大學線性代數初步

李華介

國立台灣師範大學數學系

前言

本講義主要目的是針對數學系大一學生介紹有關線性代數基本的理論, 大致上僅談論實係數的向量空間.

一般經驗上大一學生會覺得線性代數學習起來會比微積分吃力, 主要是因素可能是在高中時期學習有關線性代數部分主要著重於操作而較少論證. 這一點在大學線性代數中就較不同了. 不過經驗上也告訴我們當學生到大三大四時 (因準備研究所考試) 再重新溫習線性代數時, 會覺得它不再那麼難以親近. 由此可知大學時期的線性代數其理論並不難懂, 較大的障礙是要開始學習數學的論證. 這個障礙或許到大三大四一些數學思維較成熟時稍可解除, 不過對一些同學來說可能為時已晚, 畢竟線性代數的理論或概念與其他課程都脫不了關係. 基於這個原因寫下這份講義, 希望藉由較平易近人的方式介紹線性代數也慢慢引導熟悉數學的論證方式. 本講義希望以淺顯易懂為主旨, 而不是生動有趣. 畢竟有些事情要說明清楚就會顯得囉唆, 當然就不有趣了.

研讀本講義的同學要有心理準備, 本講義是針對數學系學生而寫, 自然偏重於整個線性代數的理論架構. 對於線性代數在其他領域的應用幾乎沒有著墨. 我們依循一貫的原則就是理論清楚了, 接下來的應用或推廣就不難了. 所以對應用有興趣的同學應再參考其他的參考書籍. 另外本講義並未提供習題, 不過在某些概念講述之後有時會提供一些問題 (Question). 這些問題幾乎是檢視觀念是否正確或是對內容是否了解, 大部分問題若觀念已清楚應可以立即回答. 所以這些問題的份量仍不及一般習題, 對熟習線性代數所給予的訓練. 針對這一點, 請欲學習好線性代數的同學務必參閱一般線性代數書籍的習題, 自行磨練.

本講義雖然主要以中文撰寫, 不過當涉及定義或專有名詞時, 為免翻譯的困擾將以英文取代. 因此將以中英夾雜較不傳統的方式顯現, 若有不便請見諒.

本講義編寫費時, 編寫完後並未經過嚴謹的校對. 疏漏在所難免, 雖不至於有理論性上嚴重的錯誤, 但讀者仍應注意不宜概括全收. 若發現錯誤, 歡迎提出寶貴的意見.

Chapter 7

Inner Product Space

在 $\mathbb{R}^m$ 中有關於 dot product 的性質, 可以推廣到一般的 vector space. 在一般的 vector space, 我們就不再用 dot product 而用所謂 *inner product* 來稱之. 對於 vector space V 若對於任意 $\mathbf{v},\mathbf{w}\in V$ 皆有 $\langle\mathbf{v},\mathbf{w}\rangle\in\mathbb{R}$ 滿足 Proposition 1.4.2 有關內積的四個性質, 即

(1) $\langle\mathbf{v},\mathbf{w}\rangle=\langle\mathbf{w},\mathbf{v}\rangle$, $\forall\mathbf{v},\mathbf{w}\in V$.

(2) $\langle\mathbf{v},\mathbf{v}\rangle\geq 0$ 且 $\langle\mathbf{v},\mathbf{v}\rangle=0$ 若且唯若 $\mathbf{v}=\mathbf{O}$.

(3) 對任意 $r\in\mathbb{R}$ 皆有 $\langle r\mathbf{v},\mathbf{v}\rangle=\langle\mathbf{v},r\mathbf{w}\rangle=r\langle\mathbf{v},\mathbf{w}\rangle$.

(4) $\langle\mathbf{u},\mathbf{v}+\mathbf{w}\rangle=\langle\mathbf{u},\mathbf{v}\rangle+\langle\mathbf{u},\mathbf{w}\rangle$, $\forall\mathbf{u},\mathbf{v},\mathbf{w}\in V$.

我們便稱 V 為 *inner product space*. 在這一章中我們將介紹 inner product space 的性質. 由於這些性質僅利用上述 inner product 的性質便可得到, 所以只要是 finite dimensional inner product space 都會有我們介紹的這些性質. 不過因為大家對較抽象的 inner product 較不熟悉, 所以這裡我們僅用大家熟悉 $\mathbb{R}^m$ 中一般的內積來處理. 另外, 由於我們將利用矩陣的乘法來處理內積, 對於 $\mathbf{v},\mathbf{w}\in\mathbb{R}^m$, 原來我們用 $\mathbf{v}\cdot\mathbf{w}$ 表示 $\mathbf{v},\mathbf{w}$ 的內積, 在這裡為了避免混淆, 我們將改用 $\langle\mathbf{v},\mathbf{w}\rangle$ 來表示 $\mathbf{v},\mathbf{w}$ 的內積. 也就是說, 若將 $\mathbf{v},\mathbf{w}$ 視為 column vectors 且用矩陣的乘法表示, 我們有 $\langle\mathbf{v},\mathbf{w}\rangle=\mathbf{v}^{\mathrm{T}}\mathbf{w}$.

7.1. Projection

我們曾經在介紹 $\mathbb{R}^2$ 上的 linear transformation 時介紹 projection. 簡單來說, 給定非零向量 $\mathbf{w}\in\mathbb{R}^2$, 我們定義 $\mathbf{v}\in\mathbb{R}^2$ 在 $\mathbf{w}$ 上的投影 $\mathrm{Proj}_{\mathbf{w}}(\mathbf{v})$ 必須在 $\mathbf{w}$ 所展成的空間中 (即 $\mathrm{Span}(\mathbf{w})$), 而且 $\mathbf{v}-\mathrm{Proj}_{\mathbf{w}}(\mathbf{v})$ 必須與 $\mathbf{w}$ 垂直, 也因此需與 $\mathrm{Span}(\mathbf{w})$ 上所有向量垂直. 依此, 我們將投影的定義推廣到一般對 $\mathbb{R}^m$ 的一個 subspace W 上的投影. 也就是說, 當 W 為 $\mathbb{R}^m$ 的 subspace, 對於 $\mathbf{v}\in\mathbb{R}^m$, 我們定義 $\mathbf{v}$ 在 W 的 projection 為 W 上的一個向量 $\mathbf{w}$, 滿足 $\mathbf{v}-\mathbf{w}$ 和 W 上所有向量垂直. 首先我們給以下的定義.

Definition 7.1.1. 給定 W 為 $\mathbb{R}^m$ 的 subspace. 令

$$W^{\perp}=\{\mathbf{v}\in\mathbb{R}^m\mid\langle\mathbf{v},\mathbf{w}\rangle=0,\forall\mathbf{w}\in W\}.$$

也就是說 $W^{\perp}$ 為 $\mathbb{R}^m$ 中和所有 W 中的向量垂直的向量所成的集合. 一般稱 $W^{\perp}$ 為 *orthogonal complement* of W.

有了 orthogonal complement 的定義我們就可以對上述的 projection 給了以下的定義.

Definition 7.1.2. 給定 W 為 $\mathbb{R}^m$ 的 subspace. 對於 $\mathbf{v}\in\mathbb{R}^m$, 若 $\mathbf{w}\in W$ 滿足 $\mathbf{v}-\mathbf{w}\in W^{\perp}$, 則稱 $\mathbf{w}$ 為 the *projection* of $\mathbf{v}$ on W.

等一下我們將說明對於任意 $\mathbf{v}\in\mathbb{R}^m$, the projection of $\mathbf{v}$ on W 一定存在, 而且唯一. 因此為了方便起見我們將此向量用 $\mathrm{Proj}_W(\mathbf{v})$ 表示. 我們將先證明唯一性, 有了唯一性, 將來我們要說明一個 $\mathbb{R}^m$ 的向量 $\mathbf{w}$ 就是 $\mathrm{Proj}_W(\mathbf{v})$, 就僅要檢查 $\mathbf{w}\in W$ 且滿足 $\mathbf{v}-\mathbf{w}\in W^{\perp}$ 就可以了.

我們先了解以下一些有關於 orthogonal complement 的性質. 假設 W 為 $\mathbb{R}^m$ 的 subspace, 並令 $\mathbf{v},\mathbf{v}'\in W^{\perp}$ 以及 $r,s\in\mathbb{R}$. 利用內積的性質 (3)(4), 對於任意 $\mathbf{w}\in W$, 我們有 $\langle r\mathbf{v}+s\mathbf{v}',\mathbf{w}\rangle = r\langle\mathbf{v},\mathbf{w}\rangle+s\langle\mathbf{v}',\mathbf{w}\rangle$. 再利用 $\mathbf{v},\mathbf{v}'\in W^{\perp}$, 即 $\langle\mathbf{v},\mathbf{w}\rangle=\langle\mathbf{v}',\mathbf{w}\rangle=0$, 得知 $\langle r\mathbf{v}+s\mathbf{v}',\mathbf{w}\rangle=0$. 此即表示對於任意 $\mathbf{v},\mathbf{v}'\in W^{\perp}$ 以及 $r,s\in\mathbb{R}$, 皆有 $r\mathbf{v}+s\mathbf{v}'\in W^{\perp}$, 得證 $W^{\perp}$ 為 $\mathbb{R}^m$ 的 subspace.

要注意 orthogonal complement 的 complement 不是指集合的補集. W 的 orthogonal complement $W^{\perp}$ 並不是 W 的補集. 甚至 $W\cap W^{\perp}$ 並不會是空集合. 這是因為 $W\cap W^{\perp}$ 也會是一個 subspace, 所以 $\mathbf{O}$ 一定在其中. 事實上我們有以下的結果.

Lemma 7.1.3. 假設 W 為 $\mathbb{R}^m$ 的 *subspace.* 則 $W\cap W^{\perp}=\{\mathbf{O}\}$.

Proof. 因 $W\cap W^{\perp}$ 為 subspace, 故知 $\mathbf{O}\in W\cap W^{\perp}$. 現假設 $\mathbf{w}\in W\cap W^{\perp}$. 由於 $\mathbf{w}\in W^{\perp}$, 對任意 $\mathbf{w}'\in W$ 皆有 $\langle\mathbf{w},\mathbf{w}'\rangle=0$. 而又 $\mathbf{w}\in W$, 故得 $\langle\mathbf{w},\mathbf{w}\rangle=0$. 因此由內積的性質 (2) 知 $\mathbf{w}=\mathbf{O}$, 得證 $W\cap W^{\perp}=\{\mathbf{O}\}$. □

現在我們可以證明 projection 的唯一性.

Proposition 7.1.4. 給定 W 為 $\mathbb{R}^m$ 的 *subspace.* 對於 $\mathbf{v}\in\mathbb{R}^m$, $\mathbf{v}$ 對於 W 的 *projection* 是唯一的.

Proof. 假設 $\mathbf{w},\mathbf{w}'\in W$ 皆為 $\mathbf{v}$ 在 W 的 projection. 也就是說 $\mathbf{w},\mathbf{w}'$ 皆滿足 $\mathbf{v}-\mathbf{w}\in W^{\perp}$ 以及 $\mathbf{v}-\mathbf{w}'\in W^{\perp}$. 由於 $W^{\perp}$ 為 subspace, 我們有 $(\mathbf{v}-\mathbf{w})-(\mathbf{v}-\mathbf{w}')\in W^{\perp}$, 亦即 $\mathbf{w}-\mathbf{w}'\in W^{\perp}$. 又因 W 為 subspace, 我們也知 $\mathbf{w}-\mathbf{w}'\in W$. 故由 Lemma 7.1.3 知 $\mathbf{w}-\mathbf{w}'=\mathbf{O}$, 即 $\mathbf{w}=\mathbf{w}'$, 證得唯一性. □

接下來我們要探討 projection 的存在性. 由於 $W^{\perp}$ 為 subspace, 我們更進一步的想要知道其 dimension. 由於 $W^{\perp}$ 與 W 有關, 我們可以預期 $\dim(W^{\perp})$ 和 $\dim(W)$ 應該有關. 事實上假設 $\dim(W)=n$ 且令 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis. 考慮 $A=\begin{bmatrix} — & \mathbf{w}_1 & — \\ & \vdots & \\ — & \mathbf{w}_n & — \end{bmatrix}$, 即 A 的 i-th row 為 $\mathbf{w}_i$ (寫成 row vector). 由於 $\mathbf{w}_i\in\mathbb{R}^m$, 故知 A 為 $n\times m$ matrix. 對於任意 $\mathbf{v}\in W^{\perp}$, 由於 $\mathbf{w}_i\in W$, 我們有 $\langle\mathbf{w}_i,\mathbf{v}\rangle=0$, $\forall i=1,\ldots,n$. 因此由矩陣乘法定義知 $A\mathbf{v}=\mathbf{O}$, 也就是說 $W^{\perp}$ 包含於 A 的 nullspace. 反之, 若 $\mathbf{v}\in\mathbb{R}^m$ 為 A 的 nullspace 中一元素, 則由 $A\mathbf{v}=\mathbf{O}$

得 $\langle \mathbf{w}_i, \mathbf{v}\rangle = 0$, $\forall i = 1,\ldots,n$. 因此對任意 $\mathbf{w} \in W$, 由於 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis, 存在 $c_1,\ldots,c_n \in \mathbb{R}$ 使得 $\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n$. 故得

$$\langle \mathbf{w}, \mathbf{v}\rangle = \langle c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n, \mathbf{v}\rangle = c_1\langle \mathbf{w}_1, \mathbf{v}\rangle + \cdots + c_n\langle \mathbf{w}_n, \mathbf{v}\rangle = 0,$$

也就是說 A 的 nullspace 包含於 $W^{\perp}$. 我們得證了以下的性質.

Lemma 7.1.5. *給定 W 為 $\mathbb{R}^m$ 的 subspace. 假設 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis. 令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ (寫成 row vector) 為 row vector 的 $n \times m$ matrix. 則 $W^{\perp}$ 為 A 的 nullspace.*

Example 7.1.6. 令 $W = \mathrm{Span}(\begin{bmatrix}1\\2\\2\\1\end{bmatrix}, \begin{bmatrix}3\\4\\2\\3\end{bmatrix})$ 我們要找出 W 的 orthogonal complement $W^{\perp}$.

首先考慮 $A = \begin{bmatrix}1 & 2 & 2 & 1\\3 & 4 & 2 & 3\end{bmatrix}$. 由 Lemma 7.1.5, 我們知道 A 的 nullspace $\{\mathbf{x} \in \mathbb{R}^4 \mid A\mathbf{x} = \mathbf{O}\}$ 即為 $W^{\perp}$. 利用 elementary row operation 將 A 的 1-st row 乘以 -3 加到 2-nd row 得 $\begin{bmatrix}1 & 2 & 2 & 1\\0 & -2 & -4 & 0\end{bmatrix}$. 再將 2-nd row 乘以 $-1/2$ 得 $\begin{bmatrix}1 & 2 & 2 & 1\\0 & 1 & 2 & 0\end{bmatrix}$. 解得 $\begin{bmatrix}-1\\0\\0\\1\end{bmatrix}, \begin{bmatrix}2\\-2\\1\\0\end{bmatrix}$ 為 A 的 nullspace, 即 $W^{\perp}$ 的一組 basis.

Question 7.1. *若 Lemma 7.1.5 中 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 僅假設 $\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n) = W$ 但不假設 linearly independent, 是否仍可得 $W^{\perp}$ 為 A 的 nullspace?*

一般來說我們習慣將 $\mathbb{R}^m$ 的向量寫成 column vector, 所以若將 Lemma 7.1.5 中 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 寫成 column vector 且分別為 A 的 column vector (此時 A 為 $m \times n$ matrix), 我們可將 Lemma 7.1.5 改寫為以下定理.

Corollary 7.1.7. *給定 W 為 $\mathbb{R}^m$ 的 subspace. 假設 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis. 令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 column vector 的 $m \times n$ matrix. 則 $W^{\perp}$ 為 A^{T} 的 nullspace.*

在 Corollary 7.1.7 中依 column space 的定義 $W = C(A)$, 故 Corollary 7.1.7 又可改寫成以下定理.

Corollary 7.1.8. *假設 A 為 $m \times n$ matrix. 則 $C(A)^{\perp} = N(A^{\mathrm{T}})$.*

現在我們可以算出 $\dim(W^{\perp})$ 和 $\dim(W)$ 之間的關係了. 事實上假設 W 為 $\mathbb{R}^m$ 的 subspace 且 $\dim(W) = n$. 利用 W 的一組 basis 為 row vectors 所形成的 $n \times m$ matrix A, 並套用 Lemma 7.1.5 我們知道 $W^{\perp}$ 會是 A 的 nullspace. 也因此 $\dim(W^{\perp})$ 會是 A 的 nullspace 的維度 (即 A 的 nullity). 依定義 A 的 rank 會是 A 的 row space (即 A 的 row vectors 所展成的 space W) 的維度, 即 $\mathrm{rank}(A) = \dim(W) = n$. 故利用 Theorem 4.4.13 我們知道 A 的 nullity 為 A 的 column 的個數 m 減去 $\mathrm{rank}(A)$, 即 $\mathrm{null}(A) = m - \mathrm{rank}(A) = m - n$. 因此我們得到以下重要的性質.

Theorem 7.1.9. 假設 W 為 $\mathbb{R}^m$ 的 *subspace.* 則 $\dim(W^\perp) = m - \dim(W)$.

利用 Theorem 7.1.9 我們很快可以推出以下的結果.

Corollary 7.1.10. 假設 W 為 $\mathbb{R}^m$ 的 *subspace.* 則 $(W^\perp)^\perp = W$.

Proof. 由 Theorem 7.1.9 我們知道 $\dim((W^\perp)^\perp) = m - \dim(W^\perp) = m - (m - \dim(W)) = \dim(W)$. 然而給定 $\mathbf{w} \in W$, 由於對任意 $\mathbf{w}' \in W^\perp$, 皆有 $\langle \mathbf{w}, \mathbf{w}' \rangle = 0$, 故得 $\mathbf{w} \in (W^\perp)^\perp$. 因此知 $W \subseteq (W^\perp)^\perp$. 故由 $\dim(W) = \dim((W^\perp)^\perp)$ 利用 Corollary 4.3.6 得證 $(W^\perp)^\perp = W$. □

這裡我們要強調一下, Corollary 7.1.10 的證明是利用 dimension 來處理. 當 W 是 infinite dimensional vector space 的 subspace 時, $(W^\perp)^\perp = W$ 未必成立. 也就是說雖然我們證明了 $W \subseteq (W^\perp)^\perp$, 但另一方向 $(W^\perp)^\perp \subseteq W$ 未必正確. 而在 finite dimensional 的情形, 還好有 dimension 可以計算, 所以才可以藉由 Corollary 4.3.6 得證 $(W^\perp)^\perp = W$.

Question 7.2. 假設 A 為 $m \times n$ *matrix.* 試證明 $C(A) = N(A^{\mathrm{T}})^\perp$.

由 Theorem 7.1.9 我們知道若 W 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 為 W 的一組 basis, 則由於 $\dim(W^\perp) = m - n$, 我們可以找到一組 $W^\perp$ 的 basis $\mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$. 我們將證明此時 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 linearly independent, 因而得知以下的結果.

Corollary 7.1.11. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 為 W 的一組 *basis* 以及 $\mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 $W^\perp$ 的一組 *basis.* 則 $\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 $\mathbb{R}^m$ 的一組 *basis.*

Proof. 首先證明 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 linearly independent. 我們用反證法, 假設存在 $c_1, \ldots, c_n, c_{n+1}, \ldots, c_m \in \mathbb{R}$ 不全為 0 滿足 $c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n + c_{n+1}\mathbf{u}_1 + \cdots + c_m\mathbf{u}_{m-n} = \mathbf{O}$. 令 $\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n$, 由於 W 為 vector space, 我們有 $\mathbf{w} \in W$. 又由於 $\mathbf{w} = -(c_{n+1}\mathbf{u}_1 + \cdots + c_m\mathbf{u}_{m-n})$ 以及 $W^\perp$ 為 vector space, 我們也有 $\mathbf{w} \in W^\perp$. 換言之, $\mathbf{w} \in W \cap W^\perp$. 因此由 Lemma 7.1.3 得證 $\mathbf{w} = \mathbf{O}$. 再由 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 為 linearly independent 以及 $c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n = \mathbf{w} = \mathbf{O}$ 得 $c_1 = \cdots = c_n = 0$. 同理得 $c_{n+1} = \cdots = c_m = 0$. 此與 $c_1, \ldots, c_n, c_{n+1}, \ldots, c_m$ 不全為 0 之假設相矛盾, 得證 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 linearly independent. 最後再由 $\dim(\mathbb{R}^m) = m$, 以及 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 $\mathbb{R}^m$ 中 m 個 linearly independent vectors, 得知 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 $\mathbb{R}^m$ 的一組 basis. □

既然 W 的一組 basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ 和 $W^\perp$ 的一組 basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_{m-n}\}$ 可形成 $\mathbb{R}^m$ 的一組 basis, 對任意 $\mathbf{v} \in \mathbb{R}^m$, 存在 $c_1, \ldots, c_n, d_1, \ldots, d_{m-n} \in \mathbb{R}$ 使得

$$\mathbf{v} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n + d_1\mathbf{u}_1 + \cdots + d_{m-n}\mathbf{u}_{m-n}.$$

此時若令 $\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n$, 則 $\mathbf{w} \in W$ 且 $\mathbf{v} - \mathbf{w} = d_1\mathbf{u}_1 + \cdots + d_{m-n}\mathbf{u}_{m-n} \in W^\perp$. 因此依定義知 $\mathbf{w}$ 即為 the projection of $\mathbf{v}$ on W. 這也證明了 projection 的存在性. 綜合以上的結果, 我們有以下的定理.

Theorem 7.1.12. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *basis* 以及 $\mathbf{u}_1,\ldots,\mathbf{u}_{m-n}$ 為 $W^\perp$ 的一組 *basis.* 對於 $\mathbf{v}\in\mathbb{R}^m$, 若 $\mathbf{v}=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n+d_1\mathbf{u}_1+\cdots+d_{m-n}\mathbf{u}_{m-n}$, 則 $\mathbf{v}$ 在 W 的 *projection* 為

$$\mathrm{Proj}_W(\mathbf{v})=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n.$$

Example 7.1.13. 我們利用 Example 7.1.6 的結果求 $\mathbf{v}=\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$ 在 $W=\mathrm{Span}(\begin{bmatrix}1\\2\\2\\1\end{bmatrix},\begin{bmatrix}3\\4\\2\\3\end{bmatrix})$ 的 projection. 已知 $\begin{bmatrix}-1\\0\\0\\1\end{bmatrix},\begin{bmatrix}2\\-2\\1\\0\end{bmatrix}$ 為 $W^\perp$ 的一組 basis, 故將 $\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$ 寫成 $\begin{bmatrix}1\\2\\2\\1\end{bmatrix},\begin{bmatrix}3\\4\\2\\3\end{bmatrix},\begin{bmatrix}-1\\0\\0\\1\end{bmatrix},\begin{bmatrix}2\\-2\\1\\0\end{bmatrix}$ 的 linear combination 得

$$\begin{bmatrix}2\\0\\1\\4\end{bmatrix}=-\begin{bmatrix}1\\2\\2\\1\end{bmatrix}+\begin{bmatrix}3\\4\\2\\3\end{bmatrix}+2\begin{bmatrix}-1\\0\\0\\1\end{bmatrix}+\begin{bmatrix}2\\-2\\1\\0\end{bmatrix}.$$

因此求出 $\mathbf{v}$ 在 W 的 projection 為 $-\begin{bmatrix}1\\2\\2\\1\end{bmatrix}+\begin{bmatrix}3\\4\\2\\3\end{bmatrix}=\begin{bmatrix}2\\2\\0\\2\end{bmatrix}$.

給定 $\mathbb{R}^m$ 的一個 subspace W, 對於 $\mathbf{v}\in\mathbb{R}^m$, 既然 $\mathrm{Proj}_W(\mathbf{v})$ 是存在且唯一的, 我們可以定義 projection on W 這樣的函數. 也就是將 Proj_W 視為是一個 $\mathbb{R}^m\to\mathbb{R}^m$ 函數且對任意 $\mathbf{v}\in\mathbb{R}^m$ 的取值就是 $\mathrm{Proj}_W(\mathbf{v})$. 我們要說明 $\mathrm{Proj}_W:\mathbb{R}^m\to\mathbb{R}^m$ 事實上是一個 linear transformation.

首先對於任意 $\mathbf{v},\mathbf{v}'\in\mathbb{R}^m$, 為了方便我們令 $\mathbf{w}=\mathrm{Proj}_W(\mathbf{v})$ 以及 $\mathbf{w}'=\mathrm{Proj}_W(\mathbf{v}')$. 也就是說 $\mathbf{w},\mathbf{w}'\in W$ 且滿足 $\mathbf{v}-\mathbf{w}\in W^\perp$ 以及 $\mathbf{v}'-\mathbf{w}'\in W^\perp$. 我們要證明 $\mathrm{Proj}_W(\mathbf{v}+\mathbf{v}')=\mathrm{Proj}_W(\mathbf{v})+\mathrm{Proj}_W(\mathbf{v}')=\mathbf{w}+\mathbf{w}'$. 換言之, 我們要證明 $\mathbf{w}+\mathbf{w}'\in W$ 且滿足 $(\mathbf{v}+\mathbf{v}')-(\mathbf{w}+\mathbf{w}')\in W^\perp$. 然而 $W,W^\perp$ 皆為 vector space, 我們自然由 $\mathbf{w},\mathbf{w}'\in W$ 以及 $\mathbf{v}-\mathbf{w},\mathbf{v}'-\mathbf{w}'\in W^\perp$ 可得 $\mathbf{w}+\mathbf{w}'\in W$ 以及 $(\mathbf{v}+\mathbf{v}')-(\mathbf{w}+\mathbf{w}')=(\mathbf{v}-\mathbf{w})+(\mathbf{v}'-\mathbf{w}')\in W^\perp$. 另外對任意 $r\in\mathbb{R}$, 我們也要證明 $\mathrm{Proj}_W(r\mathbf{v})=r\mathrm{Proj}_W(\mathbf{v})=r\mathbf{w}$. 也就是證明 $r\mathbf{w}\in W$ 以及 $r\mathbf{v}-r\mathbf{w}\in W^\perp$. 同樣的由 $W,W^\perp$ 皆為 vector space, 以及 $\mathbf{w}\in W,\mathbf{v}-\mathbf{w}\in W^\perp$, 很容易看出這是對的. 這也證明了

Proposition 7.1.14. 假設 W 為 $\mathbb{R}^m$ 的 *subspace,* 則 $\mathrm{Proj}_W:\mathbb{R}^m\to\mathbb{R}^m$ 是一個 *linear transformation.*

Question 7.3. 試說明甚麼是 $\mathrm{Proj}_W:\mathbb{R}^m\to\mathbb{R}^m$ 的 *image* 以及 *kernel.*

其實利用 Theorem 7.1.12 來求 $\mathbf{v}$ 在 W 上的投影有點麻煩. 因為我們需利用 W 的一組 basis $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 求出 $W^\perp$ 的一組 basis $\mathbf{u}_1,\ldots,\mathbf{u}_{m-n}$, 再將 $\mathbf{v}$ 寫成 $\mathbf{w}_1,\ldots,\mathbf{w}_n,\mathbf{u}_1,\ldots,\mathbf{u}_{m-n}$ 的線性組合, 才能求出 $\mathrm{Proj}_W(\mathbf{v})$. 我們希望將此步驟簡化. 首先令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 column vectors 的 $m\times n$ matrix, 則 W 為 A 的 column space $C(A)$. 依 $\mathrm{Proj}_W(\mathbf{v})$ 的定義, 我們需要

找到 $\mathbf{w}\in W$ 滿足 $\mathbf{v}-\mathbf{w}\in W^{\perp}$, 這樣 $\mathbf{w}$ 就會是 $\mathrm{Proj}_W(\mathbf{v})$ 了. 如何找到這樣的 $\mathbf{w}\in W$ 呢? 由 column space 的定義知, $\mathbf{w}\in W$ 表示存在 $\mathbf{x}\in\mathbb{R}^n$ 使得 $A\mathbf{x}=\mathbf{w}$. 至於 $\mathbf{v}-\mathbf{w}\in W^{\perp}=(C(A))^{\perp}$ 的要求, 由 Corollary 7.1.8 知此即表示 $\mathbf{v}-\mathbf{w}=\mathbf{v}-A\mathbf{x}\in N(A^{\mathrm{T}})$. 也就是說我們必須找到 $\mathbf{x}\in\mathbb{R}^n$ 使得 $A^{\mathrm{T}}(\mathbf{v}-A\mathbf{x})=\mathbf{O}$. 利用矩陣乘法性質, 此即表示 $\mathbf{x}\in\mathbb{R}^n$ 須滿足聯立方程組 $(A^{\mathrm{T}}A)\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$. 要注意 $\mathrm{Proj}_W(\mathbf{v})$ 一定存在, 所以一定存在 $\mathbf{x}\in\mathbb{R}^n$ 使得 $A\mathbf{x}=\mathrm{Proj}_W(\mathbf{v})$, 也因此這個 $\mathbf{x}$ 必滿足 $\mathbf{v}-A\mathbf{x}\in W^{\perp}$, 所以聯立方程組 $(A^{\mathrm{T}}A)\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$ 一定有解. 若能解 $(A^{\mathrm{T}}A)\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$, 則所得的解 $\mathbf{x}$ 就會使得 $A\mathbf{x}=\mathrm{Proj}_W(\mathbf{v})$ 了. 問題是要如何解 $(A^{\mathrm{T}}A)\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$ 呢? 我們有以下的定理.

Lemma 7.1.15. 假設 $A\in\mathscr{M}_{m\times n}$ 且 $\mathrm{rank}(A)=n$, 則 $A^{\mathrm{T}}A$ 是一個 $n\times n$ *invertible matrix.*

Proof. 首先由 $A\in\mathscr{M}_{m\times n}$ 且 $A^{\mathrm{T}}\in\mathscr{M}_{n\times m}$, 我們知 $A^{\mathrm{T}}A\in\mathscr{M}_{n\times n}$. 因此可以利用 Theorem 3.5.9, 來證明 $A^{\mathrm{T}}A$ 為 invertible. 也就是說證明 $(A^{\mathrm{T}}A)\mathbf{x}=\mathbf{O}$ 沒有 nontrivial solution 就等同於證得 $A^{\mathrm{T}}A$ 為 invertible matrix. 假設 $\mathbf{x}=\mathbf{u}$ 為 $(A^{\mathrm{T}}A)\mathbf{x}=\mathbf{O}$ 的一個解, 亦即 $(A^{\mathrm{T}}A)\mathbf{u}=\mathbf{O}$. 由於 $(A^{\mathrm{T}}A)\mathbf{u}=A^{\mathrm{T}}(A\mathbf{u})$, 故得 $A\mathbf{u}\in N(A^{\mathrm{T}})$. 然而 $A\mathbf{u}\in C(A)$, 所以有 $A\mathbf{u}\in C(A)\cap N(A^{\mathrm{T}})=C(A)\cap C(A)^{\perp}$. 因此利用 Lemma 7.1.3 得證 $A\mathbf{u}=\mathbf{O}$. 然而依假設 $\mathrm{rank}(A)=n$, 故由 Corollary 2.3.5 知聯立方程組 $A\mathbf{x}=\mathbf{O}$ 沒有 nontrivial solution, 也就是說 $\mathbf{u}=\mathbf{O}$. 由此知 $(A^{\mathrm{T}}A)\mathbf{x}=\mathbf{O}$ 沒有 nontrivial solution, 故 $A^{\mathrm{T}}A$ 是一個 $n\times n$ invertible matrix. $\square$

前面我們假設 A 是以 W 的一組 basis 為 column vector 所得的 $m\times n$ matrix, 所以 $\mathrm{rank}(A)=\dim(W)=n$. 故利用 Lemma 7.1.15 可得 $A^{\mathrm{T}}A$ 為 invertible. 此時只要將聯立方程組 $(A^{\mathrm{T}}A)\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$ 的兩邊乘上 $A^{\mathrm{T}}A$ 的 inverse, 即可得解為 $\mathbf{x}=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{v}$. 注意此時我們僅解得 $(A^{\mathrm{T}}A)\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$ 之解, 要將此解的左邊乘上 A 才得 $\mathrm{Proj}_W(\mathbf{v})$. 我們有以下的結論.

Theorem 7.1.16. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *basis.* 令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 *column vector* 的 $m\times n$ *matrix.* 則對於任意 $\mathbf{v}\in\mathbb{R}^m$, $\mathbf{v}$ 在 W 的 *projection* 為

$$\mathrm{Proj}_W(\mathbf{v})=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{v}.$$

Example 7.1.17. 我們要利用 Theorem 7.1.16 的結果處理 Example 7.1.13 的情形, 也就是求 $\mathbf{v}=\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$ 在 $W=\mathrm{Span}(\begin{bmatrix}1\\2\\2\\1\end{bmatrix},\begin{bmatrix}3\\4\\2\\3\end{bmatrix})$ 的投影. 首先考慮矩陣 $A=\begin{bmatrix}1&3\\2&4\\2&2\\1&3\end{bmatrix}$, 此時 $A^{\mathrm{T}}=\begin{bmatrix}1&2&2&1\\3&4&2&3\end{bmatrix}$, 故得 $A^{\mathrm{T}}A=\begin{bmatrix}10&18\\18&38\end{bmatrix}$ 以及其 inverse $(A^{\mathrm{T}}A)^{-1}=(1/28)\begin{bmatrix}19&-9\\-9&5\end{bmatrix}$. 因此由 Theorem 7.1.16 得

$$\mathrm{Proj}_W(\mathbf{v})=\frac{1}{28}\begin{bmatrix}1&3\\2&4\\2&2\\1&3\end{bmatrix}\begin{bmatrix}19&-9\\-9&5\end{bmatrix}\begin{bmatrix}1&2&2&1\\3&4&2&3\end{bmatrix}\begin{bmatrix}2\\0\\1\\4\end{bmatrix}=\begin{bmatrix}2\\2\\0\\2\end{bmatrix}.$$

和 Example 7.1.13 的結果吻合.

對於 Theorem 7.1.16 要注意的是因為 W 為 $\mathbb{R}^m$ 的 subspace, 除非 $W=\mathbb{R}^m$, 否則 $\dim(W)=n$ 會小於 m. 然而當 $W=\mathbb{R}^m$ 時, 談論對 W 的 projection 是沒有意思的, 因為此

時 $W^{\perp} = \{\mathbf{O}\}$, 所以任何 $\mathbb{R}^m$ 的向量對 $W = \mathbb{R}^m$ 的投影就是自己. 因此一般在談論投影時僅考慮 $\dim(W) = n < m$ 的情形. 也因此, 利用 W 的一組 basis 為 column vector 所成的矩陣 A, 是 $m \times n$ matrix 不會是一個方陣. 所以此時 A 和 A^{T} 皆不會是 invertible. 也因此我們不能將 $(A^{\mathrm{T}}A)^{-1}$ 寫成 $A^{-1}(A^{\mathrm{T}})^{-1}$. 也因為這樣 Theorem 7.1.16 中 $A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 絕不能寫成 $A(A^{-1}(A^{\mathrm{T}})^{-1})A^{\mathrm{T}}$, 否則會變成 identity matrix.

Theorem 7.1.16 確實簡化了 Theorem 7.1.12 求 projection 的程序. 我們只要求出 W 的一組 basis 即可, 不必再求 $W^{\perp}$ 的 basis. 由於將矩陣 $A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 乘上任何 $\mathbb{R}^m$ 的向量 $\mathbf{v}$, 就可得 $\mathrm{Proj}_W(\mathbf{v})$. 因此我們將 $A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 用 P_W 來表示, 且稱之為對於 W 的 *projection matrix*. 要注意, 依定義 P_W 事實上就是 $\mathrm{Proj}_W : \mathbb{R}^m \to \mathbb{R}^m$ 這一個 linear transformation 的 standard matrix representation. 由於一個 linear transformation 其 standard matrix representation 是唯一的, 所以雖然 W 選取不同的 basis 為 column vectors 所得的 $m \times n$ matrix A 會不同, 但由這些 A 所得的 projection matrix 都會相同.

Question 7.4. 在 *Example 7.1.17* 中對於 W 的 *projective matrix* 為何?

其實 Theorem 7.1.16 求 projection 的方法還是很繁瑣, 主要是在求 projection matrix 的過程中需求 $(A^{\mathrm{T}}A)^{-1}$ 這一個反矩陣. 將來我們會介紹另一個較簡捷可以幫助我們求 projection 的方法.

7.2. Inconsistent Systems

在這一節中我們將利用 projection 的概念處理聯立方程組無解的情況. 也就是當一個 linear system 為 inconsistent 時, 我們探討如何可找到最佳的可能解. 我們也將利用這樣的概念處理大家高中統計所學二維資料的最適合直線.

內積的性質 (2) 告訴我們除了零向量 $\mathbf{O}$ 以外, 其餘向量 $\mathbf{v} \in V$ 皆需符合 $\langle \mathbf{v}, \mathbf{v} \rangle > 0$, 所以很自然地我們可依此定義向量的長度 $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. 在我們熟悉的 $\mathbb{R}^3$ 空間中. 一個點在一個平面的投影會是該平面上距離該點最近的點. 這個性質對於推廣到一般 inner product space 的 projection 也是對的. 事實上我們有以下的結果

Proposition 7.2.1. *假設 W 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{v} \in \mathbb{R}^m$. 若 $\mathbf{w}$ 為 $\mathbf{v}$ 在 W 的投影, 則對於任意 $\mathbf{w}' \in W$ 且 $\mathbf{w}' \neq \mathbf{w}$, 皆有 $\|\mathbf{v} - \mathbf{w}'\| > \|\mathbf{v} - \mathbf{w}\|$.*

Proof. 考慮 $\mathbf{v} - \mathbf{w}' = \mathbf{v} - \mathbf{w} + \mathbf{w} - \mathbf{w}'$. 因 $\mathbf{w} = \mathrm{Proj}_W(\mathbf{v})$, 故知 $\mathbf{v} - \mathbf{w} \in W^{\perp}$. 又因 W 為 vector space, 我們有 $\mathbf{w} - \mathbf{w}' \in W$. 故得

$$\|\mathbf{v} - \mathbf{w}'\|^2 = \langle \mathbf{v} - \mathbf{w}', \mathbf{v} - \mathbf{w}' \rangle = \langle \mathbf{v} - \mathbf{w} + \mathbf{w} - \mathbf{w}', \mathbf{v} - \mathbf{w} + \mathbf{w} - \mathbf{w}' \rangle = \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle + \langle \mathbf{w} - \mathbf{w}', \mathbf{w} - \mathbf{w}' \rangle.$$

亦即 $\|\mathbf{v} - \mathbf{w}'\|^2 = \|\mathbf{v} - \mathbf{w}\|^2 + \|\mathbf{w} - \mathbf{w}'\|^2$. 又因 $\mathbf{w} \neq \mathbf{w}'$ 我們有 $\|\mathbf{w} - \mathbf{w}'\| > 0$. 得證 $\|\mathbf{v} - \mathbf{w}'\|^2 > \|\mathbf{v} - \mathbf{w}\|^2$, 即 $\|\mathbf{v} - \mathbf{w}'\| > \|\mathbf{v} - \mathbf{w}\|$. □

當 $A \in \mathscr{M}_{m \times n}$, 我們都知道聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有解, 若且唯若 $\mathbf{b} \in C(A)$. 因此, 一般在日常應用中, 當我們要處理的聯立方程組 $A\mathbf{x} = \mathbf{b}$ 是 inconsistent 時, 我們認為很有可能是 $\mathbf{b}$ 發生誤差所導致, 所以會去找 $\mathbf{b}_0$ 為所有使得聯立方程組 $A\mathbf{x} = \mathbf{b}'$ 有解的 $\mathbf{b}'$ 中與 $\mathbf{b}$ 的距離最近

的一個. 這樣的 $\mathbf{b}_0$ 所得的聯立方程組 $A\mathbf{x}=\mathbf{b}_0$ 的解, 我們認為是最佳的可能解. 然而符合 $A\mathbf{x}=\mathbf{b}'$ 有解的 $\mathbf{b}'$ 所成的集合就是 $C(A)$ 這一個 subspace, 所以依 Proposition 7.2.1 我們知道所有的 $\mathbf{b}'$ 中距離 $\mathbf{b}$ 最近的 $\mathbf{b}_0$ 應該就是 $\mathbf{b}$ 在 $C(A)$ 的 projection. 也因此由 projection 的定義知, $\mathbf{b}-\mathbf{b}_0\in C(A)^\perp=N(A^{\mathrm{T}})$. 此即表示 $A^{\mathrm{T}}(\mathbf{b}-\mathbf{b}_0)=\mathbf{O}$, 也就是

$$A^{\mathrm{T}}\mathbf{b}_0=A^{\mathrm{T}}\mathbf{b}. \tag{7.1}$$

由於我們是要解 $A\mathbf{x}=\mathbf{b}_0$ 所以代入上式推得要解

$$A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}. \tag{7.2}$$

注意式子 (7.2) 和式子 (7.1) 的不同點在於, 我們不必求出 $\mathbf{b}_0$ 再解 $A\mathbf{x}=\mathbf{b}_0$, 而是直接解 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$. 不過我們必須說明式子 (7.2) 的解確實是我們希望得到的最佳的可能解.

假設 $\mathbf{x}=\mathbf{x}_0$ 是 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 的一解, 令 $\mathbf{b}_0=A\mathbf{x}_0$ 此即表示 $\mathbf{b}_0\in C(A)$ 且 $\mathbf{b}_0-\mathbf{b}\in N(A^{\mathrm{T}})=C(A)^\perp$. 因此得證 $\mathbf{b}_0$ 就是 $\mathbf{b}$ 在 $C(A)$ 的 projection, 故由 Proposition 7.2.1 知 $\mathbf{b}_0$ 確實是所有使得聯立方程組 $A\mathbf{x}=\mathbf{b}'$ 有解的 $\mathbf{b}'$ 中距離 $\mathbf{b}$ 最近的一個. 因此若 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 有解, 則此解的確會是我們所期望的一個最佳的可能解. 現在我們面臨的問題是式子 (7.2) 一定有解嗎?

Proposition 7.2.2. *假設 $A\in\mathscr{M}_{m\times n}$ 且 $\mathbf{b}\in\mathbb{R}^m$. 聯立方程組 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 一定有解. 特別若 $\operatorname{rank}(A)=n$, 則聯立方程組有解且解唯一.*

Proof. 令 $\mathbf{b}_0$ 為 $\mathbf{b}$ 在 $C(A)$ 的 projection, 亦即 $\mathbf{b}_0\in C(A)$ 且 $\mathbf{b}-\mathbf{b}_0\in C(A)^\perp$. 因 $\mathbf{b}_0\in C(A)$ 故存在 $\mathbf{x}_0\in\mathbb{R}^n$ 使得 $A\mathbf{x}_0=\mathbf{b}_0$. 又 $\mathbf{b}-\mathbf{b}_0\in C(A)^\perp$ 而 $C(A)^\perp=N(A^{\mathrm{T}})$ (Corollary 7.1.8) 故得

$$\mathbf{O}=A^{\mathrm{T}}(\mathbf{b}-\mathbf{b}_0)=A^{\mathrm{T}}\mathbf{b}-A^{\mathrm{T}}\mathbf{b}_0=A^{\mathrm{T}}\mathbf{b}-A^{\mathrm{T}}(A\mathbf{x}_0),$$

故知 $\mathbf{x}_0$ 確實滿足 $A^{\mathrm{T}}A\mathbf{x}_0=A^{\mathrm{T}}\mathbf{b}$, 因此聯立方程組 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 一定有解.

當 $\operatorname{rank}(A)=n$, 由 Lemma 7.1.15 我們知 $A^{\mathrm{T}}A$ 為 invertible, 所以聯立方程組 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 有解且解唯一. $\square$

注意當 $\operatorname{rank}(A)<n$ 時, 聯立方程組 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 雖然有解不過其解因等同於聯立方程 $A\mathbf{x}=\operatorname{Proj}_{C(A)}(\mathbf{b})$ 的解, 故由 A 為 $m\times n$ matrix 以及 Theorem 3.4.6 知其解不唯一 (會有無窮多解). 我們特別有以下的定義.

Definition 7.2.3. 假設 $A\in\mathscr{M}_{m\times n}$ 且 $\mathbf{b}\in\mathbb{R}^m$. 考慮聯立方程組 $A\mathbf{x}=\mathbf{b}$. 我們稱聯立方程組 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 為 $A\mathbf{x}=\mathbf{b}$ 的 *normal equations*. 在 normal equations 的解集合中, 長度最短的解稱為 $A\mathbf{x}=\mathbf{b}$ 的 *least squares solution*.

特別地, 當 $\operatorname{rank}(A)=n$ 時, 聯立方程組 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 的唯一解 $\mathbf{x}=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{b}$, 就是 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution.

這裡要強調的是, 在解 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution, 我們是求 normal equation 的解, 而不是要求 $\mathbf{b}$ 在 $C(A)$ 的 projection. 所以在解出 least squares solution 後, 是不必再在解的左邊乘上 A 的. 另外即使聯立方程組 $A\mathbf{x}=\mathbf{b}$ 為 consistent, 即 $\mathbf{b}\in C(A)$, 此時 $\mathbf{b}$ 在 $C(A)$ 的 projection 就是 $\mathbf{b}$ 本身. 所以此時 $A\mathbf{x}=\mathbf{b}$ 的解和其 normal equations $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$

的解是一致的. 因此在處理 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution 的問題時, 我們可以不必擔心原方程組 $A\mathbf{x}=\mathbf{b}$ 是否有解, 直接求其 normal equations $A\mathbf{x}=\mathbf{b}$ 的解即可.

Example 7.2.4. 考慮聯立方程組 $A\mathbf{x}=\mathbf{b}$ 其中 $A=\begin{bmatrix}1&3\\2&4\\2&2\\1&3\end{bmatrix}$ 且 $\mathbf{b}=\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$, 我們要找到此方程組的 least squares solution. 此時 $A^{\mathrm{T}}A=\begin{bmatrix}10&18\\18&38\end{bmatrix}$ 以及 $A^{\mathrm{T}}\mathbf{b}=\begin{bmatrix}8\\20\end{bmatrix}$, 故得 $A\mathbf{x}=\mathbf{b}$ 的 normal equations 為

$$\begin{array}{rcr}10x_1+18x_2&=&8\\18x_1+38x_2&=&20\end{array}.$$

解得 $\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}-1\\1\end{bmatrix}$ 為 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution. 我們也可由 $\mathrm{rank}(A)=2$, 得 $A^{\mathrm{T}}A$ 為 invertible 且其 inverse 為 $(A^{\mathrm{T}}A)^{-1}=(1/28)\begin{bmatrix}19&-9\\-9&5\end{bmatrix}$. 解得

$$\begin{bmatrix}x_1\\x_2\end{bmatrix}=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{b}=\frac{1}{28}\begin{bmatrix}19&-9\\-9&5\end{bmatrix}\begin{bmatrix}8\\20\end{bmatrix}=\begin{bmatrix}-1\\1\end{bmatrix}.$$

我們可以檢查 $\begin{bmatrix}1&3\\2&4\\2&2\\1&3\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix}=\begin{bmatrix}2\\2\\0\\2\end{bmatrix}$ 確為 $\mathbf{b}$ 在 $C(A)$ 的 projection (參見 Example 7.1.17).

Question 7.5. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $W=\mathrm{Span}(\mathbf{w}_1,\dots,\mathbf{w}_n)$. 令 A 為以 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 為 *column vector* 的 $m\times n$ *matrix*. 試說明對於任意 $\mathbf{v}\in\mathbb{R}^m$ 若 $\mathbf{x}=\mathbf{x}_0\in\mathbb{R}^n$ 為 $A\mathbf{x}=\mathbf{v}$ 的 *normal equations* $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$ 的一組解, 則 $\mathbf{v}$ 在 W 的 *projection* 為 $A\mathbf{x}_0$. 又當 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 不是 *linearly independent* 時, *normal equations* 的解不唯一, 試問這會不會和 $\mathrm{Proj}_W(\mathbf{v})$ 的唯一性相衝突? 又此時對於 W 的 *projection matrix* P_W 是否仍為 $A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$?

有關 least squares solution 的應用, 最常見的就是二維資料的最適合直線. 也就是說當我們有一組二維的資料 $(x_1,y_1),\dots,(x_m,y_m)$, 我們希望找到一條直線 $y=ax+b$, 希望這些點 (x_i,y_i) 盡可能地靠近這條直線. 注意這裡 $x_1,\dots,x_m$ 以及 $y_1,\dots,y_m$ 都是給定的數, 而我們要解的不是 x,y 而是這個直線的斜率 a 以及 y 截距 b. 依聯立方程組的觀點來看, 我們希望找到 a,b 使之符合

$$\begin{array}{rcl}ax_1+b&=&y_1\\ax_2+b&=&y_2\\&\vdots&\\ax_m+b&=&y_m.\end{array}$$

換句話說我們要解聯立方程組 $A\mathbf{x}=\mathbf{b}$, 其中 $A=\begin{bmatrix}x_1&1\\x_2&1\\\vdots&\vdots\\x_m&1\end{bmatrix}$, $\mathbf{x}=\begin{bmatrix}a\\b\end{bmatrix}$, $\mathbf{b}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_m\end{bmatrix}$. 當然了, 這些給定的資料 $(x_1,y_1),\dots,(x_m,y_m)$ 不一定會使得 $A\mathbf{x}=\mathbf{b}$ 有解, 所以我們要求的是 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution, 也就是解 $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$. 注意, 因為一般要分析的二為資料是要探討

x_i, y_i 之間的關係, 所以這些資料中 $x_1,\ldots,x_m$ 是不會全相同的 (否則 $(x_1,y_1),\ldots,(x_m,y_m)$ 會同時落在一鉛直線上, 也沒甚麼好探討的了), 所以這裡 $\mathrm{rank}(A)=2$, 因此由 Proposition 7.2.2 知 normal equations $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 一定有解. 接下來我們要說明的是, 這樣所得的直線其意義為何?

給定 $a,b\in\mathbb{R}$, 對於 $i=1,\ldots,m$, 令 $y_i'=ax_i+b$. 也就是說 (x_i,y_i) 會在直線 $y=ax+b$ 上. 為方便起見令 $\mathbf{b}'=\begin{bmatrix} y_1'\\ y_2'\\ \vdots\\ y_m'\end{bmatrix}$, 此時 $\mathbf{b}'=\begin{bmatrix} x_1 & 1\\ x_2 & 1\\ \vdots & \vdots\\ x_m & 1\end{bmatrix}\begin{bmatrix} a\\ b\end{bmatrix}\in C(A)$. 又令 $\varepsilon_i=y_i'-y_i$, 此即將直線 $y=ax+b$ 代 x_i 所得的 y_i' 與實際資料中的 y_i 之誤差. 我們知道代這些 x_i 後所得的誤差越小越好, 不過由於 ε_i 會有正有負, 怕正負抵銷了影響誤差的判定, 所以我們取平方, 也就是說希望 $\varepsilon_1^2+\cdots+\varepsilon_m^2$ 的值越小越好. 然而依定義 $\varepsilon_1^2+\cdots+\varepsilon_m^2$ 就是 $\mathbf{b}'$ 和 $\mathbf{b}$ 距離的平方, 即 $\|\mathbf{b}-\mathbf{b}'\|^2$. 另一方面, 我們利用 $A\mathbf{x}=\mathbf{b}$ 解出的 least squares solution $\mathbf{x}=\begin{bmatrix} a\\ b\end{bmatrix}$, 這裡的 a,b 就是讓直線 $y=ax+b$ 所估計得的 $\mathbf{b}'$ 會是 $\mathbf{b}$ 在 $C(A)$ 的投影, 也就是說會讓 $\|\mathbf{b}-\mathbf{b}'\|$ 的值最小. 所以符合我們希望誤差 $\varepsilon_1^2+\cdots+\varepsilon_m^2$ 最小的要求. 也因此利用 least squares solution 所得的直線, 我們稱之為 *least squares line* (最適合直線, 或最小平方直線), 在統計學中稱之為 *line of regression* (迴歸直線).

Example 7.2.5. 考慮二維資料 $(-1,0),(1,1),(2,3)$ 我們要找出此資料的 least square line $y=ax+b$.

原先要解的方程組是

$$\begin{aligned} -1a+b &= 0\\ 1a+b &= 1\\ 2a+b &= 3, \end{aligned}$$

其矩陣表示法為 $\begin{bmatrix} -1 & 1\\ 1 & 1\\ 2 & 1\end{bmatrix}\begin{bmatrix} a\\ b\end{bmatrix}=\begin{bmatrix} 0\\ 1\\ 3\end{bmatrix}$. 故其 normal equations 為

$$\begin{bmatrix} -1 & 1 & 2\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix} -1 & 1\\ 1 & 1\\ 2 & 1\end{bmatrix}\begin{bmatrix} a\\ b\end{bmatrix}=\begin{bmatrix} -1 & 1 & 2\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix} 0\\ 1\\ 3\end{bmatrix},$$

即

$$\begin{aligned} 6a+2b &= 7\\ 2a+3b &= 4, \end{aligned}$$

解得 least squares solution 為 $a=13/14, b=5/7$ 故此組資料的 least squares line 為

$$y=\frac{13}{14}x+\frac{5}{7}.$$

接著我們探討幾個和 least squares line 有關的性質.

Proposition 7.2.6. 考慮二維資料 $(x_1,y_1),\ldots,(x_m,y_m)$. 令 $y=ax+b$ 為其 *least squares line* 且對於 $i=1,\ldots,m$ 令 $y_i'=ax_i+b$. 則有以下知性質:

(1) $y_1+\cdots+y_m=y_1'+\cdots+y_m'$.

(2) $x_1y_1+\cdots+x_my_m=x_1y'_1+\cdots+x_my'_m$.

(3) 令 $\overline{x}$ 和 $\overline{y}$ 分別為 $x_1,\ldots,x_m$ 以及 $y_1,\ldots,y_m$ 的平均數, 即 $\overline{x}=(x_1+\cdots+x_m)/m$, $\overline{y}=(y_1+\cdots+y_m)/m$. 則點 $(\overline{x},\overline{y})$ 在直線 $y=ax+b$ 上, 即 $\overline{y}=a\overline{x}+b$.

Proof. 為方便起見令 $\mathbf{b}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_m\end{bmatrix},\mathbf{b}'=\begin{bmatrix}y'_1\\y'_2\\\vdots\\y'_m\end{bmatrix}$ 又令 $A=\begin{bmatrix}x_1&1\\x_2&1\\\vdots&\vdots\\x_m&1\end{bmatrix}$, 則依 $\mathbf{x}=\begin{bmatrix}a\\b\end{bmatrix}$ 是 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution 得 $\mathbf{b}'=A\begin{bmatrix}a\\b\end{bmatrix}$ 為 $\mathbf{b}$ 在 $C(A)$ 的 projection. 即 $\mathbf{b}-\mathbf{b}'\in C(A)^{\perp}=N(A^{\mathrm{T}})$, 故得

$$A^{\mathrm{T}}(\mathbf{b}-\mathbf{b}')=\begin{bmatrix}x_1&\cdots&x_m\\1&\cdots&1\end{bmatrix}\begin{bmatrix}y_1-y'_1\\y_2-y'_2\\\vdots\\y_m-y'_m\end{bmatrix}=\begin{bmatrix}x_1(y_1-y'_1)+\cdots+x_m(y_m-y'_m)\\(y_1-y'_1)+\cdots+(y_m-y'_m)\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}.$$

得證 (1), (2).

令 $\overline{y'}=(y'_1+\cdots+y'_m)/m$, 由 (1) 知 $\overline{y}=\overline{y'}$. 又由於對所有 $i=1,\ldots,m$, 皆有 $y'_i=ax_i+b$, 故知 $\overline{y'}=a\overline{x}+b$. 得證 (3), 即 $\overline{y}=a\overline{x}+b$. □

最後我們說明一下, 對於二維資料我們不只能談論這資料的最適合直線, 也能談論其他的最適合曲線. 例如有些二維資料我們覺得和拋物線有關, 我們就可以試圖找一條最適合的拋物線 $y=ax^2+bx+c$, 然後代入這些二維資料得到以 a,b,c 為變數的聯立方程組, 再找出此方程組的 least squares solution. 這樣所得的拋物線就會是誤差平方和最小的最適合拋物線了. 至於更高次的多項式也可如法炮製, 我們就不詳述了.

7.3. Orthogonal Basis and Gram-Schmidt Process

我們曾經介紹兩種求 $\mathbb{R}^m$ 中的向量 $\mathbf{v}$ 在一個 subspace W 的 projection 的方法. 這兩種方法其實都有一點繁雜. 第一種方法, 即 Theorem 7.1.12, 我們必須先由 W 的 basis 找到 $W^{\perp}$ 的 basis, 再將 $\mathbf{v}$ 寫成這兩組 basis 所形成的 $\mathbb{R}^m$ 的 basis 的 linear combination, 最後再刪除 $W^{\perp}$ 的部分, 就是 $\mathbf{v}$ 在 W 的 projection 了. 這裡最麻煩的是將 $\mathbf{w}$ 寫成這一組 basis 的 linear combination, 一般來說牽涉到解聯立方程組的問題. 另外好不容易寫成 linear combination 後又要將一部分捨去, 有點多餘做虛功的感覺. 另一種方法, 即 Theorem 7.1.16, 我們必須寫下以 W 的一組 basis 為 column vectors 的矩陣 A, 然後利用它得到對 W 投影的 projection matrix. 這裡較麻煩的地方就是, 我們必須求出 $A^{\mathrm{T}}A$ 的 inverse. 上述兩種方法都牽涉到 W 的一組 basis. 不過我們知道不管選哪一組 basis, 所得的結果都會一樣. 是否有可能找到一組特殊的 basis, 可以讓上述兩種方法較繁雜的部分變簡單呢? 在這一節我們便是要回答這一個問題, 並探討如何找到這種特殊的 basis.

假設 W 為 $\mathbb{R}^n$ 的 subspace 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為其一組 basis 滿足 $\langle\mathbf{w}_i,\mathbf{w}_j\rangle=0,\forall i\neq j$. 對於任意 $\mathbf{w}\in W$, 因為 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis, 存在 $c_1,\ldots,c_n\in\mathbb{R}$ 使得 $\mathbf{w}=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n$.

一般來說我們都是利用解聯立方程組的方法找到 $c_1,\ldots,c_n$, 不過這裡由於這些 $\mathbf{w}_i$ 之間兩兩互相垂直, 我們可以利用內積求出 c_i. 事實上對於任意 $i=1,\ldots,n$, 考慮 $\langle \mathbf{w},\mathbf{w}_i\rangle$. 我們有

$$\langle \mathbf{w},\mathbf{w}_i\rangle = \langle c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n,\mathbf{w}_i\rangle = c_1\langle \mathbf{w}_1,\mathbf{w}_i\rangle+\cdots+c_n\langle \mathbf{w}_n,\mathbf{w}_i\rangle = c_i\langle \mathbf{w}_1,\mathbf{w}_i\rangle.$$

因為 $\mathbf{w}_i \neq \mathbf{O}$, 我們有 $\langle \mathbf{w}_i,\mathbf{w}_i\rangle = \|\mathbf{w}_i\|^2 \neq 0$, 故得 $c_i = \langle \mathbf{w},\mathbf{w}_i\rangle/\langle \mathbf{w}_i,\mathbf{w}_i\rangle = \langle \mathbf{w},\mathbf{w}_i\rangle/\|\mathbf{w}_i\|^2$. 特別的, 若 $\|\mathbf{w}_i\|=1$, 則 $c_i = \langle \mathbf{w},\mathbf{w}_i\rangle$. 我們有以下的定理.

Proposition 7.3.1. *假設 W 為 $\mathbb{R}^n$ 的 subspace 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為其一組 basis 滿足 $\langle \mathbf{w}_i,\mathbf{w}_j\rangle = 0, \forall i\neq j$. 則對於任意 $\mathbf{w}\in W$, 我們有*

$$\mathbf{w} = \frac{\langle \mathbf{w},\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1+\cdots+\frac{\langle \mathbf{w},\mathbf{w}_i\rangle}{\|\mathbf{w}_i\|^2}\mathbf{w}_i+\cdots+\frac{\langle \mathbf{w},\mathbf{w}_n\rangle}{\|\mathbf{w}_n\|^2}\mathbf{w}_n.$$

由於這種兩兩互相垂直的 basis, 對於寫下一個向量的 linear combination 相當的方便, 我們有以下的定義.

Definition 7.3.2. 假設 W 是 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis 滿足對於任意 $i\neq j$ 皆有 $\langle \mathbf{w}_i,\mathbf{w}_j\rangle = 0$. 則稱 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *orthogonal basis*. 若又要求 $\langle \mathbf{w}_i,\mathbf{w}_i\rangle = 1$, $\forall i=1,\ldots,n$, 則稱 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *orthonormal basis*.

要注意, 若 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 是 W 的一組 orthogonal basis, 對於所有 $i=1,\ldots,n$, 令 $\mathbf{u}_i = \mathbf{w}_i/\|\mathbf{w}_i\|$, 則 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 就會是 W 的一組 orthonormal basis. 這是因為

$$\langle \mathbf{u}_i,\mathbf{u}_j\rangle = \langle \frac{1}{\|\mathbf{w}_i\|}\mathbf{w}_i, \frac{1}{\|\mathbf{w}_j\|}\mathbf{w}_j\rangle = \frac{1}{\|\mathbf{w}_i\|}\frac{1}{\|\mathbf{w}_j\|}\langle \mathbf{w}_i,\mathbf{w}_j\rangle.$$

因此當 $i\neq j$, 我們有 $\langle \mathbf{u}_i,\mathbf{u}_j\rangle = \langle \mathbf{w}_i,\mathbf{w}_j\rangle = 0$; 而當 $i=j$, 我們有 $\langle \mathbf{u}_i,\mathbf{u}_i\rangle = \langle \mathbf{w}_i,\mathbf{w}_i\rangle/\|\mathbf{w}_i\|^2 = 1$.

由前面已知, 若能找到 W 的一組 orthogonal basis $\mathbf{w}_1,\ldots,\mathbf{w}_n$, 則我們可以很容易的將任意 W 中的向量寫成 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 的線性組合. 這似乎克服了前面所述 Theorem 7.1.12 較複雜的部份. 事實上確實如此, 若能找到 W 的一組 orthogonal basis $\mathbf{w}_1,\ldots,\mathbf{w}_n$, 則我們便可以輕易地得到 $\mathbf{v}\in\mathbb{R}^m$ 在 W 的 projection 了. 這是因為若 $\mathbf{w} = \mathrm{Proj}_W(\mathbf{v})$ 且 $\mathbf{w} = c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n$. 此時由於 $\mathbf{v}-\mathbf{w}\in W^{\perp}$, 對於所有 $i=1,\ldots,n$, 我們有 $\langle \mathbf{v},\mathbf{w}_i\rangle - \langle \mathbf{w},\mathbf{w}_i\rangle = \langle \mathbf{v}-\mathbf{w},\mathbf{w}_i\rangle = 0$. 因此得 $\langle \mathbf{v},\mathbf{w}_i\rangle = \langle \mathbf{w},\mathbf{w}_i\rangle = c_i\langle \mathbf{w}_i,\mathbf{w}_i\rangle$, 亦即 $c_i = \langle \mathbf{v},\mathbf{w}_i\rangle/\langle \mathbf{w}_i,\mathbf{w}_i\rangle$. 所以我們得到一個比 Theorem 7.1.12 更簡捷求 projection 的方法.

Theorem 7.3.3. *假設 W 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 orthogonal basis. 若 $\mathbf{v}\in\mathbb{R}^m$, 則 $\mathbf{v}$ 在 W 的 projection 為*

$$\mathrm{Proj}_W(\mathbf{v}) = \frac{\langle \mathbf{v},\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1+\cdots+\frac{\langle \mathbf{v},\mathbf{w}_n\rangle}{\|\mathbf{w}_n\|^2}\mathbf{w}_n.$$

特別的當 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 orthonormal basis, 則

$$\mathrm{Proj}_W(\mathbf{v}) = \langle \mathbf{v},\mathbf{w}_1\rangle\mathbf{w}_1+\cdots+\langle \mathbf{v},\mathbf{w}_n\rangle\mathbf{w}_n.$$

由 Theorem 7.3.3, 我們知道只要找到 W 的一組 orthogonal basis, 不只不必去解聯立方程組而且不必去求 $W^{\perp}$ 的一組 basis, 確實省去了許多麻煩. 同樣的, 我們也可利用 W 的一

組 orthogonal basis $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 更簡便的得到對於 W 的 projection matrix. 事實上令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 column vectors 的 $m\times n$ matrix, 則因

$$A^{\mathrm{T}}A=\begin{bmatrix} — & \mathbf{w}_1^{\mathrm{T}} & — \\ — & \mathbf{w}_2^{\mathrm{T}} & — \\ & \vdots & \\ — & \mathbf{w}_n^{\mathrm{T}} & — \end{bmatrix}\begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \\ | & | & & | \end{bmatrix}$$

其 (i,j)-th entry 為 $\mathbf{w}_i^{\mathrm{T}}\mathbf{w}_j=\langle\mathbf{w}_i,\mathbf{w}_j\rangle$, 故由 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 orthogonal, 我們得 $A^{\mathrm{T}}A$ 為 diagonal matrix, 也因此很容易求得 $(A^{\mathrm{T}}A)^{-1}$, 亦即

$$A^{\mathrm{T}}A=\begin{bmatrix} \|\mathbf{w}_1\|^2 & & & \\ & \|\mathbf{w}_2\|^2 & & \\ & & \ddots & \\ & & & \|\mathbf{w}_n\|^2 \end{bmatrix},\quad (A^{\mathrm{T}}A)^{-1}=\begin{bmatrix} \frac{1}{\|\mathbf{w}_1\|^2} & & & \\ & \frac{1}{\|\mathbf{w}_2\|^2} & & \\ & & \ddots & \\ & & & \frac{1}{\|\mathbf{w}_n\|^2} \end{bmatrix}.$$

這也克服了利用 Theorem 7.1.16 求 projection 較複雜的部分, 我們有以下的結果.

Theorem 7.3.4. *假設 W 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 orthogonal basis. 則對於任意 $\mathbf{v}\in\mathbb{R}^m$, 皆有 $\mathrm{Proj}_W(\mathbf{v})=P_W\mathbf{v}$, 其中 P_W 為 $m\times m$ matrix*

$$P_W=\frac{1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}}+\frac{1}{\|\mathbf{w}_2\|^2}\mathbf{w}_2\mathbf{w}_2^{\mathrm{T}}+\cdots+\frac{1}{\|\mathbf{w}_n\|^2}\mathbf{w}_n\mathbf{w}_n^{\mathrm{T}}.$$

Proof. 首先我們強調這裏我們將 $\mathbf{w}_i$ 視為 $m\times 1$ matrix, 所以 $\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}}$ 會是 $m\times m$ matrix. 令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 column vectors 的 $m\times n$ matrix. 利用 Theorem 7.1.16 我們知道 $P_W=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. 然而因 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 orthogonal, 由前述結果知

$$A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}=\begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} \frac{1}{\|\mathbf{w}_1\|^2} & & & \\ & \frac{1}{\|\mathbf{w}_2\|^2} & & \\ & & \ddots & \\ & & & \frac{1}{\|\mathbf{w}_n\|^2} \end{bmatrix}\begin{bmatrix} — & \mathbf{w}_1^{\mathrm{T}} & — \\ — & \mathbf{w}_2^{\mathrm{T}} & — \\ & \vdots & \\ — & \mathbf{w}_n^{\mathrm{T}} & — \end{bmatrix} \tag{7.3}$$

因此我們只要檢查 $m\times m$ matrix

$$\frac{1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}}+\frac{1}{\|\mathbf{w}_2\|^2}\mathbf{w}_2\mathbf{w}_2^{\mathrm{T}}+\cdots+\frac{1}{\|\mathbf{w}_n\|^2}\mathbf{w}_n\mathbf{w}_n^{\mathrm{T}} \tag{7.4}$$

確實與 (7.3) 的 $m\times m$ matrix 為相同的矩陣.

對於 $i=1,\ldots,m$, 令 $\mathbf{e}_i\in\mathbb{R}^m$, 為 i-th entry 為 1 其他 entry 為 0 的 vector. 我們知道任何 $m\times m$ matrix 右邊乘上 $\mathbf{e}_i$ 便會得到此 matrix 的 i-th column. 所以若能證明 (7.3) 和 (7.4) 這兩個矩陣的右邊乘上 $\mathbf{e}_i$ 會相等, 便證得 (7.3) 和 (7.4) 為相同的矩陣, 因而得證本定理. 然而 (7.3) 的右邊乘上 $\mathbf{e}_i$ 會等於

$$\begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} \frac{1}{\|\mathbf{w}_1\|^2} & & & \\ & \frac{1}{\|\mathbf{w}_2\|^2} & & \\ & & \ddots & \\ & & & \frac{1}{\|\mathbf{w}_n\|^2} \end{bmatrix}\begin{bmatrix} \langle\mathbf{w}_1,\mathbf{e}_i\rangle \\ \langle\mathbf{w}_2,\mathbf{e}_i\rangle \\ \vdots \\ \langle\mathbf{w}_n,\mathbf{e}_i\rangle \end{bmatrix}=\begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} \frac{\langle\mathbf{w}_1,\mathbf{e}_i\rangle}{\|\mathbf{w}_1\|^2} \\ \frac{\langle\mathbf{w}_2,\mathbf{e}_i\rangle}{\|\mathbf{w}_2\|^2} \\ \vdots \\ \frac{\langle\mathbf{w}_n,\mathbf{e}_i\rangle}{\|\mathbf{w}_n\|^2} \end{bmatrix}$$

此恰等於 (7.4) 的右邊乘上 $\mathbf{e}_i$ 的結果

$$\frac{\langle \mathbf{w}_1, \mathbf{e}_i \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + \frac{\langle \mathbf{w}_2, \mathbf{e}_i \rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2 + \cdots + \frac{\langle \mathbf{w}_n, \mathbf{e}_i \rangle}{\|\mathbf{w}_n\|^2}\mathbf{w}_n.$$

□

Question 7.6. 試由 $\mathrm{Proj}_W : \mathbb{R}^m \to \mathbb{R}^m$ 是 *linear transformation* 以及此 *linear transformation* 的 *standard matrix representation* 就是 P_W, 利用 *Theorem 7.3.3* 證明 *Theorem 7.3.4.*

我們已經知道只要找到 W 的一組 orthogonal basis, 就可以輕易求得 $\mathbf{v} \in \mathbb{R}^m$ 在 W 的 projection. 對一般 $\mathbb{R}^m$ 的 nonzero subspace, 我們將介紹一個方法找到它的一組 orthogonal basis, 也因此證明了對於一般 $\mathbb{R}^m$ 的 nonzero subspace 一定存在 orthogonal basis (以及 orthonormal basis). 這個方法就是所謂的 *Gram-Schmidt process.*

給定 $\mathbb{R}^m$ 的一個 nonzero subspace W, 且假設 $\dim(W) = n$. 首先我們說明若 $\mathbf{w}_1, \ldots, \mathbf{w}_k \in W$ 為非零向量且滿足 $\langle \mathbf{w}_i, \mathbf{w}_j \rangle = 0$, $\forall i \neq j$, 則 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 為 linearly independent. 這是因為若不是 linearly independent 表示存在 $c_1, \ldots, c_k \in \mathbb{R}$ 不全為 0 使得 $c_1\mathbf{w}_1 + \cdots + c_k\mathbf{w}_k = \mathbf{O}$. 然而對任意 $i = 1, \ldots, k$ 由於 $0 = \langle c_1\mathbf{w}_1 + \cdots + c_k\mathbf{w}_k, \mathbf{w}_i \rangle = c_i\|\mathbf{w}_i\|^2$. 也因此由 $\|\mathbf{w}_i\| \neq 0$, 得證 $c_i = 0$. 此和 $c_1, \ldots, c_k$ 不全為 0 的假設相矛盾, 故知 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 為 linearly independent. 因此要找到 W 的一組 orthogonal basis, 我們只要在 W 中找到 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 滿足 $\langle \mathbf{w}_i, \mathbf{w}_j \rangle = 0$, $\forall i \neq j$ 即可, 因為它們是 linearly independent 且 $\dim(W) = n$, 故由 Corollary 4.3.5 知它們是 W 的一組 basis. 接下來我們要說明在 W 中如何找到這樣的一組 nonzero vectors.

首先因 $W \neq \{\mathbf{O}\}$, 故可在 W 中取一 nonzero vector $\mathbf{v}_1$. 為了方便起見我們令 $\mathbf{w}_1 = \mathbf{v}_1$ 且 $W_1 = \mathrm{Span}(\mathbf{v}_1) = \mathrm{Span}(\mathbf{w}_1)$. 若 $\dim(W) = 1$, 則 $W = W_1$ 故 $\mathbf{w}_1$ 就是 W 的一個 orthogonal basis. 而若 $\dim(W) > 1$, 則因 $W_1 \subsetneq W$, 我們可以找到 nonzero vector $\mathbf{v}_2 \in W$ 且 $\mathbf{v}_2 \notin W_1$. 現在我們要利用 $\mathbf{v}_2$, 找到 $\mathbf{w}_2 \in W$ 滿足 $\mathbf{w}_2 \neq \mathbf{O}$ 且 $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0$. 很自然的, 我們會考慮 $\mathbf{w}_2 = \mathbf{v}_2 - \mathrm{Proj}_{W_1}(\mathbf{v}_2)$, 因為此時 $\mathbf{w}_2 \in W_1^\perp$, 而 $W_1 = \mathrm{Span}(\mathbf{w}_1)$, 故當然有 $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0$. 我們也要說明 $\mathbf{w}_2 \neq \mathbf{O}$. 這是因為若 $\mathbf{w}_2 = \mathbf{O}$, 會得到 $\mathbf{v}_2 = \mathrm{Proj}_{W_1}(\mathbf{v}_2) \in W_1$, 此與當初 $\mathbf{v}_2 \notin W_1$ 的假設相矛盾. 另一方面因為 $W = \mathrm{Span}(\mathbf{w}_1)$, 利用 Proposition 1.4.9 我們知 $\mathrm{Proj}_{W_1}(\mathbf{v}_2) = (\langle \mathbf{v}_2, \mathbf{w}_1 \rangle / \|\mathbf{w}_1\|^2)\mathbf{w}_1$, 所以我們知

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1.$$

另外要注意的是 $\mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2) = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$, 這是因為依 $\mathbf{w}_1, \mathbf{w}_2$ 的選取, 我們有 $\mathbf{w}_1, \mathbf{w}_2 \in \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$, 因此 $\mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2) \subseteq \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$. 然而因為 $\mathbf{v}_1, \mathbf{v}_2$ 為 linearly independent 且 $\mathbf{w}_1, \mathbf{w}_2$ 為 linearly independent, 故由 $\dim(\mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2)) = \dim(\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)) = 2$ 得證 $\mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2) = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$. 為了方便起見, 我們令 $W_2 = \mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2) = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$. 現若 $\dim(W) = 2$, 則 $W = W_2$, 故 $\mathbf{w}_1, \mathbf{w}_2$ 為 W 的一組 orthogonal basis. 而若 $\dim(W) > 2$, 則因 $W_2 \subsetneq W$, 我們可以找到 nonzero vector $\mathbf{v}_3 \in W$ 且 $\mathbf{v}_3 \notin W_2$. 現在我們要利用 $\mathbf{v}_3$, 找到 $\mathbf{w}_3 \in W$ 滿足 $\mathbf{w}_3 \neq \mathbf{O}$ 且 $\langle \mathbf{w}_1, \mathbf{w}_3 \rangle = \langle \mathbf{w}_2, \mathbf{w}_3 \rangle = 0$. 同前, 我們考慮 $\mathbf{w}_3 = \mathbf{v}_3 - \mathrm{Proj}_{W_1}(\mathbf{v}_2)$, 因為此時 $\mathbf{w}_3 \in W_2^\perp$, 而 $W_2 = \mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2)$, 故當然有 $\langle \mathbf{w}_1, \mathbf{w}_3 \rangle = \langle \mathbf{w}_2, \mathbf{w}_3 \rangle = \mathbf{O}$. 另外因 $\mathbf{v}_3 \notin W_2$, 同前面的理由我們有 $\mathbf{w}_3 \neq \mathbf{O}$. 另一方面因為 $\mathbf{w}_1, \mathbf{w}_2$ 為 W_2 的 orthogonal basis, 利用 Theorem 7.3.3

我們得
$$\mathbf{w}_3 = \mathbf{v}_3 - \mathrm{Proj}_{W_2}(\mathbf{v}_3) = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{w}_1 \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 - \frac{\langle \mathbf{v}_3, \mathbf{w}_2 \rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2.$$
最後和前面同樣的理由, 我們有 $\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3) = \mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3)$. 這樣一直下去, 我們可以得到 $W_k = \mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k) = \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k)$ 且 $\mathbf{w}_1,\ldots,\mathbf{w}_k$ 是 W_k 的一組 orthogonal basis. 現若 $k = \dim(W) = n$, 則得 $W_k = W$, 所以 $\mathbf{w}_1,\ldots,\mathbf{w}_k$ 是 W 的一組 orthogonal basis. 而若 $k < n$, 則存在 $\mathbf{v}_{k+1} \in W$ 且 $\mathbf{v}_{k+1} \not\in W_k$. 故令
$$\mathbf{w}_{k+1} = \mathbf{v}_{k+1} - \mathrm{Proj}_{W_k}(\mathbf{v}_{k+1}) = \mathbf{v}_{k+1} - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_1 \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_2 \rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_k \rangle}{\|\mathbf{w}_k\|^2}\mathbf{w}_k,$$
則得 $\mathbf{w}_{k+1} \neq \mathbf{O}$ 且 $\langle \mathbf{w}_{k+1}, \mathbf{w}_i \rangle = 0, \forall i = 1,\ldots,k$. 另外因 $\mathbf{w}_1,\ldots,\mathbf{w}_k,\mathbf{w}_{k+1} \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k,\mathbf{v}_{k+1})$, 同上可得 $\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k,\mathbf{w}_{k+1}) = \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k,\mathbf{v}_{k+1})$, 故令 $W_{k+1} = \mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k,\mathbf{w}_{k+1}) = \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k,\mathbf{v}_{k+1})$, 我們有 $\mathbf{w}_1,\ldots,\mathbf{w}_k,\mathbf{w}_{k+1}$ 為 W_{k+1} 的一組 orthogonal basis. 這樣一直下去直到得到 $W_n = W$, 這樣 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 就是 W 的一組 orthogonal basis.

上述 Gram-Schmidt process 中 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的選取事實上和我們過去找 $\mathbb{R}^m$ 的 subspace 的 basis 方法是一樣的. 差別就是我們要將 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 這組 basis 修改成 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 這一組 orthogonal basis. 因此如果一開始以給定 W 的一組 basis $\mathbf{v}_1 \ldots,\mathbf{v}_n$, 我們可以將之直接套用, 因此有以下的結果.

Theorem 7.3.5 (Gram-Schmidt Process)**.** 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 W 的一組 *basis.* 令
$$\mathbf{w}_1 = \mathbf{v}_1, \quad \mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1, \ldots$$
這樣一直下去, 即對於 $i = 1,\ldots,n-1$ 令
$$\mathbf{w}_{i+1} = \mathbf{v}_{i+1} - \frac{\langle \mathbf{v}_{i+1}, \mathbf{w}_1 \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 - \frac{\langle \mathbf{v}_{i+1}, \mathbf{w}_2 \rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{v}_{i+1}, \mathbf{w}_i \rangle}{\|\mathbf{w}_i\|^2}\mathbf{w}_i,$$
則 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *orthogonal basis.* 而且
$$\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_i) = \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_i), \quad \forall i = 1,\ldots,n.$$

Example 7.3.6. 我們要用 orthogonal basis 來處理 projection 的問題. 回顧在 Example 7.1.17 我們要求 $\mathbf{v} = \begin{bmatrix} 2\\0\\1\\4 \end{bmatrix}$ 在 $W = \mathrm{Span}(\begin{bmatrix} 1\\2\\2\\1 \end{bmatrix}, \begin{bmatrix} 3\\4\\2\\3 \end{bmatrix})$ 的投影. 首先找 W 的一組 orthogonal basis. 令
$$\mathbf{w}_1 = \begin{bmatrix} 1\\2\\2\\1 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 3\\4\\2\\3 \end{bmatrix} - \frac{18}{10}\begin{bmatrix} 1\\2\\2\\1 \end{bmatrix} = \frac{2}{5}\begin{bmatrix} 3\\1\\-4\\3 \end{bmatrix}.$$
此時 $\mathbf{w}_1,\mathbf{w}_2$ 為 W 的一組 orthogonal basis, 故利用 Theorem 7.3.3 得
$$\mathrm{Proj}_W(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + \frac{\langle \mathbf{v}, \mathbf{w}_2 \rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2 = \frac{8}{10}\begin{bmatrix} 1\\2\\2\\1 \end{bmatrix} + \frac{28/5}{28/5} \times \frac{2}{5}\begin{bmatrix} 3\\1\\-4\\3 \end{bmatrix} = \begin{bmatrix} 2\\2\\0\\2 \end{bmatrix}.$$

Example 7.3.7. 令 $\mathbf{v}_1=\begin{bmatrix}1\\1\\1\\1\end{bmatrix},\mathbf{v}_2=\begin{bmatrix}3\\1\\-1\\1\end{bmatrix},\mathbf{v}_3=\begin{bmatrix}1\\1\\3\\3\end{bmatrix}$, 我們要求 $W=\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3)$ 的一組 orthogonal basis. 首先令 $\mathbf{w}_1=\mathbf{v}_1$, 得

$$\mathbf{w}_2=\mathbf{v}_2-\frac{\langle\mathbf{v}_2,\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1=\begin{bmatrix}3\\1\\-1\\1\end{bmatrix}-\frac{4}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}=\begin{bmatrix}2\\0\\-2\\0\end{bmatrix}.$$

最後得

$$\mathbf{w}_3=\mathbf{v}_3-\frac{\langle\mathbf{v}_3,\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1-\frac{\langle\mathbf{v}_3,\mathbf{w}_2\rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2=\begin{bmatrix}1\\1\\3\\3\end{bmatrix}-\frac{8}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}-\frac{-4}{8}\begin{bmatrix}2\\0\\-2\\0\end{bmatrix}=\begin{bmatrix}0\\-1\\0\\1\end{bmatrix}.$$

對於每一個 $\mathbf{w}_i$, 我們可以除以其長度 $\|\mathbf{w}_i\|$, 得到一組 orthonormal basis

$$\mathbf{u}_1=\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix},\mathbf{u}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\\0\end{bmatrix},\mathbf{u}_3=\frac{1}{\sqrt{2}}\begin{bmatrix}0\\-1\\0\\1\end{bmatrix}.$$

假設 $A\in\mathscr{M}_{m\times n}$ 且 $\mathrm{rank}(A)=n$, 則 A 的 column vectors 形成 $C(A)$ 的一組 basis. 假設 A 的 column 分別為 $\mathbf{v}_1,\dots,\mathbf{v}_n$, 我們利用 Gram-Schmidt process 得到 $C(A)$ 的一組 orthonormal basis $\mathbf{u}_1,\dots,\mathbf{u}_n$. 由於 $\mathbf{u}_1,\dots,\mathbf{u}_n$ 是 orthonormal basis, 我們有

$$\mathbf{v}_i=\langle\mathbf{v}_i,\mathbf{u}_1\rangle\mathbf{u}_1+\cdots+\langle\mathbf{v}_i,\mathbf{u}_i\rangle\mathbf{u}_i+\cdots+\langle\mathbf{v}_i,\mathbf{u}_n\rangle\mathbf{u}_n.$$

然而在 Gram-Schmidt process, 對於 $i=1,\dots,n$, 我們都有 $\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_i)=\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_i)$, 所以由 $\mathbf{u}_{i+1},\dots,\mathbf{u}_n\in\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_i)^\perp$ 知

$$\langle\mathbf{v}_i,\mathbf{u}_{i+1}\rangle=\langle\mathbf{v}_i,\mathbf{u}_{i+2}\rangle=\cdots=\langle\mathbf{v}_i,\mathbf{u}_n\rangle=0.$$

另外由 $\mathbf{v}_i\not\in\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_{i-1})=\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_{i-1})$, 我們知 $\langle\mathbf{v}_i,\mathbf{u}_i\rangle\neq 0$. 因此若令 Q 為以 $\mathbf{u}_1,\dots,\mathbf{u}_n$ 為 column vectors 的 $m\times n$ matrix, 依矩陣乘法定義我們有 $A=QR$, 其中 R 的 (i,j)-th entry 為 $\langle\mathbf{v}_i,\mathbf{u}_j\rangle$. 故由前面所述, 當 $j>i$ 時 $\langle\mathbf{v}_i,\mathbf{u}_j\rangle=0$, 我們知 R 為 $n\times n$ upper triangular matrix (上三角矩陣), 而且對角線的位置 (i,i)-th entry 為 $\langle\mathbf{v}_i,\mathbf{u}_i\rangle\neq 0$, 我們得 $\mathrm{rank}(R)=n$, 故 R 為 invertible. 這就是所謂 A 的 QR decomposition. 我們再用一個例子來說明.

Example 7.3.8. 考慮 $A=\begin{bmatrix}1&3&1\\1&1&1\\1&-1&3\\1&1&3\end{bmatrix}$ 因 A 的 column vectors 就是 Example 7.3.7 的 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$, 我們直接套用其結果得

$$Q=\begin{bmatrix}\frac{1}{2}&\frac{1}{\sqrt{2}}&0\\\frac{1}{2}&0&-\frac{1}{\sqrt{2}}\\\frac{1}{2}&-\frac{1}{\sqrt{2}}&0\\\frac{1}{2}&0&\frac{1}{\sqrt{2}}\end{bmatrix},\quad R=\begin{bmatrix}\langle\mathbf{v}_1,\mathbf{u}_1\rangle&\langle\mathbf{v}_1,\mathbf{u}_2\rangle&\langle\mathbf{v}_1,\mathbf{u}_3\rangle\\\langle\mathbf{v}_2,\mathbf{u}_1\rangle&\langle\mathbf{v}_2,\mathbf{u}_2\rangle&\langle\mathbf{v}_2,\mathbf{u}_3\rangle\\\langle\mathbf{v}_3,\mathbf{u}_1\rangle&\langle\mathbf{v}_3,\mathbf{u}_2\rangle&\langle\mathbf{v}_3,\mathbf{u}_3\rangle\end{bmatrix}=\begin{bmatrix}2&2&4\\0&2\sqrt{2}&-\sqrt{2}\\0&0&\sqrt{2}\end{bmatrix}.$$

很容易檢查, 我們確有 $A=QR$.

QR decomposition 可以簡化我們處理 least squares solution 的問題. 假設我們要處理聯立方程組 $A\mathbf{x}=\mathbf{b}$, 其中 $A\in\mathscr{M}_{m\times n}$ 且 $\operatorname{rank}(A)=n$. 回顧意下, 要求 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution, 就是要解 normal equations $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$. 現若將 A 寫成 QR decomposition $A=QR$, 我們就是要解 $(QR)^{\mathrm{T}}QR\mathbf{x}=(QR)^{\mathrm{T}}\mathbf{b}$, 利用 transpose 和矩陣乘法性質得 $R^{\mathrm{T}}Q^{\mathrm{T}}QR\mathbf{x}=R^{\mathrm{T}}Q^{\mathrm{T}}\mathbf{b}$. 然而 Q 的 column vectors 是 $C(Q)$ 的 orthonormal basis, 前面提過 $Q^{\mathrm{T}}Q$ 會是 $n\times n$ diagonal matrix, 且其 (i,i)-th entry 為 $\|\mathbf{u}_i\|^2=1$. 也就是說 $Q^{\mathrm{T}}Q$ 是 identity matrix I_n. 因此我們可以將 normal equations 化簡成 $R^{\mathrm{T}}R\mathbf{x}=R^{\mathrm{T}}Q^{\mathrm{T}}\mathbf{b}$. 再利用 R 為 invertible, 所以 R^{T} 也是 invertible, 兩邊乘上 R^{T} 的 inverse, 我們再將 normal equations 化簡為 $R\mathbf{x}=Q^{\mathrm{T}}\mathbf{b}$. 這個聯立方程組就很好解了, 這是因為 R 為 upper triangular matrix 且對角線位置皆不為 0, 所以 R 本身就是 echelon form, 因此我們可以很快求出解.

7.4. 結論

我們介紹了在一個 inner product space 中的 subspace W 其 orthogonal complement $W^{\perp}$ 的概念, 並利用它推廣了過去在 $\mathbb{R}^2$, $\mathbb{R}^3$ 投影的概念到一般的 inner product space. 有了投影的概念, 我們可以對於一個 inconsistent 的聯立方成組提出了最佳解的概念, 即 least square solution. 我們提出了三種得到 projection 的方法, 其中最有效的方法就是利用 Gram-Schmidt process 找到要投影的 subspace 的一組 orthogonal basis (或 orthonormal basis). 找到 subspace 的 orthogonal basis 的另一個好處是讓我們很容易將此 subspace 中任意的向量, 寫成這組 orthogonal basis 的 linear combination.