

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

1) Addition: $+ : V \times V \rightarrow V$ [Takes two elements of V and returns another element of V]

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

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

④ 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

$T : V \rightarrow W$ such that:

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

$$2) T(\alpha v) = \alpha T(v) \text{ for any scalar } \alpha.$$

We can represent linear transformations as matrices. Let

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

then $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$. 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 :

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

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

Then, what linearity implies is that

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

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

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

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

Then, linearity again implies

$$A \begin{bmatrix} 5 \\ 3 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \cdot 1 + 3 \cdot 1 \\ 5 \cdot 1 + 3 \cdot 2 \\ 5 \cdot 1 + 3 \cdot 0 \end{bmatrix}$$

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 .

$$A = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}_{n \times p} \quad B = \begin{bmatrix} b_1 & \dots & b_k \end{bmatrix}_{p \times k}$$

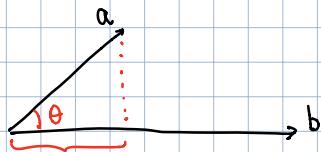
$$AB = \begin{bmatrix} a_1 \cdot b_1 & \dots & a_1 \cdot b_k \\ \vdots & & \vdots \\ a_n \cdot b_1 & \dots & a_n \cdot b_k \end{bmatrix}_{n \times k}$$

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,

$$\text{proj}_b a = \|a\| \cos(\theta) \frac{b}{\|b\|}$$

unit vector
pointing in direction of b

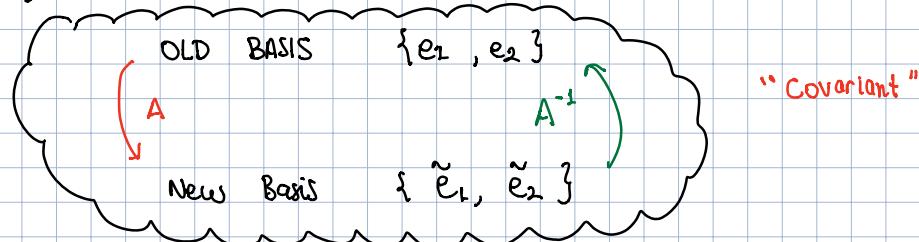
$$= \frac{a \cdot b}{\|b\|^2} b$$

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

- ④ 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.



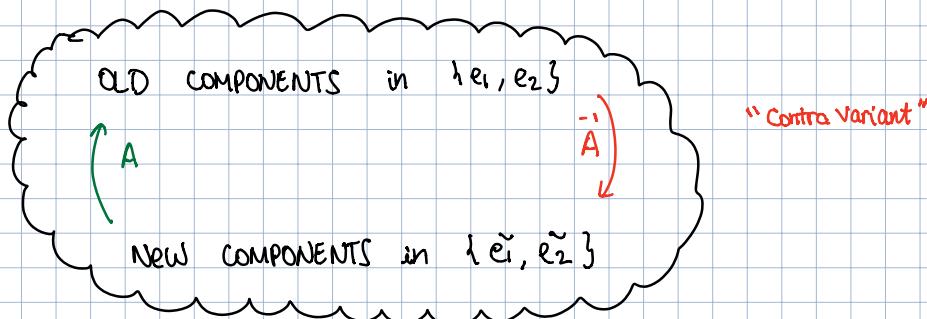
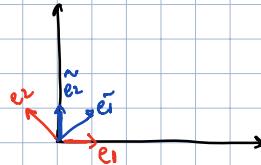
Example: . $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $e_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$ $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$

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



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

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



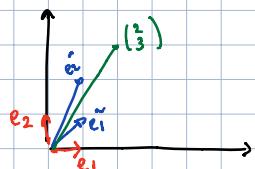
$$\text{Examples: } \cdot \quad e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad e_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

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

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 $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

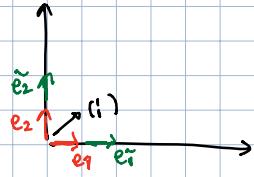
that is,

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

Does this make sense? Uhm, yes.

Consider two cases

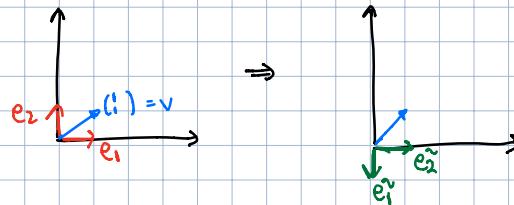
1) Pure Scale: $A = 2 I = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



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. \textcircled{R} The opposite happens in a dual space!

Detour : Vector and Matrix Differentiation Rules

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

$$\frac{\partial f(x)}{\partial x} \underset{n \times 1}{\underset{\text{partial derivatives along a column!}}{\overrightarrow{\text{}}} \underset{1 \times L}{=}} \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

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,

$$\frac{\partial f(x)}{\partial x'} \underset{1 \times n}{\underset{\text{partial derivatives along a row!}}{\overrightarrow{\text{}}} \underset{L \times 1}{=}} \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \dots & \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \\ = \left[\frac{\partial f(x)}{\partial x} \right]'$$

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

$$\frac{\partial Ax}{\partial x'} = \begin{pmatrix} \frac{\partial a_1' x}{\partial x'} \\ \vdots \\ \frac{\partial a_m' x}{\partial x'} \end{pmatrix} = \begin{pmatrix} a_1' \\ \vdots \\ a_m' \end{pmatrix} = A.$$

 we know how
to compute
each one of these

$$\begin{aligned} \frac{\partial x' A'}{\partial x} &= \left(\frac{\partial x' a_1}{\partial x} \dots \frac{\partial x' a_m}{\partial x} \right) \\ &= (a_1 \dots a_m)' = A'. \end{aligned}$$

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

$$\begin{aligned} \frac{\partial f(x)}{\partial \alpha} \Big|_{x(\alpha)} &= \frac{\partial f(x)}{\partial x_i} \frac{\partial x_i}{\partial \alpha} + \dots + \frac{\partial f(x_n)}{\partial x_n} \frac{\partial x_n}{\partial \alpha} \\ &= \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x(\alpha)} \frac{\partial x_i}{\partial \alpha} \Big|_{x(\alpha)} = \underbrace{\frac{\partial f}{\partial x_i}}_{\text{Some people use this notation}} \frac{\partial x_i}{\partial \alpha} \end{aligned}$$

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

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

Then ,

$$\begin{aligned} \cdot \frac{\partial f(x)}{\partial \alpha_{rx1}} &= \left(\begin{array}{c} \frac{\partial f(x)}{\partial \alpha_1} \\ \vdots \\ \frac{\partial f(x)}{\partial \alpha_r} \end{array} \right) \\ &= \left(\begin{array}{c} \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_1} \\ \vdots \\ \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_r} \end{array} \right) \end{aligned}$$

use
multivariate
chain rule

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

$$\begin{aligned} &= \left(\begin{array}{c} \frac{\partial x'}{\partial \alpha_1} \frac{\partial f}{\partial x} \\ \vdots \\ \frac{\partial x'}{\partial \alpha_r} \frac{\partial f}{\partial x} \end{array} \right) = \left(\begin{array}{c} \frac{\partial x'}{\partial \alpha_1} \\ \vdots \\ \frac{\partial x'}{\partial \alpha_r} \end{array} \right) \frac{\partial f}{\partial x} \\ &= \underbrace{\frac{\partial x'}{\partial \alpha_{rx1}}}^{rxn} \underbrace{\frac{\partial f}{\partial x}}_{nx1}. \end{aligned}$$

we can
transport
scalars!

$$\begin{aligned} \cdot \frac{\partial f(x)_{ix1}}{\partial \alpha'_{ixr}} &= \left(\begin{array}{ccc} \frac{\partial f(x)}{\partial \alpha_1} & \dots & \frac{\partial f(x)}{\partial \alpha_r} \end{array} \right) \\ &= \left(\begin{array}{ccc} \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_1} & \dots & \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_r} \end{array} \right) \\ &= \left(\begin{array}{c} \left(\frac{\partial f}{\partial x} \right)' \frac{\partial x}{\partial \alpha_1} \\ \dots \\ \left(\frac{\partial f}{\partial x} \right)' \frac{\partial x}{\partial \alpha_r} \end{array} \right) \\ &= \left(\frac{\partial f}{\partial x} \right)' \left(\begin{array}{c} \frac{\partial x}{\partial \alpha_1} \\ \dots \\ \frac{\partial x}{\partial \alpha_r} \end{array} \right) \\ &= \underbrace{\frac{\partial f}{\partial x}}_{ixn} \underbrace{\frac{\partial x}{\partial \alpha'}}_{nxr} \end{aligned}$$

use
multivariate
chain rule

Special cases of the previous rules are

$$\frac{\partial \alpha' x}{\partial \alpha'_{rxr}} = \frac{\partial x'}{\partial \alpha} \quad \frac{\partial \alpha' x}{\partial x} = \frac{\partial x'}{\partial \alpha} \quad a .$$

$$\frac{\partial \alpha' x}{\partial \alpha'_{rrr}} = \frac{\partial \alpha' x}{\partial x'} \quad \frac{\partial x}{\partial x'} = a' \frac{\partial x}{\partial \alpha'} .$$

This allows us to generalize to the derivative of a vector

$$\begin{aligned} \frac{\partial Ax}{\partial \alpha'_{rrr}} &= \left(\begin{array}{c} \frac{\partial (a_1' x)}{\partial \alpha'} \\ \vdots \\ \frac{\partial (a_m' x)}{\partial \alpha'} \end{array} \right) = \left(\begin{array}{c} a'_1 \frac{\partial x}{\partial \alpha'} \\ \vdots \\ a'_m \frac{\partial x}{\partial \alpha'} \end{array} \right) \\ &= \left(\begin{array}{c} a'_1 \\ \vdots \\ a'_m \end{array} \right) \frac{\partial x}{\partial \alpha'} = A \frac{\partial x}{\partial \alpha'} \end{aligned}$$

$$\begin{aligned} \frac{\partial x' A'}{\partial \alpha_{rrr}} &= \left(\begin{array}{ccc} \frac{\partial x' a_1}{\partial \alpha} & \dots & \frac{\partial x' a_m}{\partial \alpha} \end{array} \right) \\ &= \left(\begin{array}{ccc} \frac{\partial x'}{\partial \alpha} a_1 & \dots & \frac{\partial x'}{\partial \alpha} a_m \end{array} \right) \\ &= \frac{\partial x'}{\partial \alpha} (a_1 \dots a_m) = \frac{\partial x'}{\partial \alpha} A' \end{aligned}$$

Notice that

$$\frac{\partial I}{\partial x'} = \left(\begin{array}{c} \frac{\partial x_1}{\partial x'} \\ \vdots \\ \frac{\partial x_n}{\partial x'} \end{array} \right) = \left(\begin{array}{cccc} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & 1 \end{array} \right) = I.$$

Let $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$ and A is $m \times n$. Then

$$\bullet \quad \frac{\partial (z' A x)}{\partial x}_{\text{row}} = \frac{\partial (c' x)}{\partial x}, \text{ where } c = A' z$$

$$= c = A' z.$$

$$\bullet \quad \frac{\partial (z' A x)}{\partial z}_{\text{row}} = \frac{\partial (z' d)}{\partial z}, \text{ where } d = Ax$$

$$= d = Ax.$$

$$\frac{\partial (z' A x)}{\partial \alpha} = \begin{pmatrix} \frac{\partial z' A x}{\partial \alpha_1} \\ \vdots \\ \frac{\partial z' A x}{\partial \alpha_r} \end{pmatrix} = \begin{pmatrix} \frac{\partial z' A x}{\partial x'_1} \frac{\partial x'_1}{\partial \alpha_1} + \frac{\partial z' A x}{\partial z'_1} \frac{\partial z'_1}{\partial \alpha_1} \\ \vdots \\ \frac{\partial z' A x}{\partial x'_r} \frac{\partial x'_r}{\partial \alpha_r} + \frac{\partial z' A x}{\partial z'_r} \frac{\partial z'_r}{\partial \alpha_r} \end{pmatrix}$$

use the fact that
we can transpose
scalars.

$$= \begin{pmatrix} \frac{\partial x'_1}{\partial \alpha_1} \frac{\partial z' A x}{\partial x'_1} + \frac{\partial x'_1}{\partial \alpha_1} \frac{\partial z' A x}{\partial z'_1} \\ \vdots \\ \frac{\partial x'_r}{\partial \alpha_r} \frac{\partial z' A x}{\partial x'_r} + \frac{\partial x'_r}{\partial \alpha_r} \frac{\partial z' A x}{\partial z'_r} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial x'_1}{\partial \alpha_1} \frac{\partial z' A x}{\partial x'_1} \\ \vdots \\ \frac{\partial x'_r}{\partial \alpha_r} \frac{\partial z' A x}{\partial x'_r} \end{pmatrix} + \begin{pmatrix} \frac{\partial x'_1}{\partial \alpha_1} \frac{\partial z' A x}{\partial z'_1} \\ \vdots \\ \frac{\partial x'_r}{\partial \alpha_r} \frac{\partial z' A x}{\partial z'_r} \end{pmatrix}$$

$$= \frac{\partial x'_1}{\partial \alpha} A' z + \frac{\partial x'_r}{\partial \alpha} Ax.$$

As a special case we have

$$\frac{\partial x' A x}{\partial x} = \frac{\partial x' A x}{\partial x} \overset{c}{+} \frac{\partial x' A x}{\partial x} \overset{d}{+}$$

$$= \underbrace{Ax}_c + \underbrace{A' x}_d = (A + A')x$$

Notice that we can also differentiate a scalar along two dimensions

$$\begin{aligned}\frac{\partial \mathbb{Z}' A x}{\partial A_{m \times n}} &= \left(\begin{array}{ccc} \frac{\partial \mathbb{Z}' A x}{\partial A_{11}} & \dots & \frac{\partial \mathbb{Z}' A x}{\partial A_{1n}} \\ \vdots & & \\ \frac{\partial \mathbb{Z}' A x}{\partial A_{m1}} & \dots & \frac{\partial \mathbb{Z}' A x}{\partial A_{mn}} \end{array} \right) \\ &= \left(\begin{array}{ccc} z_1 x_1 & \dots & z_n x_n \\ \vdots & & \vdots \\ z_m x_1 & \dots & z_m x_n \end{array} \right) = \mathbb{Z} \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} \\ &= \mathbb{Z} \mathbf{x}'.\end{aligned}$$

Application to GMM criterion function:

Let $\mathbb{Z} \in \mathbb{R}^l$, $x \in \mathbb{R}^k$, $b \in \mathbb{R}^k$, $y \in \mathbb{R}$, and W is $l \times l$ and symmetric.
Then

$$\begin{aligned}\frac{\partial}{\partial b} \left[\underbrace{(\mathbb{Z}(y - x'b))' W (\mathbb{Z}(y - x'b))}_c \right] &= \frac{\partial c'}{\partial b} W c + \frac{\partial c'}{\partial b} W c \\ &= 2 \underbrace{\frac{\partial (\mathbb{Z}(y - x'b))'}{\partial b} W (\mathbb{Z}(y - x'b))}_{\text{because of symmetry}} \\ &= 2 \frac{\partial}{\partial b} [(\mathbb{Z}y - \mathbb{Z}x'b)' W (\mathbb{Z}y - \mathbb{Z}x'b)] \\ &= -2 \frac{\partial b' \cancel{\mathbb{Z}x} \cancel{\mathbb{Z}' W}}{\partial b} W (\mathbb{Z}(y - x'b)) \\ &= -2 x \mathbb{Z}' W (\mathbb{Z}(y - x'b))\end{aligned}$$

Now, the real criterion function uses $\mathbb{Z}_{n \times l}$, $X_{n \times k}$, and W is $l \times l$.
Then,

$$\begin{aligned}\frac{\partial}{\partial b} \left[(\mathbb{Z}' (y - xb))' W (\mathbb{Z}' (y - xb)) \right] &= -2 \frac{\partial b' X' \mathbb{Z}}{\partial b} W (\mathbb{Z}' (y - xb)) \\ &= -2 X' \mathbb{Z} W (\mathbb{Z}' y - \mathbb{Z}' xb)\end{aligned}$$

Example: Non linear GMM

$$\text{Consider the criterion function } Q_n(\theta) = \frac{1}{2} \left[\frac{1}{n} \sum_{i=1}^n g(w_i, \theta)' \underbrace{\text{A'A}}_{\text{LxL}} \underbrace{\left[\frac{1}{n} \sum_{i=1}^n g(w_i, \theta) \right]}_{\text{Lx1}} \right]$$

Symmetric

$$\frac{\partial Q_n(\theta)}{\partial \theta} = \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i, \theta)'}{\partial \theta} \right] A'A \left[\frac{1}{n} \sum_{i=1}^n g(w_i, \theta) \right]$$

$$\frac{\partial^2 Q_n(\theta)}{\partial \theta' \partial \theta} = d' \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i, \theta)'}{\partial \theta} \right] \quad d := \frac{1}{n} \sum \frac{\partial g(w_i, \theta)'}{\partial \theta}$$

+

$$\left[\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta'} \left[\frac{\partial g(w_i, \theta)'}{\partial \theta} \right] \right] C \quad C := A'A \left[\frac{1}{n} \sum_{i=1}^n g(w_i, \theta) \right]$$

$$= \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i, \theta)'}{\partial \theta} \right] A'A \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i, \theta)}{\partial \theta'} \right] \quad \text{LxL} \quad \text{LxL}$$

+
 $\underbrace{\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta'} \left[\frac{\partial g(w_i, \theta)'}{\partial \theta} \right]}$

$$\left[\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta'} \left[\frac{\partial g(w_i, \theta)'}{\partial \theta} \right] \right] A'A \left[\frac{1}{n} \sum_{i=1}^n g(w_i, \theta) \right] \quad \text{LxL}$$

We need a new operator!

(*) Property:

$$M'v = [I_d \otimes v'] \text{vec}(M)$$

where d is the number of columns of M .

$$= \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i, \theta)'}{\partial \theta} \right] A'A \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i, \theta)}{\partial \theta'} \right]$$

+

$$\frac{\partial}{\partial \theta'} \left(I_K \otimes \left(\frac{1}{n} \sum_{i=1}^n g(w_i, \theta)' A'A \right) \right) \left[\frac{1}{n} \sum_{i=1}^n \text{vec} \left(\frac{\partial g(w_i, \theta)}{\partial \theta'} \right) \right] \quad K \times L \quad K \times L$$

only affects this argument

$$\begin{aligned}
&= \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i, \theta)}{\partial \theta} \right] A' A \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i, \theta)}{\partial \theta'} \right] \\
&\quad + \\
&\quad \left[I_k \otimes \left(\frac{1}{n} \sum_{i=1}^n g(w_i, \theta)' A' A \right) \right] \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta'} \text{vec} \left(\frac{\partial g(w_i, \theta)}{\partial \theta} \right) \right]
\end{aligned}$$

and you will notice that now dimensions match, so that we get a $k \times k$ matrix.

To see why the Vec operator property holds, consider:

$$M = \begin{bmatrix} m_1 & \dots & m_d \end{bmatrix}_{d \times 1} \quad \text{and } v \in \mathbb{R}^d$$

$$\begin{aligned}
M' v &= \begin{bmatrix} m_1' \\ \vdots \\ m_d' \end{bmatrix}' v = \begin{bmatrix} m_1' v \\ \vdots \\ m_d' v \end{bmatrix} \\
&= \begin{bmatrix} v' & 0 & \dots & 0 \\ 0 & v' & \dots & 0 \\ \dots & & & \\ 0 & \dots & 0 & v' \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_d \end{bmatrix} \\
&= (I_d \otimes v') \text{ vec}(M)
\end{aligned}$$

where

$$A \otimes B = \begin{bmatrix} a_{11} B & \dots & a_{1n} B \\ \vdots & \ddots & \vdots \\ a_{m1} B & \dots & a_{mn} B \end{bmatrix}_{mp \times nq}$$