

# 线性代数预习学案

TechX 机器学习课程学术团队

June 2021

## 1 导论

作为数学的一个分支，线性代数被广泛应用于科学和工程中。学习线性代数，是理解和实现机器学习算法的基础。因此，我们为即将在这个暑假参加 TechX 机器学习课程的同学准备了这份预习学案，希望大家初次学习线性代数提供一些引导。这份学案共包含 6 个小节，除第一小节外每一小节为两天的预习内容，可在一到两小时内学习完成。多年来，互联网上积累了大量易于理解的线性代数学习材料。因此，本学案在对重要概念进行阐述的同时，也包含了一些链接，帮助大家快速找到与概念相关的优质线上资源加深理解。每一小节都在最开始列出了这一小节中的重点概念。对于基础较好的同学，可以根据自己对小节中重点概念的掌握情况，有选择地跳过一些小节。

希望这份学案能够帮助大家敲开线性代数的大门。当然，本学案绝对称不上详尽，而只是把与在八月份的课程中将涉及的机器学习算法高度相关的线性代数知识做了概述。因此，学有余力的同学可以结合一本教材进行深入学习。我们推荐麻省理工大学 Gilbert Strang 教授撰写的 [Introduction to Linear Algebra](#) 和他的[公开课](#)。

## 2 标量和向量

- 每个**标量**就是一个实数，两个标量间的运算即为两个实数间的运算。此外，标量还可与向量、矩阵相乘。
- **向量**，即“有方向的量”，可以表示为一个数组。向量可以与标量相乘，还可以与另一向量进行点乘。同时，向量也可被看成一个行数或列数为 1 的矩阵（矩阵将在第 3 小节中涉及），满足其所有性质。

### 2.1 标量

每个**标量**就是一个实数，两个标量间的运算即为两个实数间的运算。此外，标量还可与向量、矩阵相乘。

### 2.2 平面中的向量

**向量**，即“有方向的量”，可以表示为一个数组。在平面直角坐标系中，一个向量可以被几何表示为由原点出发的有向线段，如图 1 所示。我们可以用该向量终点 A 的横纵坐标来表示这个向量，即：

$$\mathbf{a} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \text{ 或行向量 } [2, 3]$$

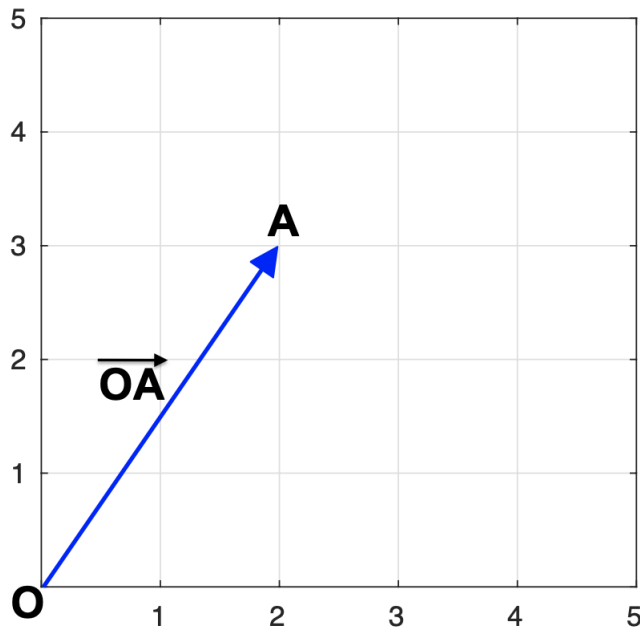


图 1: 平面中的向量

### 2.3 二维向量的运算

设  $\mathbf{u} = \begin{bmatrix} u_1 & u_2 \end{bmatrix}$ ,  $\mathbf{v} = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$  和任意标量  $c$ , 我们定义:

- 两向量的和  $\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 & u_2 + v_2 \end{bmatrix}$ .
- 标量和向量的乘积  $c\mathbf{u} = \begin{bmatrix} cu_1 & cu_2 \end{bmatrix}$ .
- 两向量的点乘积  $\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2$ .

### 2.4 高维空间中的向量

在机器学习中, 我们所遇到的往往是高维空间中的向量。例如, 在人脸识别技术中, 输入是一张图片, 包含上百个像素, 每个像素又是由红、蓝、绿三元色组成。如果我们把每个像素中的每个元素看成是向量中的一维, 则输入向量中包含成千上万个维度。向量的运算从二维很自然地引申至高维:

设  $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$ ,  $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ , 则

- 两向量的和  $\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 & u_2 + v_2 & \dots & u_n + v_n \end{bmatrix}$ .
- 标量和向量的乘积  $c\mathbf{u} = \begin{bmatrix} cu_1 & cu_2 & \dots & cu_n \end{bmatrix}$ .
- 两向量的点乘积  $\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \dots + u_nv_n$ .

### 2.5 预习拓展

- [Bilibili 视频](#): 著名 up 主 3BLUE1BROWN 的线性代数小讲堂, 直观地阐述了向量究竟是什么。
- [MIT 线性代数公开课 Lecture 1](#): 全英文, 对本小节的概念进行了更深入的探究, 并辅以其几何意义。

## 3 矩阵

- 每个**矩阵**可以被视为一个二维数组。矩阵的每一行可被看成一个行向量，每一列可被看成一个列向量。每个行向量/列向量可被看成一个行数/列数为 1 的矩阵。
- 两矩阵加减要求两矩阵拥有相同的维度。
- 矩阵可与一标量或另一矩阵相乘，与另一矩阵相乘时要求另一矩阵的行数等于此矩阵的列数。
- 矩阵的转置操作将矩阵的行和列进行了翻转。

### 3.1 矩阵

矩阵是一个长方形形状的二维数组，由行和列组成。我们称矩阵的行数和列数为该矩阵的维度。若矩阵的行数和列数相等，我们则称该矩阵为方阵。例如：

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \text{不是方阵}, B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{是方阵}$$

我们用  $a_{r,c}$  表示矩阵  $A$  第  $r$  行第  $c$  列的数，例如： $a_{1,1} = 1, b_{2,1} = 3$ 。

### 3.2 矩阵的加减

设矩阵  $A$  和矩阵  $B$  的维度均为  $m$  行  $n$  列：

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \ddots & & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix}, B = \begin{bmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,n} \\ b_{2,1} & b_{2,2} & \dots & b_{2,n} \\ \vdots & \ddots & & \vdots \\ b_{m,1} & b_{m,2} & \dots & b_{m,n} \end{bmatrix}$$

则：

$$A + B = \begin{bmatrix} a_{1,1} + b_{1,1} & a_{1,2} + b_{1,2} & \dots & a_{1,n} + b_{1,n} \\ a_{2,1} + b_{2,1} & a_{2,2} + b_{2,2} & \dots & a_{2,n} + b_{2,n} \\ \vdots & \ddots & & \vdots \\ a_{m,1} + b_{m,1} & a_{m,2} + b_{m,2} & \dots & a_{m,n} + b_{m,n} \end{bmatrix}$$

$$A - B = \begin{bmatrix} a_{1,1} - b_{1,1} & a_{1,2} - b_{1,2} & \dots & a_{1,n} - b_{1,n} \\ a_{2,1} - b_{2,1} & a_{2,2} - b_{2,2} & \dots & a_{2,n} - b_{2,n} \\ \vdots & \ddots & & \vdots \\ a_{m,1} - b_{m,1} & a_{m,2} - b_{m,2} & \dots & a_{m,n} - b_{m,n} \end{bmatrix}$$

**注意：两矩阵仅能在维度相同时进行加减运算！**

### 3.3 矩阵乘法

设矩阵  $A$  的维度为  $m$  行  $n$  列，矩阵  $B$  的维度为  $n$  行  $p$  列， $A$  的列数等于  $B$  的行数：

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \ddots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}, B = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,p} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,p} \\ \vdots & \ddots & & \vdots \\ b_{n,1} & b_{n,2} & \cdots & b_{n,p} \end{bmatrix}$$

则两矩阵的乘积  $AB$  是一个维度为  $m$  行  $p$  列的矩阵，且  $(AB)_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,n}b_{n,j}$ 。

**注意：**两矩阵仅能在第一矩阵的列数等于第二矩阵的行数时进行乘法运算，且矩阵乘法不满足交换律！

### 3.4 矩阵的转置

设矩阵  $A$  的维度为  $m$  行  $n$  列， $A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \ddots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$ ，则矩阵  $A$  的转置（记为  $A^T$ ）将矩阵  $A$  的

行和列进行了翻转，得到一个维度为  $n$  行  $m$  列的矩阵：

$$A^T = \begin{bmatrix} a_{1,1} & a_{2,1} & \cdots & a_{m,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{m,2} \\ \vdots & \ddots & & \vdots \\ a_{1,m} & a_{2,m} & \cdots & a_{m,m} \end{bmatrix}$$

### 3.5 练习

想要熟练掌握矩阵的乘法运算，练习是必不可少的。那么，大家赶紧找来纸笔，准备好迎接挑战吧！

点击[此处](#)开始练习。

## 4 线性方程组和行阶梯形

### 4.1 线性方程组

线性方程即为一次方程式（在方程式中，未知数的次数为1），其一般形式为  $a_1x_1 + a_2x_2 + \cdots + a_nx_n + b = 0$ ，其中  $x_1, x_2, \dots, x_n$  为未知数， $a_1, a_2, \dots, a_n, b$  为常数。线性方程的解集为多维空间中的一个超平面。超平面是平面在高维空间中引申出的概念。若某一线性方程中包含3个未知数，则其解集就是三维空间中的一个平面。

两个或以上的线性方程即构成线性方程组。可以想见线性方程组可以有：

- **0个解：**例如两个线性方程的解集平面平行，则三维空间中任意一点均无法同时成为两个线性方程的解，因此线性方程组无解。
- **1个解：**例如三个线性方程的解集平面有且仅有一个公共点，则三维空间中仅有这一点可以同时成为三个线性方程的解，因此线性方程组有唯一解。
- **无穷多解：**例如两个线性方程的解集平面相交产生一条交线，则这条交线上的任意一点都是这两个线性方程的解，因此线性方程组有无穷多解。

大家应该已经清楚线性方程组的解法：对线性方程组中的每一个方程式，我们用其将其中一个未知数消掉（这一过程被称为“消元”），代入到其他还未用过的线性方程中；解出一个未知数的值后，我们再将它代入之前的式子中解得更多的未知数。

## 4.2 线性方程组的矩阵表示法

每个线性方程组均可用矩阵和一行向量相乘的形式表示。假设我们有

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = b_2 \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m \end{cases} \quad (1)$$

则根据矩阵乘法的定义，令  $A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \ddots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$ ,  $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ ,  $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ , 该线性方程组等价于

$$A\mathbf{x} = \mathbf{b} \quad (2)$$

## 4.3 行阶梯形矩阵

我们希望找到解形如  $A\mathbf{x}=\mathbf{b}$  的方程式的一般方法。因此，我们希望矩阵  $A$  呈行阶梯形。

**行阶梯形矩阵：**行阶梯形矩阵满足矩阵每一行第一个非零数均严格在上面一行第一个非零数的右边。例如，矩阵  $P = \begin{bmatrix} 2 & 7 & -3 \\ 0 & 4 & -2 \\ 0 & 0 & 1 \end{bmatrix}$  和  $Q = \begin{bmatrix} 2 & 7 & -3 \\ 0 & 4 & -2 \\ 0 & 0 & 0 \end{bmatrix}$  和  $R = \begin{bmatrix} 2 & 7 & -3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$  均为行阶梯形矩阵，而矩阵  $S = \begin{bmatrix} 2 & 7 & -3 \\ 0 & 0 & 1 \\ 0 & 3 & 0 \end{bmatrix}$  则不是，因为  $S$  第三行的 3 在第二行的 1 左边而非右边。

**行阶梯形矩阵有什么好处：**我们将通过一个例子说明行阶梯形矩阵对我们解  $A\mathbf{x}=\mathbf{b}$  的帮助。设

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{2,2} & a_{2,3} \\ 0 & 0 & a_{3,3} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

则由第三行的  $a_{3,3}x_3 = b_3$  可求得  $x_3$ ；将求得的  $x_3$  代入由第二行的  $a_{2,2}x_2 + a_{2,3}x_3 = b_2$  可求得  $x_2$ ；再将求得的  $x_2$  和  $x_3$  代入由第一行的  $a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3 = b_1$  可求得  $x_1$ 。由此可见，相比于一般矩阵，当  $A$  为行阶梯形矩阵时，我们能更容易地得到线性方程组的解，所用到的方法是将未知数**从下到上**依次解得。

## 4.4 行阶梯形矩阵和线性方程组解的个数

之前已经提到，线性方程组可能有 0 个解、1 个解或无穷多解。那么，当矩阵  $A$  为行阶梯形时，我们如何判断方程组  $A\mathbf{x}=\mathbf{b}$  解的个数呢？我们考虑如下几个例子：

$$\bullet A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix} : \text{此时, 由第三行可唯一确定 } x_3 \text{ 的值, 由第二行 (和 } x_3 \text{ 的值) 可唯一确定 } x_2$$

的值, 由第一行 (和  $x_2, x_3$  的值) 可唯一确定  $x_1$  的值。因此方程组有唯一解。

$$\bullet A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix} : \text{此时, 由第三行可知 } 0 = 7, \text{ 显然不成立。因此方程组无解。}$$

$$\bullet A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 5 \\ 6 \\ 0 \end{bmatrix} : \text{此时, 由第三行可知 } 0 = 0, \text{ 显然成立。对于任意给定的 } x_3, \text{ 由第二行可确}$$

定对应的  $x_2$  的值, 再由第一行可确定对应的  $x_1$  的值。因此方程组有无穷多解。

我们可以总结出行阶梯形矩阵与线性方程组解的个数的关系: 若矩阵中包含全为零的行, 且对应的  $\mathbf{b}$  中的值不为零, 则方程组无解; 否则, 若矩阵中不全为零的行数等于未知数的个数 (即矩阵的列数), 则方程组有唯一解; 若非以上两种情况, 则方程组有无穷多解。

## 5 高斯消元

- $A\mathbf{x}=\mathbf{b}$  的增光矩阵为  $[A|\mathbf{b}]$ 。
- 初等行变换包括: 交换两行的位置; 用一个非零的数乘某一行; 把其中一行的若干倍加到另一行上。对于一个线性方程组的增广矩阵, 初等行变换不改变该线性方程组的解。
- 高斯消元是将一个矩阵通过若干次初等行变换化为行阶梯形矩阵的过程。

### 5.1 增广矩阵

对于一个线性方程组  $A\mathbf{x}=\mathbf{b}$ , 我们定义它的增广矩阵为  $[A|\mathbf{b}]$ 。例如: 若  $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix}$ , 则增

$$\text{广矩阵为 } A = \left[ \begin{array}{ccc|c} 1 & 2 & 3 & 5 \\ 0 & 4 & 5 & 6 \\ 0 & 0 & 0 & 7 \end{array} \right]。$$

### 5.2 初等行变换

为了得到线性方程组  $A\mathbf{x}=\mathbf{b}$  的解, 我们希望将  $A$  转化为行阶梯形。但在转化过程中, 我们需要方程组的解不变。

问: 在解多元一次方程组时, 我们怎样能在消元时保证方程组的解不变呢?

答: 我们可以交换两个方程式、将一个方程式等号左右两端各乘上一个相同的非零常数或将一个方程式的若干倍加至另一个方程式上。

这三种变换恰恰对应了矩阵的三种初等行变换。对于增广矩阵:

- 交换两个方程式：交换两行的位置；
- 将一个方程式等号左右两端各乘上一个相同的非零常数：用一个非零的数乘某一行；
- 将一个方程式的若干倍加至另一个方程式上：把其中一行的若干倍加到另一行上。

我们称交换两行的位置、用一个非零的数乘某一行和把其中一行的若干倍加到另一行上的变换为初等行变换。

### 5.3 高斯消元

高斯消元，是通过初等行变换把增广矩阵变为行阶梯形矩阵来求解线性方程组的过程。高斯消元的基本思路是，对于每个未知数  $x_i$ ，找到一个  $x_i$  的系数非零，但  $x_1$  到  $x_{i-1}$  的系数都是零的方程式，然后用初等行变换将其他方程式的  $x_i$  的系数全部消成零。

解下来，我们通过一个例子说明高斯消元的过程。设增广矩阵为  $A = \left[ \begin{array}{ccc|c} 1 & 2 & -1 & -6 \\ 2 & 1 & -3 & -9 \\ -1 & -1 & 2 & 7 \end{array} \right]$

- 第 1 步：将第一行的 -2 倍加到第二行上，得到  $\left[ \begin{array}{ccc|c} 1 & 2 & -1 & -6 \\ 0 & -3 & -1 & 3 \\ -1 & -1 & 2 & 7 \end{array} \right]$ ；

- 第 2 步：将第一行加到第三行上，得到  $\left[ \begin{array}{ccc|c} 1 & 2 & -1 & -6 \\ 0 & -3 & -1 & 3 \\ 0 & 1 & 1 & 1 \end{array} \right]$ ；

- 第 3 步：交换第二行和第三行，得到  $\left[ \begin{array}{ccc|c} 1 & 2 & -1 & -6 \\ 0 & 1 & 1 & 1 \\ 0 & -3 & -1 & 3 \end{array} \right]$ ；

- 第 4 步：将第二行的 3 倍加到第三行上，得到  $\left[ \begin{array}{ccc|c} 1 & 2 & -1 & -6 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 6 \end{array} \right]$ 。

此时，矩阵已被转化为行阶梯形。我们即可由  $x_3$  至  $x_1$  的顺序由下到上求得线性方程组的解。

当然，在高斯消元的过程中，有可能找不到一个  $x_i$  的系数非零，但  $x_1$  到  $x_{i-1}$  的系数都是零的方程式。若遇到这种情况，我们可以直接跳过  $x_i$ ，继续对  $x_{i+1}$  进行消元，最终也能得到行阶梯形矩阵。

### 5.4 预习拓展

- [MIT 线性代数公开课 Lecture 2](#): 全英文，对 “Matrix Elimination” 进行了更深入的探究，并结合了许多例子加深理解。

## 6 逆矩阵

- 单位矩阵（记作  $I$ ）是对角线上的数均为 1，其余位置上均为 0 的方阵；

- 矩阵  $A$  的逆矩阵记作  $A^{-1}$ ，满足  $AA^{-1} = I$ ；
- 用高斯消元可以求得一个矩阵的逆矩阵（如果存在）。

## 6.1 单位矩阵与逆矩阵

我们通常把 1 称作“单位”，是因为 1 乘任何实数的结果都为那个实数。在线性代数中，我们把乘任何一个矩阵（当然，我们需要两矩阵的维度满足矩阵乘法的要求）的结果都为那个矩阵的矩阵称为单位矩阵。单位矩阵行数和列数相等，因此是一个方阵；且其对角线上的数均为 1，其余位置上均为 0。我们将  $n$  行  $n$  列的单位矩阵记为  $I_n$ ：

$$\text{单位矩阵 } I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

需要注意的是，矩阵乘法并不满足交换律。但我们很容易得到结论：无论一个矩阵从左边还是右边乘单位矩阵，所得结果均与该矩阵相同。因此，我们可以直接说把“乘任何一个矩阵的结果都为那个矩阵”的矩阵称为单位矩阵，而无需指定单位矩阵在乘号的左边或右边。

同时，我们定义矩阵  $A$  的逆矩阵（记为  $A^{-1}$ ）为满足  $AA^{-1} = I$  且  $A^{-1}A = I$  的矩阵  $A^{-1}$ 。容易得到，一个矩阵的逆矩阵如果存在必唯一。

## 6.2 高斯消元求逆矩阵

在前面的章节中，大家已经学会了如何利用高斯消元算法求解线性方程组。利用同样的办法，我们可以求出一个矩阵的逆矩阵（如果存在）。考虑  $AA^{-1} = I$ ：

根据矩阵乘法的定义，单位矩阵的第一列由矩阵  $A$  乘矩阵  $A^{-1}$  的第一列得到。由此我们得到了一个线性方程组，并可用之前的高斯消元法求得它的解。我们可以一次求出矩阵  $A$  的逆矩阵的第一列、第二列一直到第  $n$  列，并由此得到矩阵  $A^{-1}$ 。

特别地，如果在求解时发现某个线性方程组无解（这种情况下矩阵  $A$  的行阶梯形矩阵有一行全为 0），则矩阵  $A$  的逆矩阵不存在。