第五章 矩阵计算问题

第 14 讲 线性方程组

黄定江

DaSE @ ECNU

djhuang@dase.ecnu.edu.cn

14.1.1 线性方程组引例

在工程问题中, 线性方程组描述了变量之间最基本的关系. 线性方程在各个科学分支中无处不在，例如弹性力学，电阻网络，曲线拟合等. 线性方程构成了线性代数的核心并且经常构成了优化问题的约束条件. 因为许多优化算法的迭代过程非常依赖线性方程组的解, 所以它也是许多优化算法的基础. 接下来，我们展示一个线性方程组的例子.

例 1

(三点测距问题) 三角测量是一种确定点位置的方法，给定距离到已知控制点 (锚点). 三边测量可以应用于许多不同的领域，如地理测绘，地震学，导航（例如 *GPS* 系统）等. 在图*1*中，三个测距点 $a_1, a_2, a_3 \in \mathbb{R}^2$ 的坐标是已知的，并且从点 $\boldsymbol{x} = (x_1, x_2)^{\mathrm{T}}$ 到测距点的距离为 d_1, d_2, d_3. $\boldsymbol{x}$ 的未知坐标与距离测量有关, 可以由下面非线性方程组描述

$$\|\boldsymbol{x} - \boldsymbol{a}_1\|_2^2 = d_1^2, \quad \|\boldsymbol{x} - \boldsymbol{a}_2\|_2^2 = d_2^2, \quad \|\boldsymbol{x} - \boldsymbol{a}_3\|_2^2 = d_3^2$$

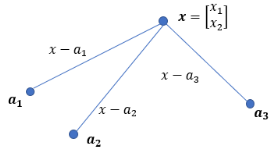


图 1: 三点测量位置图例

通过第一个方程减去另外两个方程，我们获得了两个 $\boldsymbol{x}$ 的线性方程组。

$$2\left(\boldsymbol{a}_2 - \boldsymbol{a}_1\right)^{\mathrm{T}} \boldsymbol{x} = d_1^2 - d_2^2 + \|\boldsymbol{a}_2\|_2^2 - \|\boldsymbol{a}_1\|_2^2$$

$$2\left(\boldsymbol{a}_3 - \boldsymbol{a}_1\right)^{\mathrm{T}} \boldsymbol{x} = d_1^2 - d_3^2 + \|\boldsymbol{a}_3\|_2^2 - \|\boldsymbol{a}_1\|_2^2$$

也就是说，原始非线性方程组的每个解也可以看作线性方程组的解。使用方程组标准形式 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ 可以描述为：

$$A = \begin{bmatrix} 2\left(\boldsymbol{a}_2 - \boldsymbol{a}_1\right)^{\mathrm{T}} \\ 2\left(\boldsymbol{a}_3 - \boldsymbol{a}_1\right)^{\mathrm{T}} \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} d_1^2 - d_2^2 + \|\boldsymbol{a}_2\|_2^2 - \|\boldsymbol{a}_1\|_2^2 \\ d_1^2 - d_3^2 + \|\boldsymbol{a}_3\|_2^2 - \|\boldsymbol{a}_1\|_2^2 \end{bmatrix}$$

正如先前的示例, 一般的线性方程被描述为如下的向量形式:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{1}$$

其中 $\boldsymbol{x} \in \mathbb{R}^n$ 是未知变量，$\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 是参数矩阵，$\boldsymbol{y} \in \mathbb{R}^m$ 是已知向量. 若解决先前示例中的问题，我们必须要解线性方程问题以及这些线性方程组的解是否存在？

若存在，方程组的解是否唯一呢？针对解得存在性以及唯一性，接下来我们先讨论线性方程组的基本性质，描述线性方程组所有可能的解.

根据 $\boldsymbol{A}$ 的性质，我们估计线性方程组可能没有解，或者有唯一解，或者有无限多个解. 对于有无限多的解，解得集合形成了 $\mathbb{R}^n$ 的子空间. 对于没有解的情况，我们将引进近似解的概念. 我们将充分利用子空间的各种性质, 对线性方程进行分析.

14.1.2 线性方程组 (回顾)

含有 n 个未知量 $x_1, x_2, \cdots, x_n$, n 个方程的线性方程组的一般形式为

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n. \end{cases} \tag{2}$$

若记

$$\boldsymbol{A} = (a_{ij})_{n\times n}, \boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T, \boldsymbol{b} = (b_1, b_2, \cdots, b_n)^T,$$

则方程 (2) 可表示为如下的矩阵形式：

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \tag{3}$$

方程组

$$\boldsymbol{A}\boldsymbol{x}=\mathbf{0} \tag{4}$$

称为方程组 (3) 对应的**齐次线性方程组**, 当 $\boldsymbol{b}\neq 0$ 时，方程组 (3) 称为**非齐次线性方程组**。

定义 1

矩阵 $\boldsymbol{A}=(a_{ij})^{m\times n}$, 称为方程组 (3) 的 **系数矩阵**, 矩阵 $\tilde{\boldsymbol{A}}=(\boldsymbol{A},\boldsymbol{b})$ 称为它的 **增广矩阵**. 方程组 $\boldsymbol{A}\boldsymbol{x}=\mathbf{0}$ 称为方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的 **导出组**。

定义 2

若向量 $\boldsymbol{x}^0=\left(x_1{}^0,x_2{}^0,\cdots,x_n{}^0\right)^T$ 满足方程组(3), 即 $\boldsymbol{A}\boldsymbol{x}^0=\boldsymbol{b}$, 称 $\boldsymbol{x}^0$ 为方程组(3)的**解向量**.

求解方程组 —系数矩阵为可逆方阵 (回顾)

定理 1

(克莱姆法则) 如果 $det(\boldsymbol{A}) \neq 0$, 则方程组(3)有唯一的解 $x_i = \frac{D_i}{det(\boldsymbol{A})}, (i = 1, 2, ..., n)$, 其中 $det(\boldsymbol{A})$ 是未知量 $x_1, x_2, \cdots, x_n$ 的**系数矩阵 $\boldsymbol{A}$ 的行列式**. 而 D_i 是将 $det(\boldsymbol{A})$ 中第 i 列的元素 $a_{1i}, a_{2i}, \cdots, a_{ni}$ 分别用 $b_1, b_2, \cdots, b_n$ 去替代所得的行列式。

定理 2

齐次线性方程组 $\boldsymbol{A}\mathbf{x} = \mathbf{0}$ 有非零解的必要充分条件是 $\det(\boldsymbol{A}) = 0$.

当 $\det(\boldsymbol{A}) = |a_{ij}| \neq 0$ 时, 方程组(3)的解向量为 $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$. 这种方法通常称为**逆矩阵法**

求解方程组 —一般线性方程组 (回顾)

定义 3

给定方程组

$$\bar{\boldsymbol{A}}\boldsymbol{x} = \bar{\boldsymbol{b}}. \tag{5}$$

若 $\boldsymbol{x}^0 = \left(x_1^0, x_2^0, \cdots x_n^0\right)^T$ 是方程组(5) 的解向量时，则它也是方程组(3)：$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的解向量恒成立，则方程组(3) 与方程组(5)是 **同解方程组**。

定理 3

对方程组(3)的系数矩阵 $\boldsymbol{A}$ 及右端 $\boldsymbol{b}$ 作相同的行初等变换，得到的新方程组与原方程组是同解方程组。

定义 4

设 $\boldsymbol{\eta_1}, \boldsymbol{\eta_2}, \cdots, \boldsymbol{\eta_t}$，是齐次线性方程组(4)：$\boldsymbol{Ax} = \mathbf{0}$ 的解向量组，如果 $\boldsymbol{\eta_1}, \boldsymbol{\eta_2}, \cdots, \boldsymbol{\eta_t}$，线性无关，且方程组(4)的任意解向量 $\boldsymbol{\eta}$ 都可由 $\boldsymbol{\eta_1}, \boldsymbol{\eta_2}, \cdots, \boldsymbol{\eta_t}$，线性表出，则称解向量组 $\boldsymbol{\eta_1}, \boldsymbol{\eta_2}, \cdots, \boldsymbol{\eta_t}$，为方程组(4)的一个 **基础解系**.

定理 4

设齐次线性方程组(4)的系数矩阵 $\boldsymbol{A}$ 的秩为 r, 此时

（1）方程组(4)有非零解的必要充分条件是 $r < n$.

（2）若 $r < n$，则方程组(4)一定有基础解系. 基础解系不是唯一的，但任两个基础解系必等价，且每一个基础解系所含解向量的个数都等于 $n - r$.

（3）若 $r < n$，设 $\boldsymbol{\eta_1}, \boldsymbol{\eta_2}, \cdots, \boldsymbol{\eta_{n-r}}$ 是方程组(4) 的一个基础解系，则它的一般解为

$$\boldsymbol{\eta} = \lambda_1\boldsymbol{\eta_1} + \lambda_2\boldsymbol{\eta_2} + \cdots + \lambda_{n-r}\boldsymbol{\eta_{n-r}} \tag{6}$$

其中 $\lambda_i(i = 1, 2, \cdots, n - r)$ 是数域 $\mathbb{K}$ 中的任意常数.

定理 5

方程组(3)有解的必要充分条件是：$\text{rank}(\boldsymbol{A}) = \text{rank}(\tilde{\boldsymbol{A}})$. 矩阵 $\tilde{\boldsymbol{A}}$ 为它的 **增广矩阵**

定理 6

设 $\text{rank}(\boldsymbol{A}) = \text{rank}(\tilde{\boldsymbol{A}}) = r$,$\gamma_{\mathbf{0}}$ 是非齐次方程组(3)的一个解向量（常称为**特解**）,$\eta_{\mathbf{1}}, \eta_{\mathbf{2}}, \cdots, \eta_{\mathbf{n-r}}$ 是其导出组(4)的一个基础解系，则方程组(3)的解向量均可表为：

$$\gamma = \gamma_0 + \eta = \gamma_0 + \lambda_1 \boldsymbol{\eta}_1 + \lambda_2 \boldsymbol{\eta}_2 + \cdots + \lambda_{n-r} \boldsymbol{\eta}_{n-r},$$

其中 $\lambda_i (i = 1, 2, \cdots, n-r)$ 是数域 $\mathbb{K}$ 中的任意常数（这种形式的解向量常称为一般解）.

14.1.3 线性方程组的解集与基本子空间 (回顾)

线性方程组：

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}, \qquad \boldsymbol{A} \in \mathbb{R}^{m\times n} \tag{7}$$

线性方程组的解集被定义为:

$$S \doteq \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\} \tag{8}$$

用 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n \in \mathbb{R}^n$ 表示矩阵 $\boldsymbol{A}$ 的列, i.e. $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n]$.$\boldsymbol{A}\boldsymbol{x}$ 仅仅表示矩阵 $\boldsymbol{A}$ 的列与向量 $\boldsymbol{x}$ 中各个元素的加权和:

$$\boldsymbol{A}\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n \tag{9}$$

回顾定义

$\boldsymbol{A}$ 的列空间定义为:

$$\mathsf{Col}\,(\boldsymbol{A}) = \mathbf{span}\,\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n\}\,,$$

其中 $\boldsymbol{a}_i$ 为 $\boldsymbol{A}$ 的列向量；$\boldsymbol{A}$ 的零空间定义为:

$$\mathsf{Null}\,(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = 0\} \tag{10}$$

它的维数记为 $\mathsf{nullity}\,(\boldsymbol{A})$

一个子空间 $\mathbb{S} \subset \mathbb{R}^n$ 的正交补定义为:

$$\mathbb{S}^{\perp} = \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y}^T\boldsymbol{x} = 0, \forall \boldsymbol{x} \in \mathbb{S}\} \tag{11}$$

通过定义，我们能够看出, 无论 $\boldsymbol{x}$ 的值是什么, $\boldsymbol{A}\boldsymbol{x}$ 生成了由矩阵 $\boldsymbol{A}$ 的列张成的子空间。向量 $\boldsymbol{A}\boldsymbol{x} \in \mathsf{Col}\,(\boldsymbol{A})$。若 $\boldsymbol{y} \notin \mathsf{Col}\,(\boldsymbol{A})$，则线性方程组没有解. 因此解集 $\mathbb{S}$ 为空. 等价地，线性方程组有解当且仅当 $\boldsymbol{y} \in \mathsf{Col}\,(\boldsymbol{A})$

从矩阵值域的角度，给出定理5的证明：

定理 7

方程组(3)的解存在的充分必要条件是

$$\text{rank}(\boldsymbol{A}) = \text{rank}([\boldsymbol{A}, \boldsymbol{b}]) \tag{12}$$

证明 **必要性** 设存在 $\boldsymbol{x}$ 使 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, 则 $\boldsymbol{b}$ 是 $\boldsymbol{A}$ 的列向量的线性组合, 即有 $\boldsymbol{b} \in \text{Col}(\boldsymbol{A})$. 这说明 $\text{Col}([\boldsymbol{A}, \boldsymbol{b}]) = \text{Col}(\boldsymbol{A})$, 由此即知，必有 $\text{rank}([\boldsymbol{A}, \boldsymbol{b}]) = \text{rank}(\boldsymbol{A})$.

充分性 若 $\text{rank}([\boldsymbol{A}, \boldsymbol{b}]) = \text{rank}(\boldsymbol{A})$ 成立, 则 $\boldsymbol{b} \in \text{Col}(\boldsymbol{A})$, 即 $\boldsymbol{b}$ 可表为

$$\boldsymbol{b} = \sum_{i=1}^{n} x_i a_i \tag{13}$$

这里 $\boldsymbol{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]$. 于是，令 $\boldsymbol{x} = (x_1, \cdots, x_n)^{\mathrm{T}}$, 即有 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, 从而定理得证.

线性方程组的解集与零空间:

定理 8

假定方程组(3)的解存在, 并且假定 $\boldsymbol{x}$ 是其任一给定的解, 则(3)的全部解得集合是

$$\boldsymbol{x} + Null(\boldsymbol{A}) \tag{14}$$

证明 如果 $\boldsymbol{y}$ 满足(3), 则 $\boldsymbol{A}(\boldsymbol{y}-\boldsymbol{x})=0$, 即 $(\boldsymbol{y}-\boldsymbol{x}) \in \text{Null}(\boldsymbol{A})$, 于是有 $\boldsymbol{y}=\boldsymbol{x}+(\boldsymbol{y}-\boldsymbol{x}) \in \boldsymbol{x}+\text{Null}(\boldsymbol{A})$. 反之, 如果 $\boldsymbol{y} \in x+\text{Null}(\boldsymbol{A})$, 则存在 $\boldsymbol{z} \in \text{Null}(\boldsymbol{A})$, 使 $\boldsymbol{y}=\boldsymbol{x}+\boldsymbol{z}$, 从而有 $\boldsymbol{A}\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}\boldsymbol{z}=\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$.

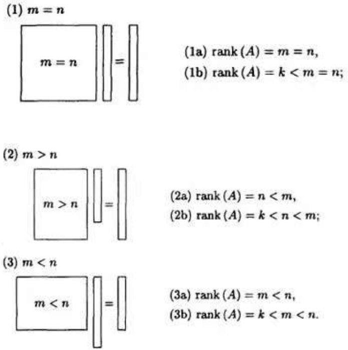
定理8告诉我们, 只要知道了(3)的一个解, 便可以用它及 $\text{Null}(\boldsymbol{A})$ 中的向量的和得到(3)的全部解. 由此可知,(3)的解要想唯一, 只有当 $\text{Null}(\boldsymbol{A})$ 中仅有零向量才行.

推论 1

(3)的解唯一的充分必要条件是 $nullity(\boldsymbol{A})=0$.

14.1.4 线性方程组的分类

按照矩阵 $\boldsymbol{A}$ 的秩与行列数的关系，可以划分为三种不同类型的方程:



我们先简略的考虑这三种情形下的解集，并在之后的章节中详细介绍

14.1.4 线性方程组的分类

超定系统

线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$ 中线性方程的个数大于未知变量的个数时，我们说 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$ 是超定的. 或者说矩阵 $\boldsymbol{A}^{m\times n}$ 的行数大于列数: $m>n$. 假设 $\boldsymbol{A}$ 是一个列满秩矩阵, 也就是说 $\mathrm{rank}(\boldsymbol{A})=n$, 则我们可以得出 $\mathsf{Null}\,(\boldsymbol{A})=0$

因此线性方程组的解要么没有解，要么有唯一解. 在超定系统中，$\boldsymbol{y}\notin\mathsf{Col}\,(\boldsymbol{A})$ 是很常见的，因此引入近似解的概念，近似解使得 $\boldsymbol{A}\boldsymbol{x}$ 与 $\boldsymbol{y}$ 在合适的度量下距离最小.(后面详细讨论)

14.1.4 线性方程组的分类

欠定系统

线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$ 中未知变量的个数小于方程组的个数时，我们说线性方程组是欠定的. 或者说 $\boldsymbol{A}^{m\times n}$ 的列数大于行数:$m<n$. 假设 $\boldsymbol{A}$ 是一个行满秩矩阵，也就是说 $\mathrm{rank}(\boldsymbol{A})=m$, $\mathsf{Col}\,(\boldsymbol{A})=\mathbb{R}^m$. 则根据定理8:

$$\mathrm{rank}(\boldsymbol{A})+\dim(\mathsf{Null}\,(\boldsymbol{A}))=n \tag{15}$$

因此 $\dim(\mathsf{Null}\,(\boldsymbol{A}))=n-m>0$. 此时线性方程组有解且有无限多个解，并且解集的维度是 $n-m$。在所有可能的解中，我们总是对具有最小范数的解很感兴趣.(后面详细讨论)

14.1.4 线性方程组的分类

方阵系统

线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$ 中线性方程的个数等于未知变量的个数时，我们说 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$ 是方阵. 或者说 $\boldsymbol{A}^{m\times n}$ 的列数等于行数:$m=n$. 如果系数矩阵是满秩的即 $\boldsymbol{A}$ 可逆, $\boldsymbol{A}^{-1}$ 唯一且有 $\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{I}$. 在这种情况下，线性方程组的解是唯一的:

$$\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{y} \tag{16}$$

注意，实际中我们几乎不会通过先求 $\boldsymbol{A}^{-1}$ 再乘以向量 $\boldsymbol{y}$ 的方式求解 $\boldsymbol{x}$. 而是通过数值方法 (比如之前学过的 LU 分解，Cholesky 分解) 来计算线性方程组非奇异系统的解.

最后我们看一个在本章中反复使用的定理:

定理 9

下面 2 条性质成立:

1. $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ *是一个列满秩矩阵 (i.e.,* $\mathrm{rank}(\boldsymbol{A}) = n$*) 当且仅当* $\boldsymbol{A}^T\boldsymbol{A}$ *是可逆的.*
2. $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ *是一个行满秩 (i.e. ,* $\mathrm{rank}(\boldsymbol{A}) = m$*) 当且仅当* $\boldsymbol{A}\boldsymbol{A}^T$ *是可逆的.*

证明.

对于第一点. 如果 $\boldsymbol{A}^T\boldsymbol{A}$ 不是可逆的, 则存在 $\boldsymbol{x} \neq 0$ 使得 $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = 0$. $\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = 0$, 因此 $\boldsymbol{A}\boldsymbol{x} = 0$. 所以 $\boldsymbol{A}$ 不是一个列满秩矩阵. 反之, 如果 $\boldsymbol{A}^T\boldsymbol{A}$ 是可逆的, 对于每个 $\boldsymbol{x} \neq 0, \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} \neq 0$, 也能推出对于每一个非零的 $\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x} \neq 0$. 第二点的证明过程与第一点的证明过程相似. $\square$

我们首先要讨论的一类特殊的线性方程组

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y},\ \boldsymbol{A}\in\mathbb{R}^{n\times n}$$

其中 $\boldsymbol{A}$ 可逆。

直接方法

- 如果 $\boldsymbol{A}$ 还具有其他特殊的结构，例如是上三角形矩阵或者下三角形矩阵，则可以直接应用前代法或者回代法。
- 如果 $\boldsymbol{A}$ 不是三角形矩阵，那么我们则可以使用高斯消去法把方程变成一个是带有上三角形式的方程。
- 但这个方法的缺点是：如果有几个方程系数矩阵相同而右侧项 $\boldsymbol{y}$ 不同，则需要每次将重做整个过程。

基于矩阵分解的方法

- 另外一种常见的求解方法是使用矩阵分解的方法。我们可以将矩阵分解成一系列特殊结构矩阵的乘积。
- 例如：正交矩阵、对角矩阵、三角形矩阵等。
- 然后我们通过更简单的方程组的等式来找到解。
- 这种方法的一个优势是，一旦我们对系数矩阵进行了分解，那么对于不同的右侧项就无需重新计算。

例 2

设矩阵 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ 的奇异值分解为 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$，其中 $\boldsymbol{U},\boldsymbol{V}$ 是正交矩阵，$\boldsymbol{\Sigma}$ 是对角矩阵且可逆。那么我们在求解方程组

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$$

时，等价求解一系列如下方程组

$$\boldsymbol{U}\boldsymbol{w}=\boldsymbol{y},\boldsymbol{\Sigma}\boldsymbol{z}=\boldsymbol{w},\ \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{z}$$

而这些方程对应的解为

$$\boldsymbol{w}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y},\boldsymbol{z}=\boldsymbol{\Sigma}^{-1}\boldsymbol{w},\boldsymbol{x}=\boldsymbol{V}\boldsymbol{z}$$

例 3

设可逆矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 有 QR 分解 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ 其中 $\boldsymbol{Q}$ 是正交矩阵 $\boldsymbol{R}$ 是主对角线均为正的上三角矩阵。那么我们在求解方程组

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

时，等价求解方程组

$$\boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{y}$$

而对于上三角矩阵的方程组，我们可以使用回代法来求解。

例 4

设矩阵 $\boldsymbol{A}$ 有 LU 分解 $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$，其中 $\boldsymbol{L}$ 是下三角矩阵，$\boldsymbol{U}$ 是上三角矩阵。那么我们在求解方程组

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

时，等价求解一系列如下方程组

$$\boldsymbol{L}\boldsymbol{z} = \boldsymbol{y}, \boldsymbol{U}\boldsymbol{x} = \boldsymbol{z}$$

对于第一个下三角方程，我们使用前代法来求解。对于第二个上三角方程，我们使用回代法来求解。

例 5

设矩阵 $\boldsymbol{A}$ 有 Cholesky 分解 $\boldsymbol{A}=\boldsymbol{L}\boldsymbol{L}^T$，其中 $\boldsymbol{L}$ 是下三角矩阵。那么我们在求解方程组

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$$

时，等价求解一系列如下方程组

$$\boldsymbol{L}\boldsymbol{z}=\boldsymbol{y},\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{z}$$

对于第一个下三角方程，我们使用前代法来求解。对于第二个上三角方程，我们使用回代法来求解。

上面考虑了方程情形，我们接下来考虑非方阵情形

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}, \boldsymbol{A}\in\mathbb{R}^{m\times n}, \boldsymbol{y}\in\mathbb{R}^{m}$$

我们可以计算 $\boldsymbol{A}$ 的奇异值分解。设 $\boldsymbol{A}=\boldsymbol{U}\tilde{\boldsymbol{\Sigma}}\boldsymbol{V}^{\mathrm{T}}$，记 $\tilde{\boldsymbol{x}}=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}, \tilde{\boldsymbol{y}}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y}$，那么我们就得到了一个关于对角矩阵的方程组

$$\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{x}}=\tilde{\boldsymbol{y}}$$

上述这个方程组是很容易计算的。但是同样可能会出现两种情况：

- 如果 $\tilde{\boldsymbol{y}}$ 最后 $m-r$ 个分量不为零。因为这 $m-r$ 个方程左边为 0，所以方程组将无解。这种情况表明 $\boldsymbol{y}$ 不在 $\boldsymbol{A}$ 的列空间中。
- 而当 $\boldsymbol{y}$ 在 $\boldsymbol{A}$ 的列空间中，那么最后 $m-r$ 个方程成立，我们可以用前面 r 个方程进行求解，即

$$\tilde{\boldsymbol{x}}_i=\frac{\tilde{\boldsymbol{y}}_i}{\sigma_i}, i=1,\ldots,r$$

  $\tilde{\boldsymbol{x}}$ 中后 $n-r$ 个分量可以取任意值。如果 $\boldsymbol{A}$ 是一个列满秩矩阵（即他的零空间为 $\{\boldsymbol{0}\}$），那么我们就会有唯一解。

我们一旦计算得到了 $\tilde{\boldsymbol{x}}$ 那么就可以通过 $\boldsymbol{x}=\boldsymbol{V}\tilde{\boldsymbol{x}}$ 来得到方程的解。

我们首先考虑一个最简单的线性方程组

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} \boldsymbol{x} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

其中 $a_{ii} \neq 0, i = 1, 2, \dots, n$。

那么就有

$$\boldsymbol{x} = \begin{pmatrix} y1/a_{11} \\ y_2/a_{22} \\ \vdots \\ y_n/a_{nn} \end{pmatrix}$$

14.3 三角形线性方程组

在前面，我们考虑了理论上如何对一个一般的线性方程组求解。

接下来，我们将将考虑如何使用计算机对一个线性方程组求解，尤其是一个规模巨大的线性方程组。

首先来考虑一个稍微简单些的情况，三角形线性方程组。

回忆在 LU 分解时，我们讲解过上三角矩阵和上三角矩阵：

定义 5

如果一个矩阵 $\boldsymbol{A}$ 主对角线以上所有元素为 *0*, 则称其为**下三角矩阵**。
如果一个矩阵 $\boldsymbol{A}$ 主对角线以下的所以元素为 *0*, 则称其为**上三角矩阵**。

定义 6

如果一个线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的系数矩阵 $\boldsymbol{A}$ 是上三角形矩阵或者下三角矩阵，我们则称其为**上三角形线性方程组**或者**下三角形线性方程组**

针对上三角形线性方程组和下三角形线性方程组，我们可以分别用两种特别的方法解出方程组的解。
接下来，我们分别介绍这两种方法。

我们利用**前代法**计算下三角形线性方程组。

注意，我们要求系数矩阵主对角线上元素均非 0。从而保证方程组有且仅有一个解。

$$\begin{pmatrix} a_{11} & & & & \\ a_{21} & a_{22} & & & \\ a_{31} & a_{32} & a_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix}$$

其中 $a_{11}, a_{22}, ..., a_{nn}$ 非 0.

通过第一个方程

$$a_{11}x_1 = b_1$$

我们可以很容易的求解 x_1。

得到 x_1 后，我们将 x_1 代入第二个方程

$$a_{21}x_1 + a_{22}x_2 = b_2$$

我们可以很容易的求解 x_2。

$\cdots\cdots$

像这样，通过求解前 $k-1$ 个方程得到前 $k-1$ 个未知量，并将其代入第 k 个方程，从而求得第 k 个未知量的解方程的方法称为前代法。

从矩阵变换的角度，在前代法的第 (k) 个循环中，我们将会遇到下面这样一个形式，此时，我们很容易的通过 $a_{kk}x_k = b_k^{(k-1)}$ 来求得 x_k。

$$\begin{pmatrix} a_{11} & & & & & & & \\ 0 & a_{22} & & & & & & \\ 0 & 0 & a_{33} & & & & & \\ \vdots & \vdots & \vdots & \ddots & & & & \\ 0 & 0 & 0 & \dots & a_{kk} & & & \\ 0 & 0 & 0 & \dots & a_{k+1k} & a_{k+1k+1} & & \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & a_{nk} & a_{nk+1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2^{(1)} \\ b_3^{(2)} \\ \vdots \\ b_k^{(k-1)} \\ b_{k+1}^{(k-1)} \\ \vdots \\ b_n^{(k-1)} \end{pmatrix}$$

此时我们将第 k 列从第 $k+1$ 行到第 n 行化为 0，同时更新 b。那么我们就能通过 $a_{k+1,k+1}x_{k+1}=b_{k+1}^{(k)}$ 求得 x_{k+1}

$$
\begin{pmatrix}
a_{11} & & & & & & & \\
0 & a_{22} & & & & & & \\
0 & 0 & a_{33} & & & & & \\
\vdots & \vdots & \vdots & \ddots & & & & \\
0 & 0 & 0 & \dots & a_{kk} & & & \\
0 & 0 & 0 & \dots & 0 & a_{k+1k+1} & & \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \ddots & \\
0 & 0 & 0 & \dots & 0 & a_{nk+1} & \dots & a_{nn}
\end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}
=
\begin{pmatrix} b_1 \\ b_2^{(1)} \\ b_3^{(2)} \\ \vdots \\ b_k^{(k-1)} \\ b_{k+1}^{(k)} \\ \vdots \\ b_n^{(k)} \end{pmatrix}
$$

Algorithm 1 前代法

$x_1 = y_1/a_{11}$
for $i = 2$ **to** n **do**
  $s = y_i$
  for $j = 1, \ldots, i-1$ **do**
    $s = s - a_{ij}x_j$
  end for
  $x_i = s/a_{ii}$
end for

图 2: 前代法算法

前代法，就从前往后 (从 x_1 往 x_n) 依次求解。
回代法则恰好相反，他是从后往前依次求解。
回代法是用于上三角形的线性方程组求解。
同样我们要求其系数矩阵对角线上元素非 0.

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

与前代法类似，在回代法第 $(n-k+1)$ 个循环内。

$$
\begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1k-1} & a_{1k} & 0 & \cdots & 0 \\
 & a_{22} & \cdots & a_{2k-1} & a_{2k} & 0 & \cdots & 0 \\
 & & \ddots & & \vdots & \vdots & & \vdots \\
 & & & a_{k-1k-1} & a_{k-1k} & 0 & \cdots & 0 \\
 & & & & a_{kk} & 0 & \cdots & 0 \\
 & & & & & a_{k+1k+1} & \cdots & 0 \\
 & & & & & & \ddots & \vdots \\
 & & & & & & & a_{nn}
\end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}
=
\begin{pmatrix} b_1^{(n-k)} \\ b_2^{(n-k)} \\ \vdots \\ b_{k-1}^{(n-k)} \\ b_k^{(n-k)} \\ b_{k+1}^{(n-k-1)} \\ \vdots \\ b_n \end{pmatrix}
$$

此时我们将第 k 列从第 1 行到第 $k-1$ 行化为 0，同时更新 b。

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k-1} & 0 & 0 & \dots & 0 \\ & a_{22} & \dots & a_{2k-1} & 0 & 0 & \dots & 0 \\ & & \ddots & & \vdots & \vdots & & \vdots \\ & & & a_{k-1k-1} & 0 & 0 & \dots & 0 \\ & & & & a_{kk} & 0 & \dots & 0 \\ & & & & & a_{k+1k+1} & \dots & 0 \\ & & & & & & \ddots & \vdots \\ & & & & & & & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1^{(n-k+1)} \\ b_2^{(n-k+1)} \\ \vdots \\ b_{k-1}^{(n-k+1)} \\ b_k^{(n-k)} \\ b_{k+1}^{(n-k-1)} \\ \vdots \\ b_n \end{pmatrix}$$

Algorithm 2 回代法

$x_n = y_n / a_{nn}$

for $i = n - 1$ **to** 1 **do**

  $s = y_i$

  for $j = i + 1, \ldots, n$ **do**

    $s = s - a_{ij} x_j$

  end for

  $x_i = s / a_{ii}$

end for

图 3: 回代法算法

14.4 三角分解解满秩方阵方程组

如果线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的系数矩阵 $\boldsymbol{A}$ 可以进行 LU 分解，则我们可以对 $\boldsymbol{A}$ 先进行 LU 分解
令 $\boldsymbol{A}=\boldsymbol{LU}$, 我们则可以得到方程组 $\boldsymbol{LUx}=\boldsymbol{b}$ 令 $\boldsymbol{Ux}=\boldsymbol{y}$ 则我们可以得到两个线性方程组。

$$\boldsymbol{Ly}=\boldsymbol{b}$$

$$\boldsymbol{Ux}=\boldsymbol{y}$$

然后利用前代法求出 $\boldsymbol{y}$, 再利用回代法求解出原线性方程组的解 $\boldsymbol{x}$

因为在计算机上求解方程，我们还需要考虑资源问题，为了节约资源，下面给出一种紧凑的求解方式。

给定矩阵 $\boldsymbol{A}$ 和向量 $\boldsymbol{b}$，我们先对 $\boldsymbol{A}$ 进行 LU 分解。并且使用 $\boldsymbol{A}$ 的上三角部分存储上三角矩阵，用下三角部分存储下三角矩阵。

比如矩阵

$$\begin{pmatrix} 3 & 2 & -1 \\ 6 & 6 & -2 \\ -3 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

就可以使用

$$\begin{pmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 2 & -1 \end{pmatrix}$$

来存储。

然后我们来重新给出 LU 分解的算法流程。

$$u_{1i} = a_{1i}, \quad i = 1, \cdots, n;$$

$$l_{i1} = a_{i1}/u_{11}, i = 2, \cdots, n;$$

对 $k = 2, \cdots$,n, 计算

$$u_{ki} = a_{ki} - \sum_{r=1}^{k-1} l_{kr} u_{ri}, \quad i = k, \cdots, n;$$

$$l_{ik} = (a_{ik} - \sum_{r=1}^{k-1} l_{kr} u_{ri})/u_{kk}, \quad k = 2, \cdots, n.$$

最后再使用前代法和回代法求出最终的解。

例 6

求解 $$\begin{pmatrix} 3 & 2 & -1 \\ 6 & 6 & -2 \\ -3 & 2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \\ -5 \end{pmatrix}$$

我们先对 $\begin{pmatrix} 3 & 2 & -1 \\ 6 & 6 & -2 \\ -3 & 2 & 0 \end{pmatrix}$ LU 分解。

(i) $\begin{pmatrix} 3 & 2 & -1 \\ 2 & 6 & -2 \\ -1 & 2 & 0 \end{pmatrix}$

(ii) $\begin{pmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 4 & -1 \end{pmatrix}$

(iii) $\begin{pmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 2 & -1 \end{pmatrix}$

(iv) $\begin{pmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 2 & -1 \end{pmatrix}$

然后再进行前代法，注意此处，我们只关心 $\boldsymbol{y}$ 不关心 $\boldsymbol{L}$ 如何变换，故 $\boldsymbol{L}$ 并不需要实际上去化为 0。

$$\begin{pmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 2 & -1 \end{pmatrix}$$

(i) 前代法第一个循环

$$\begin{pmatrix} 0 \\ -2 \\ -5 \end{pmatrix}$$

(i) 前代法第二个循环

$$\begin{pmatrix} 0 \\ -2 \\ -1 \end{pmatrix}$$

最后进行回代法，

$$\begin{pmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 2 & -1 \end{pmatrix}$$

(i) 回代法第一个循环

$$\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

(i) 回代法第二个循环

$$\begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$$

(i) 回代法第三个循环

$$\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

便得解 $(1, -1, 1)^{\mathrm{T}}$

对于正定矩阵还可以使用 Cholesky 分解来进行求解。因为和 LU 分解类似，此处不再赘述。

敏度分析问题引入

考虑如下两组线性方程组：

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2.0001 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 0 \\ 1 & 1.0001 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2.0001 \end{bmatrix} \tag{17}$$

的解都是 $\boldsymbol{x} = (1,1)^{\mathrm{T}}$。

但是如果我们对方程的常数项做一点微小的变动，求解方程组

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 0 \\ 1 & 1.0001 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \tag{18}$$

前者的解为 $\boldsymbol{x} = (2,0)^{\mathrm{T}}$，而后者的解为 $\boldsymbol{x} = (1, \frac{10000}{10001})^{\mathrm{T}} \approx (1, 0.9999)^{\mathrm{T}}$。

可以看到左边方程的解变化的非常大，而右边方程的解几乎没有变化。

在本节中，我们将分析数据中小扰动对非奇异方阵线性方程解的影响。
我们将分别讨论输入的扰动对解的影响，系数矩阵的扰动对解的影响，以及输入和系数矩阵联合扰动对解的影响。
本节我们用 2 范数度量矩阵和向量的扰动程度，由于矩阵的 2-范数是和向量的 2-范数相容的，有以下性质：

(1) $\|\boldsymbol{A}\|_2 \geq 0(\|\boldsymbol{A}\|_2 = 0 \iff \boldsymbol{A} = \boldsymbol{O})$ （正定性）；
(2) $\|c\boldsymbol{A}\|_2 = |c|\|\boldsymbol{A}\|_2$，$c$ 为实数 （齐次性）；
(3) $\|\boldsymbol{A}+\boldsymbol{B}\|_2 \leq \|\boldsymbol{A}\|_2 + \|\boldsymbol{B}\|_2$ （三角不等式）；
(4) $\|\boldsymbol{A}\boldsymbol{B}\|_2 \leq \|\boldsymbol{A}\|_2\|\boldsymbol{B}\|_2$ （相容性）.

14.5.1 输入的扰动敏感性

令 $\boldsymbol{x}$ 为线性方程 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$ 的解，其中 $\boldsymbol{A}$ 为非奇异方阵，且 $\boldsymbol{y}\neq 0$。假设我们通过施加一个小的扰动项 $\Delta\boldsymbol{y}$ 来略微改变 $\boldsymbol{y}$，并将 $\boldsymbol{x}+\Delta\boldsymbol{x}$ 称为扰动方程组的解：

$$\boldsymbol{A}(\boldsymbol{x}+\Delta\boldsymbol{x})=\boldsymbol{y}+\Delta\boldsymbol{y}$$

我们的关键问题是：如果 $\Delta\boldsymbol{y}$ 比较小的时候，$\Delta\boldsymbol{x}$ 会不会同样比较小？

上面的公式看出，以及 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$ 来看，扰动 $\Delta\boldsymbol{x}$ 本身就是线性方程组的解。

$$\boldsymbol{A}\Delta\boldsymbol{x}=\Delta\boldsymbol{y}$$

并且，由于认为 $\boldsymbol{A}$ 是可逆的，我们可以写成

$$\Delta\boldsymbol{x}=\boldsymbol{A}^{-1}\Delta\boldsymbol{y}$$

$$\Delta \boldsymbol{x} = \boldsymbol{A}^{-1} \Delta \boldsymbol{y}$$

利用该方程两边的 2-范数得出

$$\|\Delta \boldsymbol{x}\|_2 = \left\|\boldsymbol{A}^{-1} \Delta \boldsymbol{y}\right\|_2 \leq \left\|\boldsymbol{A}^{-1}\right\|_2 \|\Delta \boldsymbol{y}\|_2$$

其中 $\left\|\boldsymbol{A}^{-1}\right\|_2$ 是 $\boldsymbol{A}^{-1}$ 的谱范数。

类似地，从 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ 得出 $\|\boldsymbol{y}\|_2 = \|\boldsymbol{A}\boldsymbol{x}\|_2 \leq \|\boldsymbol{A}\|_2 \|\boldsymbol{x}\|_2$，因此

$$\|\boldsymbol{x}\|_2^{-1} \leq \frac{\|\boldsymbol{A}\|_2}{\|\boldsymbol{y}\|_2}$$

将上面两个公式相乘，我们得到

$$\frac{\|\Delta \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq \left\|\boldsymbol{A}^{-1}\right\|_2 \|\boldsymbol{A}\|_2 \frac{\|\Delta \boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}$$

这个结果将“输入项” $\boldsymbol{y}$ 的相对变化与“输出” $\boldsymbol{x}$ 的相对变化联系起来。

定义 7

设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 是可逆矩阵，称数

$$\kappa(\boldsymbol{A}) = \left\|\boldsymbol{A}^{-1}\right\|_2 \|\boldsymbol{A}\|_2,$$

是矩阵 $\boldsymbol{A}$ 的条件数。

设 σ_1, σ_n 分别是矩阵 $\boldsymbol{A}$ 的最大奇异值和最小奇异值，那么

$$\|\boldsymbol{A}\|_2 = \sigma_1, \qquad \left\|\boldsymbol{A}^{-1}\right\|_2 = 1/\sigma_n$$

因此矩阵 $\boldsymbol{A}$ 的条件数也可以定义为：

$$\kappa(\boldsymbol{A}) = \frac{\sigma_1}{\sigma_n}, 1 \leq \kappa(\boldsymbol{A}) \leq \infty.$$

大的 $\kappa(\boldsymbol{A})$ 意味着 $\boldsymbol{y}$ 上的扰动可能导致 $\boldsymbol{x}$ 上有很大的扰动，即方程对输入数据的变化非常敏感。如果 $\boldsymbol{A}$ 是奇异的，那么 $\kappa = \infty$。非常大的 $\kappa(\boldsymbol{A})$ 表明 $\boldsymbol{A}$ 接近奇异；我们说在这种情况下 $\boldsymbol{A}$ 是病态的。

我们将以上讨论总结为如下引理：

引理 1

(对于输入的敏感性) 令 $\boldsymbol{A}$ 为非奇异方阵，$\boldsymbol{x}, \Delta\boldsymbol{x}$ 满足

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

$$\boldsymbol{A}(\boldsymbol{x} + \Delta\boldsymbol{x}) = \boldsymbol{y} + \Delta\boldsymbol{y}$$

有下式成立

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq \kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}$$

其中 $\kappa(\boldsymbol{A}) = \left\|\boldsymbol{A}^{-1}\right\|_2 \|\boldsymbol{A}\|_2$ 是矩阵 $\boldsymbol{A}$ 的条件数。

14.5.2 系数矩阵中的扰动敏感性

接下来我们考虑 $\boldsymbol{A}$ 矩阵的扰动对解 $\boldsymbol{x}$ 的影响。令 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ 并且令 $\Delta\boldsymbol{A}$ 为一个扰动，满足下面等式

$$(\boldsymbol{A} + \Delta\boldsymbol{A})(\boldsymbol{x} + \Delta\boldsymbol{x}) = \boldsymbol{y}, \quad \text{对于一些}\Delta\boldsymbol{x}$$

那么有

$$\boldsymbol{A}\Delta\boldsymbol{x} = -\Delta\boldsymbol{A}(\boldsymbol{x} + \Delta\boldsymbol{x})$$

因此 $\Delta\boldsymbol{x} = -\boldsymbol{A}^{-1}\Delta\boldsymbol{A}(\boldsymbol{x} + \Delta\boldsymbol{x})$。则

$$\|\Delta\boldsymbol{x}\|_2 = \left\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}(\boldsymbol{x} + \Delta\boldsymbol{x})\right\|_2 \le \left\|\boldsymbol{A}^{-1}\right\|_2 \|\Delta\boldsymbol{A}\|_2 \|\boldsymbol{x} + \Delta\boldsymbol{x}\|_2$$

并且

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x} + \Delta\boldsymbol{x}\|_2} \le \left\|\boldsymbol{A}^{-1}\right\|_2 \|\boldsymbol{A}\|_2 \frac{\|\Delta\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_2}$$

我们再次看到条件数出现在不等式中。只有在条件数不是太大时，相对较小的扰动，即 $\frac{\|\Delta\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_2} \ll 1$ 对 $\boldsymbol{x}$ 的相对影响才较小。也就是说，$\kappa(\boldsymbol{A}) \simeq 1$。这个会在下一个引理中总结。

引理 2

(系数矩阵中的扰动敏感性) 令 $\boldsymbol{A}$ 为非奇异方阵，$\boldsymbol{x}, \Delta\boldsymbol{A}, \Delta\boldsymbol{x}$ 满足

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

$$(\boldsymbol{A} + \Delta\boldsymbol{A})(\boldsymbol{x} + \Delta\boldsymbol{x}) = \boldsymbol{y}$$

有下式成立

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x} + \Delta\boldsymbol{x}\|_2} \le \kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_2}$$

14.5.3 对 A,y 联合扰动的敏感性

我们最后考虑 $\boldsymbol{A}$ 和 $\boldsymbol{y}$ 的同时扰动对 $\boldsymbol{x}$ 的影响。令 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$，并且令 $\Delta\boldsymbol{A},\Delta\boldsymbol{y}$ 为扰动，满足下面等式

$$(\boldsymbol{A}+\Delta\boldsymbol{A})(\boldsymbol{x}+\Delta\boldsymbol{x})=\boldsymbol{y}+\Delta\boldsymbol{y}, \text{对于一些}\Delta\boldsymbol{x}$$

然后，$\boldsymbol{A}\Delta\boldsymbol{x}=\Delta\boldsymbol{y}-\Delta\boldsymbol{A}(\boldsymbol{x}+\Delta\boldsymbol{x})$，因此 $\Delta\boldsymbol{x}=\boldsymbol{A}^{-1}\Delta\boldsymbol{y}-\boldsymbol{A}^{-1}\Delta\boldsymbol{A}(\boldsymbol{x}+\Delta\boldsymbol{x})$。则

$$\begin{aligned}\|\Delta\boldsymbol{x}\|_2&=\left\|\boldsymbol{A}^{-1}\Delta\boldsymbol{y}-\boldsymbol{A}^{-1}\Delta\boldsymbol{A}(\boldsymbol{x}+\Delta\boldsymbol{x})\right\|_2\\&\le\left\|\boldsymbol{A}^{-1}\Delta\boldsymbol{y}\right\|_2+\left\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}(\boldsymbol{x}+\Delta\boldsymbol{x})\right\|_2\\&\le\left\|\boldsymbol{A}^{-1}\right\|_2\|\Delta\boldsymbol{y}\|_2+\left\|\boldsymbol{A}^{-1}\right\|_2\|\Delta\boldsymbol{A}\|_2\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2.\end{aligned}$$

接着，上式除以 $\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2$，

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2}\le\left\|\boldsymbol{A}^{-1}\right\|_2\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}\frac{\|\boldsymbol{y}\|_2}{\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2}+\kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_2}$$

但是 $\|\boldsymbol{y}\|_2 = \|\boldsymbol{A}\boldsymbol{x}\|_2 \le \|\boldsymbol{A}\|_2\|\boldsymbol{x}\|_2$，因此

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2} \le \kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2} + \kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_2}$$

下一步，我们用 $\|\boldsymbol{x}\|_2 = \|\boldsymbol{x}+\Delta\boldsymbol{x}-\Delta\boldsymbol{x}\|_2 \le \|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2 + \|\Delta\boldsymbol{x}\|_2$ 去写

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2} \le \kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}\left(1+\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2}\right) + \kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_2}$$

从中我们得到

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2}\left(1-\kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}\right) \le \kappa(\boldsymbol{A})\left(\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}+\frac{\|\Delta\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_2}\right)$$

因此

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}+\Delta\boldsymbol{x}\|_2} \le \frac{\kappa(\boldsymbol{A})}{1-\kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}}\left(\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}+\frac{\|\Delta\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_2}\right)$$

扰动的“放大因子”受 $\frac{\kappa(\boldsymbol{A})}{1-\kappa(\boldsymbol{A})\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}}$ 的制约，该值小于某个给定的 γ，如果

$$\kappa(\boldsymbol{A}) \le \frac{\gamma}{1+\gamma\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}}$$

因此，我们看到联合扰动的影响仍然由 $\boldsymbol{A}$ 的条件数控制，如下所述

引理 3

(对 A,y 扰动的敏感性) 令 $\boldsymbol{A}$ 为非奇异方阵，令 $\gamma > 1$ 已知，并且令 $\boldsymbol{x}, \Delta\boldsymbol{y}, \Delta\boldsymbol{A}, \Delta\boldsymbol{x}$ 满足下面等式

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

$$(\boldsymbol{A} + \Delta\boldsymbol{A})(\boldsymbol{x} + \Delta\boldsymbol{x}) = \boldsymbol{y} + \Delta\boldsymbol{y}$$

如果

$$\kappa(\boldsymbol{A}) \leq \frac{\gamma}{1 + \gamma\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2}}$$

那么

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x} + \Delta\boldsymbol{x}\|_2} \leq \gamma\left(\frac{\|\Delta\boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2} + \frac{\|\Delta\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_2}\right)$$

本节小节

本节的主要内容有：

(1) 回顾了线性方程组的基本概念，并对线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$, $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 进行分类。

当 $\mathrm{rank}(\boldsymbol{A})=\min\{m,n\}$ 时，可以把方程组分为以下三类：

- 满秩方阵方程组
- 列满秩超定方程组
- 行满秩欠定方程组

对 $\mathrm{rank}(\boldsymbol{A})<\min\{m,n\}$ 的情况，我们不进行讨论。

(2) 对于满秩方阵方程组，我们可以采用三角分解，并分别利用前代法和回代法进行求解。

(3) 我们对满秩方阵方程组的输入和系数矩阵变化对解的影响进行了敏度分析。

在下一节，我们将学习如何处理列满秩超定方程组和行满秩欠定方程组的情况。