Theorem 7.4.2 可以用其他的形式表達. 例如利用 Corollary 7.3.14, 我們可以得到

$$C(A) = (C(A)^{\perp})^{\perp} = N(A^{\mathrm{t}})^{\perp}.$$

另一方面, 利用 $(A^{\mathrm{t}})^{\mathrm{t}} = A$, 我們可得

$$C(A^{\mathrm{t}})^{\perp} = N(A), \quad C(A^{\mathrm{t}}) = N(A)^{\perp}.$$

回顧一下, 我們稱 $\dim(C(A))$ 為 A 的 rank, 用 $\mathrm{rank}(A)$ 表示, 而 $\dim(N(A))$ 稱為 A 的 nullity, 用 $\mathrm{nullity}(A)$ 表示. 當計算維度牽涉到 orthogonal complement 時要小心, 因為 $W^{\perp}$ 其實是和將 W 視為哪個 vector space 的 subspace 有關, 因此 Proposition 7.3.13 計算 $W^{\perp}$ 的維度其實是和 W 所在的向量空間 V (即將 W 視為 V 的 subspace) 的維度有關. 當 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$, 我們是將 $C(A)$ 視為 $\mathbb{R}^m$ 的 subspace (因 A 的每個 column vector 是在 $\mathbb{R}^m$ 中), 而將 $N(A)$ 視為 $\mathbb{R}^n$ 的 subspace (因為只有 $\mathbb{R}^n$ 的向量可以乘在 A 的右邊). 所以利用 $C(A^{\mathrm{t}}) = N(A)^{\perp}$, 我們可以得

$$\mathrm{rank}(A) = \dim(C(A^{\mathrm{t}})) = \dim(N(A)^{\perp}) = n - \dim(N(A)) = n - \mathrm{nullity}(A).$$

這與 Dimension Theorem 相吻合.

Lemma 7.4.1 還有許多的應用. 我們有以下的定理.

Proposition 7.4.3. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$. 則 $N(A^{\mathrm{t}}A) = N(A)$.

Proof. 假設 $\mathbf{v} \in N(A)$, 則由 $A\mathbf{v} = \mathbf{0}$ 得 $(A^{\mathrm{t}}A)\mathbf{v} = A^{\mathrm{t}}(A\mathbf{v}) = A^{\mathrm{t}}(\mathbf{0}) = \mathbf{0}$. 即 $\mathbf{v} \in N(A^{\mathrm{t}}A)$, 故得證 $N(A) \subseteq N(A^{\mathrm{t}}A)$. 反之, 若 $\mathbf{v} \in N(A^{\mathrm{t}}A)$, 則由 $(A^{\mathrm{t}}A)\mathbf{v} = \mathbf{0}$ 以及 Lemma 7.4.1, 可得

$$\langle A\mathbf{v}, A\mathbf{v}\rangle = \langle A^{\mathrm{t}}(A\mathbf{v}), \mathbf{v}\rangle = \langle (A^{\mathrm{t}}A)\mathbf{v}, \mathbf{v}\rangle = \langle \mathbf{0}, \mathbf{v}\rangle = 0.$$

故由內積性質得證 $A\mathbf{v} = \mathbf{0}$, 即 $\mathbf{v} \in N(A)$. 因此得 $N(A^{\mathrm{t}}A) \subseteq N(A)$. □

當 A 為 $m \times n$ matrix, 則 $A^{\mathrm{t}}A$ 就會是 $n \times n$ 方陣. 因此由 Dimension Theorem 以及 Proposition 7.4.3, 我們有

$$n - \mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{nullity}(A^{\mathrm{t}}A) = \dim(N(A^{\mathrm{t}}A)) = \dim(N(A)) = \mathrm{nullity}(A) = n - \mathrm{rank}(A)$$

因此得知 $\mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(A)$. 我們有以下的結果.

Corollary 7.4.4. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$. 則 $\mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(A)$. 特別的我們有 $\mathrm{rank}(A) = n$ 若且唯若 $A^{\mathrm{t}}A$ 為 *invertible matrix.*

Proof. 我們已經證得 $\mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(A)$. 由於 $A^{\mathrm{t}}A$ 為 $n \times n$ matrix, $\mathrm{rank}(A^{\mathrm{t}}A) = n$ 等價於 $A^{\mathrm{t}}A$ 為 invertible matrix (Theorem 3.5.2), 也因此由 $\mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(A)$ 知 $\mathrm{rank}(A) = n$ 等價於 $A^{\mathrm{t}}A$ 為 invertible matrix. □

當 W 為 $\mathbb{R}^m$ 的 subspace 時, Theorem 7.4.2 也可以幫助我們求 $\mathbf{v} \in \mathbb{R}^n$ 在 W 上的 orthogonal projection. 要注意在 Theorem 7.3.8 中找 W 的 orthogonal basis 的方法求 orthogonal projection 是適用於一般的 inner product space, 而這裡我們介紹的方法僅適用於 $\mathbb{R}^m$ 且使用 dot product.

假設 $W=\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n)\subseteq\mathbb{R}^m$ 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, Theorem 7.4.2 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 $A^{\mathrm{t}}(\mathbf{v}-A\mathbf{x})=\mathbf{0}$. 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. 我們有以下的定理.

Proposition 7.4.5. 假設 $W=\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n)\subseteq\mathbb{R}^m$ 且考慮 $\mathbb{R}^m$ 的 *dot product.* 對於任意 $\mathbf{v}\in\mathbb{R}^m$, 令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 *column vectors* 的 $m\times n$ *matrix* 且考慮聯立方程組 $(A^{\mathrm{t}}A)\mathbf{x}=A^{\mathrm{t}}\mathbf{v}$. 若 $\mathbf{x}_0$ 為此聯立方程組的一個解, 則 $\mathrm{Proj}_W(\mathbf{v})=A\mathbf{x}_0$.

特別的, 如果 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis, 則 $\mathrm{rank}(A)=\dim(W)=n$. 故利用 Corollary 7.4.4 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.

Corollary 7.4.6. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且考慮 $\mathbb{R}^m$ 的 *dot product.* 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 7.4.7. 我們要利用 Corollary 7.4.6 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 Corollary 7.4.6 得

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

這個結果和我們在 Example 7.3.10 利用 orthogonal basis 處理投影的結果一致.

對於 Corollary 7.4.6 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 $W^{\perp}=\{\mathbf{0}\}$, 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 $A^{-1}(A^{\mathrm{t}})^{-1}$. 因為這原因 Corollary 7.4.6 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 7.4.6 簡化了求 projection 的程序. 我們只要求出 W 的一組 basis 即可, 不必先求 W 的 orthogonal 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 $A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}$ 稱之為對於 W 的 *projection matrix*. 事實上它就是 $\mathrm{Proj}_W:\mathbb{R}^m\to\mathbb{R}^m$ 這一個 linear operator 用 standard ordered basis 所得的 matrix representation. 由於一個 linear transformation 固定其 ordered basis 後其 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 7.11. 在 *Example 7.4.7* 中對於 W 的 *projection matrix* 為何? 用 *Example 7.3.10* 中所得的 W 的 *orthogonal basis* 所得的 *projection matrix* 又是為何?

假設 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$ 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 $\mathbf{v}_1,\ldots,\mathbf{v}_n$, 我們利用 Gram-Schmidt process 得到 $C(A)$ 的一組 orthonormal basis $\mathbf{u}_1,\ldots,\mathbf{u}_n$. 由於 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 是 orthonormal basis, 將 $\mathbf{v}_j$ 寫成 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 的線性組合可得

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

因此若令 Q 為以 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 為 column vectors 的 $m\times n$ matrix, 依矩陣乘法定義我們可以將 A 寫成 QR, 即

$$\begin{bmatrix}| & & | & & |\\ \mathbf{v}_1 & \cdots & \mathbf{v}_j & \cdots & \mathbf{v}_n\\ | & & | & & |\end{bmatrix}=\begin{bmatrix}| & & | & & |\\ \mathbf{u}_1 & \cdots & \mathbf{u}_j & \cdots & \mathbf{u}_n\\ | & & | & & |\end{bmatrix}\begin{bmatrix}\langle\mathbf{v}_1,\mathbf{u}_1\rangle & & \langle\mathbf{v}_j,\mathbf{u}_1\rangle & & \langle\mathbf{v}_n,\mathbf{u}_1\rangle\\ \vdots & \cdots & \vdots & \cdots & \vdots\\ \langle\mathbf{v}_1,\mathbf{u}_n\rangle & & \langle\mathbf{v}_j,\mathbf{u}_n\rangle & & \langle\mathbf{v}_n,\mathbf{u}_n\rangle\end{bmatrix},$$

其中 R 是一個 $n\times n$ matrix 且其 j-th column 為 $\begin{bmatrix}\langle\mathbf{v}_j,\mathbf{u}_1\rangle\\ \vdots\\ \langle\mathbf{v}_j,\mathbf{u}_n\rangle\end{bmatrix}$, 也就是說 R 的 (i,j)-th entry 為 $\langle\mathbf{v}_j,\mathbf{u}_i\rangle=\langle\mathbf{u}_i,\mathbf{v}_j\rangle$. 在 Gram-Schmidt process, 對於 $j=1,\ldots,n$, 我們都有 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_j)=\mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_j)$ 而且 $\mathbf{u}_{j+1},\ldots,\mathbf{u}_n\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_j)^{\perp}$, 故知

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

另外由 $\mathbf{v}_j\not\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{j-1})=\mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_{j-1})$, 我們知 $\langle\mathbf{u}_j,\mathbf{v}_j\rangle\neq0$, 否則會造成 $\mathbf{v}_j=\langle\mathbf{u}_1,\mathbf{v}_j\rangle\mathbf{u}_1+\cdots+\langle\mathbf{u}_{j-1},\mathbf{v}_j\rangle\mathbf{u}_{j-1}\in\mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_{j-1})$ 之矛盾. 故由前面所述, 當 $i>j$ 時 $\langle\mathbf{u}_i,\mathbf{v}_j\rangle=0$, 我們知 R 為 $n\times n$ upper triangular matrix, 而且對角線的位置 (j,j)-th entry 為 $\langle\mathbf{u}_j,\mathbf{v}_j\rangle\neq0$, 我們得 $\mathrm{rank}(R)=n$, 故 R 為 invertible. 這就是所謂 A 的 *QR decomposition*. 我們用一個例子來說明.

Example 7.4.8. 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 7.3.12 的 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$, 我們直接套用其結果得

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

很容易檢查, 我們確有 $A=QR$.

將矩陣 A 寫成 QR decomposition 在許多應用上有其方便性. 特別是探討與 $A^{\mathrm{t}}A$ 有關的問題. 此時我們有 $A^{\mathrm{t}}A=(QR)^{\mathrm{t}}(QR)=(R^{\mathrm{t}}Q^{\mathrm{t}})(QR)=R^{\mathrm{t}}(Q^{\mathrm{t}}Q)R$. 然而 Q 的 column vectors 是 $C(A)=C(Q)$ 的一組 orthonormal basis, 很容易驗證 $Q^{\mathrm{t}}Q$ 會是 $n\times n$ diagonal matrix, 且其 (i,i)-th entry 為 $\langle\mathbf{u}_i,\mathbf{u}_i\rangle=\|\mathbf{u}_i\|^2=1$. 也就是說 $Q^{\mathrm{t}}Q$ 是 identity matrix I_n. 因此我們可得 $A^{\mathrm{t}}A=R^{\mathrm{t}}R$. 將 $A^{\mathrm{t}}A$ 寫成 $R^{\mathrm{t}}R$ 的好處是 R 是一個 upper triangular matrix 且為 invertible (注意 A 未必 invertible). 例如 Corollary 7.4.6 對 $W=C(A)$ 的 projection matrix $A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}$ 就可以寫成

$$A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}=(QR)(R^{\mathrm{t}}R)^{-1}(QR)^{\mathrm{t}}=(QR)(R^{-1}(R^{\mathrm{t}})^{-1})(R^{\mathrm{t}}Q^{\mathrm{t}})=Q(RR^{-1})((R^{\mathrm{t}})^{-1}R^{\mathrm{t}})Q^{\mathrm{t}}=QQ^{\mathrm{t}}$$

這種簡單的形式了. 再次提醒我們知 $Q^{\mathrm{t}}Q$ 是 identity matrix, 但 QQ^{t} 就未必是 identity 了. 下一節中探討解聯立方程組問題時還會看到 QR decomposition 的應用.

7.5. 聯立方程組和內積的連結

在這一節中我們將利用內積的概念處理聯立方程組無解的情況. 我們要探討一個 linear system 無解時或是有解但解不唯一時, 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所學有關統計二維資料的最適合直線.

當 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$, $\mathbf{b}\in\mathbb{R}^m$ 我們知道 $C(A)$ 可以幫助我們判斷聯立方程組 $A\mathbf{x}=\mathbf{b}$ 是否有解, 而 $N(A)$ 可幫助我們判斷 $A\mathbf{x}=\mathbf{b}$ 若有解其解是否唯一. 又 Theorem 7.4.2 告訴我們 $C(A)^{\perp}=N(A^{\mathrm{t}})$, 我們將利用這個關係來探討聯立方程組無解或解不唯一時如何處理問題.

首先我們探討無解的情形. 當 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$, 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 $A\mathbf{x}=\mathbf{b}$ 無解時, 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 7.3.16 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 orthogonal projection. 也因此由 orthogonal projection 的定義以及 Theorem 7.4.2 知, $\mathbf{b}-\mathbf{b}_0\in C(A)^{\perp}=N(A^{\mathrm{t}})$. 此即表示 $A^{\mathrm{t}}(\mathbf{b}-\mathbf{b}_0)=\mathbf{0}$, 也就是

$$A^{\mathrm{t}}\mathbf{b}_0=A^{\mathrm{t}}\mathbf{b}. \tag{7.4}$$

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

$$A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}. \tag{7.5}$$

注意式子 (7.5) 和式子 (7.4) 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 (7.5) 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 $\mathbf{b}-\mathbf{b}_0\in N(A^{\mathrm{t}})=C(A)^{\perp}$. 因此得證 $\mathbf{b}_0$ 就是 $\mathbf{b}$ 在 $C(A)$ 的 orthogonal projection, 故由 Proposition 7.3.16 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 (7.5) 一定有解嗎?

Proposition 7.5.1. *假設 $A\in \mathsf{M}_{m\times n}(\mathbb{R})$ 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. 特別地, 若 $\mathrm{rank}(A)=n$, 則聯立方程組有解且解唯一.*

Proof. 令 $\mathbf{b}_0$ 為 $\mathbf{b}$ 在 $C(A)$ 的 orthogonal 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 $C(A)^{\perp}=N(A^{\mathrm{t}})$ 故得

$$\mathbf{0}=A^{\mathrm{t}}(\mathbf{b}-\mathbf{b}_0)=A^{\mathrm{t}}\mathbf{b}-A^{\mathrm{t}}\mathbf{b}_0=A^{\mathrm{t}}\mathbf{b}-A^{\mathrm{t}}(A\mathbf{x}_0),$$

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 $\mathrm{rank}(A)=n$, 由 Corollary 7.4.4 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. □

注意當 $\mathrm{rank}(A)<n$ 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 $A\mathbf{x}=\mathrm{Proj}_{C(A)}(\mathbf{b})$ 的解, 故由 A 為 $m\times n$ matrix 以及 Theorem 3.4.5 知其解不唯一 (會有無窮多解). 有關於解不唯一的情形, 我們等一下再談.

由 Proposition 7.5.1, 我們特別有以下的定義.

Definition 7.5.2. 假設 $A\in \mathsf{M}_{m\times n}(\mathbb{R})$ 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*.

這裡要強調的是, 當求 $A\mathbf{x}=\mathbf{b}$ 無解時, 我們是求 normal equations 的解, 而不是要求 $\mathbf{b}$ 在 $C(A)$ 的 projection. 所以在解出 normal equation 的後, 是不必再在解的左邊乘上 A 的. 另外即使聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解, 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 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 的解即可.

Example 7.5.3. 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 $\mathbf{b}=\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$, 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

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

解得 $\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}-1\\1\end{bmatrix}$ 為其解. 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 7.4.7).

有關 normal equation 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 $(x_1,y_1),\ldots,(x_m,y_m)$, 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 $x_1,\ldots,x_m$ 以及 $y_1,\ldots,y_m$ 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 $(x_1,y_1),\ldots,(x_m,y_m)$ 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 normal equation 的解, 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維資料是要探討 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 7.5.1 知 normal equations $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 一定有解且解唯一. 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 $\mathbf{b}'=\begin{bmatrix}y'_1\\\vdots\\y'_m\end{bmatrix}$, 此時 $\mathbf{b}'=\begin{bmatrix}x_1&1\\\vdots&\vdots\\x_m&1\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}\in C(A)$. 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 $A\mathbf{x}=\mathbf{b}$ 求出其 normal equation $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 的解 $\mathbf{x}=\begin{bmatrix}a\\b\end{bmatrix}$, 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. 也因此這樣所得的直線, 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 7.5.4. 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 7.5.5. 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.

上一節提過 QR decomposition 可以幫忙處理這類的問題. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathrm{rank}(A) = n$, 對於聯立方程組 $A\mathbf{x} = \mathbf{b}$ 要其解 normal equations $A^{\mathrm{t}}A\mathbf{x} = A^{\mathrm{t}}\mathbf{b}$. 我們可以將 A 寫成 QR decomposition $A = QR$ 來處理. 此時我們要解 $(QR)^{\mathrm{t}}QR\mathbf{x} = (QR)^{\mathrm{t}}\mathbf{b}$, 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 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.

接下來我們探討 linear system $A\mathbf{x} = \mathbf{b}$ 有解, 但解不唯一的情形. 在很多情況當 $A\mathbf{x} = \mathbf{b}$ 解不唯一時, 我們希望能找到長度最小的解. 我們有以下的定義.

Definition 7.5.6. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathbf{b} \in \mathbb{R}^m$. 考慮聯立方程組 $A\mathbf{x} = \mathbf{b}$. 若 $\mathbf{v} \in \mathbb{R}^n$ 滿足 $A\mathbf{v} = \mathbf{b}$ 且對任意 $A\mathbf{x} = \mathbf{b}$ 的一組解 $\mathbf{w}$ 皆滿足 $\|\mathbf{v}\| \leq \|\mathbf{w}\|$, 則稱 $\mathbf{v}$ 為 linear system $A\mathbf{x} = \mathbf{b}$ 的 *minimal solution* (或稱 *least square solution*).