TOPIC 1

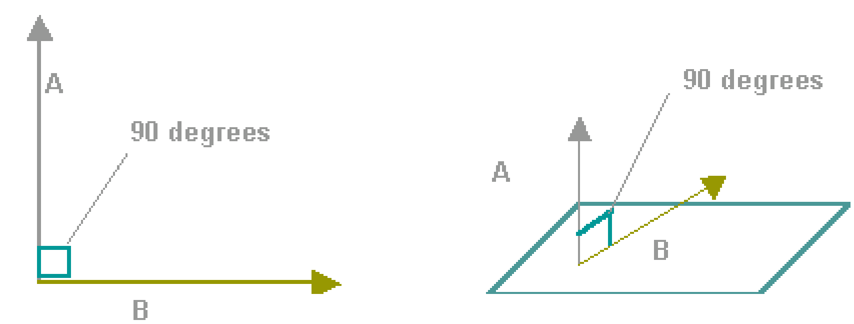
Least squares: Cholesky, QR and Householder

2 A key concept: orthogonality

Course: Math 535 (http://www.math.wisc.edu/~roch/mmids/) - Mathematical Methods in Data Science (MMiDS)
Author: Sebastien Roch (http://www.math.wisc.edu/~roch/), Department of Mathematics, University of Wisconsin-Madison
Updated: Jun 14, 2021
Copyright: © 2021 Sebastien Roch

Orthogonality plays a key role in linear algebra for data science thanks to its computational properties and its connection to the least-squares problem.

Definition (Orthogonality): Vectors $\mathbf{u}$ and $\mathbf{v}$ in V are orthogonal if their inner product satisfies $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.



(Source) (https://towardsdatascience.com/from-norm-to-orthogonality-fundamental-mathematics-for-machine-learning-with-intuitive-examples-57bb898e69f2)

Orthogonality has important implications. The following classical result will be useful below.

Lemma (Pythagoras): Let $\mathbf{u}, \mathbf{v} \in V$ be orthogonal. Then $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Proof: Using $\|\mathbf{w}\|^2 = \langle \mathbf{w}, \mathbf{w} \rangle$, we get

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + 2 \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2. \end{aligned}$$

□

Here is an application of *Pythagoras*.

Lemma (Cauchy-Schwarz): For any $\mathbf{u}, \mathbf{v} \in V$, $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$.

Proof: Let $\mathbf{q} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ be the unit vector in the direction of $\mathbf{v}$. We want to show $|\langle \mathbf{u}, \mathbf{q} \rangle| \leq \|\mathbf{u}\|$. Decompose $\mathbf{u}$ into its projection onto $\mathbf{q}$ and what's left:

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q} + \{ \mathbf{u} - \langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q} \} .$$

The two terms on the right-hand side are orthogonal, so *Pythagoras* gives

$$\|\mathbf{u}\|^2 = \|\langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q}\|^2 + \|\mathbf{u} - \langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q}\|^2 \geq \|\langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q}\|^2 = \langle \mathbf{u}, \mathbf{q} \rangle^2.$$

Taking a square root gives the claim.□

2.1 Basis expansion

To begin to see the power of orthogonality, consider the following. A list of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is orthonormal if the $\mathbf{u}_i$'s are pairwise orthogonal and each has norm 1, that is, for all i and all $j \neq i$, $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ and $\|\mathbf{u}_i\| = 1$.

Lemma (Properties of Orthonormal Lists): Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be an orthonormal list of vectors. Then:

1. $\| \sum_{j=1}^m \alpha_j \mathbf{u}_j \|^2 = \sum_{j=1}^m \alpha_j^2$ for any $\alpha_j \in \mathbb{R}$, $j \in [m]$, and
2. $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ are linearly independent.

Proof: For 1, using that $\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle$ and $\langle \beta \mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_3 \rangle = \beta \langle \mathbf{x}_1, \mathbf{x}_3 \rangle + \langle \mathbf{x}_2, \mathbf{x}_3 \rangle$ (which follow directly from the definition of the inner product), we have

$$\begin{aligned}\left\| \sum_{j=1}^m \alpha_j \mathbf{u}_j \right\|^2 &= \left\langle \sum_{i=1}^m \alpha_i \mathbf{u}_i, \sum_{j=1}^m \alpha_j \mathbf{u}_j \right\rangle \\ &= \sum_{i=1}^m \alpha_i \left\langle \mathbf{u}_i, \sum_{j=1}^m \alpha_j \mathbf{u}_j \right\rangle \\ &= \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j \left\langle \mathbf{u}_i, \mathbf{u}_j \right\rangle \\ &= \sum_{i=1}^m \alpha_i^2\end{aligned}$$

where we used orthonormality in the rightmost equation, that is, $\langle \mathbf{u}_i, \mathbf{u}_j \rangle$ is 1 if $i = j$ and 0 otherwise.

For 2, suppose $\sum_{i=1}^m \beta_i \mathbf{u}_i = \mathbf{0}$, then we must have by 1 that $\sum_{i=1}^m \beta_i^2 = 0$. That implies $\beta_i = 0$ for all i. Hence the $\mathbf{u}_i$'s are linearly independent.$\square$

Given a basis $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ of $\mathcal{U}$, we know that: for any $\mathbf{w} \in \mathcal{U}$, $\mathbf{w} = \sum_{i=1}^m \alpha_i \mathbf{u}_i$ for some α_i's. It is not immediately obvious in general how to find the α_i's. In the orthonormal case, however, it is straighforward.

Theorem (Orthonormal Expansion): Let $\mathbf{q}_1, \dots, \mathbf{q}_m$ be an orthonormal basis of $\mathcal{U}$ and let $\mathbf{u} \in \mathcal{U}$. Then

$$\mathbf{u} = \sum_{j=1}^m \langle \mathbf{u}, \mathbf{q}_j \rangle \, \mathbf{q}_j.$$

Proof: Because $\mathbf{u} \in \mathcal{U}$, $\mathbf{u} = \sum_{i=1}^m \alpha_i \mathbf{q}_i$ for some α_i. Take the inner product with $\mathbf{q}_j$ and use orthonormality:

$$\langle \mathbf{u}, \mathbf{q}_j \rangle = \left\langle \sum_{i=1}^m \alpha_i \mathbf{q}_i, \mathbf{q}_j \right\rangle = \sum_{i=1}^m \alpha_i \langle \mathbf{q}_i, \mathbf{q}_j \rangle = \alpha_j.$$

$\square$

So we've shown that working with orthonormal bases is desirable. What if we don't have one? We review the Gram-Schmidt algorithm (https://en.wikipedia.org/wiki/Gram–Schmidt_process) below, which will imply that every linear subspace has an orthonormal basis.

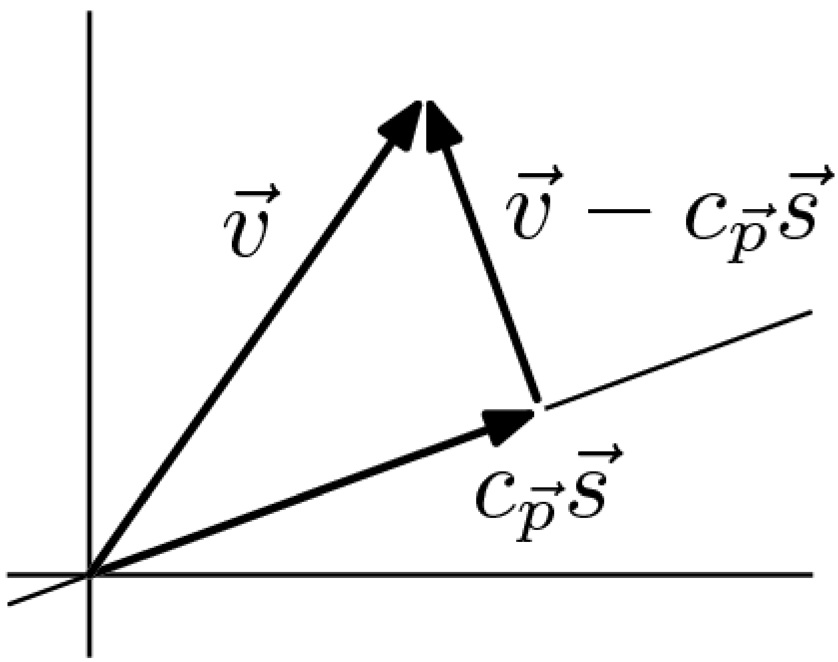
But, first, we define the orthogonal projection.

2.2 Orthogonal projection

Let's consider the following problem. We have a linear subspace $\mathcal{U} \subseteq V$ and a vector $\mathbf{v} \notin \mathcal{U}$. We want to find the vector $\mathbf{v}^*$ in $\mathcal{U}$ that is closest to $\mathbf{v}$ in 2-norm, that is, we want to solve

$$\min_{\mathbf{v}^* \in \mathcal{U}} \|\mathbf{v}^* - \mathbf{v}\|.$$

Example: Consider the two-dimensional case with a one-dimensional subspace, say $\mathcal{U} = \mathrm{span}(\mathbf{u_1})$ with $\|\mathbf{u}_1\| = 1$. The geometrical intuition is in the following figure. The solution $\mathbf{v}^*$ has the property that the difference $\mathbf{v} - \mathbf{v}^*$ makes a right angle with $\mathbf{u}_1$, that is, it is orthogonal to it.



([Source (https://commons.wikimedia.org/wiki/File:Linalg_projection_4.png)](https://commons.wikimedia.org/wiki/File:Linalg_projection_4.png))

Letting $\mathbf{v}^* = \alpha^* \mathbf{u}_1$, the geometrical condition above translates into

$$0 = \langle \mathbf{u}_1, \mathbf{v} - \mathbf{v}^* \rangle = \langle \mathbf{u}_1, \mathbf{v} - \alpha^* \mathbf{u}_1 \rangle = \langle \mathbf{u}_1, \mathbf{v} \rangle - \alpha^* \langle \mathbf{u}_1, \mathbf{u}_1 \rangle = \langle \mathbf{u}_1, \mathbf{v} \rangle - \alpha^*$$

so

$$\mathbf{v}^* = \langle \mathbf{u}_1, \mathbf{v} \rangle \mathbf{u}_1.$$

By *Pythagoras*, we then have for any $\alpha \in \mathbb{R}$

$$\begin{aligned} \|\mathbf{v} - \alpha \mathbf{u}_1\|^2 &= \|\mathbf{v} - \mathbf{v}^* + \mathbf{v}^* - \alpha \mathbf{u}_1\|^2 \\ &= \|\mathbf{v} - \mathbf{v}^* + (\alpha^* - \alpha) \mathbf{u}_1\|^2 \\ &= \|\mathbf{v} - \mathbf{v}^*\|^2 + \|(\alpha^* - \alpha) \mathbf{u}_1\|^2 \\ &\geq \|\mathbf{v} - \mathbf{v}^*\|^2. \end{aligned}$$

That confirms the optimality of $\mathbf{v}^*$. The argument in this example carries through in higher dimension, as we show next. ◁

Definition (Orthogonal Projection on an Orthonormal List): Let $\mathbf{q}_1, \ldots, \mathbf{q}_m$ be an orthonormal list. The orthogonal projection of $\mathbf{v} \in V$ on $\{\mathbf{q}_i\}_{i=1}^m$ is defined as

$$(\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v} = \sum_{j=1}^m \langle \mathbf{v}, \mathbf{q}_j \rangle\, \mathbf{q}_j.$$

⊲

Definition and Theorem (Orthogonal Projection): Let $\mathcal{U} \subseteq V$ be a linear subspace with orthonormal basis $\mathbf{q}_1, \ldots, \mathbf{q}_m$ and let $\mathbf{v} \in V$. Then $(\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v} \in \mathcal{U}$ and, for any $\mathbf{u} \in \mathcal{U}$,

$$(*) \quad \left\langle \mathbf{v} - (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v}, \mathbf{u} \right\rangle = 0$$

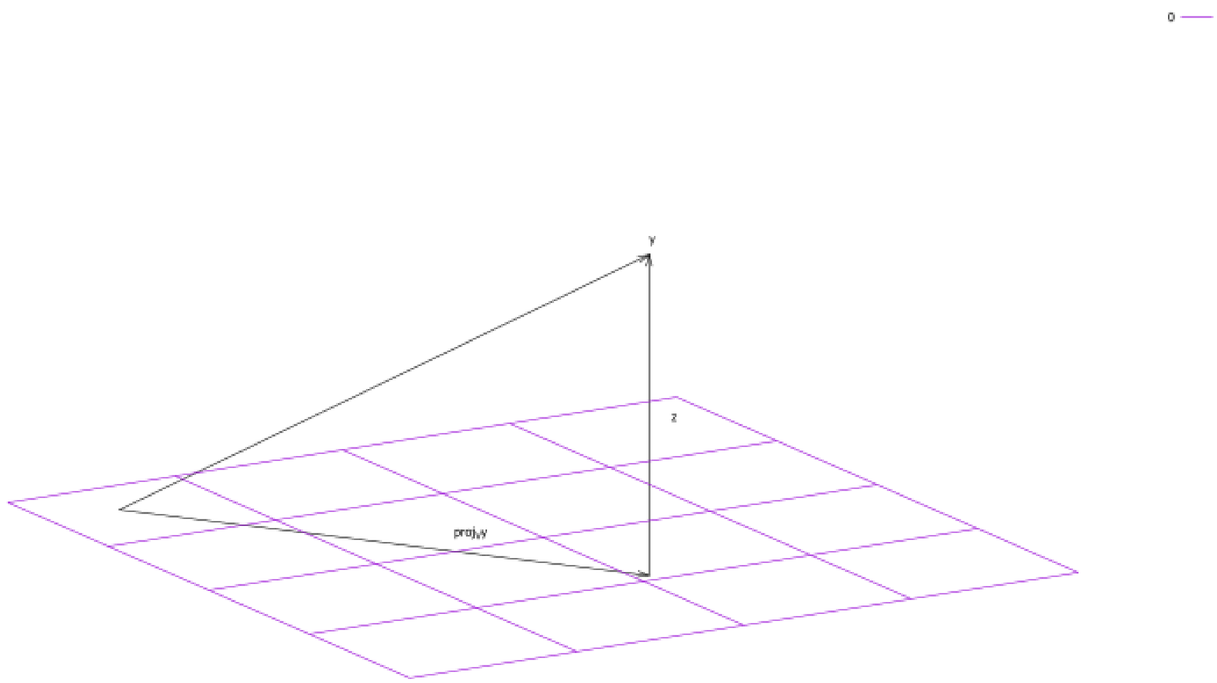
and

$$(**) \quad \|\mathbf{v} - (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v}\| \leq \|\mathbf{v} - \mathbf{u}\|.$$

Furthermore, if $\mathbf{u} \in \mathcal{U}$ and the inequality above is an equality, then $\mathbf{u} = (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\mathbf{v}$. Hence, for any orthonormal basis $\mathbf{q}'_1, \ldots, \mathbf{q}'_m$ of $\mathcal{U}$, it holds that

$$(***) \quad \mathcal{P}_{\mathcal{U}}\mathbf{v} = (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v} = (\mathcal{P}_{\{\mathbf{q}'_i\}_{i=1}^m})\,\mathbf{v},$$

where the first equality is a definition. We refer to $\mathcal{P}_{\mathcal{U}}\mathbf{v}$ as the orthogonal projection of $\mathbf{v}$ on $\mathcal{U}$.



[(Source) (https://commons.wikimedia.org/wiki/File:Ortho_projection.svg)](https://commons.wikimedia.org/wiki/File:Ortho_projection.svg)

Proof: By definition, it is immediate that $(\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v} \in \mathrm{span}(\{\mathbf{q}_i\}_{i=1}^m) = \mathcal{U}$. We first prove $(*)$. We can write any $\mathbf{u} \in \mathcal{U}$ as $\sum_{j=1}^m \alpha_j \mathbf{q}_j$ for some α_j's. Then

$$\left\langle \mathbf{v} - \sum_{j=1}^m \langle \mathbf{v}, \mathbf{q}_j \rangle\, \mathbf{q}_j, \sum_{j=1}^m \alpha_j' \mathbf{q}_j \right\rangle = \sum_{j=1}^m \langle \mathbf{v}, \mathbf{q}_j \rangle\, \alpha_j' - \sum_{j=1}^m \alpha_j' \langle \mathbf{v}, \mathbf{q}_j \rangle = 0$$

where we used the orthonormality of the $\mathbf{q}_j$'s in the leftmost equality.

To prove $(**)$, note that for any $\mathbf{u} \in \mathcal{U}$ the vector $\mathbf{u}' = (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v} - \mathbf{u}$ is also in $\mathcal{U}$. Hence by $(*)$ and *Pythagoras*,

$$\begin{aligned} \|\mathbf{v} - \mathbf{u}\|^2 &= \|\mathbf{v} - (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v} + (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v} - \mathbf{u}\|^2 \\ &= \|\mathbf{v} - (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v}\|^2 + \|(\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v} - \mathbf{u}\|^2 \\ &\geq \|\mathbf{v} - (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v}\|^2. \end{aligned}$$

Furthermore, equality holds only if $\|(\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v} - \mathbf{u}\|^2 = 0$ which holds only if $\mathbf{u} = (\mathcal{P}_{\{\mathbf{q}_i\}_{i=1}^m})\,\mathbf{v}$ by the point-separating property of the 2-norm. The rightmost equality in $(***)$ follows from swapping $\{\mathbf{q}_i'\}_{i=1}^m$ for $\{\mathbf{q}_i\}_{i=1}^m$. $\square$

The *Orthogonal Projection Theorem* implies that any $\mathbf{v} \in V$ can be decomposed into its orthonogonal projection onto $\mathcal{U}$ and a vector orthogonal to it.

Definition (Orthogonal Complement): Let $\mathcal{U} \subseteq V$ be a linear subspace. The orthogonal complement of $\mathcal{U}$, denoted $\mathcal{U}^\perp$, is defined as

$$\mathcal{U}^\perp = \{\mathbf{w} \in V \ : \ \langle \mathbf{w}, \mathbf{u} \rangle = 0, \forall \mathbf{u} \in \mathcal{U}\}.$$

$\triangleleft$

Exercise: Establish that $\mathcal{U}^\perp$ is a linear subspace. $\triangleleft$

Lemma (Orthogonal Decomposition): Let $\mathcal{U} \subseteq V$ be a linear subspace and let $\mathbf{v} \in V$. Then $\mathbf{v}$ can be decomposed as $(\mathbf{v} - \mathcal{P}_{\mathcal{U}}\mathbf{v}) + \mathcal{P}_{\mathcal{U}}\mathbf{v}$ where $(\mathbf{v} - \mathcal{P}_{\mathcal{U}}\mathbf{v}) \in \mathcal{U}^\perp$ and $\mathcal{P}_{\mathcal{U}}\mathbf{v} \in \mathcal{U}$.

Proof: Immediate consequence of the previous theorem. $\square$

The map $\mathcal{P}_{\mathcal{U}}$ is linear, that is, $\mathcal{P}_{\mathcal{U}}(\alpha\,\mathbf{x} + \mathbf{y}) = \alpha\,\mathcal{P}_{\mathcal{U}}\mathbf{x} + \mathcal{P}_{\mathcal{U}}\mathbf{y}$ for all $\alpha \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Indeed,

$$\mathcal{P}_{\mathcal{U}}(\alpha\,\mathbf{x} + \mathbf{y}) = \sum_{j=1}^m \langle \alpha\,\mathbf{x} + \mathbf{y}, \mathbf{q}_j \rangle\,\mathbf{q}_j = \sum_{j=1}^m \left\{\alpha\,\langle \mathbf{x}, \mathbf{q}_j \rangle + \langle \mathbf{y}, \mathbf{q}_j \rangle\right\}\mathbf{q}_j = \alpha\,\mathcal{P}_{\mathcal{U}}\mathbf{x} + \mathcal{P}_{\mathcal{U}}\mathbf{y}.$$

Therefore it can be encoded as an $n \times n$ matrix P. Let

$$Q = \begin{pmatrix} | & & | \\ \mathbf{q}_1 & \dots & \mathbf{q}_m \\ | & & | \end{pmatrix}$$

and note that computing

$$Q^T\mathbf{v} = \begin{pmatrix} \langle \mathbf{v}, \mathbf{q}_1 \rangle \\ \cdots \\ \langle \mathbf{v}, \mathbf{q}_m \rangle \end{pmatrix}$$

lists the coefficients in the expansion of $\mathcal{P}_{\mathcal{U}}\mathbf{v}$ over the basis $\mathbf{q}_1, \dots, \mathbf{q}_m$.

Hence we see that

$$P = QQ^T.$$

This is not to be confused with

$$Q^TQ = \begin{pmatrix} \langle \mathbf{q}_1, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{q}_1, \mathbf{q}_m \rangle \\ \langle \mathbf{q}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{q}_2, \mathbf{q}_m \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{q}_m, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{q}_m, \mathbf{q}_m \rangle \end{pmatrix} = I_{m\times m}$$

where $I_{m\times m}$ denotes the $m \times m$ identity matrix.

Exercise: Let $\mathcal{U} \subseteq V$ be a linear subspace and let $\mathbf{v} \in \mathcal{U}$. Show that $\mathcal{P}_{\mathcal{U}}\mathbf{v} = \mathbf{v}$. ◁

2.3 Gram-Schmidt

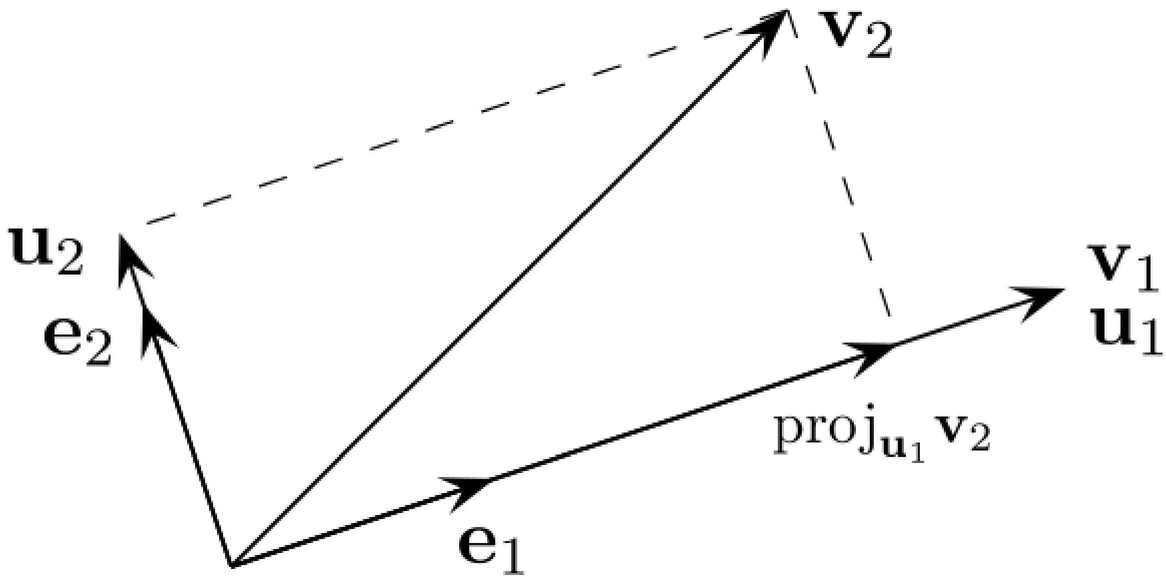
We have some business left over: constructing orthonormal bases. Let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be linearly independent. We use the Gram-Schmidt algorithm to obtain an orthonormal basis of $\mathrm{span}(\mathbf{a}_1, \ldots, \mathbf{a}_m)$. The process takes advantage of the properties of the orthogonal projection derived above. In essence we add the vectors $\mathbf{a}_i$ one by one, but only after taking out their orthogonal projection on the previously included vectors. The outcome spans the same subspace and the *Orthogonal Decomposition Lemma* ensures orthogonality.

Theorem (Gram-Schmidt): Let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be linearly independent. Then there exists an orthonormal basis $\mathbf{q}_1, \ldots, \mathbf{q}_m$ of $\mathrm{span}(\mathbf{a}_1, \ldots, \mathbf{a}_m)$.

Proof idea: Suppose first that $m = 1$. In that case, all that needs to be done is to divide $\mathbf{a}_1$ by its norm to obtain a unit vector whose span is the same as $\mathbf{a}_1$, that is, we set $\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$.

Suppose now that $m = 2$. We first let $\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$ as in the previous case. Then we subtract from $\mathbf{a}_2$ its projection on $\mathbf{q}_1$, that is, we set $\mathbf{v}_2 = \mathbf{a}_2 - \langle \mathbf{q}_1, \mathbf{a}_2 \rangle \mathbf{q}_1$. By the *Orthogonal Projection Theorem*, $\mathbf{v}_2$ is orthogonal to $\mathbf{q}_1$. Moreover, because $\mathbf{a}_2$ is a linear combination of $\mathbf{q}_1$ and $\mathbf{v}_2$, we have $\mathrm{span}(\mathbf{q}_1, \mathbf{v}_2) = \mathrm{span}(\mathbf{a}_1, \mathbf{a}_2)$. It remains to divide by the norm of the resulting vector: $\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$.

For general m, we proceed similarly but project onto the subspace spanned by the previously added vectors at each step.



(Source (https://en.wikipedia.org/wiki/Gram–Schmidt_process))

Proof: The inductive step is the following. Assume that we have constructed orthonormal vectors $\mathbf{q}_1, \dots, \mathbf{q}_{j-1}$ such that

$$U_{j-1} := \mathrm{span}(\mathbf{q}_1, \dots, \mathbf{q}_{j-1}) = \mathrm{span}(\mathbf{a}_1, \dots, \mathbf{a}_{j-1}).$$

By the *Properties of Orthonormal Lists*, $\{\mathbf{q}\}_{i=1}^{j-1}$ forms an orthonormal basis for U_{j-1}, so we can compute the orthogonal projection of $\mathbf{a}_j$ on U_{j-1} as

$$\mathcal{P}_{U_{j-1}} \mathbf{a}_j = \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i.$$

where we defined $r_{ij} = \langle \mathbf{q}_i, \mathbf{a}_j \rangle$. And we set

$$\mathbf{v}_j = \mathbf{a}_j - \mathcal{P}_{U_{j-1}} \mathbf{a}_j \quad \text{and} \quad \mathbf{q}_j = \frac{\mathbf{v}_j}{\|\mathbf{v}_j\|}.$$

Here we used that $\|\mathbf{v}_j\| > 0$: indeed otherwise $\mathbf{a}_j$ would be equal to its projection $\mathcal{P}_{U_{j-1}} \mathbf{a}_j \in \mathrm{span}(\mathbf{a}_1, \dots, \mathbf{a}_{j-1})$ which would contradict linear independence of the $\mathbf{a}_k$'s.

By the *Orthogonal Projection Theorem*, $\mathbf{q}_j$ is orthogonal to $\mathrm{span}(\mathbf{q}_1, \ldots, \mathbf{q}_{j-1})$ and, unrolling the calculations above, $\mathbf{a}_j$ can be re-written as the following linear combination of $\mathbf{q}_1, \ldots, \mathbf{q}_j$

$$\begin{aligned}
\mathbf{a}_j &= \mathcal{P}_{U_{j-1}} \mathbf{a}_j + \mathbf{v}_j \\
&= \mathcal{P}_{U_{j-1}} \mathbf{a}_j + \|\mathbf{v}_j\| \mathbf{q}_j \\
&= \mathcal{P}_{U_{j-1}} \mathbf{a}_j + \|\mathbf{a}_j - \mathcal{P}_{U_{j-1}} \mathbf{a}_j\| \mathbf{q}_j \\
&= \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i + \left\| \mathbf{a}_j - \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i \right\| \mathbf{q}_j \\
&= \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i + r_{jj} \mathbf{q}_j,
\end{aligned}$$

where we defined $r_{jj} = \left\| \mathbf{a}_j - \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i \right\| = \|\mathbf{v}_j\|$.

Hence $\mathbf{q}_1, \ldots, \mathbf{q}_j$ forms an orthonormal list with $\mathrm{span}(\mathbf{a}_1, \ldots, \mathbf{a}_j) \subseteq \mathrm{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j)$. The opposite inclusion holds by construction. Moreover, because $\mathbf{q}_1, \ldots, \mathbf{q}_j$ are orthonormal, they are linearly independent by the *Properties of Orthonormal Lists* so must form a basis of their span. So induction goes through. □

NUMERICAL CORNER We implement the Gram-Schmidt algorithm in Julia. For reasons that will become clear in a future notebook, we output the $\mathbf{q}_j$'s and r_{ij}'s, each in matrix form.

In [1]:
```
# Julia version: 1.5.1
using LinearAlgebra
```

In [2]:
```
function mmids_gramschmidt(A)
    n,m = size(A)
    Q = zeros(Float64,n,m)
    R = zeros(Float64,m,m)
    for j = 1:m
        v = A[:,j]
        for i = 1:j-1
            R[i,j] = dot(Q[:,i],A[:,j])
            v -= R[i,j]*Q[:,i]
        end
        R[j,j] = norm(v)
        Q[:,j] = v/R[j,j]
    end
    return Q,R
end
```

Out[2]:
```
mmids_gramschmidt (generic function with 1 method)
```

Let's try a simple example.

In [3]:
```
w1, w2 = [1,0,1], [0,1,1]
A = hcat(w1,w2)
```

Out[3]:
```
3×2 Array{Int64,2}:
 1  0
 0  1
 1  1
```

In [4]:
```
Q,R = mmids_gramschmidt(A);
```

In [5]:
```
Q
```

Out[5]:
```
3×2 Array{Float64,2}:
 0.707107  -0.408248
 0.0        0.816497
 0.707107   0.408248
```

In [6]:
```
R
```

Out[6]:
```
2×2 Array{Float64,2}:
 1.41421  0.707107
 0.0      1.22474
```

Exercise: Let $B \in \mathbb{R}^{n \times m}$ be a matrix. Show that there exist orthonormal bases of $\mathrm{col}(B)$ and $\mathrm{null}(B)$. $\lhd$

Exercise: Let $\mathcal{W}$ be a linear subspace of $\mathbb{R}^d$ and let $\mathbf{w}_1, \dots, \mathbf{w}_k$ be an orthonormal basis of $\mathcal{W}$. Show that there exists an orthonormal basis of $\mathbb{R}^d$ that includes the $\mathbf{w}_i$'s. $\lhd$

Lemma (Nullity): Let $B \in \mathbb{R}^{n \times m}$ be a matrix of rank $= k$. Then $\dim(\text{null}(B)) = n - k$. Put differently,

$$\dim(\text{col}(B)) + \dim(\text{null}(B)) = n.$$

For a proof, see Wikipedia (https://en.wikipedia.org/wiki/Rank%E2%80%93nullity_theorem#First_proof).