Def.- (Vector Space) A vector space over a field F is a set V together with two operations:

1) Addition : $+ : V \times V \to V$ (Takes two elements of V and returns another element of V)

2) Scalar Multiplication : $\cdot : F \times V \to V$

And must satisfy eight particular axioms (associativity of +, commutativity of +, etc.)

⊛ We refer to elements of V as vectors and elements of F as scalars.

Def.- (Linear Transformation) A linear transformation between two vector spaces V and W is a mapping $T: V \to W$ such that:

1) $T(v_1 + v_2) = T(v_1) + T(v_2)$, v_1 and $v_2 \in V$

2) $T(\alpha v) = \alpha T(v)$ for any scalar α.

We can represent linear transformations as matrices. Let

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ a_{21} & \cdots & a_{2k} \\ \vdots & \ddots & \vdots \\ a_{n1} & & a_{nk} \end{bmatrix}$$

then $A : \mathbb{R}^k \to \mathbb{R}^n$. The number of columns of a matrix encodes the number of basis vectors of the input space. In particular, every column shows where does the standard basis vectors move.

Example:

1) $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$

- $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \overbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}^{\hat{i}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

- $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \overbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}^{\hat{j}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

Then, what linearity implies is that

$$A \begin{bmatrix} 5 \\ 3 \end{bmatrix} = A\left(5\hat{i} + 3\hat{j}\right) = 5\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5\cdot 1 + 3\cdot 1 \\ 5\cdot 1 + 3\cdot 2 \end{bmatrix}$$

So matrix multiplication is performed by "projecting" the rows of a matrix onto every column.

2) $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix}$

- $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
- $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$

Then, linearity again implies

$$A \begin{bmatrix} 5 \\ 3 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 5.1 + 3.1 \\ 5.1 + 3.2 \\ 5.1 + 3.0 \end{bmatrix}$$

We can conclude that in general, matrix with matrix multiplication consist on projections of the rows of some matrix A onto the columns of some other matrix B.

$$A = \begin{bmatrix} a_1' {}_{1\times p} \\ \vdots \\ a_n {}_{1\times p} \end{bmatrix}_{n\times p} \qquad B = \begin{bmatrix} b_1 {}_{p\times 1} & \cdots & b_k {}_{p\times 1} \end{bmatrix}_{p\times k}$$

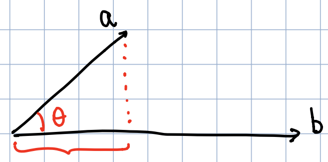
$$AB = \begin{bmatrix} a_1 \cdot b_1 & \cdots & a_1 \cdot b_k \\ \vdots & & \\ \underbrace{a_n \cdot b_1} & \cdots & a_n \cdot b_k \end{bmatrix}_{n\times k}$$

for all these operations to be well defined we need that the row vectors of A have the same size as the column vectors of B.

Dot Products and Projections:

The dot product between vectors a and b is computed as

$$a \cdot b = \|a\| \cos(\theta) \|b\|$$



This size is $\cos(\theta)\ \|a\|$

Notice that the dot product gives the size of the projection but scaled by the size of vector b. To get the value of this projection we normalize it by dividing the size of b.

Notice, however, that this yields a scalar. To make it point in the direction of b we can use b to create a unit vector in that direction. That is,

$$\text{proj}_b\, a = \|a\| \cos(\theta) \underbrace{\frac{b}{\|b\|}}_{\substack{\text{unit vector} \\ \text{pointing in direction of } b}}$$

$$= \underbrace{\frac{a \cdot b}{\|b\|^2}}_{} \, b$$

looks like univariate regression! We are computing the projection of a into the subspace created by the vector b.

⊛ Notice that $\cos\theta = \frac{a \cdot b}{\|a\| \, \|b\|}$ so it doesn't depend on the scale of a and b.
this is actually a measure of correlation: $\cos(\theta) \in [-1, 1]$.

Vector Transformation Rules:

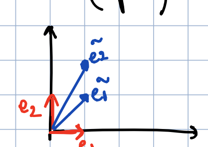
The linear mapping A is what we use to compute a new basis given by the columns of A from an old basis.

$$\begin{tikzcd} \text{OLD BASIS } \{e_1, e_2\} \arrow[d, bend right, "A"'] \\ \text{New Basis } \{\tilde{e}_1, \tilde{e}_2\} \arrow[u, bend right, "A^{-1}"'] \end{tikzcd}$$

"Covariant"

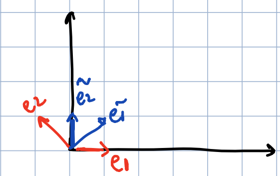
Example: • $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $e_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$ $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$

$$\tilde{e}_1 = A e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \tilde{e}_2 = A e_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$



- $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\quad e_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ $\quad A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$

$$\tilde{e}_1 = Ae_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \tilde{e}_2 = Ae_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



OLD COMPONENTS in $\{e_1, e_2\}$

$A \uparrow \qquad\qquad \downarrow A^{-1}$ "Contra Variant"

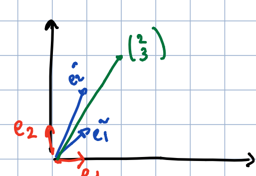
NEW COMPONENTS in $\{\tilde{e}_1, \tilde{e}_2\}$

Examples:
- $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\quad e_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$ $\quad A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$

$$\tilde{e}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \tilde{e}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Then, how do we express $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ in the $\{e_1, e_2\}$ basis into the new basis? We multiply the inverse of A, although it may seem unintuitive.

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$



So in the perspective of the new basis it looks like the vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

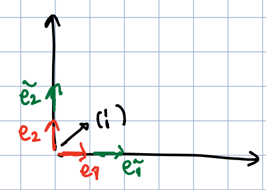
that is,

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} = 1\,\tilde{e}_1 + 1\,\tilde{e}_2 .$$

Does this make sense? Uhm, yes.

Consider two cases

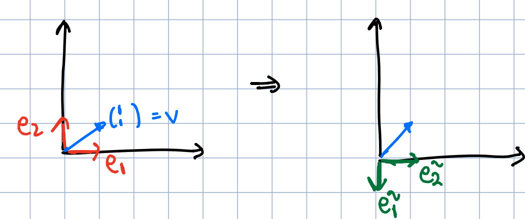
1) Pure Scale: $A = 2I = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



The vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ in the perspective of the new basis feels half of what it feels in the old basis.

$$A^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \text{ in the new basis.}$$

2) Pure rotation: $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$



$$A^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

If the basis rotates clockwise then it feels as if v rotated counterclockwise in this new basis. That is why we say that vector components are contravariant and basis are covariant. (*) The opposite happens in a dual space!

Now I hope you see the connection to this equation:

generalized inverse →

$$X^{-} y = \hat{\beta}_y$$

$$\Leftrightarrow$$

$$(X'X)^{-1} X' y = \hat{\beta}_y$$

so $\hat{\beta}_y$ is the way we express y based on the perspective given by a basis constructed using columns of X.

Vector and Matrix Differentiation Rules

I) Scalar numerator

We start by considering a scalar function or scalar field that take vectors $x \in \mathbb{R}^n$ as input. Then, define

$$\frac{\partial\, f(x)_{1\times 1}}{\partial\, x_{n\times 1}} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

partial derivatives along a column!

As a special case we have $f(x) = a'x$, therefore

$$\frac{\partial\, a'x}{\partial x} = \begin{bmatrix} \frac{\partial a'x}{\partial x_1} \\ \vdots \\ \frac{\partial a'x}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a.$$

And recall that $a'x = x'a$ because it's a scalar, so that

$$\frac{\partial\, x'a}{\partial x} = \begin{bmatrix} \frac{\partial x'a}{\partial x_1} \\ \vdots \\ \frac{\partial x'a}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a.$$

Similarly,

$$\frac{\partial\, f(x)_{1\times 1}}{\partial\, x'_{1\times n}} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \cdots & \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

$$= \left[\frac{\partial f(x)}{\partial x}\right]'$$

partial derivatives along a row!

RULE 1:

$$\boxed{\frac{\partial f(x)_{1\times 1}}{\partial x'_{1\times k}} = \left[\frac{\partial f(x)_{1\times 1}}{\partial x_{k\times 1}}\right]'}$$

II) Vector numerator

Let A be a $m\times n$ matrix, $A = \begin{pmatrix} a_1'{}_{1\times n} \\ \vdots \\ a_m'{}_{1\times n} \end{pmatrix}$ where $a_j \in \mathbb{R}^n$ for $j=1,\dots,m$.

$$\frac{\partial \, Ax_{m\times 1}}{\partial x'_{1\times n}} = \begin{pmatrix} \frac{\partial a_1'x}{\partial x'}{}_{1\times 1} \\ \vdots \\ \frac{\partial a_m'x}{\partial x'}{}_{1\times 1} \end{pmatrix} = \begin{pmatrix} a_1' \\ \vdots \\ a_m' \end{pmatrix} = A_{m\times n}$$

we know how to compute each one of these

$$\frac{\partial \, x'A'_{1\times m}}{\partial x_{n\times 1}} = \begin{pmatrix} \frac{\partial x'a_1}{\partial x} & \cdots & \frac{\partial x'a_m}{\partial x} \end{pmatrix}$$

$$= \begin{pmatrix} a_1 & \cdots & a_m \end{pmatrix} = A'_{n\times m}.$$

III) Chain Rules on I and II

Multivariate chain rule: let $f(x)$ and $x(\alpha)$, $\alpha \in \mathbb{R}$

$$\frac{\partial f(x)}{\partial \alpha}{}_{1\times 1} = \frac{\partial f(x)}{\partial x_1}\frac{\partial x_1}{\partial \alpha} + \dots + \frac{\partial f(x_n)}{\partial x_n}\frac{\partial x_n}{\alpha}$$

$$= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}{}_{1\times 1}\frac{\partial x_i}{\partial \alpha}{}_{1\times 1} \equiv \underbrace{\frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial \alpha}}$$

Some people use this notation where i denotes the dummy indices. It's also known as Einstein's summation.

Let $\alpha \in \mathbb{R}^r$ and $x = x(\alpha)$. Then

$$\frac{\partial x_{n\times 1}}{\partial \alpha'_{1\times r}} = \begin{pmatrix} \frac{\partial x_1}{\partial \alpha'}{}_{1\times r} \\ \vdots \\ \frac{\partial x_n}{\partial \alpha'} \end{pmatrix}_{n\times r}$$

What now?

Notice

- $$\underset{r\times 1}{\frac{\partial f(x)_{1\times 1}}{\partial \alpha}} = \begin{pmatrix} \frac{\partial f(x)}{\partial \alpha_1} \\ \vdots \\ \frac{\partial f(x)}{\partial \alpha_r} \end{pmatrix}$$

use multivariate chain rule

$$= \begin{pmatrix} \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_1} \\ \vdots \\ \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_r} \end{pmatrix}$$

$$= \begin{pmatrix} \left(\frac{\partial f}{\partial x}\right)' \frac{\partial x}{\partial \alpha_1} \\ \vdots \\ \left(\frac{\partial f}{\partial x}\right)' \frac{\partial x}{\partial \alpha_r} \end{pmatrix}$$

we can transpose scalars!

$$= \begin{pmatrix} \frac{\partial x'}{\partial \alpha_1} \frac{\partial f}{\partial x} \\ \vdots \\ \frac{\partial x'}{\partial \alpha_r} \frac{\partial f}{\partial x} \end{pmatrix} = \begin{pmatrix} \frac{\partial x'}{\partial \alpha_1} \\ \vdots \\ \frac{\partial x'}{\partial \alpha_r} \end{pmatrix} \frac{\partial f}{\partial x}$$

$$= \underbrace{\frac{\partial x'_{1\times n}}{\partial \alpha_{r\times 1}}}_{r\times n} \underbrace{\frac{\partial f_{1\times 1}}{\partial x_{n\times 1}}}_{n\times 1}.$$

- $$\frac{\partial f(x)_{1\times 1}}{\partial \alpha'_{1\times r}} = \begin{pmatrix} \frac{\partial f(x)}{\partial \alpha_1} & \cdots & \frac{\partial f(x)}{\partial \alpha_r} \end{pmatrix}$$

use multivariate chain rule

$$= \begin{pmatrix} \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_1} & \cdots & \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_r} \end{pmatrix}$$

$$= \begin{pmatrix} \left(\frac{\partial f}{\partial x}\right)' \frac{\partial x}{\partial \alpha_1} & \cdots & \left(\frac{\partial f}{\partial x}\right)' \frac{\partial x}{\partial \alpha_r} \end{pmatrix}$$

$$= \left(\frac{\partial f}{\partial x}\right)' \begin{pmatrix} \frac{\partial x}{\partial \alpha_1} & \cdots & \frac{\partial x}{\partial \alpha_r} \end{pmatrix}$$

$$= \underbrace{\frac{\partial f_{1\times 1}}{\partial x'_{1\times n}}}_{1\times n} \underbrace{\frac{\partial x_{n\times 1}}{\partial \alpha'_{1\times r}}}_{n\times r}$$

Special cases of the previous rules are

$$\frac{\partial\, a'x_{\,1\times1}}{\partial \alpha_{\,r\times1}} = \frac{\partial x'}{\partial \alpha}\frac{\partial a'x}{\partial x} = \frac{\partial x'}{\partial \alpha}\, a \, .$$

$$\frac{\partial\, a'x_{\,1\times1}}{\partial \alpha'_{\,1\times r}} = \frac{\partial a'x}{\partial x'}\frac{\partial x}{\partial \alpha'} = a'\frac{\partial x}{\partial \alpha'} \, .$$

Rule 2 : $\dfrac{\partial f(x(\alpha))_{\,1\times1}}{\partial \alpha_{\,r\times1}} = \dfrac{\partial x'}{\partial \alpha}\dfrac{\partial f}{\partial x}$

This allows us to generalize to the derivative of a vector

- $$\frac{\partial\, Ax_{\,m\times1}}{\partial \alpha'_{\,1\times r}} = \begin{pmatrix} \dfrac{\partial (a_1'x)}{\partial \alpha'}_{\substack{1\times1\\1\times r}} \\ \vdots \\ \dfrac{\partial (a_m'x)}{\partial \alpha'} \end{pmatrix} = \begin{pmatrix} a_1'\dfrac{\partial x}{\partial \alpha'} \\ \vdots \\ a_m'\dfrac{\partial x}{\partial \alpha'} \end{pmatrix}$$
$$= \begin{pmatrix} a_1' \\ \vdots \\ a_m' \end{pmatrix}\frac{\partial x}{\partial \alpha'} = \underset{m\times n}{A}\,\frac{\partial x}{\partial \alpha'}_{\,n\times r}$$

- $$\frac{\partial\, x'A'_{\,1\times m}}{\partial \alpha_{\,r\times1}} = \begin{pmatrix} \dfrac{\partial x'a_1}{\partial \alpha} & \cdots & \dfrac{\partial x'a_m}{\partial \alpha} \end{pmatrix}$$
$$= \begin{pmatrix} \dfrac{\partial x'}{\partial \alpha}a_1 & \cdots & \dfrac{\partial x'}{\partial \alpha}a_m \end{pmatrix}$$
$$= \frac{\partial x'}{\partial \alpha}\begin{pmatrix} a_1 & \cdots & a_m \end{pmatrix} = \frac{\partial x'}{\partial \alpha}A'$$

Notice that

$$\frac{\partial x}{\partial x'} = \begin{pmatrix} \dfrac{\partial x_1}{\partial x'} \\ \vdots \\ \dfrac{\partial x_n}{\partial x'} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \vdots & & \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I.$$

Rule 3: $$\frac{\partial \, Ax_{\,m\times 1}}{\partial \, \alpha'_{\,1\times r}} = A \, \frac{\partial x}{\partial \alpha'}$$

IV) Chain Rule on Quadratic Forms

Let $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$ and A is $m\times n$. Also let $x = x(\alpha)$, $z = z(\alpha)$. Then

- $$\frac{\partial (z'Ax)}{\partial \alpha} = \begin{pmatrix} \frac{\partial z'Ax}{\partial \alpha_1} \\ \vdots \\ \frac{\partial z'Ax}{\partial \alpha_r} \end{pmatrix} = \begin{pmatrix} \frac{\partial z'Ax}{\partial x'} \frac{\partial x}{\partial \alpha_1} + \frac{\partial z'Ax}{\partial z'} \frac{\partial z}{\partial \alpha_1} \\ \vdots \\ \frac{\partial z'Ax}{\partial x'} \frac{\partial x}{\partial \alpha_r} + \frac{\partial z'Ax}{\partial z'} \frac{\partial z}{\partial \alpha_r} \end{pmatrix}$$

use the fact that we can transpose scalars.

$$= \begin{pmatrix} \frac{\partial x'}{\partial \alpha_1} \frac{\partial z'Ax}{\partial x} + \frac{\partial z'}{\partial \alpha_1} \frac{\partial z'Ax}{\partial z} \\ \vdots \\ \frac{\partial x'}{\partial \alpha_r} \frac{\partial z'Ax}{\partial x} + \frac{\partial z'}{\partial \alpha_r} \frac{\partial z'Ax}{\partial z} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial x'}{\partial \alpha_1} \frac{\partial z'Ax}{\partial x} \\ \vdots \\ \frac{\partial x'}{\partial \alpha_r} \frac{\partial z'Ax}{\partial x} \end{pmatrix} + \begin{pmatrix} \frac{\partial z'}{\partial \alpha_1} \frac{\partial z'Ax}{\partial z} \\ \vdots \\ \frac{\partial z'}{\partial \alpha_r} \frac{\partial z'Ax}{\partial z} \end{pmatrix}$$

Trick: differentiate z' then transpose to have x' at LHS and differentiate.

$$= \frac{\partial x'}{\partial \alpha} A'z + \frac{\partial z'}{\partial \alpha} Ax.$$

As a special case we have

$$\frac{\partial x'Ax}{\partial x} = \frac{\partial x' \overbrace{Ax}^{c}}{\partial x} + \frac{\partial \overbrace{x'A}^{d'} x}{\partial x}$$

$$= \underbrace{Ax}_{c} + \underbrace{A'x}_{d} = (A + A')x$$

Rule 4: $$\frac{\partial z'Ax}{\partial \alpha} = \frac{\partial x'}{\partial \alpha} A'z + \frac{\partial z'}{\partial \alpha} Ax$$

V) <u>Second Derivative</u>

- $$\underset{k\times 1}{\frac{\partial}{\partial \alpha}} \left[\underset{1\times \ell}{z'} \underset{\ell\times\ell}{A} \underset{1\times k}{\frac{\partial x_{\ell\times 1}}{\partial \alpha'}} \right] = \frac{\partial}{\partial \alpha} \left[z'A \frac{\partial x}{\partial \alpha_1} \quad \cdots \quad z'A \frac{\partial x}{\partial \alpha_k} \right]$$

$$= \left[\frac{\partial z'}{\partial \alpha} A \frac{\partial x}{\partial \alpha_1} + \frac{\partial^2 x'}{\partial \alpha \partial \alpha_1} A' z \quad \cdots \quad \frac{\partial z'}{\partial \alpha} A \frac{\partial x}{\partial \alpha_k} + \frac{\partial^2 x'}{\partial \alpha \partial \alpha_k} A' z \right]$$

$$= \frac{\partial z'}{\partial \alpha} A \frac{\partial x}{\partial \alpha'} + \left[\frac{\partial^2 x'}{\partial \alpha \partial \alpha_1} \cdots \frac{\partial^2 x'}{\partial \alpha \partial \alpha_k} \right] \begin{bmatrix} A'z & \cdots & 0 \\ \vdots & A'z & \vdots \\ 0 & \cdots & A'z \end{bmatrix}$$

WRONG NAIVE WAY $$= \frac{\partial z'}{\partial \alpha} A \frac{\partial x}{\partial \alpha'} + \underbrace{\frac{\partial^2 x'_{1\times\ell}}{\partial \alpha'_{1\times k} \partial \alpha_{k\times 1}}} \left[I_k \otimes A'z \right]$$

Does this even make sense?

RIGHT WAY $$= \underbrace{\frac{\partial z'}{\partial \alpha} A \frac{\partial x}{\partial \alpha'}}_{k\times k} + \underset{k\times 1}{\frac{\partial}{\partial \alpha}} \underbrace{\mathrm{vec}\left(\frac{\partial x}{\partial \alpha'} \right)'}_{1\times k\ell} \underbrace{\left[I_k \otimes A'x \right]}_{k\ell\times k}$$

where

$$\mathrm{vec}\left(\frac{\partial x}{\partial \alpha'} \right) = \mathrm{vec}\left(\frac{\partial x}{\partial \alpha_1} \quad \cdots \quad \frac{\partial x}{\partial \alpha_k} \right)$$

$$= \begin{pmatrix} \frac{\partial x}{\partial \alpha_1} \\ \vdots \\ \frac{\partial x}{\partial \alpha_k} \end{pmatrix}.$$

Rule 5: $$\frac{\partial}{\partial \alpha} \left[z'A \frac{\partial x}{\partial \alpha'} \right] = \frac{\partial z'}{\partial \alpha} A \frac{\partial x}{\partial \alpha'} + \frac{\partial}{\partial \alpha} \mathrm{vec}\left(\frac{\partial x}{\partial \alpha'} \right)' \left[I_k \otimes A'z \right]$$

EXAMPLE 1: Linear GMM

Let X be $n \times k$, Z be $n \times l$, $\theta \in \mathbb{R}^k$, $y \in \mathbb{R}^n$, and W be $l \times l$ and symmetric.
The sample criterion function is

$$Q_n(\theta) = \frac{1}{2}\left[Z'(Y - X\theta)\right]' W \left[Z'(Y - X\theta)\right]$$

- $$\frac{\partial Q_n(\theta)}{\partial \theta} \underset{\substack{\text{RULE} \\ 4}}{=} \frac{\partial}{\partial \theta}\left[Z'(Y - X\theta)\right]' W \left[Z'(Y - X\theta)\right]$$

$$= \frac{\partial}{\partial \theta}\left[-\theta' X' Z + Y' Z\right] W \left[Z'(Y - X\theta)\right]$$

$$\underset{\substack{\text{RULE} \\ 3}}{=} -X' Z W Z'(Y - X\theta) = 0$$

- $$\frac{\partial Q_n(\theta)}{\partial \theta \, \partial \theta'} = \frac{\partial}{\partial \theta}\left[\theta' X' Z W Z' X - Y' Z W Z' X\right]$$

$$= X' Z W Z' X$$

Maybe something insightful can be gained if we rewrite

$$Z'(Y - X\theta) = \begin{bmatrix} z_1 & \dots & z_n \end{bmatrix}_{l \times n} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}_{n \times 1} = \sum_{i=1}^{n} z_i u_i$$

Then

- $$Q_n(\theta) = \frac{1}{2}\left[\sum_{i=1}^{n} z_i (y_i - x_i'\theta)\right]' W \left[\sum_{i=1}^{n} z_i (y_i - x_i'\theta)\right]$$

- $$\frac{\partial}{\partial \theta} Q_n(\theta) = \frac{\partial}{\partial \theta}\left[\sum_{i=1}^{n}\left(-\underset{1\times k}{\theta'}\underset{k\times 1}{x_i}\underset{1\times l}{z_i'} + y_i \underset{1\times l}{z_i'}\right)\right] W \left[\sum_{i=1}^{n} z_i (y_i - x_i'\theta)\right]$$

$$= -\left[\sum_{i=1}^{n} x_i z_i'\right] W \left[\sum_{i=1}^{n} z_i u_i\right] \qquad \text{(A)}$$

- $$\frac{\partial}{\partial \theta}\frac{\partial Q_n(\theta)}{\partial \theta'} = -\frac{\partial}{\partial \theta}\left\{\left[\sum_{i=1}^{n} (y_i - x_i'\theta) z_i'\right] W \left[\sum_{i=1}^{n} x_i' z_i\right]\right\}$$

$$= \left[\sum_{i=1}^{n} x_i z_i'\right] W \left[\sum_{i=1}^{n} x_i' z_i\right] \qquad \text{(B)}$$

Notice that

$$\sqrt{n}\,(\hat{\theta}_n - \theta) = \Big(\underbrace{\Big[\frac{1}{n}\sum_{i=1} X_i Z_i'\Big] W \Big[\frac{1}{n}\sum_{i=1}^{n} x_i' z_i\Big]}_{(B)}\Big)^{-1} \underbrace{\Big[\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\Big] W \frac{1}{\sqrt{n}}\sum_{i=1}^{n}}_{(A)} Z_i u_i + o_p(1)$$

(B) $\Downarrow$ The Hessian is closely connected to scaling a.k.a the variance.

(A) $\Downarrow$ The score or gradient is closely connected to the asymptotic distribution!

Coincidence? You will see it all eventually in this course. Be patient.

EXAMPLE 2: Non-linear GMM

Consider the criterion function $Q_n(\theta) = \frac{1}{2}\Big[\frac{1}{n}\sum_{i=1}^{n} \underbrace{g(W_i,\theta)'}_{1\times\ell}\Big] \underbrace{A'A}_{\substack{\ell\times\ell \\ \text{symmetric}}} \Big[\frac{1}{n}\sum_{i=1}^{n} \underbrace{g(W_i,\theta)}_{\ell\times 1}\Big]$

where W_i is the data (y_i, x_i', z_i').

- $$\frac{\partial Q_n(\theta)_{1\times1}}{\partial \theta_{k\times1}} = \Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)'_{1\times\ell}}{\partial\theta_{k\times1}}\Big]_{k\times\ell} A'A \Big[\frac{1}{n}\sum_{i=1}^{n} g(W_i,\theta)\Big]_{\ell\times1}$$

- $$\frac{\partial Q_n(\theta)}{\partial\theta\,\partial\theta'} = \frac{\partial}{\partial\theta}\Big\{\Big[\frac{1}{n}\sum_{i=1}^{n} g(W_i,\theta)'\Big] A'A \Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)}{\partial\theta'}\Big]\Big\}$$

$$= \Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)'}{\partial\theta}\Big] A'A \Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)}{\partial\theta'}\Big] +$$

RULE 5

$$\frac{\partial}{\partial\theta}\Big[\frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathrm{vec}\Big(\frac{\partial g(W_i,\theta)}{\partial\theta'}\Big)'}_{1\times k\ell}\Big]\Big[\underbrace{I_k}_{k\times k} \otimes A'A \frac{1}{n}\sum_{i=1}^{n} \underbrace{g(W_i,\theta)}_{\ell\times1}\Big]$$

$$= \Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)'}{\partial\theta}\Big] A'A \Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)}{\partial\theta'}\Big] +$$

$$\Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\mathrm{vec}\Big(\frac{\partial g(W_i,\theta)'}{\partial\theta}\Big)'\Big]\Big[I_k \otimes A'A \frac{1}{n}\sum_{i=1}^{n} g(W_i,\theta)\Big]$$