Determinants, Matrix Norms, Inverse Mapping Theorem

G. B. Folland

The purpose of this notes is to present some useful facts about matrices and determinants and a proof of the inverse mapping theorem that is rather different from the one in Apostol.

Notation: $M_n(\mathbb{R})$ denotes the set of all $n \times n$ real matrices.

Determinants: If $A \in M_n(\mathbb{R})$, we can consider the rows of A: $\mathbf{r}_1, \ldots, \mathbf{r}_n$. These are elements of $\mathbb{R}^n$, considered as row vectors. Conversely, given n row vectors $\mathbf{r}_1, \ldots, \mathbf{r}_n$, we can stack them up into an $n \times n$ matrix A. Thus we can think of a function f on matrix space $M_n(\mathbb{R})$ as a function of n $\mathbb{R}^n$-valued variables or vice versa:

$$f(A) \longleftrightarrow f(\mathbf{r}_1, \ldots, \mathbf{r}_n).$$

Basic Fact: *There is a unique function* $\det : M_n(\mathbb{R}) \to \mathbb{R}$ *(the* "determinant"*) with the following three properties:*

i. det *is a linear function of each row when the other rows are held fixed: that is,*

$$\det(\alpha\mathbf{a} + \beta\mathbf{b}, \mathbf{r}_2, \ldots, \mathbf{r}_n) = \alpha \det(\mathbf{a}, \mathbf{r}_2, \ldots, \mathbf{r}_n) + \beta \det(\mathbf{b}, \mathbf{r}_2, \ldots, \mathbf{r}_n),$$

and likewise for the other rows.

ii. *If two rows of A are interchanged,* $\det A$ *is multiplied by* -1*:*

$$\det(\ldots, \mathbf{r}_i, \ldots, \mathbf{r}_j, \ldots) = -\det(\ldots, \mathbf{r}_j, \ldots, \mathbf{r}_i, \ldots).$$

iii. $\det(I) = 1$, *where I denotes the $n \times n$ identity matrix.*

The uniqueness of the determinant follows from the discussion below; existence takes more work to establish. We shall not present the proof here but give the formulas. For $n = 2$ and $n = 3$ we have

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc, \qquad \det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei - afh + bfg - bdi + cdh - ceg.$$

For general n, $\det A = \sum_\sigma (\operatorname{sgn}\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}$, where the sum is over all permutations σ of $\{1, \ldots, n\}$, and $\operatorname{sgn}\sigma$ is 1 or -1 depending on whether σ is obtained by an even or odd number of interchanges of two numbers. (This formula is a computational nightmare for large n, being a sum of $n!$ terms, so it is of little use in practice. There are better ways to compute determinants, as we shall see shortly.)

An important consequence of properties (i) and (ii) is

iv. *If one row of A is the zero vector, or if two rows of A are equal, then* $\det A = 0$.

Properties (i), (ii), and (iv) tell how the determinant of a matrix behaves under the elementary row operations:

– Multiplying a row by a scalar multiplies the determinant by that scalar.
– Interchanging two rows multiplies the determinant by -1.
– Adding a multiple of one row to another row leaves the determinant unchanged, because

$$\det(\ldots,\mathbf{r}_i+\alpha\mathbf{r}_j,\ldots,\mathbf{r}_j,\ldots)=\det(\ldots,\mathbf{r}_i,\ldots,\mathbf{r}_j,\ldots)+\alpha\det(\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_j,\ldots),$$

by (i), and the last term is zero by (iv).

This gives a reasonably efficient way to compute determinants. To wit, any $n\times n$ matrix A can be row-reduced either to a matrix with an all-zero row (whose determinant is 0) or to the identity matrix (whose determinant is 1). Just keep track of what happens to the determinant as you perform these row operations, and you will have calculated $\det A$. (There are shortcuts for this procedure, but that's another story. We'll mostly be dealing with 2×2 and 3×3 matrices, for which one can just use the explicit formulas above.)

This observation is also the key to the main theoretical significance of the determinant. A matrix A is *invertible* — that is, the map $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^n$ defined by $\mathbf{f}(\mathbf{x})=A\mathbf{x}$ is invertible, where elements of $\mathbb{R}^n$ are considered as column vectors — if and only if A can be row-reduced to the identity. Indeed, in this case, if one starts with the identity matrix I and performs the same sequence of row operations on it that row-reduces A to I, the result is A^{-1}. It follows that *A is invertible if and only if $\det A\neq 0$.*

If A is invertible, its inverse can be calculated in terms of determinants. Indeed, suppose $A\in M_n(\mathbb{R})$, and let B_{ij} be the $(n-1)\times(n-1)$ matrix obtained by deleting the ith row and jth column of A. Then the ijth entry of A^{-1} is given by

$$(A^{-1})_{ij}=(-1)^{i+j}\frac{\det B_{ij}}{\det A}.$$

This formula is computationally less efficient than calculating A^{-1} by row reduction, at least when the size of A is large. However, it is theoretically important. In particular, it shows explicitly how the condition $\det A\neq 0$ comes into the picture, and since determinants are polynomial functions of the entries of a matrix, it shows that *the entries of A^{-1} are continuous functions of the entries of A, as long as $\det A\neq 0$.*

By the way, there is nothing special about the real number system here. Everything we have said works equally well if $\mathbb{R}$ is replaced by $\mathbb{C}$, or indeed by any field (except that the preceding remark about continuity only applies to fields where the notion of continuity makes sense).

Matrix Norms: It is often desirable to have a notion of the "size" of a matrix, like the norm or magnitude of a vector. One way to manufacture such a thing is simply to regard the n^2 entries of a matrix $A\in M_n(\mathbb{R})$ as the components of a vector in $\mathbb{R}^{n^2}$ and take its Euclidean norm. The resulting quantity is usually called the *Hilbert-Schmidt norm* of the matrix; it can be denoted by $\|A\|_{\text{HS}}$:

$$\|A\|_{\text{HS}}=\left[\sum_{i,j=1}^{n}|A_{ij}|^2\right]^{1/2}.$$

However, if we think of a matrix as determining a linear transformation of $\mathbb{R}^n$, namely $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, it is often better to use a different norm that is more closely related to the action of A on vectors. Namely, since the unit sphere $\{\mathbf{x} : \|\mathbf{x}\| = 1\}$ in $\mathbb{R}^n$ is compact and the function $\mathbf{x} \mapsto \|A\mathbf{x}\|$ from $\mathbb{R}^n$ to $\mathbb{R}$ is continuous on it, it has a maximum value, which we denote by $\|A\|_{\mathrm{op}}$:

$$\|A\|_{\mathrm{op}} = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|.$$

$\|A\|_{\mathrm{op}}$ is called the *operator norm* of A (the term "linear operator" being a common synonym for "linear transformation"). It is easy to check that both $\|A\|_{\mathrm{HS}}$ and $\|A\|_{\mathrm{op}}$ are indeed norms on the vector space $M_n(\mathbb{R})$ in the proper sense of the word; that is, they satisfy

$$\|cA\| = |c|\,\|A\| \quad (c \in \mathbb{R}); \qquad \|A + B\| \leq \|A\| + \|B\|, \qquad \|A\| = 0 \iff A = 0.$$

We observe that if $\mathbf{x}$ is any nonzero vector in $\mathbb{R}^n$ and $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$ is the corresponding unit vector, then

$$\|A\mathbf{x}\| = \big\|A(\|\mathbf{x}\|\mathbf{u})\big\| = \|\mathbf{x}\|\|A\mathbf{u}\| \leq \|\mathbf{x}\|\,\|A\|_{\mathrm{op}} = \|A\|_{\mathrm{op}}\|\mathbf{x}\|,$$

with equality if $\mathbf{x}$ is the unit vector that achieves the maximum in the definition of $\|A\|_{\mathrm{op}}$. In other words, $\|A\|_{\mathrm{op}}$ *is the smallest constant* C *such that* $\|A\mathbf{x}\| \leq C\|\mathbf{x}\|$ *for all* $\mathbf{x} \in \mathbb{R}^n$.

This property of the operator norm makes it very handy for calculations involving the magnitudes of vectors. On the other hand, it is often not easy to calculate $\|A\|_{\mathrm{op}}$ exactly in terms of the entries of A. The Hilbert-Schmidt norm is easier to compute, so it is good to know that the operator norm is no bigger than it:

$$\|A\|_{\mathrm{op}} \leq \|A\|_{\mathrm{HS}}.$$

(Equality is achieved for any matrix A with only one nonzero entry.) The reason is the following: Let $\mathbf{r}_1, \ldots, \mathbf{r}_n$ be the rows of A; then the jth component of $A\mathbf{x}$ is $\mathbf{r}_j \cdot \mathbf{x}$, so by the Cauchy-Schwarz inequality,

$$\|A\mathbf{x}\|^2 = \sum_1^n |(A\mathbf{x})_j|^2 = \sum_1^n |(\mathbf{r}_j \cdot \mathbf{x})|^2 \leq \sum_1^n \|\mathbf{r}_j\|^2\|\mathbf{x}\|^2 = \sum_{j=1}^n \sum_{i=1}^n |A_{ij}|^2\|\mathbf{x}\|^2 = \|A\|_{\mathrm{HS}}^2\|\mathbf{x}\|^2.$$

There's also an inequality going the other way. The jth column of A is $A\mathbf{e}_j$ where $\mathbf{e}_j$ is the jth unit coordinate vector, so

$$\|A\|_{\mathrm{HS}}^2 = \sum_{j=1}^n \|A\mathbf{e}_j\|^2 \leq \sum_{j=1}^n \|A\|_{\mathrm{op}}^2\|\mathbf{e}_j\|^2 = n\|A\|_{\mathrm{op}}^2.$$

That is, $\|A\|_{\mathrm{HS}} \leq \sqrt{n}\|A\|_{\mathrm{op}}$; equality is achieved when A is the identity matrix.

The operator and Hilbert-Schmidt norms both have the useful property that the norm of a product is at most the product of the norms:

$$\|AB\|_{\mathrm{op}} \leq \|A\|_{\mathrm{op}}\|B\|_{\mathrm{op}}, \qquad \|AB\|_{\mathrm{HS}} \leq \|A\|_{\mathrm{HS}}\|B\|_{\mathrm{HS}}.$$

For the operator norm this follows from the simple calculation

$$\|AB\mathbf{x}\| \le \|A\|_{\rm op}\|B\mathbf{x}\| \le \|A_{\rm op}\|B\|_{\rm op}\|\mathbf{x}\|,$$

and for the Hilbert-Schmidt norm it comes from the fact that the ijth entry of AB is the dot product of the ith row of A with the jth column of B together with the Cauchy-Schwarz inequality; details are left to the reader.

The Inverse Mapping Theorem (or Inverse Function Theorem): This is Theorem 13.6 in Apostol. Rather than going through the proof there, which depends on a series of preliminary results (Theorems 13.2–13.5), I'm going to present the proof in Rudin's *Principles of Mathematical Analysis*, which is perhaps more elegant and certainly shorter. It uses the following three lemmas. *In them, and throughout this proof, the norm of a matrix is always understood to be the operator norm.*

Lemma 1 (The Mean Value Inequality)**.** *Let U be an open convex set in $\mathbb{R}^n$, and let $\mathbf{f} : U \to \mathbb{R}^m$ be differentiable everywhere on U. If $\|\mathbf{Df}(\mathbf{z})\| \le C$ for all $\mathbf{z} \in U$, then $\|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})\| \le C\|\mathbf{y} - \mathbf{x}\|$ for all $\mathbf{x}, \mathbf{y} \in U$.*

This follows from Theorem 12.9 of Apostol by taking $\mathbf{a}$ to be the unit vector in the direction of $\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})$:

$$\|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})\| = \mathbf{a} \cdot [\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})] = \mathbf{a} \cdot [(\mathbf{Df}(\mathbf{z}))(\mathbf{y} - \mathbf{x})] \le \|\mathbf{a}\|\,\|\mathbf{Df}(\mathbf{z})\|\,\|\mathbf{y} - \mathbf{x}\| \le C\|\mathbf{y} - \mathbf{x}\|.$$

Lemma 2 (The Fixed Point Theorem for Contractions)**.** *Let (M, d) be a complete metric space, and let $\phi : M \to M$ be a map such that for some constant $c < 1$ we have $d(\phi(x), \phi(y)) \le cd(x, y)$ for all $x, y \in M$. Then there is a unique $x \in M$ such that $\phi(x) = x$.*

This is Theorem 4.48 in Apostol.

Lemma 3. *Suppose $A \in M_n(\mathbb{R})$ is invertible. If $\|B - A\| < 1/\|A^{-1}\|$, then B is invertible.*

Proof. Observe that $B = A + (B - A) = A[I + A^{-1}(B - A)]$. If $\mathbf{x} \ne \mathbf{0}$, we have

$$\|A^{-1}(B - A)\mathbf{x}\| \le \|A^{-1}\|\,\|B - A\|\,\|\mathbf{x}\| < \|\mathbf{x}\|;$$

hence $A^{-1}(B-A)\mathbf{x} \ne -\mathbf{x}$; hence $[I + A^{-1}(B-A)]\mathbf{x} \ne \mathbf{0}$; hence $B\mathbf{x} \ne \mathbf{0}$ since A is invertible. But for square matrices, invertibility is equivalent to the condition that $\{\mathbf{x} : B\mathbf{x} = \mathbf{0}\} = \{\mathbf{0}\}$, so B is invertible. ∎

Theorem 1 (Inverse Mapping Theorem)**.** *Let $\mathbf{f}$ be a mapping of class C^1 from an open set $S \subset \mathbb{R}^n$ into $\mathbb{R}^n$. Suppose that $\mathbf{a}$ is a point in S such that $\mathbf{Df}(\mathbf{a})$ is invertible, and let $\mathbf{b} = \mathbf{f}(\mathbf{a})$. Then there are open sets U and V in $\mathbb{R}^n$ with $\mathbf{a} \in U$ and $\mathbf{b} \in V$ such that $\mathbf{f}$ maps U one-to-one onto V. Moreover, the inverse mapping $\mathbf{g} : V \to U$, defined by the condition $\mathbf{f}(\mathbf{g}(\mathbf{y})) = \mathbf{y}$ for $\mathbf{y} \in V$, is of class C^1 on V, and $\mathbf{Dg}(\mathbf{y}) = [\mathbf{Df}(\mathbf{g}(\mathbf{y}))]^{-1}$.*

Proof. To simplify the notation, set $A = \mathbf{Df}(\mathbf{a})$; thus A is invertible. Given $\mathbf{y} \in \mathbb{R}^n$, define a function $\boldsymbol{\phi} : S \to \mathbb{R}^n$ by

$$\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x} + A^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x})). \tag{1}$$

(We really should write $\boldsymbol{\phi}_{\mathbf{y}}$ instead of $\boldsymbol{\phi}$ to indicate the dependence on $\mathbf{y}$, but we won't.) The point of $\boldsymbol{\phi}$ is that $\mathbf{y} = \mathbf{f}(\mathbf{x})$ if and only if $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x}$, so solving $\mathbf{f}(\mathbf{x}) = \mathbf{y}$ amounts to finding a fixed point of $\boldsymbol{\phi}$.

Since $\mathbf{Df}$ is continuous, there is an open ball $U \subset S$ centered at $\mathbf{a}$ such that

$$\|\mathbf{Df}(\mathbf{x}) - A\| \leq \frac{1}{2\|A^{-1}\|} \quad \text{for} \quad \mathbf{x} \in U. \tag{2}$$

Let $V = \mathbf{f}(U)$. To prove the first assertion of the theorem, we need to show that $\mathbf{f}$ is one-to-one on U and that V is open. Since

$$\mathbf{D}\boldsymbol{\phi}(\mathbf{x}) = I - A^{-1}\mathbf{Df}(\mathbf{x}) = A^{-1}(A - \mathbf{Df}(\mathbf{x})),$$

we have $\|\mathbf{D}\boldsymbol{\phi}(\mathbf{x})\| \leq \|A^{-1}\|\,\|A - \mathbf{Df}(\mathbf{x}))\| < \frac{1}{2}$, and hence, by Lemma 1,

$$\|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\| \leq \tfrac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\| \quad \text{for} \quad \mathbf{x}_1, \mathbf{x}_2 \in U. \tag{3}$$

It follows that $\mathbf{f}$ is one-to-one on U, for if $\mathbf{f}(\mathbf{x}_1) = \mathbf{y} = \mathbf{f}(\mathbf{x}_2)$ then $\mathbf{x}_1$ and $\mathbf{x}_2$ are both fixed points of $\boldsymbol{\phi}$, so $\|\mathbf{x}_1 - \mathbf{x}_2\| \leq \frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|$, which is possible only if $\mathbf{x}_1 = \mathbf{x}_2$.

Now, to show that V is open, suppose $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0) \in V$, and let B be an open ball of radius r centered at $\mathbf{x}_0$, where r is small enough so that $\overline{B} \subset U$. We will show that $\mathbf{y} \in V$ provided that $\|\mathbf{y} - \mathbf{y}_0\| < r/2\|A^{-1}\|$. Taking such a $\mathbf{y}$ as the $\mathbf{y}$ in (1), we have

$$\|\boldsymbol{\phi}(\mathbf{x}_0) - \mathbf{x}_0\| = \|A^{-1}(\mathbf{y} - \mathbf{y}_0)\| \leq \|A^{-1}\|\,\|\mathbf{y} - \mathbf{y}_0\| < \tfrac{1}{2}r.$$

Hence, by (3), if $\mathbf{x} \in \overline{B}$,

$$\|\boldsymbol{\phi}(\mathbf{x}) - \mathbf{x}_0\| \leq \|\boldsymbol{\phi}(\mathbf{x}) - \boldsymbol{\phi}(\mathbf{x}_0)\| + \|\boldsymbol{\phi}(\mathbf{x}_0) - \mathbf{x}_0\| < \tfrac{1}{2}\|\mathbf{x} - \mathbf{x}_0\| + \tfrac{1}{2}r \leq r,$$

that is, $\boldsymbol{\phi}(\mathbf{x}) \in \overline{B}$. Thus $\boldsymbol{\phi}$ maps $\overline{B}$ into itself, and by (3) it is a contraction on $\overline{B}$ since $\overline{B} \subset U$. Moreover, $\overline{B}$ is compact and hence complete. Hence, by Lemma 2, $\boldsymbol{\phi}$ has a fixed point $\mathbf{x} \in \overline{B}$. That is, there is an $\mathbf{x} \in \overline{B} \subset U$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$, and hence $\mathbf{y} \in \mathbf{f}(U) = V$, as claimed.

We have shown that $\mathbf{f} : U \to V$ is invertible; we denote its inverse by $\mathbf{g} : V \to U$.

To prove the last assertion of the theorem, suppose $\mathbf{y} \in V$ and $\mathbf{y} + \mathbf{k} \in V$, and set $\mathbf{x} = \mathbf{g}(\mathbf{y})$ and $\mathbf{h} = \mathbf{g}(\mathbf{y} + \mathbf{k}) - \mathbf{g}(\mathbf{y})$; thus $\mathbf{y} = \mathbf{f}(\mathbf{x})$ and $\mathbf{y} + \mathbf{k} = \mathbf{f}(\mathbf{x} + \mathbf{h})$. Taking this $\mathbf{y}$ as the $\mathbf{y}$ in (1), we have

$$\boldsymbol{\phi}(\mathbf{x} + \mathbf{h}) - \boldsymbol{\phi}(\mathbf{x}) = \mathbf{h} + A^{-1}[\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x} + \mathbf{h})] = \mathbf{h} - A^{-1}\mathbf{k}.$$

It follows from (3) that $\|\mathbf{h} - A^{-1}\mathbf{k}\| \leq \frac{1}{2}\|\mathbf{h}\|$, so $\|A^{-1}\mathbf{k}\| \geq \frac{1}{2}\|\mathbf{h}\|$ and hence

$$\|\mathbf{h}\| \leq 2\|A^{-1}\|\,\|\mathbf{k}\|. \tag{4}$$

By (2) and Lemma 3, $\mathbf{Df}(\mathbf{x})$ is invertible; let us denote its inverse by B. Since

$$\mathbf{g}(\mathbf{y}+\mathbf{k}) - \mathbf{g}(\mathbf{y}) - B\mathbf{k} = (\mathbf{x}+\mathbf{h}) - \mathbf{x} - B\mathbf{k} = -B\mathbf{k} + \mathbf{h} = -B[\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{Df}(\mathbf{x})\mathbf{h}],$$

(4) shows that

$$\frac{\|\mathbf{g}(\mathbf{y}+\mathbf{k}) - \mathbf{g}(\mathbf{y}) - B\mathbf{k}\|}{\|\mathbf{k}\|} \le 2\|B\|\,\|A^{-1}\|\frac{\|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{Df}(\mathbf{x})\mathbf{h}\|}{\|\mathbf{h}\|}.$$

Now let $\mathbf{k} \to \mathbf{0}$; then by (4), $\mathbf{h} \to \mathbf{0}$ too. The numerator on the right is the error term in the definition of differentiability of $\mathbf{f}$ at $\mathbf{x}$, so the quantity on the right tends to 0. Hence the quantity on the left does so too, and by definition of differentiability again, this means that $\mathbf{g}$ is differentiable at $\mathbf{y}$ and $\mathbf{Dg}(\mathbf{y}) = B = [\mathbf{Df}(\mathbf{x})]^{-1} = [\mathbf{Df}(\mathbf{g}(\mathbf{y}))]^{-1}$ as claimed. Finally, $\mathbf{g}$ is continuous (since $\mathbf{g}$ is differentiable), $\mathbf{Df}$ is continuous (since $\mathbf{f}$ is of class C^1), and the inversion map is continuous on matrix space. Therefore $\mathbf{Dg}$ is continuous and hence $\mathbf{g}$ is of class C^1. ∎

Note that the formula for $\mathbf{Dg}(\mathbf{y})$ is in accordance with the chain rule: since $\mathbf{f} \circ \mathbf{g}$ is the identity mapping, the chain rule says that

$$I = \mathbf{D}(\mathbf{f} \circ \mathbf{g})(\mathbf{y}) = \mathbf{Df}(\mathbf{g}(\mathbf{y})) \cdot \mathbf{Dg}(\mathbf{y}) \quad \text{(matrix multiplication)},$$

i.e., $\mathbf{Dg}(\mathbf{y}) = [\mathbf{Df}(\mathbf{g}(\mathbf{y}))]^{-1}$.