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 5×3 matrix, so $A\colon \mathbb{R}^3 \to \mathbb{R}^5$.

- $N(A)$ is a subspace of ______
- $C(A)$ is a subspace of ______

The transpose A^T is a ______ matrix, so $A^T\colon \quad \to$

- $C(A^T)$ is a subspace of ______
- $N(A^T)$ is a subspace of ______

Observation: Both $C(A^T)$ and $N(A)$ are subspaces of ______. Might there be a geometric relationship between the two? (No, they're not equal.) Hm...

Also: Both $N(A^T)$ and $C(A)$ 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

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

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 $V = \{\mathbf{0}\}$ 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 $V^\perp = \{\mathbf{0}\}$

Examples in $\mathbb{R}^4$:

- The orthogonal complement of $V = \{\mathbf{0}\}$ 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 $V^\perp = \{\mathbf{0}\}$

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) $V \cap V^\perp = \{\mathbf{0}\}$
(d) $V + V^\perp = \mathbb{R}^n$.

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

Also, it turns out that the expression $\mathbf{x} = \mathbf{v} + \mathbf{w}$ is unique: that is, there is only one way to write $\mathbf{x}$ 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:

$$\begin{aligned} C(A^T) &= \text{column space of } A^T \\ &= \text{span of columns of } A^T \\ &= \text{span of rows of } A. \end{aligned}$$

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

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

$$\begin{aligned} \mathbf{x} \in N(A^T) &\iff A^T\mathbf{x} = \mathbf{0} \\ &\iff (A^T\mathbf{x})^T = \mathbf{0}^T \\ &\iff \mathbf{x}^T A = \mathbf{0}^T. \end{aligned}$$

So, for an $m \times n$ matrix A, we see that: $N(A^T) = \{\mathbf{x} \in \mathbb{R}^m \mid \mathbf{x}^T A = \mathbf{0}^T\}$.
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 $\mathbf{b} \in C(A)$. So, the system of equations $A\mathbf{x} = \mathbf{b}$ does have solutions, possibly infinitely many.

Q: What is the solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ with $\|\mathbf{x}\|$ the smallest?

Problem 2: Let $\mathbf{b} \notin C(A)$. So, the system of equations $A\mathbf{x} = \mathbf{b}$ does <u>not</u> have any solutions. In other words, $A\mathbf{x} - \mathbf{b} \neq \mathbf{0}$.

Q: What is the vector $\mathbf{x}$ that minimizes the error $\|A\mathbf{x} - \mathbf{b}\|$? That is, what is the vector $\mathbf{x}$ that comes closest to being a solution to $A\mathbf{x} = \mathbf{b}$?

Orthogonal Projection

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

$$\mathbf{x} = \mathbf{v} + \mathbf{w}, \quad \text{where } \mathbf{v} \in V \text{ and } \mathbf{w} \in V^{\perp}.$$

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

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

$$\mathrm{Proj}_V + \mathrm{Proj}_{V^\perp} = I_n.$$

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

Formula: Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of $V \subset \mathbb{R}^n$. Let A be the $n \times k$ matrix

$$A = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_k \\ | & & | \end{bmatrix}.$$

Then:

$$\mathrm{Proj}_V = A(A^TA)^{-1}A^T. \tag{$*$}$$

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

(1) The distance from $\mathbf{x}$ to V is: $\|\mathrm{Proj}_{V^\perp}(\mathbf{x})\| = \|\mathbf{x} - \mathrm{Proj}_V(\mathbf{x})\|$.

(2) The vector in V that is closest to $\mathbf{x}$ is: $\mathrm{Proj}_V(\mathbf{x})$.

Derivation of $(*)$*:* Notice $\mathrm{Proj}_V(\mathbf{x})$ is a vector in $V = \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = C(A) = \mathrm{Range}(A)$, and therefore $\mathrm{Proj}_V(\mathbf{x}) = A\mathbf{y}$ for some vector $\mathbf{y} \in \mathbb{R}^k$.

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

Now, it turns out that our matrix A^TA is invertible (proof in L20), so we get $\mathbf{y} = (A^TA)^{-1}A^T\mathbf{x}$. Thus, $\mathrm{Proj}_V(\mathbf{x}) = A\mathbf{y} = A(A^TA)^{-1}A^T\mathbf{x}$. $\Diamond$

Minimum Magnitude Solution

Prop 19.6: Let $\mathbf{b} \in C(A)$ (so $A\mathbf{x} = \mathbf{b}$ has solutions). Then there exists exactly one vector $\mathbf{x}_0 \in C(A^T)$ with $A\mathbf{x}_0 = \mathbf{b}$.

And: Among all solutions of $A\mathbf{x} = \mathbf{b}$, the vector $\mathbf{x}_0$ has the smallest length.

In other words: There is exactly one vector $\mathbf{x}_0$ in the row space of A which solves $A\mathbf{x} = \mathbf{b}$ – and this vector is the solution of smallest length.

To Find $\mathbf{x}_0$: Start with any solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$. Then

$$\boxed{\mathbf{x}_0 = \mathrm{Proj}_{C(A^T)}(\mathbf{x}).}$$

Least Squares Approximation

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

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

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

$$\boxed{A^T A\mathbf{x}^* = A^T\mathbf{b}.}$$

Quadratic Forms (Intro)

Given an $m \times n$ matrix A, we can regard it as a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$. 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 $Q_A\colon \mathbb{R}^n \to \mathbb{R}$ given by:

$$\begin{aligned} Q_A(\mathbf{x}) &= \mathbf{x} \cdot A\mathbf{x} && (\cdot \text{ is the dot product}) \\ &= \mathbf{x}^T A\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

Notice that quadratic forms are <u>not</u> linear transformations!

Orthonormal Bases

Def: A basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ for a subspace V is an **orthonormal basis** if:
(1) The basis vectors are mutually orthogonal: $\mathbf{w}_i \cdot \mathbf{w}_j = 0$ (for $i \neq j$);
(2) The basis vectors are unit vectors: $\mathbf{w}_i \cdot \mathbf{w}_i = 1$. (i.e.: $\|\mathbf{w}_i\| = 1$)

Orthonormal bases are nice for (at least) two reasons:
(a) It is much easier to find the **$\mathcal{B}$-coordinates** $[\mathbf{v}]_{\mathcal{B}}$ of a vector when the basis $\mathcal{B}$ 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 $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ be an orthonormal basis for a subspace $V \subset \mathbb{R}^n$.
(a) Every vector $\mathbf{v} \in V$ can be written
$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{w}_1)\mathbf{w}_1 + \cdots + (\mathbf{v} \cdot \mathbf{w}_k)\mathbf{w}_k.$$
(b) For all $\mathbf{x} \in \mathbb{R}^n$:
$$\text{Proj}_V(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{w}_1)\mathbf{w}_1 + \cdots + (\mathbf{x} \cdot \mathbf{w}_k)\mathbf{w}_k.$$
(c) Let A be the matrix with columns $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$. 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 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$. Then the matrix
$$C = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix}$$
is an orthogonal matrix.

In fact, <u>every</u> 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 $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$:
$$C\mathbf{v} \cdot C\mathbf{w} = \mathbf{v} \cdot \mathbf{w}.$$

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 $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ for a subspace $V \subset \mathbb{R}^n$. We would like an *orthonormal* basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ for our subspace V.

Notation: We will let

$$\begin{aligned} V_1 &= \operatorname{span}(\mathbf{v}_1) \\ V_2 &= \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2) \\ &\vdots \\ V_k &= \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = V. \end{aligned}$$

Idea: Build an orthonormal basis for V_1, then for V_2, ..., up to $V_k = V$.

Gram-Schmidt Algorithm: Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for $V \subset \mathbb{R}^n$.

(1) Define $\mathbf{w}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$.

(2) Having defined $\{\mathbf{w}_1, \ldots, \mathbf{w}_j\}$, let

$$\begin{aligned} \mathbf{y}_{j+1} &= \mathbf{v}_{j+1} - \operatorname{Proj}_{V_j}(\mathbf{v}_{j+1}) \\ &= \mathbf{v}_{j+1} - (\mathbf{v}_{j+1} \cdot \mathbf{w}_1)\mathbf{w}_1 - (\mathbf{v}_{j+1} \cdot \mathbf{w}_2)\mathbf{w}_2 - \cdots - (\mathbf{v}_{j+1} \cdot \mathbf{w}_j)\mathbf{w}_j, \end{aligned}$$

and define $\mathbf{w}_{j+1} = \frac{\mathbf{y}_{j+1}}{\|\mathbf{y}_{j+1}\|}$.

Then $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ is an orthonormal basis for V.

Definiteness

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

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

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

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

Def: Let A be a symmetric matrix.

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

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

We say A is **indefinite** if there are vectors $\mathbf{x}$ for which $\mathbf{x}^T A\mathbf{x} > 0$, and also vectors $\mathbf{x}$ for which $\mathbf{x}^T A\mathbf{x} < 0$.

In other words:
- ∘ A is positive definite $\iff$ Q_A is positive definite.
- ∘ A is negative definite $\iff$ Q_A is negative definite.
- ∘ A is indefinite $\iff$ Q_A is indefinite.

The Hessian

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

$$Hf(\mathbf{a}) = \begin{bmatrix} f_{x_1x_1}(\mathbf{a}) & \cdots & f_{x_1x_n}(\mathbf{a}) \\ \cdots & \ddots & \cdots \\ f_{x_nx_1}(\mathbf{a}) & \cdots & f_{x_nx_n}(\mathbf{a}) \end{bmatrix}.$$

Note that the Hessian is a symmetric matrix. Therefore, we can also regard $Hf(\mathbf{a})$ as a quadratic form:

$$Q_{Hf(\mathbf{a})}(\mathbf{x}) = \mathbf{x}^T Hf(\mathbf{a})\,\mathbf{x} = \begin{bmatrix} x_1 \cdots x_n \end{bmatrix} \begin{bmatrix} f_{x_1x_1}(\mathbf{a}) & \cdots & f_{x_1x_n}(\mathbf{a}) \\ \cdots & \ddots & \cdots \\ f_{x_nx_1}(\mathbf{a}) & \cdots & f_{x_nx_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

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 $f\colon \mathbb{R} \to \mathbb{R}$, 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 $f\colon \mathbb{R}^n \to \mathbb{R}$ is a point $\mathbf{a} \in \mathbb{R}^n$ at which $Df(\mathbf{a}) = \mathbf{0}^T$ or $Df(\mathbf{a})$ is undefined.

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

Theorem: If $f\colon \mathbb{R}^n \to \mathbb{R}$ has a local max / local min at $\mathbf{a} \in \mathbb{R}^n$, then $\mathbf{a}$ 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 $f\colon \mathbb{R}^n \to \mathbb{R}$ is a critical point of f that is not a local max or local min.

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

(a) If $Hf(\mathbf{a})$ is positive definite, then $\mathbf{a}$ is a local min of f.
(b) If $Hf(\mathbf{a})$ is positive semi-definite, then $\mathbf{a}$ is local min or saddle point.
(c) If $Hf(\mathbf{a})$ is negative definite, then $\mathbf{a}$ is a local max of f.
(d) If $Hf(\mathbf{a})$ is negative semi-definite, then $\mathbf{a}$ is local max or saddle point.
(e) If $Hf(\mathbf{a})$ is indefinite, then $\mathbf{a}$ is a saddle point of f.

Local Extrema vs Global Extrema

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

(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 $f\colon \mathbb{R}^n \to \mathbb{R}$ 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 $f\colon \mathbb{R}^n \to \mathbb{R}$ on a region $R \subset \mathbb{R}^n$.

(1) Find the critical points of f on the <u>interior</u> of R.

(2) Find the extreme values of f on the <u>boundary</u> 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 $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function, and $S \subset \mathbb{R}^n$ be a subset.

The *restricted function* $f|_S\colon S \to \mathbb{R}^m$ 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 $\mathbf{a}$ is an extreme value of $f|_S$ on the level set $S = \{g(x_1, \ldots, x_n) = c\}$, and if $\nabla g(\mathbf{a}) \neq \mathbf{0}$, then

$$\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$$

for some constant λ.

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

This means that $\nabla f(\mathbf{a})$ is orthogonal to the level set S, so $\nabla f(\mathbf{a})$ must be a scalar multiple of the normal vector $\nabla g(\mathbf{a})$. That is, $\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$. $\square$

Motivation for Eigenvalues & Eigenvectors

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

Idea: Maybe there's an orthonormal basis $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ of $\mathbb{R}^n$ that is somehow "best suited to A" – so that with respect to the basis $\mathcal{B}$, 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 <u>orthonormal basis</u> $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ of $\mathbb{R}^n$ such that each $\mathbf{w}_1, \ldots, \mathbf{w}_n$ is an <u>eigenvector</u> of A.

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

Why is this good? Well, since $\mathcal{B}$ is a basis, every $\mathbf{w} \in \mathbb{R}^n$ can be written $\mathbf{w} = u_1\mathbf{w}_1 + \cdots + u_n\mathbf{w}_n$. (That is, the $\mathcal{B}$-coordinates of $\mathbf{w}$ are $(u_1, \ldots, u_n)$.) It then turns out that:

$$\begin{aligned} Q_A(\mathbf{w}) &= Q_A(u_1\mathbf{w}_1 + \cdots + u_n\mathbf{w}_n) \\ &= (u_1\mathbf{w}_1 + \cdots + u_n\mathbf{w}_n) \cdot A(u_1\mathbf{w}_1 + \cdots + u_n\mathbf{w}_n) \\ &= \boxed{\lambda_1(u_1)^2 + \lambda_2(u_2)^2 + \cdots + \lambda_n(u_n)^2.} \qquad \text{(yay!)} \end{aligned}$$

In other words: the quadratic form Q_A is in diagonal form with respect to the basis $\mathcal{B}$. 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 $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\} \subset \mathbb{R}^2$.
∘ The **2-sphere** (the "sphere") is $\mathbb{S}^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\} \subset \mathbb{R}^3$.
∘ The **3-sphere** is $\mathbb{S}^3 = \{(x, y, z, w) \in \mathbb{R}^4 \mid x^2 + y^2 + z^2 + w^2 = 1\} \subset \mathbb{R}^4$.
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

$$\begin{aligned}\mathbb{S}^{n-1} &= \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid (x_1)^2 + \cdots + (x_n)^2 = 1\} \\ &= \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|^2 = 1\} \subset \mathbb{R}^n.\end{aligned}$$

In other words, $\mathbb{S}^{n-1}$ consists of the **unit vectors** in $\mathbb{R}^n$.

Optimizing Quadratic Forms on Spheres

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

Solution: Let λ_{max} and λ_{min} be the largest and smallest eigenvalues of A.
∘ The maximum value of Q_A for unit vectors is λ_{max}. Any unit vector $\mathbf{w}_{\text{max}}$ which attains this maximum is an eigenvector of A with eigenvalue λ_{max}.
∘ The minimum value of Q_A for unit vectors is λ_{min}. Any unit vector $\mathbf{w}_{\text{min}}$ which attains this minimum is an eigenvector of A with eigenvalue λ_{min}.

Corollary: Let A be a symmetric $n \times n$ matrix.
(a) A is positive-definite $\iff$ the minimum value of Q_A restricted to unit vector inputs is positive (i.e., iff $\lambda_{\text{min}} > 0$).
(b) A is negative-definite $\iff$ the maximum value of Q_A restricted to unit vector inputs is negative (i.e., iff $\lambda_{\text{max}} < 0$).
(c) A is indefinite $\iff$ $\lambda_{\text{max}} > 0$ and $\lambda_{\text{min}} < 0$.

Directional First & Second Derivatives

Def: Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a function, $\mathbf{a} \in \mathbb{R}^n$ be a point.

The **directional derivative** of f at $\mathbf{a}$ in the direction $\mathbf{v}$ is:

$$D_{\mathbf{v}}f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{v}.$$

The **"directional second derivative"** of f at $\mathbf{a}$ in the direction $\mathbf{v}$ is:

$$Q_{Hf(\mathbf{a})}(\mathbf{v}) = \mathbf{v}^T Hf(\mathbf{a})\mathbf{v}.$$

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

Q: What direction $\mathbf{v}$ increases the directional derivative the most? What direction $\mathbf{v}$ decreases the directional derivative the most?

A: We've learned this: the gradient $\nabla f(\mathbf{a})$ is the direction of greatest increase, whereas $-\nabla f(\mathbf{a})$ is the direction of greatest decrease.

New Questions:

- ◦ What direction $\mathbf{v}$ increases the directional **second** derivative the most?
- ◦ What direction $\mathbf{v}$ decreases the directional **second** derivative the most?

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

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