On $\det X$, $\log\det X$ and $\log\det X^T X$

Andersen Ang

Mathématique et de Recherche opérationnelle
Faculté polytechnique de Mons
UMONS
Mons, Belgium

email: manshun.ang@umons.ac.be
homepage: *angms.science*

August 14, 2017

Overview

1. On $\det X$
   - Physical meaning of $\det X$
   - Jacobi's Formula and $\frac{\partial \det X}{\partial X} = \det X \cdot X^{-T}$
2. On $\log\det X$
   - Physical meaning of $\log\det X$
   - $\log\det X$ is concave
   - (Fast) computation of log det
3. On $\log\det X^T X$
   - Physical meaning of $\log\det X^T X$
   - Derivative of $\log\det X^T X + \delta I$
4. Summary

What is det : the volume of parallelepiped

For a square matrix $X = [x_1\, x_2 \dots x_n] \in \mathbb{R}^{n\times n}$ with linearly independent columns, $\det X$ tells the volume of the *parallelepiped* spaned by the columns x_i

$$\det X = \mathsf{vol}(\{x_1\, x_2 \dots x_n\})$$

Why? Consider the case $n = 2$: two vectors $x_1, x_2 \in \mathbb{R}^2$, we can always perform Gram-Schmidt orthogonalization to get two vectors v_1, v_2

$$\begin{aligned} v_1 &= x_1 \\ v_2 &= c_{12}v_1 + v_2^{\perp} \end{aligned}$$

where vector $v_2^{\perp}$ is orthogonal to v_1. Then

$$\begin{aligned} \det(v_1, v_2) &= \det(v_1, c_{12}v_1 + v_2^{\perp}) \\ &= \det(v_1, v_2^{\perp}) \\ &= \text{signed volume}(v_1, v_2) \end{aligned}$$

same argument applies to any n

For detail, see Hannah, John. "A geometric approach to determinants." American Mathematical Monthly (1996): 401-409.

Derivative of det$X$ - the Jacobi's Formula

For a non-singular matrix X, recall that :

- adjugate-det-inverse relationship: adj$X = \det X \cdot X^{-1}$
- adjugate-cofactor relationship : adj$X = C^T$

Jacobi's Formula gives the derivative of $\det X$ with respect to (w.r.t.) scalar x

$$
\begin{aligned}
\frac{\partial \det X}{\partial x} &= \det X \cdot \mathrm{Tr}\Big(X^{-1}\frac{\partial X}{\partial x}\Big) \qquad (1) \\
&= \mathrm{Tr}\Big(\underbrace{\det X \cdot X^{-1}}_{\mathrm{adj}X}\frac{\partial X}{\partial x}\Big) \\
&= \mathrm{Tr}\Big(\underbrace{\mathrm{adj}X}_{C^T}\frac{\partial X}{\partial x}\Big) \\
&= \Big\langle C, \frac{\partial X}{\partial x}\Big\rangle_F
\end{aligned}
$$

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

Derivative of det$X$ - the Jacobi's Formula

The equation $\dfrac{\partial \det X}{\partial x} = \left\langle C, \dfrac{\partial X}{\partial x} \right\rangle_F$ makes sense because :

- The derivative of scalar value $\det X$ w.r.t. scalar x is a scalar
- $\left\langle C, \dfrac{\partial X}{\partial x} \right\rangle_F$ is a scalar

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

$$\frac{\partial \det X}{\partial X} = \det X \cdot X^{-T}$$

For derivation, refer to previous document.

What is log det

The log-determinant of a matrix X is $\log \det X$

- X has to be square ($\because$ det)
- X has to be positive definite (pd), because
  - $\det X = \prod_i \lambda_i$
  - all eigenvalues of pd matrix are positive
  - domain of $\log$ has to be positive real number ($\log$ of negative number produces complex number which is out of context here)
- if X is positive semidefinite (psd), it is better to consider a regularized version $\log \det(X + \delta I)$, where $\delta > 0$ is a small positive number. As all eigenvalues of a psd matrix is non-negative (zero or positve), adding a small positive number δ removes all the zero eigenvalues, turns $X + \delta I$ to pd
- $\log \det$ is a concave function
- $\dfrac{\partial \log \det X}{\partial X} = X^{-T}$

Why $\log \det$

For a matrix X, the expression $\det X = \prod \lambda_i$ 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 important, one way to suppress the leading eigenvalues is to use $\log$. Hence we have

$$\log \det X = \sum \log \lambda_i = \log \lambda_1 + \log \lambda_2 + ...$$

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

$\log\det X$ is concave

For $f(X) = \log\det X$, $X \in \mathbb{S}^n_{++}$, f is concave on X.
Proof. To show f is concave on X is equivalent to show $g(t) = f(X) = f(Z + tV)$ is concave on t where $Z, V \in \mathbb{S}^n$, and dom$g = \{t | Z + tV \succ 0\} \cap \{t = 0\}$

$$
\begin{aligned}
g(t) &= \log\det(Z + tV) \\
&= \log\det\left(Z^{\frac{1}{2}}(I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})Z^{\frac{1}{2}}\right) \\
&= \log\det\left(Z^{\frac{1}{2}}(I + t\hat{V})Z^{\frac{1}{2}}\right) && \text{let } \hat{V} = Z^{-\frac{1}{2}}VZ^{-\frac{1}{2}} \\
&= \log\left[\det\left(I + t\hat{V}\right)\det Z\right] && \because \det(AB) = \det(A)\det(B) \\
&= \log\det\left(I + t\hat{V}\right) + \log\det Z \\
&= \log\det\left(QQ^T + tQ\Lambda Q^T\right) + \log\det Z && \hat{V} = Q\Lambda Q^T, QQ^T = Q^TQ = I \\
&= \log\det\left(Q(I + t\Lambda)Q^T\right) + \log\det Z \\
&= \log\det(QQ^T) + \log\det(I + t\Lambda) + \log\det Z \\
&= \log\det(I + t\Lambda) + \log\det Z && QQ^T = I, \det I = 1, \log 1 = 0 \\
&= \log\prod(1 + t\lambda_i) + \log\det Z && \det\Lambda = \prod\lambda_i \\
&= \sum\log(1 + t\lambda_i) + \log\det Z
\end{aligned}
$$

$\frac{\partial g}{\partial t} = \sum\frac{\lambda_i}{1 + t\lambda_i}$ and $\frac{\partial^2 g}{\partial t^2} = -\sum\frac{\lambda_i^2}{(1 + t\lambda_i)^2} < 0$, $g(t)$ is concave, so $f(X)$ is concave (and $-f(X)$ is convex)
Reference: appendix of the Convex Optimization book by Boyd

Application of concavity of $\log\det X$: Taylor bound

Fact : first order Taylor approximation is a global over-estimator of a concave function. That is,

$$f(x) \leq f(y) + \nabla f(y)^T (x - y)$$

As $\log\det X$ is concave, so it is upper bounded by its first order Taylor approximation.

$$\log\det X \leq \log\det Y + \left\langle Y^{-T}, (X - Y) \right\rangle_F$$

Again, Y^{-T} and $X - Y$ are matrices while $\log\det X$ and $\log\det Y$ are scalars, so the matrix inner product $\langle\, , \,\rangle$ has to be applied.

$$\log\det X \leq \log\det Y + \mathsf{Tr}\Big(Y^{-1}(X - Y)\Big)$$

Such expression can be used for minimization of $\log\det X$.

(Fast) computation of log det

Ways to compute $\log \det X$:

1. By definition.
Compute $\det X$, then take log.
For example, using Laplace cofactor expansion formula

2. Eigenvalues $(\det X = \prod \lambda_i)$
Compute the eigenvalues of X, then $\log \det X = \sum \log \lambda_i$

3. Cholesky factorization.
For X being pd, apply Cholesky decomoposition on $X = LL^T$, then with $\det L = \prod L_{ii}$, $\log \det X = 2 \sum \log L_{ii}$

There are many other methods , for example approximating $\det X$ for big matrix X.

On $\log\det X^T X$: why $\log\det X^T X$

The $\det X$ requires X to be square matrix.

For non-square X, one can try $\det X^T X$, where $X^T X$ is the Gram matrix of X and it is always psd : $y^T X^T X y = \|Xy\|_2^2 \geq 0$.

Again it is better to consider a regularied version $\log\det(X^X + \delta I)$ for removing the possibility of having $\det(X^T X + \delta I) = 0$.

Note. $\log\det X^T X$ and $\log\det(X^T X + \delta I)$ is not conave nor covex in X

Derivative of log det $X^T X + \delta I$

Let matrix $B = X^T X + \delta I$ to shorten the notation.

Again, the derivative of the scalar-valued function $\log \det B$ w.r.t. to a scalar x is a scalar, so the following expression after using chain rule is correct :

$$\frac{\partial \log \det B}{\partial x} = \frac{\partial \log \det B}{\partial \det B} \frac{\partial \det B}{\partial x} \tag{2}$$

And the following expression after using chain rule is wrong :

$$\frac{\partial \log \det B}{\partial x} = \frac{\partial \log \det B}{\partial B} \frac{\partial B}{\partial x}$$

The correct expression (with inner product operator) should be

$$\frac{\partial \log \det B}{\partial x} = \mathsf{Tr}\Big[\Big(\frac{\partial \log \det B}{\partial B}\Big)^T \frac{\partial B}{\partial x}\Big] \tag{3}$$

Derivative of log det $X^T X + \delta I$ ... 2

Consider equation (2) : $\dfrac{\partial \log \det B}{\partial x} = \dfrac{\partial \log \det B}{\partial \det B} \dfrac{\partial \det B}{\partial x}$.

$\frac{\partial \log \det B}{\partial \det B}$ is just a simple log differentiation as $\det B$ is a scalar

$$\frac{\partial \log \det B}{\partial \det B} = \frac{1}{\det B} \tag{4}$$

For $\frac{\partial \det B}{\partial x}$, apply Jacobi's Formula

$$\frac{\partial \det B}{\partial x} = \det B \, \mathsf{Tr}\left(B^{-1} \frac{\partial B}{\partial x}\right) \tag{5}$$

Put (4),(5) to (2) :

$$\frac{\partial \log \det B}{\partial x} = \mathsf{Tr}\left(B^{-1} \frac{\partial B}{\partial x}\right)$$

Derivative of log det $X^TX + \delta I$ ... 3

Put $B = X^TX + \delta I$ back ,we have

$$\frac{\partial \log\det(X^TX + \delta I)}{\partial x} = \mathsf{Tr}\Big((X^TX + \delta I)^{-1}\frac{\partial (X^TX + \delta I)}{\partial x}\Big)$$

As $\frac{\partial (X^TX+\delta I)}{\partial x} = \frac{\partial X^TX}{\partial x} + \frac{\partial \delta I}{\partial x}$ and $\frac{\partial \delta I}{\partial x}$ gives zero matrix so

$$\frac{\partial \log\det(X^TX + \delta I)}{\partial x} = \mathsf{Tr}\Big((X^TX + \delta I)^{-1}\frac{\partial X^TX}{\partial x}\Big) \tag{6}$$

Again, note that $\log\det(.)$ and x are scalars but $(X^TX + \delta I)^{-1}$ and $\dfrac{\partial X^TX}{\partial x}$ are matrices. It is the trace operator turns the matrix $(X^TX + \delta I)^{-1}\dfrac{\partial X^TX}{\partial x}$ back to scalar so that equation (6) makes sense!

Derivative of log det $X^T X + \delta I$ ... 4

Now consider another way to show the same result. Consider (3) :

$$\frac{\partial \log \det B}{\partial x} = \mathsf{Tr}\Big[\Big(\frac{\partial \log \det B}{\partial B}\Big)^T \frac{\partial B}{\partial x}\Big]$$

Since $\frac{\partial \log \det B}{\partial B} = B^{-T}$ (page5), so we have

$$\frac{\partial \log \det B}{\partial x} = \mathsf{Tr}\Big[B^{-1} \frac{\partial B}{\partial x}\Big]$$

From last page we have $\frac{\partial B}{\partial x} = \frac{\partial X^T X}{\partial x}$, so

$$\frac{\partial \log \det B}{\partial x} = \mathsf{Tr}\Big[B^{-1} \frac{\partial X^T X}{\partial x}\Big]$$

Put back $B = X^T X + \delta I$ we get the same result as last page.

Derivative of log det $X^T X + \delta I$ ... 5

So we have $\dfrac{\partial \log\det(X^T X + \delta I)}{\partial x} = \text{Tr}\Big[(X^T X + \delta I)^{-1} \dfrac{\partial X^T X}{\partial x}\Big]$

Put $x = X_{ij}$ we have $\dfrac{\partial X^T X}{\partial X_{ij}} = X^T J^{ij} + J^{ji} X$ where J^{ij} is single entry matrix. We get

$$\begin{aligned}
\frac{\partial \log\det(X^T X + \delta I)}{\partial X_{ij}} &= \text{Tr}\Big[(X^T X + \delta I)^{-1}(X^T J^{ij} + J^{ij} X)\Big] \\
&= \text{Tr}\Big[(X^T X + \delta I)^{-1} X^T J^{ij}\Big] + \text{Tr}\Big[(X^T X + \delta I)^{-1} J^{ij} X)\Big]
\end{aligned}$$

By $\text{Tr}(AJ^{ij}) = [A^T]_{ij}$ and $\text{Tr}(AJ^{ji}B) = [BA]_{ij}$, we have

$$\frac{\partial \log\det(X^T X + \delta I)}{\partial X_{ij}} = \Big[(X^T X + \delta I)^{-1} X^T\Big]^T_{ij} + \Big[X(X^T X + \delta I)^{-1}\Big]_{ij}$$

As $X^T X + \delta I$ is symmetric so

$$\frac{\partial \log\det(X^T X + \delta I)}{\partial X} = 2X(X^T X + \delta I)^{-1}$$

Last page - summary

- Physical meanings of $\det X$ and $\log\det X$
- $\log\det X$ is concave in X
- Jacobi Formula and derivatives w.r.t. matrix variable X of $\det X$, $\log\det X$ and $\log\det(X^T X + \delta I)$

End of document