Derivative of $\log$ determinant

Sungchan Yi

August 9th, 2020

이 포스트에서 증명하려는 등식은 Bishop의 PRML 책의 부록 C에서 발견한 다음 등식이다.

Invertible matrix $\mathbf{A}$에 대하여 다음이 성립한다.

$$\frac{\partial}{\partial x} \log(\det \mathbf{A}) = \mathrm{tr}\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\right) \tag{1}$$

PRML 책에서는 증명을 위해 다음 결과들을 활용하라고 친절하게 적어주었다.

Real symmetric matrix $\mathbf{A}$의 orthonormal eigenvector $\mathbf{u}_i$에 대하여 다음이 성립한다.

$$\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_j = \delta_{ij} \tag{2}$$

단, δ_{ij}는 Kronecker delta function 이다.

Real symmetric $n \times n$ matrix $\mathbf{A}$를 orthogonal matrix 로 diagonalize 하여 $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}$ 로 나타내면, ($\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_i, \dots, \lambda_n)$) $\mathbf{U}$의 i-번째 열 $\mathbf{u}_i$ 에 대하여

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \tag{3}$$

로 나타낼 수 있으며, $\mathbf{A}$가 invertible 이면 $\mathbf{A}^{-1} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^{\mathrm{T}}$ 이므로,

$$\mathbf{A}^{-1} = \sum_{i=1}^{n} \lambda_i^{-1} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}} \tag{4}$$

임을 알 수 있고, $\det \mathbf{A} = \det(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}) = \det(\mathbf{U}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Lambda}) = \det \boldsymbol{\Lambda}$ 으로부터

$$\det \mathbf{A} = \prod_{i=1}^{n} \lambda_i \tag{5}$$

이다.

주어진 내용들을 바탕으로 증명을 시도해 보자.

먼저 (1)의 좌변을 살펴보자. 당연히 (5)를 적용해야 할 것 처럼 생겼다.

$$\frac{\partial}{\partial x}\log(\det \mathbf{A}) = \frac{\partial}{\partial x}\log\prod_{i=1}^{n}\lambda_i = \frac{\partial}{\partial x}\sum_{i=1}^{n}\log\lambda_i = \sum_{i=1}^{n}\frac{\partial}{\partial x}(\log\lambda_i) = \sum_{i=1}^{n}\frac{1}{\lambda_i}\frac{\partial\lambda_i}{\partial x}$$

인 것은 쉽게 알 수 있다. 마지막 등식에는 연쇄법칙(chain rule)이 적용되어 있다. 이제 (1)의 우변을 변형할 차례이다. 곱의 미분법을 이용해 (3)의 양변을 미분해 보면,

$$\frac{\partial\mathbf{A}}{\partial x} = \sum_{i=1}^{n}\frac{\partial}{\partial x}\left(\lambda_i\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right) = \sum_{i=1}^{n}\frac{\partial\lambda_i}{\partial x}(\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}) + \sum_{i=1}^{n}\lambda_i\frac{\partial}{\partial x}\left(\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)$$

임을 알 수 있다. 이제 양변의 왼쪽에 (4)를 곱한다.

$$\begin{aligned}
\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial x} &= \left(\sum_{i=1}^{n}\lambda_i^{-1}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\left(\sum_{i=1}^{n}\frac{\partial\lambda_i}{\partial x}(\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}) + \sum_{i=1}^{n}\lambda_i\frac{\partial}{\partial x}\left(\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right)\right)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{-1}\frac{\partial\lambda_j}{\partial x}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}} + \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{-1}\lambda_j\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\frac{\partial}{\partial x}\left(\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\right)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{-1}\frac{\partial\lambda_j}{\partial x}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}} + \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{-1}\lambda_j\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_j\frac{\partial\mathbf{u}_j^{\mathrm{T}}}{\partial x} + \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{-1}\lambda_j\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\frac{\partial u_j}{\partial x}\mathbf{u}_j^{\mathrm{T}}
\end{aligned}$$

마지막 등식에는 곱의 법칙을 한 번 더 적용했다. 앞 두 항의 경우 $\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_j$ 가 포함되어 있는데, (2)에 의해 $i = j$ 인 경우만 남고 나머지는 모두 0이므로, 간단하게 할 수 있다.

$$\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial x} = \sum_{i=1}^{n}\lambda_i^{-1}\frac{\partial\lambda_i}{\partial x}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}} + \sum_{i=1}^{n}\mathbf{u}_i\frac{\partial\mathbf{u}_i^{\mathrm{T}}}{\partial x} + \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{-1}\lambda_j\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\frac{\partial u_j}{\partial x}\mathbf{u}_j^{\mathrm{T}}$$

이제 양변에 tr (대각합) 를 씌우면,

$$\mathrm{tr}\left(\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial x}\right) = \mathrm{tr}\left(\sum_{i=1}^{n}\lambda_i^{-1}\frac{\partial\lambda_i}{\partial x}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\right) + \mathrm{tr}\left(\sum_{i=1}^{n}\mathbf{u}_i\frac{\partial\mathbf{u}_i^{\mathrm{T}}}{\partial x}\right) + \mathrm{tr}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{-1}\lambda_j\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\frac{\partial u_j}{\partial x}\mathbf{u}_j^{\mathrm{T}}\right)$$

와 같이 된다. 식을 간단히 하기 위해, tr 의 중요한 성질인 $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ 를 이용하여 벡터들의 순서를 적당히 바꿔줘야 한다. 그리고 (2)의 성질을 다시 적용하여 정리한다.

$$\begin{aligned}
\mathrm{tr}\left(\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial x}\right) &= \mathrm{tr}\left(\sum_{i=1}^{n}\lambda_i^{-1}\frac{\partial\lambda_i}{\partial x}\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i\right) + \mathrm{tr}\left(\sum_{i=1}^{n}\frac{\partial\mathbf{u}_i^{\mathrm{T}}}{\partial x}\mathbf{u}_i\right) + \mathrm{tr}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{-1}\lambda_j\mathbf{u}_j^{\mathrm{T}}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}}\frac{\partial u_j}{\partial x}\right)\\
&= \mathrm{tr}\left(\sum_{i=1}^{n}\lambda_i^{-1}\frac{\partial\lambda_i}{\partial x}\mathbf{I}_n\right) + \mathrm{tr}\left(\sum_{i=1}^{n}\frac{\partial\mathbf{u}_i^{\mathrm{T}}}{\partial x}\mathbf{u}_i\right) + \mathrm{tr}\left(\sum_{i=1}^{n}\mathbf{u}_i^{\mathrm{T}}\frac{\partial u_i}{\partial x}\right)
\end{aligned}$$

이제 마지막 두 항은 곱의 미분법을 이용하여 묶어주고 (2)를 다시 적용하면

$$\operatorname{tr}\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\right) = \sum_{i=1}^{n} \frac{1}{\lambda_i}\frac{\partial \lambda_i}{\partial x} + \operatorname{tr}\left(\sum_{i=1}^{n} \frac{\partial}{\partial x}\left(\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i\right)\right) = \sum_{i=1}^{n} \frac{1}{\lambda_i}\frac{\partial \lambda_i}{\partial x} + \operatorname{tr}\left(\sum_{i=1}^{n} \frac{\partial \mathbf{I}}{\partial x}\right) = \sum_{i=1}^{n} \frac{1}{\lambda_i}\frac{\partial \lambda_i}{\partial x}$$

가 되어 원하는 결론을 얻는다.

다만 이 증명에는 약간의 문제가 있다. 증명 과정에서 (3), (4)의 성질을 이용했는데, 이 성질들을 이용하려면 $\mathbf{A}$가 diagonalizable 이라는 조건이 필요하다.[1] 하지만 이는 일반적으로 참이 아니다.

다른 방법으로 증명을 시도해 보자.

행렬 $\mathbf{A}$ 에 대해 다음 등식이 성립함을 알 수 있다.

$$\det(\mathbf{I} + \epsilon \mathbf{A}) = 1 + \epsilon \operatorname{tr} \mathbf{A} + O(\epsilon^2)$$

형태가 왠지 characteristic polynomial 과 유사하다. Characteristic polynomial 은 $p_{\mathbf{A}}(t) = \det(t\mathbf{I} - \mathbf{A})$ 로 정의되지만, characteristic polynomial 에서 t^{n-1}의 계수가 $\det(\mathbf{I} + \epsilon\mathbf{A})$ 에서 ϵ의 계수와 절댓값이 같다는 것은 쉽게 알 수 있다. 그리고 characteristic polynomial 에서 t^{n-1}의 계수는 모든 eigenvalue 들의 합이자, $\operatorname{tr} \mathbf{A}$ 와 일치한다.[2]

이제 만약 $\mathbf{A}$가 invertible 이면, $\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B}$ 을 이용하여,

$$\det(A + \epsilon \mathbf{B}) = \det \mathbf{A} \cdot \det\left(\mathbf{I} + \epsilon \mathbf{A}^{-1}\mathbf{B}\right) = \det \mathbf{A} \cdot (1 + \epsilon \operatorname{tr}(\mathbf{A}^{-1}\mathbf{B}) + O(\epsilon^2))$$

임을 알 수 있다. 따라서,

$$\frac{\partial}{\partial x}\det \mathbf{A} = \lim_{\epsilon \to 0} \frac{\det\left(\mathbf{A} + \epsilon \dfrac{\partial \mathbf{A}}{\partial x} + O(\epsilon^2)\right) - \det \mathbf{A}}{\epsilon} = \det \mathbf{A} \cdot \operatorname{tr}\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\right)$$

라는 식을 얻는다. 마지막으로 양변을 $\det \mathbf{A}$ 로 나누면 좌변이 정확히 (1)의 좌변과 일치한다.

$$\frac{\partial}{\partial x}\log(\det \mathbf{A}) = \frac{\dfrac{\partial}{\partial x}\det \mathbf{A}}{\det \mathbf{A}} = \operatorname{tr}\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\right)$$

따라서 등식 (1)은 invertible matrix $\mathbf{A}$에 대하여 성립한다. □

[1] 더 정확히는 $\mathbf{A}$의 eigenvector들로 이루어진 orthonormal basis가 존재한다는 조건이 필요하다.

[2] 증명은 triangularization 을 이용하면 간편하다.