- **Email**: chrisbrown@utexas.edu
- **EID**: chb595


# 1 Inverse Transpose

I will show that

$$\frac{\partial ln|A|}{\partial A} = (A^{-1})^T$$

We'll name $|x|$ the function $g(x)$, and $ln(x)$ the function $f(x)$, and declaring $F(x) = f(g(x))$, the chain rule says that $\frac{F'(x)}{x'} = \frac{f'(g(x))}{g'}\frac{g'(x)}{x'}$.

Thus the term on the left side of the equation, $\frac{\partial(f(g(A)))}{\partial A}$ derives to:

$$\frac{\partial f(g(A))}{\partial g(A)}\frac{\partial g(A)}{\partial A}$$

And back in the original terms:

$$\frac{\partial ln|A|}{\partial |A|}\frac{\partial |A|}{\partial A}$$

Now since $\frac{d}{dx}lnx = \frac{1}{x}$, and since $\partial[constant] = 1$, and $|A|$ is just a constant, we just get:

$$\frac{1}{|A|}\frac{\partial |A|}{\partial A}$$

The fraction on the right hand side can alternatively be notated: $\nabla_A|A|$, or written out as:

$$\begin{bmatrix} \frac{\partial|A|}{\partial A_{11}} & \frac{\partial|A|}{\partial A_{12}} & \cdots \\ \frac{\partial|A|}{\partial A_{21}} & \ddots & \cdots \\ \vdots & \vdots & \frac{\partial|A|}{\partial A_{mn}} \end{bmatrix}$$

As in Andrew Ng's course notes, we define the adjoint as $(\text{adjoint}(A))_{ij} = (-1)^{i+j}|A_{\setminus j,\setminus i}|$.

And keep that definition of the determinant: $|A| = \sum_{i=1}^{n} a_{ij}(-1)^{i+j}|A_{\setminus i,\setminus j}|$

Let's transpose the determinant in the definition of the adjoint: $(\text{adjoint}(A))_{ij} = (-1)^{i+j}|(A^T)_{\setminus i,\setminus j}|$

Which of course we can rewrite as: $(\text{adjoint}(A^T))_{ij} = (-1)^{i+j}|(A)_{\setminus i,\setminus j}|$.

And so we can redefine the determinant as: $|A| = \sum_{i=1}^{n} a_{ij}\text{adjoint}(A^T)$

And now it's relatively clear why $\nabla_A|A| = \text{adjoint}(A)$

By definition of the determinant, $\text{adjoint}(A)A = |A|I$. This can be reordered: $\text{adjoint}(A)A = |A|I \rightarrow \frac{\text{adjoint}(A)A}{|A|} = I$ and since $|A|$ is a scalar, it doesn't really matter where it goes. Then we get $\frac{\text{adjoint}(A)AA^{-1}}{|A|} = IA^{-1} \rightarrow \frac{\text{adjoint}(A)}{|A|} = A^{-1}$.

Now since $\nabla_A|A| = \text{adjoint}(A)$, we get $\frac{\nabla_A|A|}{|A|} = A^{-1}$.

# 2 Positive eigenvalues

For a matrix $A$ and any eigenvector $x$ and eigenvalue $\lambda$:

$$Ax = \lambda x$$

A (square) positive definite matrix $A$ is defined as: $x^T A x > 0$, where $x$ is any vector of the same width as $A$ (and $x$ is non-zero). We wish to show that if $x$ is an eigenvector of $A$, which is positive definite, then the corresponding eigenvalue $\lambda$ is positive.

First, multiply both sides of the eigenvector equation by $x^T$:

$$x^T A x = x^T \lambda x$$

Because we presume that $A$ is positive definite, and we are dealing with non-zero eigenvectors ($x$), we see that the left side of this equation must be greater than zero, and so must be the right side:

$$x^T \lambda x > 0$$

Now, since $\lambda$ is just a scalar number, we can move it over:

$$x^T x \lambda > 0$$

A vector multiplied by its transpose is a sum of squares:

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

We're only dealing with real numbers, so our squares must be non-negative. But we've also asserted that our vector $x$ is non-zero, so the sum of our squares must be positive (non-zero). We declare that $c = x^T x$,

$$c\lambda > 0$$

And since we've shown that $c > 0$, divide both sides by $c$, and we're done:

$$\lambda > 0$$

Of course, this works for any eigenvalue $\lambda$, since it works for the general case.

# 3 Matrix Algebra

We will show:

$$(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}$$

Begin by multiplying both sides by $(A + BD^{-1}C)$

$$(A + BD^{-1}C)^{-1}(A + BD^{-1}C) \stackrel{?}{=} (A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1})(A + BD^{-1}C)$$

The left side reduces to unity:

$$I \stackrel{?}{=} (A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1})(A + BD^{-1}C)$$

Factor out the $A^{-1}$ on the right side:

$$I \stackrel{?}{=} (A^{-1})(1 - B(D + CA^{-1}B)^{-1}CA^{-1})(A + BD^{-1}C)$$

Now distribute the $(A + BD^{-1}C)$ term.

$$I \stackrel{?}{=} (A^{-1})(A + BD^{-1}C - B(D + CA^{-1}B)^{-1}CA^{-1}A - B(D + CA^{-1}B)^{-1}CA^{-1}BD^{-1}C)$$

Simplify out $A^{-1}A$.

$$I \stackrel{?}{=} (A^{-1})(A + BD^{-1}C - B(D + CA^{-1}B)^{-1}C - B(D + CA^{-1}B)^{-1}CA^{-1}BD^{-1}C)$$

Factor out $B$ to the left, and $C$ to the right.

$$I \stackrel{?}{=} (A^{-1})(A + B(D^{-1} - (D + CA^{-1}B)^{-1} - (D + CA^{-1}B)^{-1}CA^{-1}BD^{-1})C)$$

Factor out $(D + CA^{-1}B)^{-1}$:

$$I \stackrel{?}{=} (A^{-1})(A + B(D^{-1} - (D + CA^{-1}B)^{-1}(I + CA^{-1}BD^{-1}))C)$$

And inject a $DD^{-1}$:

$$I \stackrel{?}{=} (A^{-1})(A + B(D^{-1} - (D + CA^{-1}B)^{-1}(I + CA^{-1}BD^{-1})DD^{-1})C)$$

And distribute towards the left:

$$I \stackrel{?}{=} (A^{-1})(A + B(D^{-1} - (D + CA^{-1}B)^{-1}(D + CA^{-1}BD^{-1}D)D^{-1})C)$$

The $D^{-1}D$ cancels to $I$:

$$I \stackrel{?}{=} (A^{-1})(A + B(D^{-1} - (D + CA^{-1}B)^{-1}(D + CA^{-1}B)D^{-1})C)$$

And $(D + CA^{-1}B)^{-1}(D + CA^{-1}B)$ also cancels to $I$:

$$I \stackrel{?}{=} (A^{-1})(A + B(D^{-1} - D^{-1})C)$$

I'll multiply it out, just to be clear. First, $A^{-1}$:

$$I \stackrel{?}{=} A^{-1}A + A^{-1}B(D^{-1} - D^{-1})C$$

Simplify the left term, distribute the right:

$$I \stackrel{?}{=} I + A^{-1}BD^{-1}C - A^{-1}BD^{-1}C$$

Subtract $I$ from both sides:

$$0 \stackrel{?}{=} A^{-1}BD^{-1}C - A^{-1}BD^{-1}C$$

Add $A^{-1}BD^{-1}C$ to both sides, and the equality is obvious.

$$A^{-1}BD^{-1}C = A^{-1}BD^{-1}C$$

# 4 Lagrange Can

For a can of radius $r$ and height $h$, we want to pick the $r$ and $h$ that maximize volume,

$$V = \pi r^2 h$$

while constraining surface area to a given constant.

Because 'can' suggests that the intended solid is not precisely a cylinder, I will proceed under the assumption that the solid in question is a hollow cylinder with one end missing. Of course, this only affects surface area, not volume. For simplicity, I'll assume the walls of the solid are infinitely thin.

$$A = \pi r^2 + 2\pi r h$$

Each surface could be multiplied by two for the inside and outside of the same wall, but the end result would be the same (we would just have to divide $A$ by 2 before we use it), so I will stick with counting only the outside surfaces.

The constraint means that, since we know $A$, given $r$, we can determine $h$ (or given $h$, we can determine $r$). Now we have:

$$\max_{r,h} V = \pi r^2 h \qquad \text{subject to} \qquad \pi r^2 + 2\pi r h = A$$

One constraint means one Lagrange multiplier. We set up the optimization equation:

$$\max_{r,h,\lambda} V = \pi r^2 h + \lambda(\pi r^2 + 2\pi r h - A)$$

Now we differentiate:

$$\begin{aligned}
V'_h &= \pi r^2 + 2\pi\lambda r && = 0 = r + 2\lambda \\
V'_r &= 2\pi r h + 2\pi\lambda r + 2\pi\lambda h && = 0 = rh + \lambda r + \lambda h \\
V'_\lambda &= \pi r^2 + 2\pi r h - A && = 0
\end{aligned}$$

So from the first and second equations we have:

$$\begin{aligned}
r &= -2\lambda \\
h &= \frac{-\lambda r}{r + \lambda}
\end{aligned}$$

Merging the first and second equations produces:

$$h = -2\lambda$$

And substituting these values of $r$ and $h$ into the final equation produces

$$0 = 4\pi\lambda^2 + 8\pi\lambda^2 - A = 12\pi\lambda^2 - A$$

Which rearranges into

$$12\pi\lambda^2 = A \qquad \text{and} \qquad \lambda = \sqrt{\frac{A}{12\pi}}$$

Multiplying this back into the equations for $r$ and $h$, we get

$$r = h = -2\sqrt{\frac{A}{12\pi}}$$

It's absurd to have negative dimensions, but let's continue, to derive the volume:

$$V = -\frac{2A}{3}\sqrt{\frac{A}{12\pi}}$$

Yep, it's still kind of strange, so apparently it *is* important to subtract the Lagrangian component when maximizing, instead of adding. (And now I'm perplexed why the box example in class (Feb. 6) worked out just fine by adding, even though it was a maximization problem, too.)

In which case we have:

$$\max_{r,\lambda} V = \pi r^2 h - \lambda(\pi r^2 + 2\pi r h - A)$$

And:

$$\begin{aligned}
V_h' &= \pi r^2 - 2\pi\lambda r & &= 0 = r - 2\lambda \\
V_r' &= 2\pi r h - 2\pi\lambda r - 2\pi\lambda h & &= 0 = rh - \lambda r - \lambda h \\
V_\lambda' &= -\pi r^2 - 2\pi r h + A & &= 0 = \pi r^2 + 2\pi r h - A
\end{aligned}$$

And:

$$r = h = 2\lambda$$

And substituting into the third equation gives us the same $\lambda$,

$$\lambda = \sqrt{\frac{A}{12\pi}}$$

But now of course we'll get positive $r$ and $h$ values:

$$r = h = 2\sqrt{\frac{A}{12\pi}}$$

And a reasonable volume:

$$V = \frac{2A}{3}\sqrt{\frac{A}{12\pi}}$$

For the *un*opened can scenario, $r = 2\lambda$ and $h = 4\lambda$, and $\lambda = \sqrt{\frac{A}{24\pi}}$. In which case

$$V = \pi\left(2\sqrt{\frac{A}{24\pi}}\right)^2\left(4\sqrt{\frac{A}{24\pi}}\right) = \frac{2A}{3}\sqrt{\frac{A}{24\pi}}$$

# 5 Cost + Reward function optimization

We begin with the dynamic equation (acceleration is a function of gas applied at some time):

$$\ddot{x} = u(t)$$

and initial conditions (which just say we start at the beginning, with a speed of zero):

$$x = 0$$
$$\dot{x} = 0$$

and the reward - cost function:

$$J = x(T) - \int_0^T u^2(t)dt$$

In that equation, $x(T)$ represents the distance traveled at time $T$ (the bigger the better). And $\frac{1}{2}u^2(t)$ represents the acceleration (amount of gas) applied at point $t$, and the integration of that function from time $=0$ to time $=T$ is the total amount of power ($=$ gas) used between the starting point and the end time, but squared in order to encourage a gradual application of power.

$\dot{x}$ is a vector, which consists of two scalars, which correspond to the derivatives of position and velocity:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} \begin{pmatrix} x_2 \\ \dot{x}_2 \end{pmatrix}$$