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 4.4.4 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 4.4.7. 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 4.4.8. 我們要利用 Corollary 4.4.7 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 4.4.7 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 4.3.10 利用 orthogonal basis 處理投影的結果一致。

對於 Corollary 4.4.7 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 4.4.7 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。

Corollary 4.4.7 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.*

Question 4.8. 在 *Example 4.4.8* 中對於 W 的 *projection matrix* 為何？用 *Example 4.3.10* 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 $\mathrm{Col}(A)$ 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 $\mathrm{Col}(A)$ 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 $\mathbf{v}_j\notin\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{j-1})=\mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_{j-1})$，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 4.4.9. 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 4.3.12 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 $\mathrm{Col}(A)$ 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 4.4.7 對 $W=\mathrm{Col}(A)$ 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^tQ 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。

4.5. 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 $\mathrm{Col}(A)$ 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 4.4.2 告訴我們 $\mathrm{Col}(A)^\perp = N(A^t)$，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。

4.5.1. 無解的情形. 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 $\mathbf{b} \in \mathrm{Col}(A)$。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，並稱 $\mathbf{b}_0 - \mathbf{b}$ 為 $\mathbf{b}$ 的 *error vector*。然而符合 $A\mathbf{x} = \mathbf{b}'$ 有解的 $\mathbf{b}'$ 所成的集合就是 $\mathrm{Col}(A)$ 這一個 subspace，所以依 Proposition 4.3.15 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 $\mathrm{Col}(A)$ 的 orthogonal projection。也因此由 orthogonal projection 的定義以及 Theorem 4.4.2 知：$\mathbf{b} - \mathbf{b}_0 \in \mathrm{Col}(A)^\perp = N(A^t)$。此即表示 $A^t(\mathbf{b} - \mathbf{b}_0) = \mathbf{0}$，也就是

$$A^t\mathbf{b}_0 = A^t\mathbf{b}. \tag{4.4}$$

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

$$(A^tA)\mathbf{x} = A^t\mathbf{b}. \tag{4.5}$$

注意式子 (4.5) 和式子 (4.4) 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 $A^tA\mathbf{x} = A^t\mathbf{b}$。不過我們必須說明式子 (4.5) 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 $(A^tA)\mathbf{x} = A^t\mathbf{b}$ 的一解，令 $\mathbf{b}_0 = A\mathbf{x}_0$。因

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

此即 $\mathbf{b} - \mathbf{b}_0 \in N(A^t) = \mathrm{Col}(A)^\perp$，又因 $\mathbf{b}_0 = A\mathbf{x}_0 \in \mathrm{Col}(A)$，故得證 $\mathbf{b}_0$ 就是 $\mathbf{b}$ 在 $\mathrm{Col}(A)$ 的 orthogonal projection。因此由 Proposition 4.3.15 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。所以若 $A^tA\mathbf{x} = A^t\mathbf{b}$ 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 (4.5) 一定有解嗎？

Proposition 4.5.1. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathbf{b} \in \mathbb{R}^m$。聯立方程組 $A^tA\mathbf{x} = A^t\mathbf{b}$ 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. 由 Proposition 3.7.2 我們知聯立方程組 $A^tA\mathbf{x} = A^t\mathbf{b}$ 有解，等同於 $A^t\mathbf{b} \in \mathrm{Col}(A^tA)$。然而 Corollary 4.4.5 告訴我們 $\mathrm{Col}(A^tA) = \mathrm{Col}(A^t)$，故由 $A^t\mathbf{b} \in \mathrm{Col}(A^t)$ 知 $A^t\mathbf{b} \in \mathrm{Col}(A^tA)$，因此聯立方程組 $A^tA\mathbf{x} = A^t\mathbf{b}$ 一定有解。

當 $\operatorname{rank}(A)=n$，由 Corollary 4.4.4 我們知 $A^{t}A$ 為 invertible，所以聯立方程組 $A^{t}A\mathbf{x}=A^{t}\mathbf{b}$ 有解且解唯一。 □

由 Proposition 4.5.1 我們特別有以下的定義。

Definition 4.5.2. 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 $A^{t}A\mathbf{x}=A^{t}\mathbf{b}$ 為 $A\mathbf{x}=\mathbf{b}$ 的 *normal equation*，而 normal equation 的解稱為原方程組的 *least squares solution*。

注意當 $\operatorname{rank}(A)<n$ 時，normal equation 雖然有解不過其解因等同於聯立方程 $A\mathbf{x}=\operatorname{Proj}_{\operatorname{Col}(A)}(\mathbf{b})$ 的解，故由 A 為 $m\times n$ matrix 以及 Theorem 2.5.5 知其解不唯一（會有無窮多解）。有關於解不唯一的情形，通常我們又會去找長度最小的解（稱為 *minimal least squares solution*）。這牽涉到有無窮多解時如何找到最佳解的問題，會在下一段再談，目前只探討唯一解的情況。

另外要強調的是：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 $\operatorname{Col}(A)$ 的 projection。所以在解出 normal equation 的後，不必再在解的左邊乘上 A。另外即使聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解，即 $\mathbf{b}\in\operatorname{Col}(A)$。此時 $\mathbf{b}$ 在 $\operatorname{Col}(A)$ 的 projection 就是 $\mathbf{b}$ 本身，所以此時 $A\mathbf{x}=\mathbf{b}$ 的解和其 normal equations $A^{t}A\mathbf{x}=A^{t}\mathbf{b}$ 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 $A^{t}A\mathbf{x}=A^{t}\mathbf{b}$ 的解即可。

Example 4.5.3. 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 $A^{t}A=\begin{bmatrix}10&18\\18&38\end{bmatrix}$ 以及 $A^{t}\mathbf{b}=\begin{bmatrix}8\\20\end{bmatrix}$，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 $\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}-1\\1\end{bmatrix}$ 為其解。我們也可由 $\operatorname{rank}(A)=2$，得 $A^{t}A$ 為 invertible 且其 inverse 為 $(A^{t}A)^{-1}=(1/28)\begin{bmatrix}19&-9\\-9&5\end{bmatrix}$ 解得

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

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 $\operatorname{Col}(A)$ 的 projection（參見 Example 4.4.8）。

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 $(x_1,y_1),\ldots,(x_n,y_n)$，我們希望找到一條直線 $y=mx+c$，使得這些點 (x_i,y_i) 盡可能地靠近這條直線。注意這裡 $x_1,\ldots,x_n$ 以及 $y_1,\ldots,y_n$ 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 m 以及 y 截距 c。依聯立方程組的觀點來看，我們希望找到 m,c 使之符合

$$\begin{array}{rcl} mx_1+c &=& y_1 \\ mx_2+c &=& y_2 \\ &\vdots& \\ mx_n+c &=& y_n. \end{array}$$

換句話說我們要解聯立方程組 $A\mathbf{x}=\mathbf{b}$，其中 $A=\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}$, $\mathbf{x}=\begin{bmatrix} m \\ c \end{bmatrix}$, $\mathbf{b}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$。當然了，這些給定的資料 $(x_1,y_1),\dots,(x_n,y_n)$ 只有在同一直線時，才會使得 $A\mathbf{x}=\mathbf{b}$ 有解。但一般情形並不會如此，所以我們要求的是 $A\mathbf{x}=\mathbf{b}$ 的 normal equation 的解，也就是解 $A^tA\mathbf{x}=A^t\mathbf{b}$。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 $x_1,\dots,x_n$ 是不會全相同的（否則 $(x_1,y_1),\dots,(x_n,y_n)$ 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 4.5.1 知 normal equations $A^tA\mathbf{x}=A^t\mathbf{b}$ 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 $m,c\in\mathbb{R}$，對於 $i=1,\dots,n$，令 $y_i'=mx_i+c$。也就是說 (x_i,y_i') 會在直線 $y=mx+c$ 上。為方便起見令 $\mathbf{b}'=\begin{bmatrix} y_1' \\ \vdots \\ y_n' \end{bmatrix}$，此時 $\mathbf{b}'=\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}\begin{bmatrix} m \\ c \end{bmatrix}\in\mathrm{Col}(A)$。又令 $\varepsilon_i=y_i'-y_i$，此即將直線 $y=mx+c$ 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 ε_i 可能會有正有負。因為正負會抵銷而影響誤差的判定，所以我們取平方，也就是說希望 $\varepsilon_1^2+\cdots+\varepsilon_n^2$ 的值越小越好。然而依定義 $\varepsilon_1^2+\cdots+\varepsilon_n^2$ 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 $A^tA\mathbf{x}=A^t\mathbf{b}$ 的解 $\mathbf{x}=\begin{bmatrix} m \\ c \end{bmatrix}$，這裡的 m,c 就是讓直線 $y=mx+c$ 所估計得的 $\mathbf{b}'$ 會是 $\mathbf{b}$ 在 $\mathrm{Col}(A)$ 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 $\varepsilon_1^2+\cdots+\varepsilon_n^2$ 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 4.5.4. 考慮二維資料 $(-1,0),(1,1),(2,3)$ 我們要找出此資料的 least square line $y=mx+c$。

原先要解的方程組是

$$\begin{array}{rcl} -1m+c &=& 0 \\ 1m+c &=& 1 \\ 2m+c &=& 3, \end{array}$$

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

$$\begin{bmatrix} -1 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} m \\ c \end{bmatrix}=\begin{bmatrix} -1 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix},$$

即

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

解得 least squares solution 為 $m=13/14, c=5/7$ 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 4.5.5. 考慮二維資料 $(x_1,y_1),\ldots,(x_n,y_n)$。令 $y=mx+c$ 為其 *least squares line* 且對於 $i=1,\ldots,n$ 令 $y'_i=mx_i+c$。則有以下知性質：

(1) $y_1+\cdots+y_n=y'_1+\cdots+y'_n$。

(2) $x_1y_1+\cdots+x_ny_n=x_1y'_1+\cdots+x_ny'_n$。

(3) 令 $\overline{x}$ 和 $\overline{y}$ 分別為 $x_1,\ldots,x_n$ 以及 $y_1,\ldots,y_n$ 的平均數，即 $\overline{x}=(x_1+\cdots+x_n)/n$、$\overline{y}=(y_1+\cdots+y_n)/n$。則點 $(\overline{x},\overline{y})$ 在直線 $y=mx+c$ 上，即 $\overline{y}=m\overline{x}+c$。

Proof. 為方便起見令 $\mathbf{b}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{b}'=\begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{bmatrix}$，則依 $\mathbf{b}-\mathbf{b}'\in \mathrm{Col}(A)^{\perp}=N(A^{\mathrm{t}})$，得

$$A^{\mathrm{t}}(\mathbf{b}-\mathbf{b}')=\begin{bmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} y_1-y'_1 \\ y_2-y'_2 \\ \vdots \\ y_n-y'_n \end{bmatrix}=\begin{bmatrix} x_1(y_1-y'_1)+\cdots+x_n(y_n-y'_n) \\ (y_1-y'_1)+\cdots+(y_n-y'_n) \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

得證 (1)、(2)。又因 $\overline{y'}=(y'_1+\cdots+y'_n)/n$，由 (1) 知 $\overline{y}=\overline{y'}$。又由於對所有 $i=1,\ldots,n$，皆有 $y'_i=mx_i+c$，故知 $\overline{y'}=m\overline{x}+c$。得證 (3)，即 $\overline{y}=m\overline{x}+c$。 □

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。

4.5.2. 無窮多解的情形. 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 4.5.6. 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*。

如何找到 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution 呢？假設 $\mathbf{v},\mathbf{w}$ 皆為 $A\mathbf{x}=\mathbf{b}$ 的一組解，則由 $A\mathbf{v}=\mathbf{b}$ 以及 $A\mathbf{w}=\mathbf{b}$ 可得 $A(\mathbf{v}-\mathbf{w})=\mathbf{0}$。因此若令 $\mathbf{u}=\mathbf{v}-\mathbf{w}$，則得 $\mathbf{u}\in N(A)$ 且 $\mathbf{w}=\mathbf{v}-\mathbf{u}$。反之，當我們已找到 $\mathbf{v}$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解時，對任意 $\mathbf{u}\in N(A)$，皆有 $A(\mathbf{v}-\mathbf{u})=A\mathbf{v}-A\mathbf{u}=\mathbf{b}-\mathbf{0}=\mathbf{b}$，亦即 $\mathbf{v}-\mathbf{u}$ 也是 $A\mathbf{x}=\mathbf{b}$ 的一組解。因此聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的解集合可以寫成 $\{\mathbf{v}-\mathbf{u}\mid\mathbf{u}\in N(A)\}$，其中 $\mathbf{v}$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解。也就是說當我們已找到 $A\mathbf{x}=\mathbf{b}$ 的一組解 $\mathbf{v}$，我們要找到 $\mathbf{u}\in N(A)$ 使得 $\|\mathbf{v}-\mathbf{u}\|\leq\|\mathbf{v}-\mathbf{u}'\|, \forall\mathbf{u}'\in N(A)$。此時 $\mathbf{v}-\mathbf{u}$ 就會是 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution。如何在 $N(A)$ 中找到 $\mathbf{u}$ 呢？由於 $N(A)$ 是 $\mathbb{R}^n$ 的 subspace，Proposition 4.3.15 告訴我們 $\mathbf{u}$ 就是 $\mathbf{v}$ 在 $N(A)$ 的 orthogonal projection。也因此可得 $\mathbf{v}-\mathbf{u}\in N(A)^\perp$。再由 Theorem 4.4.2 知 $N(A)^\perp=\mathrm{Col}(A^{\mathrm{t}})$，故知此 minimal solution $\mathbf{v}-\mathbf{u}\in\mathrm{Col}(A^{\mathrm{t}})$。這表示 $\mathbf{v}-\mathbf{u}$ 會在 $n\times m$ 矩陣 A^{t} 的 column space，因此存在 $\mathbf{x}'\in\mathbb{R}^m$ 使得 $\mathbf{v}-\mathbf{u}=A^{\mathrm{t}}\mathbf{x}'$。此時由 $A(\mathbf{v}-\mathbf{u})=A\mathbf{v}-A\mathbf{u}=\mathbf{b}$，得 $(AA^{\mathrm{t}})\mathbf{x}'=A(A^{\mathrm{t}}\mathbf{x}')=A(\mathbf{v}-\mathbf{u})=\mathbf{b}$。所以我們只要解聯立方程組 $(AA^{\mathrm{t}})\mathbf{x}'=\mathbf{b}$。找出的解 $\mathbf{x}'\in\mathbb{R}^m$ 由於會使得 $A^{\mathrm{t}}\mathbf{x}'\in\mathrm{Col}(A^{\mathrm{t}})$ 且符合方程組 $A(A^{\mathrm{t}}\mathbf{x}')=\mathbf{b}$，因此令 $\mathbf{x}=A^{\mathrm{t}}\mathbf{x}'$ 就會是 linear system $A\mathbf{x}=\mathbf{b}$ 的 minimal solution。現在問題是 linear system $(AA^{\mathrm{t}})\mathbf{x}'=\mathbf{b}$ 會有解嗎？

Proposition 4.5.7. 假設 $A\in \mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathbf{b}\in\mathbb{R}^m$。假設聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解，則聯立方程組 $(AA^{\mathrm{t}})\mathbf{x}'=\mathbf{b}$ 必有解。且若 $\mathbf{x}'=\mathbf{w}\in\mathbb{R}^m$ 為其一組解，則 $\mathbf{x}=A^{\mathrm{t}}\mathbf{w}$ 會是聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的 *minimal solution*。

Proof. 我們知道聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解表示 $\mathbf{b}\in\mathrm{Col}(A)$。現要證明聯立方程組 $(AA^{\mathrm{t}})\mathbf{x}'=\mathbf{b}$ 有解，我們只要證明 $\mathbf{b}\in\mathrm{Col}(AA^{\mathrm{t}})$ 即可。回顧在 Corollary 4.4.5，我們知道對任意矩陣 A 皆有 $\mathrm{Col}(A^{\mathrm{t}}A)=\mathrm{Col}(A^{\mathrm{t}})$，所以將 A 用 A^{t} 取代，自然得 $\mathrm{Col}(AA^{\mathrm{t}})=\mathrm{Col}(A)$。因此 $\mathbf{b}\in\mathrm{Col}(AA^{\mathrm{t}})$，得證聯立方程組 $(AA^{\mathrm{t}})\mathbf{x}'=\mathbf{b}$ 有解。現若 $\mathbf{w}\in\mathbb{R}^m$ 是 $(AA^{\mathrm{t}})\mathbf{x}'=\mathbf{b}$ 的一解，即 $(AA^{\mathrm{t}})\mathbf{w}=\mathbf{b}$，則由 $(AA^{\mathrm{t}})\mathbf{w}=A(A^{\mathrm{t}}\mathbf{w})$ 知 $A^{\mathrm{t}}\mathbf{w}$ 為 $A\mathbf{x}=\mathbf{b}$ 的一解。由前面的討論我們知道既然 $A^{\mathrm{t}}\mathbf{w}$ 為 $A\mathbf{x}=\mathbf{b}$ 的一解，則聯立方程組 $A\mathbf{x}=\mathbf{b}$ 所有的解應為 $A^{\mathrm{t}}\mathbf{w}+\mathbf{u}$，其中 $\mathbf{u}\in N(A)$。此時

$$\|A^{\mathrm{t}}\mathbf{w}+\mathbf{u}\|^2=\langle A^{\mathrm{t}}\mathbf{w}+\mathbf{u},A^{\mathrm{t}}\mathbf{w}+\mathbf{u}\rangle=\langle A^{\mathrm{t}}\mathbf{w},A^{\mathrm{t}}\mathbf{w}\rangle+2\langle A^{\mathrm{t}}\mathbf{w},\mathbf{u}\rangle+\langle\mathbf{u},\mathbf{u}\rangle=\|A^{\mathrm{t}}\mathbf{w}\|^2+\|\mathbf{u}\|^2.$$

這裡 $\langle A^{\mathrm{t}}\mathbf{w},\mathbf{u}\rangle=0$，因為 $A^{\mathrm{t}}\mathbf{w}\in\mathrm{Col}(A^{\mathrm{t}})=N(A)^\perp$。由此可知當 $\mathbf{u}\in N(A)$ 且 $\mathbf{u}\neq\mathbf{0}$ 時 $\|A^{\mathrm{t}}\mathbf{w}\|<\|A^{\mathrm{t}}\mathbf{w}+\mathbf{u}\|$，故 $\mathbf{x}=A^{\mathrm{t}}\mathbf{w}$ 會是聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution。 □

在 Proposition 4.5.7 中因為 A 為 $m\times n$ matrix，所以聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的解是在 $\mathbb{R}^n$ 中。而求其 minimal solution 我們要解的聯立方程組為 $AA^{\mathrm{t}}\mathbf{x}'=\mathbf{b}$ 其解是在 $\mathbb{R}^m$ 中，所以要得到 minimal solution 必須將 $AA^{\mathrm{t}}\mathbf{x}'=\mathbf{b}$ 的解左邊再乘上 A^{t}，因此知 minimal solution 會在 A^{t} 的 column space，也就是 A 的 row space。

Corollary 4.5.8. 假設 $A \in \mathrm{M}_{m\times n}(\mathbb{R})$、$\mathbf{b} \in \mathbb{R}^m$ 且聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有解。則在 A 的 *row space* 中存在唯一的向量 $\mathbf{v}$ 滿足 $A\mathbf{v} = \mathbf{b}$，且 $\mathbf{v}$ 為 $A\mathbf{x} = \mathbf{b}$ 的 *minimal solution*。

Proof. 我們已知 $A\mathbf{x} = \mathbf{b}$ 的 minimal solution 在 $\mathrm{Col}(A^{\mathrm{t}})$ 中，亦即在 A 的 row space 中。因此僅剩下要證明唯一性。亦即假設 $\mathbf{v}, \mathbf{v}' \in \mathrm{Col}(A^{\mathrm{t}})$ 且 $A\mathbf{v} = \mathbf{b}$、$A\mathbf{v}' = \mathbf{b}$，我們要證明 $\mathbf{v} = \mathbf{v}'$。由 $A\mathbf{v} = A\mathbf{v}' = \mathbf{b}$ 可得 $A(\mathbf{v} - \mathbf{v}') = \mathbf{0}$，亦即 $\mathbf{v} - \mathbf{v}' \in N(A)$。又 $\mathrm{Col}(A^{\mathrm{t}})$ 為 $\mathbb{R}^n$ 的 subspace，故 $\mathbf{v} - \mathbf{v}' \in \mathrm{Col}(A^{\mathrm{t}})$。因此得 $\mathbf{v} - \mathbf{v}' \in N(A) \cap \mathrm{Col}(A^{\mathrm{t}})$。然而 $\mathrm{Col}(A^{\mathrm{t}}) = N(A)^{\perp}$（Theorem 4.4.2）故得 $N(A) \cap \mathrm{Col}(A^{\mathrm{t}}) = N(A) \cap N(A)^{\perp} = \{\mathbf{0}\}$（Lemma 4.3.4）。因此得證 $\mathbf{v} = \mathbf{v}'$。 □

Example 4.5.9. 我們用一個簡單的例子說明 Proposition 4.5.7 以及 Corollary 4.5.8。考慮 $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix}$、$\mathbf{b} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}$。很容易看出此 linear system $A\mathbf{x} = \mathbf{b}$ 有一組解 $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$ 由於 $\mathrm{rank}(A) = 2$。我們知到此方程組有無窮多解。事實上此 linear system $A\mathbf{x} = \mathbf{b}$ 有另一組解 $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$。我們有 $\|\mathbf{v}_2\| = \sqrt{3} < \|\mathbf{v}_1\| = 2$。還有沒有長度更小的解呢？我們要找到其 minimal solution。由於 $AA^{\mathrm{t}} = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 6 & 3 \\ 1 & 3 & 2 \end{bmatrix}$，我們解 $(AA^{\mathrm{t}})\mathbf{x}' = \mathbf{b}$，即 $\begin{bmatrix} 2 & 3 & 1 \\ 3 & 6 & 3 \\ 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}$ 得 $\mathbf{x}' = \begin{bmatrix} 2 \\ -2/3 \\ 0 \end{bmatrix}$ 為其一組解。故知 $\mathbf{x} = A^{\mathrm{t}}\mathbf{x}' = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -2/3 \\ 0 \end{bmatrix} = \begin{bmatrix} 4/3 \\ -2/3 \\ 2/3 \end{bmatrix}$ 為原方程組 $A\mathbf{x} = \mathbf{b}$ 的 minimal solution。事實上 $N(A) = \mathrm{Span}(\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix})$。所以 $A\mathbf{x} = \mathbf{b}$ 所有的解為 $\mathbf{v} + r\mathbf{u}$ 的形式，其中 $\mathbf{v} = \begin{bmatrix} 4/3 \\ -2/3 \\ 2/3 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$, $r \in \mathbb{R}$。由於 $\langle \mathbf{v}, \mathbf{u} \rangle = 0$，我們有 $\|\mathbf{v} + r\mathbf{u}\|^2 = \langle \mathbf{v} + r\mathbf{u}, \mathbf{v} + r\mathbf{u} \rangle = \|\mathbf{v}\|^2 + r^2\|\mathbf{u}\|^2 = (24/9) + 3r^2$。知 $\mathbf{v}$ 確為 linear system $A\mathbf{x} = \mathbf{b}$ 之 minimal solution。

4.5.3. Minimal Least Squares Solution. Minimal least squares solution，指的是當 $A\mathbf{x} = \mathbf{b}$ 無解時我們考慮 normal equation $A^{\mathrm{t}}A\mathbf{x} = A^{\mathrm{t}}\mathbf{b}$。此時一定有解，且所得的解稱為 $A\mathbf{x} = \mathbf{b}$ 的一個 least squares solution。不過 $A^{\mathrm{t}}A\mathbf{x} = A^{\mathrm{t}}\mathbf{b}$ 有可能有無窮多解，所以我們必須找到 $A^{\mathrm{t}}A\mathbf{x} = A^{\mathrm{t}}\mathbf{b}$ 的 minimal solution。利用前述找 minimal solution 的方法，我們要先去解 $(A^{\mathrm{t}}A)(A^{\mathrm{t}}A)^{\mathrm{t}}\mathbf{y} = A^{\mathrm{t}}\mathbf{b}$。因 $(A^{\mathrm{t}}A)^{\mathrm{t}} = A^{\mathrm{t}}A$，所以我們要去解 $(A^{\mathrm{t}}A)^2\mathbf{y} = A^{\mathrm{t}}\mathbf{b}$，得到解 $\mathbf{y}_0$ 後再求 $(A^{\mathrm{t}}A)\mathbf{y}_0$，就得到 minimal least squares solution 了。

不過方程式 $(A^{\mathrm{t}}A)^2\mathbf{y} = A^{\mathrm{t}}\mathbf{b}$ 從原方程式 $A\mathbf{x} = \mathbf{b}$ 的角度來看頗複雜，有什麼較簡單形式的方程式可以取代嗎？課堂上我們討論了 QR decomposition，然而它僅適用於有唯一解的情況，在找 minimal solution 這種有無窮多解得情況根本派不上用場。

首先觀察 $(A^{\mathrm{t}}A)^2\mathbf{y} = A^{\mathrm{t}}\mathbf{b}$ 可寫成 $(A^{\mathrm{t}}AA^{\mathrm{t}})(A\mathbf{y}) = A^{\mathrm{t}}\mathbf{b}$。所以若 $\mathbf{y}_0$ 為一解，那當然 $\mathbf{z}_0 = A\mathbf{y}_0$ 會是方程組 $(A^{\mathrm{t}}AA^{\mathrm{t}})\mathbf{z} = A^{\mathrm{t}}\mathbf{b}$ 的一解，且 $A^{\mathrm{t}}\mathbf{z}_0 = (A^{\mathrm{t}}A)\mathbf{y}_0 = \mathbf{x}_0$ 就會是 minimal least squares solution。也因此我們只需處理聯立方程組 $(A^tAA^t)\mathbf{z} = A^t\mathbf{b}$，求出一解 $\mathbf{z}_0$ 後再求 $A^t\mathbf{z}_0$ 就是 normal equation $(A^tA)\mathbf{x} = A^t\mathbf{b}$ 的 minimal solution。

另一個更直接的看法是 normal equation $(A^tA)\mathbf{x} = A^t\mathbf{b}$ 一定有解，而其 minimal solution 就是所有解中唯一在 $\mathrm{Row}(A^tA)$ 中的解 $\mathbf{x}_0$（Corollary 4.5.8）。然而 $\mathrm{Row}(A^tA) = \mathrm{Col}(A^tA) = \mathrm{Col}(A^t) = \mathrm{Row}(A)$，所以必存在 $\mathbf{z}_0$ 使得 $A^t\mathbf{z}_0 = \mathbf{x}_0$，因此我們只要解

$$(A^tAA^t)\mathbf{z} = A^t\mathbf{b}$$

的一解 $\mathbf{z}_0$，便可得 $\mathbf{x}_0 = A^t\mathbf{z}_0$ 是 normal equation $(A^tA)\mathbf{x} = A^t\mathbf{b}$ 的 minimal solution。也因此 $A^t\mathbf{z}_0$ 會是原方程組 $A\mathbf{x} = \mathbf{b}$ 的 minimal least squares solution。

Exercise 4.9. 考慮 $\mathbb{R}^n$ 利用 dot product 所成的 inner product space。令 W 為 V 的 subspace。

(1) 說明對任意 $\mathbf{v} \in \mathbb{R}^n$，存在 $\mathbf{w} \in W$、$\mathbf{w}' \in W^\perp$ 使得 $\mathbf{v} = \mathbf{w} + \mathbf{w}'$ 且說明此時 $\mathbf{v}$ 在 W 的投影為 $\mathbf{w}$ 以及 $\mathbf{v}$ 在 $W^\perp$ 的投影為 $\mathbf{w}'$。

(2) 假設 P_W、$P_{W^\perp}$ 分別為對 W 和 $W^\perp$ 的投影矩陣。證明 $P_W = I_n - P_{W^\perp}$（即證明若 $\mathbf{v} \in \mathbb{R}^n$，則 $(I_n - P_{W^\perp})\mathbf{v}$ 為 $\mathbf{v}$ 在 W 的投影）。

Exercise 4.10. 考慮 $\mathbb{R}^3$ 利用 dot product 所成的 inner product space。令

$$W = \{(x,y,z) \in \mathbb{R}^3 : x - 2y + z = 0\}.$$

以下我們用兩種方法求對 W 的投影矩陣。

(1) 找出 W 的一組 basis，並利用此組 basis 得到矩陣 A 使得 $\mathrm{Col}(A) = W$。

(2) 利用 (1) 中所得的 A 寫下對 W 的投影矩陣（請將矩陣具體乘開）。

(3) 找出 $W^\perp$ 的一組 basis，並利用此組 basis 得到矩陣 B 使得 $\mathrm{Col}(B) = W^\perp$。

(4) 利用 (3) 中所得的 B 寫下對 $W^\perp$ 的投影矩陣，並利用上一題 Exercise 4.9 (2) 的結果寫下對 W 的投影矩陣.

Exercise 4.11. 假設矩陣 A 的 column vectors 從左至右依序為

$$\mathbf{w}_1 = (1,3,1,1), \mathbf{w}_2 = (1,1,1,1), \mathbf{w}_3 = (-1,5,2,2)$$

(1) 請寫下 A 的 QR decomposition。

(2) 試利用 A 的 QR decomposition 寫下對 $\mathrm{Col}(A)$ 的投影矩陣。

Exercise 4.12. 此題我們用 normal equations 以及 QR decomposition 求 inconsistent system 的 least squares solution。請使用上一題 Exercise 4.11 的結果處理以下問題：考慮聯立方程組

$$\begin{cases} x & +y & -z & = & 4 \\ 3x & +y & +5z & = & -1 \\ x & +y & +2z & = & 5 \\ x & +y & +2z & = & 1 \end{cases}$$

(1) 將此聯立方程組寫成 $A\mathbf{x} = \mathbf{b}$ 的矩陣形式並寫下其 normal equations。

(2) 利用 normal equations 求原方程組的 least squares solution 以及其 error vector。

(3) 利用對 Col(A) 的投影矩陣求出原方程組的 error vector。

(4) 利用 QR decomposition 寫下與 normal equations 等價的方程組，並利用此方程組求出原方程組的 least squares solution。

Exercise 4.13. 考慮二維資料 $\{(-3,9),(-2,6),(0,2),(1,1)\}$。以下請利用 least squares 的方法分別找到所要求最接近函數並求其 error vectors。

(a) 常數函數 (b) 一次函數 (c) 二次函數

Exercise 4.14. 考慮二維資料 $\{(x_1,y_1),\ldots,(x_n,y_n)\}$，其中 $x_1,\ldots,x_n$ 不是全等（即存在 $x_i \neq x_j$）。令其 least squares line 為 $y = mx + c$ 且令 $y'_i = mx_i + c,\ i = 1,\ldots,n$。又令 μ_x 為 $x_1,\ldots,x_n$ 的平均數，μ_y 為 $y_1,\ldots,y_n$ 的平均數，以及 $\mu_{y'}$ 為 $y'_1,\ldots,y'_n$ 的平均數。已知 $y = mx + c$ 會通過點 (μ_x,μ_y)。利用講義 Proposition 4.5.5 回答以下問題：

(1) 說明若 $x_i \neq \mu_x$，則 $m = \dfrac{(y'_i - \mu_{y'})(x_i - \mu_x)}{(x_i - \mu_x)^2}$ 並依此推得 $m = \dfrac{\sum_{i=1}^n (y'_i - \mu_{y'})(x_i - \mu_x)}{\sum_{i=1}^n (x_i - \mu_x)^2}$.

(2) 證明 $\sum_{i=1}^n (y_i - \mu_y)(x_i - \mu_x) = \sum_{i=1}^n (y'_i - \mu_{y'})(x_i - \mu_x)$ 並依此推得

$$m = \frac{\sum_{i=1}^n (y_i - \mu_y)(x_i - \mu_x)}{\sum_{i=1}^n (x_i - \mu_x)^2}.$$

Exercise 4.15. 考慮聯立方程組

$$\begin{cases} x & +2y & +z & = & 4 \\ x & -y & +2z & = & -11 \\ x & +5y & & = & 19 \end{cases}$$

令其矩陣表示法為 $A\mathbf{x} = \mathbf{b}$。

(1) 試求此方程組的解集合。

(2) 找到此方程組的 minimal solution。

(3) 將此方程組的 minimal solution 表成 A 的 row vectors 的線性組合（寫下一組即可）。

Exercise 4.16. 考慮二維資料 $\{(2,2),(2,3),(2,4)\}$，我們想用找 minimal least squares solution 的方法找到 least squares line $y = mx + c$ 其中 $m^2 + c^2$ 是最小的。

(1) 寫下 m, c 所需符合的 normal equation，並用此 normal equation 找到其 minimal solution。

(2) 直接利用 $A^{\mathrm{t}}AA^{\mathrm{t}}\mathbf{x} = A^{\mathrm{t}}\mathbf{b}$ 的方式找到 minimal least squares solution。

(3) 試證明在更多組二維資料 $\{(x_0,y_1),\ldots,(x_0,y_k)\}$ 的情形，所得的 minimal least squares line 為 $y = \dfrac{x_0\mu_y}{x_0^2+1}x + \dfrac{\mu_y}{x_0^2+1}$。