Vector Space Concepts
ECE 275A – Statistical Parameter Estimation

Ken Kreutz-Delgado
ECE Department, UC San Diego
Heuristic Concept of a Linear Vector Space

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

- There exists a natural or conventional ‘zero point’ or “origin”, the **zero vector**, $\mathbf{0}$.

- Vectors can be **added** in a symmetric **commutative** and **associative** manner to produce other vectors

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

  $$x + y + z \triangleq x + (y + z) = (x + y) + z, \quad x, y, z \text{ are vectors} \quad \text{(associativity)}$$

- Vectors can be scaled by the (symmetric) multiplication of scalars (**scalar multiplication**) to produce other vectors

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

- The 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.

- Each vector $x$ has an **additive inverse**, $-x = (-1)x$

  $$x - x \triangleq x + (-1)x = x + (-x) = 0$$
Formal Concept of a Linear Vector Space

- A Vector Space, $\mathcal{X}$, is a set 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 consider finite dimensional vector spaces $\dim \mathcal{X} = n < \infty$ and mostly give results appropriate for this restriction.
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,

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

We refer to this as a **canonical representation** of a finite-dimensional vector. We often (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 the space $\mathcal{X}$

$$\alpha_1 x_1 + \cdots + \alpha_r x_r$$

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 important physical phenomenon of interest **cannot** be modeled by linear vector spaces, the classic example being rotations of a rigid body in three dimensional space (this is because finite (i.e., non-infinitesimal) 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 $mn$-dimensional 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 often work with $\mathcal{X}'$ in lieu of $\mathcal{X}$.
Examples of Vectors – Cont.

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 a sample of $f$ taken at time $t$.) Strictly speaking, then, $f(t)$ is **not** the function $f$ itself.
- Functions are **single-valued by definition**. Therefore

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

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:

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

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

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

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

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

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

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

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

- $\mathcal{V} = \text{hermitian matrices, } A = A^H (\text{the set of complex symmetric matrices where } A^H = (A^T) = (\bar{A})^T). \text{ This is not a subspace as it is not closed under scalar multiplication (check this!). It does include the zero element.}$
Given two subsets $\mathcal{V}$ and $\mathcal{W}$ of vectors, we define their set sum by

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

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 $\mathcal{X}$
- $\mathcal{V} \cup \mathcal{W} \subset \mathcal{V} + \mathcal{W}$ where in general $\mathcal{V} \cup \mathcal{W}$ is not a subspace.

In general, we have the following ordering of subspaces,

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

where $\{0\}$ is the trivial subspace consisting only of the zero vector (additive identity) of $\mathcal{X}$. The trivial subspace has dimension zero.
By definition, \( r \) vectors \( x_1, \ldots, x_r \in \mathcal{X} \) are **linearly independent** when,

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

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.
Assume that the $r$ vectors, $x_i$, are canonically represented, $x_i \in \mathbb{F}^n$.

Then 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 $x_1, \cdots, x_r$ are linearly independent iff the associated $n \times r$ matrix $X \triangleq (x_1 \cdots x_r)$ has full column rank (equivalently, iff the null space of $X$ is trivial).
The **span** of the collection \( x_1, \cdots, x_r \in \mathcal{X} \) is the set of all linear combinations of the vectors,

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

The set \( \mathcal{V} = \text{Span} \{x_1, \cdots, x_r\} \) 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**.
The dimension of the trivial subspace is zero, \(0 = \dim \{0\}\).

If \(V\) is a subspace of \(X\), \(V \subset X\), we have \(\dim V \leq \dim X\).

In general for two arbitrary subspaces \(V\) and \(W\) of \(X\) we have,

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

and

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

Furthermore, if \(X = V + W\) then,

\[
\dim X \leq \dim V + \dim W ,
\]

with equality if and only if \(V \cap W = \{0\}\).
Independent Subspaces and Projections

- 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 \( \dim \mathcal{X} = \dim \mathcal{V} + \dim \mathcal{W} \).

  Given 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** \( \mathcal{W} \).
  - The unique component \( w \) is called **the projection of** \( x \) **onto** \( \mathcal{W} \) **along its companion space** \( \mathcal{V} \).
\[ \mathbf{x} = \mathbf{v} \oplus \mathbf{w} \]

- Projection of \( \mathbf{x} \) onto \( \mathbf{v} \) along \( \mathbf{w} \)
- Projection of \( \mathbf{x} \) onto \( \mathbf{w} \) along \( \mathbf{v} \)
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 $P_{\mathcal{V}|\mathcal{W}} + P_{\mathcal{W}|\mathcal{V}} = I$. I.e., $P_{\mathcal{V}|\mathcal{W}} = I - P_{\mathcal{W}|\mathcal{V}}$.

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

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

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

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

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

It can also be shown that the projection operators $P_{\mathcal{V}|\mathcal{W}}$ and $P_{\mathcal{W}|\mathcal{V}}$ are **linear operators**.
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

$$A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 Ax_1 + \alpha_2 Ax_2 \quad \forall x_1, x_2 \in \mathcal{X}, \forall \alpha_1, \alpha_2 \in \mathcal{F}.$$ 

- 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 vector 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 as a consequence of their linearity.

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

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

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

The **Range Space** (or **Image**),

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

The **Nullspace** (or **Kernel**),

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

- 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**.
The dimension of the range space of a linear operator $A$ is the \textbf{rank} of $A$,

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

The dimension of the nullspace of a linear operator $A$ is the \textbf{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 \textbf{intrinsic} properties of the linear operator $A$ and \textbf{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 $A \in \mathcal{F}^{m \times n}$, 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 $\nu = n - r$. 
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 typically 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. $y \in R(A)$ for all $y \in Y$ so that a solution exists for all $y$. I.e., we demand that $A$ be onto, $R(A) = Y$ or, equivalently, that $r(A) = m$. 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.

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