Calculus on $\mathcal{B}(V)$

Differentiation:
$L(t)$ a 1-parameter family of operators in $\mathcal{B}(V), \; t \in (a, b)$.
Since $\mathcal{B}(V)$ is a metric space, continuity of $L(t)$ in $t$ is well defined.
Define differentiability as well:
$L(t)$ is differentiable at $t = t_0 \in (a, b)$ if

$$L'(t_0) = \lim_{t \to t_0} \frac{L(t) - L(t_0)}{t - t_0}$$

exists in the norm on $\mathcal{B}(V)$.
For example, it is easily checked that for $L \in \mathcal{B}(V)$, $e^{tL}$ is differentiable in $t$ for all $t \in \mathbb{R}$, and

$$\frac{d}{dt} e^{tL} = Le^{tL} = e^{tL}L .$$

Similarly consider families of operators in $\mathcal{B}(V)$ depending on several real or complex parameters.
A family $L(z)$ where

$$z = x + iy \in \Omega^{open} \subset \mathbb{C} \; (x, y \in \mathbb{R})$$

is said to be holomorphic in $\Omega$ if the partial derivatives

$$\frac{\partial}{\partial x} L(z), \quad \frac{\partial}{\partial y} L(z)$$

exist and are continuous in $\Omega$, and satisfy the Cauchy-Riemann eqn.

$$\left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) L(z) = 0 \quad \text{on} \; \Omega .$$

This is equivalent to the assumption that in a neighborhood of each point $z_0 \in \Omega$, $L(z)$ is given by the $\mathcal{B}(V)$-norm convergent power series

$$L(z) = \sum_{k=0}^{\infty} \frac{1}{k!} (z - z_0)^k \left( \frac{d}{dz} \right)^k L(z_0).$$
Calculus on $\mathcal{B}(V)$

*Integration:*  
One can also integrate families of operators.

If $L(t)$ depends continuously on $t \in [a, b]$, then it can be shown using the same estimates as for $\mathbb{F}$-valued functions (and the uniform continuity of $L(t)$ since $[a, b]$ is compact) that the Riemann sums

$$
\frac{b-a}{N} \sum_{k=0}^{n-1} L \left( a + \frac{k}{N} (b-a) \right)
$$

converge in $\mathcal{B}(V)$-norm (recall $V$ is a Banach space) as $n \to \infty$ to an operator in $\mathcal{B}(V)$, denoted

$$
\int_a^b L(t)dt .
$$

More general Riemann sums than just the left-hand “rectangular rule” with equally spaced points can be used.

Many results from standard calculus carry over, including

$$
\left\| \int_a^b L(t)dt \right\| \leq \int_a^b \| L(t) \| dt
$$

which follows from

$$
\left\| \frac{b-a}{N} \sum_{k=0}^{N-1} L \left( a + \frac{k}{N} (b-a) \right) \right\| \leq \frac{b-a}{N} \sum_{k=0}^{N-1} \left\| L \left( a + \frac{k}{N} (b-a) \right) \right\| .
$$

By parameterizing paths in $\mathbb{C}$, one can define line integrals of holomorphic families of operators. We will discuss such constructions further as we need them.
Operators in Finite Dimensions

*Transposes and Adjoints*

\[ A \in \mathbb{C}^{m \times n} \]

\[ A^T \in \mathbb{C}^{n \times m} \text{ the transpose of } A: (A_{ij} = (A^T)_{ji}) \]

\[ A^H = \bar{A}^T \text{ the conjugate-transpose (or Hermitian transpose) of } A \]

One often writes \( A^* = A^H \).

But there is a subtlety here that we must be careful about. The subtlety is related to the identification of a linear operator with its representation as a matrix. We take a moment to explain this subtlety.

Recall that an inner product on \( \mathbb{C}^n \) can be represented as a matrix multiply, e.g. for the Euclidean inner product

\[ \langle x, y \rangle = y^H x \]

For \( A \in \mathbb{C}^{n \times n} \),

\[ \langle Ax, y \rangle = \langle x, A^H y \rangle \]

since

\[ y^H Ax = (A^H y)^H x \]
Caution
The notation $A^*$ is used with two different meanings (particularly when $\mathbb{F} = \mathbb{C}$).

$$L \in \mathcal{B}(V,W) \iff L^* \in \mathcal{B}(W^*,V^*)$$

In finite dim., one can choose bases of $V$ and $W$ to encode the action of $L$ as left matrix multiplication on column vectors associated with the bases components (y=Tx).

Denote such a matrix as $T$.

Then the action of $L^*$ on $W^*$ can be encoded using the corresponding dual bases using the same matrix $T$. But now the action is represented as right multiplication by $T$ on row vectors of components in in the dual basis (a=bT).

On the other hand, in the presence of an inner product, the definition

$$\langle Lv, w \rangle = \langle v, L^* w \rangle$$

identifies $L^*$ with left-multiplication by the conjugate-transpose matrix.

These two definitions are related by the identification

$$V \cong V^*$$

induced by the inner product:

$$w \in V \iff \langle \cdot, w \rangle = w^* \in V^*.$$

But the conjugation in this identification gives rise to a different representation of $L^*$. The first is a natural representation obtained through the composition of function, and the second is through the inner product.
Norms on Matrices

Commonly used norms on $\mathbb{C}^{m \times n}$.

\[ \|A\|_1 = \sum_{i,j} |a_{ij}| \quad \text{(the } \ell^1\text{-norm on } A \text{ as if it were in } \mathbb{C}^{n^2}) \]
\[ \|A\|_\infty = \max_{i,j} |a_{ij}| \quad \text{(the } \ell^\infty\text{-norm on } A \text{ as if it were in } \mathbb{C}^{n^2}) \]
\[ \|A\|_2 = \left( \sum_{i,j} |a_{ij}|^2 \right)^{\frac{1}{2}} \quad \text{(the } \ell^2\text{-norm on } A \text{ as if it were in } \mathbb{C}^{n^2}) \]

$\|A\|_2$ is called the Hilbert-Schmidt norm of $A$, or the Frobenius norm of $A$, and is often denoted $\|A\|_F$.

It is also called the Euclidean norm on $\mathbb{C}^{m \times n}$. The associated inner product is

\[ \langle A, B \rangle = \text{tr } B^H A. \]

We also have $p$-norms for matrices: let $1 \leq p \leq \infty$,

\[ \|A\|_p = \max_{\|x\|_p = 1} \|Ax\|_p \quad \left( = \max_{\|x\|_p \leq 1} \|Ax\|_p = \max_{x \neq 0}(\|Ax\|_p/\|x\|_p) \right). \]

In particular,

\[ \|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}| \quad \text{(max column sum)} \]
\[ \|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}| \quad \text{(max row sum)} \]

$\|A\|_2$ is the spectral norm. We study it in detail later.
Consistent Matrix Norms

Definitions

- Let \( \mu : \mathbb{C}^{m \times n} \to \mathbb{R}, \nu : \mathbb{C}^{n \times k} \to \mathbb{R}, \rho : \mathbb{C}^{m \times k} \to \mathbb{R} \) be norms. We say that \( \mu, \nu, \rho \) are consistent if \( \forall A \in \mathbb{C}^{m \times n} \) and \( \forall B \in \mathbb{C}^{n \times k} \),
  \[ \rho(AB) \leq \mu(A)\nu(B) \]

- A norm on \( \mathbb{F}^{m \times n} \) is called consistent if it is consistent with itself, i.e., the definition above with \( m = n = k \) and \( \rho = \mu = \nu \). So by definition a norm on \( \mathbb{F}^{m \times n} \) is consistent iff it is submultiplicative.

- A collection \( \{\nu_{m,n} : m \geq 1, n \geq 1\} \), where \( \nu_{m,n} : \mathbb{F}^{m \times n} \to \mathbb{R} \) is a norm on \( \mathbb{F}^{m \times n} \), is called a family of matrix norms.

- A family \( \{\nu_{m,n} : m \geq 1, n \geq 1\} \) of matrix norms is called consistent if
  \[ (\forall m, n, k \geq 1)(\forall A \in \mathbb{F}^{m \times n})(\forall B \in \mathbb{F}^{n \times k}) \]
  \[ \nu_{m,k}(AB) \leq \nu_{m,n}(A)\nu_{n,k}(B). \]

Facts

Let \( \{\nu_{m,n}\} \) be a consistent family of matrix norms. Then

1. \( (\forall n \geq 1) \) \( \nu_{n,n} \) is submultiplicative.

2. \( (\forall m, n \geq 1) \) \( (\forall A \in \mathbb{F}^{m \times n}) \) \( \nu_{m,n}(A) \geq \mu_{m,n}(A) \), where \( \mu_{m,n} \) is the operator norm on \( \mathbb{F}^{m \times n} \) induced by \( \nu_{n,1} \) and \( \nu_{m,1} \).
Examples

(1) For $m > 1$, let $\nu_{m,1}$ be any norm on $\mathbb{F}^m$, and $\nu_{1,1}(x) = |x|$. For $m, n \geq 1$, let $\nu_{m,n}$ be the operator norm on $\mathbb{F}^{m \times n}$ induced by $\nu_{n,1}$ and $\nu_{m,1}$ Then $\{\nu_{m,n}\}$ is a consistent family of matrix norms.

(2) (maximum row sum norm) For $m, n \geq 1$ and $A \in \mathbb{F}^{m \times n}$, let

$$\nu_{m,n}(A) = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|.$$ 

Then $\nu_{n,1}$ is the $\ell^\infty$-norm on $\mathbb{F}^n$, and $\nu_{m,n}(A)$ is the operator norm induced by the $\ell^\infty$-norms on $\mathbb{F}^n$ and $\mathbb{F}^m$, which we denoted by $|||A|||_\infty$.

(3) (maximum column sum norm) For $m, n \geq 1$ and $A \in \mathbb{F}^{m \times n}$, let

$$\nu_{m,n}(A) = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}|.$$ 

Then $\nu_{n,1}$ is the $\ell^1$-norm on $\mathbb{F}^n$, and $\nu_{m,n}(\cdot)$ is the operator norm induced by the $\ell^1$-norms on $\mathbb{F}^n$ and $\mathbb{F}^m$, which we denoted by $|||A|||_1$.

(4) ($\ell^1$-norm on $\mathbb{F}^{m \times n}$ as if it were $\mathbb{F}^{mn}$) For $m, n \geq 1$ and $A \in \mathbb{F}^{m \times n}$, let

$$\nu_{m,n}(A) = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|.$$ 

Then $\{\nu_{m,n}\}$ is a consistent family of matrix norms. We denoted $\nu_{m,n}(A)$ by $||A||_1$. Also note that $||A||_1 \geq |||A|||_1$ agrees with Fact (2) above.
(5) ($\ell^2$-norm on $\mathbb{F}^{m \times n}$ as if it were $\mathbb{F}^{mn}$, the Hilbert-Schmidt norm)

For $m, n \geq 1$ and $A \in \mathbb{F}^{m \times n}$, let

$$\nu_{m,n}(A) = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}}.$$

Then $\nu_{n,1}$ is the $\ell^2$-norm on $\mathbb{F}^n$. If $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{n \times k}$, then by the Schwarz inequality,

$$(\nu_{m,k}(AB))^2 = \sum_{i=1}^{m} \sum_{j=1}^{k} \left| \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j} \right|^2 \leq \sum_{i=1}^{m} \sum_{j=1}^{k} \left( \sum_{\ell=1}^{n} |a_{i\ell}|^2 \right) \left( \sum_{r=1}^{n} |b_{rj}|^2 \right) = (\nu_{m,n}(A) \nu_{n,k}(B))^2,$$

so $\{\nu_{m,n}\}$ is a consistent family of matrix norms.

This is not a special case of example (1): for example, for $n > 1$, $\nu_{n,n}(I) = \sqrt{n}$ but the operator norm of $I$ is 1.

Denote $\nu_{m,n}(A)$ by $\|A\|_2$.

For $A \in \mathbb{F}^{m \times n}$ and $x \in \mathbb{F}^n$, we have

$$\|Ax\|_2 \leq \|A\|_2 \cdot \|x\|_2,$$

so for $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{n \times k}$,

$$\|AB\|_2 \leq \|A\|_2 \cdot \|B\|_2.$$

Fact (2) above gives the important inequality: for $A \in \mathbb{F}^{m \times n}$, $\|A\|_2 \leq \|A\|_2$. Thus the operator norm induced by the $\ell^2$-norms on $\mathbb{F}^m$ and $\mathbb{F}^n$ is dominated by the Frobenius norm.
**Condition Number and Error Sensitivity**

Assume \( A \in \mathbb{C}^{n \times n} \) is invertible.

How sensitive is the solution to \( Ax = b \) to perturbations in \( b \in \mathbb{C}^n \) and \( A \)?

\( \| \cdot \| \) a consistent matrix norm on \( \mathbb{C}^{n \times n} \).

\[ Ax = b \quad \text{and} \quad A\hat{x} = \hat{b}. \]

How far is \( \hat{x} \) from \( x \)?

\[ e = x - \hat{x} = \text{the error vector} \]
\[ \|e\| = \text{the error} \]

\[ r = b - \hat{b} = \text{the residual vector} \]
\[ \|r\| = \text{the residual} \]

\[ \|e\|/\|x\| = \text{the relative error} \]
\[ \|r\|/\|b\| = \text{the relative residual} \]

\[ Ae = A(x - \hat{x}) = b - \hat{b} = r \]

\[ \|e\| = \|A^{-1}r\| \leq \|A^{-1}\| \|r\| \quad \text{and} \quad \|b\| \leq \|A\| \|x\| \]

so

\[ \|e\| \cdot \|b\| \leq \|A\| \cdot \|A^{-1}\| \cdot \|x\| \cdot \|r\| \]

which implies

\[ \frac{\|e\|}{\|x\|} \leq \kappa(A) \frac{\|r\|}{\|b\|}. \]

\[ \kappa(A) = \|A\| \cdot \|A^{-1}\| = \text{condition number of } A. \]
**Condition Number and Error Sensitivity**

\[ \kappa(A) = \|A\| \cdot \|A^{-1}\| = \text{condition number of } A. \]

Note that for any operator norm

\[ 1 = \|I\| = \|AA^{-1}\| \leq \|A\| \|A^{-1}\| = \kappa(A). \]

A matrix is said to be perfectly conditioned if \( \kappa(A) = 1 \), and is said to be ill-conditioned if \( \kappa(A) \) is large.

If \( \hat{x} \) is the result of a numerical algorithm for solving \( Ax = b \)
(with round-off error)
then the error \( e = x - \hat{x} \) is not computable

However, the residual \( r = b - A\hat{x} \) is computable, so we obtain an upper bound on the relative error

\[ \frac{\|e\|}{\|x\|} \leq \kappa(A) \frac{\|r\|}{\|b\|}. \]

In practice, we don’t know \( \kappa(A) \) (although we may be able to estimate it), and this upper bound may be much larger than the actual relative error.
We now assume that both $A$ and $b$ are perturbed to

$$\hat{A} = A + E \quad \text{and} \quad \hat{b} = b - r$$

and obtain a bound on the relative error $\|e\|/\|x\|$ where

$$e = x - \hat{x} \quad \text{with} \quad \hat{x} \text{ solving } \hat{A}\hat{x} = \hat{b}.$$ 

First use the Neumann Lemma to note that if

$$\|A^{-1}E\| \leq \|A^{-1}\| \|E\| < 1,$$

then

$$\|(A + E)^{-1}\| = \|(I + A^{-1}E)^{-1}A^{-1}\|$$

$$\leq \|(I + A^{-1}E)^{-1}\| \|A^{-1}\|$$

$$\leq \frac{1}{1 - \|A^{-1}\| \|E\|} \|A^{-1}\|.$$

Also, $(A + E)x = b + Ex$ and $(A + E)\hat{x} = \hat{b}$ so

$$e = x - \hat{x} = (A + E)^{-1}(Ex + r)$$

and so

$$\|e\| \leq \|(A + E)^{-1}\| \|E\| \|x\| + \|r\|$$

$$\leq \frac{1}{1 - \|A^{-1}\| \|E\|} \|A^{-1}\| \left[ \frac{\|E\|}{\|A\|} \|A\| \|x\| + \frac{\|r\|}{\|b\|} \|A\| \|x\| \right].$$

Therefore,

$$\frac{\|e\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \kappa(A) \|E\| / \|A\|} \left[ \frac{\|E\|}{\|A\|} + \frac{\|r\|}{\|b\|} \right].$$