接著我們探討 singular value 以及 singular vector 的性質。考慮 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathrm{rank}(A) = r$。假設 $A = U\Sigma V^{\mathrm{t}}$，其中 $U \in \mathsf{M}_{m\times m}(\mathbb{R})$, $V \in \mathsf{M}_{n\times n}(\mathbb{R})$ 皆為 orthogonal matrix（即 $U^{-1} = U^{\mathrm{t}}$ 且 $V^{-1} = V^{\mathrm{t}}$）而 $\Sigma \in \mathsf{M}_{m\times n}(\mathbb{R})$ 是 (i,i)-th entry 為 σ_i 其餘 entry 為 0 的 $m\times n$ 矩陣。則

$$A^{\mathrm{t}}A = (U\Sigma V^{\mathrm{t}})^{\mathrm{t}}(U\Sigma V) = (V\Sigma^{\mathrm{t}}U^{\mathrm{t}})(U\Sigma V^{\mathrm{t}}) = V(\Sigma^{\mathrm{t}}\Sigma)V^{\mathrm{t}} \tag{8.14}$$

注意 $\Sigma^{\mathrm{t}}\Sigma \in \mathsf{M}_{n\times n}(\mathbb{R})$ 為對角線 (i,i)-th entry 為 σ_i^2 的對角矩陣。所以式子 (8.14) 就是將 $A^{\mathrm{t}}A$ 做正交對角化。由於 $A^{\mathrm{t}}A$ 的 eigenvalue λ_i 是可以唯一確定的，所以 $\sigma_i = \sqrt{\lambda_i}$ 也可唯一確定。也因此若限定 $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n$，則矩陣 Σ 是唯一的。另一方面，若令矩陣 V 的 i-th column 為 $\mathbf{v}_i$，則 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為以 $A^{\mathrm{t}}A$ 的 eigenvectors 所形成的 $\mathbb{R}^n$ 之一組 orthonormal basis。同理考慮 AA^{t}，我們有 $AA^{\mathrm{t}} = U(\Sigma\Sigma^{\mathrm{t}})U^{\mathrm{t}}$，因為 $\Sigma\Sigma^{\mathrm{t}}$ 為 $m\times m$ 的對角矩陣，故若令矩陣 U 的 i-th column 為 $\mathbf{u}_i$，則 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 為以 AA^{t} 的 eigenvectors 所形成的 $\mathbb{R}^m$ 之一組 orthonormal basis。

注意，因為 $\mathrm{rank}(A) = \mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(AA^{\mathrm{t}}) = r$，從 $\Sigma^{\mathrm{t}}\Sigma$ 和 $\Sigma\Sigma^{\mathrm{t}}$ 的對角線來看可知：若將 $A^{\mathrm{t}}A$ 和 AA^{t} eigenvalue 從大排到小，它們前 r 個非零的 eigenvalue 會相同而 $A^{\mathrm{t}}A$ 和 AA^{t} 後面分別有 $n-r$、$m-r$ 個 eigenvalue 為 0。也因此 A 和 A^{t} 有相同的非零的 singular value，其餘分別有 $n-r$ 和 $m-r$ 個 singular value 為 0。

接著我們看 singular vectors。依 singular value 從大到小的順序，當 $1 \leq i \leq r$，由 $(A^{\mathrm{t}}A)\mathbf{v}_i = \sigma_i^2\mathbf{v}_i$，知 $\mathbf{v}_i \in \mathrm{Col}(A^{\mathrm{t}}) = \mathrm{Row}(A)$。因此由 $\dim(\mathrm{Row}(A)) = r$，得 $\mathbf{v}_1,\ldots,\mathbf{v}_r$ 為 $\mathrm{Row}(A)$ 的一組 orthonormal basis。又當 $r+1 \leq i \leq n$，$A(\mathbf{v}_i) = \mathbf{0}$，故 $\mathbf{v}_{r+1},\ldots,\mathbf{v}_n$ 為 $N(A)$ 的一組 orthonormal basis（驗證了 $N(A)^{\perp} = \mathrm{Col}(A^{\mathrm{t}}) = \mathrm{Row}(A)$）。同理 $\mathbf{u}_1,\ldots,\mathbf{u}_r$ 是 $\mathrm{Col}(A)$ 的一組 orthonormal basis 且 $\mathbf{u}_{r+1},\ldots,\mathbf{u}_m$ 為 $N(A^{\mathrm{t}})$ 的一組 orthonormal basis（驗證了 $N(A^{\mathrm{t}})^{\perp} = \mathrm{Col}(A)$）。另一方面，若 $A = U\Sigma V^{\mathrm{t}}$ 為 A 的 svd，則 $A^{\mathrm{t}} = V\Sigma^{\mathrm{t}}U^{\mathrm{t}}$ 就會是 A^{t} 的 svd。所以 A 的 right singular vector 就會是 A^{t} 的 left singular vector；反之亦然。

回顧在談論矩陣乘法時，我們介紹了矩陣乘法四種看法（參見 Section 2.4）。前面 singular value decomposition 的性質完全建立於矩陣乘法的結合律以及 inner product、column、row 這三種乘法的看法。而 svd 的應用，就在於第四種即 outer product 的看法。由於 A 的 svd 可寫成

$$A = U\Sigma V^{\mathrm{t}} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \mathbf{0} \\ & & \sigma_r & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right] \begin{bmatrix} \text{---}\ \mathbf{v}_1\ \text{---} \\ \vdots \\ \text{---}\ \mathbf{v}_n\ \text{---} \end{bmatrix}$$

因此利用 outer product（參見式子 (2.19)）可得

$$A = \sigma_1\mathbf{u}_1\mathbf{v}_1^{\mathrm{t}} + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^{\mathrm{t}}.$$

當我們將資料的訊息存成矩陣 A 時，就可由 A 的 singular value decomposition 依上式中 σ_i 的大小做資料的壓縮、修正或辨識等。

Singular Value Decomposition 對於一般 inner product space 之間的 linear transformation 也有相對應的版本。不過我們只談論 $\mathbb{R}^n, \mathbb{R}^m$ 上的 dot product，所以只考慮 $\mathbb{R}^n$ 到 $\mathbb{R}^m$

的 linear transformation。利用 linear transformation 給定定義域與對應域的 ordered basis 後所得的表現矩陣的方法，我們有以下的結論。

Proposition 8.4.2. 令 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation* 且假設 $\dim(\mathsf{R}(T))=r$。則在 $\mathbb{R}^n$ 與 $\mathbb{R}^m$ 上分別存在由其 *orthonormal basis* 所形成的 *ordered basis* β,γ 使得 T 的表現矩陣為

$$[T]_\beta^\gamma=\left[\begin{array}{ccc|c}\sigma_1 & & & \\ & \ddots & & \mathbf{0} \\ & & \sigma_r & \\ \hline & \mathbf{0} & & \mathbf{0}\end{array}\right]\in\mathsf{M}_{m\times n}(\mathbb{R}).$$

Proposition 8.4.2 對於 linear transformation 以及矩陣其所謂 *pseudo-inverse* 的概念相當重要。我們先談 linear transformation 的情況。

令 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 linear transformation 且假設 $\operatorname{rank}(T)=r$（即 T 的 image $\mathsf{R}(T)$ 的維度 $\dim(\mathsf{R}(T))=r$）令 $\mathbf{v}_1,\ldots,\mathbf{v}_r,\ldots,\mathbf{v}_n$ 是一組由 T 的 right singular vectors 所形成 $\mathbb{R}^n$ 上的 orthonormal basis，其中 singular values 滿足 $\sigma_1\geq\cdots\geq\sigma_r>0$ 以及 $\sigma_{r+1}=\cdots=\sigma_n=0$。且令 $\mathbf{u}_1,\ldots,\mathbf{u}_r,\cdots,\mathbf{u}_m$ 是所對應的 left singular vectors 所形成 $\mathbb{R}^n$ 上的 orthonormal basis。依此設定，$\mathbf{v}_{r+1},\ldots,\mathbf{v}_n$ 為 T 的 null space $\mathsf{N}(T)$ 的一組 orthonormal basis 且 $\mathbf{v}_1,\ldots,\mathbf{v}_r$ 為 $\mathsf{N}(T)^\perp$ 的一組 orthonormal basis。另一方面 $\mathbf{u}_1,\ldots,\mathbf{u}_r$ 為 T 的 image $\mathsf{R}(T)$ 的一組 orthonormal basis 且 $\mathbf{u}_{r+1},\ldots,\mathbf{u}_m$ 為 $\mathsf{R}(T)^\perp$ 的一組 orthonormal basis。依此設定，若將 T 限制在 $\mathsf{N}(T)^\perp=\operatorname{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_r)$ 之下，考慮 $T|_{\mathsf{N}(T)^\perp}:\mathsf{N}(T)^\perp\to\mathsf{R}(T)$，則 $T|_{\mathsf{N}(T)^\perp}$ 為 isomorphism（即一對一且映成的線性映射）。而 $T|_{\mathsf{N}(T)^\perp}$ 的反函數 $T|_{\mathsf{N}(T)^\perp}^{-1}$ 滿足

$$T|_{\mathsf{N}(T)^\perp}^{-1}(\mathbf{u}_i)=\frac{1}{\sigma_i}\mathbf{v}_i,\quad\forall i=1,\ldots,r.$$

由於一個 linear transformation 可以由定義域的一組基底所決定（參見 Theorem 6.1.8），所以利用 $\mathbf{u}_1,\ldots,\mathbf{u}_m$ 為 $\mathbb{R}^m$ 的一組基底，我們定義 $T^\dagger:\mathbb{R}^m\to\mathbb{R}^n$ 為滿足

$$T^\dagger(\mathbf{u}_i)=\begin{cases}\dfrac{1}{\sigma_i}\mathbf{v}_i, & 1\leq i\leq r;\\[2ex] \mathbf{0}, & r+1\leq i\leq m\end{cases}$$

的 linear transformation，並稱 $T^\dagger$ 為 T 的 *pseudo-inverse*。

為方便說明，我們將 T 和 $T^\dagger$ 之間關係顯示如下：

$$\begin{array}{ccccc|ccccc}
 & \mathbb{R}^n & \xrightarrow{T} & \mathbb{R}^m & & & \mathbb{R}^m & \xrightarrow{T^\dagger} & \mathbb{R}^n & \\
\mathsf{N}(T)^\perp & \left\{\begin{array}{c}\mathbf{v}_1\\ \vdots\\ \mathbf{v}_r\end{array}\right. & \longmapsto & \left.\begin{array}{c}\sigma_1\mathbf{u}_1\\ \vdots\\ \sigma_r\mathbf{u}_r\end{array}\right\} & \mathsf{R}(T) & \mathsf{R}(T) & \left\{\begin{array}{c}\mathbf{u}_1\\ \vdots\\ \mathbf{u}_r\end{array}\right. & \longmapsto & \left.\begin{array}{c}\frac{1}{\sigma_1}\mathbf{v}_1\\ \vdots\\ \frac{1}{\sigma_r}\mathbf{v}_r\end{array}\right\} & \mathsf{N}(T)^\perp \\
\mathsf{N}(T) & \left\{\begin{array}{c}\mathbf{v}_{r+1}\\ \vdots\\ \mathbf{v}_n\end{array}\right. & \longmapsto & \mathbf{0} & & \mathsf{R}(T)^\perp & \left\{\begin{array}{c}\mathbf{u}_{r+1}\\ \vdots\\ \mathbf{u}_m\end{array}\right. & \longmapsto & \mathbf{0} &
\end{array}$$

注意，從這裡可看出，雖然 T 的 singular vectors 並不唯一，但 $\mathsf{N}(T)$, $\mathsf{R}(T)$ 是由 T 決定的，所以由 $T^\dagger$ 滿足 $T^\dagger|_{\mathsf{R}(T)} = T|_{\mathsf{N}(T)^\perp}^{-1}$ 以及 $T^\dagger|_{\mathsf{R}(T)^\perp}$ 為零函數，知 $T^\dagger$ 是由 T 唯一確定的。

由上面 T, $T^\dagger$ 的關係，我們也很容易了解 $T^\dagger \circ T : \mathbb{R}^n \to \mathbb{R}^n$ 這個 linear operator。因為 $N(T)^\perp$ 的基底 $\mathbf{v}_1, \ldots, \mathbf{v}_r$ 中任一向量 $\mathbf{v}_i$ 皆滿足 $T^\dagger \circ T(\mathbf{v}_i) = \mathbf{v}_i$，所以 $T^\dagger \circ T(\mathbf{v}) = \mathbf{v}$, $\forall \mathbf{v} \in N(T)^\perp$。另一方面，因為 $N(T)$ 的基底 $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ 中任一向量 $\mathbf{v}_i$ 皆滿足 $T^\dagger \circ T(\mathbf{v}_i) = \mathbf{0}$，所以 $T^\dagger \circ T(\mathbf{v}) = \mathbf{0}$, $\forall \mathbf{v} \in N(T)$。也就是說 $T^\dagger \circ : \mathbb{R}^n \to \mathbb{R}^n$ 滿足

$$T^\dagger \circ T(\mathbf{v}) = \begin{cases} \mathbf{v}, & \forall \mathbf{v} \in N(T)^\perp; \\ \mathbf{0}, & \forall \mathbf{v} \in N(T) = (N(T)^\perp)^\perp \end{cases}$$

因此知 $T^\dagger \circ T : \mathbb{R}^n \to \mathbb{R}^n$ 是對 $\mathsf{N}(T)^\perp$ 的投影。同理，$T \circ T^\dagger : \mathbb{R}^m \to \mathbb{R}^m$ 是對 $\mathsf{R}(T)$ 的投影。我們得到以下結果：

Theorem 8.4.3. 令 $T : \mathbb{R}^n \to \mathbb{R}^m$ 為 *linear transformation* 且 $T^\dagger : \mathbb{R}^m \to \mathbb{R}^n$ 為 T 的 *pseudo-inverse*。則

- $T^\dagger \circ T : \mathbb{R}^n \to \mathbb{R}^n$ 是對 $\mathsf{N}(T)^\perp$ 的 *orthogonal projection.*
- $T \circ T^\dagger : \mathbb{R}^m \to \mathbb{R}^m$ 是對 $\mathsf{R}(T)$ 的 *orthogonal projection.*

特別的，當 $T : \mathbb{R}^n \to \mathbb{R}^m$ 為一對一，即 $\mathsf{N}(T) = \{\mathbf{0}\}$，此時 $\mathsf{N}(T)^\perp = \mathbb{R}^n$，所以 $T^\dagger \circ T = \mathrm{id}_{\mathbb{R}^n}$；而當 $T : \mathbb{R}^n \to \mathbb{R}^m$ 為映成，即 $\mathsf{R}(T) = \mathbb{R}^m$，故知 $T \circ T^\dagger = \mathrm{id}_{\mathbb{R}^m}$。而當 $T : \mathbb{R}^n \to \mathbb{R}^m$ 為 isomorphism（此時 $m = n$），由於 $T^\dagger \circ T = T \circ T^\dagger = \mathrm{id}_{\mathbb{R}^n}$，故知 $T^\dagger$ 為 T 的 inverse T^{-1}。所以我們可以說 pseudo-inverse 是反函數的推廣，使其適用於一般的 linear transformation.

接下來，我們將 $m \times n$ 矩陣視為 $\mathbb{R}^n$ 映射到 $\mathbb{R}^m$ 的 linear transformation，探討矩陣 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ 的 pseudo-inverse. 前面已知，對於 linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$，若 $\mathrm{rank}(T) = r$ 且 β, γ 分別為 $\mathbb{R}^n$, $\mathbb{R}^m$ 上 orthonormal ordered basis 使得

$$[T]_\beta^\gamma = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \mathbf{0} \\ & & \sigma_r & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right] \in \mathsf{M}_{m\times n}(\mathbb{R}),$$

則 $T^\dagger : \mathbb{R}^m \to \mathbb{R}^n$ 的表現矩陣為

$$[T^\dagger]_\gamma^\beta = \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \ddots & & \mathbf{0} \\ & & \frac{1}{\sigma_r} & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right] \in \mathsf{M}_{n\times m}(\mathbb{R}).$$

因此若 $A = U\Sigma V^t$ 為 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ 的 svd 其中 $\sigma_1, \cdots, \sigma_r > 0$ 且 $\sigma_{r+1} = \cdots = \sigma_n = 0$，則令 $\Sigma^\dagger$ 為 $n \times m$ 矩陣，其中當 $1 \le i \le r$ 時 $\Sigma^\dagger$ 的 (i,i)-th entry 為 $\dfrac{1}{\sigma_i}$，而其餘 entry 為 0。也就

是說 Σ 和 $\Sigma^\dagger$ 的關係如下：

$$\Sigma=\left[\begin{array}{ccc|c}\sigma_1 & & & \\ & \ddots & & \mathbf{0} \\ & & \sigma_r & \\ \hline & \mathbf{0} & & \mathbf{0}\end{array}\right]\in\mathsf{M}_{m\times n}(\mathbb{R}),\quad \Sigma^\dagger=\left[\begin{array}{ccc|c}\frac{1}{\sigma_1} & & & \\ & \ddots & & \mathbf{0} \\ & & \frac{1}{\sigma_r} & \\ \hline & \mathbf{0} & & \mathbf{0}\end{array}\right]\in\mathsf{M}_{n\times m}(\mathbb{R}).$$

決定了 $\Sigma^\dagger$ 後，我們定義 A 的 pseudo-inverse $A^\dagger\in\mathsf{M}_{n\times m}(\mathbb{R})$ 為

$$A^\dagger=V\Sigma^\dagger U^{\mathrm{t}}.$$

注意，前面提及一個 linear transformation 的 pseudo-inverse 是唯一的，所以雖然一個矩陣的 svd 並不唯一，但其 pseudo-inverse 是唯一的。

我們看 Example 8.4.1 中各矩陣的 pseudo-inverse.

Example 8.4.4. (1) $A_1=\begin{bmatrix}0&1\\0&0\end{bmatrix}$，已知 $A_1=\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}0&1\\1&0\end{bmatrix}$ 為 A_1 的 svd。因為 $\Sigma=\Sigma^\dagger$，可得

$$A_1^\dagger=\begin{bmatrix}0&1\\1&0\end{bmatrix}^{\mathrm{t}}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}1&0\\0&1\end{bmatrix}^{\mathrm{t}}=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}1&0\\0&1\end{bmatrix}=\begin{bmatrix}0&0\\1&0\end{bmatrix}$$

我們選在 Example 8.4.1 中 A_1 另一個 svd，即 $A_1=\begin{bmatrix}-1&0\\0&1\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}0&-1\\1&0\end{bmatrix}$. 計算其 pseudo-inverse 依然為

$$A_1^\dagger=\begin{bmatrix}0&-1\\1&0\end{bmatrix}^{\mathrm{t}}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}-1&0\\0&1\end{bmatrix}^{\mathrm{t}}=\begin{bmatrix}0&1\\-1&0\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}-1&0\\0&1\end{bmatrix}=\begin{bmatrix}0&0\\1&0\end{bmatrix}.$$

(2) $A_2=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$ 有一個 svd 為 $A_2=\begin{bmatrix}-1&0\\0&1\end{bmatrix}\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}0&1\\1&0\end{bmatrix}$，故得

$$A_2^\dagger=\begin{bmatrix}0&1\\1&0\end{bmatrix}^{\mathrm{t}}\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}-1&0\\0&1\end{bmatrix}^{\mathrm{t}}=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}-1&0\\0&1\end{bmatrix}=\begin{bmatrix}0&1\\-1&0\end{bmatrix}.$$

注意：A_2 為 invertible，所以其 pseudo-inverse 就是其 inverse，亦即 $A^\dagger=A^{-1}$。另外這裡 A_1,A_2 非 0 的 singular value 只有 1，所以 $\Sigma^\dagger=\Sigma^{\mathrm{t}}$。在此情況之下若 A 的 svd 為 $A=U\Sigma V^{\mathrm{t}}$，則 $A^\dagger=V\Sigma^\dagger U^{\mathrm{t}}=(V^{\mathrm{t}})^{\mathrm{t}}\Sigma^{\mathrm{t}}U^{\mathrm{t}}=A^{\mathrm{t}}$。所以 A_1,A_2 的 pseudo-inverse 都恰為其轉置矩陣。

(3) $A_3=\begin{bmatrix}0&1\\0&1\end{bmatrix}$ 有一個 svd 為 $A_3=\begin{bmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}\begin{bmatrix}\sqrt2&0\\0&0\end{bmatrix}\begin{bmatrix}0&1\\1&0\end{bmatrix}$，故得

$$A_3^\dagger=\begin{bmatrix}0&1\\1&0\end{bmatrix}^{\mathrm{t}}\begin{bmatrix}\frac{1}{\sqrt2}&0\\0&0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}^{\mathrm{t}}=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&0\\0&0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\-\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}=\begin{bmatrix}0&0\\\frac12&\frac12\end{bmatrix}.$$

(4) $A_4=\begin{bmatrix}1&2\\2&1\end{bmatrix}$ 有一個 svd 為 $A_4=\begin{bmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}\begin{bmatrix}3&0\\0&1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\end{bmatrix}$，故得

$$A_4^\dagger=\begin{bmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\end{bmatrix}^{\mathrm{t}}\begin{bmatrix}\frac13&0\\0&1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}^{\mathrm{t}}=\begin{bmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\end{bmatrix}\begin{bmatrix}\frac13&0\\0&1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\-\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}=\begin{bmatrix}-\frac13&\frac23\\\frac23&-\frac13\end{bmatrix}.$$

同樣的，因為 A_4 為可逆矩陣，故 $A_4^\dagger=A_4^{-1}$。

(5) $B=\begin{bmatrix}1&1\\1&-1\\1&1\end{bmatrix}$ 有一個 svd 為 $B=\begin{bmatrix}\frac{1}{\sqrt{2}}&0&\frac{1}{\sqrt{2}}\\0&1&0\\\frac{1}{\sqrt{2}}&0&-\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}2&0\\0&\sqrt{2}\\0&0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}$，故得

$$B^{\dagger}=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\frac{1}{2}&0&0\\0&\frac{1}{\sqrt{2}}&0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}&0&\frac{1}{\sqrt{2}}\\0&1&0\\\frac{1}{\sqrt{2}}&0&-\frac{1}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}\frac{1}{4}&\frac{1}{2}&\frac{1}{4}\\\frac{1}{4}&-\frac{1}{2}&\frac{1}{4}\end{bmatrix}.$$

注意，因為 $\operatorname{rank}(B)=2$，我們有 $B^{\dagger}B=I_2$。 $\sharp$

回顧一下，當給定 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$，我們可以將之視為線性映射 $T:\mathbb{R}^n\to\mathbb{R}^m$ 其中 $T(\mathbf{v})=A\mathbf{v},\forall\mathbf{v}\in\mathbb{R}^n$。此時 $\mathrm{Col}(A)=\mathsf{R}(T)$ 且 $\mathsf{N}(A)=\mathsf{N}(T)$，又因 $\mathsf{N}(A)^{\perp}=\mathrm{Col}(A^{\mathrm{t}})=\mathrm{Row}(A)$，我們將 A 和 $A^{\dagger}$ 之間關係顯示如下：

$$\begin{array}{ccccc|ccccc}
 & \mathbb{R}^n & \overset{A}{\longrightarrow} & \mathbb{R}^m & & & \mathbb{R}^m & \overset{A^{\dagger}}{\longrightarrow} & \mathbb{R}^n & \\
\mathrm{Row}(A) & \left\{\begin{matrix}\mathbf{v}_1\\ \vdots\\ \mathbf{v}_r\end{matrix}\right. & \longmapsto & \left.\begin{matrix}\sigma_1\mathbf{u}_1\\ \vdots\\ \sigma_r\mathbf{u}_r\end{matrix}\right\} & \mathrm{Col}(A) & \mathrm{Col}(A) & \left\{\begin{matrix}\mathbf{u}_1\\ \vdots\\ \mathbf{u}_r\end{matrix}\right. & \longmapsto & \left.\begin{matrix}\frac{1}{\sigma_1}\mathbf{v}_1\\ \vdots\\ \frac{1}{\sigma_r}\mathbf{v}_r\end{matrix}\right\} & \mathrm{Row}(A) \\
\mathsf{N}(A) & \left\{\begin{matrix}\mathbf{v}_{r+1}\\ \vdots\\ \mathbf{v}_n\end{matrix}\right. & \longmapsto & \mathbf{0} & & \mathrm{Col}(A)^{\perp} & \left\{\begin{matrix}\mathbf{u}_{r+1}\\ \vdots\\ \mathbf{u}_m\end{matrix}\right. & \longmapsto & \mathbf{0} &
\end{array}$$

同樣的，由於 $\mathrm{Row}(A)^{\perp}=\mathsf{N}(A)$ 我們有以下有關 Theorem 8.4.3 的矩陣版本：

Theorem 8.4.5. 設 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$ 且 $A^{\dagger}\in\mathsf{M}_{n\times m}$ 為 A 的 *pseudo-inverse*。則

- $A^{\dagger}A\in\mathsf{M}_{n\times n}(\mathbb{R})$ 是對 $\mathrm{Row}(A)$ 的投影矩陣。
- $AA^{\dagger}\in\mathsf{M}_{m\times m}(\mathbb{R})$ 是對 $\mathrm{Col}(A)$ 的投影矩陣。

特別的，當 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\operatorname{rank}(A)=n$，此時 $\mathrm{Row}(A)=\mathbb{R}^n$，所以 $A^{\dagger}A=I_n$（參見 Theorem 2.5.5）；而當 $\operatorname{rank}(A)=m$，即 $\mathrm{Col}(A)=\mathbb{R}^m$，可得 $AA^{\dagger}=I_m$（參見 Theorem 2.5.2）。又當 A 為可逆矩陣（此時 $n=m$ 且 $\operatorname{rank}(A)=n$），由於 $A^{\dagger}A=AA^{\dagger}=I_n$，故知 $A^{\dagger}$ 為 A 的反矩陣 A^{-1}（參見 Theorem 2.6.2）。所以我們可以說 pseudo-inverse 是反矩陣的推廣，使其適用於一般的矩陣。

Theorem 8.4.5 最主要的應用就是幫助我們處理聯立方程組 $A\mathbf{x}=\mathbf{b}$。回顧一下，當聯立方程組的係數矩陣 A 為可逆矩陣，則方程組 $A\mathbf{x}=\mathbf{b}$ 一定有解且有唯一的一組解 $\mathbf{x}=A^{-1}\mathbf{b}$。而當 A 不是可逆，那 $\mathbf{x}=A^{\dagger}\mathbf{b}$ 和聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的解有何關係呢？我們分成兩種情況討論：

$A\mathbf{x}=\mathbf{b}$ 有解. 此時 $\mathbf{b}\in\mathrm{Col}(A)$（參見 Lemma 2.5.1），又因 $AA^{\dagger}$ 為對 $\mathrm{Col}(A)$ 的投影，所以若將 $\mathbf{x}=A^{\dagger}\mathbf{b}$ 代入 $A\mathbf{x}$ 會得 $A\mathbf{x}=A(A^{\dagger}\mathbf{b})=(AA^{\dagger})\mathbf{b}=\mathbf{b}$。也就是說 $\mathbf{x}=A^{\dagger}\mathbf{b}$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解。然而若 $A\mathbf{x}=\mathbf{b}$ 有無窮多組解，那 $\mathbf{x}=A^{\dagger}\mathbf{b}$ 這組解有何特殊意義呢？回顧一下，當聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有無窮多組解時，其中一定有唯一的一組解是在 $\mathrm{Row}(A)$ 中，且這組解是所

有解中長度最短的一組解，稱之為 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution（參見 Corollary 4.5.8）。而在前面 $A, A^\dagger$ 的討論中，我們知道 $A^\dagger\mathbf{b}\in\mathrm{Row}(A)$，所以 $\mathbf{x}=A^\dagger\mathbf{b}$ 就是聯立方程組 $A\mathbf{x}=\mathbf{b}$ 在 $\mathrm{Row}(A)$ 中唯一的那組解，也就是說 $\mathbf{x}=A^\dagger\mathbf{b}$ 是 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution. 事實上，當 $\mathbf{x}_0$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解時，將 $\mathbf{x}_0$ 投影到 $\mathrm{Row}(A)$，得 $(A^\dagger A)\mathbf{x}_0=A^\dagger(A\mathbf{x}_0)=A^\dagger\mathbf{b}$，就是 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution。

$A\mathbf{x}=\mathbf{b}$ 無解. 此時 $\mathbf{b}\not\in\mathrm{Col}(A)$，所以我們要在 $\mathrm{Col}(A)$ 中找到一個距離 $\mathbf{b}$ 最近的向量 $\mathbf{b}_0$，此時 $A\mathbf{x}=\mathbf{b}_0$ 有解，且其解稱為方程組 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution。向量 $\mathbf{b}_0$ 事實上是 $\mathbf{b}$ 在 $\mathrm{Col}(A)$ 的投影（參見 Proposition 4.3.15），即 $\mathbf{b}_0=(AA^\dagger)\mathbf{b}=A(A^\dagger\mathbf{b})$。也就是說 $\mathbf{x}=A^\dagger\mathbf{b}$ 是 $A\mathbf{x}=\mathbf{b}_0$ 的一組解，因此也就是 $A\mathbf{x}=\mathbf{b}$ 的一組 least squares solution. 而若 $A\mathbf{x}=\mathbf{b}$ 有無窮多的 least squares solution，即 $A\mathbf{x}=\mathbf{b}_0$ 有無窮多解，此時由前面的討論我們知 $\mathbf{x}=A^\dagger\mathbf{b}$ 是所有解中長度最短的，所以它就是 $A\mathbf{x}=\mathbf{b}$ 的 minimal least squares solution。

Example 8.4.6. (1) 考慮聯立方程組

$$\text{(a)}\quad \begin{cases} x & +y & =1 \\ x & -y & =2 \\ x & +y & =1 \end{cases};\qquad \text{(b)}\quad \begin{cases} x & +y & =1 \\ x & -y & =2 \\ x & +y & =2 \end{cases}$$

為了方便起見我們分別用 $A\mathbf{x}=\mathbf{b}_1$, $A\mathbf{x}=\mathbf{b}_2$ 表示方程組 (a), (b). 很容易看出 (a) 有解但 (b) 無解。事實上若用合併的增廣矩陣 $[A|\mathbf{b}_1\,\mathbf{b}_2]$ 將係數矩陣 A 化為 echelon form，即 $\left[\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & -1 & 2 & 2 \\ 1 & 1 & 1 & 2 \end{array}\right]\rightsquigarrow\left[\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right]$ 可知 (a) 有唯一解 $\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} \frac{3}{2} \\ -\frac{1}{2} \end{bmatrix}$。而利用 Example 8.4.4 (5) 的結果, $A^\dagger=\begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix}$，確實 $\begin{bmatrix} x \\ y \end{bmatrix}=A^\dagger\mathbf{b}_1=\begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}=\begin{bmatrix} \frac{3}{2} \\ -\frac{1}{2} \end{bmatrix}$ 為 $A\mathbf{x}=\mathbf{b}_1$ 的唯一解。至於方程組 (b) 我們可利用其 normal equation $A^tA\mathbf{x}=A^t\mathbf{b}_2$，得到其 least squares solution（參見 Definition 4.5.2）。將 normal equation 的增廣矩陣 $[A^tA|A^t\mathbf{b}_2]$ 化為 echelon form，即 $\left[\begin{array}{cc|c} 3 & 1 & 5 \\ 1 & 3 & 1 \end{array}\right]\rightsquigarrow\left[\begin{array}{cc|c} 3 & 1 & 5 \\ 0 & \frac{8}{3} & \frac{-2}{3} \end{array}\right]$ 解得 $A\mathbf{x}=\mathbf{b}_2$ 的 least squares solution 為 $\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} \frac{7}{4} \\ -\frac{1}{4} \end{bmatrix}$，因而 $\begin{bmatrix} x \\ y \end{bmatrix}=A^\dagger\mathbf{b}_2=\begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}=\begin{bmatrix} \frac{7}{4} \\ -\frac{1}{4} \end{bmatrix}$ 確實是方程組 (b) 的 least squares solution.

(2) 考慮聯立方程組

$$\text{(a)}\quad \begin{cases} x & +y & +z=1 \\ x & -y & +z=1 \\ x & & +z=1 \end{cases};\qquad \text{(b)}\quad \begin{cases} x & +y & +z=1 \\ x & -y & +z=1 \\ x & & +z=2 \end{cases}$$

為了方便起見我們分別用 $A\mathbf{x}=\mathbf{b}_1$, $A\mathbf{x}=\mathbf{b}_2$ 表示方程組 (a), (b). 若用合併的增廣矩陣 $[A|\mathbf{b}_1\,\mathbf{b}_2]$ 將係數矩陣 A 化為 echelon form，即 $\left[\begin{array}{ccc|cc} 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 2 \end{array}\right]\rightsquigarrow\left[\begin{array}{ccc|cc} 1 & 1 & 1 & 1 & 1 \\ 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$

可知 (a) 有無窮多解 $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1-t \\ 0 \\ t \end{bmatrix}$；而 (b) 無解。所以我們需找到 (a) 的 minimal solution 以及 (b) 的 minimal least squares solution.

要找到 $A\mathbf{x} = \mathbf{b}_1$ 的 minimal solution，可以先找到 $AA^{\mathrm{t}}\mathbf{x} = \mathbf{b}_1$ 的一組解 $\mathbf{x}_0$，便可得 $\mathbf{x} = A^{\mathrm{t}}\mathbf{x}_0$ 為 $A\mathbf{x} = \mathbf{b}_1$ 的 minimal solution（參見 Proposition 4.5.7）。故先將增廣矩陣 $[AA^{\mathrm{t}}|\mathbf{b}_1]$ 化為 echelon form，即 $\left[\begin{array}{ccc|c} 3 & 1 & 2 & 1 \\ 1 & 3 & 2 & 1 \\ 2 & 2 & 2 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 3 & 1 & 2 & 1 \\ 0 & \frac{8}{3} & \frac{4}{3} & \frac{2}{3} \\ 0 & 0 & 0 & 0 \end{array}\right]$ 得到 $\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2} \end{bmatrix}$ 為 $AA^{\mathrm{t}}\mathbf{x} = \mathbf{b}_1$ 的一組解。也因此知 $\mathbf{x} = A^{\mathrm{t}}\mathbf{x}_0 = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$ 為方程組 (a) 的 minimal solution. 而利用 A 的 svd 可算出 $A^{\dagger} = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{2} & \frac{-1}{2} & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$，因而知 $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = A^{\dagger}\mathbf{b}_1 = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{2} & \frac{-1}{2} & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$ 確實是 $A\mathbf{x} = \mathbf{b}_1$ 的 minimal solution。

至於方程組 $A\mathbf{x} = \mathbf{b}_2$，由於其 normal equation $A^{\mathrm{t}}A\mathbf{x} = A^{\mathrm{t}}\mathbf{b}_2$ 的解都是 least squares solution。所以套用剛才方法找到 normal equation $\begin{bmatrix} 3 & 0 & 3 \\ 0 & 2 & 0 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 4 \end{bmatrix}$ 的 minimal solution 即可。亦即，若 normal equation 為 $B\mathbf{x} = \mathbf{b}_0$，先找到 $BB^{\mathrm{t}}\mathbf{x} = \mathbf{b}_0$ 的一組解，例如 $\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \\ \frac{2}{9} \end{bmatrix}$，就可得 minimal solution $B^{\mathrm{t}}\mathbf{x}_0 = \begin{bmatrix} \frac{2}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}$。另一方面，$A^{\dagger}\mathbf{b}_2 = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{2} & \frac{-1}{2} & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}$ 確實是方程組 (b) 的 minimal least squares solution. $\sharp$

Exercise 8.21. 請直接利用 Exercise 8.16 關於以下矩陣的 svd 處理。

$$C=\begin{bmatrix}1&1&1\\1&0&-1\\1&-2&1\\1&1&1\end{bmatrix},\quad C'=\begin{bmatrix}1&1&1&1\\1&0&-2&1\\1&-1&1&1\end{bmatrix}$$

(1) 將矩陣 C 寫成 rank 1 的矩陣（principle component）之和。

(2) 分別寫下 C,C' 的 pseudo-inverse.

(3) 考慮聯立方程組

$$\text{(a)}\begin{cases}x_1 & +x_2 & +x_3 & +x_4=2\\ x_1 & & -2x_3 & +x_4=-1\\ x_1 & -x_2 & +x_3 & +x_4=2\end{cases};\quad \text{(b)}\begin{cases}x_1 & +x_2 & +x_3=1\\ x_1 & & -x_3=0\\ x_1 & -x_2 & +x_3=1\\ x_1 & +x_2 & +x_3=2\end{cases}$$

若有解請寫下所有的解，並利用 pseudo-inverse 求得其 minimal solution；若無解，請寫下其 normal equation。求其解並驗證是否與利用 pseudo-inverse 求得的 least squares solution 一致.

Exercise 8.22. 考慮矩陣 $A=\begin{bmatrix}1 & 0\end{bmatrix}$, $B=\begin{bmatrix}1\\1\end{bmatrix}$。

(1) 試求 $A^\dagger$, $B^\dagger$, $(AB)^\dagger$ 以及 $(BA)^\dagger$.

(2) 驗證 $(AB)^\dagger=B^\dagger A^\dagger$ 以及 $(BA)^\dagger=A^\dagger B^\dagger$ 是否成立。

Exercise 8.23. 設 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$。證明以下成立：

(1) $(A^\dagger)^\dagger=A$.

(2) $(A^{\mathrm{t}})^\dagger=(A^\dagger)^{\mathrm{t}}$.

(3) $AA^\dagger A=A$ 且 $A^\dagger AA^\dagger=A^\dagger$.

Exercise 8.24. 假設 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$, $B\in\mathsf{M}_{n\times k}(\mathbb{R})$ 且 $\mathrm{rank}(A)=\mathrm{rank}(B)=n$. 回顧：此時 $A^{\mathrm{t}}A$ 以及 $B^{\mathrm{t}}B$ 皆為 invertible.

(1) 證明 $A^\dagger=(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}$ 且 $B^\dagger=B^{\mathrm{t}}(BB^{\mathrm{t}})^{-1}$.

(2) 證明 $(AB)^\dagger=B^\dagger A^\dagger$.