# Function differentials simplify the calculation of derivatives with respect to matrices

Joaquín Rapela*

January 1, 2022

**Abstract**

The derivative of a matrix functions (e.g., $F(X) : \mathbb{R}^{m\times p} \rightarrow \mathbb{R}^{n\times q}$) is a four-dimensional tensor (e.g., $\frac{\partial F(X);}{\partial X} \in n\times q\times m\times p$). Here we present key concepts from the differential calculus of matrix functions that allow to derive derivatives of matrix functions by only manipulating matrices, and not four-dimensional tensors. We illustrate the use of these concepts in didactic derivations of derivatives of two matrix functions, required to compute the derivative of the log-likelihood function of the probabilistic principal component analysis model.

## List of definitions, theorems, claims and subclaims



## Contents



*j.rapela@ucl.ac.uk

# 1 Introduction

The derivation of optimal statistical signal processing algorithms frequently requires the calculation of derivatives of matrix functions (e.g.; $F(X) : \mathbb{R}^{m\times p} \rightarrow \mathbb{R}^{n\times q}$) with respect to to their inputs. For example, these derivatives are used in the derivation of the expectation maximization algorithm for linear dynamical systems (Shumway et al., 2016, Chapter 6).

In general, matrix derivatives are four dimensional tensors (e.g., $\frac{\partial F(X);}{\partial X} \in n \times q \times m \times p$; Dattorro, 2015, Appendix D), which complicates their manipulation. Here we present key results from the calculus of differentials of matrix functions (Section 2; see also Magnus and Neudecker, 2019), which allow to derive derivatives of matrix functions by only manipulating matrices, and not tensors.

We then use this calculus to didactically derive two derivatives of matrix functions required to compute the derivative of the log-likelihood function of the probabilistic PCA model. This log-likelihood function and its derivative are given in Eq. 1 and 6, respectively. and the two derivatives of matrix functions required to computer Eq 6 are derived in Claims 2 and 3.

$$\ell = -\frac{N}{2}\log|2\pi C| - \frac{N}{2}\operatorname{Tr}\left[C^{-1}S\right] \tag{1}$$

$$\text{where } C = \Lambda\Lambda^\intercal + \psi I,\ \Lambda \in \mathbb{R}^{m\times p} \text{ and } S \in \mathbb{R}^{m\times m}$$

$$\frac{\partial \ell}{\partial \Lambda} = \frac{N}{2}\left(-\frac{\partial}{\partial \Lambda}\log|2\pi C| - \frac{\partial}{\partial \Lambda}\operatorname{Tr}\left[C^{-1}S\right]\right) \tag{2}$$

$$= \frac{N}{2}\left(-\frac{\partial}{\partial \Lambda}\log\left[(2\pi)^m|C|\right] - \frac{\partial}{\partial \Lambda}\operatorname{Tr}\left[C^{-1}S\right]\right) \tag{3}$$

$$= \frac{N}{2}\left(-\frac{\partial}{\partial \Lambda}\left[m\log 2\pi + \log|C|\right] - \frac{\partial}{\partial \Lambda}\operatorname{Tr}\left[C^{-1}S\right]\right) \tag{4}$$

$$= \frac{N}{2}\left(-\frac{\partial}{\partial \Lambda}\log|C| - \frac{\partial}{\partial \Lambda}\operatorname{Tr}\left[C^{-1}S\right]\right) \tag{5}$$

$$= N\left(-C^{-1}\Lambda + C^{-1}SC^{-1}\Lambda\right) \tag{6}$$

Notes:

- Eq. 2 holds due to the linearity of the derivative,
- Eq. 3 follows from the fact that if $A \in \mathbb{R}^{m\times m}$ then $|kA| = k^m|A|$,
- To obtain Eq. 4 I used properties of the logarithm,
- Eq. 5 is valid because $m\log(2\pi)$ is a constant with respect to $\Lambda$,
- To calculate Eq. 6 I used Claim 2 and Claim 3.

# 2 Matrix differentials

Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$, then the derivative of $\phi$ at the point $c$ can be defined using limits, as in Eq. 7, or using the Taylor expansion of $\phi$ at the point $c + u$, as in Eq. 8. Claim 1 proves that these definitions are equivalent. The linear term of the Taylor expansion definition is the differential of $f$ at point $c$ with increment $u$, as stated in Definition 1.

$$\lim_{u\rightarrow 0}\frac{\phi(c+u) - \phi(c)}{u} \triangleq \phi'(c) \tag{7}$$

$$\phi(c+u) = \phi(c) + \phi'(c)u + r_c(u) \tag{8}$$
$$\text{with } \lim_{u\to 0} \frac{r_c(u)}{u} = 0$$

**Claim 1** (equivalence of two definitions of the derivative). *A function $\phi : \mathbb{R} \to \mathbb{R}$ satisfies Eq. 7 if and only if it satisfies Eq. 8.*

*Proof.* $\to$:

$$\begin{aligned}\phi(c+u) &= \phi(c) + \phi'(c)u + \big(\phi(c+u) - \phi(c) - \phi'(c)u\big)\\ &= \phi(c) + \phi'(c)u + r_c(u)\end{aligned}$$

$$\text{where } r_c(u) = \phi(c+u) - \phi(c) - \phi'(c)u$$
$$\text{and } \lim_{u\to 0} \frac{r_c(u)}{u} = \lim_{u\to 0} \frac{\phi(c+u) - \phi(c)}{u} - \phi'(c) = \phi'(c) - \phi'(c) = 0 \tag{9}$$

$\leftarrow$: From Eq. 8:

$$\frac{\phi(c+u) - \phi(c)}{u} = \phi'(c) + \frac{r_c(u)}{u}$$

Then

$$\lim_{u\to 0} \frac{\phi(c+u) - \phi(c)}{u} = \phi'(c) + \lim_{u\to 0} \frac{r_c(u)}{u} = \phi'(c) + 0 = \phi'(c)$$

□

**Definition 1** (differential of a one-dimensional function). *Let $\phi : \mathbb{R} \to \mathbb{R}$, the* ***differential*** *of $\phi$ at point $c$ with increment $u$ is $\phi'(c)u$.*

Definition 2 extends the concepts of Taylor expansion, differentiability, differential and derivative to functions of the type $F : \mathbb{R}^{n\times q} \to \mathbb{R}^{m\times p}$.

**Definition 2** (differential of a matrix function). *Let $F : S \subset \mathbb{R}^{n\times q} \to \mathbb{R}^{m\times p}$, let $C$ be an interior point of $S$ (i.e.; $\exists\, r \in \mathbb{R}^+$ such that $B(C, r) \subset S$), and let $r \in \mathbb{R}^+$ such that $B(C, r) \subset S$. If there exist a real $mp \times nq$ matrix $A(C)$ such that for all $U \in \mathbb{R}^{n\times q}$ with $|U| < r$ it follows that*

$$\operatorname{vec} F(C+U) = \operatorname{vec} F(C) + A(C) \operatorname{vec} U + \operatorname{vec} R_C(U)$$

*with $\lim_{U\to 0} \frac{R_C(U)}{||U||} = 0, ||U|| = (\operatorname{Tr} U^\intercal U)^{\frac{1}{2}}$, then the function $F$ is said to be* ***differentiable*** *at $C$. The $m \times p$ matrix $dF(C; U)$ defined as*

$$\operatorname{vec} dF(C; U) = A(C) \operatorname{vec} U$$

*is called the* ***differential*** *of $F$ at $C$ with increment $U$, and the matrix $A(C)$ is called the* ***first derivative*** *of $F$ at $C$.*

We next define a few concepts necessary to prove the chain rule for differentials of matrix functions in Theorem 3.

**Definition 3** (vector function corresponding to a matrix function). *Let $F : S \subset \mathbb{R}^{n\times q} \to \mathbb{R}^{m\times p}$, the corresponding vector function $f : \operatorname{vec} S \subset \mathbb{R}^{nq} \to \mathbb{R}^{mp}$ is*

$$f(\operatorname{vec} X) = \operatorname{vec} F(X) \tag{10}$$

**Definition 4** (Jacobian matrix). *Let $F : S \subset \mathbb{R}^{n\times q} \to \mathbb{R}^{m\times p}$. The **Jacobian matrix** of $F$ at $C$ is an $mp \times nq$ matrix whose $ij$th element is the partial derivative of the $i$th component of $\operatorname{vec} F(X)$ with respect to the $j$th component of $\operatorname{vec} X$, evaluated at $X = C$, that is,*

$$DF(C) = Df(\operatorname{vec} C) \tag{11}$$

**Theorem 1** (first identification theorem for matrix functions). *Let $F : S \subset \mathbb{R}^{n\times q} \to \mathbb{R}^{m\times p}$ be a matrix function differentiable at an interior point $C$ of $S$. Then*

$$A(C) = DF(C)$$

*where $A(C)$ and $DF(C)$ are given in Definitions 2 and 4, respectively.*

*Proof.* Because $F$ is differentiable at $C$, for all $1 \leq i \leq m$ and $1 \leq j \leq p$, Eq. 12 holds.

$$\begin{aligned}
\operatorname{vec} F(C + tE_{ij}) &= \operatorname{vec} F(C) + A(C) \operatorname{vec} tE_{ij} + \operatorname{vec} R_C(tE_{ij}) && (12)\\
\operatorname{vec} F(C + tE_{ij}) - \operatorname{vec} F(C) &= A(C) \operatorname{vec} tE_{ij} + \operatorname{vec} R_C(tE_{ij}) \\
\operatorname{vec} F(C + tE_{ij}) - \operatorname{vec} F(C) &= A(C)t \operatorname{vec} E_{ij} + \operatorname{vec} R_C(tE_{ij}) && (13)\\
\lim_{t\to 0} \frac{\operatorname{vec} F(C + tE_{ij}) - \operatorname{vec} F(C)}{t} &= \lim_{t\to 0} \frac{A(C)\ t \operatorname{vec} E_{ij} + \operatorname{vec} R_C(tE_{ij})}{t} && (14)\\
DF(C) \operatorname{vec} E_{ij} &= A(C) \operatorname{vec} E_{ij} + \lim_{t\to 0} \frac{\operatorname{vec} R_C(tE_{ij})}{||tE_{ij}||} && (15)\\
DF(C) \operatorname{vec} E_{ij} &= A(C) \operatorname{vec} E_{ij} && (16)\\
DF(C) &= A(C) && (17)
\end{aligned}$$

Notes:

- Eq. 13 holds due to the linearity of the vec function.
- the left-hand side of Eq. 15 equals that of Eq. 14 by Definition 4.
- the second term in the right-hand side of Eq. 15 equals that of in Eq. 14 because $t = ||tE_{ij}||$.
- Eq. 16 follows from the fact that $\lim_{t\to 0} \frac{\operatorname{vec} R_C(tE_{ij})}{||tE_{ij}||} = \lim_{tE_{ij}\to 0} \frac{\operatorname{vec} R_C(tE_{ij})}{||tE_{ij}||} = 0$, by Definition 2.
- Eq. 17 is true because Eq. 16 holds for all $1 \leq i \leq m$ and $1 \leq j \leq p$.

□

**Theorem 2** (chain rule for Jacobians of matrix functions). *Assume that $F : S \subset \mathbb{R}^{n\times q} \to \mathbb{R}^{m\times q}$ is differentiable at an interior point $C$ of $S$. Let $T \subset \mathbb{R}^{m\times q}$ such that $F(X) \in T$, $\forall X \in S$ and assume that $G : T \to \mathbb{R}^{x\times s}$ is differentiable at the interior point $B = F(C)$ of $T$. Then the composite function $H : S \to \mathbb{R}^{r\times s}$ defined by*

$$H(X) = G(F(X)) \tag{18}$$

*and denoted by* $H = G \circ F$ *is differentiable at* $C$*, and*

$$DH(C) = DG(F(C))\ DF(C) \tag{19}$$

*Proof.*

$$\begin{aligned}
DH(C) &= DG(F(C)) \\
&= Dg(\operatorname{vec} F(C)) && (20)\\
&= Dg(f(\operatorname{vec} C)) && (21)\\
&= D(g \circ f)(\operatorname{vec} C) && (22)\\
&= Dg(f(\operatorname{vec} C))Df(\operatorname{vec} C) && (23)\\
&= Dg(\operatorname{vec} F(C))Df(\operatorname{vec} C) && (24)\\
&= DG(F(C))DF(C) && (25)
\end{aligned}$$

Notes:

- Eq. 20 follows the definition of the Jacobian matrix (Definition 4), applied to $G$.
- Eq. 21 uses the definition of the vector function corresponding to a matrix function (Eq. 10), applied to $F$.
- Eq. 22 uses the definition of composite function.
- Eq. 23 employs the chain rule for Jacobians of vector functions (Marsden and Tromba, 2003, Theorem 11: Chain Rule).
- Eq. 24 applies the definition of the vector function corresponding to a matrix function (Eq. 10) to $f$.
- Eq. 25 uses the definition of the Jacobian matrix (Eq. 4) twice, for $Dg$ and $Df$.

□

**Theorem 3** (chain rule for differentials of matrix functions)**.** *If* $F$ *is differentiable at* $C$ *and* $G$ *is differentiable at* $F(C)$*, then the differential of the composite function* $H = G \circ F$ *is*

$$dH(C; U) = dG(F(C); dF(C; U)) \tag{26}$$

*for every* $U \in \mathbb{R}^{n \times q}$*.*

*Proof.*

$$\begin{aligned}
\operatorname{vec} dH(C : U) = \operatorname{vec} d(G \circ F)(C; U) &= D(G \circ F)(C) \operatorname{vec} U && (27)\\
&= DG(F(C))DF(C) \operatorname{vec} U && (28)\\
&= DG(F(C)) \operatorname{vec} dF(C; U) && (29)\\
&= \operatorname{vec} dG(F(C); dF(C; U)) && (30)
\end{aligned}$$

□

Notes:

- Eq. 27 follows from the definition of differential in Definition 2 and Theorem 1 applied to $d(G \circ F)(C;U)$,
- Eq. 28 follows from Theorem 2,
- Eq. 29 follows from the definition of differential in Definition 2 and Theorem 1 applied to $dF(C;U)$,
- Eq. 30 follows from the definition of differential in Definition 2 and Theorem 1 applied to $dG(F(C);dF(C;U))$.

**Theorem 4** (product rule for differentials)**.**

$$d(FG)(C;U) = dF(C;U)G(C) + F(C)dG(C;U) \tag{31}$$

*Proof.* I will sketch the proof using $F, G : \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$. Let:

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$
$$U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$$
$$F(C) = \begin{bmatrix} F_{11}(C) & F_{12}(C) \\ F_{21}(C) & F_{22}(C) \end{bmatrix}$$
$$G(C) = \begin{bmatrix} G_{11}(C) & G_{12}(C) \\ G_{21}(C) & G_{22}(C) \end{bmatrix}$$

Let's first calculate the entry $[1,1]$ of the left-hand side of Eq. 31. From Definition 2 and Theorem 1 we have

$$\operatorname{vec} d(FG)(C;U) = D(FG)(C) \operatorname{vec}(U) \tag{32}$$

and from Definition 4 we have

$$[D(FG)(C)]_{ij} = \frac{\partial(\operatorname{vec}(FG(C)))_i}{\partial(\operatorname{vec} C)_j} \tag{33}$$

To compute $D(FG)(C)$ let's first write $(FG)(C) = F(C)G(C)$

$$(FG)(C) = \begin{bmatrix} F_{11}(C)G_{11}(C) + F_{12}(C)G_{21}(C) & F_{11}(C)G_{12}(C) + F_{12}(C)G_{22}(C) \\ F_{21}(C)G_{11}(C) + F_{22}(C)G_{21}(C) & F_{21}(C)G_{12}(C) + F_{22}(C)G_{22}(C) \end{bmatrix}$$

From Eq. 33, the first row of $D(FG)(C)$ is

$$\begin{aligned}
D(FG)(C)[1,:] &= [\ \frac{\partial(\operatorname{vec}(FG)(C))_1}{\partial(\operatorname{vec} C)_1}, \frac{\partial(\operatorname{vec}(FG)(C))_1}{\partial(\operatorname{vec} C)_2}, \frac{\partial(\operatorname{vec}(FG)(C))_1}{\partial(\operatorname{vec} C)_3}, \frac{\partial(\operatorname{vec}(FG)(C))_1}{\partial(\operatorname{vec} C)_4}\ ] \\
&= [\ \frac{\partial(FG)_{11}(C)}{\partial C_{11}}, \frac{\partial(FG)_{11}(C)}{\partial C_{21}}, \frac{\partial(FG)_{11}(C)}{\partial C_{12}}, \frac{\partial(FG)_{11}(C)}{\partial C_{22}}\ ] \\
&= [\ \frac{\partial F_{11}(C)}{\partial C_{11}} G_{11}(C) + F_{11}(C)\frac{\partial G_{11}(C)}{\partial C_{11}} + \frac{\partial F_{12}(C)}{\partial C_{11}} G_{21}(C) + F_{12}(C)\frac{\partial G_{21}(C)}{\partial C_{11}}, \\
&\quad \frac{\partial F_{11}(C)}{\partial C_{21}} G_{11}(C) + F_{11}(C)\frac{\partial G_{11}(C)}{\partial C_{21}} + \frac{\partial F_{12}(C)}{\partial C_{21}} G_{21}(C) + F_{12}(C)\frac{\partial G_{21}(C)}{\partial C_{21}}, \\
&\quad \frac{\partial F_{11}(C)}{\partial C_{12}} G_{11}(C) + F_{11}(C)\frac{\partial G_{11}(C)}{\partial C_{12}} + \frac{\partial F_{12}(C)}{\partial C_{12}} G_{21}(C) + F_{12}(C)\frac{\partial G_{21}(C)}{\partial C_{12}}, \\
&\quad \frac{\partial F_{11}(C)}{\partial C_{22}} G_{11}(C) + F_{11}(C)\frac{\partial G_{11}(C)}{\partial C_{22}} + \frac{\partial F_{12}(C)}{\partial C_{22}} G_{21}(C) + F_{12}(C)\frac{\partial G_{21}(C)}{\partial C_{22}}\ ]
\end{aligned}$$

From Eq. 32, the $[1,1]$ entry of $d(FG)(C;U)$ is:

$$\begin{aligned}
dFG(C;U)[1,1] =& D(FG)(C)[1,:]\operatorname{vec}(U) \\
=& \frac{\partial F_{11}(C)}{\partial C_{11}}U_{11}G_{11}(C) + F_{11}(C)\frac{\partial G_{11}(C)}{\partial C_{11}}U_{11} + \frac{\partial F_{12}(C)}{\partial C_{11}}U_{11}G_{21}(C) + F_{12}(C)\frac{\partial G_{21}(C)}{\partial C_{11}}U_{11}+ \\
& \frac{\partial F_{11}(C)}{\partial C_{21}}U_{21}G_{11}(C) + F_{11}(C)\frac{\partial G_{11}(C)}{\partial C_{21}}U_{21} + \frac{\partial F_{12}(C)}{\partial C_{21}}U_{21}G_{21}(C) + F_{12}(C)\frac{\partial G_{21}(C)}{\partial C_{21}}U_{21}+ \\
& \frac{\partial F_{11}(C)}{\partial C_{12}}U_{12}G_{11}(C) + F_{11}(C)\frac{\partial G_{11}(C)}{\partial C_{12}}U_{12} + \frac{\partial F_{12}(C)}{\partial C_{12}}U_{12}G_{21}(C) + F_{12}(C)\frac{\partial G_{21}(C)}{\partial C_{12}}U_{12}+ \\
& \frac{\partial F_{11}(C)}{\partial C_{22}}U_{22}G_{11}(C) + F_{11}(C)\frac{\partial G_{11}(C)}{\partial C_{22}}U_{22} + \frac{\partial F_{12}(C)}{\partial C_{22}}U_{22}G_{21}(C) + F_{12}(C)\frac{\partial G_{21}(C)}{\partial C_{22}}U_{22}
\end{aligned} \tag{34}$$

Having calculated the left-hand side of Eq. 31 at index $[1,1]$ in Eq. 34, let's now calculate the right-hand side or Eq. 31 at index $[1,1]$ and check if it coincides with the right-hand side of Eq. 34. From Definition 4

$$DF(C)[1,:] = \left[\frac{\partial F_{11}(C)}{\partial C_{11}}, \frac{\partial F_{11}(C)}{\partial C_{21}}, \frac{\partial F_{11}(C)}{\partial C_{12}}, \frac{\partial F_{11}(C)}{\partial C_{22}}\right] \tag{35}$$

$$DF(C)[3,:] = \left[\frac{\partial F_{12}(C)}{\partial C_{11}}, \frac{\partial F_{12}(C)}{\partial C_{21}}, \frac{\partial F_{12}(C)}{\partial C_{12}}, \frac{\partial F_{12}(C)}{\partial C_{22}}\right] \tag{36}$$

From Definition 2 and Theorem 1

$$\begin{aligned}
dF(C;U)[1,1] &= DF(C)[1,:]\operatorname{vec}U \\
&= \frac{\partial F_{11}(C)}{\partial C_{11}}U_{11} + \frac{\partial F_{11}(C)}{\partial C_{21}}U_{21} + \frac{\partial F_{11}(C)}{\partial C_{12}}U_{12} + \frac{\partial F_{11}(C)}{\partial C_{22}}U_{22} \\
dF(C;U)[1,2] &= DF(C)[3,:]\operatorname{vec}U \\
&= \frac{\partial F_{12}(C)}{\partial C_{11}}U_{11} + \frac{\partial F_{12}(C)}{\partial C_{21}}U_{21} + \frac{\partial F_{12}(C)}{\partial C_{12}}U_{12} + \frac{\partial F_{12}(C)}{\partial C_{22}}U_{22}
\end{aligned}$$

Now the first term in the right-hand side of Eq. 31 at index $[1,1]$ is

$$\begin{aligned}
(dF(C;U)G(C))[1,1] =& dF(C;U)[1,1]G_{11}(C) + dF(C;U)[1,2]G_{21}(C) \\
=& \frac{\partial F_{11}(C)}{\partial C_{11}}U_{11}G_{11}(C) + \frac{\partial F_{11}(C)}{\partial C_{21}}U_{21}G_{11}(C)+ \\
& \frac{\partial F_{11}(C)}{\partial C_{12}}U_{12}G_{11}(C) + \frac{\partial F_{11}(C)}{\partial C_{22}}U_{22}G_{11}(C)+ \\
& \frac{\partial F_{12}(C)}{\partial C_{11}}U_{11}G_{21}(C) + \frac{\partial F_{12}(C)}{\partial C_{21}}U_{21}G_{21}(C)+ \\
& \frac{\partial F_{12}(C)}{\partial C_{12}}U_{12}G_{21}(C) + \frac{\partial F_{12}(C)}{\partial C_{22}}U_{22}G_{21}(C)
\end{aligned}$$

that gives all the terms in Eq. 34 with a partial derivative in $F_{ij}(C)$. The remainder terms in Eq. 34 come from the second term in the right-had side of Eq. 31.

Thus, at index $[1,1]$, when $F, G : \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$, the left-hand side and right-hand side of Eq. 31 are equal.

□

# 3 Proofs with matrix differentials

**Claim 2** (derivative of the log determinant of $C$)**.** *Let $C = \Lambda\Lambda^\intercal + \psi I$ and $\Lambda \in \mathbb{R}^{m\times p}$ then*

$$\frac{\partial}{\partial\Lambda}\log|C| = 2C^{-1}\Lambda \tag{37}$$

*Proof.* Note the notational difference between Eq. 37 and the definition of first derivative in Definition 2. According to the statement of this claim $\frac{\partial}{\partial\Lambda}\log|C| \in \mathbb{R}^{m\times q}$ but according to Definition 2 $\frac{\partial}{\partial\Lambda}\log|C| \in \mathbb{R}^{1\times mq}$. In this proof I will use the notation from Definition 2 and I will show that the first derivative of $\log|C|$ is $A(\Lambda) = (\mathrm{vec}\, 2C^{-1}\Lambda)^\intercal$.

Define $F(\Lambda) = \Lambda\Lambda^\intercal + \psi I$ and $G(M) = \log|M|$. Then $\log|C| = (G\circ F)(\Lambda)$

$$\begin{aligned}
dF(\Lambda; d\Lambda) &= d(\Lambda\Lambda^\intercal)(\Lambda; d\Lambda) && (38)\\
&= d\Lambda(\Lambda; d\Lambda)\ \Lambda^\intercal + \Lambda\ d\Lambda^\intercal(\Lambda; d\Lambda) && (39)\\
&= (d\Lambda)\ \Lambda^\intercal + \Lambda\ (d\Lambda)^\intercal && (40)
\end{aligned}$$

and

$$dG(C; dC) = d\log|C|(C; dC) \tag{41}$$

Notes:

- Eq. 38 follows from the definition of $F(\Lambda)$ and from the fact that $\Lambda$ does not appear in $\psi I$.
- Eq. 39 results from the application of the product rule for differentials (Theorem 4).
- Eq. 40 is a consequence of sub claims 2 and 3.
- Eq. 41 stems from the definition of $G$.

$$\begin{aligned}
d\log|C|(\Lambda; d\Lambda) &= d(G\circ F)(\Lambda; d\Lambda) && (42)\\
&= dG(F(\Lambda); dF(\Lambda; d\Lambda)) && (43)\\
&= d\log|C|(F(\Lambda); dF(\Lambda; d\Lambda)) && (44)\\
&= \mathrm{Tr}(C^{-1}dF(\Lambda; d\Lambda)) && (45)\\
&= \mathrm{Tr}(C^{-1}((d\Lambda)\ \Lambda^\intercal + \Lambda\ (d\Lambda)^\intercal)) && (46)\\
&= \mathrm{Tr}(C^{-1}(d\Lambda)\Lambda^\intercal) + \mathrm{Tr}(C^{-1}\Lambda(d\Lambda)^\intercal) && (47)\\
&= 2\,\mathrm{Tr}(\Lambda^\intercal C^{-1}(d\Lambda)) && (48)\\
&= \mathrm{Tr}(2\Lambda^\intercal C^{-1}(d\Lambda)) && (49)\\
&= (\mathrm{vec}\, 2C^{-1}\Lambda)^\intercal\, \mathrm{vec}\, d\Lambda && (50)
\end{aligned}$$

Notes:

- Eq. 42 applies the definition of $F$ and $G$ given above,

- Eq. 43 follows from Theorem 3,
- Eq. 44 arises from Eq 41,
- Eq. 45 results from sub claim 1,
- Eq. 46 uses Eq. 40,
- Eq. 47 is a consequence of the linearity of the Tr,
- Eq. 48 holds because $\mathrm{Tr}(C^{-1}(d\Lambda)\Lambda^\intercal) = \mathrm{Tr}(\Lambda^\intercal C^{-1}(d\Lambda))$, due to the circular property of the trace, and because $\mathrm{Tr}(C^{-1}\Lambda(d\Lambda)^\intercal) = \mathrm{Tr}((C^{-1}\Lambda(d\Lambda)^\intercal)^\intercal) = \mathrm{Tr}((d\Lambda)\Lambda^\intercal C^{-1}) = \mathrm{Tr}(\Lambda^\intercal C^{-1}(d\Lambda))$, due to the invariance of the trace to transposition and again due to the circular property of the trace,
- Eq. 50 follows from the fact that $\mathrm{Tr}(A^\intercal B) = (\mathrm{vec}\,A)^\intercal B$.

Therefore, from the definition of first derivative in Definition 2 and Eq. 50, the first derivative of $\log |C|$ with respect to $\Lambda$ is $A(\Lambda) = (\mathrm{vec}\, 2C^{-1}\Lambda)^\intercal$, as required to complete the proof.

□

**Claim 3** (derivative of the trace of a matrix inverse times a constant matrix)**.** *Let* $C = \Lambda\Lambda^\intercal + \psi I$ *then*

$$\frac{\partial}{\partial\Lambda}\mathrm{Tr}[C^{-1}S] = -2C^{-1}SC^{-1}\Lambda \tag{51}$$

*Proof.* Note the notational difference between the statement of this claim and the definition of first derivative in Definition 2. According to the statement of this claim $\frac{\partial}{\partial\Lambda}\mathrm{Tr}[C^{-1}S] \in \mathbb{R}^{m\times q}$ but according to Definition 2 $\frac{\partial}{\partial\Lambda}\mathrm{Tr}[C^{-1}S] \in \mathbb{R}^{1\times mq}$. In this proof I will use the notation from Definition 2 and I will show that the first derivative of $\mathrm{Tr}[C^{-1}S]$ is $A(\Lambda) = (\mathrm{vec}(-2C^{-1}SC^{-1}\Lambda))^\intercal$.

$$\begin{aligned}
d\,\mathrm{Tr}[C^{-1}S](\Lambda; d\Lambda) &= \mathrm{Tr}[dC^{-1}S(\Lambda; d\Lambda)] && (52)\\
&= \mathrm{Tr}[dC^{-1}(\Lambda; d\Lambda)S] && (53)\\
&= \mathrm{Tr}[-C^{-1}dC(\Lambda; d\Lambda)C^{-1}S] && (54)\\
&= \mathrm{Tr}[-C^{-1}((d\Lambda)\Lambda^\intercal + \Lambda(d\Lambda^\intercal))C^{-1}S] && (55)\\
&= \mathrm{Tr}[-C^{-1}(d\Lambda)\Lambda^\intercal C^{-1}S] + \mathrm{Tr}[-C^{-1}\Lambda(d\Lambda^\intercal)C^{-1}S] && (56)\\
&= \mathrm{Tr}[-2\Lambda^\intercal C^{-1}SC^{-1}d\Lambda] && (57)\\
&= (\mathrm{vec}(-2C^{-1}SC^{-1}\Lambda))^\intercal d\Lambda && (58)
\end{aligned}$$

Notes:

- Eq. 52 holds due to the linearity of the differential and the trace,
- Eq. 53 is a consequence of Theorem 4 and the fact that $S$ is a constant with respect to $\Lambda$,
- Eq. 54 follows from sub claim 4,
- Eq. 55 results from Eq. 40,
- Eq. 56 stems from the linearity of the trace,
- The first term of Eq. 56 equals $\mathrm{Tr}[-\Lambda^\intercal C^{-1}SC^{-1}d\Lambda]$ due to the circularity of the trace and the second term also equals this value because, in addition, $\mathrm{Tr}[A] = \mathrm{Tr}[A^\intercal]$. Thus, the sum of these two terms equals Eq. 57,

- Eq. 58 arises from the fact that $\mathrm{Tr}[A^\intercal B] = (\mathrm{vec}\, A)^\intercal \,\mathrm{vec}\, B$.

□

**Sub Claim 1** (differential of log determinant)**.**

$$d \log |C|(C; dC) = \mathrm{Tr}(C^{-1} dC)$$

*Proof.* Define $F : S \subseteq \mathbb{R} \to \mathbb{R}^+$, where $S$ is the space of positive semidefinite matrices, as $F(X) = |X|$ and $g : \mathbb{R}^+ \to \mathbb{R}$ as $g(x) = \log x$. Then $H(C) := \log |C| = (g \circ F)(C)$ and by Theorem 3

$$d \log |C|(C; dC) = dg(F(C); dF(C; dC)) \tag{59}$$

From Definition 1

$$dg(x; dx) = g'(x)dx = \frac{dx}{x} \tag{60}$$

Next we calculate $dF(C; dC)$. Expanding the determinant along the $i$th row we obtain

$$F(X) = \sum_{j=1}^{N} (-1)^{i+j} X_{ij} M_{ij}(X)$$

where $M_{ij}(X)$ is then $ij$th minor of $X$. Then

$$\frac{\partial F(X)}{\partial X_{ik}} = (-1)^{i+k} M_{ik}(X)$$

since $X_{ik}$ does not appear in $M_{ij}(X)$, $1 \leq j \leq N$. $(-1)^{i+k} M_{ik}(X)$ is the $ik$th entry of the cofactor matrix of $X$, $\mathrm{cof}(X)$. So

$$\frac{\partial F(X)}{\partial X_{ik}} = \mathrm{cof}(X)_{ik}$$

Next from Definition 4 the Jacobian of $F(C)$ is

$$\begin{aligned}
DF(C) &= \left.\frac{\partial\, \mathrm{vec}\, F(X)}{\partial (\mathrm{vec}\, X)'}\right|_{X=C} \\
&= \left.\frac{\partial F(X)}{\partial (\mathrm{vec}\, X)'}\right|_{X=C} && (61) \\
&= \left[ \left.\frac{\partial F(X)}{\partial X_{11}}\right|_{X=C}, \left.\frac{\partial F(X)}{\partial X_{12}}\right|_{X=C}, \ldots, \left.\frac{\partial F(X)}{\partial X_{NN}}\right|_{X=C} \right] \\
&= [\mathrm{cof}(C)_{11}, \mathrm{cof}(C)_{12}, \ldots, \mathrm{cof}(C)_{NN}] \\
&= (\mathrm{vec}\, \mathrm{cof}(C)^\intercal)^\intercal = (\mathrm{vec}\, \mathrm{adj}(C))^\intercal && (62)
\end{aligned}$$

Notes:

- Eq. 61 holds because $F(X)$ is a scalar.

Now by Definition 2 and Theorem 1

$$\begin{aligned} dF(C; dC) &= \operatorname{vec} dF(C; dC) && (63) \\ &= DF(C) \operatorname{vec} dC && (64) \\ &= (\operatorname{vec} \operatorname{adj}(C))^\intercal \operatorname{vec} dC && (65) \end{aligned}$$

Notes:

- Eq. 63 holds because $dF(C; dC)$ is a scalar.
- Eq. 64 follows from Definition 2 and Theorem 1.
- Eq. 65 follows from Definition 62.

Finally,

$$\begin{aligned} d \log |C|(C; dC) &= \frac{dF(C; dC)}{F(C)} && (66) \\ &= \frac{(\operatorname{vec} \operatorname{adj}(C))^\intercal \operatorname{vec} dC}{|C|} && (67) \\ &= \left( \operatorname{vec} \frac{\operatorname{adj}(C)}{|C|} \right)^\intercal \operatorname{vec} dC && (68) \\ &= \left( \operatorname{vec} C^{-1} \right)^\intercal \operatorname{vec} dC && (69) \\ &= \operatorname{Tr} \left( C^{-1} dC \right) && (70) \end{aligned}$$

Notes:

- Eq. 66 follows from Eqs. 59 and 60.
- Eq. 67 is a due to Eq. 65.
- Eq. 68 hold by the linearity of the vec operation.
- Eq. 69 uses the linear algebra result that $\frac{\operatorname{adj}(C)}{|C|} = C^{-1}$
- Eq. 70 holds because $\operatorname{Tr}[A^\intercal B] = (\operatorname{vec} A)^\intercal \operatorname{vec} B$.

□

**Sub Claim 2** (differential of identity function)**.** *Let* $V(\Lambda) = \Lambda \in \mathbb{R}^{m \times p}$ *then* $dV(\Lambda; d\Lambda) = d\Lambda$.

*Proof.* From Definition 2 and Theorem 1

$$\operatorname{vec} dV(\Lambda; d\Lambda) = DV(\Lambda) \operatorname{vec} d\Lambda \tag{71}$$

and from Definition 4

$$DV(\Lambda) = \frac{\partial \operatorname{vec} V(\Lambda)}{\partial (\operatorname{vec} \Lambda)'} = \frac{\partial \operatorname{vec} \Lambda}{\partial (\operatorname{vec} \Lambda)'} = I_{mp \times mp} \tag{72}$$

Thus, from Eqs. 71 and 72

$$\operatorname{vec} dV(\Lambda; d\Lambda) = I_{mp \times mp} \operatorname{vec} d\Lambda = \operatorname{vec} d\Lambda \tag{73}$$

Therefore, from Eq. 73, $dV(\Lambda; d\Lambda) = d\Lambda$.

□

**Sub Claim 3** (differential of transpose function)**.** *Let* $W(\Lambda) = \Lambda^\intercal \in \mathbb{R}^{p\times m}$ *then* $dW(\Lambda; d\Lambda) = (d\Lambda)^\intercal$*.*

*Proof.* We will sketch the proof using $\Lambda \in \mathbb{R}^{2\times 2}$.

$$W(\Lambda) = \Lambda^\intercal = \begin{bmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{bmatrix}^\intercal = \begin{bmatrix}\Lambda_{11} & \Lambda_{21}\\ \Lambda_{12} & \Lambda_{22}\end{bmatrix}$$

From Definition 2 and Theorem 1

$$\operatorname{vec} dW(\Lambda; d\Lambda) = DW(\Lambda) \operatorname{vec} d\Lambda \tag{74}$$

and from Definition 4

$$\begin{aligned} DW(\Lambda) &= \frac{\partial \operatorname{vec} W(\Lambda)}{\partial(\operatorname{vec}\Lambda)'} = \frac{\partial \operatorname{vec}\Lambda^\intercal}{\partial(\operatorname{vec}\Lambda)'}\\ &= \begin{bmatrix} \frac{\partial\lambda_{11}}{\partial\lambda_{11}} & \frac{\partial\lambda_{11}}{\partial\lambda_{21}} & \frac{\partial\lambda_{11}}{\partial\lambda_{12}} & \frac{\partial\lambda_{11}}{\partial\lambda_{22}}\\ \frac{\partial\lambda_{12}}{\partial\lambda_{11}} & \frac{\partial\lambda_{12}}{\partial\lambda_{21}} & \frac{\partial\lambda_{12}}{\partial\lambda_{12}} & \frac{\partial\lambda_{12}}{\partial\lambda_{22}}\\ \frac{\partial\lambda_{21}}{\partial\lambda_{11}} & \frac{\partial\lambda_{21}}{\partial\lambda_{21}} & \frac{\partial\lambda_{21}}{\partial\lambda_{12}} & \frac{\partial\lambda_{21}}{\partial\lambda_{22}}\\ \frac{\partial\lambda_{22}}{\partial\lambda_{11}} & \frac{\partial\lambda_{22}}{\partial\lambda_{21}} & \frac{\partial\lambda_{22}}{\partial\lambda_{12}} & \frac{\partial\lambda_{22}}{\partial\lambda_{22}} \end{bmatrix}\\ &= \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0\\ 0.0 & 0.0 & 1.0 & 0.0\\ 0.0 & 1.0 & 0.0 & 0.0\\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \end{aligned} \tag{75}$$

Thus, from Eqs. 74 and 75

$$\begin{aligned} \operatorname{vec} dW(\Lambda; d\Lambda) &= \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0\\ 0.0 & 0.0 & 1.0 & 0.0\\ 0.0 & 1.0 & 0.0 & 0.0\\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \operatorname{vec} d\Lambda\\ &= \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0\\ 0.0 & 0.0 & 1.0 & 0.0\\ 0.0 & 1.0 & 0.0 & 0.0\\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \begin{bmatrix} d\Lambda_{11}\\ d\Lambda_{21}\\ d\Lambda_{12}\\ d\Lambda_{22}\end{bmatrix} = \begin{bmatrix} d\Lambda_{11}\\ d\Lambda_{12}\\ d\Lambda_{21}\\ d\Lambda_{22}\end{bmatrix} = \operatorname{vec} d\Lambda^\intercal \end{aligned} \tag{76}$$

Therefore, from Eq. 76, $dW(\Lambda; d\Lambda) = d\Lambda^\intercal$.

□

**Sub Claim 4** (differential of inverse function)**.**

$$dC^{-1}(\Lambda; d\Lambda) = -C^{-1} dC(\Lambda; d\Lambda) C^{-1} \tag{77}$$

*Proof.*

$$I = CC^{-1} \quad \text{then} \tag{78}$$

$$0 = dI(\Lambda; d\Lambda) \tag{79}$$

$$= dCC^{-1}(\Lambda; d\Lambda) \tag{80}$$

$$= dC(\Lambda; d\Lambda)\ C^{-1} + C\ dC^{-1}(\Lambda; d\Lambda) \quad \text{then} \tag{81}$$

$$dC^{-1}(\Lambda; d\Lambda) = -C^{-1} dC(\Lambda; d\Lambda) C^{-1} \tag{82}$$

Notes:

1. Eq. 79 is valid because the identity matrix, $I$, is constant with respect to $\Lambda$,
2. Eq. 80 follows from Eq. 78,
3. Eq. 81 is a consequence of Theorem 4,
4. Eq. 82 results from solving Eq. 81 for $dC^{-1}(\Lambda; d\Lambda)$.

□

# 4 Discussion

The first thing that most of of us do when we need to calculate the derivative of a function with matrix arguments is look for its formula in the Matrix Cookbook[1]. However, sometimes the formula is not there, or we may be curious about the derivation of a given formula in this cookbook.

It is fulfilling to understand concepts in exquisite levels of detail, as I tried to do it here.

# References


Dattorro, J. (2015). *Convex optimization & Euclidean distance geometry*. Lulu. com.

Magnus, J. R. and Neudecker, H. (2019). *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons.

Marsden, J. E. and Tromba, A. (2003). *Vector calculus*. Macmillan, 5 edition.

Shumway, R. H., Stoffer, D. S., and Stoffer, D. S. (2016). *Time series analysis and its applications*. Springer, 4 edition.


[1] https://www.math.uwaterloo.ca/ hwolkowi/matrixcookbook.pdf