

Section 3: Span, Subspaces and Bases

MATH241 Lecture Notes, notes pages 22–31

Notation. Scalars are plain italic (a, α, z), vectors are bold ($\mathbf{u}, \mathbf{x}$), and matrices are italic capitals (A, B ; zero matrix O , identity I). The handwritten notes underline vectors once and matrices twice.

These notes are built on top of Orhan Arikan's linear algebra course notes.

— notes page 22 —

Span and Subspace (Section 9.7 in the textbook)

Def'n Given vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ in a vector space S , the set of all their linear combinations,

$$\mathbf{u} = \alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k, \quad \text{for } \alpha_i \in \mathbb{R}, \quad 1 \leq i \leq k,$$

is called the *span* of $\mathbf{u}_1, \dots, \mathbf{u}_k$, and is written

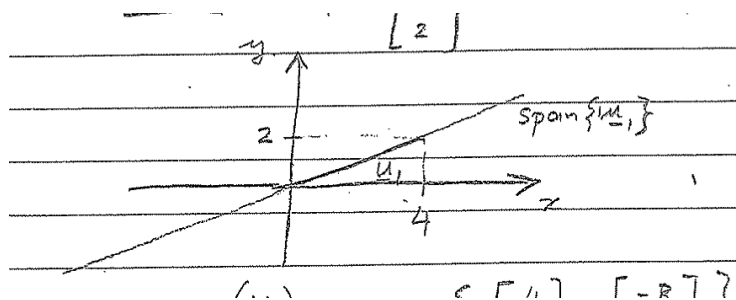
$$\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}.$$

The set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is called a *generating set* of this span.

ex's

- (i) The span of the single vector $\mathbf{u}_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \in \mathbb{R}^2$ is all of its scalar multiples. In the plane this is the straight line through the origin in the direction of $\mathbf{u}_1$:

$$\text{span}\{\mathbf{u}_1\} = \left\{ \mathbf{u} : \mathbf{u} = \alpha \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \alpha \in \mathbb{R} \right\}.$$



- (ii) Adding a second vector does not always enlarge the span. Here $\mathbf{u}_2 = -2\mathbf{u}_1$ points along the same line, so

$$\text{span} \left\{ \underbrace{\begin{bmatrix} 4 \\ 2 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} -8 \\ -4 \end{bmatrix}}_{\mathbf{u}_2} \right\} = \text{span} \left\{ \begin{bmatrix} 4 \\ 2 \end{bmatrix} \right\}.$$

Indeed, every combination of the two vectors is a multiple of $\mathbf{u}_1$:

$$\mathbf{u} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 = \alpha_1 \mathbf{u}_1 - 2\alpha_2 \mathbf{u}_1 = (\alpha_1 - 2\alpha_2) \mathbf{u}_1 = \beta \cdot \mathbf{u}_1.$$

Def'n A non-empty subset T of a vector space S is a *subspace* of S if it satisfies both of the following:

- (i) It is closed under addition: for all $\mathbf{u}$ and $\mathbf{v}$ in T , the sum $\mathbf{u} + \mathbf{v}$ is in T .
- (ii) It is closed under scalar multiplication: for all $\mathbf{u} \in T$ and $\alpha \in \mathbb{R}$, the multiple $\alpha\mathbf{u}$ is in T .

Only these two conditions need to be checked. The other axioms of a vector space are inherited from S , so a subspace is itself a vector space.

Theorem 9.7.1

Every span is a subspace.

Theorem 9.7.1

If $\mathbf{u}_1, \dots, \mathbf{u}_k$ are vectors in a vector space S , then $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a subspace of S .

Proof The span is non-empty, since it contains $\mathbf{0}$ and each $\mathbf{u}_i$. Now take two elements of it. If $\mathbf{u} \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, there are scalars $\alpha_1, \dots, \alpha_k$ with $\mathbf{u} = \alpha_1\mathbf{u}_1 + \dots + \alpha_k\mathbf{u}_k$.

— notes page 23 —

Likewise, if $\mathbf{v} \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, there are scalars $\beta_1, \dots, \beta_k$ with $\mathbf{v} = \beta_1\mathbf{u}_1 + \dots + \beta_k\mathbf{u}_k$. Adding gives

$$\mathbf{u} + \mathbf{v} = (\alpha_1 + \beta_1)\mathbf{u}_1 + \dots + (\alpha_k + \beta_k)\mathbf{u}_k \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\},$$

so the span is closed under addition. For any $c \in \mathbb{R}$,

$$c\mathbf{u} = c\alpha_1\mathbf{u}_1 + \dots + c\alpha_k\mathbf{u}_k \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\},$$

so it is closed under scalar multiplication too. Hence $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a subspace of S . ■

Linear Dependence

Def'n A set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is *linearly dependent* (LD) if at least one of the vectors can be written as a linear combination of the others. If none of them can, the set is *linearly independent* (LI). Informally, a dependent set contains a vector that is redundant.

ex's

(i) The set $\left\{ \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\mathbf{u}_2}, \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{\mathbf{u}_3} \right\}$ is linearly dependent, because $\mathbf{u}_1 = \frac{1}{2}\mathbf{u}_2 + \frac{1}{2}\mathbf{u}_3$.

(ii) The set $\left\{ \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{u}_2} \right\}$ is linearly independent, since neither vector is a multiple of the other.

(iii) The set $\left\{ \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{u}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{u}_3} \right\}$ is linearly dependent, because $\mathbf{u}_3 = 0 \cdot \mathbf{u}_1 + 0 \cdot \mathbf{u}_2$. A set containing the zero vector is always dependent.

Theorem 9.8.1

The definition of dependence is hard to test directly, because we would have to try every vector as a combination of the others. The next theorem gives an equivalent test that is easier to use: look for non-trivial ways of combining the vectors to get zero.

Theorem 9.8.1

Let $\mathbf{u}_1, \dots, \mathbf{u}_k$ be vectors in a vector space S . The set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is LD if and only if there exist scalars α_i , not all zero, such that

$$\alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k = \mathbf{0}.$$

Proof The proof has two parts, one for each direction of “if and only if”.

— notes page 24 —

proof of $(\Rightarrow)$ If $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is LD, then for some j we can write

$$\mathbf{u}_j = \sum_{i \neq j} \alpha_i \mathbf{u}_i.$$

Moving $\mathbf{u}_j$ to the other side gives $\mathbf{0} = \sum_{i \neq j} \alpha_i \mathbf{u}_i - \mathbf{u}_j$. This is a combination of all the vectors with coefficients α_i , where the coefficient of $\mathbf{u}_j$ is -1 . So not all coefficients are zero, and the combination equals $\mathbf{0}$. ■

proof of $(\Leftarrow)$ Suppose $\sum_{i=1}^k \alpha_i \mathbf{u}_i = \mathbf{0}$ with not all α_i zero. Pick some $\alpha_j \neq 0$. Then

$$\begin{aligned} \sum_{i \neq j} \alpha_i \mathbf{u}_i + \alpha_j \mathbf{u}_j &= \mathbf{0} \\ \Rightarrow \mathbf{u}_j &= \sum_{i \neq j} \left(-\frac{\alpha_i}{\alpha_j} \right) \mathbf{u}_i, \end{aligned}$$

which expresses $\mathbf{u}_j$ as a combination of the others, so the set is LD. ■

ex/ Is $\left\{ \underbrace{\begin{bmatrix} 2 \\ 0 \\ 1 \\ -3 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{u}_2}, \underbrace{\begin{bmatrix} 2 \\ 2 \\ 3 \\ 0 \end{bmatrix}}_{\mathbf{u}_3} \right\}$ linearly independent?

By Theorem 9.8.1, the set is LI exactly when $\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 = \mathbf{0}$ has only the trivial solution $\alpha_1 = \alpha_2 = \alpha_3 = 0$. Writing this out component by component gives a homogeneous linear system:

$$\left. \begin{aligned} 2\alpha_1 + 2\alpha_3 &= 0 \\ \alpha_2 + 2\alpha_3 &= 0 \\ \alpha_1 + \alpha_2 + 3\alpha_3 &= 0 \\ -3\alpha_1 + \alpha_2 &= 0 \end{aligned} \right\}$$

We reduce its augmented matrix. The columns are $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ and the zero right-hand side.

$$\begin{array}{cccc|ccc} \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 1 & 1 & 3 & 0 \\ -3 & 1 & 0 & 0 \end{bmatrix} & \xrightarrow[\text{Elimination}]{\text{Gauss}} & \begin{bmatrix} \boxed{2} & 0 & 2 & 0 \\ 0 & \boxed{1} & 2 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{0} & & & & \end{array}$$

The echelon form (reached after interchanging the last two rows) has three pivots and no free variables, so the only solution is the trivial one. Therefore $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is linearly independent.

— notes page 25 —

Theorem 9.8.4

For a linearly independent set, a combination of the vectors can be written in only one way. Two combinations that give the same vector must use the same coefficients.

Theorem 9.8.4

If $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is LI, then

$$\alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k = \beta_1 \mathbf{u}_1 + \dots + \beta_k \mathbf{u}_k \Rightarrow \alpha_i = \beta_i, \quad 1 \leq i \leq k.$$

Proof Subtract the two sides of the equation from each other:

$$\alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k = \beta_1 \mathbf{u}_1 + \dots + \beta_k \mathbf{u}_k \Rightarrow (\alpha_1 - \beta_1) \mathbf{u}_1 + \dots + (\alpha_k - \beta_k) \mathbf{u}_k = \mathbf{0}.$$

Because $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is LI, the only combination that gives $\mathbf{0}$ is the trivial one, so $\alpha_i - \beta_i = 0$, that is, $\alpha_i = \beta_i$ for $1 \leq i \leq k$. ■

Theorem 9.8.5

Orthogonal vectors point in completely different directions, so none of them can be built from the others.

Theorem 9.8.5

Every finite set of non-zero orthogonal vectors is LI.

Proof Suppose, for contradiction, that a set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ of non-zero orthogonal vectors were LD. Then for some j ,

$$\mathbf{u}_j = \sum_{i \neq j} \alpha_i \mathbf{u}_i.$$

Take the dot product of both sides with $\mathbf{u}_j$. Every term on the right vanishes, because $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for $i \neq j$:

$$\mathbf{u}_j \cdot \mathbf{u}_j = \left(\sum_{i \neq j} \alpha_i \mathbf{u}_i \right) \cdot \mathbf{u}_j = \sum_{i \neq j} \alpha_i \underbrace{(\mathbf{u}_i \cdot \mathbf{u}_j)}_0 = 0.$$

So $\mathbf{u}_j \cdot \mathbf{u}_j = 0$, which forces $\mathbf{u}_j = \mathbf{0}$. This contradicts the assumption that all the vectors are non-zero. ■

Important Observation: a set that contains the zero vector is always linearly dependent. For $U = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ with, say, $\mathbf{u}_j = \mathbf{0}$, we can take every coefficient to be 0 except the one in front of $\mathbf{0}$, which can be any non-zero number such as 5:

$$0\mathbf{u}_1 + 0\mathbf{u}_2 + \dots + 5\mathbf{0} + \dots + 0\mathbf{u}_k = \mathbf{0}.$$

This is a combination with a non-zero coefficient that equals $\mathbf{0}$.

Bases, Expansions, Dimension

Def'n A finite set of vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ in a vector space S is a *basis* for S if each vector in S can be written in exactly one way as

$$\mathbf{u} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_k \mathbf{e}_k = \sum_{i=1}^k \alpha_i \mathbf{e}_i.$$

A basis is a set of building blocks: every vector is made from them, and made in only one way.

Theorem 9.9.1

There is an easier test for a basis than checking uniqueness directly.

Theorem 9.9.1

$\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is a basis for S if $\text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_k\} = S$ and $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is LI.

Proof Let $\mathbf{u}$ be any vector in S . Since $\text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_k\} = S$, we can write $\mathbf{u} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_k \mathbf{e}_k$ for some scalars α_i .

— notes page 26 —

To conclude that $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is a basis, we still have to show that this expansion of $\mathbf{u}$ is unique. If

$$\mathbf{u} = \beta_1 \mathbf{e}_1 + \dots + \beta_k \mathbf{e}_k = \alpha_1 \mathbf{e}_1 + \dots + \alpha_k \mathbf{e}_k$$

then $\beta_i = \alpha_i$ for every i , by Theorem 9.8.4, because $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is LI. ■

ex/ In $\mathbb{R}^2$, the vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ form a basis. To see how a vector is expanded in it, take $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and look for the coefficients:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow \begin{aligned} \alpha_1 + \alpha_2 &= u_1 \\ 2\alpha_1 - \alpha_2 &= u_2. \end{aligned}$$

The augmented coefficient matrix reduces as

$$\begin{bmatrix} 1 & 1 & u_1 \\ 2 & -1 & u_2 \end{bmatrix} \xrightarrow[\text{Elimination}]{\text{Gauss}} \begin{bmatrix} 1 & 1 & u_1 \\ 0 & -3 & u_2 - 2u_1 \end{bmatrix}.$$

Back substitution gives

$$\begin{aligned} \alpha_2 &= -\frac{1}{3}u_2 + \frac{2}{3}u_1 \\ \alpha_1 &= u_1 - \alpha_2 = u_1 + \frac{1}{3}u_2 - \frac{2}{3}u_1 = \frac{1}{3}u_1 + \frac{1}{3}u_2. \end{aligned}$$

So every vector of $\mathbb{R}^2$ is expanded in this basis as

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \frac{u_1 + u_2}{3} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{2u_1 - u_2}{3} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The coefficients are unique, as they must be, because $\mathbf{e}_1$ and $\mathbf{e}_2$ are LI.

Def'n The *dimension* of a vector space counts how many independent directions it has. If the greatest number of LI vectors that can be found in S is k , with $1 \leq k < \infty$, then S is k -dimensional, and we write

$$\dim S = k.$$

If $S = \{\mathbf{0}\}$, then $\dim S = 0$. If arbitrarily many LI vectors can be found in S , then S is infinite-dimensional.

Theorem 9.9.2

The dimension can be read off from any basis.

Theorem 9.9.2

If a vector space S has a basis consisting of k vectors, then $\dim S = k$.

Proof Let $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ be the basis. It is LI, because it is a basis, so $\dim S \geq k$.

— notes page 27 —

Now suppose $\dim S > k$. Then there is an LI set $\{\mathbf{u}_1, \dots, \mathbf{u}_{k+1}\}$ of $k+1$ vectors in S . Since $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is a basis, each $\mathbf{u}_j$ can be written uniquely in terms of it:

$$\mathbf{u}_j = \sum_{i=1}^k \alpha_{ji} \mathbf{e}_i, \quad 1 \leq j \leq k+1.$$

Now consider a combination of the $\mathbf{u}_j$ that equals zero, and rewrite it in terms of the $\mathbf{e}_i$:

$$\begin{aligned} \mathbf{0} &= \beta_1 \mathbf{u}_1 + \dots + \beta_{k+1} \mathbf{u}_{k+1} = \sum_{j=1}^{k+1} \beta_j \mathbf{u}_j \\ &= \sum_{j=1}^{k+1} \beta_j \sum_{i=1}^k \alpha_{ji} \mathbf{e}_i = \sum_{i=1}^k \mathbf{e}_i \left[\sum_{j=1}^{k+1} \alpha_{ji} \beta_j \right]. \end{aligned}$$

Since the $\mathbf{e}_i$ are LI, every bracket must be zero:

$$\Rightarrow \sum_{j=1}^{k+1} \alpha_{ji} \beta_j = 0 \quad \text{for } 1 \leq i \leq k.$$

Written out in full, these are k equations in the $k+1$ unknowns $\beta_1, \dots, \beta_{k+1}$:

$$\left. \begin{aligned} \alpha_{11}\beta_1 + \alpha_{21}\beta_2 + \dots + \alpha_{k+1,1}\beta_{k+1} &= 0 \\ \alpha_{1,2}\beta_1 + \alpha_{2,2}\beta_2 + \dots + \alpha_{k+1,2}\beta_{k+1} &= 0 \\ &\vdots \\ \alpha_{1,k}\beta_1 + \alpha_{2,k}\beta_2 + \dots + \alpha_{k+1,k}\beta_{k+1} &= 0 \end{aligned} \right\} \begin{array}{l} k \text{ equations in} \\ k+1 \text{ unknowns.} \end{array}$$

A homogeneous system with more unknowns than equations always has non-trivial solutions (Theorem 8.3.4). So there are β_j , not all zero, with $\beta_1 \mathbf{u}_1 + \dots + \beta_{k+1} \mathbf{u}_{k+1} = \mathbf{0}$. This contradicts the assumption that $\{\mathbf{u}_1, \dots, \mathbf{u}_{k+1}\}$ is LI. Hence $\dim S = k$. ■

Theorem 9.9.3

Theorem 9.9.3

The dimension of $\mathbb{R}^n$ is n .

Proof Take the standard basis vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, ..., $\mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$. They are LI. Every $\mathbf{u} \in \mathbb{R}^n$

equals $\sum_{i=1}^n \alpha_i \mathbf{e}_i$ with $\alpha_i = u_i$, so $\text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_n\} = \mathbb{R}^n$. Hence $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$, and $\dim \mathbb{R}^n = n$. ■

— notes page 28 —

Theorem 9.9.4

The dimension of a span equals the number of vectors in the generating set that are really needed.

Theorem 9.9.4

For vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$, not all zero, $\dim[\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}]$ equals the greatest number of linearly independent vectors within the generating set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$.

Proof Let m be the greatest number of linearly independent vectors in $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, and number the vectors so that the first m are LI. Each of the remaining vectors $\mathbf{u}_{m+1}, \dots, \mathbf{u}_k$ must then depend linearly on $\mathbf{u}_1, \dots, \mathbf{u}_m$:

$$\mathbf{u}_i = \alpha_{i1}\mathbf{u}_1 + \dots + \alpha_{im}\mathbf{u}_m, \quad m+1 \leq i \leq k.$$

Now take any $\mathbf{v} \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, say $\mathbf{v} = \sum_{i=1}^k \beta_i \mathbf{u}_i$. Replacing each of the last $k-m$ vectors by its expression gives

$$\begin{aligned} \Rightarrow \quad \mathbf{v} &= \sum_{i=1}^m \beta_i \mathbf{u}_i + \sum_{j=m+1}^k \beta_j \mathbf{u}_j \\ \mathbf{v} &= \sum_{i=1}^m \beta_i \mathbf{u}_i + \sum_{j=m+1}^k \beta_j \left[\sum_{i=1}^m \alpha_{ji} \mathbf{u}_i \right] \\ \mathbf{v} &= \sum_{i=1}^m \left[\beta_i + \sum_{j=m+1}^k \beta_j \alpha_{ji} \right] \mathbf{u}_i \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_m\}. \end{aligned}$$

So $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\} = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$. Since $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is LI and spans this space, it is a basis for $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, and the dimension is m . ■

Orthogonal Bases

A vector space of dimension two or more has infinitely many bases. In practice we prefer *orthogonal* bases, because the coefficients of a vector are then easy to find.

Let $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ be an orthogonal basis for S . Any $\mathbf{u} \in S$ can be written as $\mathbf{u} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_k \mathbf{e}_k$. To find α_j , take the dot product with $\mathbf{e}_j$. By orthogonality, every term except the j -th vanishes:

$$\mathbf{u} \cdot \mathbf{e}_j = (\alpha_1 \mathbf{e}_1 + \dots + \alpha_k \mathbf{e}_k) \cdot \mathbf{e}_j = \alpha_j \mathbf{e}_j \cdot \mathbf{e}_j.$$

Solving for α_j gives the coefficient formula, where the denominator $\mathbf{e}_i \cdot \mathbf{e}_i = \|\mathbf{e}_i\|^2$ is the normalization factor:

Expansion in an orthogonal basis

$$\alpha_j = \frac{\mathbf{u} \cdot \mathbf{e}_j}{\mathbf{e}_j \cdot \mathbf{e}_j} \quad \Rightarrow \quad \mathbf{u} = \sum_{i=1}^k \left(\frac{\mathbf{u} \cdot \mathbf{e}_i}{\mathbf{e}_i \cdot \mathbf{e}_i} \right) \mathbf{e}_i.$$

— notes page 29 —

Dividing each basis vector by its length, $\hat{\mathbf{e}}_i = \mathbf{e}_i / \|\mathbf{e}_i\|$, turns the formula above into $\mathbf{u} = \sum_{i=1}^k (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \hat{\mathbf{e}}_i$. If the basis is *orthonormal*, that is, orthogonal with every vector of length one, then the denominators are all 1 and the coefficients are just dot products:

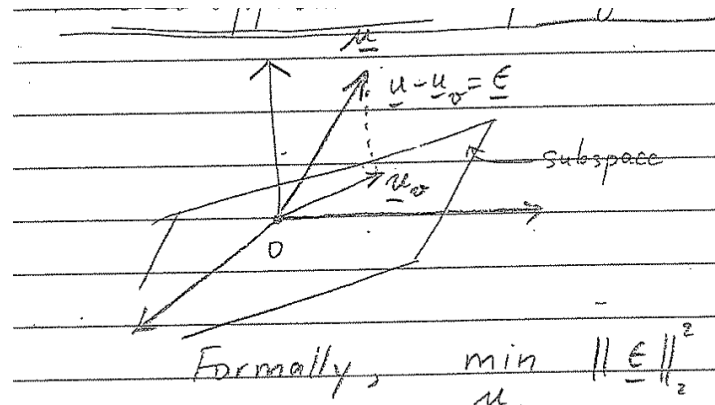
$$\alpha_i = \frac{\mathbf{u} \cdot \mathbf{e}_i}{\underbrace{\mathbf{e}_i \cdot \mathbf{e}_i}_1} = \mathbf{u} \cdot \mathbf{e}_i.$$

Best Approximation

In many engineering applications we are given a vector $\mathbf{u}$ and want its best approximation inside a subspace. Let S be a vector space and V a subspace of S , and let $\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_N\}$ be an orthonormal basis for V :

$$\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij}, \quad 1 \leq i, j \leq N, \quad \text{span}\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_N\} = V.$$

The picture shows the idea. The vector $\mathbf{u}_V$ in V should be as close to $\mathbf{u}$ as possible, so the error $\boldsymbol{\epsilon} = \mathbf{u} - \mathbf{u}_V$ should be as short as possible.



We want to minimize the squared length of the error by choosing $\mathbf{u}_V \in V$:

$$\min_{\mathbf{u}_V} \|\boldsymbol{\epsilon}\|_2^2 = \min_{\mathbf{u}_V} \|\mathbf{u} - \mathbf{u}_V\|_2^2.$$

Every $\mathbf{u}_V \in V$ can be written in the orthonormal basis as $\mathbf{u}_V = \sum_{i=1}^N \alpha_i \hat{\mathbf{e}}_i$, so choosing $\mathbf{u}_V$ means choosing the numbers α_i . The problem becomes one of minimizing a cost function J of these numbers:

$$\min_{\alpha_1, \dots, \alpha_N} \underbrace{\left\| \mathbf{u} - \sum_{i=1}^N \alpha_i \hat{\mathbf{e}}_i \right\|_2^2}_{J(\alpha_1, \dots, \alpha_N)}.$$

Expanding the squared length as a dot product, and using the fact that the basis is orthonormal, gives

$$\begin{aligned} J(\alpha_1, \dots, \alpha_N) &= \left(\mathbf{u} - \sum_{i=1}^N \alpha_i \hat{\mathbf{e}}_i \right) \cdot \left(\mathbf{u} - \sum_{i=1}^N \alpha_i \hat{\mathbf{e}}_i \right) \\ &= \mathbf{u} \cdot \mathbf{u} - 2 \sum_{i=1}^N \alpha_i \mathbf{u} \cdot \hat{\mathbf{e}}_i + \underbrace{\left(\sum_{i=1}^N \alpha_i \hat{\mathbf{e}}_i \right) \cdot \left(\sum_{i=1}^N \alpha_i \hat{\mathbf{e}}_i \right)}_{\sum_{i=1}^N \alpha_i^2}. \end{aligned}$$

— notes page 30 —

In short, J is a quadratic function of the α_i :

$$J(\alpha_1, \dots, \alpha_N) = \underbrace{\mathbf{u} \cdot \mathbf{u} - 2 \sum_{i=1}^N \alpha_i \mathbf{u} \cdot \hat{\mathbf{e}}_i + \sum_{i=1}^N \alpha_i^2}_{\text{quadratic function of the } \alpha_i}.$$

Its minimum is where all the partial derivatives are zero:

$$\frac{\partial J}{\partial \alpha_i} = -2\mathbf{u} \cdot \hat{\mathbf{e}}_i + 2\alpha_i = 0 \Rightarrow \alpha_i = \mathbf{u} \cdot \hat{\mathbf{e}}_i.$$

So the best approximation of $\mathbf{u}$ in the subspace V is

$$\mathbf{u}_V = \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \hat{\mathbf{e}}_i.$$

If $\mathbf{u}$ already lies in V , its best approximation is $\mathbf{u}$ itself.

Now let us look at the error of this best approximation:

$$\boldsymbol{\epsilon} = \mathbf{u} - \mathbf{u}_V = \mathbf{u} - \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \hat{\mathbf{e}}_i.$$

Claim The error $\boldsymbol{\epsilon}$ is orthogonal to V , that is, orthogonal to every vector in V .

Proof Any vector in V can be written $\mathbf{v} = \sum_{j=1}^N \beta_j \hat{\mathbf{e}}_j$, because $\text{span}\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_N\} = V$. Then

$$\begin{aligned} \boldsymbol{\epsilon} \cdot \mathbf{v} &= \left(\mathbf{u} - \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \hat{\mathbf{e}}_i \right) \cdot \left(\sum_{j=1}^N \beta_j \hat{\mathbf{e}}_j \right) \\ &= \sum_{j=1}^N \beta_j \mathbf{u} \cdot \hat{\mathbf{e}}_j - \sum_{i=1}^N \sum_{j=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \beta_j \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j \\ &= \sum_{j=1}^N \beta_j \mathbf{u} \cdot \hat{\mathbf{e}}_j - \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \beta_i = 0, \end{aligned}$$

where the last step uses $\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij}$. ■

Because the error is perpendicular to V , we call $\mathbf{u}_V$ the *orthogonal projection* of $\mathbf{u}$ onto V :

Orthogonal projection onto V

$$\text{proj}_V \mathbf{u} = \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \hat{\mathbf{e}}_i.$$

This formula needs an orthonormal basis of V . What if we are only given a basis that is not orthonormal?

— notes page 31 —

Answer We can use the Gram–Schmidt orthogonalization process.

The Gram-Schmidt Orthogonalization Process

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_N\}$ be any basis for a subspace V . We want to build from it an orthonormal basis $\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_N\}$ for the same subspace. The process works one vector at a time, and each new vector is made orthogonal to all the previous ones.

Step 1 Take the first vector and normalize it: $\hat{\mathbf{e}}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$. Then $\hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_1 = 1$.

Step 2 (i) Project $\mathbf{v}_2$ onto $\text{span}\{\hat{\mathbf{e}}_1\}$ to get $\mathbf{v}_{2p} = (\mathbf{v}_2 \cdot \hat{\mathbf{e}}_1) \hat{\mathbf{e}}_1$. (ii) Compute the projection error $\boldsymbol{\epsilon}_2 = \mathbf{v}_2 - \mathbf{v}_{2p}$; it is orthogonal to $\hat{\mathbf{e}}_1$. Normalize it:

$$\hat{\mathbf{e}}_2 = \frac{\boldsymbol{\epsilon}_2}{\|\boldsymbol{\epsilon}_2\|}, \quad \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_1 = 0, \quad \|\hat{\mathbf{e}}_2\| = 1.$$

Step 3 Repeat for the remaining vectors, $2 < i \leq N$. Project $\mathbf{v}_i$ onto the span of the vectors found so far, $\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_{i-1}\}$:

$$\mathbf{v}_{ip} = \sum_{j=1}^{i-1} (\mathbf{v}_i \cdot \hat{\mathbf{e}}_j) \hat{\mathbf{e}}_j.$$

The projection error is $\boldsymbol{\epsilon}_i = \mathbf{v}_i - \mathbf{v}_{ip}$. Then set

$$\hat{\mathbf{e}}_i = \frac{\boldsymbol{\epsilon}_i}{\|\boldsymbol{\epsilon}_i\|}.$$

The new vector $\hat{\mathbf{e}}_i$ is orthogonal to $\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_{i-1}$, because $\boldsymbol{\epsilon}_i$ is the error of projecting onto their span. It has length one by construction. The division by $\|\boldsymbol{\epsilon}_i\|$ is allowed because $\boldsymbol{\epsilon}_i \neq \mathbf{0}$: the $\mathbf{v}_i$ are LI, so $\mathbf{v}_i$ does not lie in the span of the previous ones. So Gram–Schmidt turns any basis into an orthonormal basis.

Gram–Schmidt formulas

$$\hat{\mathbf{e}}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \quad \hat{\mathbf{e}}_2 = \frac{\mathbf{v}_2 - (\mathbf{v}_2 \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1}{\|\mathbf{v}_2 - (\mathbf{v}_2 \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1\|}, \quad \dots, \quad \hat{\mathbf{e}}_i = \frac{\mathbf{v}_i - \sum_{j=1}^{i-1} (\mathbf{v}_i \cdot \hat{\mathbf{e}}_j)\hat{\mathbf{e}}_j}{\left\| \mathbf{v}_i - \sum_{j=1}^{i-1} (\mathbf{v}_i \cdot \hat{\mathbf{e}}_j)\hat{\mathbf{e}}_j \right\|}.$$