Inner products, lengths, and orthogonality

For any $\vec{u}, \vec{v} \in \mathbb{R}^n$, we define $\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i v_i \in \mathbb{R}$. We note that this can be expressed using conventional vector multiplication: the **dot product** is equal to $\vec{u}^T \vec{v}$ and $\vec{v}^T \vec{u}$.

Properties

Given $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, we know that:

1. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
2. $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$
3. $(c\vec{u}) \cdot \vec{w} = c(\vec{u} \cdot \vec{w})$
4. $\vec{u} \cdot \vec{u} \geq 0$

Thus it is clear that the dot product forms a linear map from $\mathbb{R}^n$ to $\mathbb{R}$. This is also evident from the fact that, as mentioned, the dot product is equivalent to matrix multiplication by a transpose.

Define the *norm* or *length* of $\vec{u}$ as $||\vec{u}|| = \sqrt{\vec{u} \cdot \vec{u}}$. We can see that scaling is linear, but the length must remain positive: $||c\vec{u}|| = |c|||\vec{u}||$. We can define a **unit vector** as a vector with length 1. We can see that for any $\vec{v} \in \mathbb{R}^n$, a unit vector is $\frac{1}{||\vec{v}||}\vec{v}$. We can also define the **distance** $dist(\vec{u}, \vec{v}) = ||\vec{u} - \vec{v}||$.

Orthogonality

Two vectors in $\mathbb{R}^2$ are *orthogonal* if the angle between them is $\frac{\pi}{2}$. We can extend this definition to $\mathbb{R}^n$ by generalising it algebraically. It is clear that, if $\vec{u}$ and $\vec{v}$ are orthogonal, then:

$$\begin{gathered} dist(\vec{u}, \vec{v}) = dist(\vec{u}, -\vec{v}) \\ \iff ||\vec{u} - \vec{v}|| = ||\vec{u} + \vec{v}|| \iff ||\vec{u} - \vec{v}||^2 = ||\vec{u} + \vec{v}||^2 \\ \iff (\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) = (\vec{u} + \vec{v}) \cdot (\vec{u} + \vec{v}) \\ \iff ||\vec{u}||^2 - 2(\vec{u} \cdot \vec{v}) + ||\vec{v}||^2 = ||\vec{u}||^2 + 2(\vec{u} \cdot \vec{v}) + ||\vec{v}||^2 \\ \iff \vec{u} \cdot \vec{v} = 0 \end{gathered}$$

Thus, we have the formal definition of orthogonality: two vectors in $\mathbb{R}^n$ are orthogonal iff their dot product is 0.

$$\boxed{\text{Theorem: } \vec{u} \cdot \vec{v} = 0 \iff ||\vec{u} \pm \vec{v}|| = ||\vec{u}|| + ||\vec{v}||}$$

Proof. By definition, $||\vec{u} \pm \vec{v}|| = ||\vec{u}|| + ||\vec{v}|| \pm 2(\vec{u} \cdot \vec{v})$. Iff $\vec{u} \cdot \vec{v} = 0$, we are left with the original statement. $\square$

We can also define orthogonality between vectors and vector spaces. Given $W \subset \mathbb{R}^n$, $\vec{z} \in \mathbb{R}^n$ is orthogonal to W ($\vec{z} \perp W$) iff $\vec{z} \perp \vec{v} \; \forall \; \vec{v} \in W$. We define the **orthogonal complement** to W as $W^\perp = \{\vec{z} \in \mathbb{R}^n : \vec{z} \perp W\}$. It can be shown that $W^\perp$ is a vector subspace.

A remark: given $\vec{A}_{m \times n}, \vec{u} \in \mathbb{R}^n, \vec{v} \in \mathbb{R}^m$, $(\vec{A}\vec{u}) \cdot \vec{v} = \vec{u} \cdot \vec{A}^T \vec{v}$, even though the dot products are happening in different dimensions!

Proof. $(\vec{A}\vec{u}) \cdot \vec{v} = \vec{u}^T((\vec{A}^T \vec{v}) = \vec{u} \cdot \vec{A}^T \vec{v}$ $\square$

We can use this result to show that

Theorem: If the matrix $\vec{U}$ defines a map from orthogonal columns, that is, $\vec{U}\vec{U}^T = \vec{I}$, then $||\vec{U}\vec{x}|| = ||\vec{x}||$, and $\vec{U}\vec{x} \cdot \vec{U}\vec{y} = \vec{x} \cdot \vec{y}$. Also, $\vec{x} \perp \vec{y} \iff \vec{U}\vec{x} \perp \vec{U}\vec{y}$.

Proof. The third and first properties follow directly from the second and their proofs are trivial. To prove the second property, we can use the definition of the dot product: $\vec{U}\vec{x} \cdot \vec{U}\vec{y} = (\vec{U}\vec{x})^T \vec{U}\vec{y} = \vec{x}^T \vec{U}^T \vec{U}\vec{y} = \vec{x}^T \vec{y} = \vec{x} \cdot \vec{y}$. $\square$

Orthogonal projections

Given $W \subset \mathbb{R}^n, \vec{y} \in \mathbb{R}^n$, we define an *orthogonal projection of $\vec{y}$ in W* as a vector $\hat{y} \in W$ such that $\vec{y} - \hat{y} \perp W$. We will see that $proj_W(\vec{y})$ is the "closest" point of W to $\vec{y}$.

Theorem: Any $\vec{y} \in \mathbb{R}^n$ can be written uniquely as $\vec{y} = \hat{y} + \vec{z}$ where $\hat{y} \in W, \vec{z} = \vec{y} - \hat{y} \perp W$.

Proof. Let $\{\vec{u}_1, \vec{u}_2, \cdots \vec{u}_p\}$ be an orthonormal basis for W. Define $proj_W(\vec{y}) = \hat{y} = \sum_{i=1}^{p} \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i$. It is obvious that $\hat{y} \in W$. Now, $\vec{y} - \hat{y} \perp W \iff \vec{y} - \hat{y} \perp \vec{u}_j \; \forall \; 1 \leq j \leq p$. We can apply the dot product: $(\vec{y} - \hat{y}) \cdot \vec{u}_j = \vec{y} \cdot \vec{u}_j - \left(\sum_{i=1}^{p} \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i\right) \cdot \vec{u}_j = \vec{y} \cdot \vec{u}_j - \vec{y} \cdot \vec{u}_j = 0$.

Proof of uniqueness: assume that there is another representation of $\vec{y}$ in this form, that is, $\vec{y} = \hat{y}' + \vec{z'}, \hat{y}' \in W, \vec{z'} = \vec{y} - \hat{y}' \in W^\perp$. Then, $\hat{y} + \vec{z} = \hat{y}' + \vec{z'}$. Define, in two ways, $\vec{v} = \hat{y} - \hat{y}' = \vec{z'} - \vec{z}$. We can see that the first definition of $\vec{v}$ is in W, and the second definition is in $W^{\perp}$. Therefore, $\vec{v} \cdot \vec{v} = 0$, and so $\vec{v} = 0$. Therefore, $\hat{y}$ and $\vec{z}$ are both unique. $\square$

We can note that if $\vec{y} \in W$, then $proj_W(\vec{y}) = \vec{y}$, and so $\vec{y} = proj_W(\vec{y}) = \sum_{i=1}^{p} \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i$. This goes back to the definition of $\vec{y}$ as a linear combination of orthonormal vectors.

Theorem: $||\vec{y} - \hat{y}|| \leq ||\vec{y} - \vec{v}|| \; \forall \; \vec{v} \in W$. Equality holds iff $\hat{y} = \vec{v}$.

Proof. Let $\vec{v} \in W, \vec{v} \neq \hat{y}$. Then, since $\vec{y} - \hat{y} \perp W$, we know that $(\vec{y} - haty) \perp (\hat{y} - \vec{v})$ as $(\hat{y} - \vec{v}) \in W$. Therefore, $||\vec{y} - \vec{v}||^2 = ||\vec{y} - \hat{y} + \hat{y} - \vec{v}||^2 = ||\vec{y} - \hat{y}||^2 + ||\hat{y} - \vec{v}||^2$. So $||\vec{y} - \hat{y}|| \leq ||\vec{y} - \vec{v}||$. $\square$

Theorem: Any $W \in \mathbb{R}^n$ has an orthonormal basis.

Proof. This is not a complete proof, but the presentation of the Gram-Schmidt algorithm for generating an orthonormal basis for any subspace of $mathbbR^n$.

Let $\{\vec{x}_1, \vec{x}_2, \cdot \vec{x}_p\}$ be a basis for W. We can create another set of vectors $\{\vec{v}_1, \vec{v}_2, \cdot \vec{v}_p\}$ that spans W and is orthonormal. We set $\vec{v}_1 = \vec{x}_1$ and $\vec{v}_2 = \vec{x}_2 - proj_{\vec{v}_1}(\vec{x}_2)$. These vectors are by definition orthogonal. Now, we can set $\vec{v}_3 = \vec{x}_3 - proj_{span \vec{v}_1, \vec{v}_2}(\vec{x}_3)$, which is clearly orthogonal to both $\vec{v}_1$ and $\vec{v}_2$. We continue this pattern up to $\vec{v}_p$.

In other words:

$$\vec{v}_1 = \vec{x}_1$$

$$\vec{v}_2 = \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1$$

$$\vec{v}_3 = \vec{x}_3 - \left(\frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 + \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2 \right)$$

$$\vec{v}_p = \sum_{i=1}^{p-1} \frac{\vec{x}_p \cdot \vec{v}_i}{\vec{v}_i \cdot \vec{v}_i} \vec{v}_i$$

The Gram-Schmidt theorem proves that this new set also forms a basis for W. $\square$

Important ideas in the proof:

1. $\{\vec{v}_1, \vec{v}_2, \cdot \vec{v}_i\} \, 1 \leq i \leq p$ is orthogonal
2. $span \vec{v}_1, \vec{v}_2, \cdot \vec{v}_i = span \vec{x}_1, \vec{x}_2, \cdot \vec{x}_i 1 \leq i \leq p$

QR factorisation

Theorem: Let $\overrightarrow{A}_{m\times p}$ be a matrix with linearly independent columns. Then, there exist two matrices $\overrightarrow{Q}_{m\times p}$ and $\overrightarrow{R}_{p\times p}$ such that $\overrightarrow{Q}\overrightarrow{Q}^T = \overrightarrow{I}_m$ (that is, the columns of $\overrightarrow{Q}$ are orthonormal), $\operatorname{col}\overrightarrow{A} = \operatorname{col}\overrightarrow{Q}$, $\overrightarrow{R}$ is upper triangular with positive diagonal entries and invertible, and $\overrightarrow{A} = \overrightarrow{Q}\overrightarrow{R}$.

Since the columns of $\overrightarrow{A}$ are linearly independent, we know that they form a basis for $\operatorname{col}\overrightarrow{A}$. We can form $\overrightarrow{Q}$ from the orthonormal basis for $\operatorname{col}\overrightarrow{A}$ using Gram-Schmidt, where each column of $\overrightarrow{Q}\overrightarrow{q}_i = \frac{\overrightarrow{v}_i}{||\overrightarrow{v}_i||}$ where $\overrightarrow{v}_i$ are the columns of the orthonormal basis. Since the columns of $\overrightarrow{Q}$ are orthonormal, we can compute $\overrightarrow{R} = \overrightarrow{Q}^T\overrightarrow{A}$.