Finite Dimensional Hilbert Spaces and Linear Inverse Problems

ECE 174 – Lecture Supplement – Spring 2009

Ken Kreutz-Delgado
Electrical and Computer Engineering
Jacobs School of Engineering
University of California, San Diego

Report Number ECE174LSHS-S2009V1.0

November 12, 2013

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.[1]

The properties of *scalar multiplication,* αx, of a vector $x \in \mathcal{X}$ by a scalar $\alpha \in \mathcal{F}$ are also given on page 160 of the textbook. The scalars used in this course are either the field of real numbers, $\mathcal{F} = \mathbb{R}$, or the field of complex numbers, $\mathcal{F} = \mathbb{C}$.[2] Unless stated otherwise, in the remainder of this memo we assume that the scalars are from the field of complex numbers.

In this course we essentially consider only *finite dimensional* vector spaces $\dim \mathcal{X} = n < \infty$ and give results appropriate for this restriction.[3] Any vector x in an n-dimensional vector

[1] *Matrix Analysis and Applied Linear Algebra,* C.D. Meyer, SIAM, 2000.

[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 advanced textbooks, including *Linear Operator Theory in Engineering and Science,* A.W. Naylor and G.R. Sell,

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,

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

We refer to this as a *canonical representation* of a finite-dimensional vector. We often assume that vectors in an n-dimensional vector space are canonically represented by $n \times 1$ column vectors. For example, the space, $\mathbb{R}^{m\times n}$ of real $m \times n$ matrices forms an mn-dimensional vector space. An element $A \in \mathbb{R}^{m\times n}$ can be given a canonical representation as an $mn \times 1$ column vector by "stacking" its columns, $\text{stack}(A) \in \mathbb{R}^{mn}$. As this procedure can obviously be undone, we have $\mathbb{R}^{mn} \approx \mathbb{R}^{m\times n}$.

The complex conjugate of a scalar α is denoted by $\bar{\alpha}$. The complex conjugate of a matrix A (respectively a vector x) is denoted by $\bar{A}$ (respectively $\bar{x}$) 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) $A^H = (\bar{A})^T = \overline{(A)^T}$ ($x^H = (\bar{x})^T = \overline{(x)^T}$). A matrix is said to be *hermitian* if $A^H = A$.[4]

Linear Vector Subspace. 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. 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 only needs to check that every linear combination of vectors in $\mathcal{V}$ yields a vector in $\mathcal{V}$.[5] This latter property is called the property of *closure of the subspace* $\mathcal{V}$ 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 subsets $\mathcal{V}$ and $\mathcal{W}$ of vectors, we define their *sum* by[6]

$$\mathcal{V} + \mathcal{W} = \{v + w \;|v \in \mathcal{V} \text{ and } w \in \mathcal{W}\} \;.$$

If the sets $\mathcal{V}$ and $\mathcal{W}$ are in addition both *subspaces* of $\mathcal{X}$, then $\mathcal{V} \cap \mathcal{W}$ and $\mathcal{V} + \mathcal{W}$ are also subspaces of $\mathcal{X}$, and $\mathcal{V} \cup \mathcal{W} \subset \mathcal{V} + \mathcal{W}$.[7] In general,

$$\{0\} \subset \mathcal{V} \cap \mathcal{W} \subset \mathcal{V} + \mathcal{W} \subset \mathcal{X} \,,$$

Springer-Verlag, 1982, which is available in a relatively inexpensive paperback edition. Some of the necessary extensions are briefly mentioned as we proceed through our development.

[4]Note that a real hermitian matrix is symmetric, $A^T = A$.

[5]Two important special cases of this test are that if the zero element does not belong to the subset or if multiplication of a vector by the scalar -1 violates closure then the subset is obviously not a subset.

[6]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.

[7]In general $\mathcal{V} \cup \mathcal{W}$ is *not* a subspace.

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 $x_1, \cdots, x_r \in \mathcal{X}$ are *linearly independent* when,

$$\alpha_1 x_1 + \cdots \alpha_r x_r = 0 \;\text{ if and only if }\; \alpha_1 = \cdots \alpha_r = 0\,.$$

Note that this definition can be written in matrix-vector form as,[8]

$$X\alpha = \begin{pmatrix} x_1 & \cdots & x_r \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_r \end{pmatrix} = 0 \quad \Longleftrightarrow \quad \alpha = 0\,.$$

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,

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

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 $\mathcal{V}$.[9] 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 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.*

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

$$0 \leq \dim(\mathcal{V} \cap \mathcal{W}) \leq \dim(\mathcal{V} + \mathcal{W}) \leq \dim \mathcal{X}\,.$$

[8] Assuming that $x_i \in \mathcal{F}^n$, the resulting matrix $X = (x_1 \cdots x_r)$ is $n \times r$.

[9] Because of the assumption that we are working with finite dimensional spaces, we can ignore the distinction between a Hamel basis and other types of of bases. The interested reader can find an excellent discussion of this matter in Sections 4.6, 4.7, and 5.17 of the textbook by Naylor and Sell, *op. cit.*

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.[10]

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,*

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

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

[10] E.g., see the textbook by Naylor and Sell, *op. cit.*

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 $\mathcal{F}$.[11]

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 in particular 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 $A \in \mathcal{F}^{m \times n}$.[12]

Every linear operator has two natural vector subspaces associated with it. The *range space,*

$$\mathcal{R}(A) \triangleq A(\mathcal{X}) \triangleq \{y \,|\, y = Ax,\ x \in \mathcal{X}\} \subset \mathcal{Y}\,,$$

and the *nullspace,*

$$\mathcal{N}(A) = \{x \,|Ax = 0\} \in \mathcal{X}\,.$$

Given our assumption that all vector spaces are finite dimensional, both $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are closed subspaces.[13]

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.*

[11] For example, $\mathcal{X}$ and $\mathcal{Y}$ must be both real vectors spaces, or must be both complex vector spaces.

[12] 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 $A \in \mathbb{R}^{m \times n}$.

[13] The definition of a closed set is given in Section 2.1.1 of Moon and Stirling, *op. cit.* Conditions for closure of the range and null spaces of a linear operator can be found in Naylor and Sell, *op. cit.* If we take the operator A to be *bounded* when the vector spaces are infinite dimension, a property which is always satisfied in the finite dimensional situation, then $\mathcal{N}(A)$ is closed, while $\mathcal{R}(A)$ in general is not closed. However, every finite dimensional subspace is closed, so that if $\mathcal{R}(A)$ is finite dimensional, then it is closed. Sufficient conditions for $\mathcal{R}(A) = A(\mathcal{X}) \subset \mathcal{Y}$ to be finite dimensional are that either $\mathcal{X}$ or $\mathcal{Y}$ be finite dimensional. Thus the results in this note are also valid for the case when A is *bounded* and *either* $\mathcal{X}$ or $\mathcal{Y}$ are finite dimensional. A sufficient condition for this to be true is that $\mathcal{X}$ and $\mathcal{Y}$ are *both* finite dimensional; which is precisely the condition assumed at the outset of this note.

Linear Forward and Inverse Problem. Given a linear mapping between two vector spaces $A : \mathcal{X} \to \mathcal{Y}$ 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 am m-dimensional vector y in the codomain, the problem of determining an n-dimensional vector x in the domain for which $Ax = y$ is known as an *inverse problem. The inverse problem 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,* $\mathcal{R}(A) = \mathcal{Y}$.[14] Equivalently, $r(A) = m$. If this condition is met, we say that the system $y = Ax$ is *robustly consistent,* else the system is *generically 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.[15] 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, 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 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 ϵ-ball or neighborhood of a point x, $N_\epsilon(x) = \{y \,|\, \|y - x\| \leq \epsilon\}$. There are innumerable norms that one can define on a given vector space,

[14] It it not enough to merely require consistency for a given y because even the tiniest error or misspecification in y can render the problem inconsistent.

[15] 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*

the most commonly used ones being the *1-norm, the 2-norm, and the ∞-norm.* In this course we focuss on the *weighted 2-norm,* $\|x\| = \sqrt{x^H \Omega x}$, where the *weighting matrix* Ω 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 convergent 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. Given a vector space $\mathcal{X}$ over the field of scalars $\mathcal{F}$, an inner product is an $\mathcal{F}$-valued binary operator on $\mathcal{X} \times \mathcal{X}$,

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

The inner product has the following three properties:

1. *Linearity in the second argument.*[16]
2. *Real positive-definiteness of* $\langle x, x \rangle$ for all $x \in \mathcal{X}$.[17]
3. *Conjugate-symmetry,* $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for all $x, y \in \mathcal{X}$.[18]

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 one uses in an inner product space. If the resulting normed

[16] Although the developments in the Meyer textbook, this supplement, and the class lecture assume 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 is irrelevant when $\mathcal{F} = \mathbb{R}$, in which case the inner product is linear in both arguments (i.e., in the *real* case the inner product is *bilinear*). When the vector space is complex, most mathematicians (but not all, including Meyer) and many 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.

[17] I.e., $0 \leq \langle x, x \rangle \in \mathbb{R}$ for any vector x, and $0 = \langle x, x \rangle$ iff $x = 0$.

[18] which for a real vector space reduces to the property of symmetry, $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in \mathcal{X}$.

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 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 matrix*[19] Ω 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* (or *Cartesian*) *inner-product* and the *standard* (or *Cartesian*) *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 induced norm satisfies the (generalized) *Pythagorean Theorem*

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

More generally, an inner-product induced norm satisfies the *Parallelogram Law*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

If the vectors x and y are orthogonal, $\langle x, y \rangle = 0$, then $\|x + y\|^2 = \|x - y\|^2$ and the Parallelogram Law reduces to the Pythagorean Theorem. Note that the Parallelogram Law implies the useful inner-product induced norm inequality

$$\|x + y\|^2 \leq 2 \left(\|x\|^2 + \|y\|^2 \right).$$

This inequality and the Parallelogram Law do *not* hold for the norm on a general (non-inner product) normed vector space.[20]

Another important result is that the inner-product induced norm satisfies 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 α. As a consequence of the C-S inequality, one can meaningfully define the angle θ between two vectors in a Hilbert space by

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

[19] Also known as the *metric tensor* in physics.

[20] It is a remarkable fact that if a norm satisfies the Parallelogram Law, then the norm *must* be a norm induced by a unique inner product and that inner product can be inferred from the norm. See the Wikipedia article "Parallelogram Law."

since as a consequence of the C-S inequality we must have $0 \leq \cos\theta \leq 1$. Note that this definition is consistent with our earlier definition of orthogonality between two vectors.

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 (in the finite dimensional case assumed here) obeys the property $\mathcal{V}^{\perp\perp} = \mathcal{V}$. The orthogonal complement $\mathcal{V}^\perp$ is *unique* and is a subspace in its own right for which[21]

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

Thus a subspace $\mathcal{V}$ and its orthogonal complement $\mathcal{V}^\perp$ are complementary subspaces.[22] 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 $P_\mathcal{V} = P_{\mathcal{V}|\mathcal{V}^\perp}$. 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 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 finite-dimensional Hilbert spaces, we must have that[23]

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

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 call the resulting optimal vector v the *least-squares estimate* of

[21] Some modification of the subsequent discussion is required in the case of infinite-dimensional Hilbert spaces. See Naylor and Sell, *op. cit.* The modifications required here are $\mathcal{V}^{\perp\perp} = \overline{\mathcal{V}}$ and $\mathcal{X} = \overline{\mathcal{V}} \oplus \mathcal{V}^\perp$ where $\overline{\mathcal{V}}$ denotes the *closure* of the subspace $\mathcal{V}$.

[22] Thus $\mathcal{V}^\perp$ is *more* than a complementary subspace to $\mathcal{V}$; it is *the orthogonally* complementary subspace to $\mathcal{V}$.

[23] In this case, assuming the boundedness of A, the required infinite dimensional modification is $\mathcal{Y} = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^\perp$. The second expression is unchanged.

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

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

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 a consequence of the Pythagorean theorem.[24] *Thus the error is minimized 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,

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

for appropriately chosen norms on the domain and codomain. The solution to this problem, $\hat{x}_\beta$, is a function of the *regularization parameter* β. The choice of the precise value of the regularization parameter β is often a nontrivial problem. When the problem is posed as a mapping between Hilbert spaces (so that induced norms are used on the domain and codomain) the limiting solution,

$$\hat{x} = \lim_{\beta \to 0} \hat{x}_\beta \,,$$

is called the *minimum norm least-squares* solution.[25] 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 A^+ is itself a linear operator. Note that in the special case when A is square and full-rank, it must be the case that $A^+ = A^{-1}$ showing that the pseudoinverse is a generalized inverse.

The pseudoinverse solution, $\hat{x}$, 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 $\|e\|^2 = \|y - Ax\|^2$,

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

[24] Note that the vector $v - v_0$ must be in the subspace $\mathcal{V}$.

[25] Often this is called the *weighted-minimum norm, weighted-least-squares solution* as we generally assume weighted inner products (and hence weighted induced norms) on the domain and codomain spaces. This is because the standard terminology is to take "minimum norm least-squares" solution to mean that only standard (Cartesian or un-weighted) inner products hold on the domain and codomain. As we admit the richer class of weighted norms, our terminology has a richer meaning.

Because $Ax \in \mathcal{R}(A)$ we see that *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 $\mathcal{R}(A) \subset \mathcal{Y}$, $\hat{y} = P_{\mathcal{R}(A)}y$. As discussed above, the Orthogonality Condition determines a least-squares solution $\hat{y} = Ax'$ from the geometric condition

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

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

One can write any *particular* least-squares solution, $x' \in \arg\min_x \|y - Ax\|^2$, $\hat{y} = Ax'$, as

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

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

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

$\hat{x}$ is unique, i.e., independent of the particular choice of least-squares solution x'. least-squares solution $\hat{x} \in \mathcal{N}(A)^\perp$ 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 unique least squares solution $\hat{x}$ 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

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

The *primary geometric condition* (1) ensures that x' is a least-squares solution. The *additional* requirement that x' satisfy the *secondary geometric condition* (2) then ensures that x' is the unique minimum norm least-squares solution. We now want to make the move from the insightful geometric conditions (1) and (2) to *equivalent 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 Ω and $\mathcal{Y}$ has a weighted inner-product with weighting matrix W, it is shown in class lecture that the adjoint operator can be determined to be unique and is given by the linear operator

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

The adjoint A^* is a "companion" operator associated with the operator A.[26] 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 is discussed in class lecture that the adjoint operator gives a description of $\mathcal{R}(A)^\perp$ and $\mathcal{N}(A)^\perp$ which is symmetrical to that of $\mathcal{N}(A)$ and $\mathcal{R}(A)$,[27]

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

It can be shown that $A^{**} = A$ showing that the four fundamental subspaces of A^* are (with an appropriate renaming) 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

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

Again, note that the relationships between A and A^*, and their associated four fundamental subspaces, are entirely symmetrical.[28]

In class we discuss the fact that if P is an *orthogonal* projection operator then, in addition to being idempotent, it must also be *self-adjoint,*

$$P = P^* \,.$$

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

[26] In the infinite dimensional case, the adjoint of a bounded linear operator is itself a bounded linear operator.

[27] In the infinite dimensional case, taking A to be bounded, the second expression is modified to $\mathcal{N}(A)^\perp = \overline{\mathcal{R}(A^*)}$.

[28] When generalized to the general infinite-dimensional case, but still assuming that A is a bounded linear operator, these relations must be modified by replacing the range spaces by their closures. However, as discussed in an earlier footnote, as long as A is bounded *and* either $\mathcal{X}$ or $\mathcal{Y}$ is finite dimensional then the shown relationships remain true. Every linear operator A encountered in the homework and exam problems given in this course will satisfy this latter condition.

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,*

$$A^*Ax = A^*y\,. \tag{3}$$

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,

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

The vector λ is a *nuisance parameter* (which can be interpreted as a *Lagrange multiplier*) which is usually solved for 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 simultaneously to determine both λ 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 `pinv(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 the pseudoinverse A^+ is that a candidate matrix M is the unique 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 1×1 (scalar) matrix $A = 0$ has pseudoinverse $A^+ = 0^+ = 0$. As discussed below, an alternative way to obtain the pseudoinverse when standard (unweighted) inner products are assumed on the domain and codomain is to obtain the *Singular Value Decomposition* (SVD) of A, which in Matlab is obtained by the simple function call `svd(A)`.

Completing the Square. An alternative, non-geometric way to determine a weighted least-squares solution to a full column-rank linear inverse problem is by *completing the square.* Consider the quadratic form

$$\ell(x) = x^H \Pi x - 2\mathrm{Re}\, x^H B y + y^H W y \tag{5}$$

where Π is an $n \times n$ hermitian, positive-definite matrix, $\Pi = \Pi^H > 0$, W is an $m \times m$ hermitian, positive-definite matrix and B is $n \times m$. With the positive-definite assumptions on Π and W, the quadratic form $\ell(x)$ is a real-valued function of the complex vector $x \in \mathbb{C}^n$.

With some minor algebra, one can rewrite the quadratic form as

$$\ell(x) = \left(x - \Pi^{-1}By\right)^H \Pi \left(x - \Pi^{-1}By\right) + y^H \left(W - B^H\Pi^{-1}B\right) y. \tag{6}$$

Thus for all x we have that

$$\ell(x) \geq y^H \left(W - B^H\Pi^{-1}B\right) y$$

with equality if and only if $x = \Pi^{-1}By$. Thus we have proved that

$$\hat{x} = \Pi^{-1}By = \arg\min_x \ell(x). \tag{7}$$

It is straightforward to apply this result to the full column-rank, weighted least-squares problem. Noting that

$$\begin{aligned}
\ell(x) &= \|y - Ax\|_W^2 = (y - Ax)^H W (y - Ax) \\
&= x^H A^H W A x - x^H A^H W y - y^H W A x + y^H W y \\
&= x^H \Pi x - x^H B y - y^H B^H x + y^H W y \\
&\quad \text{(Setting } \Pi \triangleq A^H W A \text{ and } B \triangleq A^H W\text{)} \\
&= x^H \Pi x - 2\mathrm{Re}\, x^H B y + y^H W y \\
&= \left(x - \Pi^{-1}By\right)^H \Pi \left(x - \Pi^{-1}By\right) + y^H \left(W - B^H\Pi^{-1}B\right) y
\end{aligned}$$

it is evident that the weighted least-squares estimate of x is given by

$$\hat{x} = \Pi^{-1}By = \left(A^H W A\right)^{-1} A^H W y.$$

Use of Derivatives. A third method of finding the least-squares solution to a full column-rank linear inverse problem is provided by taking the derivative of the the weighted least-squares loss function and setting the resulting expression to zero. Solving the resulting equation then yields the weighted least-squares solution, which is unique under the full column-rank assumption. This method is discussed in the course lectures and in a lecture Supplement.

The Singular Value Decomposition (SVD). Given a complex $m \times n$ matrix A, we can view it as a linear mapping between the unweighted finite dimensional Hilbert spaces $\mathcal{X} = \mathbb{C}^n$ and $\mathcal{Y} = \mathbb{C}^m$, $A : \mathcal{X} \to \mathcal{Y}$. Note that with the use of the unweighted (standard) inner products the adjoint of A is given by its hermitian transpose, $A^* = A^H$.

The rank of the matrix A is denoted by $r = \mathrm{r}(A)$, which is the dimension of the range of A, $\dim(\mathcal{R}(A)) = r$ and the dimension of the range of the adjoint of A, $\dim\left(\mathcal{R}(A^H)\right) = r$. The nullity $\nu = n - r$ of the matrix A denotes the dimension of the nullspace of A, $\dim(\mathcal{N}(A)) = \nu$, and $\mu = m - r$ denotes the dimension of the null space of the adjoint of A, $\dim\left(\mathcal{N}(A^H)\right) = \mu$ (this latter nullspace is also called the left-nullspace of A.)

The *Singular Value Decomposition* (SVD) of A is given by the factorization[29]

$$A = U\Sigma V^H = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} = U_1 S V_1^H = \sum_{i=1}^{r} \sigma_i u_i v_i^H \tag{8}$$

where U is $m \times m$, V is $n \times n$, Σ is $m \times n$, S is $r \times r$, U_1 is $m \times r$, U_2 is $m \times \mu$, V_1 is $n \times r$, V_2 is $n \times \nu$, u_i denotes a column of U, and v_i denotes a column of V. There are precisely r (nonzero) real singular values of A rank ordered as

$$\sigma_1 \geq \cdots \geq \sigma_r > 0\,,$$

and

$$S = \mathrm{diag}\,(\sigma_1 \cdots \sigma_r)\ .$$

The columns of U form an orthonormal basis for $\mathcal{Y} = \mathbb{C}^m$, $\langle u_i, u_j \rangle = u_i^H u_j = \delta_{i,j}$, $i, j = 1, \cdots m$. Equivalently, U is a unitary matrix, $U^H = U^{-1}$. The r columns of U_1 form an orthonormal basis for $\mathcal{R}(A)$ and the μ columns of U_2 form an orthonormal basis for $\mathcal{N}(A^H)$.

The columns of V form an orthonormal basis for $\mathcal{X} = \mathbb{C}^n$, $\langle v_i, v_j \rangle = v_i^H v_j = \delta_{i,j}$, $i, j = 1, \cdots n$, or, equivalently, $V^H = V^{-1}$. The r columns of V_1 form an orthonormal basis for $\mathcal{R}(A^H)$ and the ν columns of V_2 form an orthonormal basis for $\mathcal{N}(A)$.

With the SVD at hand the Moore-Penrose pseudoinverse can be determined as

$$A^+ = V_1 S^{-1} U_1^H = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^H\,. \tag{9}$$

[29] See page 412 of the textbook by Meyer or the useful survey paper "The Singular Value Decomposition: Its Computation and Some Applications," V.C. Klema & A.J. Laub, *IEEE Transactions on Automatic Control,* Vol. AC-25, No. 2, pp. 164-176, April 1980.

Note that (9) does indeed obeys the four Moore-Penrose necessary and sufficient conditions for A^+ to be the pseudoinverse of A,

$$(AA^+)^H = AA^+\,, \qquad (A^+A)^H = A^+A\,, \qquad AA^+A = A\,, \qquad A^+AA^+ \;=\; A^+$$

which can be used as a check that one has correctly computed the pseudoinverse. The various projection operators can be determined as

$$\begin{aligned} P_{\mathcal{R}(A)} &= U_1U_1^H = AA^+ = I - P_{\mathcal{N}(A^H)}\,, \\ P_{\mathcal{N}(A^H)} &= U_2U_2^H = I - P_{\mathcal{R}(A)}\,, \\ P_{\mathcal{R}(A^H)} &= V_1V_1^H = A^+A = I - P_{\mathcal{N}(A)}\,, \\ P_{\mathcal{N}(A)} &= V_2V_2^H = I - P_{\mathcal{R}(A^H)}\,. \end{aligned}$$

Because it is possible to construct the projection operators in more than one way, one can test for correctness by doing so and checking for consistency of the answers. This is also another way to determine that the pseudoinverse has been correctly determined.

The columns of U are the m eigenvectors of (AA^H) and are known as the *left singular vectors* of A. Specifically, the columns of U_1 are the r eigenvectors of (AA^H) having associated nonzero eigenvalues $\sigma_i^2 > 0$, $i = 1, \cdots r$, while the columns of U_2 are the remaining μ eigenvectors of (AA^H), which have zero eigenvalues.

The n columns of V are the eigenvectors of (A^HA) and are known as the *right singular vectors* of A. Specifically, the columns of V_1 are the r eigenvectors of (A^HA) having associated nonzero eigenvalues $\sigma_i^2 > 0$, $i = 1, \cdots r$, while the ν columns of V_2 are the eigenvectors of (A^HA), which have zero eigenvalues.