MATH2040A Week 9 Tutorial Notes

1 Inner Product

1.1 Inner Product Space

A vector space has no structure to talk about anything other than linear relations. However, we also want to talk about the "size" of a vector and the "angle" between two vectors, in a way that is similar to the (usual) Euclidean space. To do so, we simple add more structures to a (plain) vector space, which leads to the following definition of *(real / complex) inner product space*:

Definition 1.1. An inner product space is a vector space $V = (V, F, +, \cdot)$ with real ($F = \mathbb{R}$) or complex ($F = \mathbb{C}$) scalar field[1], additionally equipped with a binary operation $\langle \cdot, \cdot \rangle : V \times V \to F$ that satisfies the following requirements:

- for each $y \in V$, $x \mapsto \langle x, y \rangle$ is a linear map
- for every $x \in V$, $\langle x, x \rangle \geq 0$ and equality holds if and only if $x = 0$ (in particular, $\langle x, x \rangle$ is real)
- for every $x, y \in V$, $\langle x, y \rangle = \overline{\langle y, x \rangle}$, where $\overline{z}$ is the (complex) conjugate on F

The binary map $\langle \cdot, \cdot \rangle$ is an *inner product* of the space V.

Typical examples of inner products are

- the (usual) dot product $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$ is an inner product of $\mathbb{R}^n$
- for each $a < b$, $\langle f, g \rangle = \int_a^b f(t)g(t)\,\mathrm{d}t$ is an inner product of $C^0([a, b], \mathbb{R})$, the space of real-valued continuous functions on $[a, b]$

and their complex analogs.

1.2 Normed Space

Given an inner product $\langle \cdot, \cdot \rangle$, we can induce a notion of "size" on vectors $\| x \| = \sqrt{\langle x, x \rangle} \geq 0$, which you can show that the following properties hold:

- for every $x \in V$, $\| x \| \geq 0$, and $\| x \| = 0$ if and only if $x = 0$
- for every $x \in V$ and $c \in F$, $\| cx \| = | c | \| x \|$ where $| c |$ is the modulus / absolute value of scalar c
- (triangle inequality) for every $x, y \in V$, $\| x + y \| \leq \| x \| + \| y \|$

Definition 1.2. A normed space is a real (or complex) vector space equipped with a map $\| \cdot \| : V \to F$ that satisfies the above 3 conditions. The map $\| \cdot \|$ is a *norm* of the space V.

So the inner product naturally induces a norm.

Given a norm $\| \cdot \|$, we can define a notion of "distance" between vectors (as the size of their difference), a notion of "neighbourhood" (consisting of vectors with small distances), and many more concepts from usual Euclidean spaces that you are familiar with.

[1] While in principle we *can* talk about arbitrary scalar field, it requires some (strong) conditions on the scalar field to make sense of the conditions. See this question on MSE and this question on MO. For simplicity, in this course we will only consider real or complex inner product spaces.

1.3 Properties of Inner Product and Norm

Here are some properties of inner product and the induced norm that are easy to verify:

- $\| x \|^2 = \langle x, x \rangle \geq 0$ and equality holds if and only if $x = 0$
- $\left\langle \sum_j c_j x_j, \sum_k d_k y_k \right\rangle = \sum_j \sum_k c_j \overline{d_k} \langle x_j, y_k \rangle$. In particular, for each $x \in V$, $y \mapsto \langle x, y \rangle$ is conjugate linear: $\langle x, ay + bz \rangle = \overline{a} \langle x, y \rangle + \overline{b} \langle x, z \rangle$
- $\langle x, y \rangle = 0$ for all y if and only if $x = 0$. Equivalently, $\langle x, y \rangle = \langle z, y \rangle$ for all y if and only if $x = z$.
- (Cauchy–Schwartz inequality) $| \langle x, y \rangle | \leq \| x \| \| y \|$
- (Pythagorean theorem) If $\langle v, w \rangle = 0$, $\| v + w \|^2 = \| v \|^2 + \| w \|^2$
- (Parallelogram law) $\| x + y \|^2 + \| x - y \|^2 = 2 \| x \|^2 + 2 \| y \|^2$

There are many uses of Cauchy–Schwartz inequality that you may have already seen in high school.

2 Gram–Schmidt Process

Definition 2.1. In an inner product space,

- a set S of vectors is *orthogonal* if every distinct $x, y \in S$, $\langle x, y \rangle = 0$
- a vector v is a *unit vector* if $\| v \| = 1$
- a set S is *orthonormal* if it is an orthogonal set of unit vectors. If $S = \{ v_1, v_2, \ldots \}$, this means that $\langle v_j, v_k \rangle = \delta_{jk}$

If $\beta = \{ e_1, \ldots, e_n \}$ is an orthogonal basis, then we have a simple computation of the representation of each vector: $v = \sum \frac{\langle v, e_j \rangle}{\| e_j \|^2} e_j$. If β is also orthonormal, this simplifies to $v = \sum \langle v, e_j \rangle e_j$. This property of orthonormal basis greatly simplifies many computations (e.g. computing coordinate / matrix representation).

Given a (countable[2]) linearly independent set $S = \{ v_1, v_2, \ldots \}$, *Gram–Schmidt process* can convert it to a linearly independent orthogonal set S' that has the same span as S:

- $w_1 = v_1$
- for $j \geq 2$, $w_j = v_j - \sum_{k<j} \frac{\langle v_j, w_k \rangle}{\| w_k \|^2} w_k = v_j - (\text{ projection of } v_j \text{ onto Span}(\{ w_1, \ldots, w_{j-1} \}))$

then $S' = \{ w_1, w_2, \ldots \}$. As noted in lecture,

- $w_j = 0$ if $\{ v_1, \ldots, v_j \}$ is linearly dependent, so $w_j \neq 0$ if S is linearly independent
- S' is orthogonal and so is also linearly independent (as $0 \notin S'$)
- Span(S') = Span(S), and in particular S' is an orthongonal basis of Span(S)

Typically, we also normalize S' to $S'' = \{ e_1, e_2, \ldots \}$ with $e_j = w_j / \| w_j \|$, so that S'' is orthonormal.

The same process can also be applied to linearly *dependent* S, as long as we skip all v_j and w_j that are zero.

3 Exercises

1. (Textbook Sec. 6.1 Q20)

   Let V be an inner product space over F. Prove the *polarization identities*: For all $x, y \in V$,

   (a) $\langle x, y \rangle = \frac{1}{4} \| x + y \|^2 - \frac{1}{4} \| x - y \|^2$ if $F = \mathbb{R}$

   (b) $\langle x, y \rangle = \frac{1}{4} \sum_{k=1}^{4} i^k \left\| x + i^k y \right\|^2$ if $F = \mathbb{C}$

[2] It is possible to consider *uncountably* infinite set, although this would require more machineries.

Solution:

(a) For $x, y \in V$,

$$\begin{aligned}
&\frac{1}{4} \| x+y \|^2 - \frac{1}{4} \| x-y \|^2 \\
&= \frac{1}{4} (\langle x+y,\, x+y \rangle - \langle x-y,\, x-y \rangle) \\
&= \frac{1}{4} (\langle x,\, x \rangle + \langle x,\, y \rangle + \langle y,\, x \rangle + \langle y,\, y \rangle - \langle x,\, x \rangle + \langle x,\, y \rangle + \langle y,\, x \rangle - \langle y,\, y \rangle) \\
&= \frac{1}{2} (\langle x,\, y \rangle + \langle y,\, x \rangle) \\
&= \langle x,\, y \rangle
\end{aligned}$$

(b) For $x, y \in V$,

$$\begin{aligned}
\frac{1}{4} \sum_{k=1}^{4} i^k \left\| x + i^k y \right\|^2 &= \frac{1}{4} \sum_{k=1}^{4} i^k \left\langle x + i^k y,\, x + i^k y \right\rangle \\
&= \frac{1}{4} \sum_{k=1}^{4} i^k \left(\langle x,\, x \rangle + \left\langle x,\, i^k y \right\rangle + \left\langle i^k y,\, x \right\rangle + \left\langle i^k y,\, i^k y \right\rangle \right) \\
&= \frac{1}{4} \sum_{k=1}^{4} i^k \left(\langle x,\, x \rangle + i^{-k} \langle x,\, y \rangle + i^k \langle y,\, x \rangle + \langle y,\, y \rangle \right) \\
&= \frac{1}{4} \left(\sum_{k=1}^{4} i^k \right) (\langle x,\, x \rangle + \langle y,\, y \rangle) + \frac{1}{4} \sum_{k=1}^{4} \langle x,\, y \rangle + \frac{1}{4} \sum_{k=1}^{4} i^{2k} \langle y,\, x \rangle \\
&= \langle x,\, y \rangle
\end{aligned}$$

Note

A result attributed to von Neumann (sometimes together with Fréchet and Jordan) states that a norm is induced from an inner product if and only if it satisfies the parallelogram law, in which case the inner product can be recovered from the polarization identity.

With the same approach, you can show the following result: on $n \geq 3$, if $\omega \neq \pm 1$ satisfies $\omega^n = 1$, then $\langle x,\, y \rangle = \frac{1}{n} \sum_{k=1}^{n} \omega^k \left\| x + \omega^k y \right\|^2$ on a complex inner product space.

2. Let V be a *complex* inner product space. Find all $T \in L(V)$ such that $\langle Tv,\, v \rangle = 0$ for all $v \in V$.

Solution: Let $T \in L(V)$ be one such map.
Let $x \in V$, $y = Tx \in V$. Then

$$\begin{aligned}
0 &= \langle T(x+y),\, x+y \rangle \\
&= \langle Tx,\, x \rangle + \langle Ty,\, y \rangle + \langle Tx,\, y \rangle + \langle Ty,\, x \rangle \\
&= \langle Tx,\, y \rangle + \langle Ty,\, x \rangle \\
0 &= \langle T(x+iy),\, x+iy \rangle \\
&= \langle Tx,\, x \rangle + \langle T(iy),\, iy \rangle + \langle Tx,\, iy \rangle + \langle T(iy),\, x \rangle \\
&= -i \langle Tx,\, y \rangle + i \langle Ty,\, x \rangle
\end{aligned}$$

which implies that $\| Tx \|^2 = \langle Tx,\, y \rangle = 0$, and so $Tx = 0$.
As $x \in V$ is arbitrary, $T = 0$.

It is easy to verify that $T = 0$ satisfies the condition, so the zero map is the only linear map that satisfies the condition.

Note

You can see that such operator must be "close to zero" in some sense: if λ is an eigenvalue of T with an associated eigenvector $v \neq 0$, then $0 = \langle Tv, v \rangle = \lambda \| v \|^2$, so $\lambda = 0$. This means that such operator cannot have any eigenvalue other than 0. Of course, this is not sufficient to claim that $T = 0$ (e.g. T may not have any eigenvalue at all).

For arguably a more pragmatic approach (that utilizes the concept of *self-adjoint*, which should be covered in lecture soon), see this answer on MSE.

This result does not hold if the space is over real numbers: consider $V = \mathbb{R}^2$ equipped with the usual structures, and $T \in L(V)$ is the rotation by $\pi/2$ radian, or more precisely, T is the left multiplication by the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

3. Let V be an inner product space, $x, y \in V$, $\lambda \in [0, 1]$, $z = \lambda x + (1 - \lambda)y$.
   Show that $\| x \|^2 \| z - y \| + \| y \|^2 \| z - x \| = \| x - y \| (\| z \|^2 + \| x - z \| \| y - z \|)$.

Solution:

$$
\begin{aligned}
& \| x - y \| (\| z \|^2 + \| x - z \| \| y - z \|) \\
&= \| x - y \| \Big(\| \lambda x + (1 - \lambda)y \|^2 + \| (1 - \lambda)(x - y) \| \| \lambda(y - x) \| \Big) \\
&= \| x - y \| \Big(\| \lambda x + (1 - \lambda)y \|^2 + \lambda(1 - \lambda) \| x - y \|^2 \Big) \\
&= \| x - y \| \Big((\lambda^2 + \lambda(1 - \lambda)) \| x \|^2 + ((1 - \lambda)^2 + \lambda(1 - \lambda)) \| y \|^2 \Big) \\
&= \| x \|^2 \| \lambda(x - y) \| + \| y \|^2 \| (1 - \lambda)(x - y) \| \\
&= \| x \|^2 \| z - y \| + \| y \|^2 \| z - x \|
\end{aligned}
$$

Note

In the context of Euclidean geometry (with $V = \mathbb{R}^2$ with the usual structures), this is Stewart's theorem that you may have learned in high school (especially if you did math competitions), which states that in a triangle ABC with D being a point on BC, $| AB |^2 | CD | + | AC |^2 | BD | = | BC | (| AD |^2 + | BD | | CD |)$.

If $\lambda = 1/2$, this is Apollonius's theorem, which states that in a triangle ABC with median AD, $| AB |^2 + | AC |^2 = 2(| AD |^2 + | BD |^2)$.

4. Let V be an inner product space, $n \in \mathbb{Z}^+$, $S = \{ v_1, \ldots, v_n \} \subseteq V$ be linearly independent. Show that there exists $w \in V$ such that $\langle w, v_j \rangle = 1$ for each j.

Solution: We will find one such vector w in Span(S). For simplicity, we will instead work with orthonormal $S' = \{ e_1, \ldots, e_n \}$ by applying Gram–Schmidt process on S, which by the property of the process has the same span Span(S') = Span(S).

A vector $w = \sum c_k e_k \in$ Span(S') with $c_1, \ldots, c_n \in F$ satisfies the condition if and only if the coefficients c_k satisfy $1 = \langle w, v_j \rangle = \sum_k \langle e_k, v_j \rangle c_k$ for each j. So, such vector $w \in$ Span(S') exists if and only if the equation

$$Lc = s$$

has a solution for c, with column vectors $c = \begin{pmatrix} c_1 & \dots & c_n \end{pmatrix}^\mathsf{T}$, $s = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}^\mathsf{T}$, and matrix $L_{jk} = \langle\, e_k ,\, v_j \,\rangle$.

By the definition of Gram–Schmidt process, $e_1 = a_1^{-1} v_1$ and $e_j = a_j^{-1}(v_j - \sum_{k=1}^{j-1} \langle\, v_j ,\, e_k \,\rangle e_k)$ for $j \geq 2$ with $a_1 = \|\, v_1 \,\| > 0$, $a_j = \left\|\, v_j - \sum_{k=1}^{j-1} \langle\, v_j ,\, e_k \,\rangle e_k \,\right\| > 0$, with positivity due to S being linearly independent.
This implies that $v_1 = a_1 e_1$ and $v_j = a_j e_j + \sum_{k=1}^{j-1} \langle\, v_j ,\, e_k \,\rangle e_k$ for $j \geq 2$. In particular, for each j we have $v_j \in \operatorname{Span}(\ \{\, e_1, \dots, e_j \,\}\ )$.
Since $S' = \{\, e_1, \dots, e_n \,\}$ is orthonormal, for all j, k, $L_{jj} = \langle\, e_j ,\, v_j \,\rangle = a_j > 0$ and $L_{jk} = \langle\, e_k ,\, v_j \,\rangle = 0$ whenever $k > j$.
This implies that L is a lower triangular matrix with positive diagonal entries, and so L is invertible.
Therefore, there exists a (unique) solution $c = \begin{pmatrix} c_1 & \dots & c_n \end{pmatrix}^\mathsf{T}$ that solves $Lc = s$, and so such vector w exists.

Note

The same argument also shows that on any given $s_1, \dots, s_n \in F$ there exists a *unique* vector $w \in \operatorname{Span}(\ S\ ) \subseteq V$ such that $\langle\, w ,\, v_j \,\rangle = s_j$. We can compute the coefficients iteratively as $c_j = a_j^{-1}(s_j - \sum_{k=1}^{j-1} c_k \langle\, v_j ,\, e_k \,\rangle)$ for all j (with usual convention for empty sum). The expressions being computationally simple is exactly due to L being lower triangular.

If we abuse notation and write $e = \begin{pmatrix} e_1 & \dots & e_n \end{pmatrix}$ as a row vector, then $w = ec = eL^{-1}s$, which indicates that s is the coordinate representation of w in the basis defined by eL^{-1}.

More concretely, as noted in the proof, $v_j = \sum_{k=1}^{n} \overline{L_{jk}} e_k = \sum_{k=1}^{n} e_k R_{kj}$ with $R = L^*$ being the conjugate transpose of L which is upper triangular, so on any given orthonormal basis α (of any finite dimensional subspace of V that contains $\operatorname{Span}(\ S\ )$) and $A = \begin{pmatrix} [v_1]_\alpha & \dots & [v_n]_\alpha \end{pmatrix}$, $A = QR$ with $Q = \begin{pmatrix} [e_1]_\alpha & \dots & [e_n]_\alpha \end{pmatrix}$ which satisfies $Q^*Q = I_n$ (this is the QR decomposition of A). The vector w can then be represented as $[w]_\alpha = Qc = QL^{-1}s = A(R^*R)^{-1}s$, or in a basis-free way $w = \sum_k s_k (\sum_j e_j (L^{-1})_{jk}) = \sum ((R^*R)^{-1}s)_j v_j$.