MATH2040A Week 11 Tutorial Notes

This tutorial session is mostly a review session.

1 Gram–Schmidt Process

Recall that the *Gram–Schmidt process* converts a list $v_1, \dots, v_n$ of linearly independent vectors into a list of orthogonal vectors $w_1, \dots, w_n$ with

- $w_1 = v_1$, $e_1 = w_1 / \| w_1 \|$
- for $j \geq 2$, $w_j = v_j - \sum_{k<j} \frac{\langle v_j, w_k \rangle}{\| w_k \|^2} w_k = v_j - \sum_{k<j} \langle v_j, e_k \rangle e_k$, $e_j = w_j / \| w_j \|$

The output of Gram–Schmidt process has the following properties, which are easy to show and sometimes handy:

- for each m, if $v_1, \dots, v_m$ are linearly independent, then so are $w_1, \dots, w_m$, which are also orthogonal and $\mathrm{Span}(\{ v_1, \dots, v_m \}) = \mathrm{Span}(\{ w_1, \dots, w_m \})$
- $v_j - w_j$ is the orthogonal projection of v_j onto $\mathrm{Span}(\{ v_1, \dots, v_{j-1} \})$
- $v_j = w_j + \sum_{k<j} c_k w_k$ for some $c_1, \dots, c_{j-1} \in F$. In particular, $\langle v_j, w_j \rangle = 1$, $\langle v_j, e_j \rangle = 1/ \| w_j \| > 0$
- for each j, on $\beta = \{ v_1, \dots, v_j \}$, $\gamma = \{ w_1, \dots, w_j \}$, $[\mathrm{Id}_{\mathrm{Span}(\beta)}]_\beta^\gamma$ is upper triangular and all diagonal entries are 1

2 Orthogonal Complement

Recall that the *orthogonal complement* of a set $S \subseteq V$ is the set $S^\perp = \{ v \in V \mid \langle v, u \rangle = 0, \forall u \in S \}$. The basic properties are:

- $S^\perp = \mathrm{Span}(S)^\perp$ is a subspace
- if $S_1 \subseteq S_2$, $S_1^\perp \supseteq S_2^\perp$
- $(S^\perp)^\perp \supseteq S$, and if U is a finite dimensional subspace, $(U^\perp)^\perp = U$
- if $U_1, U_2 \subseteq V$ are subspaces, $(U_1 + U_2)^\perp = U_1^\perp \cap U_2^\perp$. If furthermore V is finite dimensional, $(U_1 \cap U_2)^\perp = U_1^\perp + U_2^\perp$
- if U is a finite dimensional subspace, $V = U \oplus U^\perp$. In particular, if V is also finite dimensional, $\dim(V) = \dim(U) + \dim(U^\perp)$

3 Orthogonal Projection

Recall that for a finite dimensional[1] subspace $U \subseteq V$ with orthonormal basis $\alpha = \{ e_1, \dots, e_n \}$, the *orthogonal projection* P_U onto U is the linear map defined by $P_U(v) = \sum \langle v, e_j \rangle e_j$ for $v \in V$. The basic properties of P_U are

- P_U is a projection onto U along $U^\perp$. In particular, this means that

[1] It is possible to consider orthogonal projection for infinite dimensional subspace, although additional conditions on U are needed.

- P_U is idempotent: $P_U^2 = P_U$
- $\mathsf{R}(\, P_U \,) = U$ and $\mathsf{N}(\, P_U \,) = U^\perp$
- U is the set of fixed points of P_U: $U = \{\, v \in V \mid P_U(v) = v \,\}$

- $\| \, P_U(v) \, \| \leq \| \, v \, \|$
- for each $v \in V$, $P_U(v)$ is the unique minimizer of $\| \, u - v \, \|$ on U, that is, $P_U(x) \in U$ is the optimal approximation of v on U

4 Exercises

1. Equip $\mathbb{R}^n$ with the usual inner product $\langle \, x, \, y \, \rangle = \sum_{i=1}^n x_i y_i$.
Suppose that $v_0 = (1, 1, \ldots, 1), v_1 = (x_1, \ldots, x_n) \in \mathbb{R}^n$ are linearly independent.
Find the optimal approximation of $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ in $U = \text{Span}(\, \{ \, v_0, v_1 \, \} \,)$ and represent it with v_0, v_1, $m_x = \frac{1}{n} \sum x_i$, $m_y = \frac{1}{n} \sum y_i$, $m_{xx} = \frac{1}{n} \sum x_i^2$, $m_{xy} = \frac{1}{n} \sum x_i y_i$.

Solution: We apply Gram–Schmidt process on v_0, v_1:

- $w_0 = v_0$ with $\| \, w_0 \, \|^2 = n$
- $w_1 = v_1 - \frac{\langle \, v_1, \, w_0 \, \rangle}{\| \, w_0 \, \|^2} w_0 = (x_1, \ldots, x_n) - \frac{\sum x_i}{n}(1, \ldots, 1) = v_1 - m_x v_0$ with $\| \, w_1 \, \|^2 = \sum (x_i - m_x)^2 = n(m_{xx} - m_x^2)$

so the orthogonal projection $P_U(y)$ of y is

$$\begin{aligned} P_U(y) &= \frac{\langle \, y, \, w_0 \, \rangle}{\| \, w_0 \, \|^2} w_0 + \frac{\langle \, y, \, w_1 \, \rangle}{\| \, w_1 \, \|^2} w_1 \\ &= \frac{\sum y_i}{n} v_0 + \frac{\sum x_i y_i - m_x \sum y_i}{n(m_{xx} - m_x^2)} (v_1 - m_x v_0) \\ &= m_y v_0 + \frac{m_{xy} - m_x m_y}{m_{xx} - m_x^2} (v_1 - m_x v_0) = \frac{m_{xx} m_y - m_x m_{xy}}{m_{xx} - m_x^2} v_0 + \frac{m_{xy} - m_x m_y}{m_{xx} - m_x^2} v_1 \end{aligned}$$

Note

This is the equation for the (simple linear) regression line which minimizes the square error $\sum (y_i - L(x_i))^2$ on data $(x_1, y_1), \ldots, (x_n, y_n)$.

Solution: Alternatively, we can use the result from the last tutorial session exercise.
The Gram matrix of $\{ \, v_0, v_1 \, \}$ is

$$\begin{aligned} G &= \begin{pmatrix} \langle \, v_0, \, v_0 \, \rangle & \langle \, v_1, \, v_0 \, \rangle \\ \langle \, v_0, \, v_1 \, \rangle & \langle \, v_1, \, v_1 \, \rangle \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} = n \begin{pmatrix} 1 & m_x \\ m_x & m_{xx} \end{pmatrix} \\ \text{so} \quad G^{-1} &= \frac{1}{n} \frac{1}{m_{xx} - m_x^2} \begin{pmatrix} m_{xx} & -m_x \\ -m_x & 1 \end{pmatrix} \end{aligned}$$

Also, $\begin{pmatrix} \langle \, y, \, v_0 \, \rangle & \langle \, y, \, v_1 \, \rangle \end{pmatrix}^\mathsf{T} = \begin{pmatrix} \sum y_i & \sum x_i y_i \end{pmatrix}^\mathsf{T} = n \begin{pmatrix} m_y & m_{xy} \end{pmatrix}^\mathsf{T}$.
Thus, $G^{-1} \begin{pmatrix} \langle \, y, \, v_0 \, \rangle & \langle \, y, \, v_1 \, \rangle \end{pmatrix}^\mathsf{T} = \frac{1}{m_{xx} - m_x^2} \begin{pmatrix} m_{xx} m_y - m_x m_{xy} & m_{xy} - m_x m_y \end{pmatrix}^\mathsf{T}$.
By the result from the last tutorial session exercise,

$$P_U(y) = \frac{m_{xx} m_y - m_x m_{xy}}{m_{xx} - m_x^2} v_0 + \frac{m_{xy} - m_x m_y}{m_{xx} - m_x^2} v_1$$

2. Let $U \subseteq V$ be a finite dimensional subspace of an inner product space V, and $x \in V$.
Show that $y = P_U(x) \in U$ is the unique vector in U such that $\mathrm{Re}\langle\, x-y,\, u-y \,\rangle \le 0$ for all $u \in U$, where $\mathrm{Re}(z) = \frac{1}{2}(z+\overline{z})$ is the real part of the complex number $z \in \mathbb{C}$.

In another word, $y = P_U(x)$ is the only element in U such that $x-y$ and $u-y$ "do not form an acute angle" for any $u \in U$.

Solution: By property of orthogonal projection, $x - y = x - P_U(x) \in U^\perp$. Also, for all $u \in U$, $u - y \in U$. This implies that $\mathrm{Re}\langle\, x-y,\, u-y \,\rangle = \mathrm{Re}(0) = 0 \le 0$.

Let $y' \in U \setminus \{y\}$. Then $y - y' \in U$ and is nonzero.
This implies that $\langle\, x-y',\, y-y' \,\rangle = \langle\, x-y,\, y-y' \,\rangle + \langle\, y-y',\, y-y' \,\rangle = \|\, y-y' \,\|^2 > 0$.
In particular, $\mathrm{Re}\langle\, x-y',\, y-y' \,\rangle > 0$.

Combined, $y = P_U(x)$ is the only element in U such that $\mathrm{Re}\langle\, x-y,\, u-y \,\rangle \le 0$ for all $u \in U$.

Solution: Here is a proof that uses the characterization of orthogonal projection being the unique optimal approximation.

$y = P_U(x)$ is the optimal approximation of x in U if and only if $\|\, y-x \,\| < \|\, u-x \,\|$ for all $u \in U \setminus \{y\}$.
As U is a subspace, this is equivalent to

$$\begin{aligned} \|\, y-x \,\|^2 &< \|\, \lambda u + (1-\lambda)y - x \,\|^2 \\ &= \|\, y-x+\lambda(u-y) \,\|^2 \\ &= \|\, y-x \,\|^2 - 2\lambda \mathrm{Re}\langle\, x-y,\, u-y \,\rangle + \lambda^2 \|\, u-y \,\|^2 \end{aligned}$$

$$\text{or equivalently, } \mathrm{Re}\langle\, x-y,\, u-y \,\rangle < \frac{\lambda}{2}\|\, u-y \,\|^2$$

for all $u \in U \setminus \{y\}$ and $\lambda \in (0,1]$.
As $\|\, u-y \,\| > 0$ for all $u \in U \setminus \{y\}$, this is equivalent to $\mathrm{Re}\langle\, x-y,\, u-y \,\rangle \le 0$ for all $u \in U \setminus \{y\}$.
Trivially, $\mathrm{Re}\langle\, x-y,\, u-y \,\rangle \le 0$ holds on $u = y$ as well, so this is equivalent to $\mathrm{Re}\langle\, x-y,\, u-y \,\rangle \le 0$ for all $u \in U$.
Therefore, for $y \in U$, $y = P_U(x)$ if and only if $\mathrm{Re}\langle\, x-y,\, u-y \,\rangle \le 0$ for all $u \in U$.

Note

Using this proof, the same conclusion can be shown to hold true for a more general class of set U that is not necessarily a subspace, as long as U is still sufficiently "nice".

3. Let V be a finite dimensional inner product space with a basis $\beta = \{\, v_1, \dots, v_n \,\}$ and corresponding Gram matrix $G \in F^{n\times n}$ as defined by $G_{jk} = \langle\, v_k,\, v_j \,\rangle$ for all j,k, $f \in V^* = L(V,F)$, $U = \mathsf{N}(\, f \,)$.
Represent $\left\{\, [v]_\beta \;\middle|\; v \in U^\perp \,\right\}$ with $a = \begin{pmatrix} \overline{f(v_1)} & \dots & \overline{f(v_n)} \end{pmatrix}^\mathsf{T} \in F^n$ and the Gram matrix G.

Solution: Suppose $a = 0_{F^n}$. Then $f = 0$ on a basis β of V and so $f = 0$ on V.
This implies that $U = \mathsf{N}(\, f \,) = V$, $U^\perp = \{0\}$ and so $\left\{\, [v]_\beta \;\middle|\; v \in U^\perp \,\right\} = \{0_{F^n}\} = \mathrm{Span}\big(\, \{\, G^{-1}a \,\} \,\big)$.
Thus, in the following argument we may assume that $a \neq 0_{F^n}$.

Since $a \neq 0_{F^n}$, we must have $f \neq 0$, and so $\mathrm{rank}(f) = 1$. By dimensional theorem and the property of orthogonal complement, $\dim(U^\perp) = \dim(V) - \dim(U) = \dim(V) - \mathrm{nullity}(f) = \mathrm{rank}(f) = 1$.
To find $U^\perp$, it then suffices to find a nonzero vector in $U^\perp$.

Let $w = \sum (G^{-1}a)_j v_j \in V$. Since $a \neq 0_{F^n}$, $[w]_\beta = G^{-1}a \neq 0_{F^n}$, so $w \neq 0$.
Let $u = \sum c_j v_j \in U$ with $c_1, \dots, c_n \in F$.

Then $\langle\, w,\, u \,\rangle = \sum (G^{-1}a)_k \overline{c_j} \,\langle\, v_k,\, v_j \,\rangle = \sum \overline{c_j} G_{jk} (G^{-1}a)_k = \sum \overline{c_j} a_j = \overline{\sum c_j f(v_j)} = \overline{f(u)} = 0$.
As $u \in U$ is arbitrary, $w \in U^\perp$.

These imply that $U^\perp = \mathrm{Span}(\ \{\ w\ \}\ )$, and so $\left\{\ [v]_\beta \;\middle|\; v \in U^\perp\ \right\} = \mathrm{Span}\big(\ \{\ G^{-1}a\ \}\ \big)$.

Note

w is constructed by noting the fact that $\langle\, w,\, u \,\rangle = [u]_\beta^* G [w]_\beta$.

Since $(U_1 \cap U_2)^\perp = U_1^\perp + U_2^\perp$, with $f_1, \ldots, f_m \in V^*$ we can show that $\left\{\ [v]_\beta \;\middle|\; v \in (\ \bigcap \mathsf{N}(\ f_k\ )\ )^\perp\ \right\} = \mathsf{R}\big(\ G^{-1}A^*\ \big)$ with $A_{jk} = f_j(v_k)$.

4. Equip $\mathsf{P}(\mathbb{R})$ with inner product $\langle\, f,\, g \,\rangle = \int_{-1}^1 f(x)g(x)e^x \,\mathrm{d}x$.
Suppose $p_0, p_1, \ldots \in \mathsf{P}(\mathbb{R})$ are orthogonal, and p_n has degree $\deg(p_n) = n$ and leading coefficient 1 for each n.
Show that for $n \geq 1$, $p_{n+1} = (x - a_n)p_n - b_n p_{n-1}$ with $a_n = \frac{\langle\, x,\, p_n^2 \,\rangle}{\langle\, 1,\, p_n^2 \,\rangle}$, $b_n = \frac{\langle\, x^n,\, p_n \,\rangle}{\langle\, x^{n-1},\, p_{n-1} \,\rangle}$.

Solution: For each n, since $p_0, \ldots, p_n \in \mathsf{P}_n(\mathbb{R})$ are $n+1 = \dim(\mathsf{P}_n(\mathbb{R}))$ vectors with distinct degree, they are linearly independent and so form a basis of $\mathsf{P}_n(\mathbb{R})$.

Let $n \geq 1$.

As $\deg(xp_n) = n+1$, $xp_n \in \mathrm{Span}(\ \{\ p_0, \ldots, p_{n+1}\ \}\ )$, so $xp_n = \sum_{i=0}^{n+1} c_{ni} p_i$ with $c_{ni} = \langle\, xp_n,\, p_i \,\rangle / \|\ p_i\ \|^2$.
Since for $i < n-1$, $\deg(xp_i) = i+1 < n$, we must have $\langle\, xp_n,\, p_i \,\rangle = \int xp_n(x)p_i(x)w(x) = \langle\, p_n,\, xp_i \,\rangle = 0$, so $c_{ni} = 0$ for all such i.
Also, as p_{n-1}, p_n all have degree less than $n+1$, and xp_n, p_{n+1} have leading coefficient 1, we must have $c_{n,n+1} = 1$.
This implies that $xp_n = p_{n+1} + c_{n,n}p_n + c_{n,n-1}p_{n-1}$, so $p_{n+1} = (x - a_n)p_n - b_n p_{n-1}$ with $a_n = c_{nn} = \langle\, xp_n,\, p_n \,\rangle / \|\ p_n\ \|^2 = \langle\, x,\, p_n^2 \,\rangle / \langle\, 1,\, p_n^2 \,\rangle$ and $b_n = c_{n,n-1} = \langle\, xp_n,\, p_{n-1} \,\rangle / \|\ p_{n-1}\ \|^2$.

It remains to show that $b_n = \langle\, x^n,\, p_n \,\rangle / \langle\, x^{n-1},\, p_{n-1} \,\rangle$.
We first show that $\langle\, x^m,\, p_m \,\rangle \neq 0$ for all $m \geq 0$, so that the expression makes sense.
Since $x^m \in \mathrm{Span}(\ \{\ p_0, \ldots, p_m\ \}\ )$, we have $x^m = \sum_{i=0}^m \frac{\langle x^m, p_i \rangle}{\|\ p_i\ \|^2} p_i = \frac{\langle x^m, p_m \rangle}{\|\ p_m\ \|^2} p_m + \sum_{i=0}^{m-1} \frac{\langle x^m, p_i \rangle}{\|\ p_i\ \|^2} p_i$.
As $\deg(p_i) = i < m$ for all $i < m$, $\sum_{i=0}^{m-1} \frac{\langle x^m, p_i \rangle}{\|\ p_i\ \|^2} p_i$ has degree at most $m-1$.
Since $\deg(x^m) = m > m-1$, this implies that $\frac{\langle x^m, p_m \rangle}{\|\ p_m\ \|^2} \neq 0$ and so $\langle\, x^m,\, p_m \,\rangle \neq 0$.

Taking inner product with x^{n-1} on the recurrence relation, we have

$$\begin{aligned}
\langle\, x^{n-1},\, p_{n+1} \,\rangle &= \langle\, x^{n-1},\, xp_n \,\rangle - a_n \langle\, x^{n-1},\, p_n \,\rangle - b_n \langle\, x^{n-1},\, p_{n-1} \,\rangle \\
&= \langle\, x^n,\, p_n \,\rangle - a_n \langle\, x^{n-1},\, p_n \,\rangle - b_n \langle\, x^{n-1},\, p_{n-1} \,\rangle
\end{aligned}$$

Since $\deg(x^{n-1}) = n-1$, we have $x^{n-1} \in \mathrm{Span}(\ \{\ p_0, \ldots, p_{n-1}\ \}\ )$, and so by orthogonality $\langle\, x^{n-1},\, p_n \,\rangle = \langle\, x^{n-1},\, p_{n+1} \,\rangle = 0$.
This implies that $\langle\, x^n,\, p_n \,\rangle = b_n \langle\, x^{n-1},\, p_{n-1} \,\rangle$, so $b_n = \langle\, x^n,\, p_n \,\rangle / \langle\, x^{n-1},\, p_{n-1} \,\rangle$.

Note

The first three polynomials are $p_0 = 1$, $p_1 \approx x - 0.3130$, $p_2 \approx x^2 - 0.2688x - 0.2897$.

Using the recurrence relation, we can also show that $b_n = \frac{\|\ p_n\ \|^2}{\|\ p_{n-1}\ \|^2} > 0$. Note that $a_n = \frac{\langle xp_n, p_n \rangle}{\|\ p_n\ \|^2}$, so for the iteration we only need to compute $\|\ p_n\ \|^2$ and $\langle\, xp_n,\, p_n \,\rangle$ for each p_n.

While the inner product is defined with a specific weight e^x, we only need to use the property $\langle\, fg,\, h \,\rangle = \langle\, f,\, gh \,\rangle$ for polynomials f, g, h, so this result also holds for a general class of orthogonal polynomials (that are defined by inner product of the form $\langle\, f,\, g \,\rangle = \int f(x)g(x)w(x)$).