# Jacobi's Formula for the Derivative of a Determinant

Jacobi's formula is $d\,\det(B) = \mathrm{Trace}(\mathrm{Adj}(B)\, dB)$ in which Adj(B) is the *Adjugate* of the square matrix $B$ and $dB$ is its differential. This formula will be derived and then applied to …

- the rôle of the Wronskian in the solution of linear differential equations,
- the derivative of a simple eigenvalue, and
- inverses of nearly singular matrices.

Certain definitions and formulas will be taken for granted. Given an n-by-n matrix $B = \{ß_{ij}\}$, its *Classical Adjoint* or ( better ) *Adjugate* $\mathrm{Adj}(B) = A = \{\alpha_{ij}\}$ is defined thus:

$$\alpha_{ij} = (-1)^{i-j} \det(\text{ B without its } j^{th} \text{ row and } i^{th} \text{ column })$$

so that $AB = BA = \det(B)I$. In other words, $\alpha_{ij}$ is the *cofactor* of $ß_{ji}$ in $\det(B)$, so $\alpha_{ij}$ is a polynomial function of the elements of $B$ but independent of $ß_{jk}$ and $ß_{ki}$ for all $k$; and $\sum_k \alpha_{ik} ß_{kj} = \sum_k ß_{ik} \alpha_{kj} = \det(B)$ if $i = j$ but $0$ otherwise. Rank(B) is the biggest dimension of a submatrix of $B$ whose determinant is nonzero; $\mathrm{Rank}(B) = n$ just when $\det(B) \neq 0$, in which case $\mathrm{Adj}(B) = \det(B)B^{-1}$; and $\mathrm{Adj}(B) \neq O$ just when $\mathrm{Rank}(B) \geq n-1$. If $B$ is differentiable and $\mathrm{Rank}(B) = n$ then $d(B^{-1}) = -B^{-1}(dB)B^{-1}$. Also needed is the function $\mathrm{Trace}(B) = \sum_j ß_{jj}$, and the fact that $\mathrm{Trace}(PQ) = \mathrm{Trace}(QP)$. For proofs of these formulas see any text on matrices or linear algebra.

## Proof of Jacobi's Formula:

In $\det(B) = \sum_k ß_{ik} \alpha_{ki}$, each element $ß_{ij}$ of $B$ appears linearly multiplied by its cofactor $\alpha_{ji}$, so $\partial\det(B)/\partial ß_{ij} = \alpha_{ji}$; this leads quickly to Jacobi's Formula

$$d\,\det(B) = \sum_j \sum_i (\partial\det(B)/\partial ß_{ij}) dß_{ij} = \sum_j \sum_i \alpha_{ji} dß_{ij} = \mathrm{Trace}(\mathrm{Adj}(B)\, dB).$$

## The Wronskian:

Consider square matrix solutions $X(\tau)$ of a linear differential equation $dX/d\tau = L(\tau)X$ with a piecewise continuous coefficient matrix $L(\tau)$. Because $L(\tau)$ is *not* assumed to *commute* with $L(\theta)$ when $\theta \neq \tau$ ( i.e., $L(\tau)L(\theta) \neq L(\theta)L(\tau)$ ), $\exp(\int_o^\tau L(\theta)\, d\theta)$ need *not* be a solution $X(\tau)$ of the differential equation. None the less a linear differential equation for the *Wronskian* $\det(X(\tau))$ can be found and solved to prove assertions about *Fundamental Solutions* found in many texts about differential equations. Begin with the observation that solutions $X(\tau)$ satisfy

$$d\,\det(X)/d\tau = \mathrm{Trace}(\mathrm{Adj}(X)\, dX/d\tau) = \mathrm{Trace}(\mathrm{Adj}(X)\, L\, X) = \mathrm{Trace}(X\, \mathrm{Adj}(X)\, L)$$
$$= \det(X)\, \mathrm{Trace}(L),$$

which can be solved for

$$\det(X(\tau)) = \exp\left( \int_o^\tau \mathrm{Trace}(L(\theta))\, d\theta \right) \det(X(0)).$$

This implies that $\det(X(\tau))$ is nonzero for *all* $\tau$ if nonzero for *any* $\tau$. Consequently $X(\tau)$ is invertible for all $\tau$ if invertible for any $\tau$, in which case $X(\tau)$ is called a *Fundamental Solution* since every vector solution $x$ of $dx/d\tau = Lx$ has the form $x = Xc$ for some constant vector $c$. To verify this claim, observe that $X^{-1}x$ must be constant because

$$Xd(X^{-1}x)/dt = X(dX^{-1}/dt)x + dx/dt = -(dX/dt)X^{-1}x + Lx = -LXX^{-1}x + Lx = o.$$

( Strictly speaking, $\det(X(\tau))$ is the Wronskian of the columns of $X(\tau)$ .)

Fundamental solutions $X(\tau)$ of $dX/d\tau = LX$ with their nonvanishing Wronskians figure in the solution of the non-homogeneous linear differential equation $dx/d\tau = Lx + f$ by the *Method of Variation of Parameters*. The idea is to substitute $x = Xp$ into the last differential equation and solve it for the parameter vector $p$. This substitution yields $dp(\tau)/d\tau = X(\tau)^{-1}f(\tau)$, whence follows $x(\tau) = X(\tau)( c + \int^{\tau} X(\theta)^{-1}f(\theta)\, d\theta )$ for any constant vector $c$.

Exercise: Given smooth scalar functions $h(\tau)$ and $g(\tau)$, the homogeneous second-order differential equation $y'' - hy' - gy = 0$ can, in principle, be solved for scalar solutions $y(\tau)$. The Wronskian of any two solutions $y_1$ and $y_2$ is

$$W(y_1(\tau), y_2(\tau)) := y_1(\tau)y_2'(\tau) - y_1'(\tau)y_2(\tau) .$$

Show that $W = \exp( \int h(\tau)\, d\tau + \text{constant} )$. Provided it is nonzero, it appears as a divisor in expressions for solutions $z(\tau)$ of $z'' - hz' - gz = e(\tau)$ obtained in textbooks by variation of parameters. Find those expressions, and rederive them after first converting the scalar second-order differential equations to 2-vector first-order differential equations.

## Adjugates of Singular Matrices:

Almost all n-by-n matrices are nonsingular because the singular matrices $B$, satisfying the equation $\det(B) = 0$, lie in a hypersurface of dimension $n^2 - 1$ in the $n^2$-dimensional space of n-by-n matrices. Among the singular matrices, almost all have rank $n-1$; the matrices of rank less than $n-1$ lie in a hypersurface of dimension $n^2 - 4$ embedded in the hypersurface of singular matrices.

To see why the foregoing assertions are true, consider a matrix $C$ of rank $n-2$, and suppose for the sake of simplicity that its first $n-2$ columns are linearly independent. These columns can be embedded in a basis of $n$ columns for the n-dimensional space of n-columns, and then the last two columns of $C$ can be any two columns whose expressions in that basis have zeros for their last two elements. If any one of those four zero elements were replaced by a nonzero then Rank(C) would increase to $n - 1$. Thus, the matrices $C$ of rank $n-2$ ( or less ) must satisfy four equations in $n^2$ unknowns, whereas the matrices of rank $n-1$ ( or less ) need satisfy just one equation $\det(\dots) = 0$. It turns out that the $(n^2 - 1)$-dimensional hypersurface of singular matrices intersects itself in the $(n^2 - 4)$-dimensional hypersurface of matrices of rank $n-2$ or less, but we shall not prove nor need this,

Let $\text{Rank}(B) = n-1$, so $A = \text{Adj}(B) \neq O$. But since $BA = O$, all the columns of $A$ must lie in the one-dimensional null-space of $B$, and similarly for the rows of $A$, so $A = vu^T \neq O$ where $Bv = o$ and $u^TB = o^T$. In other words,

> almost all singular matrices $B$ have adjugates of rank one; $\text{Adj}(B) = vu^T \neq O$
> for suitable eigenvectors $v$ and $u^T$ belonging to $B$'s eigenvalue $0$.

The exceptional n-by-n singular matrices $B$ have $\text{rank}(B) < n-1$ and $\text{Adj}(B) = O$.

## The Derivative of a Simple Eigenvalue:

Suppose ß is a simple eigenvalue of a matrix B . Replacing B by B – ßI allows us to assume that ß = 0 for the sake of simplicity; in other words we assume that 0 is a simple eigenvalue of B . This means that det(αI – B) vanishes when α = 0 , but because 0 is a simple eigenvalue the derivative d det(αI – B)/dα ≠ 0 when α = 0 . By Jacobi's formula,

$$d\,\det(\alpha I - B)/d\alpha = \mathrm{Trace}(\mathrm{Adj}(\alpha I - B)I)\ ,$$

so we infer that Trace(Adj(B)) ≠ 0 . This implies that $\mathrm{Adj}(B) = vu^T$ is of rank one, not zero, with eigenvectors v and $u^T$ belonging to B 's eigenvalue 0 ; and $u^Tv = \mathrm{Trace}(\mathrm{Adj}(B)) \ne 0$ .

In general, if $u^T$ is a row eigenvector and v a column eigenvector belonging to the same *simple* eigenvalue ß of a matrix, then $u^Tv \ne 0$ . This is important because it allows us to compute the differential of this simple eigenvalue using its eigenvectors as follows:

$$0 = \det(ßI - B)\ ,\quad \text{so differentiate this equation to get} \ldots$$
$$0 = d\,\det(ßI - B) = \mathrm{Trace}(\ \mathrm{Adj}(ßI - B)(dß - dB)\ )$$
$$= \mathrm{Trace}(\ (vu^T)(dß - dB)\ )\ ,\quad \text{where } Bv = ßv \text{ and } u^TB = ßu^T \text{ and } u^Tv \ne 0\ ,$$
$$= u^Tv\, dß - u^T(dB)v\ .$$

Therefore $dß = u^T(dB)v/u^Tv$ provides the derivative of a simple eigenvalue ß of B .

## The Derivative of the Adjugate:

We need the differential of A = Adj(B) . This is obtained easily when B is nonsingular, in which case $A = B^{-1}\det(B)$ and, since $d(B^{-1}) = -B^{-1}(dB)B^{-1}$ , it soon follows that d Adj(B) = S(B, dB) where, for n-by-n matrices Z ,

$$S(B, Z) := \big(\ \mathrm{Trace}(\mathrm{Adj}(B)\ Z)\ \mathrm{Adj}(B) - \mathrm{Adj}(B)\ Z\ \mathrm{Adj}(B)\ \big)/\det(B)$$

provided det(B) ≠ 0 . This formula is a little misleading because it involves division by det(B) and degenerates into 0/0 when B is singular. In fact Adj(B) is a polynomial function of the elements of B , so its differential d Adj(B) = S(B, dB) is also a polynomial function of the elements of B and linear in dB regardless of whether B is singular.

This polynomial derivative of the adjugate figures in the determinant's second differential

$$d^2\det(B) = d\ \mathrm{Trace}(\mathrm{Adj}(B)dB) = \mathrm{Trace}(\ d(\mathrm{Adj}(B)dB)\ )$$
$$= \mathrm{Trace}(\ S(B, dB)\ dB + \mathrm{Adj}(B)d^2B\ )\ ,$$

and therefore figures also in the third term of the Taylor Series ( for any n-by-n Z )

$$\det(B + Z\tau) = \det(B) + \mathrm{Trace}(\ \mathrm{Adj}(B)Z\ )\tau + \mathrm{Trace}(S(B,Z)Z)\tau^2/2 + \ldots\ .$$

## Inverses of Nearly Singular Matrices:

Our final application of Jacobi's formula is a description of the behavior of inverses of matrices close to almost any singular matrix B . In particular, assume $\mathrm{Adj}(B) = vu^T \ne O$ , as is so for almost all singular matrices B , and let Z be a matrix with B 's dimensions. Now, $\mathrm{Trace}(\mathrm{Adj}(B)Z) = u^TZv$ is nonzero for almost all Z , and for those Z

$$\det(B + Z\tau) = u^TZv\tau + \mathrm{Trace}(S(B,Z)Z)\tau^2/2 + \ldots$$

is nonzero for all sufficiently small nonzero $\tau$ . For those $\tau$ and Z and B we find that

$$(B + Z\tau)^{-1} = Adj(B + Z\tau)/det(B + Z\tau)$$
$$= ( vu^T + S\tau + ... )/(u^TZv\tau + Trace(SZ)\tau^2/2 + ...)$$
$$= vu^T/(\tau\, u^TZv) + \big(u^TZvS - Trace(SZ)\, vu^T/2\big)/(u^TZv)^2 + ...$$

where S = S(B, Z) . Thus, as $\tau$ approaches zero and $B + Z\tau$ approaches almost any singular matrix B along almost any fixed direction Z in matrix space, $(B + Z\tau)^{-1}$ approaches infinity along a fixed direction parallel to $vu^T = Adj(B)$ in matrix space; in fact $(B + Z\tau)^{-1}$ is approximated by its leading term $vu^T/(\tau\, u^TZv)$ with ever smaller relative error, as $\tau$ approaches 0 , and with absolute error $(B + Z\tau)^{-1} - vu^T/(\tau\, u^TZv)$ that approaches a constant

$$S(B,Z)/(u^TZv) - Trace(S(B,Z)Z/2)\, vu^T/(u^TZv)^2 .$$

This last expression looks worse than it is, but its reduction to a manageable form must be deferred to another time. The important fact is that inversion maps almost all matrices in a sufficiently tiny ball, centered about almost any singular matrix B , to two narrow fingers reaching in from infinity in directions nearly parallel to known matrices ±Adj(B) of rank 1 .

Why should we care? One reason is that attempts to invert matrices B numerically usually yield approximations to $B^{-1}$ that are instead very much like $(B + Z\tau)^{-1}$ for some tiny $Z\tau$ comparable to roundoff in the elements of B . If B is very nearly ( or exactly ) singular, the computed inverse is recognizable as a matrix nearly proportional to Adj(B) of rank 1 ; the factor of proportionality is unpredictable except that it is huge, bigger than 1/( roundoff in B ) .

Another reason to care is that we often compute eigenvectors by solving a singular set of linear equations $(ßI - B)v \approx o$ for a nonzero vector v given a good approximation to the eigenvalue ß . In effect, we solve $(ßI - B)v = e$ for v with some tiny uncontrolled e comparable to roundoff. The foregoing analysis tells us that the computed solution v is likely to point in the direction of the desired eigenvector unless B nearly has a non-simple eigenvalue near ß .

"Likely", "usually", "most", "almost all", "nearly" … are weasel words that offend many a pure mathematician who prefers "always", "every", "all" … and takes "exactly" for granted. Applied mathematicians encounter situations often where the weasel words are the best that can be expected. In fact, the weasel words above are provably the best that can be expected in their contexts, but that is a story for another day.