

# Notes on Linear Algebra

Paul F. Roysdon, Ph.D.

## I. MATRIX & VECTOR PROPERTIES

### A. Vector Properties

First define a vector space  $\mathcal{V}$ , where  $\mathcal{V}_0$  is a vector subspace of  $\mathcal{V}$ , if and only if (iff):

- $\mathcal{V}_0$  is non-empty
- $\mathcal{V}_0$  is closed under multiplication
- $\mathcal{V}_0$  is closed under addition

A basis of subspace  $\mathcal{V}_0$  is a linearly-independent spanning set of  $\mathcal{V}$ .

The following properties of a vector space hold  $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$  and  $a, b \in \mathbb{R}$ :

$$\begin{aligned}\mathbf{u} + (\mathbf{v} + \mathbf{w}) &= (\mathbf{u} + \mathbf{v}) + \mathbf{w} \\ \mathbf{u} + \mathbf{v} &= \mathbf{v} + \mathbf{u} \\ \exists \mathbf{0} \in \mathcal{V} : \mathbf{v} + \mathbf{0} &= \mathbf{v}, \forall \mathbf{v} \in \mathcal{V} \\ \forall \mathbf{v} \in \mathcal{V}, \exists (-\mathbf{v}) \in \mathcal{V} : \mathbf{v} + (-\mathbf{v}) &= \mathbf{0} \\ (ab)\mathbf{v} &= a(b\mathbf{v}) \\ 1\mathbf{v} &= \mathbf{v} \\ a(\mathbf{u} + \mathbf{v}) &= a\mathbf{u} + a\mathbf{v} \\ (a + b)\mathbf{v} &= a\mathbf{v} + b\mathbf{v}\end{aligned}$$

### B. Products of Vectors

The *scalar product* (also called the *inner product* or *dot product*) of two vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{m \times 1}$  is defined as

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \cdot \mathbf{v} = \sum_{i=1}^m u_i v_i$$

The Euclidean norm of a vector is the scalar product of a vector with itself

$$\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$$

the multiplicative property is

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} = \mathbf{u}^T \cdot \mathbf{v} = \mathbf{v}^T \cdot \mathbf{u} \neq \mathbf{u} \cdot \mathbf{v}^T$$

the associative property is

$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$$

The vector  $\mathbf{u}$  is a *unit vector* if  $\|\mathbf{u}\| = 1$ . A vector has both magnitude and direction, with magnitude defined by  $\|\mathbf{u}\|$  and direction defined by  $\frac{\mathbf{u}}{\|\mathbf{u}\|}$ .

The angle  $\alpha$  between two vectors  $\mathbf{u}$  and  $\mathbf{v}$  is defined as

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\alpha)$$

Two vectors are orthogonal if and only if  $\mathbf{u} \cdot \mathbf{v} = 0$ .

A function  $\mathcal{V} \mapsto \mathbb{R}$  is a norm iff,  $\forall \mathbf{v} \in \mathcal{V}$ , the all three properties must be satisfied

1.  $\|\mathbf{v}\| \geq 0$  and,  $\|\mathbf{v}\| = 0 \iff \mathbf{v} = \mathbf{0}$
2.  $\|c\mathbf{v}\| = |c| \cdot \|\mathbf{v}\|$ ,  $c \in \mathbb{R}$
3.  $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$

Some useful vector norms are

$$\begin{aligned}\|\mathbf{u}\|_p &= \left( \sum_{i=1}^m |u_i|^p \right)^{1/p}, \quad p \geq 1 \\ \|\mathbf{u}\|_1 &= \sum_{i=1}^m |u_i| \\ \|\mathbf{u}\|_2 &= \sqrt{\mathbf{u}^T \mathbf{u}} \\ \|\mathbf{u}\|_\infty &= \max_i \{|u_i|\}\end{aligned}$$

The Cauchy-Schwartz Inequality for  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{m \times 1}$ :

$$|\mathbf{u}^T \mathbf{v}| \leq \|\mathbf{u}\|_2 \cdot \|\mathbf{v}\|_2$$

The *vector product* (also called the *outer-product* or *cross-product*) of two vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{3 \times 1}$  is defined as

$$\begin{aligned}\mathbf{u} \times \mathbf{v} &= \begin{vmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \\ &= \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\end{aligned}\tag{1}$$

where  $|\cdot|$  denotes a determinant, and  $i, j$ , and  $k$  are unit vectors pointing along the principal axes of the reference frame. Notice  $\mathbf{u} \times \mathbf{u} = \mathbf{0}$  and  $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$ .

It is sometimes convenient to express the vector product of eqn. (1) in the matrix form

$$\mathbf{u} \times \mathbf{v} = \mathbf{U} \mathbf{v}$$

where

$$\mathbf{U} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}$$

This skew symmetric form of  $\mathbf{u}$  is convenient in analysis. The matrix  $\mathbf{U}$  will be denoted  $\mathbf{U} = [\mathbf{u} \times]$ . The vector product has

the following properties:

$$\begin{aligned}
\mathbf{u} \times \mathbf{u} &= \mathbf{0} \\
\mathbf{u} \times \mathbf{v} &= -\mathbf{v} \times \mathbf{u} \\
\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) &\neq (\mathbf{u} \times \mathbf{v}) \times \mathbf{w} \\
\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) &= (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w} \\
\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) &= \mathbf{v} \cdot (\mathbf{w} \times \mathbf{u}) \\
&= \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) \\
(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{w} \times \mathbf{z}) &= (\mathbf{u} \cdot \mathbf{w})(\mathbf{v} \cdot \mathbf{z}) - (\mathbf{v} \cdot \mathbf{w})(\mathbf{u} \cdot \mathbf{z}).
\end{aligned}$$

### C. Products of Matrices

Matrix multiplication has several useful formulas, where  $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{R}^{m \times m}$ . The associative property is

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}).$$

The distributive property is

$$\begin{aligned}
(\mathbf{A} + \mathbf{B})\mathbf{C} &= \mathbf{AC} + \mathbf{BC} \\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{ABC} \dots)^{-1} &= \dots \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{A}^\top)^{-1} &= (\mathbf{A}^{-1})^\top \\
(\mathbf{A} + \mathbf{B})^\top &= \mathbf{A}^\top + \mathbf{B}^\top \\
(\mathbf{AB})^\top &= \mathbf{B}^\top \mathbf{A}^\top \\
(\mathbf{ABC} \dots)^\top &= \dots \mathbf{C}^\top \mathbf{B}^\top \mathbf{A}^\top \\
(\mathbf{A}^H)^{-1} &= (\mathbf{A}^{-1})^H \\
(\mathbf{A} + \mathbf{B})^H &= \mathbf{A}^H + \mathbf{B}^H \\
(\mathbf{AB})^H &= \mathbf{B}^H \mathbf{A}^H \\
(\mathbf{ABC} \dots)^H &= \dots \mathbf{C}^H \mathbf{B}^H \mathbf{A}^H.
\end{aligned}$$

Matrix multiplication does not commute:  $\mathbf{AB} \neq \mathbf{BA}$ . Matrix division is undefined. When  $\mathbf{A} \in \mathbb{R}^{n \times m}$  and  $\mathbf{B} \in \mathbb{R}^{m \times p}$ , then their product  $\mathbf{C} = \mathbf{AB} \in \mathbb{R}^{n \times p}$  is defined as  $c_{i,j} = \sum_{k=1}^m a_{i,k} b_{k,j}$ . Note that the number of columns in  $\mathbf{A}$  and the number of rows in  $\mathbf{B}$ , referred to as the inner dimension of the product, must be identical; otherwise the two matrices cannot be multiplied.

A square matrix  $\mathbf{A}$  is an orthogonal matrix *iff*  $\mathbf{AA}^\top = \mathbf{A}^\top \mathbf{A}$  is a diagonal matrix. A square matrix  $\mathbf{A}$  is an orthonormal matrix if and only if  $\mathbf{AA}^\top = \mathbf{A}^\top \mathbf{A} = \mathbf{I}$ . Let  $\mathbf{R}_a^b$  represent the vector transformation from reference frame  $a$  to reference frame  $b$ . The inverse transformation from reference frame  $b$  to reference frame  $a$  can be shown to be  $\mathbf{R}_b^a = (\mathbf{R}_a^b)^{-1} = (\mathbf{R}_a^b)^\top$ . Then  $(\mathbf{R}_a^b)^\top \mathbf{R}_a^b = \mathbf{R}_a^b (\mathbf{R}_a^b)^\top = \mathbf{I}$ . This shows that  $\mathbf{R}_a^b$  and  $\mathbf{R}_b^a$  are an orthonormal matrix.

### D. Matrix Norms

A matrix norm is a mapping which fulfills 1-3

1.  $\|\mathbf{A}\| \geq 0$  and,  $\|\mathbf{A}\| = 0 \iff \mathbf{A} = \mathbf{0}$
2.  $\|c\mathbf{A}\| = |c| \cdot \|\mathbf{A}\|$ ,  $c \in \mathbb{R}$
3.  $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$

An induced norm is a matrix norm induced by a vector norm

$$\|\mathbf{A}\| = \sup\{\|\mathbf{Ax}\| \mid \|\mathbf{x}\| = 1\}$$

where  $\|\cdot\|$  on the left side is the induced matrix norm, while  $\|\cdot\|$  on the right side denotes the vector norm. For induced norms it holds that

1.  $\|\mathbf{I}\| = 1$
2.  $\|\mathbf{Ax}\|_p \leq \|\mathbf{A}\|_p \cdot \|\mathbf{x}\|_p$ ,  $\forall \mathbf{A}, \mathbf{x}$
3.  $\|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \cdot \|\mathbf{B}\|_p$ ,  $\forall \mathbf{A}, \mathbf{B}$

where eq. (2) is known as the Submultiplicative Property of Induced Norms. Induced norms have the following properties:

$$\begin{aligned}
\|\mathbf{A}\|_p &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p} \\
&= \max_{\mathbf{x} \neq \mathbf{0}} \left\| \mathbf{A} \frac{\mathbf{x}}{\|\mathbf{x}\|_p} \right\|_p \\
&= \max_{\|\mathbf{y}\|_p = 1} \|\mathbf{Ay}\|_p
\end{aligned}$$

Some useful matrix norms are

$$\begin{aligned}
\|\mathbf{A}\|_p &= \left( \sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^p \right)^{1/p} \\
\|\mathbf{A}\|_1 &= \max_{1 \leq j \leq n} \sum_{i=1}^m |A_{ij}| \\
\|\mathbf{A}\|_2 &= \sqrt{\lambda_{\max}(\mathbf{A}^\top \mathbf{A})} \\
\|\mathbf{A}\|_\infty &= \max_{1 \leq i \leq m} \sum_{j=1}^n |A_{ij}| \\
\|\mathbf{A}\|_F &= \sqrt{\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2} = \sqrt{\text{Tr}(\mathbf{AA}^H)} \\
\|\mathbf{A}\|_{\max} &= \max_{ij} |A_{ij}|
\end{aligned}$$

where  $\lambda_{\max}$  is the maximum eigenvalue of  $(\mathbf{A}^\top \mathbf{A})$ .

### E. Condition Number

The condition number is a measure of how sensitive a function is to errors in the input. A problem with a low condition number is said to be well-conditioned, while a problem with a high condition number is said to be ill-conditioned.

The condition number,  $\kappa(\mathbf{A})$ , for the general matrix  $\mathbf{A}$ , is

$$\kappa(\mathbf{A}) = \frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})}$$

where  $\sigma_{\max}(\mathbf{A})$  and  $\sigma_{\min}(\mathbf{A})$  are the maximum and minimum singular values of  $\mathbf{A}$ .

If  $\mathbf{A}$  is a symmetric, positive definite matrix,

$$\kappa(\mathbf{A}) = \frac{|\lambda_{\max}(\mathbf{A})|}{|\lambda_{\min}(\mathbf{A})|}$$

where  $\lambda_{\max}(\mathbf{A})$  and  $\lambda_{\min}(\mathbf{A})$  are the maximum and minimum eigenvalues of  $\mathbf{A}$ .

If  $\mathbf{A}$  is unitary, then

$$\kappa(\mathbf{A}) = 1$$

If the condition number is  $\epsilon$  (some small value) larger than one, the matrix is well conditioned, and therefore its inverse can be computed with good accuracy.

#### F. Trace Operator

The trace of a square matrix  $\mathbf{A} \in \mathbb{R}^{m \times m}$  is defined as

$$\text{Tr}(\mathbf{A}) = \sum_{i=1}^m a_{ii}$$

Notice:

$$\begin{aligned} \text{Tr}(\mathbf{A}) &= \text{Tr}(\mathbf{A}^\top) \\ \text{Tr}(\mathbf{A} + \mathbf{B}) &= \text{Tr}(\mathbf{A}) + \text{Tr}(\mathbf{B}) \\ \text{Tr}(\mathbf{AB}) &= \text{Tr}(\mathbf{BA}) \\ \text{Tr}(\mathbf{ABC}) &= \text{Tr}(\mathbf{BCA}) = \text{Tr}(\mathbf{CAB}) \\ \mathbf{v}^\top \mathbf{v} &= \text{Tr}(\mathbf{vv}^\top) \\ \text{Tr}(\mathbf{vv}^\top) &= \text{Tr}(\mathbf{v}^\top \mathbf{v}) = \|\mathbf{v}\|^2 \\ \text{Tr}(\mathbf{A}) &= \sum_{i=1}^m \lambda_i, \quad \lambda_i = \text{eig}(\mathbf{A}) \end{aligned}$$

where  $\mathbf{v} \in \mathbb{R}^{m \times 1}$ , and the eigenvalues  $\boldsymbol{\lambda} \in \mathbb{R}^{m \times 1}$ .

#### G. Definiteness of Matrices

For the symmetric matrix  $\mathbf{A} \in \mathbb{R}^{m \times m}$  and vector  $\mathbf{x} \in \mathbb{R}^{m \times 1}$ , the scalar mapping  $\mathbf{x}^\top \mathbf{A} \mathbf{x}$  is called a *quadratic form*, and has the following properties:

- 1)  $\mathbf{A}$  is positive definite *iff*  $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$ ,  $\forall \mathbf{x} \neq 0$ .
- 2)  $\mathbf{A}$  is positive semi-definite *iff*  $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0$ ,  $\forall \mathbf{x} \neq 0$ .
- 3)  $\mathbf{A}$  is negative definite *iff*  $\mathbf{x}^\top \mathbf{A} \mathbf{x} < 0$ ,  $\forall \mathbf{x} \neq 0$ .
- 4)  $\mathbf{A}$  is negative semi-definite *iff*  $\mathbf{x}^\top \mathbf{A} \mathbf{x} \leq 0$ ,  $\forall \mathbf{x} \neq 0$ .

where *iff* is defined as “if and only if”. The designer is often interested in the sign of the scalar output,  $\mathbf{x}^\top \mathbf{A} \mathbf{x}$ . When the matrix  $\mathbf{A}$  fails to have any of the above properties, then the matrix  $\mathbf{A}$  is sign indefinite.

#### H. Rank of Matrices

The rank of a matrix follows the following rules:

For  $\mathbf{A} \in \mathbb{R}^{p \times q}$

$$\begin{aligned} \text{rank}(\mathbf{A}) &= \text{rank}(\mathbf{A}^\top) \\ \text{rank}(\mathbf{A}^\top \mathbf{A}) &= \text{rank}(\mathbf{A} \mathbf{A}^\top) = \text{rank}(\mathbf{A}) \\ \text{rank}(\mathbf{AB}) &\leq \text{rank}(\mathbf{A}) \\ \text{rank}(\mathbf{AB}) &\leq \text{rank}(\mathbf{B}) \\ \text{rank}(\mathbf{A} + \mathbf{B}) &\leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \\ \text{rank}(\mathbf{A}) &= \min\{p, q\} \end{aligned}$$

Sylvester's inequality: for  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{B} \in \mathbb{R}^{n \times r}$

$$\begin{aligned} \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n &\leq \text{rank}(\mathbf{AB}) \\ &\leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\} \end{aligned}$$

and

$$\begin{aligned} |\text{rank}(\mathbf{A}) - \text{rank}(\mathbf{B})| &\leq \text{rank}(\mathbf{A} + \mathbf{B}) \\ &\leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \end{aligned}$$

For full-rank  $\mathbf{A}$

$$\begin{aligned} \mathbf{A} &\in \mathbb{R}^{n \times n}, \quad \text{rank}(\mathbf{A}) = n \\ \mathbf{A} &\in \mathbb{R}^{n \times m}, \quad \text{rank}(\mathbf{A}) = m, \quad m > n \\ \mathbf{A} &\in \mathbb{R}^{m \times n}, \quad \text{rank}(\mathbf{A}) = n, \quad m > n \end{aligned}$$

Elementary transformations do not change rank, e.g. swap rows or columns, addition operations on rows, multiplication operations by non-zero values for rows or columns.

#### I. Independence and Determinants

Given vectors  $\mathbf{u}_i \in \mathbb{R}^n$  for  $i = 1, \dots, m$ , the vectors are linearly dependent if there exists  $\alpha_i \in \mathbb{R}$ , not all zero, such that  $\sum_{i=1}^m \alpha_i \mathbf{u}_i = 0$ . If the summation is only zero when all  $\alpha_i = 0$ , then the set of vectors is linearly independent. A set of  $m > n$  vectors in  $\mathbb{R}^n$  is always linearly dependent.

A set of  $n$  vectors  $\mathbf{u}_i \in \mathbb{R}^n$ ,  $i = 1, \dots, n$  can be arranged as a square matrix  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ . Note that a matrix can always be decomposed into its component column (or row) vectors. The rank of a matrix, denoted  $\text{rank}(\mathbf{U})$ , is the number of independent column (or row) vectors in the matrix. If this set of vectors is linearly independent (i.e.,  $\text{rank}(\mathbf{U}) = n$ ), then we say that the matrix  $\mathbf{U}$  is nonsingular (or of full rank).

A convenient tool for checking whether a square matrix is nonsingular is the determinant. The determinant of  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , denoted  $|\mathbf{A}|$  or  $\det(\mathbf{A})$ , is a scalar real number that can be computed either as

$$|\mathbf{A}| = \sum_{j=1}^n a_{kj} c_{kj}$$

which uses row expansion; or,

$$|\mathbf{A}| = \sum_{k=1}^n a_{kj} c_{kj}$$

which uses column expansion. For a square matrix  $\mathbf{A}$ , the *cofactor* associated with  $a_{kj}$  is

$$c_{kj} = (-1)^{k+j} M_{kj}$$

where  $M_{kj}$  is the *minor* associated with  $a_{kj}$ .  $M_{kj}$  is defined as the determinant of the matrix formed by dropping the  $k$ -th row and  $j$ -th column from  $\mathbf{A}$ . The determinant of  $\mathbf{A} \in \mathbb{R}^{1 \times 1}$  is (the scalar)  $\mathbf{A}$ . The determinant of  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is written in terms of the summation of the determinants of matrices in  $\mathbb{R}^{(n-1) \times (n-1)}$ . This dimension reduction process continues until it involves only scalars, for which the computation is straightforward.

If  $|\mathbf{A}| \neq 0$ , then  $\mathbf{A}$  is nonsingular and the vectors forming the rows (and columns) of  $\mathbf{A}$  are linearly independent. If  $|\mathbf{A}| = 0$ , then  $\mathbf{A}$  is singular and the vectors forming the rows (and columns) of  $\mathbf{A}$  are linearly dependent.

Determinants have the following useful properties: For  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$

- 1)  $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$
- 2)  $|\mathbf{A}| = |\mathbf{A}^\top|$
- 3) If any row or column of  $\mathbf{A}$  is entirely zero, then  $|\mathbf{A}| = 0$ .

- 4) If any two rows (or columns) of  $\mathbf{A}$  are linearly dependent, then  $|\mathbf{A}| = 0$ .
- 5) Interchanging two rows (or two columns) of  $\mathbf{A}$  reverses the sign of the determinant.
- 6) Multiplication of a row of  $\mathbf{A}$  by  $\alpha \in \mathbb{R}$ , yields  $\alpha|\mathbf{A}|$ .
- 7) A scaled version of one row can be added to another row without changing the determinant.

A summary of the properties are

$$\begin{aligned}
\det(\mathbf{A}) &= \prod_i \lambda_i, \quad \lambda_i = \text{eig}(\mathbf{A}) \\
\det(c\mathbf{A}) &= c^n \det(\mathbf{A}), \quad \text{if } \mathbf{A} \in \mathbb{R}^{n \times n} \\
\det(\mathbf{A}^\top) &= \det(\mathbf{A}) \\
\det(\mathbf{AB}) &= \det(\mathbf{A}) \det(\mathbf{B}) \\
\det(\mathbf{A}^{-1}) &= 1/\det(\mathbf{A}) \\
\det(\mathbf{A}^n) &= \det(\mathbf{A})^n \\
\det(\mathbf{I} + \mathbf{uv}^\top) &= 1 + \mathbf{u}^\top \mathbf{v}, \quad \text{for } \mathbf{u}, \mathbf{v} \in \mathbb{R}^{n \times 1}
\end{aligned}$$

For  $n = 2$ :

$$\det(\mathbf{I} + \mathbf{A}) = 1 + \det(\mathbf{A}) + \text{Tr}(\mathbf{A})$$

For  $n = 3$ :

$$\det(\mathbf{I} + \mathbf{A}) = 1 + \det(\mathbf{A}) + \text{Tr}(\mathbf{A}) + \frac{1}{2} \text{Tr}(\mathbf{A})^2 - \frac{1}{2} \text{Tr}(\mathbf{A}^2)$$

#### J. Matrix Inversion

For  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , with  $|\mathbf{A}| \neq 0$ , we denote the inverse of  $\mathbf{A}$  by  $\mathbf{A}^{-1}$  which has the property that

$$\mathbf{A}^{-1} \mathbf{A} = \mathbf{A} \mathbf{A}^{-1} = \mathbf{I}.$$

When  $\mathbf{A}$  is a square nonsingular matrix,

$$\mathbf{A}^{-1} = \frac{\mathbf{C}^\top}{|\mathbf{A}|}$$

where  $\mathbf{C}$  is the cofactor matrix for  $\mathbf{A}$  and  $\mathbf{C}^\top$  is called the adjoint of  $\mathbf{A}$ . The matrix inverse has the following properties:

$$\begin{aligned}
(\mathbf{A}^{-1})^{-1} &= \mathbf{A} \\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1} \mathbf{A}^{-1} \\
|\mathbf{A}^{-1}| &= \frac{1}{|\mathbf{A}|} \\
(\mathbf{A}^\top)^{-1} &= (\mathbf{A}^{-1})^\top = \mathbf{A}^{-\top} \\
(\mu \mathbf{A})^{-1} &= \frac{1}{\mu} \mathbf{A}^{-1}.
\end{aligned}$$

The inverse of an orthonormal matrix is the same as its transpose. A more complete treatment of matrix inversion is provided in Section V.

#### K. Matrix Inversion Lemma

Two forms of the Matrix Inversion Lemma are presented. The Lemma is useful in least squares and Kalman filter derivations. Each lemma can be proved by direct multiplication.

*Lemma 1.1:* Given four matrices  $\mathbf{P}_1$ ,  $\mathbf{P}_2$ ,  $\mathbf{H}$ , and  $\mathbf{R}$  of compatible dimensions, if  $\mathbf{P}_1$ ,  $\mathbf{P}_2$ ,  $\mathbf{R}$ , and  $(\mathbf{H}^\top \mathbf{P}_1 \mathbf{H} + \mathbf{R})$  are all invertible and

$$\mathbf{P}_2^{-1} = \mathbf{P}_1^{-1} + \mathbf{H} \mathbf{R}^{-1} \mathbf{H}^\top$$

then

$$\mathbf{P}_2 = \mathbf{P}_1 - \mathbf{P}_1 \mathbf{H} (\mathbf{H}^\top \mathbf{P}_1 \mathbf{H} + \mathbf{R})^{-1} \mathbf{H}^\top \mathbf{P}_1.$$

*Lemma 1.2:* Given four matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , and  $\mathbf{D}$  of compatible dimensions, if  $\mathbf{A}$ ,  $\mathbf{C}$ , and  $\mathbf{A} + \mathbf{BCD}$  are invertible, then

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{D} \mathbf{A}^{-1} \mathbf{B} + \mathbf{C}^{-1}) \mathbf{D} \mathbf{A}^{-1}$$

The equivalence of the two forms is shown by defining:  $\mathbf{A} = \mathbf{P}_1^{-1}$ ,  $\mathbf{B} = \mathbf{H}$ ,  $\mathbf{C} = \mathbf{R}^{-1}$ ,  $\mathbf{D} = \mathbf{H}^\top$ , and requiring  $\mathbf{A} + \mathbf{BCD} = \mathbf{P}_2^{-1}$ .

#### L. Eigenvalues and Eigenvectors

For  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , the set of scalars  $\lambda_i \in \mathbb{C}$  and (nonzero) vectors  $\mathbf{x}_i \in \mathbb{C}^n$  satisfying  $\mathbf{A} \mathbf{x}_i = \lambda_i \mathbf{x}_i$  or  $(\lambda_i \mathbf{I} - \mathbf{A}) \mathbf{x}_i = \mathbf{0}_n$  are the eigenvalues and eigenvectors of  $\mathbf{A}$ .

We are only interested in nontrivial solutions (i.e., solution  $\mathbf{x}_i = \mathbf{0}$  is not of interest). Nontrivial solutions exist only if  $(\lambda \mathbf{I} - \mathbf{A})$  is a singular matrix. Therefore, the eigenvalues of  $\mathbf{A}$  are the values of  $\lambda$  such that  $|\lambda \mathbf{I} - \mathbf{A}| = 0$ . This yields an  $n$ -th order polynomial in  $\lambda$ .

If  $\mathbf{A}$  is a symmetric matrix, then all of its eigenvalues and eigenvectors are real. If  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are eigenvectors of symmetric matrix  $\mathbf{A}$  and their eigenvalues are not equal (i.e.,  $\lambda_i \neq \lambda_j$ ), then the eigenvectors are orthogonal (i.e.  $\mathbf{x}_i \cdot \mathbf{x}_j = 0$ ).

A square matrix  $\mathbf{A}$  is *idempotent* if and only if  $\mathbf{A} \mathbf{A} = \mathbf{A}$ . Idempotent matrices are sometimes also called *projection* matrices. Idempotent matrices have the following properties:

- 1)  $\text{rank}(\mathbf{A}) = \text{Tr}(\mathbf{A})$ ;
- 2) the eigenvalues of  $\mathbf{A}$  are all either 0 or 1;
- 3) the multiplicity of 1 as an eigenvalue is the  $\text{rank}(\mathbf{A})$ ;
- 4)  $\mathbf{A}(\mathbf{I} - \mathbf{A}) = (\mathbf{I} - \mathbf{A})\mathbf{A} = \mathbf{0}$ ; and,
- 5)  $\mathbf{A}^\top$ ,  $(\mathbf{I} - \mathbf{A})$  and  $(\mathbf{I} - \mathbf{A}^\top)$  are idempotent.

#### M. The Special Case 2x2

Consider the matrix  $\mathbf{A} \in \mathbb{R}^{2 \times 2}$

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

the determinant and trace are

$$\begin{aligned}
\det(\mathbf{A}) &= A_{11} A_{22} - A_{12} A_{21} \\
\text{Tr}(\mathbf{A}) &= A_{11} + A_{22},
\end{aligned}$$

the eigenvalues are

$$\begin{aligned}
\lambda^2 - \lambda \cdot \text{Tr}(\mathbf{A}) + \det(\mathbf{A}) &= 0 \\
\lambda_1 &= \frac{\text{Tr}(\mathbf{A}) + \sqrt{\text{Tr}(\mathbf{A})^2 - 4\det(\mathbf{A})}}{2} \\
\lambda_2 &= \frac{\text{Tr}(\mathbf{A}) - \sqrt{\text{Tr}(\mathbf{A})^2 - 4\det(\mathbf{A})}}{2} \\
\lambda_1 + \lambda_2 &= \text{Tr}(\mathbf{A}) \\
\lambda_1 \lambda_2 &= \det(\mathbf{A}),
\end{aligned}$$

the eigenvectors are

$$\mathbf{v}_1 \propto \begin{bmatrix} A_{12} \\ \lambda_1 A_{11} \end{bmatrix}$$

$$\mathbf{v}_2 \propto \begin{bmatrix} A_{12} \\ \lambda_2 A_{11} \end{bmatrix},$$

and the inverse is

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}.$$

## II. MATRIX EXPONENTIAL

The Matrix Exponential may be computed by many methods; provided here are two such examples. The first two sections provide a background on finite series and Taylor series expansion.

### A. Finite Series

We define a finite series as

$$(\mathbf{X}^n - \mathbf{I})(\mathbf{X} - \mathbf{I})^{-1} = \mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \dots + \mathbf{X}^{n-1}.$$

### B. Taylor Series

Consider some scalar function  $f(\mathbf{x})$  which takes the vector  $\mathbf{x}$  as an argument. Taylor series expansion around  $\mathbf{x}_0$  is defined as

$$f(\mathbf{x}) \cong f(\mathbf{x}_0) + \mathbf{g}(\mathbf{x}_0)^\top (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^\top \mathbf{H}(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0),$$

where  $\mathbf{g}(\mathbf{x}_0) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}_0}$  and  $\mathbf{H}(\mathbf{x}_0) = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top}|_{\mathbf{x}_0}$ .

### C. Power Series

The power series expansion of the scalar exponential function is

$$e^{at} = \sum_{n=0}^{\infty} \frac{1}{n!} (at)^n = 1 + at + \frac{(at)^2}{2!} + \frac{(at)^3}{3!} + \dots$$

for  $a, t \in \mathbb{R}$ . Extension of this power series to matrix arguments serves as a definition of the matrix exponential

$$e^{\mathbf{A}} \equiv \sum_{n=0}^{\infty} \frac{1}{n!} \mathbf{A}^n = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \dots$$

$$e^{-\mathbf{A}} \equiv \sum_{n=0}^{\infty} \frac{1}{n!} (-1)^n \mathbf{A}^n = \mathbf{I} - \mathbf{A} + \frac{\mathbf{A}^2}{2!} - \frac{\mathbf{A}^3}{3!} + \dots$$

$$e^{\mathbf{A}t} \equiv \sum_{n=0}^{\infty} \frac{1}{n!} (\mathbf{A}t)^n = \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2!} + \frac{(\mathbf{A}t)^3}{3!} + \dots \quad (3)$$

$$\ln(\mathbf{I} + \mathbf{A}) \equiv \sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{n} \mathbf{A}^n = \mathbf{A} - \frac{\mathbf{A}^2}{2} + \frac{\mathbf{A}^3}{3} + \dots$$

where  $t \in \mathbb{R}$ ,  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , and  $\mathbf{I}$  is the identity matrix in  $\mathbb{R}^{n \times n}$ . Note that by the definition of the matrix exponential, it is always true that a matrix commutes with its exponential:

$$\mathbf{A}e^{\mathbf{A}t} = e^{\mathbf{A}t}\mathbf{A}.$$

Also,  $e^{\mathbf{A}t}|_{t=0} = \mathbf{I}$ .

For  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ , some useful properties of the exponential are

$$e^{\mathbf{A}}e^{\mathbf{B}} = e^{\mathbf{A}+\mathbf{B}}, \quad \text{if } \mathbf{AB} = \mathbf{BA}$$

$$(e^{\mathbf{A}})^{-1} = e^{-\mathbf{A}}$$

$$\frac{d}{dt}e^{t\mathbf{A}} = \mathbf{A}e^{t\mathbf{A}} = e^{t\mathbf{A}}\mathbf{A}, \quad t \in \mathbb{R}$$

$$\frac{d}{dt}\text{Tr}(e^{t\mathbf{A}}) = \text{Tr}(\mathbf{A}e^{t\mathbf{A}})$$

$$\det(e^{\mathbf{A}}) = e^{\text{Tr}(\mathbf{A})}$$

Some useful trigonometric functions are

$$\sin(\mathbf{A}) \equiv \sum_{n=0}^{\infty} \frac{(-1)^n \mathbf{A}^{2n+1}}{(2n+1)!} = \mathbf{A} - \frac{\mathbf{A}^3}{3!} + \frac{\mathbf{A}^5}{5!} - \dots$$

$$\cos(\mathbf{A}) \equiv \sum_{n=0}^{\infty} \frac{(-1)^n \mathbf{A}^{2n}}{(2n)!} = \mathbf{I} - \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^4}{4!} - \dots$$

Power series expansion is usually not the best numeric technique for the computation of matrix exponentials; however when the structure of the  $\mathbf{A}$  matrix is appropriate, power series methods are one approach for determining closed form solutions for the matrix exponential of  $\mathbf{A}$ .

### D. Laplace Transform

The formula

$$\mathbf{F}e^{\mathbf{A}t} = \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\}$$

is derived by taking the Laplace transform of both sides of eqn. (3)

$$\begin{aligned} \mathcal{L}\{e^{\mathbf{A}t}\} &= \mathcal{L}\left\{\mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2!} + \frac{(\mathbf{A}t)^3}{3!} + \dots\right\} \\ &= \frac{1}{s}\mathbf{I} + \frac{1}{s^2}\mathbf{A} + \frac{1}{s^3}\mathbf{A}^2 + \frac{1}{s^4}\mathbf{A}^3 + \dots \\ &= (s\mathbf{I} - \mathbf{A})^{-1} \\ \mathbf{A}e^{\mathbf{A}t} &= \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\}. \end{aligned}$$

This derivation has used the fact that

$$(s\mathbf{I} - \mathbf{A})^{-1} = \left(\frac{1}{s}\mathbf{I} + \frac{1}{s^2}\mathbf{A} + \frac{1}{s^3}\mathbf{A}^2 + \frac{1}{s^4}\mathbf{A}^3 + \dots\right)$$

which can be shown by direct multiplication of both sides of the equation by  $(s\mathbf{I} - \mathbf{A})$ .

## III. KRONECKER AND VEC OPERATORS

### A. The Kronecker Product

The Kronecker product of an  $m \times n$  matrix  $\mathbf{A}$  and an  $r \times q$  matrix  $\mathbf{B}$ , is an  $mr \times nq$  matrix,  $\mathbf{A} \otimes \mathbf{B}$  defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \dots & A_{1n}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \dots & A_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}\mathbf{B} & A_{m2}\mathbf{B} & \dots & A_{mn}\mathbf{B} \end{bmatrix}$$

The Kronecker product has the following properties

$$\begin{aligned}
\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) &= \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C} \\
\mathbf{A} \otimes \mathbf{B} &\neq \mathbf{B} \otimes \mathbf{A} \\
\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) &= (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} \\
(\alpha_A \mathbf{A} \otimes \alpha_B \mathbf{B}) &= \alpha_A \alpha_B (\mathbf{A} \otimes \mathbf{B}) \\
(\mathbf{A} \otimes \mathbf{B})^\top &= \mathbf{A}^\top \otimes \mathbf{B}^\top \\
(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= (\mathbf{AC}) \otimes \mathbf{BD} \\
(\mathbf{A} \otimes \mathbf{B})^{-1} &= \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \\
(\mathbf{A} \otimes \mathbf{B})^+ &= \mathbf{A}^+ \otimes \mathbf{B}^+ \\
\text{rank}(\mathbf{A} \otimes \mathbf{B}) &= \text{rank}(\mathbf{A}) \text{rank}(\mathbf{B}) \\
\text{Tr}(\mathbf{A} \otimes \mathbf{B}) &= \text{Tr}(\mathbf{A}) \text{Tr}(\mathbf{B}) = \text{Tr}(\mathbf{\Lambda}_A \otimes \mathbf{\Lambda}_B) \\
\det(\mathbf{A} \otimes \mathbf{B}) &= \det(\mathbf{A})^{\text{rank}(\mathbf{B})} \det(\mathbf{B})^{\text{rank}(\mathbf{A})} \\
\text{eig}(\mathbf{A} \otimes \mathbf{B}) &= \text{eig}(\mathbf{B} \otimes \mathbf{A}) \text{ if } \mathbf{A}, \mathbf{B} \text{ are square} \\
\text{eig}(\mathbf{A} \otimes \mathbf{B}) &= \text{eig}(\mathbf{A}) \text{eig}(\mathbf{B})^\top \text{ if } \mathbf{A}, \mathbf{B} \text{ symm. \& sq.} \\
\text{eig}(\mathbf{A} \otimes \mathbf{B}) &= \text{eig}(\mathbf{A}) \otimes \text{eig}(\mathbf{B})
\end{aligned}$$

where  $\mathbf{\Lambda}_A$  denotes the diagonal matrix with the eigenvalues of  $\mathbf{A}$ .

#### B. The Vec Operator

The vec-operator applied on a matrix  $\mathbf{A}$  stacks the columns into a vector, i.e. for a  $2 \times 2$  matrix

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \mapsto \text{vec}(\mathbf{A}) = \begin{bmatrix} A_{11} \\ A_{21} \\ A_{12} \\ A_{22} \end{bmatrix}$$

Properties of the vec-operator include

$$\begin{aligned}
\text{vec}(\mathbf{AXB}) &= (\mathbf{B}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{X}) \\
\text{Tr}(\mathbf{A}^\top \mathbf{B}) &= \text{vec}(\mathbf{A})^\top \text{vec}(\mathbf{B}) \\
\text{vec}(\mathbf{A} + \mathbf{B}) &= \text{vec}(\mathbf{A}) + \text{vec}(\mathbf{B}) \\
\text{vec}(\alpha \mathbf{A}) &= \alpha \text{vec}(\mathbf{A}) \\
\mathbf{a}^\top \mathbf{X} \mathbf{B} \mathbf{X}^\top \mathbf{c} &= \text{vec}(\mathbf{X})^\top (\mathbf{B} \otimes \mathbf{ca}^\top) \text{vec}(\mathbf{X}) \quad (4)
\end{aligned}$$

The proof for Eq. 4 is provided in the appendix.

#### IV. DERIVATIVES OF MATRICES

This section covers differentiation of a number of expressions with respect to a matrix  $\mathbf{X}$ . Note that it is always assumed that  $\mathbf{X}$  has no special structure, i.e. that the elements of  $\mathbf{X}$  are independent (e.g. not symmetric, Toeplitz, positive definite). See Section IV-H for differentiation of structured matrices. The basic assumptions can be written as

$$\frac{\partial \mathbf{X}_{kl}}{\partial \mathbf{X}_{ij}} = \delta_{ik} \delta_{lj}$$

for vector forms,

$$\left[ \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right]_i = \frac{\partial x_i}{\partial y_i}, \quad \left[ \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right]_i = \frac{\partial x}{\partial y_i}, \quad \left[ \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right]_{ij} = \frac{\partial x_i}{\partial y_j}$$

The following rules are general and very useful when deriving the differential of an expression:

$$\begin{aligned}
\partial \mathbf{A} &= 0 \quad (\mathbf{A} \text{ is a constant}) \\
\partial(\alpha \mathbf{X}) &= \alpha \partial \mathbf{X} \\
\partial(\mathbf{X} + \mathbf{Y}) &= \partial \mathbf{X} + \partial \mathbf{Y} \\
\partial(\text{Tr}(\mathbf{X})) &= \text{Tr}(\partial \mathbf{X}) \\
\partial(\mathbf{XY}) &= (\partial \mathbf{X})\mathbf{Y} + \mathbf{X}(\partial \mathbf{Y}) \\
\partial(\mathbf{X} \circ \mathbf{Y}) &= (\partial \mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial \mathbf{Y}) \\
\partial(\mathbf{X} \otimes \mathbf{Y}) &= (\partial \mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes (\partial \mathbf{Y}) \\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}(\partial \mathbf{X})\mathbf{X}^{-1} \\
\partial(\det(\mathbf{X})) &= \text{Tr}(\text{adj}(\mathbf{X})\partial \mathbf{X}) \\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X}) \text{Tr}(\mathbf{X}^{-1} \partial \mathbf{X}) \\
\partial(\ln(\det(\mathbf{X}))) &= \text{Tr}(\mathbf{X}^{-1} \partial \mathbf{X}) \\
\partial \mathbf{X}^\top &= (\partial \mathbf{X})^\top \\
\partial \mathbf{X}^H &= (\partial \mathbf{X})^H
\end{aligned}$$

#### A. Derivatives of a Determinant

General Form:

$$\begin{aligned}
\frac{\partial \det(\mathbf{Y})}{\partial x} &= \det(\mathbf{Y}) \text{Tr} \left[ \mathbf{Y}^{-1} \frac{\partial \mathbf{Y}}{\partial x} \right] \\
\sum_k \frac{\partial \det(\mathbf{Y})}{\partial X_{ik}} X_{jk} &= \delta_{ij} \det(\mathbf{X}) \\
\frac{\partial^2 \det(\mathbf{Y})}{\partial x^2} &= \det(\mathbf{Y}) \left\{ \text{Tr} \left[ \mathbf{Y}^{-1} \frac{\partial^2 \mathbf{Y}}{\partial x^2} \right] \right. \\
&\quad + \text{Tr} \left[ \mathbf{Y}^{-1} \frac{\partial \mathbf{Y}}{\partial x} \right] \text{Tr} \left[ \mathbf{Y}^{-1} \frac{\partial \mathbf{Y}}{\partial x} \right] \\
&\quad \left. - \text{Tr} \left[ \left( \mathbf{Y}^{-1} \frac{\partial \mathbf{Y}}{\partial x} \right) \left( \mathbf{Y}^{-1} \frac{\partial \mathbf{Y}}{\partial x} \right) \right] \right\}
\end{aligned}$$

Linear Forms:

$$\begin{aligned}
\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} &= \det(\mathbf{X})(\mathbf{X}^{-1})^\top \\
\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} &= \delta_{ij} \det(\mathbf{X}) \\
\frac{\partial \det(\mathbf{AXB})}{\partial \mathbf{X}} &= \det(\mathbf{AXB})(\mathbf{X}^{-1})^\top \\
&= \det(\mathbf{AXB})(\mathbf{X}^\top)^{-1}
\end{aligned}$$

Square Forms: If  $\mathbf{X}$  is square and invertible, then

$$\frac{\partial \det(\mathbf{X}^\top \mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = 2 \det(\mathbf{X}^\top \mathbf{A} \mathbf{X}) \mathbf{X}^{-\top}$$

If  $\mathbf{X}$  is not square but  $\mathbf{A}$  is symmetric, then

$$\frac{\partial \det(\mathbf{X}^\top \mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = 2 \det(\mathbf{X}^\top \mathbf{A} \mathbf{X}) \mathbf{A} \mathbf{X} (\mathbf{X}^\top \mathbf{A} \mathbf{X})^{-1}$$

If  $\mathbf{X}$  is not square and  $\mathbf{A}$  is not symmetric, then

$$\begin{aligned}
\frac{\partial \det(\mathbf{X}^\top \mathbf{A} \mathbf{X})}{\partial \mathbf{X}} &= 2 \det(\mathbf{X}^\top \mathbf{A} \mathbf{X}) \\
&\quad \left( \mathbf{A} \mathbf{X} (\mathbf{X}^\top \mathbf{A} \mathbf{X})^{-1} + \mathbf{A}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{A}^\top \mathbf{X})^{-1} \right)
\end{aligned}$$

Other Nonlinear Forms: Some special cases are

$$\begin{aligned}\frac{\partial \ln(\det(\mathbf{X}^\top \mathbf{X}))}{\partial \mathbf{X}} &= 2(\mathbf{X}^+)^\top \\ \frac{\partial \ln(\det(\mathbf{X}^\top \mathbf{X}))}{\partial \mathbf{X}^+} &= -2\mathbf{X}^\top \\ \frac{\partial \ln(\det(\mathbf{X}))}{\partial \mathbf{X}} &= (\mathbf{X}^{-1})^\top = (\mathbf{X}^\top)^{-1} \\ \frac{\partial \det(\mathbf{X}^k)}{\partial \mathbf{X}} &= k \det(\mathbf{X}^k) \mathbf{X}^{-\top}\end{aligned}$$

### B. Derivatives of an Inverse

The basic identity is

$$\frac{\partial \mathbf{Y}^{-1}}{\partial x} = \mathbf{Y}^{-1} \frac{\partial \mathbf{Y}}{\partial x} \mathbf{Y}^{-1}$$

from which it follows

$$\begin{aligned}\frac{\partial (\mathbf{X}^{-1})_{kl}}{\partial X_{ij}} &= (\mathbf{X}^{-1})_{ki} (\mathbf{X}^{-1})_{jl} \\ \frac{\partial \mathbf{a}^\top \mathbf{X}^{-1} \mathbf{b}}{\partial \mathbf{X}} &= -\mathbf{X}^{-\top} \mathbf{a} \mathbf{b}^\top \mathbf{X}^{-\top} \\ \frac{\partial \det(\mathbf{X}^{-1})}{\partial \mathbf{X}} &= -\det(\mathbf{X}^{-1}) (\mathbf{X}^{-1})^\top \\ \frac{\partial \text{Tr}(\mathbf{A} \mathbf{X}^{-1} \mathbf{B})}{\partial \mathbf{X}} &= -(\mathbf{X}^{-1} \mathbf{B} \mathbf{A} \mathbf{X}^{-1})^\top \\ \frac{\partial \text{Tr}((\mathbf{X} + \mathbf{A})^{-1})}{\partial \mathbf{X}} &= -((\mathbf{X} + \mathbf{A})^{-1} (\mathbf{X} + \mathbf{A})^{-1})^\top\end{aligned}$$

We also have the following result: Let  $\mathbf{A}$  be an  $n \times n$  invertible square matrix,  $\mathbf{W}$  be the inverse of  $\mathbf{A}$ , and  $J(\mathbf{A})$  is an  $n \times n$  -variate and differentiable function with respect to  $\mathbf{A}$ , then the partial differentials of  $J$  with respect to  $\mathbf{A}$  and  $\mathbf{W}$  satisfy

$$\frac{\partial J}{\partial \mathbf{A}} = -\mathbf{A}^{-\top} \frac{\partial J}{\partial \mathbf{W}} \mathbf{A}^{-\top}$$

### C. Derivatives of Eigenvalues

The following hold,

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}} \sum \text{eig}(\mathbf{X}) &= \frac{\partial}{\partial \mathbf{x}} \text{Tr}(\mathbf{X}) = \mathbf{I} \\ \frac{\partial}{\partial \mathbf{x}} \prod \text{eig}(\mathbf{X}) &= \frac{\partial}{\partial \mathbf{x}} \det(\mathbf{X}) = \det(\mathbf{X}) \mathbf{X}^{-\top}\end{aligned}$$

If  $\mathbf{A}$  is real and symmetric,  $\lambda_i$  and  $\mathbf{v}_i$  are distinct eigenvalues and eigenvectors of  $\mathbf{A}$  with  $\mathbf{v}_i^\top \mathbf{v}_i = 1$ , then

$$\begin{aligned}\partial \lambda_i &= \mathbf{v}_i^\top \partial(\mathbf{A}) \mathbf{v}_i \\ \partial \mathbf{v}_i &= (\lambda_i \mathbf{I} - \mathbf{A})^+ \partial(\mathbf{A}) \mathbf{v}_i\end{aligned}$$

### D. Derivatives of Matrices, Vectors and Scalar Forms

TODO: Add text here...

First Order:

$$\begin{aligned}\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} &= \frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \\ \frac{\partial \mathbf{a}^\top \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} &= \mathbf{a} \mathbf{b}^\top \\ \frac{\partial \mathbf{a}^\top \mathbf{X}^\top \mathbf{b}}{\partial \mathbf{X}} &= \mathbf{b} \mathbf{a}^\top \\ \frac{\partial \mathbf{a}^\top \mathbf{X} \mathbf{a}}{\partial \mathbf{X}} &= \frac{\partial \mathbf{a}^\top \mathbf{X}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} \mathbf{a}^\top \\ \frac{\partial \mathbf{X}}{\partial X_{ij}} &= \mathbf{J}^{ij} \\ \frac{\partial (\mathbf{X} \mathbf{A})_{ij}}{\partial X_{mn}} &= \delta_{im} (\mathbf{A})_{nj} = (\mathbf{J}^{mn} \mathbf{A})_{ij} \\ \frac{\partial (\mathbf{X}^\top \mathbf{A})_{ij}}{\partial X_{mn}} &= \delta_{in} (\mathbf{A})_{mj} = (\mathbf{J}^{nm} \mathbf{A})_{ij}\end{aligned}$$

Second Order:

$$\begin{aligned}\frac{\partial}{\partial X_{ij}} \sum_{klmn} X_{kl} X_{mn} &= 2 \sum_{kl} X_{kl} \\ \frac{\partial \mathbf{b}^\top \mathbf{X}^\top \mathbf{X} \mathbf{c}}{\partial \mathbf{X}} &= \mathbf{X} (\mathbf{b} \mathbf{c}^\top + \mathbf{c} \mathbf{b}^\top) \\ \frac{\partial (\mathbf{B} \mathbf{x} + \mathbf{b})^\top \mathbf{C} (\mathbf{D} \mathbf{x} + \mathbf{d})}{\partial \mathbf{x}} &= \mathbf{B}^\top \mathbf{C} (\mathbf{D} \mathbf{x} + \mathbf{d}) + \mathbf{D}^\top \mathbf{C}^\top (\mathbf{B} \mathbf{x} + \mathbf{b}) \\ \frac{\partial (\mathbf{X}^\top \mathbf{B} \mathbf{X})_{kl}}{\partial X_{ij}} &= \delta_{lj} (\mathbf{X}^\top \mathbf{B})_{ki} + \delta_{kj} (\mathbf{B} \mathbf{X})_{il} \\ \frac{\partial (\mathbf{X}^\top \mathbf{B} \mathbf{X})}{\partial \mathbf{X}_{ij}} &= \mathbf{X}^\top \mathbf{B} \mathbf{J}^{ij} + \mathbf{J}^{ji} \mathbf{B} \mathbf{X}, \quad (\mathbf{J}^{ij})_{kl} = \delta_{ik} \delta_{jl} \\ \frac{\partial \mathbf{x}^\top \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} &= (\mathbf{B}^\top + \mathbf{B}) \mathbf{x} \\ \frac{\partial \mathbf{b}^\top \mathbf{X}^\top \mathbf{D} \mathbf{X} \mathbf{c}}{\partial \mathbf{X}} &= \mathbf{D}^\top \mathbf{X} \mathbf{b} \mathbf{c}^\top + \mathbf{D} \mathbf{X} \mathbf{c} \mathbf{b}^\top \\ \frac{\partial}{\partial \mathbf{X}} (\mathbf{X} \mathbf{b} + \mathbf{c})^\top \mathbf{D} (\mathbf{X} \mathbf{b} + \mathbf{c}) &= (\mathbf{D} + \mathbf{D}^\top) (\mathbf{X} \mathbf{b} + \mathbf{c}) \mathbf{b}^\top\end{aligned}$$

Assume  $\mathbf{W}$  is symmetric, then

$$\begin{aligned}\frac{\partial}{\partial \mathbf{s}} (\mathbf{x} - \mathbf{A} \mathbf{s})^\top \mathbf{W} (\mathbf{x} - \mathbf{A} \mathbf{s}) &= -2 \mathbf{A}^\top \mathbf{W} (\mathbf{x} - \mathbf{A} \mathbf{s}) \\ \frac{\partial}{\partial \mathbf{x}} (\mathbf{x} - \mathbf{s})^\top \mathbf{W} (\mathbf{x} - \mathbf{s}) &= 2 \mathbf{W} (\mathbf{x} - \mathbf{s}) \\ \frac{\partial}{\partial \mathbf{s}} (\mathbf{x} - \mathbf{s})^\top \mathbf{W} (\mathbf{x} - \mathbf{s}) &= -2 \mathbf{W} (\mathbf{x} - \mathbf{s}) \\ \frac{\partial}{\partial \mathbf{x}} (\mathbf{x} - \mathbf{A} \mathbf{s})^\top \mathbf{W} (\mathbf{x} - \mathbf{A} \mathbf{s}) &= 2 \mathbf{W} (\mathbf{x} - \mathbf{A} \mathbf{s}) \\ \frac{\partial}{\partial \mathbf{A}} (\mathbf{x} - \mathbf{A} \mathbf{s})^\top \mathbf{W} (\mathbf{x} - \mathbf{A} \mathbf{s}) &= -2 \mathbf{W} (\mathbf{x} - \mathbf{A} \mathbf{s}) \mathbf{s}^\top\end{aligned}$$

Higher-order and Nonlinear:

$$\begin{aligned}\frac{\partial(\mathbf{X}^n)_{kl}}{\partial \mathbf{X}_{ij}} &= \sum_{r=0}^{n-1} (\mathbf{X}^r \mathbf{J}^{ij} \mathbf{X}^{n-1-r})_{kl} \\ \frac{\partial}{\partial \mathbf{X}} \mathbf{a}^\top \mathbf{X}^n \mathbf{b} &= \sum_{r=0}^{n-1} (\mathbf{X}^r)^\top \mathbf{a} \mathbf{b}^\top (\mathbf{X}^{n-1-r})^\top \\ \frac{\partial}{\partial \mathbf{X}} \mathbf{a}^\top (\mathbf{X}^n)^\top \mathbf{X} \mathbf{b} &= \sum_{r=0}^{n-1} [\mathbf{X}^{n-1-r} \mathbf{a} \mathbf{b}^\top (\mathbf{X}^n)^\top \mathbf{X}^r \\ &\quad + (\mathbf{X}^r)^\top \mathbf{X}^n \mathbf{a} \mathbf{b}^\top (\mathbf{X}^{n-1-r})^\top]\end{aligned}$$

Gradient and Hessian: Using the above we have the gradient and Hessian

$$\begin{aligned}f &= \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{x} \\ \nabla_{\mathbf{x}} f &= \frac{\partial f}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^\top) \mathbf{x} + \mathbf{b} \\ \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^\top} &= \mathbf{A} + \mathbf{A}^\top\end{aligned}$$

#### E. Derivatives of Traces

Assume  $F(\mathbf{X})$  to be a differentiable function of each of the elements of  $\mathbf{X}$ . It then holds that

$$\frac{\partial \text{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^\top$$

where  $f(\cdot)$  is the scalar derivative notation of  $F(\cdot)$ .

First Order:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}) &= \mathbf{I} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X} \mathbf{A}) &= \mathbf{A}^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A} \mathbf{X} \mathbf{B}) &= \mathbf{A}^\top \mathbf{B}^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A} \mathbf{X}^\top \mathbf{B}) &= \mathbf{B} \mathbf{A} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^\top \mathbf{A}) &= \mathbf{A} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A} \mathbf{X}^\top) &= \mathbf{A} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A} \otimes \mathbf{X}) &= \text{Tr}(\mathbf{A}) \mathbf{I}\end{aligned}$$

Second Order:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^\top \mathbf{X}) &= \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X} \mathbf{X}^\top) = 2\mathbf{X} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^2) &= 2\mathbf{X}^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^2 \mathbf{B}) &= (\mathbf{X} \mathbf{B} + \mathbf{B} \mathbf{X})^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^\top \mathbf{B} \mathbf{X}) &= \mathbf{B} \mathbf{X} + \mathbf{B}^\top \mathbf{X} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{B} \mathbf{X} \mathbf{X}^\top) &= \mathbf{B} \mathbf{X} + \mathbf{B}^\top \mathbf{X} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X} \mathbf{X}^\top \mathbf{B}) &= \mathbf{B} \mathbf{X} + \mathbf{B}^\top \mathbf{X} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X} \mathbf{B} \mathbf{X}^\top) &= \mathbf{X} \mathbf{B}^\top + \mathbf{X} \mathbf{B} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{B} \mathbf{X}^\top \mathbf{X}) &= \mathbf{X} \mathbf{B}^\top + \mathbf{X} \mathbf{B} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^\top \mathbf{X} \mathbf{B}) &= \mathbf{X} \mathbf{B}^\top + \mathbf{X} \mathbf{B} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A} \mathbf{X} \mathbf{B} \mathbf{X}) &= \mathbf{A}^\top \mathbf{X}^\top \mathbf{B}^\top + \mathbf{B}^\top \mathbf{X}^\top \mathbf{A}^\top\end{aligned}\tag{5}$$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{B}^\top \mathbf{X}^\top \mathbf{C} \mathbf{X} \mathbf{B}) &= \mathbf{C}^\top \mathbf{X} \mathbf{B} \mathbf{B}^\top + \mathbf{C} \mathbf{X} \mathbf{B} \mathbf{B}^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^\top \mathbf{B} \mathbf{X} \mathbf{C}) &= \mathbf{B} \mathbf{X} \mathbf{C} + \mathbf{B}^\top \mathbf{X} \mathbf{C}^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A} \mathbf{X} \mathbf{B} \mathbf{X}^\top \mathbf{C}) &= \mathbf{A}^\top \mathbf{C}^\top \mathbf{X} \mathbf{B}^\top + \mathbf{C} \mathbf{A} \mathbf{X} \mathbf{B} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}[(\mathbf{A} \mathbf{X} \mathbf{B} + \mathbf{C})(\mathbf{A} \mathbf{X} \mathbf{B} + \mathbf{C})^\top] &= 2\mathbf{A}^\top (\mathbf{A} \mathbf{X} \mathbf{B} + \mathbf{C}) \mathbf{B}^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A} \otimes \mathbf{X}) &= \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}) \text{Tr}(\mathbf{X}) = 2 \text{Tr}(\mathbf{X}) \mathbf{I}\end{aligned}$$

Higher-order:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X}^k) &= k(\mathbf{X}^{k-1})^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A} \mathbf{X}^k) &= \sum_{r=0}^{k-1} (\mathbf{X}^r \mathbf{A} \mathbf{X}^{k-r-1})^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{B}^\top \mathbf{X}^\top \mathbf{C} \mathbf{X} \mathbf{X}^\top \mathbf{C} \mathbf{X} \mathbf{B}) &= \mathbf{C} \mathbf{X} \mathbf{X}^\top \mathbf{C} \mathbf{X} \mathbf{B} \mathbf{B}^\top \\ &\quad + \mathbf{C}^\top \mathbf{X} \mathbf{B} \mathbf{B}^\top \mathbf{X}^\top \mathbf{C}^\top \mathbf{X} \\ &\quad + \mathbf{C} \mathbf{X} \mathbf{B} \mathbf{B}^\top \mathbf{X}^\top \mathbf{C} \mathbf{X} \\ &\quad + \mathbf{C}^\top \mathbf{X} \mathbf{X}^\top \mathbf{C}^\top \mathbf{X} \mathbf{B} \mathbf{B}^\top\end{aligned}$$

Other: Consider

$$\frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{A} \mathbf{X}^{-1} \mathbf{B}) = -(\mathbf{X}^{-1} \mathbf{B} \mathbf{A} \mathbf{X}^{-1})^\top = -\mathbf{X}^{-\top} \mathbf{A}^\top \mathbf{B}^\top \mathbf{X}^{-\top}.$$



Assume  $\mathbf{B}$  and  $\mathbf{C}$  to be symmetric, then

$$\begin{aligned} \frac{\partial}{\partial \mathbf{X}} \text{Tr}[(\mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} \mathbf{A}] &= \\ &- (\mathbf{C} \mathbf{X} (\mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1}) (\mathbf{A} + \mathbf{A}^\top) (\mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}[(\mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{B} \mathbf{X})] &= \\ &- 2 \mathbf{C} \mathbf{X} (\mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{B} \mathbf{X} (\mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} \\ &+ 2 \mathbf{B} \mathbf{X} (\mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr}[(\mathbf{A} + \mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{B} \mathbf{X})] &= \\ &- 2 \mathbf{C} \mathbf{X} (\mathbf{A} + \mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{B} \mathbf{X} (\mathbf{A} + \mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} \\ &+ 2 \mathbf{B} \mathbf{X} (\mathbf{A} + \mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1} \end{aligned}$$

#### F. Derivatives of Vector Norms

Two-norm:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} \|\mathbf{x} - \mathbf{a}\|_2 &= \frac{\mathbf{x} - \mathbf{a}}{\|\mathbf{x} - \mathbf{a}\|_2} \\ \frac{\partial}{\partial \mathbf{x}} \frac{\mathbf{x} - \mathbf{a}}{\|\mathbf{x} - \mathbf{a}\|_2} &= \frac{\mathbf{I}}{\|\mathbf{x} - \mathbf{a}\|_2} - \frac{(\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})^\top}{\|\mathbf{x} - \mathbf{a}\|_2^3} \\ \frac{\partial \|\mathbf{x}\|_2^2}{\partial \mathbf{x}} &= \frac{\partial \|\mathbf{x}^\top \mathbf{x}\|_2}{\partial \mathbf{x}} = 2\mathbf{x} \end{aligned}$$

#### G. Derivatives of Matrix Norms

Frobenius norm:

$$\frac{\partial}{\partial \mathbf{X}} \|\mathbf{X}\|_F^2 = \frac{\partial}{\partial \mathbf{X}} \text{Tr}(\mathbf{X} \mathbf{X}^H) = 2\mathbf{X}$$

Note that this is also a special case of the result in eqn. (5)

#### H. Derivatives of Structured Matrices

Assume that the matrix  $\mathbf{A}$  has some structure, i.e. symmetric, toeplitz, etc. In that case the derivatives of the previous section does not apply in general. Instead, consider the following general rule for differentiating a scalar function  $f(\mathbf{A})$

$$\frac{df}{dA_{ij}} = \sum_{kl} \frac{\partial f}{\partial A_{kl}} \frac{\partial A_{kl}}{\partial A_{ij}} = \text{Tr} \left[ \left[ \frac{\partial f}{\partial \mathbf{A}} \right]^\top \frac{\partial \mathbf{A}}{\partial A_{ij}} \right]$$

The matrix differentiated with respect to itself is in this text referred to as the *structure matrix* of  $\mathbf{A}$  and is defined simply by

$$\frac{\partial \mathbf{A}}{\partial A_{ij}} = \mathbf{S}^{ij}$$

If  $\mathbf{A}$  has no special structure we have simply  $\mathbf{S}^{ij} = \mathbf{J}^{ij}$ , that is, the structure matrix is simply the single-entry matrix. Many structures have a representation in single-entry matrices.

*The Chain Rule:* Sometimes the objective is to find the derivative of a matrix which is a function of another matrix. Let  $\mathbf{U} = f(\mathbf{X})$ , the goal is to find the derivative of the function  $g(\mathbf{U})$  with respect to  $\mathbf{X}$ :

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}}$$

Then the Chain Rule can then be written the following way:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^M \sum_{l=1}^N \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}}$$

Using matrix notation, this can be written as:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(\mathbf{U})}{\partial \mathbf{X}_{ij}} = \text{Tr} \left[ \left( \frac{\partial g(\mathbf{U})}{\partial \mathbf{U}} \right)^\top \frac{\partial \mathbf{U}}{\partial \mathbf{X}_{ij}} \right].$$

*Symmetric:* If  $\mathbf{A}$  is symmetric, then  $\mathbf{S}^{ij} = \mathbf{J}^{ij} + \mathbf{J}^{ji} - \mathbf{J}^{ij} \mathbf{J}^{ij}$  and therefore

$$\frac{df}{d\mathbf{A}} = \left[ \frac{\partial f}{\partial \mathbf{A}} \right] + \left[ \frac{\partial f}{\partial \mathbf{A}} \right]^\top - \text{diag} \left[ \frac{\partial f}{\partial \mathbf{A}} \right].$$

That is, e.g.,

$$\begin{aligned} \frac{\partial \text{Tr}(\mathbf{A} \mathbf{X})}{\partial \mathbf{X}} &= \mathbf{A} + \mathbf{A}^\top - (\mathbf{A} \circ \mathbf{I}). \\ \frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} &= \det(\mathbf{X}) (2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I})) \\ \frac{\partial \ln \det(\mathbf{X})}{\partial \mathbf{X}} &= 2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I}) \end{aligned}$$

*Diagonal:* If  $\mathbf{X}$  is diagonal, then

$$\frac{\partial \text{Tr}(\mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = \mathbf{A} \circ \mathbf{I}$$

*Toeplitz:* Like symmetric matrices and diagonal matrices, Toeplitz matrices have a special structure which should be taken into account when the derivative with respect to a matrix with Toeplitz structure.

$$\begin{aligned} \frac{\partial \text{Tr}(\mathbf{A} \mathbf{T})}{\partial \mathbf{T}} &= \frac{\partial \text{Tr}(\mathbf{T} \mathbf{A})}{\partial \mathbf{T}} \\ &= \begin{bmatrix} \text{Tr}(\mathbf{A}) & \text{Tr}([\mathbf{A}^\top]_{n1}) & \text{Tr}([\mathbf{A}^\top]_{1n} \mathbf{A}_{n-1,2}) & \cdots & \mathbf{A}_{n1} \\ \text{Tr}([\mathbf{A}^\top]_{1n}) & \text{Tr}(\mathbf{A}) & \ddots & \ddots & \vdots \\ \text{Tr}([\mathbf{A}^\top]_{1n} \mathbf{A}_{2,n-1}) & \ddots & \ddots & \ddots & \vdots \\ \mathbf{A}_{1n} & \cdots & \text{Tr}([\mathbf{A}^\top]_{1n} \mathbf{A}_{2,n-1}) & \text{Tr}([\mathbf{A}^\top]_{1n}) & \text{Tr}([\mathbf{A}^\top]_{1n} \mathbf{A}_{n-1,2}) \\ \vdots & \vdots & \vdots & \vdots & \text{Tr}(\mathbf{A}) \end{bmatrix} \\ &\equiv \alpha(\mathbf{A}) \end{aligned}$$

As it can be seen, the derivative  $\alpha(\mathbf{A})$  also has a Toeplitz structure. Each value in the diagonal is the sum of all the diagonal valued in  $\mathbf{A}$ , the values in the diagonals next to the main diagonal equal the sum of the diagonal next to the main diagonal in  $\mathbf{A}^\top$ . This result is only valid for the unconstrained Toeplitz matrix. If the Toeplitz matrix also is symmetric, the same derivative yields

$$\frac{\partial \text{Tr}(\mathbf{A} \mathbf{T})}{\partial \mathbf{T}} = \frac{\partial \text{Tr}(\mathbf{T} \mathbf{A})}{\partial \mathbf{T}} = \alpha(\mathbf{A}) + \alpha(\mathbf{A})^\top - \alpha(\mathbf{A}) \circ \mathbf{I}.$$

### I. Derivatives with Respect to Scalars

In the case where the vectors  $\mathbf{u}$  and  $\mathbf{v}$ , the matrix  $\mathbf{A}$ , and the scalar  $\mu$  are functions of a scalar quantity  $s$ , the derivative with respect to  $s$  has the following properties:

$$\begin{aligned}\frac{d}{ds}(\mathbf{u} + \mathbf{v}) &= \frac{d\mathbf{u}}{ds} + \frac{d\mathbf{v}}{ds} \\ \frac{d}{ds}(\mathbf{u} \cdot \mathbf{v}) &= \frac{d\mathbf{u}}{ds} \cdot \mathbf{v} + \mathbf{u} \cdot \frac{d\mathbf{v}}{ds} \\ \frac{d}{ds}(\mathbf{u} \times \mathbf{v}) &= \frac{d\mathbf{u}}{ds} \times \mathbf{v} + \mathbf{u} \times \frac{d\mathbf{v}}{ds} \\ \frac{d}{ds}(\mu\mathbf{v}) &= \frac{d\mu}{ds}\mathbf{v} + \mu\frac{d\mathbf{v}}{ds} \\ \frac{d}{ds}\mathbf{A}^{-1} &= -\mathbf{A}^{-1}\left(\frac{d}{ds}\mathbf{A}\right)\mathbf{A}^{-1}.\end{aligned}$$

### J. Derivatives with Respect to Vectors

Using the convention that gradients of scalar functions are defined as row vectors,

$$\begin{aligned}\frac{d}{d\mathbf{v}}(\mathbf{u} \cdot \mathbf{v}) &= \frac{d}{d\mathbf{v}}(\mathbf{v} \cdot \mathbf{u}) = \mathbf{u}^\top \\ \frac{d}{d\mathbf{v}}(\mathbf{A}\mathbf{v}) &= \mathbf{A} \\ \frac{d}{d\mathbf{v}}(\mathbf{v}^\top \mathbf{A}) &= \mathbf{A}^\top \\ \frac{d}{d\mathbf{v}}(\mathbf{v}^\top \mathbf{A}\mathbf{v}) &= \mathbf{v}^\top(\mathbf{A} + \mathbf{A}^\top) \\ &= 2\mathbf{v}^\top \mathbf{A}, \text{ if } \mathbf{A} \text{ is symmetric.}\end{aligned}$$

### K. Derivatives with Respect to Matrices

The derivative of the scalar  $\mu$  with respect to the matrix  $\mathbf{A}$  is defined by

$$\frac{d\mu}{d\mathbf{A}} = \begin{bmatrix} \frac{d\mu}{da_{11}} & \frac{d\mu}{da_{12}} & \cdots & \frac{d\mu}{da_{1n}} \\ \frac{d\mu}{da_{21}} & \frac{d\mu}{da_{22}} & \cdots & \frac{d\mu}{da_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d\mu}{da_{m1}} & \frac{d\mu}{da_{m2}} & \cdots & \frac{d\mu}{da_{mn}} \end{bmatrix}.$$

For the scalar operation defined by the trace, the following differentiation formulas are useful:

$$\begin{aligned}\frac{d(\text{Tr}(\mathbf{A}\mathbf{B}))}{d\mathbf{A}} &= \mathbf{B}^\top \quad (\mathbf{A}\mathbf{B} \text{ must be square}) \\ \frac{d(\text{Tr}(\mathbf{A}\mathbf{B}\mathbf{A}^\top))}{d\mathbf{A}} &= 2\mathbf{A}\mathbf{B}^\top \quad (\mathbf{B} \text{ must be symmetric}).\end{aligned}$$

## V. INVERSES OF MATRICES

### A. Basics

1) *Definition:* The inverse  $\mathbf{A}^{-1}$  of a matrix  $\mathbf{A} \in \mathbb{C}^{n \times n}$  is defined

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

where  $\mathbf{I}$  is the  $n \times n$  identity matrix. If  $\mathbf{A}^{-1}$  exists,  $\mathbf{A}$  is said to be *nonsingular*. Otherwise,  $\mathbf{A}$  is said to be *singular*.

2) *Cofactors and Adjoint:* The *submatrix* of a matrix  $\mathbf{A}$ , denoted  $[\mathbf{A}]_{ij}$  is an  $(n-1) \times (n-1)$  matrix obtained by deleting the  $i$ -th row and the  $j$ -th column of  $\mathbf{A}$ . The  $(i, j)$  *cofactor* of a matrix is defined as

$$\text{cof}(\mathbf{A}, i, j) = (-1)^{i+j} \det([\mathbf{A}]_{ij}).$$

The *matrix of cofactors* can be created from the cofactors

$$\text{cof}(\mathbf{A}) = \begin{bmatrix} \text{cof}(\mathbf{A}, 1, 1) & \cdots & \text{cof}(\mathbf{A}, 1, n) \\ \vdots & \text{cof}(\mathbf{A}, i, j) & \vdots \\ \text{cof}(\mathbf{A}, n, 1) & \cdots & \text{cof}(\mathbf{A}, n, n) \end{bmatrix}.$$

The *adjoint* matrix is the transpose of the cofactor matrix

$$\text{adj}(\mathbf{A}) = (\text{cof}(\mathbf{A}))^\top.$$

3) *Determinant:* The *determinant* of a matrix  $\mathbf{A} \in \mathbb{C}^{n \times n}$  is defined as

$$\begin{aligned}\det(\mathbf{A}) &= \sum_{j=1}^n (-1)^{j+1} A_{1j} \det([\mathbf{A}]_{1j}) \\ &= \sum_{j=1}^n A_{1j} \text{cof}(\mathbf{A}, 1, j).\end{aligned}$$

4) *Construction:* The inverse matrix can be constructed, using the adjoint matrix, by

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \cdot \text{adj}(\mathbf{A}).$$

5) *Condition number:* See Section I-E.

### B. Exact Relations

1) *Basics:*

$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

2) *The Woodbury identity:* The Woodbury identity comes in many variants.

$$\begin{aligned}(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^\top)^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^\top\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^\top\mathbf{A}^{-1} \\ (\mathbf{A} + \mathbf{U}\mathbf{B}\mathbf{V})^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}\mathbf{A}^{-1}\end{aligned}$$

If  $\mathbf{P}, \mathbf{R}$  are positive definite, then

$$(\mathbf{P}^{-1} + \mathbf{B}^\top\mathbf{R}^{-1}\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{R}^{-1} = \mathbf{P}\mathbf{B}^\top(\mathbf{B}\mathbf{P}\mathbf{B}^\top + \mathbf{R})^{-1}.$$

3) *The Kailath Variant:*

$$(\mathbf{A} + \mathbf{B}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$$

4) *Sherman-Morrison:*

$$(\mathbf{A} + \mathbf{b}\mathbf{c}^\top)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{b}\mathbf{c}^\top\mathbf{A}^{-1}}{1 + \mathbf{c}^\top\mathbf{A}^{-1}\mathbf{b}}$$

5) *The Searle Set of Identities:* The following set of identities are common and useful

$$\begin{aligned}(\mathbf{I} + \mathbf{A}^{-1})^{-1} &= \mathbf{A}(\mathbf{A} + \mathbf{I})^{-1} \\ (\mathbf{A} + \mathbf{B}\mathbf{B}^\top)^{-1}\mathbf{B} &= \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{B}^\top\mathbf{A}^{-1}\mathbf{B})^{-1} \\ (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} &= \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B} = \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A} \\ \mathbf{A} - \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A} &= \mathbf{B} - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B} \\ \mathbf{A}^{-1} + \mathbf{B}^{-1} &= \mathbf{A}^{-1}(\mathbf{A} + \mathbf{B})\mathbf{B}^{-1} \\ (\mathbf{I} + \mathbf{A}\mathbf{B})^{-1} &= \mathbf{I} - \mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}\mathbf{B} \\ (\mathbf{I} + \mathbf{A}\mathbf{B})^{-1}\mathbf{A} &= \mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}\end{aligned}$$

6) *Rank-1 update of inverse of inner product:* Denote  $\mathbf{A} = (\mathbf{X}^\top \mathbf{X})^{-1}$  and that  $\mathbf{X}$  is extended to include a new column vector such that  $\tilde{\mathbf{X}} = [\mathbf{X} \mathbf{v}]$ .

$$(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} = \begin{bmatrix} \mathbf{A} + \frac{\mathbf{A} \mathbf{X}^\top \mathbf{v} \mathbf{v}^\top \mathbf{X} \mathbf{A}^\top}{\mathbf{v}^\top \mathbf{v} - \mathbf{v}^\top \mathbf{X} \mathbf{A} \mathbf{X}^\top \mathbf{v}} & \frac{-\mathbf{A} \mathbf{X}^\top \mathbf{v}}{\mathbf{v}^\top \mathbf{v} - \mathbf{v}^\top \mathbf{X} \mathbf{A} \mathbf{X}^\top \mathbf{v}} \\ \frac{-\mathbf{v}^\top \mathbf{X} \mathbf{A}^\top}{\mathbf{v}^\top \mathbf{v} - \mathbf{v}^\top \mathbf{X} \mathbf{A} \mathbf{X}^\top \mathbf{v}} & \frac{1}{\mathbf{v}^\top \mathbf{v} - \mathbf{v}^\top \mathbf{X} \mathbf{A} \mathbf{X}^\top \mathbf{v}} \end{bmatrix}.$$

7) *Rank-1 update of Moore-Penrose Inverse:* **TODO: Add text here...**

### C. Implication on Inverses

If  $\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} + \mathbf{B}^{-1}$  then  $\mathbf{A} \mathbf{B}^{-1} \mathbf{A} = \mathbf{B} \mathbf{A}^{-1} \mathbf{B}$ . Assume  $\mathbf{P}$  and  $\mathbf{R}$  to be positive definite and invertible, then

$$(\mathbf{P}^{-1} + \mathbf{B}^\top \mathbf{R}^{-1} \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{R}^{-1} = \mathbf{P} \mathbf{B}^\top (\mathbf{B} \mathbf{P} \mathbf{B}^\top + \mathbf{R})^{-1}$$

### D. Approximations

The following identity is known as the *Neuman series* of a matrix, which holds when  $|\lambda_i| < 1$  for all eigenvalues  $\lambda_i$

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{n=0}^{\infty} \mathbf{A}^n$$

which is equivalent to

$$(\mathbf{I} + \mathbf{A})^{-1} = \sum_{n=0}^{\infty} (-1)^n \mathbf{A}^n.$$

When  $|\lambda_i| < 1$  for all eigenvalues  $\lambda_i$ , it holds that  $\mathbf{A} \mapsto 0$  for  $n \mapsto 1$ , and the following approximations holds

$$\begin{aligned} (\mathbf{I} - \mathbf{A})^{-1} &\approx \mathbf{I} + \mathbf{A} + \mathbf{A}^2 \\ (\mathbf{I} + \mathbf{A})^{-1} &\approx \mathbf{I} - \mathbf{A} + \mathbf{A}^2 \end{aligned}$$

The following approximation holds when  $\mathbf{A}$  large and symmetric

$$\mathbf{A} - \mathbf{A}(\mathbf{I} + \mathbf{A})^{-1} \mathbf{A} \approx \mathbf{I} - \mathbf{A}^{-1}.$$

If  $\sigma^2$  is small compared to  $\mathbf{Q}$  and  $\mathbf{M}$  then

$$(\mathbf{Q} + \sigma^2 \mathbf{M})^{-1} \approx \mathbf{Q}^{-1} - \sigma^2 \mathbf{Q}^{-1} \mathbf{M} \mathbf{Q}^{-1}.$$

### E. Generalized Inverse

1) *Definition:* A generalized inverse matrix of the matrix  $\mathbf{A}$  is any matrix  $\mathbf{A}^-$  such that

$$\mathbf{A} \mathbf{A}^- \mathbf{A} = \mathbf{A}.$$

The matrix  $\mathbf{A}^-$  is not unique.

### F. Pseudo Inverse

1) *Definition:* The pseudo inverse (or Moore-Penrose inverse) of a matrix  $\mathbf{A}$  is the matrix  $\mathbf{A}^+$  that fulfills

$$\begin{aligned} I \quad & \mathbf{A} \mathbf{A}^+ \mathbf{A} = \mathbf{A} \\ II \quad & \mathbf{A}^+ \mathbf{A} \mathbf{A}^+ = \mathbf{A}^+ \\ III \quad & \mathbf{A} \mathbf{A}^+ \text{ symmetric} \\ IV \quad & \mathbf{A}^+ \mathbf{A} \text{ symmetric} \end{aligned}$$

The matrix  $\mathbf{A}^+$  is unique and does always exist. Note that in case of complex matrices, the symmetric condition is substituted by a condition of being Hermitian.

2) *Properties:* Assume  $\mathbf{A}^+$  to be the pseudo-inverse of  $\mathbf{A}$ , then

$$\begin{aligned} (\mathbf{A}^+)^+ &= \mathbf{A} \\ (\mathbf{A}^\top)^+ &= (\mathbf{A}^+)^{\top} \\ (\mathbf{A}^H)^+ &= (\mathbf{A}^+)^H \\ (\mathbf{A}^*)^+ &= (\mathbf{A}^+)^* \\ (\mathbf{A}^+ \mathbf{A}) \mathbf{A}^H &= \mathbf{A}^H \\ (\mathbf{A}^+ \mathbf{A}) \mathbf{A}^\top &\neq \mathbf{A}^\top \\ (c\mathbf{A})^+ &= (1/c) \mathbf{A}^+ \\ \mathbf{A}^+ &= (\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top \\ \mathbf{A} &= \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^+ \\ (\mathbf{A}^\top \mathbf{A})^+ &= \mathbf{A}^+ (\mathbf{A}^\top)^+ \\ (\mathbf{A} \mathbf{A}^\top)^+ &= (\mathbf{A}^\top)^+ \mathbf{A}^+ \\ \mathbf{A}^+ &= (\mathbf{A}^H \mathbf{A})^+ \mathbf{A}^H \\ \mathbf{A}^+ &= \mathbf{A}^H (\mathbf{A} \mathbf{A}^H)^+ \\ (\mathbf{A}^H \mathbf{A})^+ &= \mathbf{A}^+ (\mathbf{A}^H)^+ \end{aligned}$$

$$(\mathbf{A} \mathbf{A}^H)^+ = (\mathbf{A}^H)^+ \mathbf{A}^+$$

$$(\mathbf{A} \mathbf{B})^+ = (\mathbf{A}^+ \mathbf{A} \mathbf{B})^+ (\mathbf{A} \mathbf{B} \mathbf{B}^+)^+$$

$$f(\mathbf{A}^H \mathbf{A}) - f(0) \mathbf{I} = \mathbf{A}^+ [f(\mathbf{A} \mathbf{A}^H) - f(0) \mathbf{I}] \mathbf{A}$$

$$f(\mathbf{A} \mathbf{A}^H) - f(0) \mathbf{I} = \mathbf{A} [f(\mathbf{A}^H \mathbf{A}) - f(0) \mathbf{I}] \mathbf{A}$$

where  $\mathbf{A} \in \mathbb{C}^{m \times n}$ .

Assume  $\mathbf{A}$  to have full rank, then

$$(\mathbf{A} \mathbf{A}^+)(\mathbf{A} \mathbf{A}^+) = \mathbf{A} \mathbf{A}^+$$

$$(\mathbf{A}^+ \mathbf{A})(\mathbf{A}^+ \mathbf{A}) = \mathbf{A}^+ \mathbf{A}$$

$$\text{Tr}(\mathbf{A} \mathbf{A}^+) = \text{rank}(\mathbf{A} \mathbf{A}^+)$$

$$\text{Tr}(\mathbf{A}^+ \mathbf{A}) = \text{rank}(\mathbf{A}^+ \mathbf{A})$$

For two matrices it hold that

$$(\mathbf{A} \mathbf{B})^+ = (\mathbf{A}^+ \mathbf{A} \mathbf{B})^+ (\mathbf{A} \mathbf{B} \mathbf{B}^+)^+$$

$$(\mathbf{A} \otimes \mathbf{B})^+ = \mathbf{A}^+ \otimes \mathbf{B}^+$$

3) *Construction:* Assume that  $\mathbf{A}$  has full rank, then

$$\mathbf{A} \quad n \times n \quad \text{Square} \quad \text{rank}(\mathbf{A}) = n \implies \mathbf{A}^+ = \mathbf{A}^{-1}$$

$$\mathbf{A} \quad n \times m \quad \text{Broad} \quad \text{rank}(\mathbf{A}) = n \implies \mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^{-1}$$

$$\mathbf{A} \quad n \times m \quad \text{Tall} \quad \text{rank}(\mathbf{A}) = m \implies \mathbf{A}^+ = (\mathbf{A} \mathbf{A}^\top)^{-1} \mathbf{A}^\top$$

The so-called “broad version” is also known as *right inverse* and the “tall version” as the *left inverse*.

Assume  $\mathbf{A}$  does not have full rank, i.e.  $\mathbf{A}$  is  $n \times m$  and  $\text{rank}(\mathbf{A}) = r < \min(n, m)$ . The pseudo inverse  $\mathbf{A}^+$  can be constructed from the singular value decomposition  $\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{V}^\top$ , by

$$\mathbf{A}^+ = \mathbf{V}_r \mathbf{D}_r^{-1} \mathbf{U}_r^\top$$

where  $\mathbf{V}_r$ ,  $\mathbf{D}_r$ , and  $\mathbf{U}_r$  are the matrices with the degenerated rows and columns deleted. Alternatively consider there exist two matrices  $\mathbf{C} \in \mathbb{R}^{n \times r}$  and  $\mathbf{D} \in \mathbb{R}^{r \times m}$  of rank  $r$ , such that  $\mathbf{A} = \mathbf{C} \mathbf{D}$ . Using these matrices it holds that

$$\mathbf{A}^+ = \mathbf{D}^\top (\mathbf{D} \mathbf{D}^\top)^{-1} (\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{C}^\top$$

## VI. SOLUTIONS TO LINEAR EQUATIONS

### A. Existence in Linear Systems

Consider the linear system  $\mathbf{Ax} = \mathbf{b}$ , and the augmented matrix  $[\mathbf{A} \ \mathbf{b}]$ ,  $\mathbf{A} \in \mathbb{R}^{p \times q}$ ,  $\mathbf{b} \in \mathbb{R}^p$ ,  $\mathbf{x} \in \mathbb{R}^q$ , only one of the following is valid:

Condition	Solution
$\text{rank}([\mathbf{A} \ \mathbf{b}]) \geq \text{rank}(\mathbf{A})$	0 solutions exist
$\text{rank}([\mathbf{A} \ \mathbf{b}]) = \text{rank}(\mathbf{A}) = q$	1 solution exists
$\text{rank}([\mathbf{A} \ \mathbf{b}]) = \text{rank}(\mathbf{A}) < q$	$\infty$ solutions exist

### B. Standard Square

Assume  $\mathbf{A}$  is square and invertible, then

$$\mathbf{Ax} = \mathbf{b} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

### C. Degenerated Square

Assume  $\mathbf{A}$  is  $n \times n$  but of rank  $r < n$ . In that case, the system  $\mathbf{Ax} = \mathbf{b}$  is solved by

$$\mathbf{x} = \mathbf{A}^+\mathbf{b}$$

where  $\mathbf{A}^+$  is the pseudo-inverse of the rank-deficient matrix.

### D. Cramer's rule

The equation  $\mathbf{Ax} = \mathbf{b}$  where  $\mathbf{A}$  is square, has exactly one solution  $\mathbf{x}$  if the  $i^{th}$  element in  $\mathbf{x}$  can be found as

$$x_i = \frac{|\mathbf{B}|}{|\mathbf{A}|}$$

where  $\mathbf{B}$  equals  $\mathbf{A}$ , but the  $i^{th}$  column in  $\mathbf{A}$  has been substituted by  $\mathbf{b}$ .

### E. Over-determined Rectangular

Assume  $\mathbf{A}$  to be  $n \times m$ ,  $n > m$  (tall) and  $\text{rank}(\mathbf{A}) = m$ , then

$$\mathbf{Ax} = \mathbf{b} \implies \mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} = \mathbf{A}^+\mathbf{b}$$

that is if there exists a solution  $\mathbf{x}$  at all! If there is no solution, the following can be useful:

$$\mathbf{Ax} = \mathbf{b} \implies \mathbf{x}_{\min} = \mathbf{A}^+\mathbf{b}$$

Now  $\mathbf{x}_{\min}$  is the vector  $\mathbf{x}$  which minimizes  $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ . The matrix  $\mathbf{A}^+$  is the pseudo-inverse of  $\mathbf{A}$ .

### F. Under-determined Rectangular

Assume  $\mathbf{A}$  is  $n \times m$  and  $n < m$  (broad) and  $\text{rank}(\mathbf{A}) = n$ .

$$\mathbf{Ax} = \mathbf{b} \implies \mathbf{x}_{\min} = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^{-1} \mathbf{b}$$

The equation have many solutions  $\mathbf{x}$ . But  $\mathbf{x}_{\min}$  is the solution which minimizes  $\|\mathbf{Ax} - \mathbf{b}\|_2^2$  and also the solution with the smallest norm  $\|\mathbf{x}\|_2^2$ . The same holds for a matrix version: Assume  $\mathbf{A}$  is  $n \times m$ ,  $\mathbf{X}$  is  $m \times n$  and  $\mathbf{B}$  is  $n \times n$ , then

$$\mathbf{AX} = \mathbf{B} \implies \mathbf{X}_{\min} = \mathbf{A}^+\mathbf{B}$$

The equation may have many solutions  $\mathbf{X}$ . But  $\mathbf{X}_{\min}$  is the solution which minimizes  $\|\mathbf{AX} - \mathbf{B}\|_2^2$  and also the solution with the smallest norm  $\|\mathbf{X}\|_2^2$ .

Similar but different: Assume  $\mathbf{A}$  is square  $n \times n$  and the matrices  $\mathbf{B}_0$ ,  $\mathbf{B}_1$  are  $n \times N$ , where  $N > n$ , then if  $\mathbf{B}_0$  has maximal rank

$$\mathbf{AB}_0 = \mathbf{B}_1 \implies \mathbf{A}_{\min} = \mathbf{B}_1 \mathbf{B}_0^\top (\mathbf{B}_0 \mathbf{B}_0^\top)^{-1}$$

where  $\mathbf{A}_{\min}$  denotes the matrix which is optimal in a least square sense. An interpretation is that  $\mathbf{A}$  is the linear approximation which maps the columns vectors of  $\mathbf{B}_0$  into the columns vectors of  $\mathbf{B}_1$ .

### G. Linear form and zeros

$$\mathbf{Ax} = \mathbf{0}, \forall \mathbf{x} \implies \mathbf{A} = \mathbf{0}$$

### H. Square form and zeros

If  $\mathbf{A}$  is symmetric, then

$$\mathbf{x}^\top \mathbf{Ax} = \mathbf{0}, \forall \mathbf{x} \implies \mathbf{A} = \mathbf{0}$$