On $\det \boldsymbol{X}$, $\operatorname{logdet}\boldsymbol{X}$ and $\operatorname{logdet}\boldsymbol{X}^\top\boldsymbol{X}$

Andersen Ang

ECS, Uni. Southampton, UK

andersen.ang@soton.ac.uk

Homepage angms.science

Version: November 11, 2023
First draft: August 14, 2017

Content

What is $\det$: the volume of parallelepiped

- Given: a square matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n] \in \mathbb{R}^{n \times n}$ with linearly independent columns.
- $\det \boldsymbol{X}$ = the volume of the *parallelepiped* spaned by the columns $\{\boldsymbol{x}_i\}$

$$\det \boldsymbol{X} = \mathsf{vol}\Big(\{\boldsymbol{x}_1\, \boldsymbol{x}_2 \ldots \boldsymbol{x}_n\}\Big).$$

- **Why**: consider case $n = 2$: two vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^2$, we can always perform a Gram-Schmidt orthogonalization to get two vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$

$$\begin{cases} \boldsymbol{v}_1 = \boldsymbol{x}_1 \\ \boldsymbol{v}_2 = c_{12}\boldsymbol{v}_1 + \boldsymbol{v}_2^{\perp} \qquad \boldsymbol{v}_2^{\perp} \perp \boldsymbol{v}_1 \end{cases} \implies \begin{aligned} \det[\boldsymbol{v}_1, \boldsymbol{v}_2] &= \det[\boldsymbol{v}_1, c_{12}\boldsymbol{v}_1 + \boldsymbol{v}_2^{\perp}] \\ &= \det[\boldsymbol{v}_1, \boldsymbol{v}_2^{\perp}] \\ &= \text{signed volume}(\boldsymbol{v}_1, \boldsymbol{v}_2) \end{aligned}$$

- Hannah, John, “A geometric approach to determinants” *American Mathematical Monthly*, 1996: 401-409.

Derivative of $\det \boldsymbol{X}$ - the Jacobi's Formula

- For a non-singular matrix $\boldsymbol{X}$, recall:
  - adjugate-det-inverse relationship: $\mathrm{adj}\boldsymbol{X} = \det \boldsymbol{X} \cdot \boldsymbol{X}^{-1}$
  - adjugate-cofactor relationship: $\mathrm{adj}\boldsymbol{X} = \boldsymbol{C}^\top$
  - Therefore, $\det \boldsymbol{X} \cdot \boldsymbol{X}^{-1} = \mathrm{adj}\boldsymbol{X} = \boldsymbol{C}^\top$
- Jacobi's formula gives the derivative of $\det \boldsymbol{X}$ with respect to (w.r.t.) scalar x

$$\begin{aligned}
\frac{\partial \det \boldsymbol{X}}{\partial x} &= \det \boldsymbol{X} \cdot \mathrm{Tr}\left(\boldsymbol{X}^{-1} \frac{\partial \boldsymbol{X}}{\partial x}\right) && \text{Jacobi's formula} \\
&= \mathrm{Tr}\left(\underbrace{\det \boldsymbol{X} \cdot \boldsymbol{X}^{-1}}_{\boldsymbol{C}^\top} \frac{\partial \boldsymbol{X}}{\partial x}\right) && c\mathrm{Tr}(\boldsymbol{A}) = \mathrm{Tr}(c\boldsymbol{A}) \\
&= \left\langle \boldsymbol{C}, \frac{\partial \boldsymbol{X}}{\partial x} \right\rangle && \text{definition of matrix inner product}
\end{aligned}$$

  So the derivative of $\det \boldsymbol{X}$ equals to the matrix inner product of cofactor and derivative of $\boldsymbol{X}$

- The equation $\frac{\partial \det \boldsymbol{X}}{\partial x} = \left\langle \boldsymbol{C}, \frac{\partial \boldsymbol{X}}{\partial x} \right\rangle$ makes sense:
  - The derivative of scalar value $\det \boldsymbol{X}$ w.r.t. scalar x is a scalar
  - $\left\langle \boldsymbol{C}, \frac{\partial \boldsymbol{X}}{\partial x} \right\rangle$ is a scalar

  The derivative of $\det \boldsymbol{X}$ w.r.t. matrix $\boldsymbol{X}$ is a matrix. The expression is

$$\frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \cdot \boldsymbol{X}^{-T}, \text{ is a matrix}$$

  For derivation, refer to previous document.

What is $\text{logdet}\boldsymbol{X}$, the log-determinant of a matrix $\boldsymbol{X}$?

- $\boldsymbol{X}$ has to be square ($\because \det$)
- $\boldsymbol{X}$ has to be positive definite (pd), because
  - $\det \boldsymbol{X} = \prod_i \lambda_i$
  - all eigenvalues of pd matrix are strictly positive
  - $\text{dom} \log = \mathbb{R}_{++}$, $\log$ of negative number produces complex number which is out of context
- if $\boldsymbol{X}$ is positive semidefinite (psd), it is better to consider $\text{logdet}(\boldsymbol{X} + \delta \boldsymbol{I})$, where $\delta > 0$. As all eigenvalues of a psd matrix is non-negative (zero or positive), adding a small positive number δ removes all the zero eigenvalues, turns $\boldsymbol{X} + \delta I$ to pd
- logdet is a concave function
- $\dfrac{\partial \text{logdet}\boldsymbol{X}}{\partial \boldsymbol{X}} = \boldsymbol{X}^{-T}$
- Why logdet: "equalizing" eigenvalues
  - For a matrix $\boldsymbol{X}$, the expression $\det \boldsymbol{X} = \prod \lambda_i$ is dominated by the leading eigenvalues. Such domination is not a problem if we only care about the leading eigenvalues.
  - When all the eigenvalues are equally important, a way to suppress the leading eigenvalues is to use $\log$. Hence we have

$$\text{logdet}\boldsymbol{X} = \sum \log \lambda_i = \log \lambda_1 + \log \lambda_2 + \cdots$$

    Again, since the input of $\log$ should not be negative nor zero, the matrix $\boldsymbol{X}$ here should be pd to make $\lambda_i > 0$

$f(\boldsymbol{X}) = \mathrm{logdet}\boldsymbol{X}$ is concave on $\boldsymbol{X} \in \mathbb{S}^n_{++}$

- Idea of the proof: by the equivalence

$$\Big\{ f \text{ is concave on } \boldsymbol{X} \Big\} \iff \left\{ \begin{array}{c} g(t) = f(\boldsymbol{X}) = f(\boldsymbol{Z} + t\boldsymbol{V}) \text{ is concave on } t \\ \mathrm{dom} g = \Big\{ t | \boldsymbol{Z} + t\boldsymbol{V} \succ \boldsymbol{0} \Big\} \cap \Big\{ t = 0 \Big\}, \quad \boldsymbol{Z} \in \mathbb{S}^n, \boldsymbol{V} \in \mathbb{S}^n \end{array} \right\}$$

- **Proof**

$$\begin{aligned}
g(t) &= \mathrm{logdet}(\boldsymbol{Z} + t\boldsymbol{V}) \\
&= \mathrm{logdet}\Big(\boldsymbol{Z}^{1/2} (\boldsymbol{I} + t\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2}) \boldsymbol{Z}^{1/2} \Big) \\
&= \mathrm{logdet}\Big(\boldsymbol{Z}^{1/2} (\boldsymbol{I} + t\hat{\boldsymbol{V}}) \boldsymbol{Z}^{1/2} \Big) && \text{let } \hat{\boldsymbol{V}} = \boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2} \\
&= \log \Big[\det \Big(\boldsymbol{I} + t\hat{\boldsymbol{V}} \Big) \det \boldsymbol{Z} \Big] && \det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{A}) \det(\boldsymbol{B}) \\
&= \mathrm{logdet}\Big(\boldsymbol{I} + t\hat{\boldsymbol{V}} \Big) + \mathrm{logdet}\boldsymbol{Z} \\
&= \mathrm{logdet}\Big(\boldsymbol{Q}\boldsymbol{Q}^\top + t\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\top \Big) + \mathrm{logdet}\boldsymbol{Z} && \hat{\boldsymbol{V}} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\top, \boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I} \\
&= \mathrm{logdet}\Big(\boldsymbol{Q}(\boldsymbol{I} + t\boldsymbol{\Lambda})\boldsymbol{Q}^\top \Big) + \mathrm{logdet}\boldsymbol{Z} \\
&= \mathrm{logdet}\Big(\boldsymbol{Q}^\top\boldsymbol{Q}(\boldsymbol{I} + t\boldsymbol{\Lambda}) \Big) + \mathrm{logdet}\boldsymbol{Z} && \det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{A}) \det(\boldsymbol{B}) \\
&= \mathrm{logdet}(\boldsymbol{Q}^\top\boldsymbol{Q}) + \mathrm{logdet}(\boldsymbol{I} + t\boldsymbol{\Lambda}) + \mathrm{logdet}\boldsymbol{Z} \\
&= \mathrm{logdet}(\boldsymbol{I} + t\boldsymbol{\Lambda}) + \mathrm{logdet}\boldsymbol{Z} && \boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}, \det \boldsymbol{I} = 1, \log 1 = 0 \\
&= \log \prod (1 + t\lambda_i) + \mathrm{logdet}\boldsymbol{Z} && \det \boldsymbol{\Lambda} = \prod \lambda_i \\
&= \sum \log(1 + t\lambda_i) + \mathrm{logdet}\boldsymbol{Z} \\
\frac{\partial g}{\partial t} &= \sum \frac{\lambda_i}{1 + t\lambda_i} \\
\frac{\partial^2 g}{\partial t^2} &= -\sum \frac{\lambda_i^2}{(1 + t\lambda_i)^2} < 0 \implies g(t) \text{ is concave} \iff f(\boldsymbol{X}) \text{ is concave}
\end{aligned}$$

Application of concavity of $\operatorname{logdet} \boldsymbol{X}$: Taylor bound

- Fact: first order Taylor approximation is a global over-estimator of a concave function.
$$f(\boldsymbol{X}) \;\leq\; f(\boldsymbol{Y}) + \langle \nabla f(Y), \boldsymbol{X} - \boldsymbol{Y} \rangle.$$
- As $\operatorname{logdet} \boldsymbol{X}$ is concave, so
$$\operatorname{logdet} \boldsymbol{X} \;\leq\; \operatorname{logdet} \boldsymbol{Y} + \langle \boldsymbol{Y}^{-\top}, \boldsymbol{X} - \boldsymbol{Y} \rangle.$$
Again, $\boldsymbol{Y}^{-\top}$ and $\boldsymbol{X} - \boldsymbol{Y}$ are matrices while $\operatorname{logdet} \boldsymbol{X}$ and $\operatorname{logdet} Y$ are scalars, so the matrix inner product $\langle \, , \, \rangle$ has to be applied.
$$\begin{aligned} \operatorname{logdet} \boldsymbol{X} &\leq \operatorname{logdet} \boldsymbol{Y} + \operatorname{Tr}\Big(\boldsymbol{Y}^{-1} (\boldsymbol{X} - \boldsymbol{Y}) \Big) \\ &= \operatorname{logdet} \boldsymbol{Y} + \operatorname{Tr}\Big(\boldsymbol{Y}^{-1} \boldsymbol{X} - \boldsymbol{I} \Big) \\ &= \operatorname{logdet} \boldsymbol{Y} + \operatorname{Tr}(\boldsymbol{Y}^{-1} \boldsymbol{X}) - n \end{aligned}$$
Such expression can be used for minimization of $\operatorname{logdet} \boldsymbol{X}$.

(Fast) Ways to compute $\text{logdet}\boldsymbol{X}$

- By definition.
  Compute $\det \boldsymbol{X}$, then take log.
  For example, using Laplace cofactor expansion formula

- Eigenvalues ($\det \boldsymbol{X} = \prod \lambda_i$)
  Compute the eigenvalues of $\boldsymbol{X}$, then $\text{logdet}\boldsymbol{X} = \sum \log \lambda_i$

- Cholesky factorization.
  For $\boldsymbol{X}$ being pd, apply Cholesky decomoposition on $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{L}^\top$
  Then with $\det \boldsymbol{L} = \prod L_{ii}$, we compute $\text{logdet}\boldsymbol{X} = 2\sum \log L_{ii}$

- Approximating the determinant $\det \boldsymbol{X}$ for $\boldsymbol{X}$ with large dimension
  e.g. for million-by-million matrix, we have to use random matrix theory to guess the eigenvalue of $\boldsymbol{X}$

Why $\text{logdet}\boldsymbol{X}^\top\boldsymbol{X}$

- $\det \boldsymbol{X}$ assumes $\boldsymbol{X}$ is a square matrix.
- For non-square $\boldsymbol{X}$, one can try $\det \boldsymbol{X}^\top\boldsymbol{X}$, where $\boldsymbol{X}^\top\boldsymbol{X}$ is the Gram matrix of $\boldsymbol{X}$ and it is always psd: $y^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{y} = \|\boldsymbol{X}\boldsymbol{y}\|_2^2 \geq 0$.
- Again it is better to consider a regularied version $\text{logdet}(\boldsymbol{X}^\top\boldsymbol{X} + \delta\boldsymbol{I})$ for removing the possibility of having $\det(\boldsymbol{X}^\top\boldsymbol{X} + \delta\boldsymbol{I}) = 0$.
- $\text{logdet}\boldsymbol{X}^\top\boldsymbol{X}$ and $\text{logdet}(\boldsymbol{X}^\top\boldsymbol{X} + \delta I)$ is not concave nor convex in $\boldsymbol{X}$

Derivative of $\text{logdet}(\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})$ ... method 1 (scalar chain rule)

- To shorten the notation, let $\boldsymbol{B} = \boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I}$.
- Again, the derivative of the scalar-valued function $\text{logdet}\boldsymbol{B}$ w.r.t. to a scalar x is a scalar, so using chain rule gives

$$\frac{\partial \text{logdet}\boldsymbol{B}}{\partial x} = \frac{\partial \text{logdet}\boldsymbol{B}}{\partial \det \boldsymbol{B}} \frac{\partial \det \boldsymbol{B}}{\partial x}. \tag{1}$$

  And the following expression is wrong:

$$\frac{\partial \text{logdet}\boldsymbol{B}}{\partial x} = \frac{\partial \text{logdet}\boldsymbol{B}}{\partial \boldsymbol{B}} \frac{\partial \boldsymbol{B}}{\partial x}.$$

- Consider (1): $\det B$ is a scalar so $\frac{\partial \text{logdet}\boldsymbol{B}}{\partial \det \boldsymbol{B}}$ is just simple log differentiation and hence $\frac{\partial \text{logdet}\boldsymbol{B}}{\partial \det \boldsymbol{B}} = \frac{1}{\det \boldsymbol{B}}$.

  For $\frac{\partial \det \boldsymbol{B}}{\partial x}$, apply Jacobi's formula gives $\frac{\partial \det \boldsymbol{B}}{\partial x} = \det \boldsymbol{B}\, \text{Tr}\Big(\boldsymbol{B}^{-1} \frac{\partial \boldsymbol{B}}{\partial x}\Big)$.

  Put ■, ■ to (1) gives $\frac{\partial \text{logdet}\boldsymbol{B}}{\partial x} = \text{Tr}\Big(\boldsymbol{B}^{-1} \frac{\partial \boldsymbol{B}}{\partial x}\Big)$, now put $\boldsymbol{B} = \boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I}$ back gives

$$\begin{aligned}
\frac{\partial \text{logdet}(\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})}{\partial x} &= \text{Tr}\Big((\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})^{-1} \frac{\partial (\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})}{\partial x}\Big) = \text{Tr}\Big((\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})^{-1} \Big(\frac{\partial \boldsymbol{X}^\top \boldsymbol{X}}{\partial x} + \underbrace{\frac{\partial \delta \boldsymbol{I}}{\partial x}}_{=\boldsymbol{0}}\Big)\Big) \\
&= \text{Tr}\Big((\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})^{-1} \frac{\partial \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}\Big).
\end{aligned}$$

Derivative of $\operatorname{logdet}(\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})$ ... method 2 (inner product)

- Using inner product, we have

$$\frac{\partial \operatorname{logdet} \boldsymbol{B}}{\partial x} = \operatorname{Tr}\Big[\Big(\frac{\partial \operatorname{logdet} \boldsymbol{B}}{\partial \boldsymbol{B}}\Big)^\top \frac{\partial \boldsymbol{B}}{\partial x}\Big]$$

- Since $\dfrac{\partial \operatorname{logdet} \boldsymbol{B}}{\partial \boldsymbol{B}} = \boldsymbol{B}^{-\top}$, hence

$$\frac{\partial \operatorname{logdet} \boldsymbol{B}}{\partial x} = \operatorname{Tr}\Big[\boldsymbol{B}^{-1} \frac{\partial \boldsymbol{B}}{\partial x}\Big]$$

- From last page $\dfrac{\partial \boldsymbol{B}}{\partial x} = \dfrac{\partial \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}$, so

$$\frac{\partial \operatorname{logdet} \boldsymbol{B}}{\partial x} = \operatorname{Tr}\Big[\boldsymbol{B}^{-1} \frac{\partial \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}\Big]$$

- Now put $\boldsymbol{B} = \boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I}$ back gives

$$\frac{\partial \operatorname{logdet}(\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})}{\partial x} = \operatorname{Tr}\Big((\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})^{-1} \frac{\partial \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}\Big).$$

Same result as the one using chain rule.

Continue from $\dfrac{\partial \mathrm{logdet}(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})}{\partial x} = \mathrm{Tr}\Big[(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})^{-1}\dfrac{\partial \boldsymbol{X}^\top\boldsymbol{X}}{\partial x}\Big]$

- Put $x = X_{ij}$ gives $\dfrac{\partial \boldsymbol{X}^\top\boldsymbol{X}}{\partial X_{ij}} = \boldsymbol{X}^\top\boldsymbol{J}^{ij} + \boldsymbol{J}^{ji}\boldsymbol{X}$ where $\boldsymbol{J}^{ij}$ is single entry matrix. We get

$$\begin{aligned}\frac{\partial \mathrm{logdet}(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})}{\partial X_{ij}} &= \mathrm{Tr}\Big[(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})^{-1}(\boldsymbol{X}^\top\boldsymbol{J}^{ij}+\boldsymbol{J}^{ij}\boldsymbol{X})\Big] \\ &= \mathrm{Tr}\Big[(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{J}^{ij}\Big] + \mathrm{Tr}\Big[(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})^{-1}\boldsymbol{J}^{ij}\boldsymbol{X}\Big].\end{aligned}$$

- By $\mathrm{Tr}(\boldsymbol{A}\boldsymbol{J}^{ij}) = [\boldsymbol{A}^\top]_{ij}$ and $\mathrm{Tr}(\boldsymbol{A}\boldsymbol{J}^{ji}\boldsymbol{B}) = [\boldsymbol{B}\boldsymbol{A}]_{ij}$, we have

$$\frac{\partial \mathrm{logdet}(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})}{\partial X_{ij}} = \Big[(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})^{-1}\boldsymbol{X}^\top\Big]^\top_{ij} + \Big[\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})^{-1}\Big]_{ij}.$$

- As $\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I}$ is symmetric so

$$\frac{\partial \mathrm{logdet}(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})}{\partial X} = 2\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X}+\delta\boldsymbol{I})^{-1}.$$

Last page - summary

- Physical meanings of $\det \boldsymbol{X}$ and $\text{logdet}\boldsymbol{X}$
- $\text{logdet}\boldsymbol{X}$ is concave in $\boldsymbol{X}$
- Jacobi Formula and derivatives w.r.t. matrix variable $\boldsymbol{X}$ of $\det \boldsymbol{X}$, $\text{logdet}\boldsymbol{X}$ and $\text{logdet}(\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})$
- $\dfrac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \cdot \boldsymbol{X}^{-\top}$ which implies $\dfrac{\partial \text{logdet}\boldsymbol{X}}{\partial \boldsymbol{X}} = \boldsymbol{X}^{-\top}$
- $\dfrac{1}{2} \dfrac{\partial \text{logdet}(\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})}{\partial \boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \delta \boldsymbol{I})^{-1}$

End of document