

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 elements. 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.

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 \( \mathcal{V} \cap \mathcal{W} = \{0\} \).

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, and is the subject matter of courses which are usually taught in graduate school.\(^8\)

\(^7\)Assuming that \( x_1 \in \mathbb{F}^n \), the resulting matrix \( X = (x_1 \cdots x_r) \) is \( n \times r \).

\(^8\)E.g., see the textbook by Naylor and Sell, op. cit.
Independent Subspaces and Projection Operators. Two subspaces, \( \mathcal{V} \) and \( \mathcal{W} \), of a vector space \( \mathcal{X} \) are independent or disjoint when \( \mathcal{V} \cap \mathcal{W} = \{0\} \). In this case we have

\[
\dim (\mathcal{V} + \mathcal{W}) = \dim \mathcal{V} + \dim \mathcal{W}.
\]

If \( \mathcal{X} = \mathcal{V} + \mathcal{W} \) for two independent subspaces \( \mathcal{V} \) and \( \mathcal{W} \) we say that \( \mathcal{V} \) and \( \mathcal{W} \) are companion subspaces and we write,

\[\mathcal{X} = \mathcal{V} \oplus \mathcal{W}.\]

In this case \( \dim \mathcal{X} = \dim \mathcal{V} + \dim \mathcal{W} \). For two companion subspaces \( \mathcal{V} \) and \( \mathcal{W} \) any vector \( x \in \mathcal{X} \) can be written uniquely as

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

The unique component \( v \) is called the projection of \( x \) onto \( \mathcal{V} \) along its companion space \( \mathcal{W} \). Similarly, the unique component \( w \) is called the projection of \( x \) onto \( \mathcal{W} \) along its companion space \( \mathcal{V} \).

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

\[P_{\mathcal{V}|\mathcal{W}} x \triangleq v \quad \text{and} \quad P_{\mathcal{W}|\mathcal{V}} x = w.
\]

Obviously,

\[P_{\mathcal{V}|\mathcal{W}} + P_{\mathcal{W}|\mathcal{V}} = I.
\]

Furthermore it is straightforward to show that \( P_{\mathcal{V}|\mathcal{W}} \) and \( P_{\mathcal{W}|\mathcal{V}} \) are idempotent,

\[P_{\mathcal{V}|\mathcal{W}}^2 = P_{\mathcal{V}|\mathcal{W}} \quad \text{and} \quad P_{\mathcal{W}|\mathcal{V}}^2 = P_{\mathcal{W}|\mathcal{V}}.
\]

In addition, it can be shown that the projection operators \( P_{\mathcal{V}|\mathcal{W}} \) and \( P_{\mathcal{W}|\mathcal{V}} \) are linear operators.

Linear Operators and Matrices. Consider a function \( A \) which maps between two vector spaces \( \mathcal{X} \) and \( \mathcal{Y} \), \( A : \mathcal{X} \to \mathcal{Y} \). The “input space” \( \mathcal{X} \) is called the domain while the “output space” is called the codomain. The mapping or operator \( A \) is said to be linear when,

\[A (\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 A x_1 + \alpha_2 A x_2 \quad \forall x_1, x_2 \in \mathcal{X}, \forall \alpha_1, \alpha_2 \in \mathcal{F}.
\]

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 \( \mathcal{F} \).\(^{9}\)

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

\(^{9}\)That is, \( \mathcal{X} \) and \( \mathcal{Y} \) must be both real vectors spaces, or must be both complex vector spaces.
an $m \times n$ matrix representation over the field $\mathcal{F}$. Henceforth we assume that any such linear mapping $A$ is an $m \times n$ matrix and we write $A \in \mathcal{F}^{m \times n}$.\[^{10}\] Note in particular that projection operators on finite dimensional vector spaces must have matrix representations.

Every linear operator has two natural vector subspaces associated with it. The range space,
\[ \mathcal{R}(A) = \{ y \mid y = Ax, \ x \in \mathcal{X} \} \subset \mathcal{Y}, \]
and the nullspace,
\[ \mathcal{N}(A) = \{ x \mid Ax = 0 \} \in \mathcal{X}. \]
The dimension of the range space is called the rank of $A$,
\[ r(A) = \text{rank}(A) = \dim \mathcal{R}(A), \]
while the dimension of the nullspace is called the nullity of $A$,
\[ \nu(A) = \text{nullity}(A) = \dim \mathcal{N}(A), \]
When solving a system of linear equations $y = Ax$ a solution exists when $y \in \mathcal{R}(A)$, in which case we say that the problem is consistent.

**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, $y = Ax$, is called the forward problem. Note that the forward problem is always well-posed in that knowing $A$ and given $x$ one can construct $y$ by straightforward matrix-vector multiplication.

Given a vector $y$ in the codomain, the problem of determining an $x$ in the domain for which $Ax = y$ is known an the inverse problem. The inverse problem is said to be well-posed if and only if the following three conditions are true for the $m \times n$ 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, $\mathcal{R}(A) = \mathcal{Y}$.\[^{11}\] Equivalently, $r(A) = m$. If this condition is met, we say that the system $y = Ax$ is consistent, else the system is inconsistent.

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

3. The solution $x$ does not depend sensitively on the value of $y$.\[^{12}\] I.e., we demand that $A$ be numerically well-conditioned.

\[^{10}\]Thus if $\mathcal{X}$ and $\mathcal{Y}$ are both complex we write $A \in \mathbb{C}^{m \times n}$ while if they are both real we write $A \in \mathbb{R}^{m \times n}$.

\[^{11}\]It is not enough to merely require consistency for a given $y$ because even the tiniest error or miss-specification in $y$ can render the problem inconsistent.

\[^{12}\]So that a slight error or miss-specification in $y$ does not cause a huge error in the solution $x$. If such a sensitive dependency exists, we say that $A$ is numerically ill-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 assume that the inverse problem is numerically well–conditioned and focuss 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, Banach 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 $x, y \in \mathcal{X}$.

The existence of a norm gives us a measure of size, $\text{size}(x) = \| x \|$, a measure of distance, $d(x, y) = \| x - y \|$, and a well–defined $\epsilon$–ball or neighborhood of a point $x$, $N_\epsilon(x) = \{ y \mid \| y - x \| \leq \epsilon \}$. As discussed in Section 5.1 of Meyer, there are innumerable norms that one can define on a given vector space, the most commonly used ones being the 1–norm, the 2–norm, and the $\infty$–norm. In this course we focuss on the *weighted 2-norm*, $\| x \| = \sqrt{x^H \Omega x}$, where the *weighting matrix* $\Omega$ is hermitian and positive–definite.

A *Banach Space* is a *complete* normed linear vector space. Completeness is a technical condition which is the requirement that every so–called Cauchy sequence is a convergent sequence. As this property 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 considered in this course are perforce Banach Spaces.

**Inner Product Space, Hilbert Space.** It is of great convenience, both conceptually and computationally, to work in a space which has well–defined concept of orthogonality, and of “angle” in general. For this reason, when possible, one attempts to define an *inner product* on the vector space of interest and then work with the associated norm induced by this inner product. As discussed on page 286 of Meyer, given a vector space $\mathcal{X}$ over the field of scalars $\mathcal{F}$, 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.**\(^{13}\)

2. **Real positive–definiteness of** \(\langle x, x \rangle\) **for all** \(x \in \mathcal{X}\).

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. 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, so the distinction between a general inner product space and a Hilbert space will not be an issue in this course.

On 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 matrix* \(\Omega\) is hermitian and positive–definite. Note that the corresponding induced norm is the weighted 2–norm mentioned above, \(\|x\| = \sqrt{x^H \Omega x}\). When \(\Omega = I\) we call the resulting inner product and induced norm the *standard inner–product* and the *standard 2–norm* respectively.

Given an inner product, one then can define orthogonality of two vectors \(x \perp y\) by the requirement that \(\langle x, y \rangle = 0\). Given two orthogonal vectors, it is straightforward to show that the (generalized) *Pythagorean theorem* holds,

\[
x \perp y \iff \|x + y\|^2 = \|x\|^2 + \|y\|^2.
\]

Another important result is the *Cauchy–Schwarz (C–S) inequality,*

\[
|\langle x, y \rangle| \leq \|x\| \|y\| \quad \text{for all} \ x, y \in \mathcal{X},
\]
with equality holding if and only if \(y = \alpha x\) for some scalar \(\alpha\). As a consequence of the C–S inequality, one can define the angle \(\theta\) between two vectors in a Hilbert space by,

\[
\cos \theta = \frac{|\langle x, y \rangle|}{\|x\| \|y\|}.
\]

\(^{13}\)Although Meyer, and the class lecture, assumes linearity in the *second* argument, this choice is arbitrary and many books take the inner product to alternatively be linear in the first argument. This distinction is only meaningful when \(\mathcal{F} = \mathbb{C}\) and does not occur when \(\mathcal{F} = \mathbb{R}\), in which case the inner product is linear in both arguments. When the vector space is complex, most mathematicians (but not Meyer) and engineers (but not this class) tend to define linearity in the first argument, while most physicists and controls engineers tend to define linearity in the second argument. Serious confusion can occur if you do not take care to determine which definition is the case.
Note that as a consequence of the C–S inequality we have $0 \leq \cos \theta \leq 1$.

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 $\mathcal{V}$ are disjoint $\mathcal{W}$, $\mathcal{V} \cap \mathcal{W} = \{0\}$. Given a subspace $\mathcal{V}$ of $\mathcal{X}$, one defines the orthogonal complement $\mathcal{V}^\perp$ to be all vectors in $\mathcal{X}$ which are perpendicular to $\mathcal{V}$. The orthogonal complement $\mathcal{V}^\perp$ is unique and is a subspace in its own right for which

$$\mathcal{X} = \mathcal{V} \oplus \mathcal{V}^\perp.$$

Thus a subspace $\mathcal{V}$ and its orthogonal complement $\mathcal{V}^\perp$ are complementary subspaces.\(^\text{14}\) Note that it must therefore be the case that

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

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 $P_\mathcal{V} = P_{\mathcal{V}^\perp \mathcal{V}}$. 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 more simply denotes the orthogonal projection operator by $P = P_\mathcal{V}$. Of course, as is the case for all projection operators, an orthogonal projection operator is idempotent.

Recall our discussion above on the range and null spaces of a linear mapping $A : \mathcal{X} \to \mathcal{Y}$. If $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces, we must have that

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

and

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

The 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 $A$.

**Projection Theorem, Orthogonality Principle.** Suppose we are given a vector $x$ in a Hilbert space $\mathcal{X}$ and are asked to find the best approximation, $v$, to $x$ in a subspace $\mathcal{V}$ in the sense that the norm of the error $e = x - v$, $\|e\| = \|x - v\|$, is to be minimized over all possible vectors $v \in \mathcal{V}$. We will call the resulting optimal vector $v$ the least–squares estimate of $x$ in $\mathcal{V}$, because it is understood that 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

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

\(^\text{14}\)Thus $\mathcal{V}^\perp$ is more than a complementary subspace to $\mathcal{V}$; it is the orthogonally complementary subspace to $\mathcal{V}$. 
must be orthogonal to \( \mathcal{V} \). We have that,

\[
\|e\|^2 = \|x - v\|^2 = \|(x - v_0) + (v_0 - v)\|^2 = \|x - v_0\|^2 + \|v_0 - v\|^2 \geq \|x - v_0\|^2,
\]
as a consequence of the Pythagorean theorem.\(^\text{15}\) Thus the error is a minimum when \( v = v_0 \). 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 known as the Orthogonality Principle, \( \langle x - v_0, v \rangle = 0 \) for all \( v \in \mathcal{V} \).

We can obtain a generalized solution to an ill–posed inverse problem \( Ax = y \) by looking for the unique solution to the regularized problem,

\[
\min_x \|y - Ax\|^2 + \beta\|x\|^2, \quad \beta > 0.
\]

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. When the problem is posed in a Hilbert space (so that the norm induced by the inner product is used) the limiting solution,

\[
\hat{x} = \lim_{\beta \to 0} \hat{x}_\beta,
\]
is called the minimum norm least–squares solution. It is also known as the pseudoinverse solution and the operator \( A^+ \) which maps \( y \) to this solution, \( \hat{x} = A^+y \) is called the pseudoinverse of \( A \).

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

\[
\hat{x} = \arg\min_{x'} \|x'\| \quad x' \in \arg\min_x \|y - Ax\|^2 \bigg\}.
\]

Because \( Ax \in \mathcal{R}(A) \) we see that any particular least–squares solution, \( x' \), to the inverse problem \( y = Ax \) provides a value \( x' \) such that \( \hat{y} = Ax' \) is the optimal least–squares estimate of \( y \) in \( \mathcal{R}(A) \subset \mathcal{Y} \), \( \hat{y} = P_{\mathcal{R}(A)}y \). As discussed above, the Orthogonality Condition determines a least–squares solution from the geometric condition

\[
\text{Geometric Condition for a Least–Squares Solution: } e = y - Ax' \perp \mathcal{R}(A) \tag{1}
\]

which can be equivalently written as \( e = y - Ax' \in \mathcal{R}(A)^\perp \).

One can write any particular least–squares solution, \( x' \), as

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

\(^{15}\)Note that the vector \( v - v_0 \) must be in the subspace \( \mathcal{V} \).
where \( \hat{x} = P_{\mathcal{N}(A)^\perp}x' \in \mathcal{N}(A)^\perp \) and the null space component \( x'_{\text{null}} = P_{\mathcal{N}(A)}x' \in \mathcal{N}(A) \) is a homogeneous solution \( Ax'_{\text{null}} = 0 \). Note that

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

The minimum norm least–squares solution \( \hat{x} \) is the unique least–squares solution in the subspace \( \mathcal{N}(A)^\perp \). 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 \( x' \perp \mathcal{N}(A) \), or equivalently that

\[ \text{Geometric Condition for a Minimum Norm LS Solution: } x' \in \mathcal{N}(A)^\perp \] (2)

We want to move from the insightful geometric conditions to algebraic expressions which allow us to analytically solve for the minimum norm least–squares solution \( \hat{x} \). To accomplish this we need to introduce the concept of the adjoint operator \( A^* \).

**Adjoint Operator, Four Fundamental Subspaces of a Linear Operator.** Given a linear operator \( A : \mathcal{X} \to \mathcal{Y} \) which maps between two finite dimensional Hilbert spaces, its adjoint operator \( A^* : \mathcal{Y} \to \mathcal{X} \) is defined by the condition,

\[ \langle y, Ax \rangle = \langle A^* y, x \rangle \quad \text{for all } x \in \mathcal{X}, \ y \in \mathcal{Y}. \]

If \( \mathcal{X} \) has a weighted inner–product with weighting matrix \( \Omega \) and \( \mathcal{Y} \) has a weighted inner–product with weighting matrix \( W \), it was shown in class lecture that the adjoint operator can be determined to be unique and given by,

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

The adjoint \( A^* \) is a “companion” linear operator associated with the linear operator \( A \). Note that if the standard inner product is used on both the domain and codomain we have \( A^* = A^H \) and if furthermore the Hilbert spaces are real we have \( A^* = A^T \). Thus the deeper meaning of the transpose of a real matrix \( A \) is that it forms the adjoint (“companion”) operator to \( A \) when \( A \) is viewed as a linear mapping between two real Hilbert spaces each having the standard inner product.

It was shown in class that the importance of the adjoint operator is that it gives a description of \( \mathcal{R}(A)^\perp \) and \( \mathcal{N}(A)^\perp \) which is symmetrical to that of \( \mathcal{N}(A) \) and \( \mathcal{R}(A) \),

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

It can be shown that \( A^{**} = A \) showing that the four fundamental subspaces of \( A^* \) are identical to those of \( A \), and thus the geometric relationships between \( A \) and \( A^* \) are entirely...
symmetrical. We can view the operator $A$ and its adjoint $A^*$ (or, equivalently, the operator $A^*$ and its adjoint $A^{**} = A$) as being companion matrices which have associated with them the four fundamental subspaces $\mathcal{R}(A)$, $\mathcal{R}(A^*)$, $\mathcal{N}(A)$, and $\mathcal{N}(A^*)$ which are related as,

$$A : \mathcal{X} \to \mathcal{Y}, \quad A^* : \mathcal{Y} \to \mathcal{X},$$

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

$$\mathcal{X} = \mathcal{R}(A^*) \oplus \mathcal{N}(A), \quad \mathcal{R}(A^*) = \mathcal{N}(A)^\perp,$$

$$\mathcal{Y} = \mathcal{R}(A) \oplus \mathcal{N}(A^*), \quad \mathcal{R}(A) = \mathcal{N}(A^*)^\perp.$$}

Again, note that the relationship between $A$ and $A^*$, and their associated four fundamental subspaces, is entirely symmetrical.

In class it was shown that if $P$ is an orthogonal projection operator, then it must be self–adjoint,

$$P = P^*.$$  

An operator is an orthogonal projection operator if and only if it is both idempotent and self–adjoint.

**The Minimum Norm Least–Squares Solution and the Pseudoinverse.** Combining the geometric conditions (1) and (2) for optimality of a minimum least–squares solution with the characterization given above of the four fundamental subspaces provided by an operator $A$ and its adjoint $A^*$ we determine the conditions for optimality to be

$$e = y - Ax \in \mathcal{N}(A^*) \quad \text{and} \quad x \in \mathcal{R}(A^*).$$

The first optimality condition is equivalent to $0 = A^* e = A^* (y - Ax)$ yielding the so–called normal equations,

$$A^* Ax = A^* y. \quad (3)$$

The normal equations are a consistent set of equations, every solution of which constitutes a least–squares solution to the inverse problem $y = Ax$. The second optimality condition is equivalent to,

$$x = A^* \lambda, \quad \lambda \in \mathcal{Y}. \quad (4)$$

The vector $\lambda$ is a nuisance parameter (which can be interpreted as a Lagrange multiplier) which is usually solved as an intermediate solution on the way to determining the optimal minimum norm least–square solution $\hat{x}$. The two sets of equations (3) and (4) can be solved in concert to determine both $\lambda$ and the optimal solution $\hat{x}$.

In principle, solving for the pseudoinverse solution $\hat{x}$ corresponds to determining the pseudoinverse operator $A^+$ which yields the solution as $\hat{x} = A^+ y$. The pseudoinverse $A^+$ is unique and linear. Although it is not always numerically advisable to determine the pseudoinverse $A^+$ itself, special–case closed form expressions for $A^+$ and general numerical procedures for constructing $A^+$ do exist. In Matlab the command `pinverse(A)` numerically
constructs the pseudoinverse matrix for \( A \) assuming that the standard inner product holds on the domain and codomain of \( A \). As shown in class, closed form expressions exist for \( A^+ \) for the following two special cases:

\[
\text{rank}(A) = n : \quad A^+ = (A^*A)^{-1}A^*,
\]
\[
\text{rank}(A) = m : \quad A^+ = A^*(AA^*)^{-1}.
\]

In both these cases if \( A \) is square one obtains \( A^+ = A^{-1} \). Generally it is not numerically sound to construct \( A^+ \) using these closed form expressions, although they are quite useful for analysis purposes.

From knowledge of the pseudoinverse \( A^+ \), one obtains the orthogonal projection operators onto the subspaces \( \mathcal{R}(A) \) and \( \mathcal{R}(A^*) \) respectively as

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

As a consequence, we also have that the orthogonal projection operators onto the subspaces \( \mathcal{N}(A^*) \) and \( \mathcal{N}(A) \) are respectively given by

\[
P_{\mathcal{N}(A^*)} = I - AA^+ \quad \text{and} \quad P_{\mathcal{N}(A)} = I - A^+A.
\]

One way to identify a candidate pseudoinverse \( A^+ \) is that a matrix \( M \) is the pseudoinverse for \( A \) if and only if it satisfies the four Moore–Penrose (M–P) pseudoinverse conditions:

\[
(MA)^* = MA, \quad (AM)^* = AM, \quad AMA = A, \quad MAM = M.
\]

For instance, using these conditions one can ascertain that the pseudoinverse of the 1 \( \times \) 1 (scalar) matrix \( A = 0 \) has pseudoinverse \( A^+ = 0^+ = 0 \). As discussed in class, an alternative way to obtain the pseudoinverse is to obtain the Singular Value Decomposition (SVD) of \( A \), which in Matlab is obtained by the simple function call \texttt{svd(A)}. 