2 Derivatives

This section is covering 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 2.8 for differentiation of structured matrices. The basic assumptions can be written in a formula as

$$\frac{\partial X_{kl}}{\partial X_{ij}} = \delta_{ik}\delta_{lj} \tag{32}$$

that is for e.g. vector forms,

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

The following rules are general and very useful when deriving the differential of an expression ([19]):

$$\begin{aligned}
\partial\mathbf{A} &= 0 \qquad (\mathbf{A} \text{ is a constant}) && (33)\\
\partial(\alpha\mathbf{X}) &= \alpha\partial\mathbf{X} && (34)\\
\partial(\mathbf{X}+\mathbf{Y}) &= \partial\mathbf{X}+\partial\mathbf{Y} && (35)\\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial\mathbf{X}) && (36)\\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial\mathbf{X})\mathbf{Y}+\mathbf{X}(\partial\mathbf{Y}) && (37)\\
\partial(\mathbf{X}\circ\mathbf{Y}) &= (\partial\mathbf{X})\circ\mathbf{Y}+\mathbf{X}\circ(\partial\mathbf{Y}) && (38)\\
\partial(\mathbf{X}\otimes\mathbf{Y}) &= (\partial\mathbf{X})\otimes\mathbf{Y}+\mathbf{X}\otimes(\partial\mathbf{Y}) && (39)\\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} && (40)\\
\partial(\det(\mathbf{X})) &= \mathrm{Tr}(\mathrm{adj}(\mathbf{X})\partial\mathbf{X}) && (41)\\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (42)\\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (43)\\
\partial\mathbf{X}^T &= (\partial\mathbf{X})^T && (44)\\
\partial\mathbf{X}^H &= (\partial\mathbf{X})^H && (45)
\end{aligned}$$

2.1 Derivatives of a Determinant

2.1.1 General form

$$\frac{\partial\det(\mathbf{Y})}{\partial x} = \det(\mathbf{Y})\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right] \tag{46}$$

$$\sum_k \frac{\partial\det(\mathbf{X})}{\partial X_{ik}}X_{jk} = \delta_{ij}\det(\mathbf{X}) \tag{47}$$

$$\begin{aligned}
\frac{\partial^2\det(\mathbf{Y})}{\partial x^2} = \det(\mathbf{Y})\Bigg[&\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial\frac{\partial\mathbf{Y}}{\partial x}}{\partial x}\right]\\
&+\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right]\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right]\\
&-\mathrm{Tr}\left[\left(\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right)\left(\mathbf{Y}^{-1}\frac{\partial\mathbf{Y}}{\partial x}\right)\right]\Bigg]
\end{aligned} \tag{48}$$

2.1.2 Linear forms

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(\mathbf{X}^{-1})^T \tag{49}$$

$$\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} = \delta_{ij} \det(\mathbf{X}) \tag{50}$$

$$\frac{\partial \det(\mathbf{AXB})}{\partial \mathbf{X}} = \det(\mathbf{AXB})(\mathbf{X}^{-1})^T = \det(\mathbf{AXB})(\mathbf{X}^T)^{-1} \tag{51}$$

2.1.3 Square forms

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

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{AX})\mathbf{X}^{-T} \tag{52}$$

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

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{AX})\mathbf{AX}(\mathbf{X}^T\mathbf{AX})^{-1} \tag{53}$$

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

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = \det(\mathbf{X}^T\mathbf{AX})(\mathbf{AX}(\mathbf{X}^T\mathbf{AX})^{-1} + \mathbf{A}^T\mathbf{X}(\mathbf{X}^T\mathbf{A}^T\mathbf{X})^{-1}) \tag{54}$$

2.1.4 Other nonlinear forms

Some special cases are (See [9, 7])

$$\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})|}{\partial \mathbf{X}} = 2(\mathbf{X}^+)^T \tag{55}$$

$$\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})}{\partial \mathbf{X}^+} = -2\mathbf{X}^T \tag{56}$$

$$\frac{\partial \ln |\det(\mathbf{X})|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^T = (\mathbf{X}^T)^{-1} \tag{57}$$

$$\frac{\partial \det(\mathbf{X}^k)}{\partial \mathbf{X}} = k\det(\mathbf{X}^k)\mathbf{X}^{-T} \tag{58}$$

2.2 Derivatives of an Inverse

From [27] we have the basic identity

$$\frac{\partial \mathbf{Y}^{-1}}{\partial x} = -\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\mathbf{Y}^{-1} \tag{59}$$

from which it follows

$$\frac{\partial(\mathbf{X}^{-1})_{kl}}{\partial X_{ij}} = -(\mathbf{X}^{-1})_{ki}(\mathbf{X}^{-1})_{jl} \tag{60}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}}{\partial \mathbf{X}} = -\mathbf{X}^{-T}\mathbf{a}\mathbf{b}^T\mathbf{X}^{-T} \tag{61}$$

$$\frac{\partial \det(\mathbf{X}^{-1})}{\partial \mathbf{X}} = -\det(\mathbf{X}^{-1})(\mathbf{X}^{-1})^T \tag{62}$$

$$\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^T \tag{63}$$

$$\frac{\partial \mathrm{Tr}((\mathbf{X}+\mathbf{A})^{-1})}{\partial \mathbf{X}} = -((\mathbf{X}+\mathbf{A})^{-1}(\mathbf{X}+\mathbf{A})^{-1})^T \tag{64}$$

From [32] we 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

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

2.3 Derivatives of Eigenvalues

$$\frac{\partial}{\partial \mathbf{X}}\sum \mathrm{eig}(\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}) = \mathbf{I} \tag{65}$$

$$\frac{\partial}{\partial \mathbf{X}}\prod \mathrm{eig}(\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\det(\mathbf{X}) = \det(\mathbf{X})\mathbf{X}^{-T} \tag{66}$$

If $\mathbf{A}$ is real and symmetric, λ_i and $\mathbf{v}_i$ are distinct eigenvalues and eigenvectors of $\mathbf{A}$ (see (276)) with $\mathbf{v}_i^T\mathbf{v}_i = 1$, then [33]

$$\partial\lambda_i = \mathbf{v}_i^T\partial(\mathbf{A})\mathbf{v}_i \tag{67}$$

$$\partial\mathbf{v}_i = (\lambda_i\mathbf{I}-\mathbf{A})^+\partial(\mathbf{A})\mathbf{v}_i \tag{68}$$

2.4 Derivatives of Matrices, Vectors and Scalar Forms

2.4.1 First Order

$$\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \tag{69}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T \tag{70}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T \tag{71}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T \tag{72}$$

$$\frac{\partial \mathbf{X}}{\partial X_{ij}} = \mathbf{J}^{ij} \tag{73}$$

$$\frac{\partial(\mathbf{X}\mathbf{A})_{ij}}{\partial X_{mn}} = \delta_{im}(\mathbf{A})_{nj} = (\mathbf{J}^{mn}\mathbf{A})_{ij} \tag{74}$$

$$\frac{\partial(\mathbf{X}^T\mathbf{A})_{ij}}{\partial X_{mn}} = \delta_{in}(\mathbf{A})_{mj} = (\mathbf{J}^{nm}\mathbf{A})_{ij} \tag{75}$$

2.4.2 Second Order

$$\frac{\partial}{\partial X_{ij}} \sum_{klmn} X_{kl} X_{mn} = 2 \sum_{kl} X_{kl} \tag{76}$$

$$\frac{\partial \mathbf{b}^T \mathbf{X}^T \mathbf{X} \mathbf{c}}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{b}\mathbf{c}^T + \mathbf{c}\mathbf{b}^T) \tag{77}$$

$$\frac{\partial (\mathbf{B}\mathbf{x}+\mathbf{b})^T \mathbf{C} (\mathbf{D}\mathbf{x}+\mathbf{d})}{\partial \mathbf{x}} = \mathbf{B}^T \mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d}) + \mathbf{D}^T \mathbf{C}^T (\mathbf{B}\mathbf{x}+\mathbf{b}) \tag{78}$$

$$\frac{\partial (\mathbf{X}^T \mathbf{B} \mathbf{X})_{kl}}{\partial X_{ij}} = \delta_{lj} (\mathbf{X}^T \mathbf{B})_{ki} + \delta_{kj} (\mathbf{B}\mathbf{X})_{il} \tag{79}$$

$$\frac{\partial (\mathbf{X}^T \mathbf{B} \mathbf{X})}{\partial X_{ij}} = \mathbf{X}^T \mathbf{B} \mathbf{J}^{ij} + \mathbf{J}^{ji} \mathbf{B} \mathbf{X} \qquad (\mathbf{J}^{ij})_{kl} = \delta_{ik}\delta_{jl} \tag{80}$$

See Sec 9.7 for useful properties of the Single-entry matrix $\mathbf{J}^{ij}$

$$\frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{x} \tag{81}$$

$$\frac{\partial \mathbf{b}^T \mathbf{X}^T \mathbf{D} \mathbf{X} \mathbf{c}}{\partial \mathbf{X}} = \mathbf{D}^T \mathbf{X} \mathbf{b} \mathbf{c}^T + \mathbf{D} \mathbf{X} \mathbf{c} \mathbf{b}^T \tag{82}$$

$$\frac{\partial}{\partial \mathbf{X}} (\mathbf{X}\mathbf{b}+\mathbf{c})^T \mathbf{D} (\mathbf{X}\mathbf{b}+\mathbf{c}) = (\mathbf{D}+\mathbf{D}^T)(\mathbf{X}\mathbf{b}+\mathbf{c})\mathbf{b}^T \tag{83}$$

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

$$\frac{\partial}{\partial \mathbf{s}} (\mathbf{x} - \mathbf{A}\mathbf{s})^T \mathbf{W} (\mathbf{x} - \mathbf{A}\mathbf{s}) = -2\mathbf{A}^T \mathbf{W} (\mathbf{x} - \mathbf{A}\mathbf{s}) \tag{84}$$

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x} - \mathbf{s})^T \mathbf{W} (\mathbf{x} - \mathbf{s}) = 2\mathbf{W}(\mathbf{x} - \mathbf{s}) \tag{85}$$

$$\frac{\partial}{\partial \mathbf{s}} (\mathbf{x} - \mathbf{s})^T \mathbf{W} (\mathbf{x} - \mathbf{s}) = -2\mathbf{W}(\mathbf{x} - \mathbf{s}) \tag{86}$$

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x} - \mathbf{A}\mathbf{s})^T \mathbf{W} (\mathbf{x} - \mathbf{A}\mathbf{s}) = 2\mathbf{W}(\mathbf{x} - \mathbf{A}\mathbf{s}) \tag{87}$$

$$\frac{\partial}{\partial \mathbf{A}} (\mathbf{x} - \mathbf{A}\mathbf{s})^T \mathbf{W} (\mathbf{x} - \mathbf{A}\mathbf{s}) = -2\mathbf{W}(\mathbf{x} - \mathbf{A}\mathbf{s})\mathbf{s}^T \tag{88}$$

As a case with complex values the following holds

$$\frac{\partial (a - \mathbf{x}^H \mathbf{b})^2}{\partial \mathbf{x}} = -2\mathbf{b}(a - \mathbf{x}^H \mathbf{b})^* \tag{89}$$

This formula is also known from the LMS algorithm [14]

2.4.3 Higher-order and non-linear

$$\frac{\partial (\mathbf{X}^n)_{kl}}{\partial X_{ij}} = \sum_{r=0}^{n-1} (\mathbf{X}^r \mathbf{J}^{ij} \mathbf{X}^{n-1-r})_{kl} \tag{90}$$

For proof of the above, see B.1.3.

$$\frac{\partial}{\partial \mathbf{X}} \mathbf{a}^T \mathbf{X}^n \mathbf{b} = \sum_{r=0}^{n-1} (\mathbf{X}^r)^T \mathbf{a} \mathbf{b}^T (\mathbf{X}^{n-1-r})^T \tag{91}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathbf{a}^T(\mathbf{X}^n)^T\mathbf{X}^n\mathbf{b} = \sum_{r=0}^{n-1}\Big[\mathbf{X}^{n-1-r}\mathbf{a}\mathbf{b}^T(\mathbf{X}^n)^T\mathbf{X}^r + (\mathbf{X}^r)^T\mathbf{X}^n\mathbf{a}\mathbf{b}^T(\mathbf{X}^{n-1-r})^T\Big] \quad (92)$$

See B.1.3 for a proof.
Assume $\mathbf{s}$ and $\mathbf{r}$ are functions of $\mathbf{x}$, i.e. $\mathbf{s} = \mathbf{s}(\mathbf{x}), \mathbf{r} = \mathbf{r}(\mathbf{x})$, and that $\mathbf{A}$ is a constant, then

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{s}^T\mathbf{A}\mathbf{r} = \left[\frac{\partial \mathbf{s}}{\partial \mathbf{x}}\right]^T\mathbf{A}\mathbf{r} + \left[\frac{\partial \mathbf{r}}{\partial \mathbf{x}}\right]^T\mathbf{A}^T\mathbf{s} \quad (93)$$

$$\frac{\partial}{\partial \mathbf{x}}\frac{(\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x})}{(\mathbf{B}\mathbf{x})^T(\mathbf{B}\mathbf{x})} = \frac{\partial}{\partial \mathbf{x}}\frac{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x}} \quad (94)$$

$$= 2\frac{\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{B}\mathbf{x}} - 2\frac{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}\mathbf{B}^T\mathbf{B}\mathbf{x}}{(\mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x})^2} \quad (95)$$

2.4.4 Gradient and Hessian

Using the above we have for the gradient and the Hessian

$$f = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} \quad (96)$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} + \mathbf{b} \quad (97)$$

$$\frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}^T} = \mathbf{A} + \mathbf{A}^T \quad (98)$$

2.5 Derivatives of Traces

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

$$\frac{\partial \mathrm{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T$$

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

2.5.1 First Order

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}) = \mathbf{I} \quad (99)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{A}) = \mathbf{A}^T \quad (100)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}) = \mathbf{A}^T\mathbf{B}^T \quad (101)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^T\mathbf{B}) = \mathbf{B}\mathbf{A} \quad (102)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{A}) = \mathbf{A} \quad (103)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^T) = \mathbf{A} \quad (104)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\otimes\mathbf{X}) = \mathrm{Tr}(\mathbf{A})\mathbf{I} \quad (105)$$

2.5.2 Second Order

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2) = 2\mathbf{X}^T \quad (106)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2\mathbf{B}) = (\mathbf{XB}+\mathbf{BX})^T \quad (107)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{BX}) = \mathbf{BX}+\mathbf{B}^T\mathbf{X} \quad (108)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{BXX}^T) = \mathbf{BX}+\mathbf{B}^T\mathbf{X} \quad (109)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XX}^T\mathbf{B}) = \mathbf{BX}+\mathbf{B}^T\mathbf{X} \quad (110)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XBX}^T) = \mathbf{XB}^T+\mathbf{XB} \quad (111)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{BX}^T\mathbf{X}) = \mathbf{XB}^T+\mathbf{XB} \quad (112)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{XB}) = \mathbf{XB}^T+\mathbf{XB} \quad (113)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXBX}) = \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T+\mathbf{B}^T\mathbf{X}^T\mathbf{A}^T \quad (114)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XX}^T) = 2\mathbf{X} \quad (115)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{B}^T\mathbf{X}^T\mathbf{CXB}) = \mathbf{C}^T\mathbf{XBB}^T+\mathbf{CXBB}^T \quad (116)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{X}^T\mathbf{BXC}\right] = \mathbf{BXC}+\mathbf{B}^T\mathbf{XC}^T \quad (117)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXBX}^T\mathbf{C}) = \mathbf{A}^T\mathbf{C}^T\mathbf{XB}^T+\mathbf{CAXB} \quad (118)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\Big[(\mathbf{AXB}+\mathbf{C})(\mathbf{AXB}+\mathbf{C})^T\Big] = 2\mathbf{A}^T(\mathbf{AXB}+\mathbf{C})\mathbf{B}^T \quad (119)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\otimes\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X})\mathrm{Tr}(\mathbf{X}) = 2\mathrm{Tr}(\mathbf{X})\mathbf{I} \quad (120)$$

See [7].

2.5.3 Higher Order

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^k) = k(\mathbf{X}^{k-1})^T \quad (121)$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^k) = \sum_{r=0}^{k-1}(\mathbf{X}^r\mathbf{AX}^{k-r-1})^T \quad (122)$$

$$\begin{aligned}\tfrac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{B}^T\mathbf{X}^T\mathbf{CXX}^T\mathbf{CXB}\right] &= \mathbf{CXX}^T\mathbf{CXBB}^T \\ &\quad +\mathbf{C}^T\mathbf{XBB}^T\mathbf{X}^T\mathbf{C}^T\mathbf{X} \\ &\quad +\mathbf{CXBB}^T\mathbf{X}^T\mathbf{CX} \\ &\quad +\mathbf{C}^T\mathbf{XX}^T\mathbf{C}^T\mathbf{XBB}^T \end{aligned} \quad (123)$$

2.5.4 Other

$$\frac{\partial}{\partial \mathbf{X}}\text{Tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B}) = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^T = -\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T} \tag{124}$$

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

$$\frac{\partial}{\partial \mathbf{X}}\text{Tr}\Big[(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{A}\Big] = -(\mathbf{C}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1})(\mathbf{A}+\mathbf{A}^T)(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \tag{125}$$

$$\frac{\partial}{\partial \mathbf{X}}\text{Tr}\Big[(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{X})\Big] = -2\mathbf{C}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} +2\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \tag{126}$$

$$\frac{\partial}{\partial \mathbf{X}}\text{Tr}\Big[(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{X})\Big] = -2\mathbf{C}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{B}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} +2\mathbf{B}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \tag{127}$$

See [7].

$$\frac{\partial \text{Tr}(\sin(\mathbf{X}))}{\partial \mathbf{X}} = \cos(\mathbf{X})^T \tag{128}$$

2.6 Derivatives of vector norms

2.6.1 Two-norm

$$\frac{\partial}{\partial \mathbf{x}}||\mathbf{x}-\mathbf{a}||_2 = \frac{\mathbf{x}-\mathbf{a}}{||\mathbf{x}-\mathbf{a}||_2} \tag{129}$$

$$\frac{\partial}{\partial \mathbf{x}}\frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|_2} = \frac{\mathbf{I}}{\|\mathbf{x}-\mathbf{a}\|_2} - \frac{(\mathbf{x}-\mathbf{a})(\mathbf{x}-\mathbf{a})^T}{\|\mathbf{x}-\mathbf{a}\|_2^3} \tag{130}$$

$$\frac{\partial||\mathbf{x}||_2^2}{\partial \mathbf{x}} = \frac{\partial||\mathbf{x}^T\mathbf{x}||_2}{\partial \mathbf{x}} = 2\mathbf{x} \tag{131}$$

2.7 Derivatives of matrix norms

For more on matrix norms, see Sec. 10.4.

2.7.1 Frobenius norm

$$\frac{\partial}{\partial \mathbf{X}}||\mathbf{X}||_\text{F}^2 = \frac{\partial}{\partial \mathbf{X}}\text{Tr}(\mathbf{X}\mathbf{X}^H) = 2\mathbf{X} \tag{132}$$

See (248). Note that this is also a special case of the result in equation 119.

2.8 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 $f(\mathbf{A})$

$$\frac{df}{dA_{ij}} = \sum_{kl}\frac{\partial f}{\partial A_{kl}}\frac{\partial A_{kl}}{\partial A_{ij}} = \text{Tr}\left[\left[\frac{\partial f}{\partial \mathbf{A}}\right]^T\frac{\partial \mathbf{A}}{\partial A_{ij}}\right] \tag{133}$$

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

$$\frac{\partial \mathbf{A}}{\partial A_{ij}} = \mathbf{S}^{ij} \tag{134}$$

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 singleentry matrices, see Sec. 9.7.6 for more examples of structure matrices.

2.8.1 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 $\mathbf{X}$:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}} \tag{135}$$

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

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^{M}\sum_{l=1}^{N} \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}} \tag{136}$$

Using matrix notation, this can be written as:

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = \mathrm{Tr}\Big[\big(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\big)^T \frac{\partial \mathbf{U}}{\partial X_{ij}} \Big]. \tag{137}$$

2.8.2 Symmetric

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

$$\frac{df}{d\mathbf{A}} = \left[\frac{\partial f}{\partial \mathbf{A}}\right] + \left[\frac{\partial f}{\partial \mathbf{A}}\right]^T - \mathrm{diag}\left[\frac{\partial f}{\partial \mathbf{A}}\right] \tag{138}$$

That is, e.g., ([5]):

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}^T - (\mathbf{A} \circ \mathbf{I}), \text{ see } (142) \tag{139}$$

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I})) \tag{140}$$

$$\frac{\partial \ln\det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I}) \tag{141}$$

2.8.3 Diagonal

If $\mathbf{X}$ is diagonal, then ([19]):

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} \circ \mathbf{I} \tag{142}$$

2.8.4 Toeplitz

Like symmetric matrices and diagonal matrices also Toeplitz matrices has a special structure which should be taken into account when the derivative with respect to a matrix with Toeplitz structure.

$$
\begin{aligned}
& \frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} \\
= \quad & \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} \\
= \quad & \begin{bmatrix}
\mathrm{Tr}(\mathbf{A}) & \mathrm{Tr}([\mathbf{A}^T]_{n1}) & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) & \cdots & A_{n1} \\
\mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A}) & \ddots & \ddots & \vdots \\
\mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \ddots & \ddots & \ddots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) \\
\vdots & \ddots & \ddots & \ddots & \mathrm{Tr}([\mathbf{A}^T]_{n1}) \\
A_{1n} & \cdots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A})
\end{bmatrix} \\
\equiv \quad & \boldsymbol{\alpha}(\mathbf{A})
\end{aligned}
\tag{143}
$$

As it can be seen, the derivative $\boldsymbol{\alpha}(\mathbf{A})$ 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 $\mathbf{A}^T$. This result is only valid for the unconstrained Toeplitz matrix. If the Toeplitz matrix also is symmetric, the same derivative yields

$$
\frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} = \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} = \boldsymbol{\alpha}(\mathbf{A}) + \boldsymbol{\alpha}(\mathbf{A})^T - \boldsymbol{\alpha}(\mathbf{A}) \circ \mathbf{I} \tag{144}
$$