3 Gaussian elimination (row reduction)

Let A be an $n \times k$ matrix and $\mathbf{b}$ is an $n \times 1$ vector. We wish to solve the equation

$$A\mathbf{x} = \mathbf{b} \tag{3.1}$$

where $\mathbf{x} \in \mathbf{R}^k$. One can write it out as follows

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1k}x_k & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2k}x_k & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \dots & + & a_{nk}x_k & = & b_n \end{array} \tag{3.2}$$

with

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \qquad \text{and} \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

but we will use the concise matrix notation. It is expected that everybody can easily move back and forth between the matrix notation (3.1) and (3.2) when needed. If you are uncertain, write up a few examples for yourself. It is important that the variables follow each other in the same order in each equation of (3.2). For example if the equations are

$$2x_3 - 5x_2 = 3$$

$$x_2 + 5x_1 - x_4 = 5$$

then first you should rewrite them in an ordered way as

$$\begin{array}{ccccccccc} & - & 5x_2 & + & 2x_3 & & & = & 3 \\ 5x_1 & + & x_2 & & & - & x_4 & = & 5 \end{array}$$

before you read off the matrix A

$$A = \begin{pmatrix} 0 & -5 & 2 & 0 \\ 5 & 1 & 0 & -1 \end{pmatrix}$$

In a first course the Gaussian elimination is usually presented in the language of the equations, but here we proceed faster, and we omit the x_j variables, and work immediately with the "augmented" matrix $[A|\mathbf{b}]$. The goal is to bring this matrix into an echelon form by **elementary operations**, which include

- Swapping rows
- Multiplying any row by a nonzero number
- Subtracting any row from any other one (but leave the first mentioned row unchanged)

The last two steps are usually used together, so some books gives the following list of elementary operations:

- Swapping rows
- Subtracting the nonzero multiple of any row from any other one. (but leave the first mentioned row unchanged)

These two sets are equivalent operations, except that the second one does not produce ones in the pivot positions, just some nonzero numbers. But this minor difference is irrelevant; the upper triangular system can easily be backsolved in both cases and of course the result is the same. For definiteness, here we use the first set of operations, i.e. we insist on ones in the pivot positions (if not one, then divide the whole row by the pivot). This is not the convention that [Demko] uses (see Section 8), however in several sections of 1502 this convention was used.

There is a slightly more important inconsistency. Unfortunately there are two almost identical echelon forms in the literature and even the books [Demko] and [HH] use different convention. The difference is whether one eliminates the nonzero elements above the pivots as well or not. This is not a very important issue, but it can be misleading, so let me fix it. I will use the terminology borrowed from Prof. Springarn's MATH1502 lecture. If you attended a different lecture, adjust your terminology.

Definition 3.1 *A matrix M is in* **row-echelon** *form if*

- *The first nonzero entry in every nonzero row is a one (we refer to them as pivot elements or leading ones). The corresponding x_j variables are called pivot variables.*
- *The leading ones progress to the right*
- *All entries in the same column and* **below** *a leading one are zero*
- *Any all-zero rows are grouped at the bottom of the matrix*

I assume that everybody knows how to use the elementary operations to bring any matrix into the row-echelon form (see [D] Sec 1.2, 1.3 and Appendix 8 for details). Just in nutshell, recall that you have to start the process from the left with the first column. Choose a nonzero element (pivot) in the first column, bring it into the first row by swapping two rows (if needed) and divide the first row through by this pivot, so that you get 1. Then use this 1 to eliminate all nonzero entries in the first column below this 1 by substracting appropriate multiple of the first row from each other rows. The resulting matrix has only the leading one in the first column. From now on you will always leave the first row untouched. Now you proceed to the second column. Choose a nonzero pivot element in this column among the elements in the second row or below (you forget about the first row!). If there is no such element, don't do anything with the second column. If there is one, then bring it into the second row by swapping, divide the second row by it so that you see the leading one and use this 1 to eliminate all nonzero elements *below* it. Do not go up to kill the possible nonzero element in the first row above the pivot. Then continue with the third column etc.

Once $M = [A|\mathbf{b}]$ is transformed into row-echelon form, it can be backsolved. Recall the following facts:

- The equation $A\mathbf{x} = \mathbf{b}$ has no solution if there is a pivot in the last column of the row echelon form of $[A|\mathbf{b}]$ (indicating that there is a row $0 \cdot x_1 + \ldots + 0 \cdot x_k = 1$ in the reduced system, which excludes any solution). For this purpose you can equally consider the reduced row-echelon form.

- The full solution of the equation can be obtained by choosing all the non-pivot variables free. Then we can backsolve all pivot variables in terms of the nonpivot variables. One has to proceed backwards.

When backsolving, we first express each pivot variable in terms of free variables and pivot variables of higher index. But when backsolving, those pivot variables are already expressed in terms of the free variables, hence eventually this pivot variable can also be expressed in terms of the free variables. However, this requires some extra calculation. It is more convenient if we compress this calculation as well into the reduction procedure. This leads to

Definition 3.2 *A matrix M is in* **reduced row-echelon** *form if*

- *The first nonzero entry in every nonzero row is a one (we refer to them as pivot elements or leading ones). The corresponding x_j variables are called pivot variables.*

- *The leading ones progress to the right*

- *Every leading one is the only nonzero element in its column*

- *Any all-zero rows are grouped at the bottom of the matrix*

The only difference between the row-echelon and the reduced row-echelon form is that in the latter we use the leading ones to eliminate the nonzero elements above them as well. It has the advantage, that when backsolving, each pivot variable is immediately expressed in terms of the nonpivot (free) variables.

The book [D] considers the row-echelon matrices as the final result of the elimination. However, [D] is not very clear, see the Appendix 8 which cleans it up and it can be read independently.

The book [HH] considers the reduced row-echelon form as the final result and calls it "echelon" form.

Here is an example of these two echelon forms. The matrix

$$\left(\begin{array}{cccc|c} \underline{1} & 2 & 3 & 4 & 3 \\ 0 & \underline{1} & 2 & 1 & 2 \\ 0 & 0 & 0 & \underline{1} & 6 \end{array}\right)$$

is in row-echelon form but not in row reduced echelon form (the pivots are underlined). The pivot variables are x_1, x_2, x_4, as the pivot columns are the first, second and fourth ones. The only free variable is x_3. When backsolving, the last row implies $x_4 = 6$. Then the last but one row gives

$$x_2 = 2 - 2x_3 - x_4$$

It is because there remained a 1 in the fourth column and second row, we see x_4 in the expression for x_2. We have to substitute the value $x_4 = 6$ to get

$$x_2 = -4 - 2x_3$$

Finally, from the first pivot we see

$$x_1 = 3 - 2x_2 - 3x_3 - 4x_4$$

Again, we have to substitute the previously obtained results for x_4 and x_2 to get

$$x_1 = 3 - 2(-4 - 2x_3) - 3x_3 - 24 = -13 + x_3$$

Eventually we have

$$\begin{aligned} x_1 &= -13 + x_3 \\ x_2 &= -4 - 2x_3 \\ x_3 &= x_3 \\ x_4 &= 6 \end{aligned}$$

as the full solution. The funny line $x_3 = x_3$ indicates that x_3 is a free variable. Of course you can rewrite the solution into the vector form

$$\mathbf{x} = \begin{pmatrix} -13 \\ -4 \\ 0 \\ 6 \end{pmatrix} + t \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} \tag{3.3}$$

where t is any real number.

Now what if we reduce the matrix further into the reduced row echelon form? This means first subtracting appropriate multiple of the last pivot row from all other rows to eliminate the nonzero elements above it, then continue with the last but one pivot row etc. (The order is important, you must start from the end). This gives

$$\left(\begin{array}{cccc|c} \underline{1} & 2 & 3 & 4 & 3 \\ 0 & \underline{1} & 2 & 1 & 2 \\ 0 & 0 & 0 & \underline{1} & 6 \end{array}\right) \Longrightarrow \left(\begin{array}{cccc|c} \underline{1} & 0 & -1 & 0 & -13 \\ 0 & \underline{1} & 2 & 0 & -4 \\ 0 & 0 & 0 & \underline{1} & 6 \end{array}\right)$$

and the solution (3.3) can be read off immediately without further calculations.

A USEFUL REMARK: Suppose you have to solve the equation $A\mathbf{x} = \mathbf{b}$ several times for various $\mathbf{b}$ vectors, but for the same A matrix (for example this problem will arise at computing the inverse matrix in Section 4.6). Notice that the elimination steps are determined purely by the matrix A. One has to perform the prescribed steps on the vector $\mathbf{b}$ as well, but it influences only the last column of the echelon matrices. Hence it is advisable to keep the echelon form of A for further use and also one should record the steps made, since the same steps will have to be performed on each vector $\mathbf{b}$. For more details, see the Section 8.2 in the Appendix.

SUMMARY: In this section we gave an algorithm how to solve $A\mathbf{x} = \mathbf{b}$ in full generality, i.e. decide if it is solvable and find all possible solutions. In the future this algorithm will be the key behind several theorems. Here let us just draw one of the most important conclusion which was actually needed in Lemma 2.12 when we proved the existence of a basis:

Theorem 3.3 *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k$ be vectors in $\mathbf{R}^n$. If $k > n$ (i.e. you have more vectors than their size), then these vectors are linearly dependent.*

Proof: We want to show that the zero vector can be written as a nontrivial linear combination of the given vectors. Form the $n \times k$ matrix $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_k \end{bmatrix}$ from these vectors by putting them into thte columns of A. A nontrivial linear combination means that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_k\mathbf{v}_k = \mathbf{0}$$

for some nonzero vector $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$. In other words it means that the equation $A\mathbf{x} = \mathbf{0}$ has a nonzero solution (recall that a vector is nonzero if at least one of its entry is nonzero). But this is clear from the Gaussian elimination: since you have more columns than rows, there is always a free variable, which you can choose to be nonzero to ensure that the whole vector $\mathbf{x}$ is not the zero vector. $\square$