

Constructing an Explicit, Always Invertible Matrix

Jianlin Su

March 1, 2019

From the article "[Appreciation of the Identity \$\det\(\exp\(\mathbf{A}\)\) = \exp\(\text{Tr}\(\mathbf{A}\)\)\$](#) ", we learned that the matrix $\exp(\mathbf{A})$ is always invertible, and its inverse is simply $\exp(-\mathbf{A})$. The issue is that $\exp(\mathbf{A})$ is primarily a theoretical definition; writing it this way has little practical value because it requires calculating every power $\mathbf{A}^n$.

Are there any concrete examples? Yes. In this post, we will construct an explicit matrix that is always invertible.

The idea is actually very simple. Suppose $\mathbf{x}$ and $\mathbf{y}$ are two k -dimensional column vectors. Then $\mathbf{xy}^\top$ is a $k \times k$ matrix. Let us consider:

$$\begin{aligned}\exp(\mathbf{xy}^\top) &= \sum_{n=0}^{\infty} \frac{(\mathbf{xy}^\top)^n}{n!} \\ &= \mathbf{I} + \mathbf{xy}^\top + \frac{\mathbf{xy}^\top \mathbf{xy}^\top}{2} + \frac{\mathbf{xy}^\top \mathbf{xy}^\top \mathbf{xy}^\top}{6} + \dots\end{aligned}\tag{1}$$

Notice that:

$$\mathbf{y}^\top \mathbf{x} = \langle \mathbf{x}, \mathbf{y} \rangle\tag{2}$$

is actually just a scalar. Therefore, we can simplify the expression further:

$$\begin{aligned}\exp(\mathbf{xy}^\top) &= \mathbf{I} + \mathbf{xy}^\top \left(1 + \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{2} + \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{6} + \dots \right) \\ &= \mathbf{I} + \mathbf{xy}^\top \left(\frac{e^{\langle \mathbf{x}, \mathbf{y} \rangle} - 1}{\langle \mathbf{x}, \mathbf{y} \rangle} \right)\end{aligned}\tag{3}$$

Now this matrix is very concrete. It can be calculated quite easily because it only involves scalar exponential operations. The term $(e^x - 1)/x$ in the parentheses has a removable discontinuity at $x = 0$; when $x = 0$, its value is 1.

According to the identity $\det(\exp(\mathbf{A})) = \exp(\text{Tr}(\mathbf{A}))$, the determinant of this matrix is:

$$\det(\exp(\mathbf{xy}^\top)) = e^{\langle \mathbf{x}, \mathbf{y} \rangle}\tag{4}$$

Its inverse matrix is:

$$\exp(-\mathbf{xy}^\top) = \mathbf{I} - \mathbf{xy}^\top \left(\frac{1 - e^{-\langle \mathbf{x}, \mathbf{y} \rangle}}{\langle \mathbf{x}, \mathbf{y} \rangle} \right)\tag{5}$$

which is also an explicit result.

Of course, a general matrix has k^2 independent parameters, whereas the matrix constructed here from two vectors has only $2k$ parameters. In terms of expressive power, this is likely insufficient. To enhance the expressivity, one could consider multiplying several such matrices together:

$$\exp(\mathbf{x}_1 \mathbf{y}_1^\top) \exp(\mathbf{x}_2 \mathbf{y}_2^\top) \exp(\mathbf{x}_3 \mathbf{y}_3^\top) \dots\tag{6}$$

Note that this is generally not equal to:

$$\exp \left(\boldsymbol{x}_1 \boldsymbol{y}_1^\top + \boldsymbol{x}_2 \boldsymbol{y}_2^\top + \boldsymbol{x}_3 \boldsymbol{y}_3^\top + \dots \right) \quad (7)$$

Well, after all this derivation, what is its actual use?

Uh... I don't know what it's useful for yet; just treat it as something to appreciate.

(Actually, my original intention was to use a constructive method to directly build invertible neural networks. Based on this ready-made matrix, constructing an invertible fully connected network is not difficult, but I haven't figured out how to generalize it to convolutional layers. Once I have settled that generalization, I will come back to discuss the applications of this matrix!)

When reprinting, please include the original article address:
<https://kexue.fm/archives/6407>

For more detailed reprinting matters, please refer to: "Scientific Space FAQ"