**Transpose & Dot Product**

**Def:** The *transpose* of an \( m \times n \) matrix \( A \) is the \( n \times m \) matrix \( A^T \) whose columns are the rows of \( A \).

So: The columns of \( A^T \) are the rows of \( A \). The rows of \( A^T \) are the columns of \( A \).

**Example:** If \( A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \), then \( A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \).

**Convention:** From now on, vectors \( \mathbf{v} \in \mathbb{R}^n \) will be regarded as “columns” (i.e.: \( n \times 1 \) matrices). Therefore, \( \mathbf{v}^T \) is a “row vector” (a \( 1 \times n \) matrix).

**Observation:** Let \( \mathbf{v}, \mathbf{w} \in \mathbb{R}^n \). Then \( \mathbf{v}^T \mathbf{w} = \mathbf{v} \cdot \mathbf{w} \). This is because:

\[
\mathbf{v}^T \mathbf{w} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = v_1w_1 + \cdots + v_nw_n = \mathbf{v} \cdot \mathbf{w}.
\]

Where theory is concerned, the key property of transposes is the following:

**Prop 18.2:** Let \( A \) be an \( m \times n \) matrix. Then for \( \mathbf{x} \in \mathbb{R}^n \) and \( \mathbf{y} \in \mathbb{R}^m \):

\[
(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A^T \mathbf{y}).
\]

Here, \( \cdot \) is the dot product of vectors.

**Extended Example**

Let \( A \) be a \( 5 \times 3 \) matrix, so \( A : \mathbb{R}^3 \to \mathbb{R}^5 \).

- \( \text{null}(A) \) is a subspace of ______
- \( \text{range}(A) \) is a subspace of ______

The transpose \( A^T \) is a _____ matrix, so \( A^T : ___ \to ___ \)

- \( \text{range}(A^T) \) is a subspace of ______
- \( \text{null}(A^T) \) is a subspace of ______

**Observation:** Both \( \text{range}(A^T) \) and \( \text{null}(A) \) are subspaces of ______. Might there be a geometric relationship between the two? (No, they’re not equal.) Hm...

**Also:** Both \( \text{null}(A^T) \) and \( \text{range}(A) \) are subspaces of ______. Might there be a geometric relationship between the two? (Again, they’re not equal.) Hm...
Orthogonal Complements

**Def:** Let $V \subset \mathbb{R}^n$ be a subspace. The **orthogonal complement** of $V$ is the set

$$V^\perp = \{ x \in \mathbb{R}^n \mid x \cdot v = 0 \text{ for every } v \in V \}.$$  

So, $V^\perp$ consists of the vectors which are orthogonal to every vector in $V$.

**Fact:** If $V \subset \mathbb{R}^n$ is a subspace, then $V^\perp \subset \mathbb{R}^n$ is a subspace.

**Examples in $\mathbb{R}^3$:**
- The orthogonal complement of $V = \{0\}$ is $V^\perp = \mathbb{R}^3$
- The orthogonal complement of $V = \{z\text{-axis}\}$ is $V^\perp = \{xy\text{-plane}\}$
- The orthogonal complement of $V = \{xy\text{-plane}\}$ is $V^\perp = \{z\text{-axis}\}$
- The orthogonal complement of $V = \mathbb{R}^3$ is $V^\perp = \{0\}$

**Examples in $\mathbb{R}^4$:**
- The orthogonal complement of $V = \{0\}$ is $V^\perp = \mathbb{R}^4$
- The orthogonal complement of $V = \{w\text{-axis}\}$ is $V^\perp = \{xyz\text{-space}\}$
- The orthogonal complement of $V = \{zw\text{-plane}\}$ is $V^\perp = \{xy\text{-plane}\}$
- The orthogonal complement of $V = \{xyz\text{-space}\}$ is $V^\perp = \{w\text{-axis}\}$
- The orthogonal complement of $V = \mathbb{R}^4$ is $V^\perp = \{0\}$

**Prop 19.3-19.4-19.5:** Let $V \subset \mathbb{R}^n$ be a subspace. Then:

(a) $\dim(V) + \dim(V^\perp) = n$
(b) $(V^\perp)^\perp = V$
(c) $V \cap V^\perp = \{0\}$
(d) $V + V^\perp = \mathbb{R}^n$.

Part (d) means: “Every vector $x \in \mathbb{R}^n$ can be written as a sum $x = v + w$ where $v \in V$ and $w \in V^\perp$.”

Also, it turns out that the expression $x = v + w$ is unique: that is, there is only one way to write $x$ as a sum of a vector in $V$ and a vector in $V^\perp$. 
Meaning of $C(A^T)$ and $N(A^T)$

**Q:** What does $C(A^T)$ mean? Well, the columns of $A^T$ are the rows of $A$. So:

\[
C(A^T) = \text{column space of } A^T = \text{span of columns of } A^T = \text{span of rows of } A.
\]

For this reason: We call $C(A^T)$ the **row space** of $A$.

**Q:** What does $N(A^T)$ mean? Well:

\[
x \in N(A^T) \iff A^T x = 0 \iff (A^T x)^T = 0^T \iff x^T A = 0^T.
\]

So, for an $m \times n$ matrix $A$, we see that: $N(A^T) = \{x \in \mathbb{R}^m \mid x^T A = 0^T\}$.

For this reason: We call $N(A^T)$ the **left null space** of $A$.

**Relationships among the Subspaces**

**Theorem:** Let $A$ be an $m \times n$ matrix. Then:

- $C(A^T) = N(A)^\perp$
- $N(A^T) = C(A)^\perp$

**Corollary:** Let $A$ be an $m \times n$ matrix. Then:

- $C(A) = N(A^T)^\perp$
- $N(A) = C(A^T)^\perp$

**Prop 18.3:** Let $A$ be an $m \times n$ matrix. Then $\text{rank}(A) = \text{rank}(A^T)$.

**Motivating Questions for Reading**

**Problem 1:** Let $b \in C(A)$. So, the system of equations $Ax = b$ does have solutions, possibly infinitely many.

Q: What is the solution $x$ of $Ax = b$ with $\|x\|$ the smallest?

**Problem 2:** Let $b \notin C(A)$. So, the system of equations $Ax = b$ does not have any solutions. In other words, $Ax - b \neq 0$.

Q: What is the vector $x$ that minimizes the error $\|Ax - b\|$? That is, what is the vector $x$ that comes closest to being a solution to $Ax = b$?
Orthogonal Projection

**Def:** Let \( V \subset \mathbb{R}^n \) be a subspace. Then every vector \( x \in \mathbb{R}^n \) can be written uniquely as

\[
x = v + w, \text{ where } v \in V \text{ and } w \in V^\perp.
\]

The **orthogonal projection** onto \( V \) is the function \( \text{Proj}_V : \mathbb{R}^n \rightarrow \mathbb{R}^n \) given by: \( \text{Proj}_V(x) = v \). (Note that \( \text{Proj}_{V^\perp}(x) = w \).)

**Prop 20.1:** Let \( V \subset \mathbb{R}^n \) be a subspace. Then:

\[
\text{Proj}_V + \text{Proj}_{V^\perp} = I_n.
\]

Of course, we already knew this: We have \( x = v + w = \text{Proj}_V(x) + \text{Proj}_{V^\perp}(x) \).

**Formula:** Let \( \{v_1, \ldots, v_k\} \) be a basis of \( V \subset \mathbb{R}^n \). Let \( A \) be the \( n \times k \) matrix

\[
A = \begin{bmatrix}
v_1 \\
\vdots \\
v_k
\end{bmatrix}.
\]

Then:

\[
\text{Proj}_V = A(A^T A)^{-1} A^T. \quad (\ast)
\]

**Geometry Observations:** Let \( V \subset \mathbb{R}^n \) be a subspace, and \( x \in \mathbb{R}^n \) a vector.

1. The distance from \( x \) to \( V \) is: \( \|\text{Proj}_{V^\perp}(x)\| = \|x - \text{Proj}_V(x)\| \).
2. The vector in \( V \) that is closest to \( x \) is: \( \text{Proj}_V(x) \).

*Derivation of \((\ast)\):* Notice \( \text{Proj}_V(x) \) is a vector in \( V = \text{span}(v_1, \ldots, v_k) = C(A) = \text{Range}(A) \), and therefore \( \text{Proj}_V(x) = Ay \) for some vector \( y \in \mathbb{R}^k \).

Now notice that \( x - \text{Proj}_V(x) = x - Ay \) is a vector in \( V^\perp = C(A)^\perp = N(A^T) \), which means that \( A^T(x - Ay) = 0 \), which means \( A^T x = A^T Ay \).

Now, it turns out that our matrix \( A^T A \) is invertible (proof in L20), so we get \( y = (A^T A)^{-1} A^T x \).

Thus, \( \text{Proj}_V(x) = Ay = A(A^T A)^{-1} A^T x \). \( \diamond \)
Minimum Magnitude Solution

Prop 19.6: Let $b \in C(A)$ (so $Ax = b$ has solutions). Then there exists exactly one vector $x_0 \in C(A^T)$ with $Ax_0 = b$.

And: Among all solutions of $Ax = b$, the vector $x_0$ has the smallest length.

In other words: There is exactly one vector $x_0$ in the row space of $A$ which solves $Ax = b$ – and this vector is the solution of smallest length.

To Find $x_0$: Start with any solution $x$ of $Ax = b$. Then

$$x_0 = \text{Proj}_{C(A^T)}(x).$$

Least Squares Approximation

Idea: Suppose $b \notin C(A)$. So, $Ax = b$ has no solutions, so $Ax - b \neq 0$.

We want to find the vector $x^*$ which minimizes the error $\|Ax^* - b\|$. That is, we want the vector $x^*$ for which $Ax^*$ is the closest vector in $C(A)$ to $b$.

In other words, we want the vector $x^*$ for which $Ax^* - b$ is orthogonal to $C(A)$. So, $Ax^* - b \in C(A)^\perp = N(A^T)$, meaning that $A^T(Ax^* - b) = 0$, i.e.:

$$A^T A x^* = A^T b.$$ 

Quadratic Forms (Intro)

Given an $m \times n$ matrix $A$, we can regard it as a linear transformation $T: \mathbb{R}^n \to \mathbb{R}^m$. In the special case where the matrix $A$ is a symmetric matrix, we can also regard $A$ as defining a “quadratic form”:

Def: Let $A$ be a symmetric $n \times n$ matrix. The quadratic form associated to $A$ is the function $Q_A: \mathbb{R}^n \to \mathbb{R}$ given by:

$$Q_A(x) = x \cdot Ax \quad (\cdot \text{ is the dot product})$$

$$= x^T A x = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Notice that quadratic forms are not linear transformations!
Orthonormal Bases

Def: A basis \( \{w_1, \ldots, w_k\} \) for a subspace \( V \) is an **orthonormal basis** if:

1. The basis vectors are mutually orthogonal: \( w_i \cdot w_j = 0 \) (for \( i \neq j \));
2. The basis vectors are unit vectors: \( w_i \cdot w_i = 1 \). (i.e.: \( \|w_i\| = 1 \))

Orthonormal bases are nice for (at least) two reasons:

(a) It is much easier to find the \( B \)-coordinates \( [v]_B \) of a vector when the basis \( B \) is orthonormal;
(b) It is much easier to find the **projection matrix** onto a subspace \( V \) when we have an orthonormal basis for \( V \).

**Prop:** Let \( \{w_1, \ldots, w_k\} \) be an orthonormal basis for a subspace \( V \subset \mathbb{R}^n \).

(a) Every vector \( v \in V \) can be written
\[
y = (v \cdot w_1)w_1 + \cdots + (v \cdot w_k)w_k.
\]
(b) For all \( x \in \mathbb{R}^n \):
\[
\text{Proj}_V(x) = (x \cdot w_1)w_1 + \cdots + (x \cdot w_k)w_k.
\]
(c) Let \( A \) be the matrix with columns \( \{w_1, \ldots, w_k\} \). Then \( A^T A = I_k \), so:
\[
\text{Proj}_V = A(A^T A)^{-1}A^T = AA^T.
\]

Orthogonal Matrices

Def: An **orthogonal matrix** is an invertible matrix \( C \) such that
\[
C^{-1} = C^T.
\]

**Example:** Let \( \{v_1, \ldots, v_n\} \) be an orthonormal basis for \( \mathbb{R}^n \). Then the matrix
\[
C = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}
\]
is an orthogonal matrix.

In fact, **every** orthogonal matrix \( C \) looks like this: the columns of any orthogonal matrix form an orthonormal basis of \( \mathbb{R}^n \).

Where theory is concerned, the key property of orthogonal matrices is:

**Prop 22.4:** Let \( C \) be an orthogonal matrix. Then for \( v, w \in \mathbb{R}^n \):
\[
Cv \cdot Cw = v \cdot w.
\]
Gram-Schmidt Process

Since orthonormal bases have so many nice properties, it would be great if we had a way of actually manufacturing orthonormal bases. That is:

**Goal:** We are given a basis \(\{v_1, \ldots, v_k\}\) for a subspace \(V \subset \mathbb{R}^n\). We would like an *orthonormal* basis \(\{w_1, \ldots, w_k\}\) for our subspace \(V\).

**Notation:** We will let

\[
V_1 = \text{span}(v_1) \\
V_2 = \text{span}(v_1, v_2) \\
\vdots \\
V_k = \text{span}(v_1, \ldots, v_k) = V.
\]

**Idea:** Build an orthonormal basis for \(V_1\), then for \(V_2, \ldots\), up to \(V_k = V\).

**Gram-Schmidt Algorithm:** Let \(\{v_1, \ldots, v_k\}\) be a basis for \(V \subset \mathbb{R}^n\).

1. Define \(w_1 = \frac{v_1}{\|v_1\|}\).
2. Having defined \(\{w_1, \ldots, w_j\}\), let

\[
y_{j+1} = v_{j+1} - \text{Proj}_{V_j}(v_{j+1}) \\
= v_{j+1} - (v_{j+1} \cdot w_1)w_1 - (v_{j+1} \cdot w_2)w_2 - \cdots - (v_{j+1} \cdot w_j)w_j,
\]

and define \(w_{j+1} = \frac{y_{j+1}}{\|y_{j+1}\|}\).

Then \(\{w_1, \ldots, w_k\}\) is an orthonormal basis for \(V\).
Definiteness

Def: Let \( Q : \mathbb{R}^n \to \mathbb{R} \) be a quadratic form.

We say \( Q \) is **positive definite** if \( Q(x) > 0 \) for all \( x \neq 0 \).

We say \( Q \) is **negative definite** if \( Q(x) < 0 \) for all \( x \neq 0 \).

We say \( Q \) is **indefinite** if there are vectors \( x \) for which \( Q(x) > 0 \), and also vectors \( x \) for which \( Q(x) < 0 \).

Def: Let \( A \) be a symmetric matrix.

We say \( A \) is **positive definite** if \( Q_A(x) = x^T A x > 0 \) for all \( x \neq 0 \).

We say \( A \) is **negative definite** if \( Q_A(x) = x^T A x < 0 \) for all \( x \neq 0 \).

We say \( A \) is **indefinite** if there are vectors \( x \) for which \( x^T A x > 0 \), and also vectors \( x \) for which \( x^T A x < 0 \).

In other words:

\( \circ \) \( A \) is positive definite \( \iff \) \( Q_A \) is positive definite.

\( \circ \) \( A \) is negative definite \( \iff \) \( Q_A \) is negative definite.

\( \circ \) \( A \) is indefinite \( \iff \) \( Q_A \) is indefinite.

The Hessian

Def: Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a function. Its **Hessian** at \( a \in \mathbb{R}^n \) is the symmetric matrix of second partials:

\[
H_f(a) = \begin{bmatrix}
  f_{x_1 x_1}(a) & \cdots & f_{x_1 x_n}(a) \\
  \vdots & \ddots & \vdots \\
  f_{x_n x_1}(a) & \cdots & f_{x_n x_n}(a)
\end{bmatrix}.
\]

Note that the Hessian is a symmetric matrix. Therefore, we can also regard \( H_f(a) \) as a quadratic form:

\[
Q_{H_f(a)}(x) = x^T H_f(a) x = [x_1 \cdots x_n] \begin{bmatrix}
  f_{x_1 x_1}(a) & \cdots & f_{x_1 x_n}(a) \\
  \vdots & \ddots & \vdots \\
  f_{x_n x_1}(a) & \cdots & f_{x_n x_n}(a)
\end{bmatrix} [x_1 \cdots x_n].
\]

In particular, it makes sense to ask whether the Hessian is positive definite, negative definite, or indefinite.
Single-Variable Calculus Review

Recall: In calculus, you learned that for a function \( f : \mathbb{R} \to \mathbb{R} \), a critical point is a point \( a \in \mathbb{R} \) where \( f'(a) = 0 \) or \( f'(a) \) does not exist.

You learned that if \( f(x) \) has a local min/max at \( x = a \), then \( x = a \) is a critical point. Of course, the converse is false: critical points don’t have to be local minima or local maxima (e.g., they could be inflection points.)

You also learned the “second derivative test.” If \( x = a \) is a critical point for \( f(x) \), then \( f''(a) > 0 \) tells us that \( x = a \) is a local min, whereas \( f''(a) < 0 \) tells us that \( x = a \) is a local max.

It would be nice to have similar statements in higher dimensions:

Critical Points & Second Derivative Test

Def: A critical point of \( f : \mathbb{R}^n \to \mathbb{R} \) is a point \( a \in \mathbb{R}^n \) at which \( Df(a) = 0^T \) or \( Df(a) \) is undefined.

In other words, each partial derivative \( \frac{\partial f}{\partial x_i}(a) \) is zero or undefined.

Theorem: If \( f : \mathbb{R}^n \to \mathbb{R} \) has a local max / local min at \( a \in \mathbb{R}^n \), then \( a \) is a critical point of \( f \).

N.B.: The converse of this theorem is false! Critical points do not have to be a local max or local min – e.g., they could be saddle points.

Def: A saddle point of \( f : \mathbb{R}^n \to \mathbb{R} \) is a critical point of \( f \) that is not a local max or local min.

Second Derivative Test: Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a function, and \( a \in \mathbb{R}^n \) be a critical point of \( f \).

(a) If \( Hf(a) \) is positive definite, then \( a \) is a local min of \( f \).
(b) If \( Hf(a) \) is positive semi-definite, then \( a \) is local min or saddle point.
(c) If \( Hf(a) \) is negative definite, then \( a \) is a local max of \( f \).
(d) If \( Hf(a) \) is negative semi-definite, then \( a \) is local max or saddle point.
(e) If \( Hf(a) \) is indefinite, then \( a \) is a saddle point of \( f \).
Local Extrema vs Global Extrema

Finding Local Extrema: We want to find the local extrema of a function $f: \mathbb{R}^n \to \mathbb{R}$.

(i) Find the critical points of $f$.
(ii) Use the Second Derivative Test to decide if the critical points are local maxima / minima / saddle points.

**Theorem:** Let $f: \mathbb{R}^n \to \mathbb{R}$ be a function. If $R \subset \mathbb{R}^n$ is a closed and bounded region, then $f$ has a global max and a global min on $R$.

Finding Global Extrema: We want to find the global extrema of a function $f: \mathbb{R}^n \to \mathbb{R}$ on a region $R \subset \mathbb{R}^n$.

(1) Find the critical points of $f$ on the interior of $R$.
(2) Find the extreme values of $f$ on the boundary of $R$. (Lagrange mult.)

Then:
- The largest value from Steps (1)-(2) is a global max value.
- The smallest value from Steps (1)-(2) is a global min value.

Lagrange Multipliers (Constrained Optimization)

**Notation:** Let $f: \mathbb{R}^n \to \mathbb{R}^m$ be a function, and $S \subset \mathbb{R}^n$ be a subset.

The *restricted function* $f|_S: S \to \mathbb{R}^m$ is the same exact function as $f$, but where the domain is restricted to $S$.

**Theorem:** Suppose we want to optimize a function $f(x_1, \ldots, x_n)$ constrained to a level set $S = \{g(x_1, \ldots, x_n) = c\}$.

If $a$ is an extreme value of $f|_S$ on the level set $S = \{g(x_1, \ldots, x_n) = c\}$, and if $\nabla g(a) \neq 0$, then

$$\nabla f(a) = \lambda \nabla g(a)$$

for some constant $\lambda$.

**Reason:** If $a$ is an extreme value of $f|_S$ on the level set $S$, then $D_v f(a) = 0$ for all vectors $v$ that are tangent to the level set $S$. Therefore, $\nabla f(a) \cdot v = 0$ for all vectors $v$ that are tangent to $S$.

This means that $\nabla f(a)$ is orthogonal to the level set $S$, so $\nabla f(a)$ must be a scalar multiple of the normal vector $\nabla g(a)$. That is, $\nabla f(a) = \lambda \nabla g(a)$. □
Motivation for Eigenvalues & Eigenvectors

We want to understand a quadratic form $Q_A(x)$, which might be ugly and complicated.

Idea: Maybe there’s an orthonormal basis $B = \{w_1, \ldots, w_n\}$ of $\mathbb{R}^n$ that is somehow “best suited to $A$” – so that with respect to the basis $B$, the quadratic form $Q_A$ looks simple.

What do we mean by “basis suited to $A$”? And does such a basis always exist? Well:

**Spectral Theorem:** Let $A$ be a symmetric $n \times n$ matrix. Then there exists an orthonormal basis $B = \{w_1, \ldots, w_n\}$ of $\mathbb{R}^n$ such that each $w_1, \ldots, w_n$ is an eigenvector of $A$.

i.e.: There is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of $A$.

Why is this good? Well, since $B$ is a basis, every $w \in \mathbb{R}^n$ can be written $w = u_1w_1 + \cdots + u_nw_n$. (That is, the $B$-coordinates of $w$ are $(u_1, \ldots, u_n)$.)

It then turns out that:

$$Q_A(w) = Q_A(u_1w_1 + \cdots + u_nw_n)$$
$$= (u_1w_1 + \cdots + u_nw_n) \cdot A(u_1w_1 + \cdots + u_nw_n)$$
$$= \lambda_1(u_1)^2 + \lambda_2(u_2)^2 + \cdots + \lambda_n(u_n)^2.$$ (yay!)

In other words: the quadratic form $Q_A$ is in diagonal form with respect to the basis $B$. We have made $Q_A$ look as simple as possible!

Also: the coefficients $\lambda_1, \ldots, \lambda_n$ are exactly the eigenvalues of $A$.

**Corollary:** Let $A$ be a symmetric $n \times n$ matrix, with eigenvalues $\lambda_1, \ldots, \lambda_n$.

(a) $A$ is positive-definite $\iff$ all of $\lambda_1, \ldots, \lambda_n$ are positive.

(b) $A$ is negative-definite $\iff$ all of $\lambda_1, \ldots, \lambda_n$ are negative.

(c) $A$ is indefinite $\iff$ there is a positive eigenvalue $\lambda_i > 0$ and a negative eigenvalue $\lambda_j < 0$.

**Useful Fact:** Let $A$ be any $n \times n$ matrix, with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n.$$

**Cor:** If any one of the eigenvalues $\lambda_j = 0$ is zero, then $\det(A) = 0$. 
What is a (Unit) Sphere?

- The **1-sphere** (the “unit circle”) is \( S^1 = \{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \} \subset \mathbb{R}^2 \).
- The **2-sphere** (the “sphere”) is \( S^2 = \{ (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1 \} \subset \mathbb{R}^3 \).
- The **3-sphere** is \( S^3 = \{ (x, y, z, w) \in \mathbb{R}^4 \mid x^2 + y^2 + z^2 + w^2 = 1 \} \subset \mathbb{R}^4 \).

Note that the 3-sphere is *not* the same as the unit ball \( \{ x^2 + y^2 + z^2 \leq 1 \} \).

- The \((n-1)\)-sphere is the set
  \[
  S^{n-1} = \{ (x_1, \ldots, x_n) \in \mathbb{R}^n \mid (x_1)^2 + \cdots + (x_n)^2 = 1 \}
  = \{ x \in \mathbb{R}^n \mid \| x \|^2 = 1 \} \subset \mathbb{R}^n.
  \]
In other words, \( S^{n-1} \) consists of the unit vectors in \( \mathbb{R}^n \).

Optimizing Quadratic Forms on Spheres

**Problem:** Optimize a quadratic form \( Q_A : \mathbb{R}^n \to \mathbb{R} \) on the sphere \( S^{n-1} \subset \mathbb{R}^n \).
That is, what are the maxima and minima of \( Q_A(w) \) subject to the constraint that \( \| w \| = 1 \)?

**Solution:** Let \( \lambda_{\text{max}} \) and \( \lambda_{\text{min}} \) be the largest and smallest eigenvalues of \( A \).

- The maximum value of \( Q_A \) for unit vectors is \( \lambda_{\text{max}} \). Any unit vector \( w_{\text{max}} \) which attains this maximum is an eigenvector of \( A \) with eigenvalue \( \lambda_{\text{max}} \).
- The minimum value of \( Q_A \) for unit vectors is \( \lambda_{\text{min}} \). Any unit vector \( w_{\text{min}} \) which attains this minimum is an eigenvector of \( A \) with eigenvalue \( \lambda_{\text{min}} \).

**Corollary:** Let \( A \) be a symmetric \( n \times n \) matrix.

- \( A \) is positive-definite \( \iff \) the minimum value of \( Q_A \) restricted to unit vector inputs is positive (i.e., iff \( \lambda_{\text{min}} > 0 \)).
- \( A \) is negative-definite \( \iff \) the maximum value of \( Q_A \) restricted to unit vector inputs is negative (i.e., iff \( \lambda_{\text{max}} < 0 \)).
- \( A \) is indefinite \( \iff \lambda_{\text{max}} > 0 \) and \( \lambda_{\text{min}} < 0 \).
**Directional First & Second Derivatives**

**Def:** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a function, \( a \in \mathbb{R}^n \) be a point.

The **directional derivative** of \( f \) at \( a \) in the direction \( v \) is:

\[
D_v f(a) = \nabla f(a) \cdot v.
\]

The “**directional second derivative**” of \( f \) at \( a \) in the direction \( v \) is:

\[
Q_{Hf(a)}(v) = v^T Hf(a)v.
\]

That is: the quadratic form whose associated matrix is the Hessian \( Hf(a) \).

**Q:** What direction \( v \) increases the directional derivative the most? What direction \( v \) decreases the directional derivative the most?

**A:** We’ve learned this: the gradient \( \nabla f(a) \) is the direction of greatest increase, whereas \( -\nabla f(a) \) is the direction of greatest decrease.

**New Questions:**
- What direction \( v \) increases the directional **second** derivative the most?
- What direction \( v \) decreases the directional **second** derivative the most?

**Answer:** The (unit) directions of minimum and maximum second derivative are (unitized) eigenvectors of \( Hf(a) \), and so they are *mutually orthogonal*.

The max/min values of the directional second derivative are the max/min eigenvalues of \( Hf(a) \).