BIOMATH 208 Week 2

Aidan Jan

January 16, 2025

Multilinear Algebra

Motivation

two very different types of vectors arise in imaging:

- Tangents to curves/surfaces: Important for describing geometry and anatomy

- Image gradients: Important for optimization and feature identification.

If we chose a new coordinate system to represent our data, these vectors transform differently.

Next, we will apply a linear transform, a rotation matrix A, to some imaging data and see how these objects change.
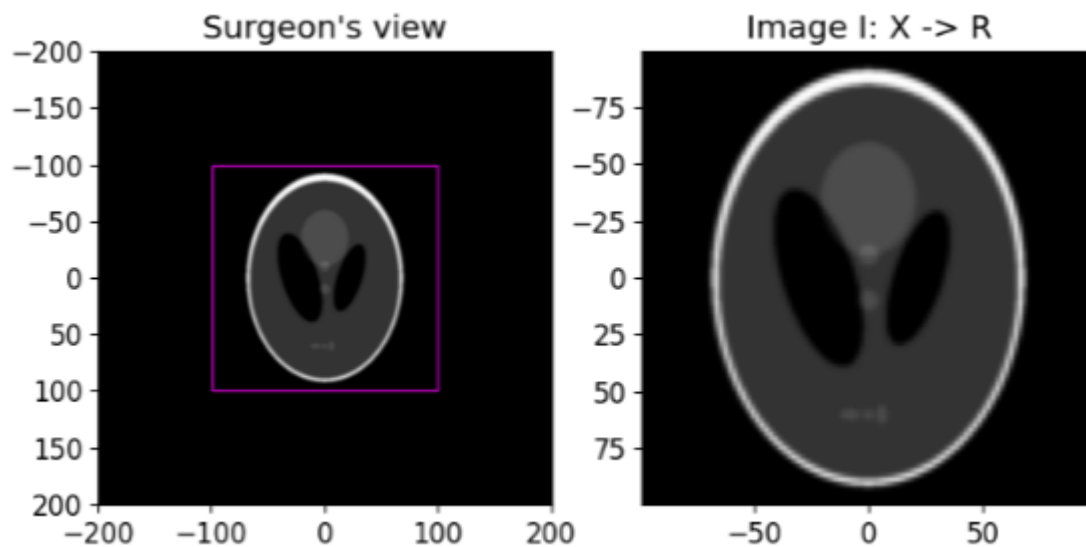
Then we'll apply a different linear transform, a rotation and scale B, to some imaging data, and see how the objects change.

Example: Brain Image

Consider this brain image represented by ellipses:
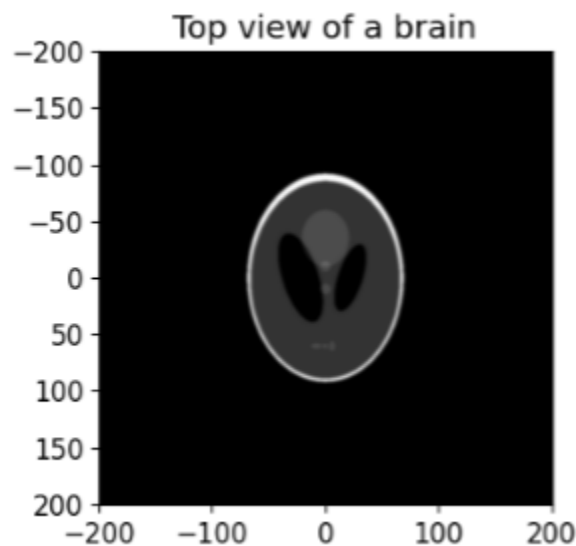
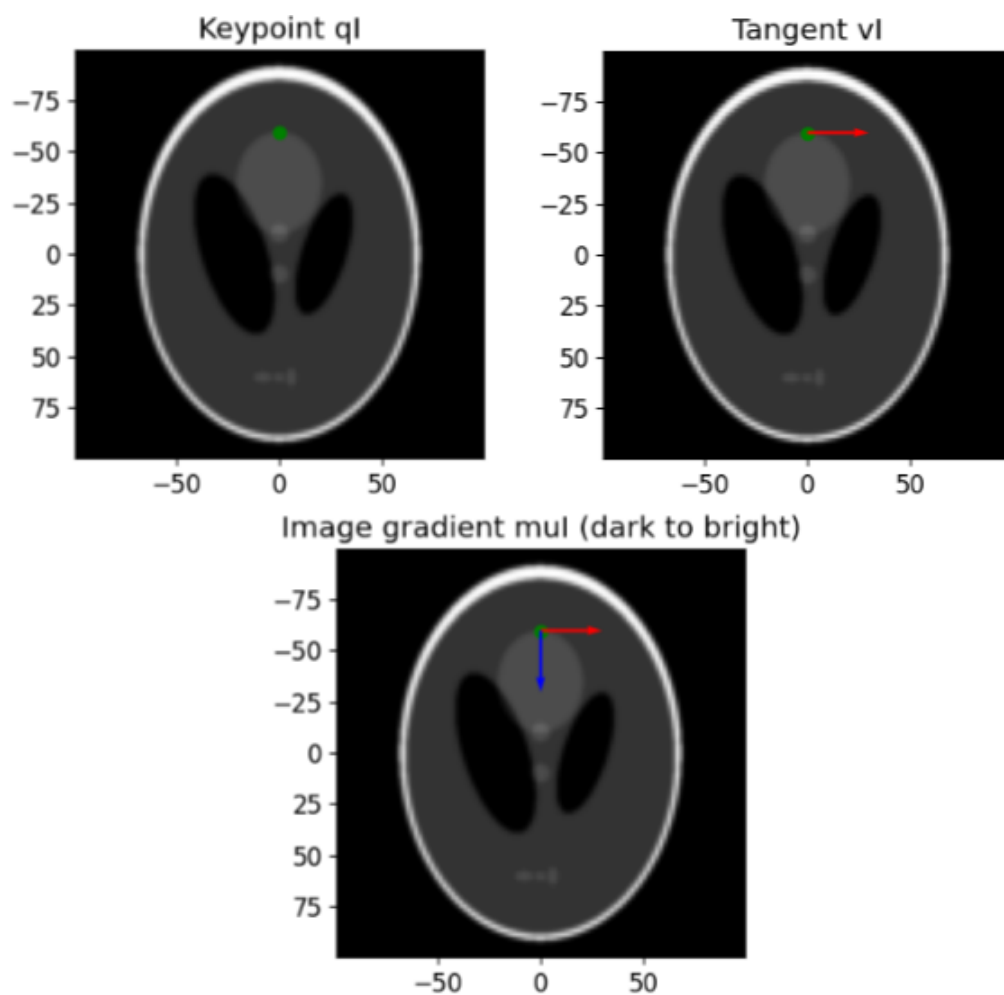Now, assume this is an image of a patient's brain, and a surgeon is standing directly over it.
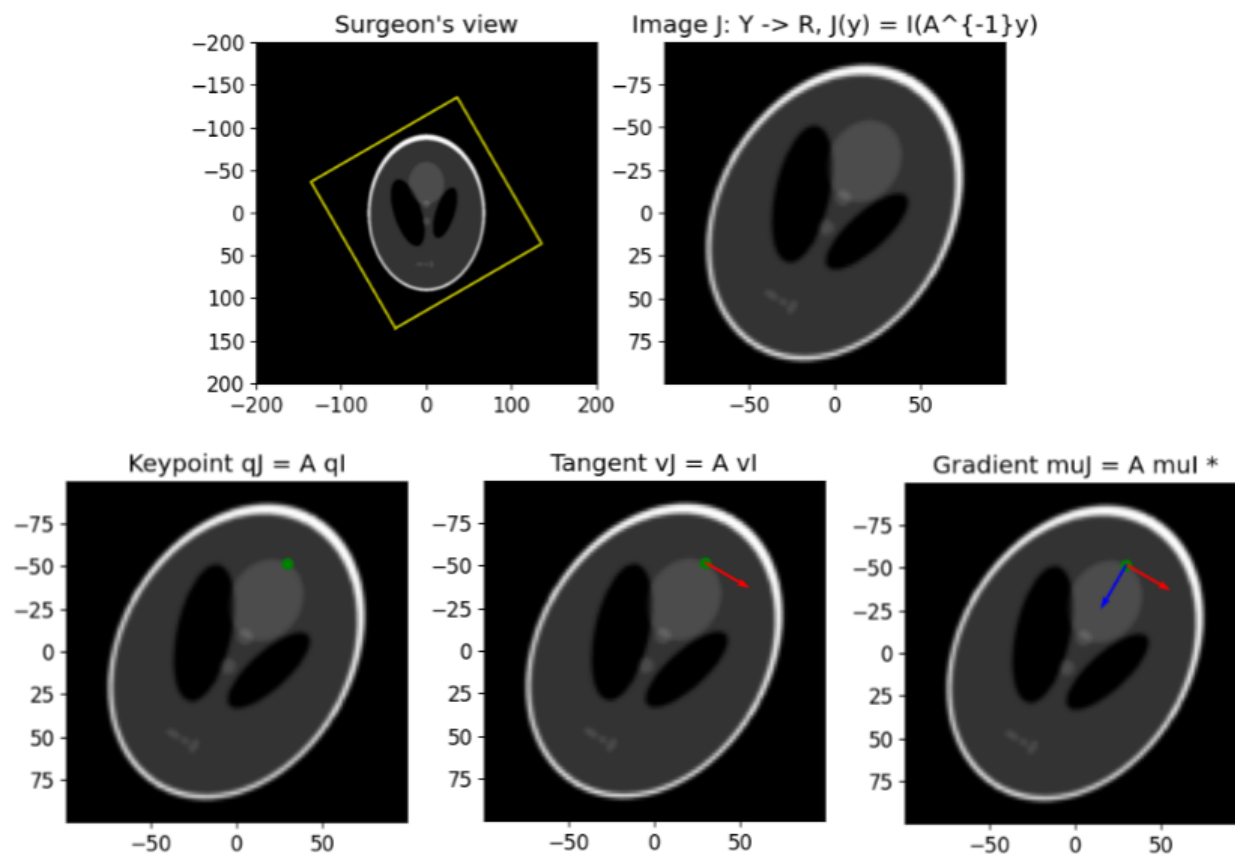
- This is like zooming into the image, thus, Image $I\colon X \to R$

With this transformation, image gradients, tangents, and points are all preserved.

Now, suppose the surgeon turns his head, so he is viewing the head at a different angle:

- A point in the domain of I is rotated to a point in the domain of J with the linear map $A : X \to Y$.

In this case, points and vectors are moved based on the rotation matrix multiplied. However, this can be expressed in linear algebra as a change of basis, and thus, points and vectors are preseved relative to the new coordinate system.

Finally, consider if the surgeon moved his head lower, so now he is not looking at the patient's brain from straight top-down, but rather at an angle.

Surgeon's view

Image K: Z -> R, K(z) = I(B^{-1}z)

Keypoint qK = B ql

Tangent vK = B mul

Gradient muK = B^{-T} mul
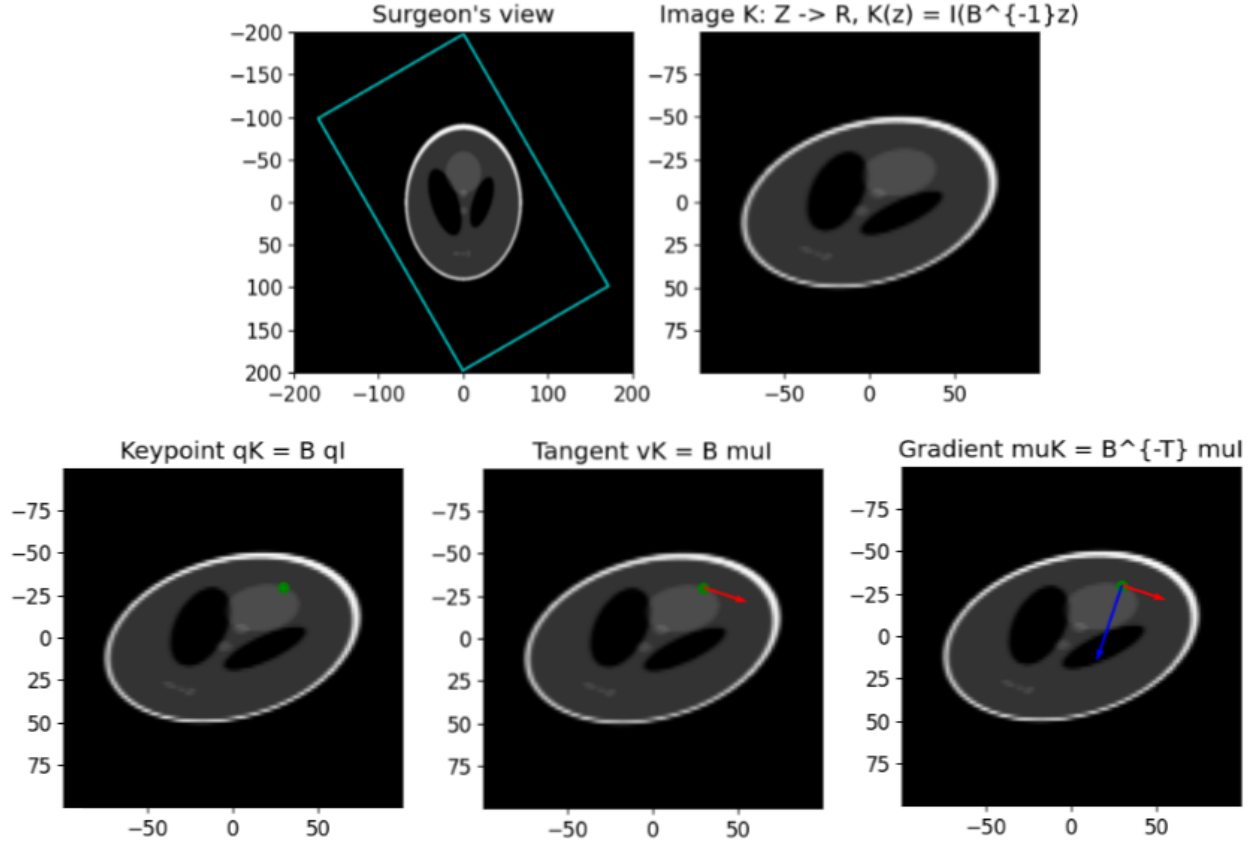
Now, notice that the point may be at the same place relative to the new basis, but the gradient and tangent vectors have shifted.

Explanation

A tangent v connects two nearby points, q_0, q_1. We can define $v = q_1 - q_0$. If we transform the points, $q \to Aq$, we transform the vector.

$$Aq_1 - Aq_0 = A(q_1 - q_0) = Av$$

A gradient points from dark to bright in an image. If we transform the image, $I(x) \to I(A^{-1}x)$, we transform the gradient. Let the gradient at q, $g = \nabla l(q)$ be a column vector, then the chain rule says:

$$\nabla[I(A^{-1}x)]_{x=q} = A^{-T}\nabla I(A^{-1}x)_{x=q}$$

so that the transformed gradient is $A^{-T}g$.

Details

What is the k-th component of $\nabla I(A^{-1}x)$? Let $B = A^{-1}$ and the components of B be written as B_{ij}.

Let (x, y) map to $(B_{00}x + B_{01}y, B_{10}x + B_{11}y)$. To find the gradient, we take the derivative.

$$\frac{\partial}{\partial x}I(B_{00}x + B_{01}y, B_{10}x + B_{11}y)$$

By chain rule, . . .

$$= \partial_0 I(B_{00}x + B_{01}y, B_{10}x + B_{11}y)B_{00} + \partial_1 I(B_{00}x + B_{01}y, B_{10}x + B_{11}y)B_{10}$$

∂_0 means "partial derivative with respect to the 0th argument"

Now, we take the derivative in terms of y.

$$\frac{\partial}{\partial y} I(B_{00}x + B_{01}y, B_{10}x + B_{11}y)$$
$$= \partial_0 I(B_{00}x + B_{01}y, B_{10}x + B_{11}y)B_{01} + \partial_1 I(B_{00}x + B_{01}y, B_{10}x + B_{11}y)B_{11}$$

What we have essentially done here is:

$$\begin{pmatrix} B_{00} & B_{10} \\ B_{01} & B_{11} \end{pmatrix} \cdot \begin{pmatrix} \partial_0 I(B_{00}x + B_{01}y, B_{10}x + B_{11}y) \\ \partial_1 I(B_{00}x + B_{01}y, B_{10}x + B_{11}y) \end{pmatrix}$$
$$= B^T \cdot \nabla I(B_{00}x + B_{01}y, B_{10}x + B_{11}y)$$
$$= A^{-T} \nabla I(B_{00}x + B_{01}y, B_{10}x + B_{11}y)$$

A^{-T} is the inverse and transpose of A. This is equal to B^T, since by our definition, $B = A^{-1}$.

Definition of Vector Space

A set V with two operations $+$ (plus) and $\cdot$ (times, i.e., scalar multiplication) is a vector space if the operations satisfy CANI-ADDU. Here, let $u, v, w \in V$ and $a, b \in \mathbb{R}$.

1. **C**ommutativity: $u + v = v + u$

2. **A**ssociativity: $(u + v) + w = u + (v + w) = u + v + w$

   - $+$ needs to be a function of two variables

3. **N**eutral element: $\exists O \in V$ such that $O + u = u + O = u$

   - This is often the zero vector, or the origin.

4. **I**nverse element: $\exists(-v)$ such that $v + (-v) = (-v) + v = 0$

5. **A**ssociativity: $a \cdot (b \cdot u) = (a \cdot b) \cdot u$, where $(b \cdot u) \to \mathbb{R} \times V \to V$, and $(a \cdot b) \to \mathbb{R} \times \mathbb{R} \to \mathbb{R}$

6. **D**istributivity over $+$ on V: $a \cdot (u + v) = a \cdot u + a \cdot v$

7. **D**istributivity over $\cdot$ on $\mathbb{R}$: $(a + b) \cdot u = a \cdot u + b \cdot u$

8. **U**nitary element: $\exists 1$ such that $1 \cdot u = u$.

Example Vector Spaces

Example: Real Numbers

Real numbers form a vector space with the usual definition of $+$ and $\cdot$.

- a single number, not pairs or triples
- not "direction and magnitude" but still obey vector space axioms

Example: The Cartesian Plane

Set of ordered pairs, $\mathbb{R} \times \mathbb{R}$. $(a, b) \to$ a vector in the cartesion plane.

We can define $+$ and $\cdot$ as the following:

- $+$: $(a, b) + (c, d) = (a + c, b + d)$, where all $a, b, c, d, a + c, b + d \in \mathbb{R}$.
- $\cdot$: $c \cdot (a, b) = (c \cdot a, c \cdot b)$, where $a, b, c, c \cdot a, c \cdot b \in \mathbb{R}$.

Using these rules, we can build a more complex vector space, using definitions from a simpler one.

Example: Arrows on paper with tail at a fixed origin

A vector is an arrow you can draw on paper. We can define + and · as the following:

- +: "translate the second arrow so its tail lines up with the tip of the first arrow, draw a new arrow from the origin to the tip of the second." Basically, add the vectors.

- ·: "shrink or grow the vector by this factor, and if the factor is negative, reflect it over the origin"

Example: Functions from $\mathbb{R} \to \mathbb{R}$

Let f, g be functions $\to$ choose a set of functions with properties we like

- Define $f + g$ by evaluating any point $x \in \mathbb{R}$, define $(f + g)(x) = f(x) + g(x)$. f and g are functions $\mathbb{R} \to \mathbb{R}$.

- Define $a \cdot f$ by evaluating at any point $x \in \mathbb{R}$, define $(a \cdot f)(x) = a \cdot f(x)$, where $(a \cdot f \in V)$ and $a \cdot f(x) \in \mathbb{R}$.

Example: Probabilities

Let p represent the probability that an event happens, and q represent the probability that another event happens.

- +: $p + q$ might be bigger than 1, which is not a probability!

- ·: $(-1) \cdot p$ is less than 0, which is also not a probability!

However, if we use logits, $\log \frac{p}{1-p}$, we can define + and · without any problems. (This will be covered at the end of the course.)

Example: Arrows anywhere on paper

The idea with this one is that we do not necessarily have a fixed origin. Because of this, we do not have a good definition of + and ·. Or at least, that is when we imagine the vectors are on a flat sheet of paper.

Consider if the vectors were drawn on a sphere.

- If two vectors are tangent to the same point, we can add them tip to tail.

- If they are not, then there is no way to add the vectors. We would consider them to be vectors in different vector spaces.

Linear Algebra Definitions

- Linear Independence: A set of vectors v_i are linearly independent if

$$\sum a_i v_i = 0 \iff a_i = 0 \forall i$$

    Otherwise they are linearly dependent.

- Basis: A linear independent set of vectors B is called a basis if any vector in V can be written as a unique linear combination of its elements.

- Dimension: The number of vectors in this set is called the dimension of the vector space.

- Coordinates: If the dimension is finite, we write basis elements as e_i, then

$$v = \sum_i v^i e_i$$

    for $v^i \in \mathbb{R}$. The v^i are the coordinates of the vector with respect to this basis.

- Einstein summation notation: In expressions like the above we will usually remove the summation sign and write

$$v = v^i e_i$$

Any pair of indices with one up and one down is summed.

Color Convention

We will use the convention

- vectors in red

- scalars in black

The previous example will be written

$$\textcolor{red}{v} = v^i \textcolor{red}{e_i}$$

Example: Polynomials

Linear Maps

A linear map $L : U \to V$, is a map between two vector spaces (U, V) that is compatible with addition and scalar multiplication. For $\textcolor{red}{u}, \textcolor{red}{v} \in U$, and $a \in \mathbb{R}$, this means:

1. $+$: $L(\textcolor{red}{u} + \textcolor{red}{v}) = L(\textcolor{red}{u}) + L(\textcolor{red}{v})$

   - "Adding the inputs is the same as adding the outputs"

2. $\cdot$: $L(a \cdot \textcolor{red}{u}) = a \cdot L(\textcolor{red}{u})$

Linear Maps are a vector space.

Definition: The vector space of linear maps:
The set of linear maps from U to V is a vector space, $\mathrm{Hom}(U, V)$ (homomorphisms). We define $+$ and $\cdot$ for each input, borrowing the definition from V. Let $\mathrm{Hom}(U, V) = H, A, B \in H, \textcolor{red}{u} \in U$, and $a \in \mathbb{R}$.

- $+$: $A +_H B$, where $+_H$ is a homomorphism, $(A +_H B)(\textcolor{red}{u}) = A(\textcolor{red}{u}) +_V B(\textcolor{red}{u})$

- $\cdot$: $a \cdot_H A$, $(a \cdot_H A)(\textcolor{red}{u}) = a \cdot_V A(\textcolor{red}{u})$

- e.g., we can add matrices elementwise, and we can scale them elementwise.

Definition: Endomorphism:
The set of linear maps $\mathrm{Hom}(V, V)$ are called endomorphisms.

Covectors

A covector is the vector space $\mathrm{Hom}[V, \mathbb{R}]$, denoted V^* are called covectors. V^* is called the dual space (or just dual) of V. V is called primal.

Example: Row and Column Vectors

Often we write the components of vectors in V as a column. Then the components of V^* are written as a row, and act on the left using matrix multiplication.

Example: Derivatives

Partial derivatives of a function are covectors, and are often written as a row.

Color Convention Pt. 2

We will expand our convention:

- vectors in red
- covectors in blue
- scalars in black.

Vectors will generally be denoted with Roman letters near the end of the alphabet (e.g., u, v, w, x, y, z).

Covectors will generally be denoted with Greek letters (e.g., $\mu, \upsilon, \omega, \chi$)

Scalars will generally be denoted with Roman letters near the beginning of the alphabet (e.g., a, b, c).

A Basic for Covectors

Definition: Dual Basis:
Once a basis for V is fixed (e_i), we can uniquely define a basis of V^*, ϵ^i by the relationship

$$\epsilon^i(e_j) = \delta^i_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Recovering Vector Coordinates

If a vector is written in coordinates as $v = v^i e_i$, then the i-th coordinate is found by

$$v^i = \epsilon^i(v)$$

Proof:

$$\begin{aligned}
\epsilon^i(v) &= \epsilon^i(v^j e_j) \\
&= v^j \cdot \epsilon^i(e_j) \\
&= v^j \delta^i_j \qquad && \text{Definition of dual basis} \\
&= v^i && \text{All terms are 0 except for the } i\text{-th term}
\end{aligned}$$

Recovering Covector Coordinates

If a covector is written in coordinates as $\mu = \mu^i \epsilon^i$, then the i-th coordinate is found by:

$$\mu^i = \mu(e_i)$$

Proof:

$$\begin{aligned}
\mu(e_i) &= (\mu_j e^j)(e_i) \\
&= \mu_j \epsilon^j(e_i) \\
&= \mu_j \delta^j_i \qquad && \text{Definition of dual} \\
&= \mu_i
\end{aligned}$$

Dual of the dual

The vector space of linear maps from $V^* \to \mathbb{R}$ is denoted V^{**}, and is isomorphic to V.

If $\mu \in V^*$, and $w \in V^{**} \simeq V$, then we define

$$w(\mu) \doteq \mu(w)$$

We may use this notation when we consider a sequence of maps acting from right to left. Often we will leave out parentheses.

Example: Dual Basis for Ordered Pairs

Let $V = \mathbb{R}^2$, and a basis be given by $e_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Considering V^* as row vectors, find the dual basis.

Essentially, we are solving for ϵ^0 and ϵ^1.

$$\begin{aligned} \epsilon^0(e_0) &= \delta^0_0 \\ &= 1 = \epsilon^{00} \cdot 1 + \epsilon^{01} \cdot 0 \\ &= \epsilon^{00} \end{aligned}$$

The multiplication by 1 and 0 are the components of e_0.

$$\begin{aligned} \epsilon^0(e_1) &= \delta^0_1 \\ &= 0 = \epsilon^{00} \cdot 1 + \epsilon^{01} \cdot 1 \\ &= \epsilon^{00} + \epsilon^{01} = 0 \end{aligned}$$

In this case, we multiply by 1 for both ϵ terms since the components of e_1 are both 1.

$$\begin{aligned} \epsilon^1(e_0) &= \delta^1_0 \\ &= 0 = \epsilon^{10} \cdot 1 + \epsilon^{11} \cdot 0 \\ &= \epsilon^{10} \end{aligned}$$

$$\begin{aligned} \epsilon^1(e_1) &= \delta^1_1 \\ &= 1 = \epsilon^{10} \cdot 1 + \epsilon^{11} \cdot 1 \\ &= \epsilon^{10} + \epsilon^{11} = 1 \end{aligned}$$

At this point, we have four linear equations and four variables, $\epsilon^{00}, \epsilon^{01}, \epsilon^{11}, \epsilon^{10}$. Solving, we get:

$$\begin{aligned} \epsilon^0 &= (1, -1) \\ \epsilon^1 &= (0, 1) \end{aligned}$$

Adjoint of a Linear Map

Definition:

- If $L : U \to V$, with $v \in U$ and $\mu \in V^*$, then we can define the map $L^* : V^* \to U^*$ implicitly by:

$$\mu(L(v)) \doteq (L^*(\mu))(v)$$

- If f is a function from U to V, then Df (the derivative) is a linear map from U to V, and Df^* is a linear map from V^* to U^*. The latter is used for "backpropagation" when training neural networks.

Example: Adjoint of a Matrix

Let U, V be vector spaces, $u \in U, v \in V^*$, and $A \in \text{Hom}(U, V)$. Let's identify A^*, using matrix multiplication:

- Let u be columns, and v be rows.

$$\mu(Au) \to \mathbb{R}$$

- Notice that A can be applied on the left of the colum input, or ght right of the row, due to how matrix multiplication works.

- "adjoint of A acts to its left on a row"

Now, let both u and v be columns. Then,

$$\mu^T(Au) = (A^T\mu)^T v$$

- "adjoint of A acts with its transpose to the right on a column"

Coordinates of a linear map

Definition:

- If L is a linear map from U to V, we can express its coordinates using the bases of U and V^*.

$$L^i_j = \epsilon^i_V[L(e^U_j)]$$

To build a linear map from coordinates, we can write:

$$L = L^i_j e^V_i \epsilon^j_U$$

Example: Matrix Vector Multiplication

What is the i-th coordinate of $L(v)$?

$$\begin{aligned}
\epsilon^i(L(v)) &= \epsilon^i(L(v^j e_j)) \\
&= v^j \epsilon^i(L(e_j)) \\
&= L^i_j v^j
\end{aligned}$$

Note that for any value mentioned here, exponents are rows, subscripts are columns. This can be written as:

$$\begin{pmatrix} L^0_0 & L^0_1 \\ L^1_0 & L^1_1 \end{pmatrix} \begin{pmatrix} v^0 \\ v^1 \end{pmatrix} = \begin{pmatrix} L^0_0 v^0 + L^0_1 v^1 \\ L^1_0 v^0 + L^1_1 v^1 \end{pmatrix}$$

Coordinates of an adjoint map

Definition:

- If $L*$ is the adjoint of L and a map from V^* to U^*, we can express its coordinates using the bases of V and U^*.

$$[L^*]^i_j = [L^*(\epsilon^i_V)](e^U_j) = \epsilon^i_V[L(e^j_U)] = L^i_j$$

- These are the same coordinates as L, even though this is a map between different vector spaces!

To build an adjoint map from coordinates, we can write:

$$L^* = L^i_j \epsilon^j_U e^V_i$$

We act from right to left, and identify e_i with an element of V^*.

Example: Matrix covector multiplication

What is the i-th coordinate of $L^*(\mu)$?

$$
\begin{aligned}
L^*(\mu)(e_i) &= \mu(L(e_i)) \\
&= \mu_j \epsilon^j(L(e_i)) \\
&= \mu_j L_i^j
\end{aligned}
$$

Notice that:

- This is the same set of formulas as for vectors.

- Instead of multiplying across columns (like vectors), we are multiplying across rows. Specifically, this is similar to multiplication with the transpose, or matrix multiplication to the right of a row vector.

Change of basis

Let A and B be two bases, with elements e_i^A and e_i^B. Their elements can be related by a linear map L, with

$$
L(e_i^A) = e_i^B, \qquad L^{-1}(e_i^B) = e_i^A
$$

The corresponding relationship for covectors is

$$
L^*(\epsilon_i^B) = \epsilon_i^A, \qquad L^{-*}(\epsilon_i^A) = \epsilon_i^B
$$

Components of Change of Basis

- The components are the same in either bases, $L_j^i = \epsilon_A^i e_j^B$

Proof:

In the basis A:
$$
[L_A]_j^i = \epsilon_A^i(L(e_j^A)) = \epsilon_A^i(e_j^B)
$$

In the basis B:
$$
[L_B]_j^i = \epsilon_B^i(L(e_j^B)) = (L^*(\epsilon_B^i)(e_j^B)) = \epsilon_A^i(e_j^B)
$$

Similarly for the inverse:
$$
[L^{-1}]_j^i = \epsilon_B^i(e_j^A)
$$

Components of Basis Vectors

We have that $e_i^B = L_i^j e_j^A$ and $\epsilon_A^i = L_j^i \epsilon_B^j$.

Components of a vector under change of basis

If basis elements change with the map L, vector coordinates change with L^{-1}.

Proof:

Let a^i be components of v in basis A, and b^i be components in basis B, then

$$
b^i = \epsilon_B^i(v) = \epsilon_B^i(a^j e_j^A) = \epsilon_B^i(e_j^A)a^j = (L^{-1})_j^i a^j
$$

- If the basis vectors grow, the components have to shrink to give the same vector.

Components of a covector under change of basis

Relative to vectors, components of covectors transform using the inverse transpose of L.

Proof:

Let α_i be components of μ in basis A, and β_i be components in basis B, then,

$$\mu = \alpha_i \epsilon_A^i = \beta_i \epsilon_B^i$$
$$\beta_i = \mu(e_i^B) = \alpha_j \epsilon_A^j(e_i^B) = \alpha_j L_i^j$$

Note: If L is unitary, these two are the same.

Visualizing a Change of Basis for Vectors

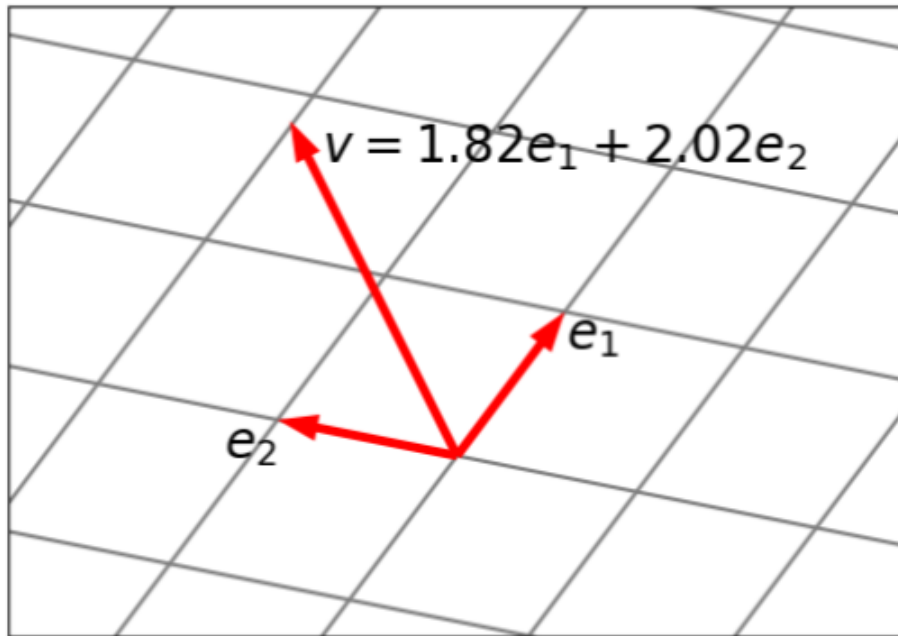First, let's draw some vectors on a coordinate grid. Notice that the two basis vectors are **not** orthonormal.

Components

$$v = 3.64e_1 + 2.02e_2$$
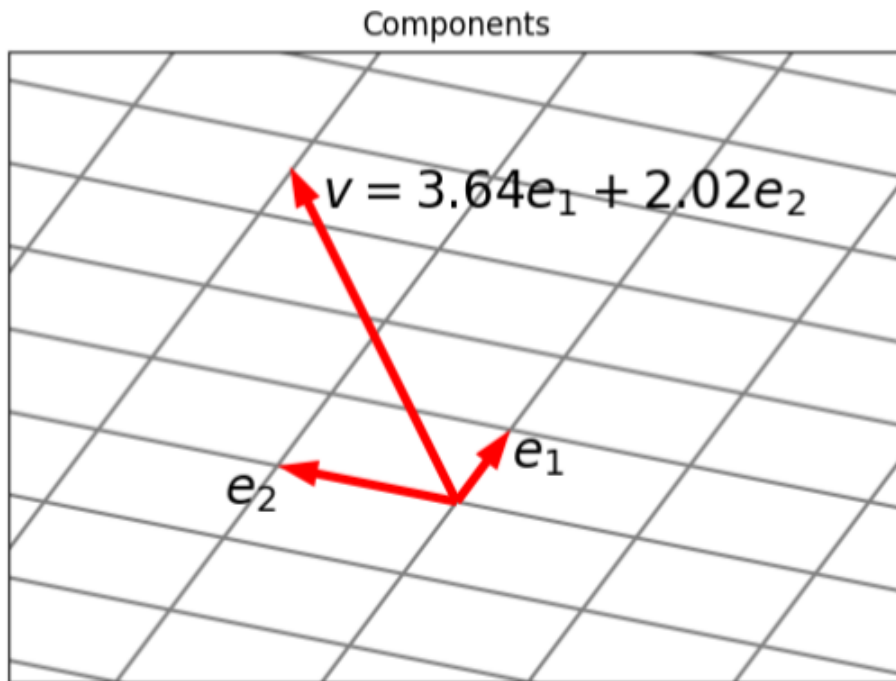
However, we can still define v in terms of e_1 and e_2.

For an example, let's stretch the basis vector e_1 by a factor of 2. Note that now although v is the same vector, its coordinates has changed.

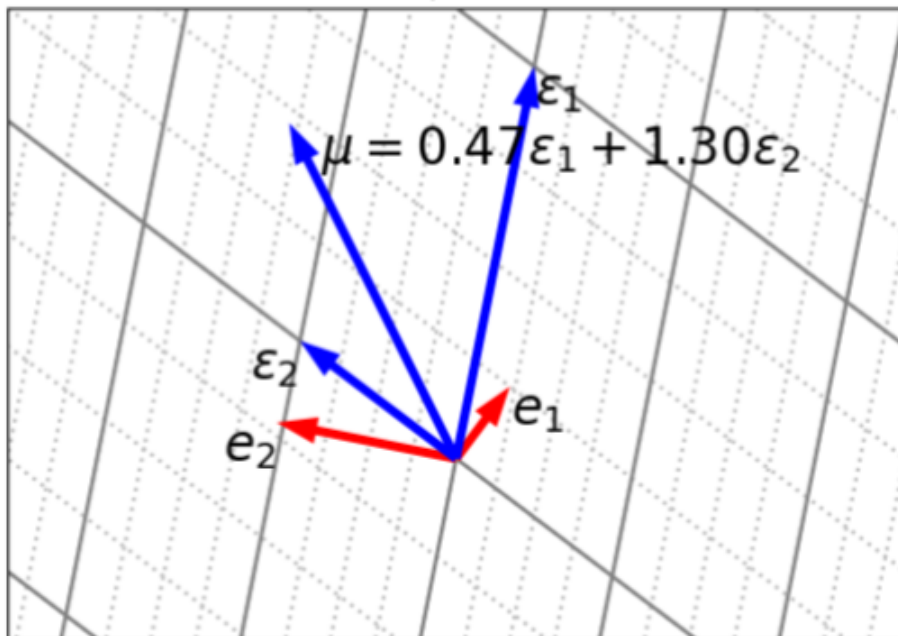Scale e_1 by 2, component scales by 0.5

$$v = 1.82e_1 + 2.02e_2$$

Now, let's add some covectors. (Note that you **cannot** add vectors and covectors.)
We will also add a coordinate plane, but this time in terms of ϵ_1 and ϵ_2.
Additionally, note that vector v is now covector μ. Essentially, that vector is expressed in two different basis, but are the same vector.

Components

$$\mu = 0.47\varepsilon_1 + 1.30\varepsilon_2$$

Now, suppose we stretch e_1 again.

Scale e_1 by 2, component scales by 2

$$\mu = 0.93\varepsilon_1 + 1.30\varepsilon_2$$
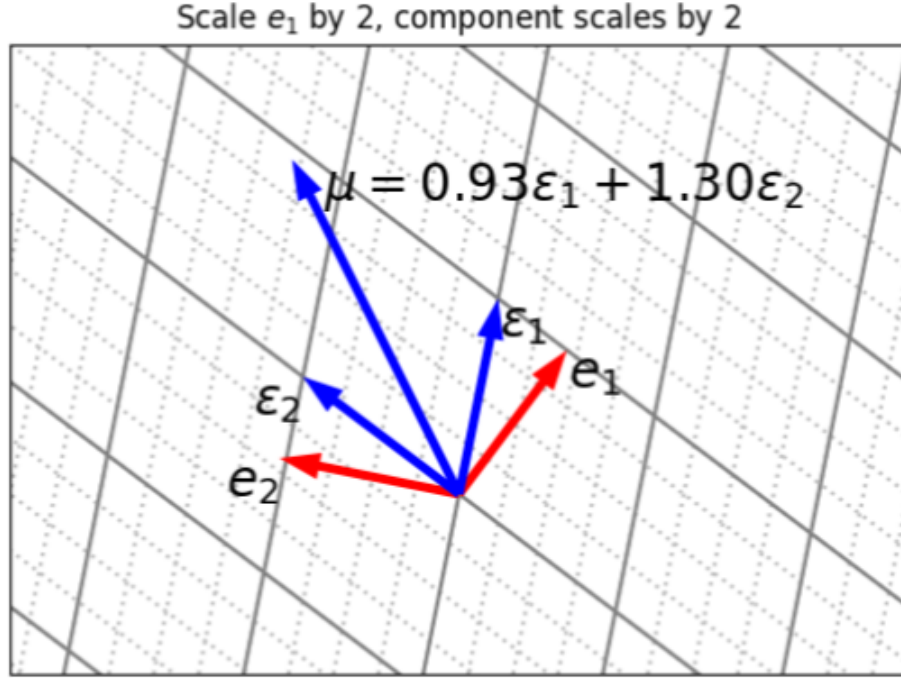
Notice now that this time, ϵ_1 has also changed length.

Summary of Vector and Covector identities

Consider vector spaces, U, V, with $u \in U$, $v \in V$, $\mu \in U^*$, and $v \in V^*$.

	Vectors	Covectors $\mathrm{Hom}(V, \mathbb{R})$
Colors	Red	Blue
Symbols	Roman $(u, v, \ldots)$	Greek $(\mu, \nu, \ldots)$
Basis vectors	e_i	ϵ^i
Dual Basis Def.		$\epsilon^i(e_j) = \delta^i_j$
Coordinates	$u^i = \epsilon^i(u)$	$\mu_i = \mu(e_i)$
Expression	$u = u^i e_i$	$\mu = \mu_i \epsilon^i$
Linear Maps	$L : U \to V$	$L^* : V^* \to U^*$
Adjoint Def.		$\nu(L(u)) \doteq (L^*(\nu))(u)$
Map Coordinates	$L^i_j = \epsilon^i_V(L(e^U_j))$	$(L^*)^i_j = L^i_j$
Map Expression	$L = L^i_j e^V_i \otimes \epsilon^j_U$	$L^* = L^i_j \epsilon^j_U \otimes e^V_i$
	acts partially on U	... on V^*
Coordinate action	$(L(u))^i = L^i_j u^j$	$(L^*(\nu))_i = L^j_i \mu_j$
	matrix multiplication	... with transpose
Basis change $(V = U)$	$L(e^A_i) \doteq e^B_i$	$L^*(\epsilon^i_B) = \epsilon^i_A$
Coordinate change matrix mult.	$u^i_B = [L^{-1}]^i_j u^j_A$	$\mu^B_i = L^j_i \mu^A_j$
	... with inverse	... with transpose

Multilinear Maps

Definition: ((p, q) Tensor)

- A (p, q) tensor is a multilinear map that inputs $p \in \mathbb{N}$ different covectors in V^*, and $q \in \mathbb{N}$ different vectors in V, and outputs a real number.

- Multilinear means it is linear in each input individually, with all the other inputs fixed.

14

- Given a basis of V and a dual basis of V^*, we can write coordinates

$$T = T^{a_1,\ldots,a_p}_{b_1,\ldots,b_q} e_{a_1} \otimes \cdots \otimes e_{a_p} \otimes \epsilon^{b_1} \otimes \cdots \otimes \epsilon^{b_q}$$

where a and b are sequences taking values in $\{0, d-1\}$ where d is the dimension of V.

Example: A (1, 1) Tensor

We can find components $T^i_j = T(\epsilon^i, e_j)$, and write $T = T^i_j e_i \otimes \epsilon^j$.

Its action on a pair of inputs μ, v is:

$$\begin{aligned}
T(\mu, v) &= T(\mu_i \epsilon^i, v^i e_j) \\
&= \mu_i T(\epsilon^i, v^j e_j) \\
&= \mu_i v^j T(\epsilon^i, e_j) \\
&= \mu_i v^j T^i_j
\end{aligned}$$

A linear map can be thought of as a (1, 1) tensor if we allow it to act "partially" on one vector.

Covectors and vectors as tensors

- A covector is an example of a (0, 1) tensor, because it inputs one vector, and outputs a real number.

- A vector is an example of a (1, 0) tensor, because it inputs one covector (thinking of it in V^{**}, and outputs a real number.)

- A scalor in $\mathbb{R}$ is technically a (0, 0) tensor.

Inner Products

Definition: An inner product is a (0, 2) tensor (which means 2 vector input, real number output), $g : V \times V \to \mathbb{R}$, often written $g(\cdot, \cdot) = \langle \cdot, \cdot \rangle$. It must have the following properties:

1. Symmetry: $g(u, v) = g(v, u)$

2. Non degeneracy: $g(u, u) \geq 0$, and it is only equal to 0, if $\mu = \vec{0}$

Lengths and Angles

- **Definition: Length**: We define $\sqrt{g(u, u)} = |u|$ as the length of vector u.

- **Definition: Angle**: The angle between u and v is given by $\cos(\theta) = g\left(\frac{u}{|u|}, \frac{v}{|v|}\right)$

Note that $g(u, v) = |u||v| \cdot \cos(\theta) = u \cdot v$.

Euclidian Distance

Often the specification of a dot product is ignored, and is chosen implicitly as "multiply the components and add them up". This corresponds to:

$$g(u, v) = \delta_{ij} u^i v^i = \sum_i u^i v^i$$

This depends on a choice of basis!

- We cannot write that sum down in Einstein convention, because there is no up/down index!

- We can 'fix' the definition by adding the components of a (0, 2) tensor, $T_{ij} = \delta_{ij}$, where T_{ij} is like an identity matrix.

Example: Polynomials on [0, 1]

Let V be the space of order $d-1$ polynomials from $[0,1] \to \mathbb{R}$, with basis vectors $e_i(x) = x^i$. Let $f, h \in V$ and define

$$\langle f, h \rangle = \int_0^1 f(x)h(x)\mathrm{d}x$$

This integral is basically multiply the components (f and h), and add them up (the integral)

What are the components?

$$g_{ij} = g(e_i, e_j)$$
$$= \int_0^1 x^i x^j \mathrm{d}x$$
$$= \int_0^1 x^{i+j}\mathrm{d}x$$
$$= \frac{1}{i+j+1}x^{i+j+1}\Big|_0^1$$
$$= \frac{1}{i+j+1}$$

In this example, we get $g_{ij} = \frac{1}{i+j+1}$, which is not the identity matrix.

Change of Basis for Inner Products

An inner product matrix transforms under a change of basis differently than a linear map.

Proof:

For a metric g,

$$g_{ij}^B = g(e_i^B, e_j^B)$$
$$= g(L_i^k e_k^A, L_j^l e_l^A)$$
$$= L_i^k L_j^l g(e_k^A, e_l^A)$$
$$= L_i^k L_j^l g_{kl}$$

But for a linear map M,

$$(M^B)_j^i = \epsilon_B^i(M(e_j^B))$$
$$= \epsilon_B^i(M(L_j^k e_k^A))$$
$$= L_j^k \epsilon_B^i M(e_k^A)$$
$$= (L^{-1})_l^i \epsilon_n^l(M(L_j^k e_k^m))$$
$$= (L^{-1})_l^i L_j^k \epsilon_A^l(M(e_k^A))$$
$$= (L^{-1})_l^i L_j^k M_k^l$$

Notice that when you change components, these turn out to be very different.

Example: The Mahalanobis Distance

Choosing an inverse covariance matrix as a metric distance is called the Mahalanobis distance.

Principal Component Analysis (PCA) allows us to find a basis in which this metric is the identity matrix.

- E.g., find a new basis, such that in this basis, the metric is the identity

The Musical Map (something we use an inner product for)

Definition: The Musical Map: Given an inner product, g, we define the flat map $\flat : V \to V^*$ by

$$\flat(u)(v) = g(u, v)$$

In coordinates, we write

$$
\begin{aligned}
\flat(u)_i &= \flat(u)(e_i) \\
&= g(u, e_i) \\
&= g(u^j e_j, e_i) \\
&= u^j g(e_j, e_i) \\
&= u^j g_{ij}
\end{aligned}
$$

It lowers indices from up to down (like a flat lowers a musical note).

The sharp map $\sharp : V^* \to V$ is defined as its inverse.

Uses of the Musical Map

1. Convert between covectors (gradients) and vectors (parameter perturbations) for "natural gradient descent". In complicated spaces, we can specify this instead of the inner product.