Proof of the infinite differentiability of e^X when X is a square matrix

The idea behind the proof is to use the theorem that says that if a series converges absolutely then it is the derivative of the series found by integrating term-by-term (Corollary 5.3.2 in Marsden and Hoffman). This means that if we can prove the uniform convergence of the series obtained by a pointwise partial differentiation of some order, then (holding all but the last of the variables of differentiation constant), we see that that series is the derivative of the series obtained by omitting that very last differentiation. Working backwards, through the successive differentiations then allows us to justify the term-by-term differentiation, provided we can prove uniform convergence for each of the intervening series.

To prove this uniform convergence we will use the Weierstrass M-test. To do this we need to examine the general term in a general term-by-term partial derivative of the series for e^X. So suppose $X \in \mathcal{M}(n, \mathbb{R})$; That is, suppose X is an $n \times n$ matrix with real coefficients. Notice that X can be thought of as a vector with n^2 entries $x_{i,j}$ so if we are going to talk about a partial derivative of X or of X^m then these partial derivatives will look like the following types of expressions:

$$\frac{\partial}{\partial x_{i,j}} X^3 \,,$$

or

$$\frac{\partial^5}{\partial x_{i_1,j_1}^2 \partial x_{i_2,j_2} \partial x_{i_3,j_3}^2} X^m \,.$$

When the order of the derivative is high this will become prohibitively cumbersome to write, so we will use the following more compact notation: We let $\mathbf{k}$ be an $n \times n$ matrix of non-negative integers (the orders of the partial derivatives with respect to the various entries in X), and we let $|\mathbf{k}|$ denote the sum of the entries in $\mathbf{k}$. Thus the second of the two derivatives listed above would be written as

$$\frac{\partial^{|\mathbf{k}|}}{\prod_{i,j} \partial x_{i,j}^{\mathbf{k}_{i,j}}} X^m ,$$

where $\mathbf{k}$ is the matrix all of whose entries are zeros, except for a 1 in the (i_2, j_2) position and a 2 in the (i_1, j_1) and (i_3, j_3) positions. To simplify the notation even further, we will write this, in turn, as

$$\frac{\partial^{|\mathbf{k}|}}{\partial x^{\mathbf{k}}} X^m .$$

We will now try to estimate the size of this derivative in terms of the norm $\|X\|$. To calculate the partial derivatives by the method discussed in class, we need a unit "vector" in the direction of each of the "axes" in the vector space $\mathcal{M}(n, \mathbb{R})$. These are the matrices $E_{i,j}$ defined as follows: $E_{i,j}$ is the $n \times n$ matrix with 1 in the (i, j)-position and 0 everywhere else. Then

$$\frac{\partial^{|\mathbf{k}|}}{\partial x^{\mathbf{k}}} X^m = \frac{\partial^{|\mathbf{k}|}}{\prod_{i,j} \partial x_{i,j}^{\mathbf{k}_{i,j}}} X^m$$

$$= \frac{\partial^{|\mathbf{k}|}}{\prod_{i,j} \partial t_{i,j}^{\mathbf{k}_{i,j}}} \left(X + \sum_{i,j} t_{i,j} E_{i,j} \right)^m_{|t_{1,1}=t_{1,2}=\cdots=t_{n,n}=0} .$$

Now consider all the terms we get when we expand

$$\left(X + \sum_{i,j} t_{i,j} E_{i,j} \right)^m$$

and organize them into a polynomial in the variables $t_{i,j}$ with matrix coefficients. When we do the differentiation and put all the $t_{i,j}$'s to zero, the only term that does not disappear is the term in $\prod_{i,j} t_{i,j}^{\mathbf{k}_{i,j}}$. The coefficient $C_{\mathbf{k}}$ of this term is found by observing that it is a matter of picking $t_{1,1}E_{1,1}$ in $\mathbf{k}_{1,1}$ of the factors, and $t_{1,2}E_{1,2}$ in $\mathbf{k}_{1,2}$ of the factors, and so on. The number of different ways this set of choices can be made is

$$\frac{m!}{\prod_{i,j}(\mathbf{k}_{i,j}!)(m - |\mathbf{k}|)!} .$$

Each of the resulting choices corresponds to a product of the matrices

$$X^{m-|\mathbf{k}|} \prod_{i,j} E_{i,j}^{\mathbf{k}_{i,j}} ,$$

though not in general in that order. In other words, $C_{\mathbf{k}}$ is the sum of

$$\frac{m!}{\prod_{i,j}(\mathbf{k}_{i,j}!)(m-|\mathbf{k}|)!} .$$

matrix products of this form. When we take the norm of a product of matrices the order is unimportant, so we conclude for each term in the coefficient $C_{\mathbf{k}}$ that its norm is no more than

$$\|X\|^{m-|\mathbf{k}|} \prod_{i,j} \|E_{i,j}\|^{\mathbf{k}_{i,j}} = \|X\|^{m-|\mathbf{k}|} .$$

Thus the norm of $C_{\mathbf{k}}$ is bounded above by

$$\frac{m!}{\prod_{i,j}(\mathbf{k}_{i,j}!)(m-|\mathbf{k}|)!} \|X\|^{m-|\mathbf{k}|} .$$

When we perform the differentiation

$$\frac{\partial^{|\mathbf{k}|}}{\partial t^{\mathbf{k}}} C_{\mathbf{k}} \prod_{i,j} t_{i,j}^{\mathbf{k}_{i,j}} ,$$

the differentiation brings out a factor $\prod_{i,j}(\mathbf{k}_{i,j}!)$. Thus

$$\begin{aligned} \left\| \frac{\partial^{|\mathbf{k}|}}{\partial t^{\mathbf{k}}} (X^m) \right\| &= \left\| \frac{\partial^{|\mathbf{k}|}}{\partial t^{\mathbf{k}}} C_{\mathbf{k}} \prod_{i,j} t_{i,j}^{\mathbf{k}_{i,j}} \right\| \\ \leq \prod_{i,j}(\mathbf{k}_{i,j}!) & \frac{m!}{\prod_{i,j}(\mathbf{k}_{i,j}!)(m-|\mathbf{k}|)!} \|X\|^{m-|\mathbf{k}|} . \\ &= \frac{m!}{(m-|\mathbf{k}|)!} \|X\|^{m-|\mathbf{k}|} . \end{aligned}$$

We are now ready to prove the main theorem of this discussion:

Theorem: *Suppose X is and $n \times n$ matrix with $\|X\| \leq \sigma < M$ for some positive real number M, and that $\sum_{m=0}^{\infty} |a_n| M^m$ converges. Then $\sum_{m=0}^{\infty} a_n X^m$ converges uniformly on $\|X\| \leq \sigma$ and the sum is infinitely differentiable on $\|X\| < \sigma$.*

Proof: For then, using the calculations on the preceding page,

$$\left\| \frac{\partial^{|\mathbf{k}|}}{\partial t^{\mathbf{k}}} \left(X^m\right) \right\| \leq \frac{m!\, \sigma^{m-|\mathbf{k}|}}{(m-|\mathbf{k}|)!}$$

$$\leq m^{|\mathbf{k}|} \left(\frac{\sigma}{M}\right)^{m-|\mathbf{k}|} M^{m-|\mathbf{k}|}.$$

But

$$\sum_{m=0}^{\infty} \left[\left(\frac{\sigma}{M}\right)^{m-|\mathbf{k}|} \frac{m^{|\mathbf{k}|}}{M^{|\mathbf{k}|}} \right] a_m\, M^m$$

converges, since

$$\left[\left(\frac{\sigma}{M}\right)^{m-|\mathbf{k}|} \frac{m^{|\mathbf{k}|}}{M^{|\mathbf{k}|}} \right] \to 0$$

as $m \to \infty$. The uniform convergence now follows by the Weierstrass M-test. This proves that each of the partial derivatives is continuous, and using the comments at the beginning of the discussion, it follows that on the region $\|X\| < \sigma$ the series sums to the appropriate partial derivative of the function $\sum_{m=0}^{\infty} a_n X^m$.

$\square$

Note that this theorem can be used to prove that e^X is infinitely differentiable, for on $\|X\| \leq \sigma$, $\sum_{m=0}^{\infty} 1/n!\, M^m$ converges. Since in this case this is true for any M, it follows that the infinite differentiability is valid for all X.

The theorem can also be used for all sorts of other functions defined by series in powers of a matrix. For example, we could define

$$\log(I+X) = X - \frac{1}{2!}X^2 + \frac{1}{3!}X^3 - \frac{1}{4!}X^4 + \cdots$$

and prove that this is infinitely differentiable on $\|X\| < 1$.