第 5 章 线性最小二乘问题

在第 4 章中，通过描述子匹配，我们已经得到了图像 I_1 和 I_2 中特征点对应点对关系集合 $\mathcal{S}=\left\{\mathbf{x}_i \leftrightarrow \mathbf{x}_i^{'}\right\}_{i=1}^{P}$，其中 $\mathbf{x}_i$ 是来自 I_1 的特征点，$\mathbf{x}_i^{'}$ 是来自 I_2 的特征点，$\mathbf{x}_i \leftrightarrow \mathbf{x}_i^{'}$ 表示 $\mathbf{x}_i$ 与 $\mathbf{x}_i^{'}$ 是一对对应的特征点，P 为 I_1 和 I_2 中具有对应关系的特征点点对的个数。根据全景拼接问题的描述，图像 I_1 和 I_2 的所有对应点可以通过同一个线性几何变换 H 关联起来。在没有任何其他先验知识的情况下，我们把 H 考虑成平面之间最具有普适性的线性变换，射影变换，则 $\forall \mathbf{x}_i \leftrightarrow \mathbf{x}_i^{'} \in \mathcal{S}$，

$$c\mathbf{x}_i^{'} = H_{3\times 3}\mathbf{x}_i \tag{5-1}$$

其中 $c \neq 0$ 是一个与 $\mathbf{x}_i^{'}$ 有关的常数，$H=\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$ 是一个 3×3 的表达平面上射影变换的 8 自由度矩阵。由于 $\mathbf{x}_i$、$\mathbf{x}_i^{'}$ 都是从图像上检测到的特征点，因此它们都是平面上的正常点（非无穷远点），因此可假定式 5-1 中的 $\mathbf{x}_i$、$\mathbf{x}_i^{'}$ 都是规范化齐次坐标形式（如果不是，则可以先转化为规范化齐次坐标形式），即 $\mathbf{x}_i=\left(x_i, y_i, 1\right)$，$\mathbf{x}_i^{'}=\left(x_i^{'}, y_i^{'}, 1\right)$。则式 5-1 可进一步变为，

$$c\begin{bmatrix} x_i^{'} \\ y_i^{'} \\ 1 \end{bmatrix}_i^{'} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{5-2}$$

这样，从每一个点对关系 $\mathbf{x}_i \leftrightarrow \mathbf{x}_i^{'}$ 中都可以得到一个形如式 5-2 的关于 H 的等式。那么，点对关系集合 $\mathcal{S}$ 中共有 P 个元素，因此可以得到 P 个形如式 5-2 的等式。接下去的任务就是要从这 P 个等式中解出 H。根据具体处理方式的不同，这个问题可以建模为齐次线性最小二乘问题或者非齐次线性最小二乘问题。接下去我们就对这两个问题详加阐述。

在本章后面的推导过程中，会遇到函数或自变量的表达中包含矩阵或向量的求导问题，如果读者对这些内容不是很熟悉的话，请参见本书附录 C。

5.1 齐次线性最小二乘问题

5.1.1 问题定义

给定一个点对关系 $\mathbf{x}_i \leftrightarrow \mathbf{x}_i^{'}$，我们得到了一个形如式 5-2 的方程。对这个方程左右展开得到，

$$\begin{cases} h_{11}x_i + h_{12}y_i + h_{13} = cx_i^{'} \\ h_{21}x_i + h_{22}y_i + h_{23} = cy_i^{'} \\ h_{31}x_i + h_{32}y_i + h_{33} = c \end{cases} \tag{5-3}$$

将式 5-3 中第一式和第三式的左右两边相除、第二式和第三式的左右两边相除，得到，

$$\begin{cases} \dfrac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}} = x_i^{'} \\ \dfrac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}} = y_i^{'} \end{cases} \tag{5-4}$$

把式 5-4 从形式上整理得到，

$$\begin{pmatrix} x_i & y_i & 1 & 0 & 0 & 0 & -x_i x_i^{'} & -y_i x_i^{'} & -x_i^{'} \\ 0 & 0 & 0 & x_i & y_i & 1 & -x_i y_i^{'} & -y_i y_i^{'} & -y_i^{'} \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{pmatrix} = \mathbf{0} \tag{5-5}$$

从式 5-5 中可以看出，由一对点对关系 $\mathbf{x}_i \leftrightarrow \mathbf{x}_i^{'}$，我们可以得到两个方程。如果有 4 对点对关系 $\left\{\mathbf{x}_i \leftrightarrow \mathbf{x}_i^{'}\right\}_{i=1}^{4}$ 的话，便可以得到 8 个线性方程，写成矩阵形式即为，

$$A_{8\times9}\mathbf{h}_{9\times1} = \mathbf{0} \tag{5-6}$$

其中，$A_{8\times9}$ 是方程组的系数矩阵，$\mathbf{h}_{9\times1} = \left(h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}\right)^T$。在一般情况下，$rank\left(A_{8\times9}\right) = 8$，则齐次线性方程组 5-6 的解空间中存在(9-8)=1 个线性无关的解向量[1]，这个解向量便对应于我们最终需要的射影矩阵 H。因此，从理论上来说，两个平面间的射影变换关系可以由 4 个有效对应点对唯一确定。我们强调是“有效”对应点对，指的是 $\left\{\mathbf{x}_i\right\}_{i=1}^{4}$ 和 $\left\{\mathbf{x}_i^{'}\right\}_{i=1}^{4}$ 中不能存在三点共线的情况。

但在估计平面间的射影变换时，我们得到的点对关系集合 $\mathcal{S} = \left\{\mathbf{x}_i \leftrightarrow \mathbf{x}_i^{'}\right\}_{i=1}^{P}$ 中的元素数量往往会远多于 4 对，即 $P > 4$，这时问题会变成什么形式呢？显然，这时从 P 个点对中，我们可以得到 $2P$ 个线性方程，其矩阵形式为，

$$A_{2P\times9}\mathbf{h}_{9\times1} = \mathbf{0} \tag{5-7}$$

其中，$A_{2P\times9}$ 是方程组的系数矩阵，$\mathbf{h}_{9\times1} = \left(h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}\right)^T$。在一般情况下，$rank\left(A_{2P\times9}\right) = 9$，因此根据齐次线性方程组解的理论[1]，方程组 5-7 只有平凡的零解。然而零解对于我们的问题来说没有意义，我们并不需要零解。我们希望能在最小二乘意义之下找

到一个适合于方程组 5-7 的非零解$\mathbf{h}^*$。同时注意到，在我们的问题中，解向量$\mathbf{h}^*$实际上代表了射影变换矩阵，而由射影变换的性质可知，对于任意实数$k \neq 0$，$k\mathbf{h}^*$与$\mathbf{h}^*$会表达相同的射影变换。因此，不失一般性，我们可以约束$\left\|\mathbf{h}^*\right\|_2^2 = 1$。这样，我们的问题就被建模为，

$$\mathbf{h}^* = \arg\min_{\mathbf{h}} \left\|A\mathbf{h}\right\|_2^2, \text{subject to} \left\|\mathbf{h}\right\|_2^2 = 1, A \in \mathbb{R}^{2P\times 9}, \mathbf{h} \in \mathbb{R}^{9\times 1} \tag{5-8}$$

其中，$P > 4$，$rank\left(A\right) = 9$。

我们可以以一种更加普适的表达方式来描述形如 5-8 所代表的一类问题，

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left\|A\mathbf{x}\right\|_2^2, \text{subject to} \left\|\mathbf{x}\right\|_2^2 = 1, A \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^{n\times 1} \tag{5-9}$$

其中，$rank\left(A\right) = n$。我们将在 5.1.2 中讲述如何解式 5-9 所定义的这个优化问题。

5.1.2 问题的求解

式 5-9 的求解问题是一个典型的带有等式约束的求函数最小值点的问题。目标函数$f\left(\mathbf{x}\right) = \left\|A\mathbf{x}\right\|_2^2$与等式约束函数$g\left(\mathbf{x}\right) = 1 - \left\|\mathbf{x}\right\|_2^2$关于优化变量 $\mathbf{x}$ 都有连续的一阶偏导数。因此，我们可以用拉格朗日乘子法来找出$f\left(\mathbf{x}\right)$在等式约束$g\left(\mathbf{x}\right) = 0$下所有可能的极值点。

构造拉格朗日函数，

$$L\left(\mathbf{x}, \lambda\right) = f\left(\mathbf{x}\right) + \lambda g\left(\mathbf{x}\right) = \left\|A\mathbf{x}\right\|_2^2 + \lambda\left(1 - \left\|\mathbf{x}\right\|_2^2\right) \tag{5-10}$$

根据拉格朗日乘子法的原理，首先要找出$L\left(\mathbf{x}, \lambda\right)$的驻点。设$\left(\mathbf{x}_0, \lambda_0\right)$是$L\left(\mathbf{x}, \lambda\right)$的一个驻点，则它必须要满足，

$$\begin{cases} \dfrac{\partial L}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}_0, \lambda=\lambda_0} = \mathbf{0} \\ \dfrac{\partial L}{\partial \lambda}\Big|_{\mathbf{x}=\mathbf{x}_0, \lambda=\lambda_0} = 0 \end{cases} \Rightarrow \begin{cases} \dfrac{\partial\left(\mathbf{x}^T A^T A\mathbf{x} + \lambda\left(1 - \mathbf{x}^T\mathbf{x}\right)\right)}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}_0, \lambda=\lambda_0} = \mathbf{0} \\ \dfrac{\partial\left(\mathbf{x}^T A^T A\mathbf{x} + \lambda\left(1 - \mathbf{x}^T\mathbf{x}\right)\right)}{\partial \lambda}\Big|_{\mathbf{x}=\mathbf{x}_0, \lambda=\lambda_0} = 0 \end{cases} \Rightarrow \begin{cases} A^T A\mathbf{x}_0 = \lambda_0\mathbf{x}_0 \\ \mathbf{x}_0^T\mathbf{x}_0 = 1 \end{cases} \tag{5-11}$$

即λ_0是$A^T A$的特征值，$\mathbf{x}_0$是$A^T A$对应于特征值λ_0的单位特征向量。显然，满足这样条件的“$\left(\mathbf{x}_0, \lambda_0\right)$”并不是唯一的。设集合$\mathcal{S} = \left\{\left(\mathbf{x}_i, \lambda_i\right) : | A^T A\mathbf{x}_i = \lambda_i\mathbf{x}_i, \mathbf{x}_i^T\mathbf{x}_i = 1\right\}$，则$\mathcal{S}$表示$L\left(\mathbf{x}, \lambda\right)$的所有驻点。设集合$\mathcal{C} = \left\{\mathbf{x}_i : | \left(\mathbf{x}_i, \lambda_i\right) \in \mathcal{S}\right\}$，则$\mathcal{C}$表示$f\left(\mathbf{x}\right)$在等式约束$g\left(\mathbf{x}\right) = 0$下所有可能的极值点。接下来，我们要在$\mathcal{C}$中挑选出能使$f\left(\mathbf{x}\right)$取得最小值（在等式约束$g\left(\mathbf{x}\right) = 0$下）的点。

若$\mathbf{x}_i \in \mathcal{C}$，则，

$$f\left(\mathbf{x}_i\right) = \left\|A\mathbf{x}_i\right\|_2^2 = \mathbf{x}_i^T A^T A\mathbf{x}_i = \mathbf{x}_i^T \lambda_i\mathbf{x}_i = \lambda_i \tag{5-12}$$

则可知 $f(\mathbf{x})$ 的最小值为 $\min\{\lambda_i\}$[1]，即为 A^TA 最小的特征值。而 $f(\mathbf{x})$ 能取到这个最小值（在等式约束 $g(\mathbf{x})=0$ 下）的点为 A^TA 的对应于其最小特征值的单位特征向量。

5.2 非齐次线性最小二乘问题

5.2.1 问题定义

我们知道，表达平面间射影变换的矩阵 H 虽然有 9 个元素，但它只有 8 个自由度。在 5.1 节的实际处理中，我们以限定“向量化之后的 H 为 9 维单位向量”的方式（式 5-8）把它的自由度限定为 8。本节我们用另外一种思路来限定 H 的自由度。

假设 H 中的元素 $h_{ij}\neq 0$，则在求解 H 的过程中可以把 h_{ij} 固定为一个常数 $c\neq 0$。不失一般性，假设 $h_{33}\neq 0$，我们固定 h_{33} 为 $h_{33}=1$。这样，给定一对对应点对 $\mathbf{x}_i^{'}=(x_i^{'},y_i^{'},1)$ 和 $\mathbf{x}_i=(x_i,y_i,1)$，它们之间的关系可表达为，

$$c\begin{bmatrix} x_i^{'} \\ y_i^{'} \\ 1 \end{bmatrix}=\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix}\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{5-13}$$

对式 5-13 左右展开得到，

$$\begin{cases} h_{11}x_i+h_{12}y_i+h_{13}=cx_i^{'} \\ h_{21}x_i+h_{22}y_i+h_{23}=cy_i^{'} \\ h_{31}x_i+h_{32}y_i+1=c \end{cases} \tag{5-14}$$

将式 5-14 中第一式和第三式的左右两边相除、第二式和第三式的左右两边相除，得到，

$$\begin{cases} \dfrac{h_{11}x_i+h_{12}y_i+h_{13}}{h_{31}x_i+h_{32}y_i+1}=x_i^{'} \\ \dfrac{h_{21}x_i+h_{22}y_i+h_{23}}{h_{31}x_i+h_{32}y_i+1}=y_i^{'} \end{cases} \tag{5-15}$$

把式 5-15 从形式上整理得到，

[1] $\{\lambda_i\}$ 表示由 A^TA 的所有特征值构成的集合。

$$\begin{pmatrix} x_i & y_i & 1 & 0 & 0 & 0 & -x_i x_i' & -y_i x_i' \\ 0 & 0 & 0 & x_i & y_i & 1 & -x_i y_i' & -y_i y_i' \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{pmatrix} = \begin{pmatrix} x_i' \\ y_i' \end{pmatrix} \tag{5-16}$$

从式 5-16 中可以看出，由一对点对关系 $\mathbf{x}_i \leftrightarrow \mathbf{x}_i'$，我们可以得到两个方程。如果有 4 对点对关系 $\{\mathbf{x}_i \leftrightarrow \mathbf{x}_i'\}_{i=1}^4$ 的话，便可以得到 8 个线性方程，写成矩阵的形式为，

$$A_{8\times8}\mathbf{h}_{8\times1} = \mathbf{b}_{8\times1} \tag{5-17}$$

其中，$A_{8\times8}$ 是方程组的系数矩阵，$\mathbf{h}_{8\times1} = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32})^T$，$\mathbf{b}_{8\times1} = (x_1', y_1', x_2', y_2', x_3', y_3', x_4', y_4')^T$。在一般情况下（$\{\mathbf{x}_i\}_{i=1}^4$ 中以及 $\{\mathbf{x}_i'\}_{i=1}^4$ 中都不能有三点共线），$rank(A_{8\times8}) = rank([A_{8\times8}; \mathbf{b}]) = 8$，则方程组 5-17 有唯一解[1]，从这个解向量我们就可以相应得到最终的射影矩阵 H。因此，从理论上来说，通过两个平面内 4 个有效对应点对，我们便可以唯一确定这两个平面间的射影变换关系。

但一般情况下，我们的点对关系集合 $\mathcal{S} = \{\mathbf{x}_i \leftrightarrow \mathbf{x}_i'\}_{i=1}^P$ 中的元素数量远多于 4 对，即 $P > 4$。这时从 P 个点对中，可以得到 $2P$ 个线性方程，其矩阵形式为，

$$A_{2P\times8}\mathbf{h}_{8\times1} = \mathbf{b}_{2P\times1} \tag{5-18}$$

其中，$A_{2P\times8}$ 是方程组的系数矩阵，$\mathbf{h}_{8\times1} = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32})^T$，$\mathbf{b}_{2P\times1}$ 为非零常数向量。在一般情况下，方程组 5-18 的系数矩阵的秩 $rank(A_{2P\times8}) = 8$，而其增广矩阵的秩 $rank([A_{2P\times8}; \mathbf{b}]) = 9$，因此根据线性方程组解的理论[1]，方程组 5-18 无解。既然从理论上来说，方程组 5-18 无解，我们只能退而求其次，希望能在最小二乘意义之下找到一个适合于方程组 5-18 的 $\mathbf{h}^*$。这样，我们的问题就被建模为，

$$\mathbf{h}^* = \arg\min_{\mathbf{h}} \|A\mathbf{h} - \mathbf{b}\|_2^2, A \in \mathbb{R}^{2P\times8}, \mathbf{h} \in \mathbb{R}^{8\times1}, \mathbf{b} \in \mathbb{R}^{2P\times1} \tag{5-19}$$

其中，$P > 4$，$rank(A) = 8$。

我们可以以一种更加普适的表达方式来描述形如 5-19 所代表的一类问题，

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2^2, A \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^{n\times1}, \mathbf{b} \neq \mathbf{0} \in \mathbb{R}^{m\times1} \tag{5-20}$$

其中，$rank(A) = n$。我们将在 5.2.2 中讲述如何解式 5-20 所定义的这个优化问题。

5.2.2 问题的求解

求式 5-20 最优解的问题是一个典型的无约束凸优化问题。我们可以首先来证明该问题的目标函数，

$$f\left(\mathbf{x}\right)=\left\|A\mathbf{x}-\mathbf{b}\right\|_2^2, A\in\mathbb{R}^{m\times n},\mathbf{x}\in\mathbb{R}^{n\times 1},\mathbf{b}\neq\mathbf{0}\in\mathbb{R}^{m\times 1} \tag{5-21}$$

为凸函数，这个证明作为练习请读者来完成。然后，我们来找到目标函数 $f(\mathbf{x})$的驻点。如果 $\mathbf{x}_s$ 为 $f(\mathbf{x})$的驻点，那么 $\mathbf{x}_s$ 需要满足，

$$\begin{aligned}\nabla f\left(\mathbf{x}\right)|_{\mathbf{x}=\mathbf{x}_s}&=\frac{d\left\|A\mathbf{x}-\mathbf{b}\right\|_2^2}{d\mathbf{x}}|_{\mathbf{x}=\mathbf{x}_s}\\&=\frac{d\left(\left(A\mathbf{x}-\mathbf{b}\right)^T\left(A\mathbf{x}-\mathbf{b}\right)\right)}{d\mathbf{x}}|_{\mathbf{x}=\mathbf{x}_s}\\&=\frac{d\left(\mathbf{x}^TA^TA\mathbf{x}-2\mathbf{x}^TA^T\mathbf{b}+\mathbf{b}^T\mathbf{b}\right)}{d\mathbf{x}}|_{\mathbf{x}=\mathbf{x}_s}\\&=2A^TA\mathbf{x}-2A^T\mathbf{b}|_{\mathbf{x}=\mathbf{x}_s}=\mathbf{0}\end{aligned} \tag{5-22}$$

因此，

$$\mathbf{x}_s=\left(A^TA\right)^{-1}A^T\mathbf{b} \tag{5-23}$$

我们需要注意的是，式 5-23 要成立的话，A^TA必须要可逆才可以。事实上，在我们这个问题中，由于我们要求$rank\left(A\right)=n$，即 A 是列满秩矩阵，则可以证明 A^TA一定是可逆的，这个证明作为练习请读者来完成。待优化的目标函数 $f(\mathbf{x})$为凸函数，则它的驻点一定也是全局最小值点[2]。因此，问题 5-20 的最优解就是$\mathbf{x}^*=\mathbf{x}_s=\left(A^TA\right)^{-1}A^T\mathbf{b}$。

5.2.3 基于奇异值分解原理的求解方法

本节将介绍非齐次线性最小二乘问题的另外一种解法：基于奇异值分解的方法。该方法同 5.2.2 中介绍的方法相比，有两个优越之处：1）要用 5.2.2 节中介绍的方法来解非齐次线性最小二乘问题时，问题中的系数矩阵必须是列满秩矩阵，如式 5-20 中，$rank\left(A_{m\times n}\right)=n$，而本节介绍的方法并不需要待解问题满足这个附加条件；2）从计算机算法实现的角度来说，本节中介绍的基于奇异值分解的方法所产生的解会具有更高的数值精度[3]。如果读者对矩阵奇异值分解的基本内容不太熟悉的话，可参见本书附录 D。

首先再来梳理一下我们要解决的问题。我们想要解如下线性方程组，

$$A_{m\times n}\mathbf{x}_{n\times 1}=\mathbf{b}_{m\times 1},\mathbf{b}\neq\mathbf{0} \tag{5-24}$$

方程组 5-24 的解会出现的情况无外乎以下三种：1）$rank\left(A\right)=rank\left(\left[A;\mathbf{b}\right]\right)=n$，此时方程组有唯一解，我们要把这个解找出来；2）$rank\left(A\right)=rank\left(\left[A;\mathbf{b}\right]\right)<n$，此时方程组有无穷多组解，我们要找到其中一个来解决我们手里的实际问题；3）$rank\left(A\right)\neq rank\left(\left[A;\mathbf{b}\right]\right)$，此时方程

组无解，我们要在最小二乘意义之下找到一个“最适合”该方程组的解。不难看出，对以上三种情况的处理都可以归结为求解如下问题，

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|A\mathbf{x}-\mathbf{b}\|_2^2, A \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^{n\times 1}, \mathbf{b} \neq \mathbf{0} \in \mathbb{R}^{m\times 1} \tag{5-25}$$

需要注意的是，式 5-25 所定义的问题同式 5-20 不同，后者有一个额外的要求 $rank(A)=n$ 而前者没有。下面我们就来看看如何具体来求解问题 5-25。

对 A 进行奇异值分解得到，

$$A = U_{m\times m}\Sigma_{m\times n}V_{n\times n}^T \tag{5-26}$$

其中 U 和 V 为正交矩阵。假设 $rank(A)=r$，则，

$$\Sigma_{m\times n} = \begin{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(m-r)\times r} & \boldsymbol{O}_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n} \tag{5-27}$$

其中 $\sigma_1,\sigma_2,...,\sigma_r > 0$ 为矩阵 A 的奇异值。进一步有，

$$\begin{aligned} A\mathbf{x}-\mathbf{b} &= U\Sigma V^T\mathbf{x}-\mathbf{b} \\ &= U\left(\Sigma V^T\mathbf{x}\right) - U\left(U^T\mathbf{b}\right) \\ &= U\left(\Sigma V^T\mathbf{x} - U^T\mathbf{b}\right) \\ &\triangleq U\left(\Sigma\mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1}\right) \end{aligned} \tag{5-28}$$

其中 $\mathbf{y}_{n\times 1} = V^T\mathbf{x}, \mathbf{c}_{m\times 1} = U^T\mathbf{b}$。由于 U 是正交矩阵，它可以保持向量长度，因此，

$$\|A\mathbf{x}-\mathbf{b}\|_2 = \left\|U\left(\Sigma\mathbf{y}-\mathbf{c}\right)\right\|_2 = \left\|\Sigma\mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1}\right\|_2 \tag{5-29}$$

我们的最终目的是要找到 $\mathbf{x}^* = \arg\min_{\mathbf{x}} \|A\mathbf{x}-\mathbf{b}\|_2$。由于有 5-29 式，我们可以间接地先找出最优的 $\mathbf{y}^* = \arg\min_{\mathbf{y}} \left\|\Sigma\mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1}\right\|_2$，再根据 $\mathbf{y}^* = V^T\mathbf{x}^*$ 解出 $\mathbf{x}^*$ 即可。

由于，

$$\Sigma\mathbf{y}_{n\times 1} = \begin{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(m-r)\times r} & \boldsymbol{O}_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sigma_1 y_1 \\ \sigma_2 y_2 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m\times 1} \tag{5-30}$$

所以，

$$\Sigma \mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1} = \begin{bmatrix} \sigma_1 y_1 - c_1 \\ \sigma_2 y_2 - c_2 \\ \vdots \\ \sigma_r y_r - c_r \\ -c_{r+1} \\ \vdots \\ -c_m \end{bmatrix}_{m\times 1} \tag{5-31}$$

根据式 5-31，只要让 $y_i = \dfrac{c_i}{\sigma_i}, 1 \le i \le r$（此时 $y_{r+1}, \dots, y_n$ 可以是任意值），则 $\left\| \Sigma \mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1} \right\|_2$ 便可取到最小长度 $\left(\sum_{i=r+1}^{m} c_i^2 \right)^{1/2}$。满足这个要求的一个“最简单”的 $\mathbf{y}_{n\times 1}^*$ 可以为，

$$\mathbf{y}_{n\times 1}^* = \begin{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_r} \end{bmatrix} & \boldsymbol{O}_{r\times(m-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(m-r)} \end{bmatrix}_{n\times m} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}_{m\times 1} = \begin{bmatrix} c_1/\sigma_1 \\ c_2/\sigma_2 \\ \vdots \\ c_r/\sigma_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1} \triangleq \Sigma^+ \mathbf{c}_{m\times 1} \tag{5-32}$$

其中，Σ^+ 代表了对矩阵 Σ 做转置并把非零对角元取倒数的操作。需要强调的是，当 $r < n$ 时，最优的 $\mathbf{y}^*$ 是不唯一的，式 5-32 中给出的 $\mathbf{y}^*$ 只是满足条件 $y_i = \dfrac{c_i}{\sigma_i}(1 \le i \le r)$ 的最优 $\mathbf{y}^*$ 中的一个。当然，如果 $r = n$，即系数矩阵 $A_{m\times n}$ 是列满秩矩阵，则最优 $\mathbf{y}^*$ 是唯一的。

有了 $\mathbf{y}^*$ 之后，可以自然得出 $\mathbf{x}^*$，

$$\mathbf{x}^* = V\mathbf{y}^* = V\Sigma^+\mathbf{c} = V\Sigma^+U^T\mathbf{b} \tag{5-33}$$

其中，$A^+ \triangleq V\Sigma^+U^T$ 称为 A 的 Moore-Penrose 广义逆。

5.3 习题

（1） 式 5-12 中出现的 λ_i 有没有可能是负数？为什么？

（2）请证明式 5-20 中的优化问题中，目标函数 $f(\mathbf{x}) = \left\| A\mathbf{x} - \mathbf{b} \right\|_2^2, A \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^{n\times 1}, \mathbf{b} \ne \mathbf{0} \in \mathbb{R}^{m\times 1}$ 为凸函数。提示：由于 $f(\mathbf{x})$ 二阶可微，我

们只需要证明该函数的定义域为凸集并且它的 Hessian 矩阵为半正定矩阵[2]即可。

（3） 有矩阵 $A \in \mathbb{R}^{m \times n}$ 且 $rank\left(A\right) = n$，请证明矩阵 $A^T A$ 必为可逆矩阵。

参考文献

[1] 李世栋，乐经良，冯卫国，王纪林，线性代数，科学出版社，2000 年。

[2] Stephen Boyd, Lieven Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[3] Gene H. Golub, Charles F. Van Loan, Matrix Computations, John Hopkins Univ Press, Baltimore, 1983.