W. Arveson

I will elaborate on some of these arguments in detail, since they illustrate how one handles such issues in a coordinate-free way. Notice that almost everything to follow reduces to the problem of making appropriate estimates. There are no rules for making good estimates; it is something you learn only by doing it enough times yourself, in your own way.

**Exercise 2.** The proofs that $\|AB\| \leq \|A\| \cdot \|B\|$ and $\|1\| = 1$ are straightforward. For matrices $A, A_0, B, B_0 \in M_n(\mathbb{R})$ we have $AB - A_0B_0 = (A - A_0)B + A_0(B - B_0)$ and therefore

$$\|AB - A_0B_0\| \leq \|A - A_0\| \cdot \|B\| + \|A_0\| \cdot \|B - B_0\|.$$ 

We have to show that if $A_1, A_2, \ldots$ and $B_1, B_2, \ldots$ are sequences that converge, respectively, to $A_0$ and $B_0$, then $A_nB_n$ converges to $A_0B_0$. For that we estimate as follows:

$$\|A_nB_n - A_0B_0\| \leq \|B_n\| \cdot \|A_n - A_0\| + \|A_0\| \cdot \|B_n - B_0\|.$$ 

From the triangle inequality we know that $\|B_n\| - \|B_0\| \leq \|B_n - B_0\| \to 0$ as $n \to \infty$, and therefore the sequence of norms $\|B_n\|$ is bounded. Choosing $M$ large enough that $\|B_n\| \leq M$ for every $n = 1, 2, \ldots$ we conclude from the previous inequality that

$$\|A_nB_n - A_0B_0\| \leq M \cdot \|A_n - A_0\| + \|A_0\| \cdot \|B_n - B_0\| \to 0,$$

as $n \to \infty$. [Note: of course there is an “$\epsilon$-$\delta$” proof of joint continuity of the function $f(A, B) = AB$ that is based on the estimates above. It might be useful for you to reformulate the above argument in those terms]

**Exercise 3.** It follows from the first inequality of Exercise 2 and an obvious induction that $\|A^p\| \leq \|A\|^p$ for every $p = 1, 2, \ldots$. Fix a matrix $A \in M_n(\mathbb{R})$ satisfying $\|A\| < 1$, and consider the partial sums of the “geometric series”

$$S_n = 1 + A + A^2 + \cdots + A^n, \quad n = 0, 1, 2, \ldots.$$ 

For every $n, k = 1, 2, \ldots$ we can estimate the norm of $S_{n+k} - S_n$ as follows

$$\|S_{n+k} - S_n\| \leq \sum_{r=n+1}^{n+k} \|A^r\| \leq \sum_{r=n+1}^{n+k} \|A\|^r \leq \|A\|^n \sum_{r=0}^{\infty} \|A\|^r = \frac{\|A\|^{n+1}}{1 - \|A\|}.$$ 

Thus $\{S_n\}$ is a Cauchy sequence. Since $M_n(\mathbb{R})$ is a complete metric space, $S_n$ must converge to a matrix $B$ as $n \to \infty$. As in the discussion of the geometric series in freshman calculus, we cancel terms to find that

$$S_n(1 - A) = (1 - A)S_n = 1 - A^{n+1},$$
and since \( \|A^n\| \leq \|A\|^{n+1} \to 0 \) as \( n \to \infty \), it follows [from continuity of the maps \( X \mapsto X(1 - A) \) and \( X \mapsto (1 - A)X \)] that
\[
B(1 - A) = \lim_{n \to \infty} S_n(1 - A) = \lim_{n \to \infty} (1 - A)S_n = (1 - A)B = 1
\]

Notice that we also have the following estimate, telling us how close \((1 - A)^{-1}\)

is to \(1\) in terms of \(\|A\|\) when \(\|A\| < 1\):

\[
(A) \quad \| (1 - A)^{-1} - 1 \| = \| B - 1 \| = \| \sum_{r=1}^{\infty} A^r \| \leq \sum_{r=1}^{\infty} \| A \|^r = \| A \| \frac{1}{1 - \| A \|}.
\]

**Exercise 4.** Note that Exercise 3 implies that every \( C \in M_n(\mathbb{R}) \) with \(\| 1 - C \| < 1\) is invertible.

Let \( A \) be an invertible \( n \times n \) matrix. We have to exhibit a positive number \( \epsilon \) with the property that every matrix \( B \) satisfying \( \| A - B \| \leq \epsilon \) is invertible. The Hint shows that for every \( B \in M_n(\mathbb{R}) \) we have
\[
\| 1 - A^{-1}B \| = \| A^{-1}(A - B) \| \leq \| A^{-1} \| \cdot \| A - B \|.
\]
So given any \( B \) satisfying \( \| A - B \| \leq \frac{1}{2\| A^{-1} \|} \) we will have \( \| 1 - A^{-1}B \| \leq 1/2 < 1 \). Exercise 3 implies that \( A^{-1}B \) must be invertible, hence \( B = A(A^{-1}B) \) is invertible. Thus we can take \( \epsilon = \frac{1}{2\| A^{-1} \|} \).

**Exercise 5.** Consider the function \( f(A) = A^{-1} \) defined on \( GL(n) \). For fixed \( A \in GL(n) \), the result of Exercise 4 implies that \( A + X \) will be invertible whenever \( \|X\| \) is sufficiently small. We now show that
\[
\lim_{\|X\| \to 0} (A + X)^{-1} = A^{-1},
\]
by estimating the norm of \((A + X)^{-1} - A^{-1}\) as follows. Assuming that \(\|X\|\) is small enough that \( A + X \) is invertible, we have
\[
(A + X)^{-1} - A^{-1} = (A(1 + A^{-1}X))^{-1} - A^{-1} = (1 + A^{-1}X)^{-1}A^{-1} - A^{-1}
\]
\[
= ((1 + A^{-1}X)^{-1} - 1)A^{-1}.
\]
Notice that this formula, together with Exercise 3, implies that \( A + X \) is invertible whenever \( \|X\| \) is small enough that \(\|A^{-1}X\| < 1\); for example, \(\|X\| < 1/\|A^{-1}\|\) is small enough. For such \( X \) we can use the estimate \((A)\) above as follows:
\[
\| (1 + A^{-1}X)^{-1} - 1 \| = \| (1 - (A^{-1}X))^{-1} - 1 \| \leq \| A^{-1}X \| \frac{1}{1 - \| A^{-1}X \|}.
\]
Since \(\|A^{-1}X\| \leq \|A^{-1}\| \cdot \|X\| \to 0 \) as \( \|X\| \to 0 \), the above inequalities imply that \(\|1 - A^{-1}X\| \to 0 \) as \( \|X\| \to 0 \), and consequently \(\|(A + X)^{-1} - A^{-1}\| \to 0 \) as \(\|X\| \to 0\).
Exercise 6. Let \( f(A) = A^{-1} \) be the inversion function defined on \( GL(n) \). We have seen in Exercise 4 that the domain of \( f \) is an open set in \( M_n(\mathbb{R}) \). Now we have to show that \( D_Af \) exists for every fixed \( A \in GL(n) \), and that \( A \mapsto D_Af \) is a continuous function from \( GL(n) \) to linear operators on \( M_n(\mathbb{R}) \). I’ll first show that all directional derivatives exist and will compute an explicit formula for the directional derivatives of \( f \) at a point \( A \in GL(n) \), defined by

\[
D_Af(X) = \frac{d}{dt} f(A + tX)|_{t=0} = \lim_{t \to 0} \frac{1}{t} (f(A + tX) - f(A)),
\]

for an arbitrary \( X \in M_n(\mathbb{R}) \). Once that has been accomplished we will have the formula we need (it will turn out to be \( D_Af(X) = -A^{-1}X A^{-1} \)), and then we will show that this linear operator \( X \mapsto D_Af(X) \) does indeed satisfy the criterion for being the derivative of \( f \) at \( A \in GL(n) \). At that point it will be easy to check that the map \( A \mapsto D_Af \) is continuous.

Fix \( A \in GL(n) \) and \( X \in M_n(\mathbb{R}) \). Since \( GL(n) \) is open, \( A + tX \) will be invertible provided that \( |t| \) is sufficiently small. For such small \( t \) we can use the formulas of Exercise 5 to write

\[
f(A + tX) - f(A) = (A + tX)^{-1} - A^{-1} = ((1 + tA^{-1}X)^{-1} - 1)A^{-1},
\]

and therefore

\[
\frac{1}{t} (f(A + tX) - f(A)) = \frac{1}{t} ((1 + tA^{-1}X)^{-1} - 1)A^{-1}.
\]

Note that \( \|tA^{-1}X\| \leq |t| \cdot \|A^{-1}\| \cdot \|X\| \) can be made as small as we like by choosing \( |t| \) small enough, and in particular for small enough \( |t| \) we will have \( \|tA^{-1}X\| < 1 \). For such \( t \) we can expand \((1 + tA^{-1}X)^{-1}\) into a convergent “geometric series” as in Exercise 3, and after subtracting \( 1 \) from that expression we obtain

\[
(1 + tA^{-1}X)^{-1} - 1 = (1 - (-tA^{-1}X))^{-1} - 1 = \sum_{r=1}^{\infty} (-t)^r(A^{-1}X)^r.
\]

Thus for all nonzero \( t \) we have

\[
\frac{1}{t} ((1 + tA^{-1}X)^{-1} - 1) = \sum_{r=1}^{\infty} (-1)^r t^{r-1} (A^{-1}X)^r
\]

\[
= -A^{-1}X + \sum_{r=2}^{\infty} (-1)^r t^{r-1} (A^{-1}X)^r = -A^{-1}X + R_t,
\]

where \( R_t = \sum_{r=2}^{\infty} (-1)^r t^{r-1} (A^{-1}X)^r \). Notice that \( \|R_t\| \to 0 \) as \( |t| \to 0 \). That is because of the estimate

\[
\|R_t\| = \| \sum_{r=2}^{\infty} (-1)^r t^{r-1} (A^{-1}X)^r \| = \sum_{r=2}^{\infty} |t|^{r-1} \| (A^{-1}X)^r \| \leq \sum_{r=2}^{\infty} |t|^{r-1} \| A^{-1}X \|^r = \frac{|t| \cdot \|A^{-1}X\|^2}{1 - |t| \cdot \|A^{-1}X\|}.
\]
since the last term of the preceding string of inequalities obviously tends to zero as \(|t| \to 0\). Thus we have proved that
\[
\lim_{t \to 0} \frac{1}{t} \left( (1 + tA^{-1}X)^{-1} - 1 \right) = -A^{-1}X.
\]
In view of formula (B) above, this argument shows that all directional derivatives of the function \(f\) exist at every point of the domain of \(f\), and that that they are given by the explicit formula
\[
(C) \quad \lim_{t \to 0} \frac{1}{t} (f(A + tX) - f(A)) = -A^{-1}XA^{-1}.
\]
This tells us what \(D_A f\) should be, namely \(D_A f(X) = -A^{-1}XA^{-1}\), for \(A \in GL(n)\) and arbitrary \(X \in M_n(\mathbb{R})\). Formula (C) is a "noncommutative" counterpart of the formula from freshman calculus which asserts that the differential of the function \(f(x) = x^{-1}\) at \(x = a\) \((a \neq 0)\) is given by \(df_a(h) = -a^{-2}h\).

Now that we know what \(D_A f\) must be, it is relatively easy to finish the proof. First, we must show that for fixed \(A \in GL(n)\), the linear operator \(X \mapsto -A^{-1}XA^{-1}\) satisfies the definition of \(D_A f\), namely that
\[
f(A + X) = f(A) - A^{-1}XA^{-1} + o(\|X\|), \quad X \in M_n(\mathbb{R}).
\]
In completely explicit terms, the assertion of the preceding line is that
\[
(D) \quad \lim_{\|X\| \to 0} \frac{\|f(A + X) - f(A) + A^{-1}XA^{-1}\|}{\|X\|} = 0,
\]
and (D) is what must be proved.

At this point, if you look back carefully through the estimates we have done above to prove (C) (where \(tX\) was used instead of \(X\)), you will see that the same arguments can be used to prove the somewhat more general statement (D). Thus, the estimates required for proving (D) have already been developed in proving the existence of directional derivatives (C) and calculating their value. It is instructive to actually carry out the estimates required to prove (D) as variations of the estimates we have made above; I will leave that for you to enjoy on your own time.

Finally, the continuity of \(D_A f\) in \(A\) amounts to showing that the function \(Df\) that takes \(A \in GL(n)\) to the linear operator \(X \mapsto D_A f(X) = -A^{-1}XA^{-1}\) is continuous. The space \(\mathcal{L}(M_n(\mathbb{R}))\) of all linear operators on \(M_n(\mathbb{R})\) is just another vector space of finite dimension \([its\ dimension\ is\ n^4]\), and a convenient norm on \(\mathcal{L}(M_n(\mathbb{R}))\) is the operator norm associated with the norm we have been using on \(M_n(\mathbb{R})\), namely
\[
\|L\| = \sup_{\|X\| \leq 1} \|L(X)\|.
\]
If \(A_k\) is a sequence in \(GL(n)\) that converges to \(A \in GL(n)\), then the operator norms of the differences \(D_{A_k} f - D_A f\) are given by
\[
\|D_{A_k} f - D_A f\| = \sup_{\|X\| \leq 1} \| - A_k^{-1}XA_k^{-1} + A^{-1}XA^{-1}\|.
\]
In fact, a straightforward application of the result of Exercise 5 shows that the right side of the preceding expression must tend to zero when \(A_k \to A\). Conclusion: The function \(f(A) = A^{-1}\) is of class \(C^1\).