Normed Vector Spaces

Some of the exercises in these notes are part of Homework 5. If you find them difficult let me know.

In these notes, all vector spaces are either real or complex. Let K denote either $\mathbb{R}$ or $\mathbb{C}$.

1 Normed vector spaces

Definition 1 *Let V be a vector space over K. A **norm** in V is a map $x \mapsto \|x\|$ from V to the set of non-negative real numbers such that*

1. *$\|x\| = 0$ if and only if $x = 0$.*

2. *$\|\alpha x\| = |\alpha| \|x\|$ for all $\alpha \in K$, $x \in V$.*

3. *$\|x + y\| \le \|x\| + \|y\|$ for all $x, y \in V$.*

A *normed vector space* is a real or complex vector space in which a norm has been defined. Formally, one says that a normed vector space is a pair $(V, \|\cdot\|)$ where V is a vector space over K and $\|\cdot\|$ is a norm in V, but then one usually uses the usual abuse of language and refers to V as being the normed space. Sometimes (frequently?) one has to consider more than one norm at the same time; then one uses sub-indices on the norm symbol: $\|x\|_1$, for example. When dealing with several normed spaces it is also customary to refer to the norm of a space denoted by V by the symbol $\|\cdot\|_V$. Other symbols for norms include $|\cdot|$ and $\||\cdot\||$.

Exercise 1 *Let $(V, \|\cdot\|)$ be a normed vector space. Prove*

$$\big|\|x\| - \|y\|\big| \le \|x - y\|$$

for all $x, y \in V$.

Here are some examples of norms for some common finite dimensional spaces. Proving that each one of them is a norm is left as an exercise; some are officially assigned as exercises.

1. Let $1 \le p < \infty$. In $K^n = \mathbb{R}^n$ or $\mathbb{C}^n$ define $\|x\|_p$ if $x = (x_1, \ldots, x_n) \in K^n$ by

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

Then $\|\cdot\|_p$ is a norm in K^n. Proving that this is so for the case $p = 1$ is trivial. For $1 < p$ it is immediate to prove that properties 1 and 2 of the definition above hold. The hard one is 3, the *triangle inequality*. To prove it I will (with your help) prove first the following inequality that, in a somewhat more general context, is known as *Hölder's inequality:* Let $x(x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in K^n$ and assume that $1 < p < \infty$. Let $p' = p/(p-1)$. Notice that $p' > 1$ and $\dfrac{1}{p} + \dfrac{1}{p'} = 1$. Then

$$\sum_{j=1}^{n} |x_j|\, |y_j| \le \|x\|_p \|y\|_{p'} \tag{1}$$

To prove it, I will need a Calculus inequality which I leave as an exercise:

Exercise 2 *With p, p' as above prove that for all $a, b \ge 0$ one has*

$$ab \le \frac{1}{p} a^p + \frac{1}{p'} b^{p'}. \tag{2}$$

Hints: One approach considers the function $\phi : (0, \infty) \to (0, \infty)$ defined by $\phi(x) = \frac{1}{p} x^p + \frac{1}{p'} x^{-p'}$. A bit of calculus shows that $\phi(x) \geq 1$ for all $x > 0$ and the inequality follows applying ϕ to $x = a^{1-1/p} b^{-1/p}$. In proving (2) one may assume that $a > 0, b > 0$, the case $a = 0$ or $b = 0$ being somewhat very trivial.

With (2 proved, notice that if $\|x\|_p = 0$ or $\|y\|_{p'} = 0$, then (1) is trivial, so assume both $\|x\|_p, \|y\|_{p'}$ are positive. Let $a_i = |x_i|/\|x\|_p$, $b_i = |y_i|/\|y\|_p$ and by (2)

$$a_i b_i \leq \frac{1}{p} a_i^p + \frac{1}{p'} b_i^{p'}.$$

Adding from $i = 1$ to n gives

$$\frac{\sum_{i=1}^{n} |x_i| \, |y_i|}{\|x\|_p \|y\|_{p'}} \leq \frac{1}{p} + \frac{1}{p'} = 1.$$

This proves (1). With (1) proved, let again $x(x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in K^n$. For the computation to follow, it helps to notice that $p'(p-1) = p$. We have

$$
\begin{aligned}
\|x+y\|_p^p &= \sum_{i=1}^{n} |x_i + y_i|^p = \sum_{i=1}^{n} |x_i + y_i| |x_i + y_i|^{p-1} \leq \sum_{i=1}^{n} (|x_i| + |y_i|) \, |x_i + y_i|^{p-1} \\
&\leq \sum_{i=1}^{n} |x_i| |x_i + y_i|^{p-1} + \sum_{i=1}^{n} |y_i| |x_i + y_i|^{p-1} \\
&\leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^{p'(p-1)} \right)^{1/p'} + \left(\sum_{i=1}^{n} |y_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^{p'(p-1)} \right)^{1/p'} \\
&= \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^p \right)^{1/p'} + \left(\sum_{i=1}^{n} |y_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^p \right)^{1/p'} \\
&= = \|x\|_p \|x+y\|_p^{p/p'} + \|y\|_p \|x+y\|_p^{p/p'} = (\|x\|_p + \|y\|_p) \|x+y\|_p^{p-1}.
\end{aligned}
$$

Dividing by $\|x+y\|_p^{p-1}$, the triangle inequality follows.

2. Let $x = (x_1, \ldots, x_n) \in K^n$. We define $\|x\|_\infty$ by

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Proving that this is a norm is fairly trivial.

3. Let V, W be normed vector spaces over K. I'll use the same symbol for the norm in V as in W; this shouldn't cause any confusion (one hopes). If $T \in L(V, W)$ define

$$\|T\| = \sup\{\|Tx\| \; : \; x \in V, \|x\| \leq 1\}.$$

The norm $\|Tx\|$ is the norm in W; $\|x\|$ is the norm in V. Some people would write

$$\|T\| = \sup\{\|Tx\|_W \; : \; x \in V, \|x\|_V \leq 1\}.$$

As an exercise below shows, this norm can be infinite, but only if both W, V are infinite dimensional. To cover all cases we define

$$B(V, W) = \{T \in L(V, W) \; : \; \|T\| < \infty\}.$$

Then $B(V, W)$ is a normed vector space over K.

Exercise 3 *Let ℓ^1 be the space of all absolutely convergent sequences of real numbers. That is $a \in \ell^1$ iff $a = \{a_n\}$ with $\sum_{n=1}^{\infty} |a_n| < \infty$. If $a = \{a_n\} \in \ell^1$ we define $\|a\|_1 = \sum_{n=1}^{\infty} |a_n|$.*

1. *Prove $(\ell^1, \|\cdot\|_1)$ is a normed vector space.*

2. *Define $S : \ell^1 \to \ell^1$ by*

$$S(a_1, a_2, a_3, \ldots) = (a_2, a_3, \ldots).$$

Prove $\|S\| = 1$.

3. *Let $V = \{\{a_n\} \ : \ a_n \in \mathbb{R}$ for $n \in \mathbb{N}, \exists N \in \mathbb{N}(\ not\ always\ the\ same\ N), a_n = 0$ if $n > N\}$. Then V is a subspace of ℓ^1, a normed space with the norm of ℓ^1. Define $T : V \to \ell^1$ by $T\{a_n\} = \{na_n\}$. Prove $\|T\| = \infty$.*

Exercise 4 *Let V, W be K-vector spaces.*

1. *Prove that $B(V, W)$ is a vector space over K and $\| \cdot \|$ is a norm for $B(V, W)$. Prove that the following definitions of $\|T\|$ are equivalent (in the sense of giving an identical value):*

$$\|T\| = \sup\{\|Tx\| \ : \ x \in V, \|x\| = 1\}.$$

and

$$\|T\| = \inf\{M \in \mathbb{R} \ : \ \|Tx\| \le M\|x\|, x \in V\}.$$

2. *Prove: If (at least) one of V, W is finite dimensional, then $B(V, W) = L(V, W)$. To do this, use Theorem 2 below!*

Exercise 5 *Let V be a normed vector space over K. Write $B(V)$ for $B(V, V)$. Prove that in addition of being a normed vector space, $B(V)$ satisfies:*

1. *$\|I\| = 1$.*

2. *$\|ST\| \le \|S\| \|T\|$ for all $S, T \in B(V)$.*

If V is a normed space, then defining $d(x, y) = \|x - y\|$ for $x, y \in V$, it is immediate that d is a distance function for V. One always considers V as a metric space with this distance. All the metric notions are thus valid in V. If $x \in V$, and $r > 0$ we will denote by $B(x, r)$ the open disk centered at x of radius r;

$$B(x, r) = \{y \in V \ : \ \|y - x\| < r\}.$$

A subset U of V is thus open iff for each $x \in U$, there is $r > 0$ such that $B(x, r) \subset U$. If $x \in V$, $A \subset V$ we define the translate of A by x by $x + A = \{x + a : \alpha \in A\}$. Then

$$B(x, r) = x + B(0, r).$$

Exercise 6 *Let V be a normed vector space. prove the following maps are homeomorphisms of V onto V:*

1. *For each $a \in V$, the map $x \mapsto x + a$.*

2. *The map $x \mapsto -x$.*

3. *for each $\alpha \in K$, $\alpha \ne 0$, the map $x \mapsto \alpha x$.*

Talking of continuity, here is an immediate consequence of Exercise 1, but it is good to have it written down.

Exercise 7 *If V is a normed vector space, the map $x \mapsto \|x\| : V \to \mathbb{R}$ is continuous.*

In a metric space, in particular in a normed vector space, all topological notions can be defined in terms of sequences. In a normed space V a sequence $\{x_n\}$ converges to $x \in V$ if and only if $\lim_{n\to\infty} \|x_n - x\| = 0$. It is a Cauchy sequence iff for every $\epsilon > 0$ there is N such that $\|x_n - x + m\| < \epsilon$ whenever $n, m \ge N$. The space is said to be *complete* iff all Cauchy sequences converge.

Definition 2 *A **Banach space** is a normed vector space that is complete.*

Exercise 8 *Let V be a normed vector space. As in $\mathbb{R}$ or $\mathbb{C}$ it makes sense to consider series. If $x_n \in V$ for $n = 1, 2, 3, \dots$ we say that the series $\sum_{n=1}^{\infty} x_n$ converges in V iff there is $x \in V$ such that*

$$\lim_{n \to \infty} \left\| \sum_{k=1}^{n} x_k - x \right\| = 0.$$

*The element x is then unique and called the **sum** of the series. We say that the series $\sum_{n=1}^{\infty} x_n$ **converges absolutely** iff $\sum_{n=1}^{\infty} \|x_n\| < \infty$. Prove: A normed vector space is complete (and hence a Banach space) if and only if every absolutely convergent series converges.*

Definition 3 *Assume V is a vector space and let $\| \cdot \|_1, \| \cdot \|_2$ be two norms for V. We say they are equivalent iff there exist positive constants a, b such that*

$$a\|x\|_1 \leq \|x\|_2 \leq b\|x\|_1 \tag{3}$$

for all $x \in V$

Exercise 9 *Prove that equivalence of norms is indeed an equivalence relation for the set of all norms of a vector space V.*

We have the following simple result.

Theorem 1 *Let V be a vector space over K and let $\| \cdot \|_1, \| \cdot \|_2$ be two norms for V. The following statements are equivalent.*

1. *The norms $\| \cdot \|_1$ and $\| \cdot \|_2$ are equivalent.*

2. *If we denote by $B_1(x, r)$ the ball of radius r with respect to $\| \cdot \|_1$ and by $B_2(x, r)$ the ball of radius r with respect to $\| \cdot \|_2$, then for every $r > 0$ there is $\rho > 0$ such that $B_1(0, \rho) \subset B_2(0, r)$, and for each $r > 0$ there is $\rho > 0$ such that $B_2(0, \rho) \subset B_1(0, r)$.*

3. *A set is open with respect to $\| \cdot \|_1$ if and only if it is open with respect to $\| \cdot \|_2$.*

4. *A sequence $\{x_n\}$ in V converges to a point $x \in V$ with respect to $\| \cdot \|_1$ if and only if it converges to the same point x with respect to $\| \cdot \|_2$.*

5. *A sequence $\{x_n\}$ in V converges to 0 with respect to $\| \cdot \|_1$ if and only if it converges to 0 with respect to $\| \cdot \|_2$.*

Proof. (**a**) $\Rightarrow$ (**b**) Assume the norms are equivalent, so that there exists $a, b > 0$ so (3) holds for all $x \in V$. Let $r > 0$ be given and take $\rho = r/b$. If $x \in B_1(0, \rho)$, then $\|x\|_1 < r/b$, thus $\|x\|_2 \leq b\|x\|_1 < r$ and $x \in B_2(0, r)$. Similarly, taking $\rho = r/a$ we see that $x \in B_2(0, \rho)$ implies $x \in B_1(0, r)$.
(**b**) $\Rightarrow$ (**c**) Assume (*b*) and let U be open with respect to $\| \cdot \|_1$. let $x \in U$. There is $r > 0$ such that

$$B_1(x, r) = x + B_1(0, r) \subset U.$$

Buy (*b*) there is $\rho > 0$ such that $B_2(0, \rho) \subset B_1(0, r)$. Thus

$$B_2(x, \rho) = x + B_2(0, \rho) \subset x + B_1(0, r) \subset U.$$

It follows that U is open with respect to $\| \cdot \|_2$. Similarly one proves that open in $\| \cdot \|_2$ implies open in $\| \cdot \|_1$.
(**c**) $\Rightarrow$ (**d**) Since the notion of convergence can be described exclusively in terms of open sets, this is trivial.
(**d**) $\Rightarrow$ (**e**) Absolutely trivial.
(**e**) $\Rightarrow$ (**d**) In a normed space, a sequence $\{x_n\}$ converges to x if and only if $\{x_n - x\}$ converges to 0. The implication is now immediate.
(**d**) $\Rightarrow$ (**a**) Assume (*e*). Suppose the first inequality of (3) does not hold; that is suppose there is no $a > 0$ such that $a\|x\|_1 \leq \|x\|_2$ for all $x \in V$. In this case for each $n \in \mathbb{N}$ there is $x_n \in V$ such that $frac1n\|x_n\|_1 > \|x_n\|_2$. In particular, $\|x_n\|_1 \neq 0$; setting $y_n = x_n/\|x_n\|_1$. Then $\|y_n\|_2 \leq \frac{1}{n}$, so $y_n \to 0$ in norm 2. But $\|y_n\|_1 = 1$, so with respect to norm 1, $\{y_n\}$ does not converge to 0, contradicting the assumption. Similarly one sees that there must be a $b > 0$ so the second inequality of (3) holds. ∎

We now come to a famous result. But first another very simple construction/exercise. Let V be a finite dimensional vector space over K and let $\{e_1, \ldots, e_n\}$ be a basis of V. We will define the $\|\cdot\|_1$ norm of V (which depends on the basis) by: if $x = \sum_{i=1}^n \xi e_i \in V$, set $\|x\|_1 = \sum_{i=1}^n |\xi_i|$. With this norm V behaves exactly like K^n ($\mathbb{R}^n$ or $\mathbb{C}^n$) with the usual metric, in particular the Heine-Borel Theorem holds: Closed and bounded subsets are compact. A complete proof of the Heine Borel Theorem, in this context, in all of its gory details, is given as an Appendix to these notes. Here I might amuse (or annoy) you with a digression about boundedness. In a metric space to be bounded means very little. If (X, d) is a metric, one can always define $d'(x, y) = \min(d(x, y), 1)$ and get an equivalent metric in which ALL sets are bounded. But it means more in a normed space; if S is bounded with respect to one norm, it will clearly be bounded with respect to all norms equivalent to that norm. So being bounded is a stronger property in normed spaces than it is in metric spaces. This was the digression.

Here's the promised theorem.

Theorem 2 *Let V be a finite dimensional K-vector space. **All** norms of V are equivalent.*

Proof. Let $\|\cdot\|$ be a norm in V. Assume V is of dimension n and let $\{e_1, \ldots, e_n\}$ be a basis of V. It will suffice to prove $\|\cdot\|$ is equivalent to $\|\cdot\|_1$. In one direction it is very easy. Let

$$b = \max(\|e_1\|, \ldots, \|e_n\|).$$

Then if $x \in V$, writing $x = \sum_{i=1}^n \xi e_i$ we have

$$\|x\| = \left\|\sum_{i=1}^n \xi e_i\right\| \leq \sum_{i=1}^n |\xi| \|e_i\| \leq b\|x\|_1.$$

For the converse inequality, we prove first that if we consider V as a metric space with the $\|\cdot\|_1$ norm, then the map $\phi : x \mapsto \|x\|$ is continuous. In fact, let $x_0 \in V$ and let $\epsilon > 0$ be given. let $\delta = \epsilon/b$. If $\|x - x_0\|_1 < \delta$ then, by what we proved

$$|\phi(x) - \phi(x_0)| = \big|\|x\| - \|x_0\|\big| \leq \|x - x_0\| \leq b\|x - x_0\|_1 < b\delta = \epsilon.$$

We continue considering V as a metric space in the $\|\cdot\|_1$ norm. In this norm, the set $S = \{x \in V : \|x\|_1 = 1\}$ is closed and bounded, hence compact (as mentioned above). Thus the continuous function ϕ assumes a minimum value on S; there is $x_0 \in S$ such that $\phi(x_0) \leq \phi(x)$ for all $x \in S$, Since $0 \notin S$, we see that $x_0 \neq 0$, hence $\phi(x_0) > 0$. Let $a = \phi(x_0)$, so $a > 0$. Then what we proved is that $a \leq \|x\|$ for all $x \in V$ with $\|x\|_1 = 1$. If $x \in V$ and $x \neq 0$, let $y = x/\|x\|_1$, so $\|y\|_1 = 1$. Then

$$a \leq \|y\| = \left\|\frac{1}{\|x\|_1}x\right\| = \frac{\|x\|}{\|x\|_1};$$

that is, we proved $a\|x\|_1 \leq \|x\|$ for all $x \in V$, $x \neq 0$. Since this last inequality is also trivially true for $x = 0$, we are done. $\blacksquare$

The following theorem is now very easy:

Theorem 3 *Let V be a normed vector space. If $\dim V < \infty$, then V is complete.*

Proof. Let $\{x_n\}$ be a Cauchy sequence in V. As is well known, Cauchy sequences are bounded, so $\{x_n\}$ can be seen as being included in some closed and bounded subset of V, hence inside a compact set, hence it has a convergent subsequence. Cauchy sequences that have convergent subsequences converge. $\blacksquare$

A sort of nice application of all this is:

Theorem 4 *Let V be a normed vector space and let W be a finite dimensional subspace of V. Then W is a closed subset of V.*

Proof. We use the easily proved fact that a subset of a metric space that is complete in the metric is closed. Because W is finite dimensional, the norm of W is equivalent to the $\|\cdot\|_1$ norm for W (with respect to any basis one may care to choose), hence W is complete, hence closed. $\blacksquare$

Here are a few more facts about normed spaces at a general level that are good to know. Heine Borel fails to be true if the dimension is infinite. That is, in every infinite normed space there are closed and bounded sets that are

not compact. If the space is not only infinite dimensional but also complete, then any basis has to be uncountable. This makes algebraic bases less than useful when discussing infinite dimensional normed spaces. That last result is a consequence of the so called Baire category theorem; I say so called because it can be stated in a very simple way, without all the nonsense about sets of the first or second category. It states: *Let X be a **complete** normed vector space and let A_n be a sequence of closed subsets of X with empty interior. Then $\bigcup_{n=1}^{\infty} A_n$ also has empty interior; in particular, it can't be X.* But I want to use all this to look at matrices.

2 Series of Matrices

In this section I will assume the field is $\mathbb{C}$. If you are still a bit stuck emotionally in the late 18th to early 19th century, and find complex numbers very disturbing, you may replace the word "complex" by the word "real," and the symbol $\mathbb{C}$ by the symbol $\mathbb{R}$, in all occurrences. We consider $\mathbb{C}^n$ as a normed space, with the Euclidean norm; which we denote with the same symbol that we use for the absolute value:

$$|z| = \sqrt{|z_1|^2 + \cdots + |z_n|^2} = \sqrt{|x_1|^2 + |y_1|^2 + \cdots + |x_n|^2 + |y_n|^2}$$

if $z = (z_1, \ldots, z_n) = (x_1 + iy_1, \ldots, x_n + iy_n)$.

We will consider $M_n(\mathbb{C}) = L(\mathbb{C}^n) = B(\mathbb{C}^n)$ as a normed vector space with the norm $\|A\|$ of a matrix A defined by

$$\|A\| = \sup\{|Az| \,:\, z \in \mathbb{C}^n, |z| \leq 1\};$$

in other words, the operator norm.

We now have the following quite simple theorem:

Theorem 5 *Assume f is a complex valued function of a complex variable defined by a power series of radius of convergence $r > 0$, with $r = \infty$ allowed; specifically, assume*

$$f(z) = \sum_{k=0}^{\infty} a_k z^k$$

for $z \in \mathbb{C}$, $|z| < r$. If $A \in M_n(\mathbb{C})$ and $\|A\| < r$. Then $\sum_{k=0}^{\infty} a_k A^k$ converges in $M_n(\mathbb{C})$.

Proof. If $\|A\| < r$, then the series $\sum_{k=0}^{\infty} a_k \|A\|^k$ converges absolutely; the result is thus a consequence of Exercise 8 and the completeness of the finite dimensional space $M_n(\mathbb{C})$. ∎

Because of the equivalence of all norms in a finite dimensional space one has that convergence in one norm implies convergence in any other norm. A matrix norm one can use is

$$\|(a_{ij})\| = \max_{1 \leq i,j \leq n} |a_{ij}|$$

Convergence in this norm means that for every (i, j), the sequence formed by taking the (i, j)-th entry of $\sum_{k=0}^{m} a_k A^k$ for $m = 0, 1, 2, \ldots$ converges to some complex number b_{ij}. Or, to be more specific: If

$$\sum_{k=0}^{m} a_k A^k = (b_{ij}^{(m)})_{1 \leq i,j \leq n},$$

then $\{\sum_{k=0}^{m} a_k A^k\}$ converges to $B = (b_{ij})$ in one, hence all, matrix norms, if and only if $\lim_{m \to \infty} b_{ij}^{(m)} = b_{ij}$ for all i, j, $1 \leq i, j \leq n$.

A case of particular interest of Theorem 5 is the case in which the radius of convergence of the series defining f is ∞; in this case f can be applied to all matrices. Thus, for all matrices $A \in M_n(C)$ we have defined e^A, $\sin A$, $\cos A$, etc.

Exercise 10 *Which of these statements are true for all matrices $A \in M_n(\mathbb{C})$?*

1. $\det(e^A) = e^{\det A}$.

2. e^A is invertible and $\left(e^A\right)^{-1} = e^{-A}$.

3. $(e^A)^k = e^{kA}$ for $k = 1, 2, \ldots$.

4. $\lambda \in \sigma(A)$ if and only if $e^\lambda \in \sigma(e^A)$.

Well, I'll suppose we gave five minutes of thought to Exercise 10. The first property is rather obviously false, what is true however is that $\det(e^A) = e^{\operatorname{tr}A}$. I defined the trace previously for operators; a basis (hence a matrix) is needed to compute it. The operator definition might be easiest to use if one wants to show that similar matrices have the same trace, but not much easier. Let $A, B \in M_n(F)$ (here F could again be an arbitrary field) and suppose $B = U^{-1}AU$, where U is invertible. For matrices, the trace is the sum of the diagonal elements; if we write $A = (a_{ij})$, $B = (b_{ij})$, $U = (u_{ij})$, $U^{-1} = (v_{ij})$, then

$$\operatorname{tr}(B) = \sum_{i=1}^{n} b_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} u_{ik}a_{kj}v_{ji} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} v_{ji}u_{ik}a_{kj} = \sum_{j=1}^{n}\sum_{k=1}^{n}\left(\sum_{i=1}^{n} v_{ji}u_{ik}\right)a_{kj}$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} \delta_{jk}a_{kj} = \sum_{j=1}^{n} a_{jj} = \operatorname{tr}(A).$$

There are a number of ways of seeing that the trace is also the sum of all eigenvalues, counted by their multiplicity. Fr example, having seen that it is invariant under similarity, one can take the matrix to Jordan form, where the trace is precisely that sum because the diagonal is occupied by the eigenvalues. Here is a more direct way. Suppose you use the following expression for the determinant of a matrix $A \in M_n(F)$:

$$\det(A) = \sum_{\sigma \in \mathfrak{S}_n} \epsilon(\sigma)a_{1\sigma(1)}\cdots a_{n\sigma(n)}.$$

Replace A by $\lambda I - A$ and ask yourself what terms in this expression for the determinant will have a power of λ^{n-1}? In computing the determinant by this method you form $n!$ products; for each product you choose an element of the first row, one from the second, and so forth, making always sure that no two chosen elements fall in the same column. To get a power of λ^{n-1} you have to select **at least** the diagonal element $\lambda - a_{jj}$ from each row but one (otherwise you do not have enough powers of λ); since the element of this one exceptional row cannot be under any of the chosen elements, it also has to be diagonal. To be brief, the λ^{n-1} powers come from the same source as the λ^n power, the one term of the sum corresponding to the identity element of $\mathfrak{S}_n$; to

$$\prod_{j=1}^{n}(\lambda - a_{jj}).$$

The other $n! - 1$ terms of the sum have powers of $\lambda < n - 1$¿ If we expand $\prod_{j=1}^{n}(\lambda - a_{jj})$ we see it has the form

$$\prod_{j=1}^{n}(\lambda - a_{jj}) = \lambda^n - (a_{11} + \cdots + a_{nn})\lambda^{n-1} + \text{ lower order terms} = \lambda^n - \operatorname{tr}(A)\lambda^{n-1} + \text{ lower order terms.}$$

The conclusion is that the characteristic polynomial of A has the form $p(\lambda) = \lambda^n - \operatorname{tr}(A)\lambda^{n-1} + $ lower order terms. If $\sigma(A) \subset F$, then it also has the form

$$p(\lambda) = \prod_{j=1}^{n}(\lambda - \lambda_j) = \lambda^n - (\lambda_1 + \cdots + \lambda_n)\lambda^{n-1} + \text{ lower order terms;}$$

it follows that $\operatorname{tr}(A) = \sum_{j=1}^{n} \lambda_j$; where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, counted according to their multiplicities.
An example(Calculations done by Maple)
For the rest of the questions, here is a more detailed exercise.

Exercise 11 *Let $A \in M_n(\mathbb{C})$. Consider A as an operator in $L(\mathbb{C}^n)$.*

1. *Show that if V is a subspace invariant for A, then it is also invariant for e^A. This is not terribly difficult, but could have a slight twist at the end.*

2. *Show that if λ is an eigenvalue of A, then e^λ is an eigenvalue of e^A.*

Is the converse true, are all eigenvalues of e^A of the form e^λ? Is it even true?

   If all eigenvalues of A are distinct; that is, if A has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, it is trivially true from this last exercise; then e^A has the eigenvalues $e^{\lambda_1}, \ldots, e^{\lambda_n}$ and since it can't have more than n eigenvalues, those are all the eigenvalues. Otherwise, things are a bit more complicated. One proof, maybe there is a simpler one, is to write

$$\mathbb{C}^n = \bigoplus_{i=1}^r V_i,$$

where V_i is the generalized eigenspace corresponding to the eigenvalue λ_i of A. Restrict e^A to V_i. In this space it has the eigenvalue e^{λ_i}. Can it have any other? Let us suppose that there is $v \in V_i$ such that $e^A v = \mu v$. Now consider the subspace W of V_i generated by $\{v, Av, A^2 v, \ldots\}$ (Of course, there is only a finite number of linearly independent vectors in this list). This space is rather clearly invariant for A, hence also for e^A. For e^A it is an eigenspace; we have for $k \in \mathbb{N} \cup \{0\}$, $e^A(A^k v) = A^k e^A v = \mu A^k v$, so in W e^A behaves like $e^\mu I$; thus has only μ as eigenvalue. But A must also have an eigenvalue in W. Since $W \subset V_i$, the only eigenvalue A can have is λ_i, and it now follows that $\mu = e^{\lambda_i}$. The conclusion is that the generalized eigenspace corresponding to the eigenvalue λ_i of A is a subspace of the one of e^{λ_i} of e^A.

   I wrote that the generalized eigenspace corresponding to e^{λ_i} of e^A contains the one of A corresponding to λ_i. Could it be larger? You may answer this question for yourself by doing the following amusingly simple exercise.

Exercise 12 *In $\mathbb{C}^2$ consider the operator of matrix*

$$A = \begin{pmatrix} 0 & 2\pi \\ -2\pi & 0 \end{pmatrix}$$

Evaluate e^A and determine $\sigma(A)$ and $\sigma(e^A)$.

   I conclude this section with an example. All calculations done by Maple.

EXAMPLE

   Consider the matrix

$$A = \begin{pmatrix} 3 & 5 & 1 \\ 24 & 5 & -2 \\ -39 & -15 & -13 \end{pmatrix}.$$

Let us try to find e^{tA}, $t \in \mathbb{R}$. Finding it directly takes heroic efforts. Here are some powers of A:

$$A^0 \;=\; I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$A^2 \;=\; \begin{pmatrix} 90 & 25 & -20 \\ 270 & 175 & 40 \\ 30 & -75 & 160 \end{pmatrix}$$

$$A^3 \;=\; \begin{pmatrix} 1650 & 875 & 300 \\ 3450 & 1625 & -600 \\ -7950 & -2625 & -1900 \end{pmatrix}$$

$$A^4 \;=\; \begin{pmatrix} 14250 & 8125 & -4000 \\ 72750 & 34375 & 8000 \\ -12750 & -24375 & 22000 \end{pmatrix}$$

Apart from the fact that the entries are getting larger and larger, do you see any pattern emerging? I don't. Perhaps if one computed a few more powers one would see something, perhaps not.

However, the Jordan normal form of A is

$$
J = \begin{pmatrix} 15 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 1 & -10 \end{pmatrix}
$$

Now

1. There is an invertible matrix U such that $A = UJU^{-1}$, and

2. From the Jordan form one gets at once the $S + N$ decomposition; in fact, $J = S + N$ where

$$
S = \begin{pmatrix} 15 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 0 & -10 \end{pmatrix} \qquad N = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}
$$

Notice that $SN = NS$, as it should be. Because of this, and because $N^2 = 0$, we have

$$
\begin{aligned}
e^{tJ} &= \sum_{k=0}^{\infty} \frac{t^k}{k!}(S+N)^k = \sum_{k=0}^{\infty} \frac{t^k}{k!}(S^k + kS^{k-1}N) = \sum_{k=0}^{\infty} \frac{t^k}{k!}S^k + \sum_{k=1}^{\infty} \frac{t^k}{(k-1)!}S^{k-1}N \\
&= \sum_{k=0}^{\infty} \frac{t^k}{k!}S^k + \sum_{k=0}^{\infty} \frac{t^k}{(k)!}S^k tN = e^{tS}(1+tN) = \begin{pmatrix} e^{15t} & 0 & 0 \\ 0 & e^{-10t} & 0 \\ 0 & 0 & e^{-10t} \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & t & 1 \end{pmatrix} \\
&= \begin{pmatrix} e^{15t} & 0 & 0 \\ 0 & e^{-10t} & 0 \\ 0 & te^{-10t} & e^{-10t} \end{pmatrix}
\end{aligned}
$$

(In this computation we used that $k/k! = 0$ if $k = 0$; otherwise it is $1/(k-1)!$)

A matrix that has the property that $A = UJU^{-1}$ is

$$
U = \begin{pmatrix} -1/3 & 0 & 1 \\ -1 & 0 & -2 \\ 1 & 1 & -3 \end{pmatrix}
$$

(See my notes on Jordan forms to see how such a U can be obtained.) Then

$$
\begin{aligned}
e^{tA} &= Ue^{tJ}U^{-1} \\
&= \begin{pmatrix} -\frac{1}{3} & 0 & 1 \\ -1 & 0 & -2 \\ 1 & 1 & -3 \end{pmatrix}\begin{pmatrix} e^{15t} & 0 & 0 \\ 0 & e^{-10t} & 0 \\ 0 & te^{-10t} & e^{-10t} \end{pmatrix}\begin{pmatrix} -\frac{6}{5} & -\frac{3}{5} & 0 \\ 3 & 0 & 1 \\ \frac{3}{5} & -\frac{1}{5} & 0 \end{pmatrix} \\
&= \begin{pmatrix} \frac{2}{5}e^{15t} + 3te^{-10t} + \frac{3}{5}e^{-10t} & \frac{1}{5}e^{15t} - \frac{1}{5}e^{-10t} & te^{-10t} \\ \frac{6}{5}e^{15t} - 6te^{-10t} - \frac{6}{5}e^{-10t} & \frac{3}{5}e^{15t} + \frac{2}{5}e^{-10t} & -2te^{-10t} \\ -\frac{6}{5}e^{15t} + \frac{6}{5}e^{-10t} - 9te^{-10t} & -\frac{3}{5}e^{15t} + \frac{3}{5}e^{-10t} & e^{-10t} - 3te^{-10t} \end{pmatrix}
\end{aligned}
$$

3 Hilbert Spaces, and the return to Halmos

We continue assuming $K = \mathbb{C}$ or $\mathbb{R}$.

Definition 4 *An inner **product space**, also known as a **pre-Hilbert space**, is a real or complex vector space V in which there is defined a map assigning to each pair of vectors x, y of V a real number denoted by (x, y) in the case of a real vector space, a complex number also denoted by (x, y) for the complex variety, such that the following properties hold:*

1. *$(x, x) \in \mathbb{R}$ and $(x, x) \geq 0$ for all $x \in V$.*

2. *$(x, x) = 0$ if and only if $x = 0$.*

3. *$(y, x) = \overline{(x, y)}$ for all $x, y \in V$. Here $\bar{z}$ denotes the complex conjugate of z if z is a complex number. For real inner product spaces, this property reads as $(x, y) = (y, x)$.*

4. *$(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z)$ for all $x, y, z \in V$, $\alpha, \beta \in K$.*

Here are a few obvious consequences of this definition. In the real case, the map $V \times V \to \mathbb{R} : (x, y) \mapsto (x, y)$ (here (x, y) is an ordered pair in its first appearance, the value the inner product assigns to the pair (x, y) in its second appearance) is a symmetric bilinear map;

$$(x, \alpha y + \beta z) = (\alpha y + \beta z, x) = \alpha(y, x) + \beta(z, x) = \alpha(x, y) + \beta(x, z).$$

In the complex case the map is linear in the first component, conjugate linear in the second:

$$(x, \alpha y + \beta z) = \overline{(\alpha y + \beta z, x)} = \overline{\alpha(y, x) + \beta(z, x)} = \bar{\alpha}\overline{(y, x)} + \bar{\beta}\overline{(x, z)} = \bar{\alpha}(x, y) + \bar{\beta}(x, z).$$

For a while one can avoid dividing into real and complex cases by just keeping in mind that complex conjugation restricted to $\mathbb{R}$ is the identity. We have for all $x, y, z, w \in V$, $\alpha, \beta, \gamma, \delta \in K$,

$$(\alpha x + \beta y, \gamma z + \delta w) = \alpha\bar{\gamma}(x, z) + \alpha\bar{\delta}(x, w) + \beta\bar{\gamma}(y, z) + \beta\bar{\delta}(y, w).$$

For future reference let us notice that if $\alpha = \gamma$, $\beta = \delta$, $x = z$, $y = w$, this becomes

$$(\alpha x + \beta y, \alpha x + \beta y) = |\alpha|^2(x, x) + 2\mathrm{Re}\,\left(\alpha\bar{\beta}(x, y)\right) + |\beta|^2(y, y). \tag{4}$$

The reason for the appearance of $2\mathrm{Re}\,\left(\alpha\bar{\beta}(x, y)\right)$ is that

$$\alpha\bar{\beta}(x, y) + \beta\bar{\alpha}(y, x) = \alpha\bar{\beta}(x, y) + \bar{\alpha}\beta\overline{(x, y)} = \alpha\bar{\beta}(x, y) + \overline{\alpha\bar{\beta}(x, y)} = 2\mathrm{Re}\,\left(\alpha\bar{\beta}(x, y)\right).$$

Here are two typical examples of inner product spaces:

1. $\mathbb{C}^n = \{z = (z_1, \ldots, z_n) : z_1, \ldots, z_n \in \mathbb{C}\}$ becomes an inner product space with the dot product:

$$(z, w) = z \cdot \bar{w} = \sum_{j=1}^{n} z_j \bar{w}_j.$$

If we replace $\mathbb{C}$ by $\mathbb{R}$, we get $\mathbb{R}^n$ as a real inner product space.

2. Let $a, b \in \mathbb{R}$, $a < b$, and let $V = \{f : [a, b] \to \mathbb{C} : f \text{ is continuous}\}$. For $f, g \in V$ define

$$(f, g) = \int_a^b f(x)\bar{g}(x)\, dx.$$

Every inner product space becomes a normed space defining $\|x\| = (x, x)^{1/2}$. This is well defined because (x, x) is always a non-negative real number. ALL norm properties are trivially satisfied, except the triangle inequality

which is not so trivially satisfied. To prove it holds we show first that the following Cauchy-Schwarz inequality holds for all $x, y \in V$. Assuming, of course, V is an inner product space.

$$|(x, y)|^2 \leq (x, x)(y, y). \tag{5}$$

Here is a quick proof[1]. Let $x, y \in V$. If $x = 0$, then $(x, y) = 0$ for all $y \in V$. In fact, that follows from the fact that $y \mapsto (x, y)$ is conjugate linear. Similarly, $y = 0$ implies $(x, y) = 0$. We may thus assume $x \neq 0 \neq y$, and then $x, x) > 0$, $(y, y) > 0$. Let $\lambda \in K$. Then

$$0 \leq (x + \lambda y, x + \lambda y) = (x, x) + 2\mathrm{Re}\ \left(\bar{\lambda}(x, y)\right) + |\lambda|^2 (y, y).$$

This is true for every λ in the field. If we take $\lambda = -(x, y)/(y, y)$ we get that

$$\bar{\lambda}(x, y) = -\overline{(x, y)}(x, y)/(y, y) = -\frac{|(x, y)|^2}{(y, y)}, \quad |\lambda|^2 (y, y) = \frac{|(x, y)|^2}{(y, y)^2}(y, y) = \frac{|(x, y)|^2}{(y, y)}$$

and we obtained

$$0 \leq (x, x) + 2\mathrm{Re}\ \left(\bar{\lambda}(x, y)\right) + |\lambda|^2 (y, y) = (x, x) - 2\frac{|(x, y)|^2}{(y, y)} + \frac{|(x, y)|^2}{(y, y)} = (x, x) - \frac{|(x, y)|^2}{(y, y)}.$$

(5) follows. We can also write (5) in the form

$$|(x, y)| \leq \|x\|\,\|y\| \tag{6}$$

for all $x, y \in V$. It is now easy to prove the triangle inequality. Recall that for every $z \in \mathbb{C}$, $|\Re z| \leq |z|$. Let $x, y \in V$.

$$\begin{aligned} \|x + y\|^2 &= (x + y, x + y) = (x, x) + 2\Re(x, y) + (y, y) = \|x\|^2 + 2\Re(x, y) + \|y\|^2 \leq \|x\|^2 + 2|(x, y)| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2. \end{aligned}$$

The result follows taking square roots.

Definition 5 *A **Hilbert space** is an inner product space that is complete in the norm; briefly, a complete inner product space.*

It follows that all Hilbert spaces are Banach spaces. By Theorem 3, all finite dimensional inner product spaces are Hilbert spaces.

The inner product generalizes the dot product and it allows us to talk of angles. One could use it to define, in general, the angle between vectors x, y by $\cos \angle(x, y) = (x, y)/(\|x\|\,\|y\|)$, assuming neither x nor y is 0. I'm not sure how much this is used, except for one case, the $90°$ angle.

Definition 6 *Let V be an inner product space, and let $x, y \in V$. We say x and y are mutually **orthogonal**, and write $x \perp y$, iff $(x, y) = 0$. If $S \subset V$, we also define the set $^\perp$ by*

$$S^\perp = \{x \in V : x \perp y \,\forall\, y \in S\}.$$

Notice that $S \cap S^\perp = \{0\}$ (if $S \neq \emptyset$); in fact $x \perp x$ clearly implies $x = 0$. The reader, assuming there is one, should have no trouble proving the following additional properties of orthogonality:

Exercise 13 *Let V be an inner product space.*

1. *If $S \subset V$, then $S^\perp$ is a subspace of V.*

2. *If $S \subset V$, then $\subset S^{\perp\perp} := \left(S^\perp\right)^\perp$.*

3. *$V^\perp = \{0\}$, $\{0\}^\perp = V$.*

---

[1]Halmos has an even quicker cute proof, based on Bessel's inequality.

I do not want to go to deep into the vastness of an infinite dimensional space, even if it is a really friendly space like a Hilbert space. But we will go at least as far as Halmos, probably a bit more. As mentioned before, bases are not terribly useful in infinite dimensional spaces; they are too small, one needs too many linear combinations. But in Hilbert spaces there is something almost as good, a set of vectors known by some authors as a complete orthonormal set, by others as a maximal orthonormal set, or an orthonormal basis. The key word is *orthonormal*, so we begin with it.

Definition 7 *Let V be an inner product space. A subset $\mathcal{O}$ of V is said to be an* **orthonormal set** *or an* **orthonormal system** *iff $(x,x) = 1$ for all $x \in \mathcal{O}$ (the normal part) and $(x,y) = 0$ for all $x, y \in \mathcal{O}$, $x \neq y$ (the ortho part),*

To see why such sets may be a substitute for a basis, we notice first:

Lemma 6 *If $\mathcal{O}$ is an orthonormal subset of the inner product space V, then $\mathcal{O}$ is linearly independent.*

Proof. Let $x_1, \ldots, x_m \in \mathcal{O}$ and assume $c_1, \ldots, c_m \in K$ such that $c_1 x_1 + \cdots + c_m x_m = 0$. Then for $j = 1, \ldots, m$,

$$0 = (0, x_j) = (c_1 x_1 + \cdots + c_m x_m, x_j) = \sum_{i=1}^{m} c_i(x_i, x_j) = c_j.$$

∎

Being linearly independent, one might assume that one can complete such a system to span and get a basis. That is sort of possible. It is particularly simple in the finite dimensional case (it also works in the countable case). One can always transform a finite or countable linearly independent set of vectors into an orthonormal set spanning the same space. The procedure is known as the *Gram-Schmidt* procedure and it works as follows. Let $\{x_1, \ldots, x_n\}$ be a linearly independent set of vectors in an inner product space V. The procedure produces an orthonormal set $\{v_1, \ldots, v_n\}$ with the property that the span of $\{v_1, \ldots, v_k\}$ coincides with that of $\{x_1, \ldots, x_k\}$ for all $k = 1, \ldots, n$. If $\{x_1, \ldots, x_n\}$ was a basis of V, we have found an orthonormal basis of V. Orthonormal bases are nice; if $\{v_1, \ldots, v_n\}$ is an orthonormal set and $x = \sum_{i=1}^{n} c_i v_i$, then $c_i = (x, v_i)$.

Here is how Gram-Schmidt goes. It is an inductive procedure; it starts by setting $v_1 = x_1/\|x_1\|$. Assume now that for some $k \geq 1$, $\{v_1, \ldots, v_k\}$ is orthonormal and the span of $\{v_1, \ldots, v_k\}$ coincides with that of $\{x_1, \ldots, x_k\}$. If $k = n$ we are done, so assume $k < n$. Define first

$$w_{k+1} = x_{k+1} - \sum_{j=1}^{k} (x_{k+1}, v_j) v_j.$$

We see that $(w_{k+1}, v_j) = 0$ for $j = 1, \ldots, k$. Set $v_{k+1} = w_{k+1}/\|w_{k+1}\|$.

Thanks to Gram-Schmidt we get the following theorem

Theorem 7 *Let V be a finite dimensional inner product space. Then V has an orthonormal basis.*

It is customary to index orthonormal sets, write $\mathcal{O} = \{v_\alpha\}_{\alpha \in A}$, where A is an index set. In the finite case it may look like $\mathcal{O} = \{v_1, \ldots, v_n\}$; in the countable case like $\mathcal{O} = \{v_n\}_{n \in \mathbb{N}}$ or $\mathcal{O} = \{v_n\}_{n \in \mathbb{Z}}$. Alternatives for the last two denotations are $\mathcal{O} = \{v_n\}_{n=1}^{\infty}$, $\mathcal{O} = \{v_n\}_{n=-\infty}^{\infty}$.

We have

Theorem 8 *(Bessel's inequality) Let $\mathcal{O}$ be an orthonormal set (finite or not) in the inner product space V; let $v_1, \ldots, v_m \in \mathcal{O}$ and let $x \in V$. Then*

$$\sum_{i=1}^{m} |(x, v_i)|^2 \leq \|x\|^2.$$

Proof.

$$0 \leq \left(x - \sum_{i=1}^{m} (x, v_i) v_i, x - \sum_{i=1}^{m} (x, v_i) v_i \right) = (x, x) - \sum_{i=1}^{m} \overline{(x, v_i)}(x, v_i) - \sum_{i=1}^{m} (x, v_i)\overline{(x, v_i)} + \sum_{i,j=1}^{m} (x, v_i)\overline{(x, v_j)}(v_i, v_j)$$

$$= \|x\|^2 - \sum_{i=1}^{m} |(x, v_i)|^2 - \sum_{i=1}^{m} |(x, v_i)|^2 + \sum_{i=1}^{m} |(x, v_i)|^2 = \|x\|^2 - \sum_{i=1}^{m} |(x, v_i)|^2.$$

The result follows.   ∎

For those who are not afraid to look infinity into the eye, here are some fun facts about adding an infinite number of numbers. I am posing this as an exercise; it would be a good exercise in an Introductory Analysis course.

Exercise 14 *Let A be a set (an index set) and for every $\alpha \in A$ assume given a **non-negative** real number a_α. If F is a finite subset of A, it is clear what $\sum_{\alpha \in F} a_\alpha$ is. For example, if $F = \{\alpha_1, \ldots, \alpha_m\}$, then $\sum_{\alpha \in F} a_\alpha = \sum_{j=1}^{m} a_{\alpha_j}$. One defines*

$$\sum_{\alpha \in A} a_\alpha = \sup\{\sum_{\alpha \in F} a_\alpha \; : \; F \subset A, F \text{ finite }\}.$$

Incidentally, one also defines $\sum_{\alpha \in \emptyset} a_\alpha = 0$.

1. *Let $B \subset C \subset A$. Prove $\displaystyle\sum_{\alpha \in B} a_\alpha \leq \sum_{\alpha \in C} a_\alpha$*

2. *If $F \subset A$ is finite there are apparently two ways to define $\sum_{\alpha \in F} a_\alpha$; the usual way and as the sup of the usual sum over all finite subsets of F. Prove both ways coincide. This should be trivial, since $F \subset F$.*

3. *Let $B, C \subset A$, $B \cap C = \emptyset$. Prove $\sum_{\alpha \in B \cup C} a_\alpha = \sum_{\alpha \in B} a_\alpha + \sum_{\alpha \in C} a_\alpha$.*

4. *Assume that $C = \{\alpha \in A : a_\alpha > 0\}$ is countable and let $\{a_1, \alpha_2, \ldots\}$ is any enumeration of the elements of the set C. Then*

$$\sum_{\alpha \in A} a_\alpha = \sum_{\alpha \in C} a_\alpha = \lim_{n \to \infty} \sum_{k=1}^{n} a_{\alpha_k}.$$

   The limit on the right hand side of the second equation above is usually written in the form $\sum_{k=1}^{\infty} a_{\alpha_k}$.[

5. *$\sum_{\alpha \in A} a_\alpha < \infty$ if and only if $C = \{\alpha \in A : a_\alpha > 0\}$ is countable and $\sum_{\alpha \in C} a_\alpha < \infty$.*

If we feel comfortable with these infinite, perhaps even uncountable, sums we get the following immediate consequences of Bessel's inequality:

Theorem 9 *Let $\{v_\alpha\}_{\alpha \in A}$ be an orthonormal set in the inner product space V. Then, for every $x \in V$,*

$$\sum_{\alpha \in A} |(x, v_\alpha)|^2 \leq \|x\|^2 \tag{7}$$

Consequently, for every $x \in V$, the set $\{\alpha \in A : (x, v_\alpha) \neq 0\}$ is countable, possibly finite.

Definition 8 *An orthonormal set $\{v_\alpha\}_{\alpha \in A}$ is said to be **complete** or **maximal** iff it is not the proper subset of an orthonormal set. In other words, if $\{v_\alpha\}_{\alpha \in A} \subset \mathcal{O} \subset V$, and $\mathcal{O} \neq \{v_\alpha\}_{\alpha \in A}$, then $\mathcal{O}$ is not orthonormal.*

The following theorem is proved in Halmos in the finite dimensional case. I give the proof in general, but we'll only use it in the finite dimensional case. So choose the proof you prefer.

Theorem 10 *Let V be a Hilbert space[2] and let $\{v_\alpha\}_{\alpha \in A}$ be an orthonormal set in V. The following assertions are equivalent.*

1. *$\{v_\alpha\}_{\alpha \in A}$ is complete.*

2. *If $x \in V$ and $(x, v_\alpha) = 0$ for all $\alpha \in A$, then $x = 0$.*

3. *The subspace spanned by $\{v_\alpha\}_{\alpha \in A}$ is dense in V.*

4. *For every $x \in V$, if $\{\alpha_1, \alpha_2, \ldots\}$ is an enumeration of $\{\alpha \in A \, (v_a, x) \neq 0\}$, then*

$$\lim_{n \to \infty} \left\| x - \sum_{k=1}^{n} (x, v_{\alpha_k}) v_{\alpha_k} \right\| = 0.$$

---

[2]That is, an inner product space that is complete. All finite inner product spaces are Hilbert spaces.

5. *For every $x, y \in V$, one has*

$$\sum_{\alpha \in a} (x, v_a)(y, v_\alpha) = (x, y)$$

in the sense that if $\{\alpha_1, \alpha_2, \ldots\}$ is an enumeration of $\{\alpha \in A \, (v_a, x) \neq 0\} \cup \{\alpha \in A \, (v_a, y) \neq 0\}$, then

$$\sum_{k=1}^{\infty} (x, v_{a_1})(y, v_{\alpha_i}) = (x, y)$$

6. *For every $x \in V$, Bessel's inequality becomes an equality called Parseval's equality,*

$$\sum_{\alpha \in A} |(x, v_a)|^2 = \|x\|^2.$$

For the general case we need to turn Halmos a bit upside down, sort of. First we prove

Lemma 11 *Assume V is a Hilbert space and C is a closed convex[3] non-empty subset of V. Then C contains a unique element of minimum norm.*

Proof. Let $\sigma = \inf\{\|x\| : x \in C\}$. We have to prove there exists a unique $x \in C$ with $\|x\| = \sigma$. There is $\{x_n\}$ in C such that $\|x_1\| \geq \|x_2\| \geq \cdots$ and $\lim_{n \to \infty} \|x_n\| = \sigma$. Unfortunately in infinite dimensions Heine Borel is not on the job, bounded does not imply sequential compact, so it isn't obvious at first glance that $\{x_n\}$ converges. here is the trick; it is where convexity comes in for the first time. By convexity, $(x_n + x_m)/2 \in C$ for all n, m, thus

$$\sigma^2 \leq \|\frac{1}{2}(x_n + x_m)\|^2 = \frac{1}{4}\|x_n\|^2 + \frac{1}{4}\|x_m\|^2 + \frac{1}{2}\text{Re}\,(x_n, x_m),$$

so

$$2\text{Re}\,(x_n, x_m) \geq 4\sigma^2 - \|x_n\|^2 - \|x_m\|^2.$$

Thus

$$\|x_n - x_m\|^2 = \|x_n\|^2 + \|x_m\|^2 - 2\text{Re}\,(x_n, x_m) \leq 2\|x_n\|^2 + 2\|x_m\|^2 - 4\sigma^2 \to 0$$

as $n, m \to \infty$. Or, if you want to be more precise, given $\epsilon > 0$ there is N such that

$$\sigma^2 \leq \|x_n\|^2 < \sigma^2 + \frac{\epsilon^2}{4}$$

if $n \geq N$. If $m, n \geq N$, then

$$\|x_n - x_m\|^2 \ll \leq 2\left(\sigma^2 + \frac{e^2}{4}\right) + 2\left(\sigma^2 + \frac{e^2}{4}\right) - 4\sigma^2 = \epsilon^2,$$

proving $\{x_n\}$ is a Cauchy sequence. Isn't this a clever argument? Because we are assuming completeness, there is $x \in V$ such that $x = \lim_{n \to \infty} x_n$. By Exercise 7, we see that $\|x\| = \sigma$. Since C is closed, $x \in C$.

To see that x is unique, assume now that $y \in C$, $\|y\| = \sigma$, $y \neq x$. We use a fact that is true in inner product spaces, but not in all normed spaces; namely, that if two points are equidistant from the origin, all the points between them (i.e., on the line segment joining the points) are strictly closer to the origin. Let $z = \frac{1}{2}(x + y)$ Then $z \in C$ and we have

$$\sigma^2 \leq \|z\|^2 = \frac{1}{4}(\|x\|^2 + \|y\|^2 + 2\text{Re}\,(x, y)) = \frac{1}{2}(\sigma^2 + \text{Re}\,(x, y))$$

so that $\text{Re}\,(x, y) = \sigma^2$. Then

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\text{Re}\,(x, y) = \sigma^2 + \sigma^2 - 2\sigma^2 = 0.$$

■

We did not really prove that if $z = (x + y)/2$ and $\|x\| = \|y\|$, that then $\|z\| < \|x\|$ since we used the fact that $z \in C$ and there fore $\|z\| \geq \sigma = \|x\|$. One doesn't really need that assumption; here is an alternative proof that

---

[3]A subset C of a real or complex vector space V is said to be *convex* iff whenever $x, y \in C$ and $s \in \mathbb{R}$, $0 \leq s \leq 1$, then $(1-s)x+sy \in C$.

$\|x\| = \|y\| = \sigma$ implies $z = (x + y)/2$ satisfies $\|z\| < \sigma$. which would be a contradiction if σ is the inf of the set of norms of elements of a set C and $z \in C$. First of all, $\|z\| \leq (1/2)(\|x\| + \|y\|) = \sigma$, so $\|z\| \leq \sigma$. So far, only the triangle property of the norm was used, which holds in all normed spaces. Assume next $\|z\| = \sigma$. Then

$$\begin{aligned} (x, z) &= (x, \tfrac{1}{2}(x + y)) = \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}(x, y) = \tfrac{1}{2}\sigma^2 + \tfrac{1}{2}(x, y), \\ (y, z) &= (y, \tfrac{1}{2}(x + y)) = \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}(y, x) = \tfrac{1}{2}\sigma^2 + \tfrac{1}{2}\overline{(x, y)}. \end{aligned}$$

Adding,

$$2\sigma^2 = 2\|z\|^2 = (x + y, z) = \sigma^2 + \text{Re }(x, y)$$

so that Re $(x, y) = \sigma^2$, Then, as before, $\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\text{Re }(x, y) = \sigma^2 + \sigma^2 - 2\sigma^2 = 0$, so that $x = y$. It follows that $\|z\| = \sigma$ implies $x = y$, hence $x = z = y$. If $x \neq y$ we must have $\|z\| < \sigma$.

Corollary 12 *Let C be a closed, non empty, convex subset of the Hilbert space V. For every $x \in V$ there exists a unique $P_C x \in C$ such that*

$$\|x - P_C x\| = dist(x, C) = \inf\{\|x - y\| : y \in C\}.$$

Proof. The existence and uniqueness of $P_C x \in C$ is immediate from Lemma 11; the set $C_1 = x - C = \{x - y : y \in C\}$ is convex, closed, hence has a unique element of minimum norm z; set $P_c x = x - z$. We need to see that $(x, x - P_C x) = 0$. ∎

Suppose now M is a closed subspace of the Hilbert space V. If $\dim V < \infty$, then all subspaces are closed because, in general, finite dimensional subspaces are closed. But in the infinite dimensional case, this is not so. Then M is a closed, convex, non-empty subset of V and we have the map $P_M : V \to M$ defined by $P_M x \in M$ is the unique point satisfying $\|x - P_M x\| = dist(x, M)$. We can also see that $P_M x$ is also the only point of M such that $x - P_M x \perp M$. That it satisfies this is proved using a calculus 1 approach. Let $y \in M$ and consider the function $f : \mathbb{R} \to \mathbb{C}$ defined by

$$f(t) = \|x - P_M x + ty\|^2 = \|x - P_M x\|^2 + 2t\text{Re }[(x - P_M x, y)] + t^2\|y\|^2.$$

By the definition of $P_M x$, this function achieves a minimum at $t = 0$, thus $f'(0) = 0$, which works out to Re $(x - P_M x, y) = 0$. If $K = \mathbb{R}$, this is the same as $(x - P_M x, y) = 0$. If $K = \mathbb{C}$ we repeat the argument with y replaced by iy to get Im $(x - P_M x, y) = \text{Re }(x - P_M x, iy) = 0$. This proves, since $y \in M$ was arbitrary, that $x - P_M x \in M^\perp$. Assume now that there is $z \in M$ such that $x - z \in M^\perp$. For every $y \in M$ we then have $x - z \perp z - y$ so that

$$\|x - y\|^2 = \|(x - z) + (z - y)\|^2 = \|x - z\|^2 + \|z - y\|^2 \geq \|x - z\|^2.$$

Because y is arbitrary, this shows that $\|x - z\| \leq dist(x, M)$. thus $\|x - z\| = dist(x, M)$; by uniqueness of $P_M x$ we get $z = P_M x$.

This is a very geometric idea; the minimum distance of a point to a subspace is realized by the one and only z in the subspace such that the vector from x to z is perpendicular to the subspace.

But there is more. Let $x, y \in V$, let $\alpha, \beta \in K$ and let $z_1 = P_M x$, $z_2 = P_M y$. Then $\alpha z_1 + \beta z_2 \in M$ and

$$(\alpha x + \beta y) - (\alpha P_M x + \beta P_M y) = \alpha(x - P_M x) + \beta(y - P_M y) \in \alpha M^\perp + \beta M^\perp \subset M^\perp,$$

proving that $P_M(\alpha x + \beta y) = \alpha P_M x + \beta P_M y$. It may be a good idea to put some, or all of this, in the form of a theorem.

Theorem 13 *Let M be a closed subspace of the Hilbert space M. There exists a linear map $P_M : V \to M$ uniquely determined by either one of the two properties:*

1. $\|x - P_M x\| = dist(x, M)$ for all $x \in V$, or

2. $x - P_M x \perp M$ for all $x \in V$.

The map P_M when restricted to M is the identity.

The only thing not explicitly proved is that $P_M|_M = I$, a fact which is quite obvious since, if $x \in M$, the element realizing the distance to M is x itself.

We can think of P_M as mapping V into V, with range M, and consider $I - P_M : V \to V$. Being linear, the range $N = R(I - P_M)$ of $I - P_M$ is also a subspace of V. It is also fairly immediate to see that if $x \in M$ and $y \in N$, then $(x, y) = 0$. In fact, $y = z - P_M z$ for some $z \in V$, thus $y \perp M$. So $M \subset N^{\perp}$, $N \subset M^{\perp}$. Now let $x \in M^{\perp}$. We can write $x = P_M x + (I - P_M)x$; taking inner product with $P_M x$, since $(x, P_M x) = ((I - P_M)x, P_M x) = 0$ because $P_M x \in M$ and $(I - P_M)x \in M^{\perp}$, we see that $\|P_M x\|^2 = 0$. That is: If $x \in M^{\perp}$, then $P_M x = 0$. But then $x = (I - P_M)x \in N$. This proves $M^{\perp} \subset N$. Similarly $N^{\perp} \subset M$. We proved

$$M^{\perp} = (I - P_M)(V), \qquad M^{\perp\perp} = (I - P_M)(V)^{\perp} = M.$$

Obviously $V = M \oplus (I - P_M)(V) = M \oplus M^{\perp}$, since $M \cap M^{\perp} = \{0\}$ and $I = P_M + (I - P_M)$. We proved.

Theorem 14 *Let M be a closed subspace of the Hilbert space V. Then*

$$V = M \oplus M^{\perp}, \qquad M^{\perp\perp} = M.$$

Moreover $M^{\perp}$ is the range of $I - P_M$, P_M as before.

It is time for an example. The space ℓ^2 consists of all sequences $a = \{a_n\}$ of complex numbers such that

$$\sum_{n=1}^{\infty} |a_n|^2 < \infty.$$

For example, the sequence $\{1, 1/2, 1/3, \ldots\} = \{1/n\}$ is in ℓ^2. And many more. ℓ^2 is a complex vector space with the obvious operations. **Claim:** If $a = \{a_n\}$, $b = \{b_n\} \in \ell^2$, then the series

$$\sum_{n=1}^{\infty} a_n \bar{b}_n$$

is absolutely convergent. In fact, for every n we have by Cauchy Schwarz

$$\sum_{k=1}^{n} |a_k \bar{b}_k| \leq \left(\sum_{k=1}^{n} |a_k|^2 \right)^{1/2} \left(\sum_{k=1}^{n} |b_k|^2 \right)^{1/2} \left(\sum_{k=1}^{\infty} |a_k|^2 \right)^{1/2} \left(\sum_{k=1}^{\infty} |b_k|^2 \right)^{1/2}$$

showing that the finite sums of the absolute values are uniformly bounded, and proving the claim. It makes thus sense to define, if $a = \{a_n\}$, $b = \{b_n\} \in \ell^2$,

$$(a, b) = \sum_{n=1}^{\infty} a_n \bar{b}_n.$$

One verifies easily that this makes ℓ^2 into an inner product space. It is a Hilbert space; proving completeness requires no more than Introductory Analysis 1 skills. Now consider the set M of all eventually zero sequences of complex numbers; $a = \{a_n\} \in M$ if and only if there exists N depending on a) such that $a_n = 0$ if $n > N$. It is clear that $M \subset \ell^2$; it is clear that it is a subspace of ℓ^2. It is also clear that M is not closed since $M \neq \ell^2$ but $\bar{M} = \ell^2$.

We are almost ready to prove Theorem 10. We still need:

Lemma 15 *Let M be a subspace of V. Then $M^{\perp\perp} = \bar{M}$ (the closure of M. In particular, a subspace M is dense in V if and only if $M^{\perp} = \{0\}$.*

Proof. I'll be a bit sketchy, but all details are quite easy. One shows that if M is a subspace, then $\bar{M}$ is also a subspace and $M^{\perp} = \bar{M}^{\perp}$. Thus, of course,

$$M^{\perp\perp} = (\bar{M})^{\perp\perp} = \bar{M}.$$

If $\bar{M} = V$, then $M^\perp = \bar{M}^\perp = V^\perp = \{0\}$. Conversely, if $M^\perp = \{0\}$, then $\bar{M}^\perp = \{0\}$, hence $\bar{M} = \bar{M}^{\perp\perp} = \{0\}^\perp = V$.
■

And now *Proof of Theorem 10.*

1 $\Rightarrow$ **2**. Assume $\{v_\alpha\}_{\alpha \in A}$ is complete. If $0 \neq x \in V$ and $(x, v_\alpha) = 0$ for all $\alpha \in A$, let $\mathcal{O} = \{v_\alpha\}_{\alpha \in A} \cup \{x/\|x\|\}$. This is an orthonormal system of which $\{v_\alpha\}_{\alpha \in A}$ is a proper subset, contradicting the completeness of $\{v_\alpha\}_{\alpha \in A}$.

2 $\Rightarrow$ **3**. Let M be the subspace spanned by the $\{v_\alpha\}_{\alpha \in A}$; in other words the set of all finite linear combinations of the v_α's. If $x \perp M$, then $(x, v + \alpha) = 0$ for all $\alpha \in A$, thus $x = 0$. That is, $M^\perp = \{0\}$, thus $\bar{M} = V$.

3 $\Rightarrow$ **4**. Let M be the subspace spanned by the $\{v_\alpha\}_{\alpha \in A}$ so that by hypothesis $\bar{M} = V$, thus $M^\perp = \{0\}$ by Lemma 15. Let $x \in V$. By Theorem 9, we know that the set $\{\alpha \in A : (x, v_a) \neq 0\}$ is countable; let $\{a_1, \alpha_2, \ldots\}$ be an enumeration of this set (we may assume the set is infinite). Set $x_n = \sum_{k=1}^n (x, v_{\alpha_k}) v_{\alpha_k}$ for $n \in \mathbb{N}$. Claim: $\{x_n\}$ is a Cauchy sequence in V. In fact, if $m < n$ (as one may assume)

$$\|x_n - x_m\|^2 = \left\| \sum_{k=m+1}^n (x, v_{\alpha_k}) v_{\alpha_k} \right\|^2 = \sum_{k=n+1}^m |(x, v_{\alpha_k})|^2$$

and the claim follows from Bessel's inequality. By completeness, let $y = \lim_{n \to \infty} x_n$. So far the hypothesis was not used at all. Now we see that $(x - y, v_\alpha) = 0$ for all $\alpha \in A$; this is clear if $\alpha \neq \alpha_n$, $n \in \mathbb{N}$ (sort of clear; not to hard to proof) and because $(x_n, v_{\alpha_k}) = (x, v_{\alpha_k})$ if $k \leq n$, we also get it to hold for $\alpha = \alpha_n$. Rather than invoking point 1, we say that the $x - y \in M^\perp = \{0\}$.

4 $\Rightarrow$ **5**. By Theorem 9, we know that the set $\{\alpha \in A : (x, v_a) \neq 0 \text{ or } (y, v_\alpha) \neq 0\}$ is countable; let $\{a_1, \alpha_2, \ldots\}$ be an enumeration of this set (we may assume the set is infinite). We now have setting $x_n = \sum_{k=1}^n (x, v_{\alpha_k}) v_{\alpha_k}$, $y_n = \sum_{k=1}^n (y, v_{\alpha_k}) v_{\alpha_k}$ that

$$(x_n, y_n) = \sum_{k=1}^n (x, v_{\alpha_k})(y, v_{\alpha_k})$$

for all n. Now

$$|(x, y) - (x_n, y_n)| = |(x, y - y_n) + (x - x_n, y) + (x - x_n, y_n - y)| \leq \|x\|\|x - x_n\| + \|x - x_n\|\|y\| + \|x - x_n\|\|y - y_n\| \to 0$$

as $n \to \infty$ by the hypothesis. The result follows.

5 $\Rightarrow$ **6**. This is immediate; take $x = y$ in 5.

6 $\Rightarrow$ **1**. If $\{v_\alpha\}$ is not complete, there has to exist a unit vector $x \in V$ orthogonal to all the v_α's. Then Parseval's equality becomes the nonsense $0 = 1$. ■

Some final words before returning to the narrow confines of finite dimensional spaces. In a Hilbert space, a complete orthonormal set plays the role of a basis. There always exist such sets (if you believe the axiom of choice, which I do). Nice Hilbert spaces are those in which a countable complete orthonormal set exists, and a vast array of spaces appearing in applications have a countable orthonormal basis. Many properties that hold in Hilbert spaces fail to hold in general Banach spaces. For example, in a Hilbert space V, given a closed subspace M, there always exists a closed subspace N (for example $N = M^\perp$) such that $V = M \oplus N$. This is not true in general Banach spaces. This ends our excursion into the infinite dimensional world, it is time to return to Halmos.

Appendix: Heine Borel

I start with some generalities. Assume (X, d) is a metric space. As we all know, a subset C of X is *compact* iff every open cover admits a finite subcover. The set C is said to be *sequentially compact* iff every sequence has a convergent subsequence (with limit in the set). There are more general topological spaces than metric spaces, in which the two concepts are not equivalent, but in metric spaces one does have that a subset C is compact if and only if it is sequentially compact. This is important because the definition with open covers is a good one to use if one knows that a set is compact, while the sequential compactness definition can be better to use when verifying that a set is compact.

I will prove this equivalence, which is not at all trivial. For the easy part of it I will use another rather immediate equivalent version of compactness, usually known as FIP (finite intersection property). Here it is:

Theorem 16 *Let C be a subset of the metric space X. The following two statements are equivalent:*

1. *C is compact.*

2. *C satisfies: Whenever $\{F_\alpha\}_{\alpha \in A}$ is a family of closed subsets of C such that the intersection of any finite subfamily is not empty (that is, for every finite subset B of A, $\bigcap_{\alpha \in B} F_\alpha \neq \emptyset$) then the intersection of the whole family is not empty ($\bigcap_{\alpha \in A} F_\alpha \neq \emptyset$). A set having this property is said to satisfy FIP.*

Proof. C is **not** compact if and only if there exists a family $\{U_\alpha\}_{\alpha \in A}$ of open subsets of C such that $C = \bigcup_{\alpha \in A} U_\alpha$ but $C \neq \bigcup_{\alpha \in B} U_\alpha$ for all finite subsets B of A. Going to complements, setting $F_\alpha = C \backslash U_\alpha$ (and vice versa) this is equivalent to the existence of a family $\{F_\alpha\}_\alpha \in A$ of closed subsets of C such that $\bigcap_{\alpha \in A} F_\alpha = \emptyset$ but $\bigcap_{\alpha \in B} F_\alpha \neq \emptyset$ for all finite subsets B of A. ∎

The most common use of this result, especially in the case of metric spaces, is through this corollary.

Corollary 17 *Let C be a compact subset of the metric space X and for each $n \in \mathbb{N}$ let F_n be a **non-empty** closed subset of C such that $F_1 \supset F_2 \supset \cdots$. Then $\bigcap_{n=1}^{\infty} F_n \neq \emptyset$.*

Proof. Since C is compact, it satisfies FIP. If $B \subset \mathbb{N}$ is finite, let $N = \max_{n \in B} n$. Then $\bigcap_{n \in B} F_n = F_N \neq \emptyset$. By FIP, $\bigcap_{n=1}^{\infty} F_n \neq \emptyset$. ∎

Now it is easy to prove the promised equivalence. Well, only one direction is easy. The other direction is quite tough.

Theorem 18 *Let C be a subset of the metric space X. Then C is compact if and only if it is sequentially compact.*

Proof. Assume first that C is compact. Let $\{x_n\}$ be a sequence of points of C. For each $n \in \mathbb{N}$ set $F_n = \{x_k : \kappa \geq n\}$. This is a decreasing sequence of non-empty sets, thus $\{\bar{F}_n\}$ (where a bar atop a set indicates its closure in X) is a decreasing sequence of non-empty closed sets. Since C is compact, it is closed, hence $\bar{F}_n \subset C$ for all $n \in \mathbb{N}$. By Corollary 17, $\bigcap_{n=1}^{\infty} \bar{F}_n \neq \emptyset$; let $z \in \bigcap_{n=1}^{\infty} \bar{F}_n$. We now construct a strictly increasing sequence $\{n_k\}$ of positive integers inductively as follows. Since z is in the closure of F_1 there are elements of F_1 arbitrarily close to z; we'll be modest and content ourselves with an element at distance less than 1 from z. That is, there is $n_1 \geq 1$ such that $d(x_{n_1}, z) < 1$. Assume found for some $k \geq 1$ integers $n_1, \ldots, n_k$ such that $1 \leq n_1 < \cdots < n_k$ and such that $d(x_{n_j}, z) < 1/j$ for $j = 1, \ldots, k$. This, of course, was done for $k = 1$. Because $z \in \bar{F}_{n_k+1}$, there is $n \geq n_k + 1$ such that $d(x_n, z) < 1/(k+1)$; set $n_{k+1} = n$ so $n_{k+1} \geq n_k + 1 > n_k$ and $d(x_{n_{k+1}}, z) < 1/(k+1)$. This completes the construction of the sequence $\{n_k\}$. Since it is strictly increasing, $\{x_{n_k}\}$ is a subsequence of $\{x_n\}$; since $d(x_{n_k}, z) < 1/k$ for all $k \in \mathbb{N}$, it is clear the subsequence converges to z. Finally, because C is closed, $z \in C$.

Conversely, assume C is sequentially compact. The proof that C is compact is a bit devious. First we prove: C is separable; that is, there exists a countable subset D of C such that $C = \bar{D}$ (the closure of D). We construct it as follows. For each $\delta > 0$, the following process must terminate in a finite number of steps: Pick a point $x_1 \in C$. Assuming you picked points $x_1, \ldots, x_k$ in C, pick a point $x_{k+1} \in C$ such that $d(x_{k+1}, x_j) \geq \delta$ for $j = 1, \ldots, k$. It has to terminate because otherwise we would get a sequence $\{x_k\}$ such that $d(x_k, x_j) \geq \delta$ for all $j \neq k$; such a sequence has NO subsequence that can be a Cauchy sequence, hence no convergent subsequences. Thus for every $\delta > 0$ there is a finite set $D_\delta \subset \mathbb{C}$ such that every point in C is at distance $< \delta$ from a point of D_δ. Let $D = \bigcup_{n=1}^{\infty} D_{1/n}$. Then D, as a countable union of finite sets, is countable. If $x \in C$ and $\epsilon > 0$, select n so $1/n < \epsilon$. There is $y \in D_{1/n}$ such that $d(x, y) < 1/n > \epsilon$. It follows that $y \in \bar{D}$.

For the next trick we let D be a countable dense subset; that is a countable subset whose closure is C and we consider the family

$$\mathcal{U} = \{B(y, r) : y \in D, r \in \mathbb{Q}, r > 0\}.$$

Because both D and $\mathbb{Q}$ are countable, this family is countable. Incidentally, $B(x, r)$ if $x \in X$, $r > 0$, is defined by $B(x, r) = \{y \in X : d(y, x) < r\}$.

With this family we can prove: Every open cover of C contains a **countable** subcover. We are still possibly in the infinite realm, but by less. So assume that $\{U_\alpha : \alpha \in A\}$ is an open cover of C. Let

$$\mathcal{U}' = \{B(y, r) : y \in D, r \in \mathbb{Q}, \exists \alpha \in A, B(y, r) \subset U_\alpha\}$$

The family $\mathcal{U}'$, being a subset of a countable set, is countable. In addition, it covers C; if $x \in C$, there is $\alpha \in A$ so $x \in U_\alpha$. Since U_α is open, there is $\rho > 0$ such that $B(x, \rho) \subset U_\alpha$. By the density of $\mathbb{Q}$ in $\mathbb{R}$, there is $r \in \mathbb{Q}$,

$0 < r < \rho/2$. Since D is dense, there is $y \in D$, $d(y,x) < r$. Then $x \in B(y,r) \subset B(x,\rho) \subset U_\alpha$. Then $x \in B(y,r)$ and $B(y,r) \in \mathcal{U}'$. Now that we proved that $\mathcal{U}'$ is a covering, for each set $B(y,r) \in \mathcal{U}'$ we select $\alpha_{y,r} \in A$ such that $B(y,r) \subset U_{\alpha_{y,r}}$. The family $\{U_{\alpha_{y,r}} : B(y,r) \in \mathcal{U}'\}$ is a countable subfamily of $\{U_\alpha : \alpha in A\}$ and covers C.

Well, we finally arrived at the last step of this lengthy argument. It suffices to prove because of what we just did: Every countable open cover of C contains a finite subcover. So assume $\{U_n : n \in \mathbb{N}\}$ is a countable open cover of C. We go by contradiction. Assuming there is no finite subcover, then for every $n \in \mathbb{N}$, there is $x_n \in C$ such that $x_n \notin \bigcup_{j=1}^n U_j$. By hypothesis, the sequence $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$; let $x = \lim_{k\to\infty} x_{n_k}$. Because the U_n's cover, there is $n \in \mathbb{N}$ such that $x \in U_n$. Since U_n is open, there is K such that $x_{n_k} \in U_n$ if $k \geq K$. However, if $k \geq n$, then $n_k \geq n$ and $x_{n_k} \notin U_n$. We get a contradiction since we can take $k \geq \max(K,n)$.

We are done! ∎

Now it is relatively easy to prove the Heine Borel theorem in the way I want to use it to prove Theorem 2.

Theorem 19 *Let V be a finite dimensional **real** vector space, let $n = \dim V$, and let $\{e_1, \ldots, e_n\}$ be a basis of V. Consider V as a metric space with the metric that comes from the norm defined by*

$$\|x\| = \sum_{i=1}^n |\xi_i|, \qquad if \quad x = \sum_{i=1}^n \xi_i e_i.$$

If $C \subset V$ and C is closed and bounded, then C is compact.

Proof. Here is the idea of the proof. I will first prove it if the dimension is 1. For dimension 1, I will use the standard definition of compactness. Then I will use the equivalence between compactness and sequential compactness to reduce the case of n dimensions, $n > 1$, to the case $n = 1$. To achieve this reduction I need to use what I think is a fairly obvious result, namely:

Lemma 20 *Let $x_k = \sum_{j=1}^n \xi_{kj} e_j$. The sequence $\{x_k\}$ converges if and only the n-sequences of real numbers $\{x i_{kj}\}_{k=1}^\infty$ $(j = 1, \ldots, n)$ converge. In this case, if $\lim_{k\to\infty} \xi_{kj} = \xi_j$ for $j = 1, \ldots, n$, then $\lim_{k\to\infty} x_k = \sum_{j=1}^n \xi_j e_j$.*

I will not insult my probably non-existent readership by adding a proof of this lemma. Let's get on with the proof. Assume $C \subset V$, and C is closed and bounded.

Case $n = 1$. This clearly becomes equivalent to proving: If C is a closed and bounded subset of $\mathbb{R}$, then C is compact. It is clearly so because $V = \{\xi e_1 : \xi \in \mathbb{R}\}$ and the only role e_1 plays here is the role of a nuisance. If we ignore it, we are in $\mathbb{R}$, with open, closed, etc., defined just as is usual in $\mathbb{R}$. So assume now C is a closed and bounded subset of $\mathbb{R}$. Because C is bounded, there is a closed interval $[a,b]$, $-\infty < a < b < \infty$, such that $C \subset [a,b]$. Using the rather simple result that a closed subset of a compact set is compact, if we can prove $[a,b]$ compact, it will follow that C is compact. We prove $[a,b]$ compact.

Assume thus that $\mathcal{U}$ is an open covering of $[a,b]$; a large, most likely infinite, possibly uncountable, family of sets covering $[a,b]$. We consider the following interesting subset of $[a,b]$: Let

$$S = \{x \in [a,b] : \exists \text{ a finite number of sets } U_1, \ldots, U_m \in \mathcal{U} \text{ s.t. } [a,x] \subset \bigcup_{j=1}^m U_j\}.$$

The goal, of course, is to prove $b \in S$, Starting modestly, we notice that $S \neq \emptyset$. In fact, since $\mathcal{U}$ covers $[a,b]$, there is $U \in \mathcal{U}$, $a \in U$; then $\{a\} = [a,a] \subset U$, so $a \in S$. In addition, as a subset of $[a,b]$, S is bounded above (by b, for example), so that we can talk of the least upper bound or supremum of S. Let $\sigma = \sup S$. First we notice that $\sigma \in S$. In fact, this holds clearly if $\sigma = a$, so assume $\sigma > a$ (actually, this isn't really needed). Because b is an upper bound of S, we have $\sigma \leq b$, thus $\sigma \in [a,b]$, hence there is $U \in \mathcal{U}$ such that $\sigma \in U$. Because U is open, there is $\delta > 0$ such that $(\sigma - \delta, \sigma + \delta) \subset U$. Now $\sigma - \delta < \sigma$, so $\sigma - \delta$ is not an upper bound of S, so that there is $x \in S$, $\sigma - \delta < x \leq \sigma$. because $x \in S$, there is a finite number of sets of $\mathcal{U}$, say $U_1, \ldots, U_m$ such that $[a,x] \subset U_1 \cup \cdots \cup U_m$. Add to this the set U such that $\sigma \in U$, and it follows that $[a,\sigma] \subset U_1 \cup \cdots \cup U_m \cup U$, proving $\sigma \in S$. But the same construction also proves $\sigma = b$. If $\sigma < b$, with U, δ as above, there will be $y \in (\sigma, \sigma + \delta)$ such that $y < b$ and then $[a,y] \subset U_1 \cup \cdots \cup U_m \cup U$. This implies $y \in S$ and since $y > \sigma$ it contradicts the definition of σ. Thus $\sigma \in S$, proving there is a finite subcover of $[a,b]$ that can be extracted from $\mathcal{U}$. This concludes the proof in the case $n = 1$.

For the general case, let C be a closed and bounded subset of V where again $\dim V = n$. We prove C is sequentially compact. Let $\{x_k\}$ be a sequence in k and write $x_k = \sum_{j=1}^n \xi_{kj} e_j$ with $\xi_{kj} \in \mathbb{R}$, $j = 1, \ldots, n$, $k = 1, 2, 3, \ldots$. To avoid subscripts of subscripts of subscripts, we will use a somewhat non canonical notation. The sequence $\{\xi_{k1}\}$ is a bounded sequence of real numbers, thus contained in some interval $[a, b]$ of $\mathbb{R}$; since $[a, b]$ is compact it has a convergent subsequence. Instead of writing $\{\xi_{k_\ell,1}\}$ for this subsequence, I'll write it as $\{\xi_{\phi_1(k),1}\}$, where $\phi_1 : \mathbb{N} \to \mathbb{N}$ is strictly increasing, so $1 \le \phi_1(1) < \phi_1(2) < \cdots$. Say $\lim_{k\to\infty} \xi_{\phi_1(k),1} = \xi_1$. Now consider $\{\xi_{\phi_1(k),2}\}$. This is again a bounded subsequence of $\mathbb{R}$, hence has a convergent subsequence; there is thus $\psi : \mathbb{N} \to \mathbb{N}$ strictly increasing such that $\{\xi_{\phi_1(\psi(k)),2}\}$ converges, say to ξ_2. Set $\phi_2 = \phi_1 \circ \psi$. Continuing this way we get n sequences $\{\xi_{\phi_j(k),j}\}_{k=1}^\infty$, such that for $j > 1$, $\{\xi_{\phi_j(k),j-1}\}_{k=1}^\infty$ is a subsequence of $\{\xi_{\phi_{j-1}(k),j=1}\}_{k=1}^\infty$; moreover $\lim_{k\to\infty} \xi_{\phi_j(k),j} = \xi_j$ for $j = 1, \ldots, n$. The last choice of indices produces a sequence $\{\xi_{\phi_n(k),n}\}_{k=1}^\infty$ with the property that $\{\xi_{\phi_n(k),j}\}_{k=1}^\infty$ is a subsequence of $\{\xi_{\phi_j(k),j}\}_{k=1}^\infty$ for $j = 1, \ldots, n$, hence $\lim_{k\to\infty} \xi_{\phi_n(k),j} = \xi_j$ for $j = 1, \ldots, n$. It follows that

$$\lim_{k\to\infty} x_{\phi_n(k)} = \sum_{j=1}^n \xi_{\phi_n(k),j} e_j = \sum_{j=1}^n \xi_j e_j.$$

This proves C is sequentially compact, hence compact. ∎

This was Heine Borel for a real vector space. What if the space is complex? Every complex vector space V can be considered as a real vector space by just restricting scalar multiplication to real scalars. If $\{e_1, \ldots, e_n\}$ is a basis of V as a complex vector space, it is immediate to verify that $\{e_1, \ldots, e_n\} \cup \{i\, e_1, \ldots, i\, e_n\}$ (where $i = \sqrt{-1}$) is a basis for V as a real space so that $\dim_{\mathbb{R}} V = 2 \dim_{\mathbb{C}} V$. The $\|\cdot\|_1$ norm of complex V with respect to the basis $\{e_1, \ldots, e_n\}$ is easily seen to be equivalent (without need to appeal to Theorem 2; that would be circular reasoning) to the $\|\cdot\|_1$ norm of real V with respect to the basis $\{e_1, \ldots, e_n\} \cup \{i\, e_1, \ldots, i\, e_n\}$.