MAU 23205 Lecture 7

John Stalker

Trinity College Dublin

29 September 2021

Matrix valued sequences, series, functions (1/4)

Most of what you would expect to be true of matrix valued sequences, series and functions, based on the results for the real valued case, is true. In definitions, statements of theorems, and proofs you replace the absolute value sign, $| \quad |$ with the norm sign $\| \quad \|$. In the case of vectors in $\mathbf{R}^n$, either row or column vectors, the norm is the Euclidean one.

$$\|\mathbf{x}\| = \sqrt{\sum_{j=1}^{n} |x_j|^2}.$$

Any function q satisfying the following conditions could be used instead of $\| \quad \|$:

- $q(\mathbf{x}) \geq 0$ and $q(\mathbf{x}) > 0$ unless $\mathbf{x} = \mathbf{0}$.
- $q(\alpha \mathbf{x}) = |\alpha| q(\mathbf{x})$.
- $q(\mathbf{x} + \mathbf{y}) \leq q(\mathbf{x}) + q(\mathbf{y})$.

Matrix valued sequences, series, functions (2/4)

For $m \times n$ matrices it is more convenient to make the following odd looking choice:

$$\|A\| = \max_{\substack{\|\mathrm{x}\|\leq 1\\ \|\mathrm{y}\|\leq 1}} |\mathbf{x}^T A\mathbf{y}| = \max_{\substack{x_1^2+\cdots+x_m^2\leq 1\\ y_1^2+\cdots+y_n^2\leq 1}} \left|\sum_{i=1}^{m}\sum_{j=1}^{n} x_i a_{i,j} y_j\right|.$$

The maximum exists. In the case $m = 1$ or $n = 1$ it agrees with the Euclidean norm on vectors. It has all the properties mentioned earlier:

- $\|A\| \geq 0$ and $\|A\| > 0$ unless $A = O$.
- $\|\alpha A\| = |\alpha|\|A\|$.
- $\|A + B\| \leq \|A\| + \|B\|$.

It is the only choice satisfying those properties and

$$\|AB\| \leq \|A\|\|B\|, \qquad \|I\| = 1, \qquad \left\|A^T\right\| = \|A\|.$$

Matrix valued sequences, series, functions (3/4)

There are only two complications compared to the real case.

- You need to be slightly careful about not changing the order of multiplication. For example,

$$\lim_{n\to\infty} A_n B_n = \left(\lim_{n\to\infty} A_n\right)\left(\lim_{n\to\infty} B_n\right)$$

provided the limits on the right exist. That wouldn't be true if you reversed the order of multiplication on only one side.

- You need to distinguish between the concepts of non-zero and invertible. For example,

$$\lim_{n\to\infty} A_n = B \Leftrightarrow \lim_{n\to\infty} (A_n - B) = O$$

but the condition for

$$\lim_{n\to\infty} A_n B_n^{-1} = \left(\lim_{n\to\infty} A_n\right)\left(\lim_{n\to\infty} B_n\right)^{-1}$$

to hold is that both limits on the right exist and $\lim_{n\to\infty} B_n$ is invertible.

Matrix valued sequences, series, functions (4/4)

Sequences, series, functions, etc. converge in the sense described above if and only if all the components, a.k.a entries, of the vector or matrix converge. The definitions in terms of norms are more suited to proving theorems than checking examples. It's the theorems you generally use to establish convergence, not the definition or the componentwise criterion.

Derivatives (1/4)

The derivative of $f \colon \mathbf{R} \to \mathbf{R}$ at $\mathbf{x} \in \mathbf{R}$ is equal to $v \in \mathbf{R}$ if and only if the four equivalent conditions below are satisfied. If they are then we say f is differentiable at x and write $f'(x) = v$.

- $\lim_{y \to x} (y - x)^{-1}(f(y) - f(x)) = v$.
- $\lim_{y \to x} (f(y) - f(x))(y - x)^{-1} = v$.
- For all $\epsilon > 0$ there is a $\delta > 0$ such that if $0 < |x - y| < \delta$ then $|f(y) - f(x) - v(y - x)| < \epsilon |x - y|$.
- For all $\epsilon > 0$ there is a $\delta > 0$ such that if $0 < |x - y| < \delta$ then $|f(y) - f(x) - (y - x)v| < \epsilon |x - y|$.

The first and second conditions differ only in the order of multiplication. So do the third and fourth. We can also define the derivative if f is just defined near x, not everywhere. I will skip the details.

Derivatives (2/4)

Which of these make sense for a function from $m \times n$ matrices to $p \times q$ matrices? The first two conditions still make sense if $m = n = 1$, i.e. for matrix valued functions of a real variable. $(y - x)^{-1}(f(y) - f(x))$ and $(f(y) - f(x))(y - x)^{-1}$ are both defined in this context, for any p and q. The derivative v is then a $p \times q$ matrix. The first two conditions are either meaningless or mostly uninteresting except when $m = n = 1$.

The third condition makes sense, i.e. the matrix operations are well defined, if $n = q$ and v is a $p \times m$ matrix. It's mostly uninteresting except when $n = q = 1$, i.e. for functions from column vectors to column vectors.

The fourth condition makes sense if $m = p$ and v is an $n \times q$ matrix. It's mostly uninteresting except when $m = p = 1$, i.e. for functions from row vectors to row vectors.

Derivatives (3/4)

In the overlapping "interesting" cases, i.e. scalar or vector valued functions of a real variable, all the "interesting" definitions agree. All bets are off in other cases.
For matrix valued functions of a real variable all the expected rules for differentiation apply, provided the algebraic operations are defined, except the product rule must be written

$$(AB)'(x) = A'(x)B(x) + A(x)B'(x)$$

and the quotient rule in either of the two forms

$$\left(AB^{-1}\right)'(x) = A'(x)B(x)^{-1} - A(x)B(x)^{-1}B'(x)B(x)^{-1},$$
$$\left(A^{-1}B\right)'(x) = A(x)^{-1}B'(x) - A(x)^{-1}A'(x)A(x)^{-1}B(x).$$

Derivatives (4/4)

For functions from column vectors to column vectors the expected rules of differentiation apply, provided the algebraic operations are defined. The form of the chain rule is

$$(\mathbf{f} \circ \mathbf{g})'(\mathbf{x}) = (\mathbf{f}' \circ \mathbf{g})(\mathbf{x})\mathbf{g}'(\mathbf{x}).$$

For functions from row vectors to row vectors the expected rules of differentiation apply, provided the algebraic operations are defined, except the form of the chain rule is

$$(\mathbf{f} \circ \mathbf{g})'(\mathbf{x}) = \mathbf{g}'(\mathbf{x})(\mathbf{f}' \circ \mathbf{g})(\mathbf{x}).$$

Componentwise differentiation

A matrix valued function of a real variable is differentiable if and only if each component/entry is differentiable as a real valued function and the derivative can be computed componentwise, i.e. the i'th row, j'th column of the derivative is the derivative of the i'th row, j'th column.

If $\mathbf{f}$ is a function from $\mathbf{R}^n$ to $\mathbf{R}^m$ whose derivative at the point $\mathbf{x} \in \mathbf{R}^n$ is the $m \times n$ matrix A then the i'th row, j'th column of A is

$$a_{i,j} = \frac{\partial f_i}{\partial x_j}(\mathbf{x}).$$

The existence of these partial derivatives is necessary *but not sufficient* for the differentiability of $\mathbf{f}$ at $\mathbf{x}$. The existence of continuous partial derivatives near $\mathbf{x}$ is a necessary and sufficient condition for $\mathbf{f}$ to be continuously differentiable near $\mathbf{x}$.