

# Deep Learning & Applied AI

Matrix meta-mechanics

Emanuele Rodolà  
[rodola@di.uniroma1.it](mailto:rodola@di.uniroma1.it)



# Matrix manipulation

The definition of matrix product gives rise to some **alternative viewpoints** that are often useful for practical manipulation of matrices.

These notes cover a few useful “meta-mechanics” of matrix products and other operations that are frequently encountered.

These results should be **checked!**

Do **not** trust everything blindly at the first exposure. After the check, you can go blindfolded.

These notes may be updated as we go on with the course.

# Transpose and inverse

A matrix  $\mathbf{A}$  is **symmetric** if:

$$\mathbf{A} = \mathbf{A}^{\top}$$

If the matrix is a product  $\mathbf{A} = \mathbf{BC}$ , the transpose applies as follows:

$$(\mathbf{BC})^{\top} = \mathbf{C}^{\top} \mathbf{B}^{\top}$$

The same holds for the inverse:

$$(\mathbf{BC})^{-1} = \mathbf{C}^{-1} \mathbf{B}^{-1}$$

A matrix  $\mathbf{A}$  is **orthogonal** if:

$$\mathbf{A}^{-1} = \mathbf{A}^{\top},$$

Thus,  $\mathbf{A}^{\top} \mathbf{A} = \mathbf{I}$  whenever  $\mathbf{A}$  is orthogonal.

# Products

Matrix-vector product:

$$\mathbf{X}\mathbf{y} = \begin{pmatrix} | & | & \cdots \\ \mathbf{x}_1 & \mathbf{x}_2 & \\ | & | & \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \end{pmatrix} = y_1 \begin{pmatrix} | \\ \mathbf{x}_1 \\ | \end{pmatrix} + y_2 \begin{pmatrix} | \\ \mathbf{x}_2 \\ | \end{pmatrix} + \cdots$$

# Products

Matrix-vector product:

$$\mathbf{X}\mathbf{y} = \begin{pmatrix} | & | & \cdots \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots \\ | & | & \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \end{pmatrix} = y_1 \begin{pmatrix} | \\ \mathbf{x}_1 \\ | \end{pmatrix} + y_2 \begin{pmatrix} | \\ \mathbf{x}_2 \\ | \end{pmatrix} + \cdots$$

Vector-matrix product; it's just a transposed version of the above:

$$\mathbf{z}^\top \mathbf{A} = (\mathbf{A}^\top \mathbf{z})^\top$$

# Products

Matrix-vector product:

$$\mathbf{X}\mathbf{y} = \begin{pmatrix} | & | & \cdots \\ \mathbf{x}_1 & \mathbf{x}_2 & \\ | & | & \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \end{pmatrix} = y_1 \begin{pmatrix} | \\ \mathbf{x}_1 \\ | \end{pmatrix} + y_2 \begin{pmatrix} | \\ \mathbf{x}_2 \\ | \end{pmatrix} + \cdots$$

Vector-matrix product; it's just a transposed version of the above:

$$\mathbf{z}^\top \mathbf{A} = (\mathbf{A}^\top \mathbf{z})^\top$$

Matrix-matrix product:

$$\mathbf{X}\mathbf{Y} = \begin{pmatrix} \text{---} & \mathbf{x}_1^\top & \text{---} \\ \text{---} & \mathbf{x}_2^\top & \text{---} \\ & \vdots & \end{pmatrix} \begin{pmatrix} | & | & \cdots \\ \mathbf{y}_1 & \mathbf{y}_2 & \\ | & | & \end{pmatrix} =$$

# Products

Matrix-vector product:

$$\mathbf{X}\mathbf{y} = \begin{pmatrix} | & | & \cdots \\ \mathbf{x}_1 & \mathbf{x}_2 & \\ | & | & \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \end{pmatrix} = y_1 \begin{pmatrix} | \\ \mathbf{x}_1 \\ | \end{pmatrix} + y_2 \begin{pmatrix} | \\ \mathbf{x}_2 \\ | \end{pmatrix} + \cdots$$

Vector-matrix product; it's just a transposed version of the above:

$$\mathbf{z}^\top \mathbf{A} = (\mathbf{A}^\top \mathbf{z})^\top$$

Matrix-matrix product:

$$\begin{aligned} \mathbf{X}\mathbf{Y} &= \begin{pmatrix} - & \mathbf{x}_1^\top & - \\ - & \mathbf{x}_2^\top & - \\ & \vdots & \end{pmatrix} \begin{pmatrix} | & | & \cdots \\ \mathbf{y}_1 & \mathbf{y}_2 & \\ | & | & \end{pmatrix} = \begin{pmatrix} | & | & \cdots \\ \mathbf{X}\mathbf{y}_1 & \mathbf{X}\mathbf{y}_2 & \\ | & | & \end{pmatrix} \\ &= \begin{pmatrix} - & \mathbf{x}_1^\top \mathbf{Y} & - \\ - & \mathbf{x}_2^\top \mathbf{Y} & - \\ & \vdots & \end{pmatrix} \end{aligned}$$

# Products

Matrix-matrix product:

$$\mathbf{XY} = \begin{pmatrix} \text{---} & \mathbf{x}_1^\top & \text{---} \\ \text{---} & \mathbf{x}_2^\top & \text{---} \\ & \vdots & \end{pmatrix} \begin{pmatrix} | & | & \cdots \\ \mathbf{y}_1 & \mathbf{y}_2 & \\ | & | & \end{pmatrix} = \begin{pmatrix} \cdots & & \\ \vdots & \mathbf{x}_i^\top \mathbf{y}_j & \vdots \\ \cdots & & \end{pmatrix}$$



# Products

Matrix-matrix product:

$$\mathbf{XY} = \begin{pmatrix} \text{---} & \mathbf{x}_1^\top & \text{---} \\ \text{---} & \mathbf{x}_2^\top & \text{---} \\ & \vdots & \end{pmatrix} \begin{pmatrix} | & | & \cdots \\ \mathbf{y}_1 & \mathbf{y}_2 & \\ | & | & \end{pmatrix} = \begin{pmatrix} \cdots & & \\ \vdots & \mathbf{x}_i^\top \mathbf{y}_j & \vdots \\ \cdots & & \end{pmatrix}$$

Vector-vector product ([inner](#)):

$$\mathbf{x}^\top \mathbf{y} = \alpha$$

# Products

Matrix-matrix product:

$$\mathbf{XY} = \begin{pmatrix} \text{---} & \mathbf{x}_1^\top & \text{---} \\ \text{---} & \mathbf{x}_2^\top & \text{---} \\ & \vdots & \end{pmatrix} \begin{pmatrix} | & | & \cdots \\ \mathbf{y}_1 & \mathbf{y}_2 & \\ | & | & \end{pmatrix} = \begin{pmatrix} \cdots & & \\ \vdots & \mathbf{x}_i^\top \mathbf{y}_j & \vdots \\ \cdots & & \end{pmatrix}$$

Vector-vector product ([inner](#)):

$$\mathbf{x}^\top \mathbf{y} = \alpha$$

Vector-vector product ([outer](#)):

$$\mathbf{xy}^\top = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} \begin{pmatrix} y_1 & y_2 & \cdots \end{pmatrix} = \begin{pmatrix} | & | & \cdots \\ y_1 \mathbf{x} & y_2 \mathbf{x} & \\ | & | & \end{pmatrix}$$

For example, a matrix full of ones is just  $\mathbf{11}^\top$ .

# Diagonal matrices

Matrix-vector product:

$$\begin{pmatrix} d_1 & & \\ & d_2 & \\ & & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} d_1 x_1 \\ d_2 x_2 \\ \vdots \end{pmatrix}$$

# Diagonal matrices

Matrix-vector product:

$$\begin{pmatrix} d_1 & & \\ & d_2 & \\ & & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} d_1 x_1 \\ d_2 x_2 \\ \vdots \end{pmatrix}$$

Matrix-matrix product:

$$\begin{pmatrix} d_1 & & \\ & d_2 & \\ & & \ddots \end{pmatrix} \begin{pmatrix} - & \mathbf{x}_1^\top & - \\ - & \mathbf{x}_2^\top & - \\ & \vdots & \end{pmatrix} = \begin{pmatrix} - & d_1 \mathbf{x}_1^\top & - \\ - & d_2 \mathbf{x}_2^\top & - \\ & \vdots & \end{pmatrix}$$

# Diagonal matrices

Matrix-vector product:

$$\begin{pmatrix} d_1 & & \\ & d_2 & \\ & & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} d_1 x_1 \\ d_2 x_2 \\ \vdots \end{pmatrix}$$

Matrix-matrix product:

$$\begin{pmatrix} d_1 & & \\ & d_2 & \\ & & \ddots \end{pmatrix} \begin{pmatrix} \text{---} & \mathbf{x}_1^\top & \text{---} \\ \text{---} & \mathbf{x}_2^\top & \text{---} \\ & \vdots & \end{pmatrix} = \begin{pmatrix} \text{---} & d_1 \mathbf{x}_1^\top & \text{---} \\ \text{---} & d_2 \mathbf{x}_2^\top & \text{---} \\ & \vdots & \end{pmatrix}$$

From the other side:

$$\begin{pmatrix} \begin{array}{c} | \\ \mathbf{y}_1 \\ | \end{array} & \begin{array}{c} | \\ \mathbf{y}_2 \\ | \end{array} & \cdots \end{pmatrix} \begin{pmatrix} d_1 & & \\ & d_2 & \\ & & \ddots \end{pmatrix} = \begin{pmatrix} d_1 \begin{array}{c} | \\ \mathbf{y}_1 \\ | \end{array} & d_2 \begin{array}{c} | \\ \mathbf{y}_2 \\ | \end{array} & \cdots \end{pmatrix}$$

# Trace

The trace of  $\mathbf{A}$  is the sum of its diagonal elements:

$$\text{tr}(\mathbf{A}) = \sum_i a_{ii}$$

It is a **linear** mapping, since:

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$$

$$\text{tr}(a\mathbf{A}) = a \text{tr}(\mathbf{A})$$

# Trace

The trace of  $\mathbf{A}$  is the sum of its diagonal elements:

$$\text{tr}(\mathbf{A}) = \sum_i a_{ii}$$

It is a **linear** mapping, since:

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$$

$$\text{tr}(a\mathbf{A}) = a \text{tr}(\mathbf{A})$$

It is invariant to the transpose:

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^\top)$$

# Trace

The trace of  $\mathbf{A}$  is the sum of its diagonal elements:

$$\text{tr}(\mathbf{A}) = \sum_i a_{ii}$$

It is a **linear** mapping, since:

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$$

$$\text{tr}(a\mathbf{A}) = a \text{tr}(\mathbf{A})$$

It is invariant to the transpose:

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^\top)$$

It is invariant to **cyclic permutations**:

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB}) = \text{tr}(\mathbf{BCA})$$



# Norms

The (squared) **Frobenius** norm for a matrix  $\mathbf{X}$  is:

$$\|\mathbf{X}\|_F^2 = \text{tr}(\mathbf{X}\mathbf{X}^\top) = \text{tr}(\mathbf{X}^\top\mathbf{X})$$

# Norms

The (squared) **Frobenius** norm for a matrix  $\mathbf{X}$  is:

$$\|\mathbf{X}\|_F^2 = \text{tr}(\mathbf{X}\mathbf{X}^\top) = \text{tr}(\mathbf{X}^\top\mathbf{X})$$

If  $\mathbf{X}$  is a vector, this reduces to:

$$\|\mathbf{x}\|_2^2 = \text{tr}(\underbrace{\mathbf{x}\mathbf{x}^\top}_{n \times n}) = \text{tr}(\underbrace{\mathbf{x}^\top\mathbf{x}}_{1 \times 1}) = \mathbf{x}^\top\mathbf{x}$$

# Norms

The (squared) **Frobenius** norm for a matrix  $\mathbf{X}$  is:

$$\|\mathbf{X}\|_F^2 = \text{tr}(\mathbf{X}\mathbf{X}^\top) = \text{tr}(\mathbf{X}^\top\mathbf{X})$$

If  $\mathbf{X}$  is a vector, this reduces to:

$$\|\mathbf{x}\|_2^2 = \text{tr}(\underbrace{\mathbf{x}\mathbf{x}^\top}_{n \times n}) = \text{tr}(\underbrace{\mathbf{x}^\top\mathbf{x}}_{1 \times 1}) = \mathbf{x}^\top\mathbf{x}$$

For example, for matrices  $\mathbf{A}$  and  $\mathbf{B}$  we can derive the distance:

$$\begin{aligned}\|\mathbf{A} - \mathbf{B}\|_F^2 &= \text{tr}((\mathbf{A} - \mathbf{B})^\top(\mathbf{A} - \mathbf{B})) \\ &= \text{tr}(\mathbf{A}^\top\mathbf{A}) - 2\text{tr}(\mathbf{A}^\top\mathbf{B}) + \text{tr}(\mathbf{B}^\top\mathbf{B}),\end{aligned}$$

where we used the linearity of the trace and its invariance to transposition.

# Ones

A vector **1** of **ones** can be used to calculate sums easily.

Sum up the elements of **A** along each row:

$$\mathbf{A}\mathbf{1}$$

Sum up along each column:

$$\mathbf{1}^\top \mathbf{A} = (\mathbf{A}^\top \mathbf{1})^\top$$

# Ones

A vector **1** of ones can be used to calculate sums easily.

Sum up the elements of **A** along each row:

$$\mathbf{A}\mathbf{1}$$

Sum up along each column:

$$\mathbf{1}^\top \mathbf{A} = (\mathbf{A}^\top \mathbf{1})^\top$$

Sum up all the elements of **A**:

$$\mathbf{1}^\top \mathbf{A} \mathbf{1}$$

# Ones

A vector **1** of ones can be used to calculate sums easily.

Sum up the elements of **A** along each row:

$$\mathbf{A}\mathbf{1}$$

Sum up along each column:

$$\mathbf{1}^\top \mathbf{A} = (\mathbf{A}^\top \mathbf{1})^\top$$

Sum up all the elements of **A**:

$$\mathbf{1}^\top \mathbf{A} \mathbf{1}$$

Note the following relationship:

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = \text{tr}(\mathbf{x}^\top \mathbf{A} \mathbf{x}) = \text{tr}(\mathbf{x} \mathbf{x}^\top \mathbf{A}) = \text{tr}(\mathbf{X} \mathbf{A})$$

# Ones

A vector **1** of **ones** can be used to calculate sums easily.

Sum up the elements of **A** along each row:

$$\mathbf{A}\mathbf{1}$$

Sum up along each column:

$$\mathbf{1}^\top \mathbf{A} = (\mathbf{A}^\top \mathbf{1})^\top$$

Sum up all the elements of **A**:

$$\mathbf{1}^\top \mathbf{A} \mathbf{1}$$

Note the following relationship:

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = \text{tr}(\mathbf{x}^\top \mathbf{A} \mathbf{x}) = \text{tr}(\mathbf{x} \mathbf{x}^\top \mathbf{A}) = \text{tr}(\mathbf{X} \mathbf{A})$$

Constructing the matrix  $\mathbf{x} \mathbf{x}^\top$  from the vector  $\mathbf{x}$  is also called **lifting**.

## Permutation matrices

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$



# Permutation matrices

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

Permutation matrices are **orthogonal**.

Their product is still a permutation matrix.

# Permutation matrices

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

Permutation matrices are **orthogonal**.

Their product is still a permutation matrix.

Their convex combination is a **doubly stochastic** matrix:

$$\alpha \mathbf{P} + (1 - \alpha) \mathbf{Q} = \mathbf{D} \quad \text{with } \alpha \in [0, 1]$$

that is, we get  $\mathbf{D}\mathbf{1} = \mathbf{1}$  and  $\mathbf{D}^\top \mathbf{1} = \mathbf{1}$ .

# Gradients of traces

The following expression appears frequently in practice:

$$\nabla \text{tr}(\mathbf{A}) = \nabla \sum_i a_{ii}$$

which requires the computation of the partial derivatives:

$$\frac{\partial}{\partial a_{ij}} \sum_i a_{ii}$$

A common pitfall is the following **invalid** operation:

$$\underbrace{\nabla \text{tr}(\mathbf{A})}_{\text{gradient of a scalar function}} \Rightarrow \text{tr}(\underbrace{\nabla \mathbf{A}}_{\text{undefined}})$$

Also observe that  $\nabla \text{tr}(\mathbf{A})$  is a matrix, while  $\text{tr}(\cdots)$  is a scalar.

# Suggested reading

For a review of matrix calculus, read Chapters 0.0 – 0.2 of the book:

R. Horn & C. Johnson, "Matrix Analysis - 2nd ed".  
Cambridge University Press, 2013