for some coefficient-vector a. On the other hand, Π_k can also be defined as $\operatorname{null} D^{k+1}$, i.e., as the collection of all real-valued functions that are $k+1$-times continuously differentiable and have their $(k+1)$st derivative identically zero.

(2.16) Remark: The nullspace $\operatorname{null} A$ of the linear map $A : X \to Y$ consists exactly of the solutions to the *homogeneous* equation

$$A? = 0.$$

The linear equation $A? = y$ is readily associated with a *homogeneous* linear equation, namely the equation

$$[A, y]? = 0,$$

with

$$[A, y] : X \times \mathbb{F} : (z, \alpha) \mapsto Az + y\alpha.$$

If $Ax = y$, then $(x, -1)$ is a nontrivial element of $\operatorname{null}[A, y]$. Conversely, if $(z, \alpha) \in \operatorname{null}[A, y]$ *and* $\alpha \neq 0$, then $z/(-\alpha)$ is a solution to $A? = y$. Hence, for the construction of solutions to linear equations, it is sufficient to know how to solve *homogeneous* linear equations, i.e., how to construct the nullspace of a linear map.

2.15 For each of the following systems of linear equations, determine A and y of the equivalent vector equation $A? = y$.
(a) $\begin{array}{rcrcr} 2x_1 & - & 3x_2 & = & 4 \\ 4x_1 & + & 2x_2 & = & -6 \end{array}$; (b) $\begin{array}{rcrcr} 2u_1 & - & 3u_2 & = & 4 \\ 4u_1 & + & 2u_2 & = & -6 \end{array}$; (c) $\begin{array}{rcrcrcr} & & & & -4c & = & 16 \\ 2a & + & 3b & & & = & 9 \end{array}$.

2.16 For each of the following A and y, write out a system of linear equations equivalent to the vector equations $A? = y$.
(a) $A = \begin{bmatrix} 2 & 3 \\ 6 & 4 \\ \mathrm{e} & -2 \end{bmatrix}$, $y = (9, -\sqrt{3}, 1)$; (b) $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}$, $y = (10, 10)$. (c) $A = [] \in \mathbb{R}^{0\times 3}$, $y = () \in \mathbb{R}^0$.

Inverses

We have agreed to think of the *matrix* $A \in \mathbb{F}^{m\times n}$ as the *column map* $[A(:, 1), \ldots, A(:, n)]$, i.e., as the linear map $\mathbb{F}^n \to \mathbb{F}^m : a \mapsto Aa := \sum_j A(:, j)a_j$. For this reason, it is also customary to refer to the range $\operatorname{ran} A$ of a matrix A as the **column space** of that matrix, while the range $\operatorname{ran} A^{\mathrm{t}}$ of its transpose is known as its **row space**. Further, we have found that, in these terms, the *matrix product* AB is also the *composition* $A \circ B$, i.e.,

$$(A \circ B)a = A(B(a)) = (AB)a = \sum_j (AB)(:, j)a_j.$$

In these terms, the identity map id_n on $\mathbb{F}^n$ corresponds to the **identity matrix** $[e_1, e_2, \ldots, e_n]$, hence the name for the latter.

(2.17) Proposition: The inverse of a linear map is again a linear map.

Proof: Let $A \in L(X, Y)$ be invertible and $y, z \in Y$. By additivity of A, $A(A^{-1}y + A^{-1}z) = A(A^{-1}y) + A(A^{-1}z) = y + z$. Hence, applying A^{-1} to both sides, we get $A^{-1}y + A^{-1}z = A^{-1}(y + z)$, thus A^{-1} is additive. Also, from $A(\alpha A^{-1}y) = \alpha A(A^{-1}y) = \alpha y$, we conclude that $\alpha A^{-1}y = A^{-1}(\alpha y)$, hence A^{-1} is homogeneous. □

Thus, if $A \in \mathbb{F}^{n\times n}$ is invertible (as a linear map from $\mathbb{F}^n$ to $\mathbb{F}^n$), then also its inverse is a linear map (from $\mathbb{F}^n$ to $\mathbb{F}^n$), hence a square matrix of order n. We call it the **inverse** matrix for A, and denote it by A^{-1}. Being the inverse for A, it is both a right and a left inverse for A, i.e., it satisfies

$$A^{-1}A = \mathrm{id}_n = AA^{-1}.$$

More generally, we would call $A \in \mathbb{F}^{m\times n}$ invertible if there were $B \in \mathbb{F}^{n\times m}$ so that

$$AB = \mathrm{id}_m \quad \text{and} \quad BA = \mathrm{id}_n.$$

However, we will soon prove (cf. (3.17)) that this can only happen when $m = n$.

We will also soon prove (cf. (3.16)Theorem below) the *pigeonhole principle for square matrices*, i.e., that a linear map from $\mathbb{F}^n$ to $\mathbb{F}^n$ is 1-1 if and only if it is onto. In other words, if $A, B \in \mathbb{F}^{n\times n}$ and, e.g., $AB = \mathrm{id}_n$, hence A is onto, then A must also be 1-1, hence invertible, and therefore its right inverse must be its inverse, therefore we must also have $BA = \mathrm{id}_n$. In short:

(2.18) Amazing Fact: If $A, B \in \mathbb{F}^{n\times n}$ and $AB = \mathrm{id}_n$, then also $BA = \mathrm{id}_n$.

To me, this continues to be one of the most remarkable results in basic Linear Algebra. Its proof uses nothing more than the identification of matrices with linear maps (between coordinate spaces) and the numerical process called *elimination*, for solving a homogeneous linear system $A? = 0$, i.e., for constructing $\mathrm{null}\, A$.

In preparation, and as an exercise in invertible matrices, we verify the following useful fact about elementary matrices.

(2.19) Proposition: For $x, y \in \mathbb{F}^n$ and $\alpha \in \mathbb{F}$, the elementary matrix

$$E_{y,z}(\alpha) = \mathrm{id}_n + \alpha y z^{\mathrm{t}}$$

is invertible if and only if $1 + \alpha z^{\mathrm{t}} y \neq 0$, and, in that case

$$E_{y,z}(\alpha)^{-1} = E_{y,z}\Big(\frac{-\alpha}{1 + \alpha z^{\mathrm{t}} y}\Big). \tag{2.20}$$

Proof: We compute $E_{y,z}(\alpha)E_{y,z}(\beta)$ for arbitrary α and β. Since

$$\alpha y z^{\mathrm{t}}\ \beta y z^{\mathrm{t}} = \alpha\beta\ (z^{\mathrm{t}} y)\ y z^{\mathrm{t}},$$

we conclude that

$$E_{y,z}(\alpha)E_{y,z}(\beta) = (\mathrm{id}_n + \alpha y z^{\mathrm{t}})(\mathrm{id}_n + \beta y z^{\mathrm{t}}) = \mathrm{id}_n + (\alpha + \beta + \alpha\beta(z^{\mathrm{t}} y)) y z^{\mathrm{t}}.$$

In particular, since the factor $(\alpha + \beta + \alpha\beta(z^{\mathrm{t}} y))$ is symmetric in α and β, we conclude that

$$E_{y,z}(\alpha)E_{y,z}(\beta) = E_{y,z}(\beta)E_{y,z}(\alpha).$$

Further, if $1 + \alpha z^{\mathrm{t}} y \neq 0$, then the choice

$$\beta = \frac{-\alpha}{1 + \alpha z^{\mathrm{t}} y}$$

will give $\alpha + \beta + \alpha\beta(z^{\mathrm{t}} y) = 0$, hence $E_{y,z}(\beta)E_{y,z}(\alpha) = E_{y,z}(\alpha)E_{y,z}(\beta) = \mathrm{id}_n$. This proves that $E_{y,z}(\alpha)$ is invertible, with its inverse given by (2.20).

Conversely, assume that $1 + \alpha z^{\mathrm{t}} y = 0$. Then $y \neq 0$, yet

$$E_{y,z}(\alpha) y = y + \alpha(z^{\mathrm{t}} y) y = 0,$$

showing that $E_{y,z}(\alpha)$ is not 1-1 in this case, hence not invertible. □

2.17 Prove: *If two matrices commute (i.e., $AB = BA$), then they are square matrices, of the same order.*

2.18 Give a noninvertible 2-by-2 matrix without any zero entries.

2.19 Prove that the matrix $A := \begin{bmatrix} 1 & 2 \\ 4 & -1 \end{bmatrix}$ satisfies the equation $A^2 = 9\,\mathrm{id}_2$. Use this to show that A is invertible, and to write down the matrix A^{-1}.

2.20 Prove: *The matrix $A := \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if and only if $ad \neq bc$, in which case $\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}/(ad-bc)$ is its inverse.*

2.21 Consider the map $f : \mathbb{C} \to \mathbb{R}^2 : z = a + \mathrm{i}b \mapsto \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Show that f is a 1-1 linear map when we think of $\mathbb{C}$ as a vector space over the real scalar field.

2.22 Let $A, B \in L(X)$. Show that $(AB)^2 = A^2B^2$ can hold without necessarily having $AB = BA$. Show also that $(AB)^2 = A^2B^2$ implies that $AB = BA$ in case both A and B are invertible.

2.23 Give an example of matrices A and B, for which both AB and BA are defined and for which $AB = \mathrm{id}$, but neither A nor B is invertible.

2.24 Prove: *If A and C are invertible matrices, and B has as many rows as does A and as many columns as does C, then also $[A, B; 0, C]$ is invertible and*

$$[A, B; 0, C]^{-1} = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BC^{-1} \\ 0 & C^{-1} \end{bmatrix}.$$

2.25 Use (2.19)Proposition to prove the **Sherman-Morrison Formula**: *If $A \in \mathbb{F}^{n\times n}$ is invertible and $y, z \in \mathbb{F}^n$ are such that $\alpha := 1 + z^{\mathrm{t}}A^{-1}y \neq 0$, then $A + yz^{\mathrm{t}}$ is invertible, and*

$$(A + yz^{\mathrm{t}})^{-1} = A^{-1} - \alpha^{-1}A^{-1}yz^{\mathrm{t}}A^{-1}.$$

(Hint: $A + yz^{\mathrm{t}} = A(\,\mathrm{id} + (A^{-1}y)z^{\mathrm{t}})$.)

2.26 Prove the **Woodbury** generalization of the Sherman-Morrison Formula: *if A and $\mathrm{id} + D^{\mathrm{t}}AC$ are invertible, then so is $A + CD^{\mathrm{t}}$, and*

$$(A + CD^{\mathrm{t}})^{-1} = A^{-1} - A^{-1}C(\,\mathrm{id} + D^{\mathrm{t}}A^{-1}C)^{-1}D^{\mathrm{t}}A^{-1}.$$

2.27 T/F

(a) If $A, B \in L(X, Y)$ are both invertible, then so is $A + B$.

(b) If $AB = 0$ for $A, B \in \mathbb{F}^{n\times n}$, then $B = 0$.

(c) If A and B are matrices with $AB = \mathrm{id}_m$ and $BA = \mathrm{id}_n$, then $B = A^{-1}$.

(d) If $A = \begin{bmatrix} B & C \\ 0 & 0 \end{bmatrix}$ with both A and B square matrices and 0 standing for zero matrices of the appropriate size, then $A^n = \begin{bmatrix} B^n & B^{n-1}C \\ 0 & 0 \end{bmatrix}$ for all n.

(e) If $A \in \mathbb{R}^{m\times n}$ and $A^{\mathrm{t}}A = 0$, then $A = 0$.

(f) If the matrix product AB is defined, then $(AB)^{\mathrm{t}} = A^{\mathrm{t}}B^{\mathrm{t}}$.

(g) If A is an invertible matrix, then so is A^{t}, and $(A^{\mathrm{t}})^{-1} = (A^{-1})^{\mathrm{t}}$.

(h) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ is an elementary matrix.

(i) If Y is a subset of some vector space X, x, y, z are particular elements of X, and x and $2y - 3x$ are in Y, but $3y - 2x$ and y are not, then Y cannot be a linear subspace.

3. Elimination, or: The determination of null A and ran A

Elimination and Backsubstitution

Elimination has as its goal an efficient description of the solution set for the *homogeneous* linear system $A? = 0$, i.e., of the nullspace of the matrix A. It is based on the following observation:

(3.1) Lemma: If B is obtained from A by subtracting some multiple of some row of A from some *other* row of A, then $\operatorname{null} B = \operatorname{null} A$.

Proof: Assume, more specifically, that B is obtained from A by subtracting α times row k from row i, for some $k \neq i$. Then, by (2.10)Example,

$$B = E_{i,k}(-\alpha)\, A,$$

with $E_{i,k}(-\alpha) = \mathrm{id}_m - \alpha e_i {e_k}^{\mathrm{t}}$. Consequently, $\operatorname{null} B \supset \operatorname{null} A$, and this holds even if $i = k$.

However, since $i \neq k$, we have ${e_k}^{\mathrm{t}} e_i = 0$, hence, for any α, $1 + \alpha({e_k}^{\mathrm{t}} e_i) = 1 \neq 0$. Therefore, by (2.19), also

$$E_{i,k}(\alpha)\, B = A,$$

hence also $\operatorname{null} B \subset \operatorname{null} A$. □

One solves the homogeneous linear system $A? = 0$ by **elimination**. This is an *inductive* process, and it results in a classification of the unknowns as *free* or *bound*. A **bound** unknown has associated with it a **pivot row** or **pivot equation** which determines this unknown uniquely once all later unknowns are determined. Any unknown without a pivot equation is a **free** unknown; its value can be chosen arbitrarily. We call the jth *column* of A bound (free) if the jth unknown is bound (free). The classification proceeds inductively, from the first to the last unknown or column, i.e., for $k = 1, 2, \ldots$, with the kth step as follows.

At the beginning of the kth **elimination step**, we have in hand a matrix B, called the **working-array**, which is **equivalent** to our initial matrix A in that $\operatorname{null} B = \operatorname{null} A$. Further, we have already classified the first $k-1$ unknowns as either bound or free, with each bound unknown associated with a particular row of B, its *pivot row*, and this row having a nonzero entry at the position of its associated bound unknown and zero entries for all previous unknowns. All other rows of B are nonpivot rows; they do not involve the unknowns already classified, i.e., they have nonzero entries only for unknowns not yet classified. (Note that, with the choice $B := A$, this description also fits the situation at the beginning of the first step.) We now classify the kth unknown or column and, correspondingly, change B, as follows:

bound case: We call the kth unknown or column **bound** (some would say **basic**) in case we can find some nonpivot row $B(h,:)$ for which $B(h,k) \neq 0$. We pick one such row and call it the **pivot row** for the kth unknown. Further, we use it to eliminate the kth unknown from all the remaining nonpivot rows $B(i,:)$ by the calculation

$$B(i,:) \leftarrow B(i,:) - \frac{B(i,k)}{B(h,k)} B(h,:).$$

free case: In the contrary case, we call the kth unknown or column **free** (some would say **nonbasic**). No action is required in this case, since none of the nonpivot rows involves the kth unknown.

By (3.1)Lemma, the changes (if any) made in B will not change $\operatorname{null} B$. This finishes the kth elimination step.

For future reference, here is a formal description of the entire algorithm. This description relies on a sequence p to keep track of which row, if any, is used as pivot row for each of the unknowns. If row h is the pivot row for the kth unknown, then $p(k) = h$ after the kth elimination step. Since p is initialized to have all its entries equal to 0, this means that, at any time, the rows k not yet used as pivot rows are exactly those for which $p(k) = 0$.

(3.2) Elimination Algorithm:
input: $A \in \mathbb{F}^{m\times n}$.
$B \leftarrow A$, $p \leftarrow (0,\ldots,0) \in \mathbb{R}^n$.
for $k = 1{:}n$, **do**:
  for some $h \notin \operatorname{ran} p$ with $B(h,k) \neq 0$, **do**:
    $p(k) \leftarrow h$
    for all $i \notin \operatorname{ran} p$, **do**:

$$B(i,:) \leftarrow B(i,:) - \frac{B(i,k)}{B(h,k)}\, B(h,:)$$

    enddo
  enddo
enddo
output: B, p, and, possibly, `free` $\leftarrow$ `find(`p`==0)`, `bound` $\leftarrow$ `find(`p`>0)`.

Note that nothing is done at the kth step if there is no $h \notin \operatorname{ran} p$ with $B(h,k) \neq 0$, i.e., if $B(h,k) = 0$ for all $h \notin \operatorname{ran} p$. In particular, $p(k)$ will remain 0 in that case.

A numerical example: We start with

$$A := \begin{bmatrix} 0 & 2 & 0 & 2 & 5 & 4 & 0 & 6 \\ 0 & 1 & 0 & 1 & 2 & 2 & 0 & 3 \\ 0 & 2 & 0 & 2 & 5 & 4 & -1 & 7 \\ 0 & 1 & 0 & 1 & 3 & 2 & -1 & 4 \end{bmatrix}, \qquad p = (0,0,0,0,0,0,0,0).$$

The first unknown is free. We take the second row as pivot row for the second unknown and eliminate it from the remaining rows, to get

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \underline{\mathbf{1}} & 0 & 1 & 2 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 1 \end{bmatrix}, \qquad p = (0,2,0,0,0,0,0,0).$$

Thus the third unknown is free as is the fourth, but the fifth is not, since there are nonzero entries in the fifth column of some nonpivotal row, e.g., the first row. We choose the first row as pivot row for the fifth unknown and use it to eliminate this unknown from the remaining nonpivot rows, i.e., from rows 3 and 4. This gives

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & \underline{\mathbf{1}} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 1 & 2 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \qquad p = (0,2,0,0,1,0,0,0).$$

The sixth unknown is free, but there are nonzero entries in the seventh column of the remaining nonpivot rows, so the seventh unknown is bound, with, e.g., the fourth row as its pivot row. We use that row to eliminate the seventh unknown from the remaining nonpivot row. This gives

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 1 & 2 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \underline{\mathbf{-1}} & 1 \end{bmatrix}, \qquad p = (0,2,0,0,1,0,4,0).$$

With that, there are no nontrivial nonpivot rows left. In particular, the eighth unknown is free, hence we have already in hand the final array.

Hence, altogether `bound` $= (2,5,7)$ ($=$ `find(p>0)`) and `free` $= (1,3,4,6,8)$ ($=$ `find(p==0)`). □

After the n steps of this elimination process (which started with $B = A$), we have in hand a matrix B with $\operatorname{null} B = \operatorname{null} A$ and with each unknown classified as bound or free. The two increasing sequences, `bound` and `free`, containing the indices of the bound and free unknowns respectively, will be much used in the sequel. Each bound unknown has associated with it a particular row of B, its pivot row. All nonpivot rows of B (if any) are entirely zero.

Neat minds would reorder the rows of B, listing first the pivot rows in order, followed by the nonpivot rows and, in this way, obtain a **row echelon form** for A. In any case, in determining $x \in \operatorname{null} B$, we only have to pay attention to the pivot rows. This means that we can determine a particular element x of $\operatorname{null} B = \operatorname{null} A$ by *backsubstitution*, i.e., from its last entry to its first as follows:

For $k = n{:}-1{:}1$, if the kth unknown is bound, i.e., $k \in$ `bound`, determine x_k from its pivot equation (since that equation only involves $x_k, \cdots, x_n$); else, pick x_k arbitrarily (as then the kth unknown is free, i.e., $k \in$ `free`).

Here is a more formal description, for future reference.

(3.3) Backsubstitution Algorithm:
input: $B \in \mathbb{F}^{m \times n}$ and p (both as output from (3.2)), $z \in \mathbb{F}^n$.
$x \leftarrow z$
for $k = n{:}-1{:}1$, **do**:
  if $p(k) \neq 0$, **then** $x_k \leftarrow -\left(\sum_{j>k} B(p(k), j) x_j\right) / B(p(k), k)$ **endif**
enddo
output: x, which is the unique solution of $A? = 0$ satisfying $x_i = z_i$ for all i with $p(i) = 0$.

Notice that the value of every free unknown is arbitrary and that, once these are chosen somehow, then the bound unknowns are uniquely determined by the requirement that we are seeking an element of $\operatorname{null} B = \operatorname{null} A$. In other words, the general element of $\operatorname{null} B$ has exactly as many degrees of freedom as there are free unknowns. Since there are #`free` unknowns, $\operatorname{null} B$ is said to be 'of dimension #`free`'.

In particular, for any k, the kth entry, x_k, of an $x \in \operatorname{null} B$ can be nonzero only in one of two ways: (a) the kth unknown is free, i.e., $k \in$ `free`; (b) the kth unknown is bound, but $x_j \neq 0$ for some $j > k$. It follows that x_k can be the rightmost nonzero entry of such an x only if the kth unknown is free. Conversely, if the kth unknown is free, and x is the element of $\operatorname{null} B = \operatorname{null} A$ computed by setting $x_k = 1$ and setting all other free entries equal to 0, then x_k is necessarily the rightmost nonzero entry of x (since all free entries to the right of it were chosen to be zero, thus preventing any bound entry to the right of it from being nonzero).

This proves

(3.4) Observation: There exists $x \in \operatorname{null} A$ with rightmost nonzero entry x_k if and only if the kth unknown is free.

This simple observation gives a *characterization* of the sequence `free` entirely in terms of the nullspace of the matrix A we started with. This implies that *the classification into free and bound unknowns or columns is independent of all the details of the elimination*. More than that, since, for any 1-1 matrix M with m columns, $\operatorname{null}(MA) = \operatorname{null} A$, it implies that, for any such matrix MA, we get exactly the same sequences `free` and `bound` as we would get for A. This is the major reason for the uniqueness of a more disciplined echelon form, the 'really reduced row echelon form', to be discussed in the next section.

Since $A(:, k) \in \operatorname{ran} A(:, [1{:}k-1])$ if and only if there is some $x \in \operatorname{null} A$ whose rightmost nonzero entry is its kth, we have the following reformulation of (3.4)Observation and consequences.

(3.5) Corollary:

(i) The kth column of A is free if and only if it is a weighted sum of the columns strictly to the left of it, i.e., $A(:,k) \in \operatorname{ran} A(:,[1{:}k-1])$.

(ii) $A(:,[1{:}k])$ is 1-1 if and only if all its columns are bound.

(iii) $\operatorname{null} A$ is nontrivial if and only if there are free columns.

Perhaps the most widely used consequence of (iii) here is the following. If there are more unknowns than equations, then there are not enough equations to go around, i.e., some unknowns must be free, therefore there are nontrivial solutions to our homogeneous equation $A? = 0$. We remember this fundamental result of elimination in the following form:

(3.6) Theorem: Any matrix with more columns than rows has a nontrivial nullspace.

3.1 Determine the bound and free columns for each of the following matrices A.

(a) $0 \in \mathbb{R}^{m\times n}$; (b) $[e_1,\ldots,e_n] \in \mathbb{R}^{n\times n}$; (c) $[e_1, 0, e_2, 0] \in \mathbb{R}^{6\times 4}$; (d) $\begin{bmatrix} 2 & 2 & 5 & 6 \\ 1 & 1 & -2 & 2 \end{bmatrix}$; (e) $\begin{bmatrix} 0 & 2 & 1 & 4 \\ 0 & 0 & 2 & 6 \\ 1 & 0 & -3 & 2 \end{bmatrix}$; (f) $[x][y]^{\mathrm{t}}$, with $x = (1,2,3,4) = y$.

3.2 (3.5)Corollary assures you that $y \in \operatorname{ran} A$ if and only if the last column of $[A, y]$ is free. Use this fact to determine, for each of the following y and A, whether or not $y \in \operatorname{ran} A$.

(a) $y = (\pi, 1-\pi)$, $A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix}$; (b) $y = e_2$, $A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 3 & -4 \\ 3 & 4 & -8 \end{bmatrix}$; (c) $y = e_2$, $A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 3 & -4 \\ 3 & 4 & -7 \end{bmatrix}$.

3.3 Prove (3.1)Lemma directly, i.e., without using (2.19)Proposition. (Hint: Prove that $\operatorname{null} B \supset \operatorname{null} A$. Then prove that also A is obtainable from B by the same kind of step, hence also $\operatorname{null} A \supset \operatorname{null} B$.)

3.4 Prove: *If M and A are matrices for which MA is defined and, furthermore, M is 1-1, then $MA? = 0$ has exactly the same free and bound unknowns as does $A? = 0$.*

The really reduced row echelon form and other reduced forms

The construction of the *really reduced row echelon form* takes elimination four steps further, none of which changes the nullspace:

(i) When the hth pivot row is found, and it is not the hth row, then it is exchanged with the current hth row to make it the hth row. (This keeps things neat; all the rows not yet used as pivot rows lie below all the rows already picked as pivot rows.)

(ii) Each pivot row is divided by its **pivot element**, i.e., by its left-most nonzero entry. (This helps with the elimination of the corresponding unknown from other rows: if $B(h,k)$ is the pivot element in question (i.e., $\texttt{bound}(h) = k$, i.e., x_k is the hth bound unknown), then, after this normalization, one merely subtracts $B(i,k)$ times $B(h,:)$ from $B(i,:)$ to eliminate the kth unknown from row i.)

(iii) One eliminates each bound unknown from *all* rows (other than its pivot row), i.e., also from pivot rows belonging to earlier bound unknowns, and not just from the rows not yet used as pivot rows. For real efficiency, though, this additional step should be carried out after elimination is completed; it starts with the elimination of the *last* bound unknown, proceeds to the second-last bound unknown, etc., and ends with the *second* bound unknown (the first bound unknown was eliminated from all other rows already).

The resulting matrix B is called the **reduced row echelon form for** A, and this is written:

$$B = \operatorname{rref}(A).$$

However, it turns out to be very neat to add the following final step:

(iv) Remove all rows that are entirely zero, thus getting the matrix

$$R := B(1{:}\#\texttt{bound}, :) =: \operatorname{rrref}(A)$$

called the *really reduced row echelon form* of A.

Here is a formal description (in which we talk about *the* rrref for A even though we prove its *uniqueness* only later, in (3.12)):

(3.7) Definition: We say that R is the **really reduced row echelon form for** $A \in \mathbb{F}^{m\times n}$ and write $R = \operatorname{rrref}(A)$, in case $R \in \mathbb{F}^{r\times n}$ for some r and there is a strictly increasing r-sequence `bound` (provided by the `MATLAB` function `rref` along with $\operatorname{rref}(A)$) so that the following is true:

1. R is a **row echelon form for** A: This means that (i) $\operatorname{null} R = \operatorname{null} A$; and (ii) for each $k = \texttt{bound}(i)$, $R(i,:)$ is the pivot row for the kth unknown, i.e., $R(i,:)$ is the unique row in R for which $R(i,k)$ is the first (or, leftmost) nonzero entry.

2. R is **really reduced** or normalized, in the sense that $R(:,\texttt{bound})$ is the identity matrix, i.e., for each i, the pivot element $R(i,\texttt{bound}(i))$ equals 1 and is the only nonzero entry in its column, and R has only these $r = \#\texttt{bound}$ rows.

A numerical example, continued: For the earlier numerical example, the rref and the rrref would look like this:

$$\texttt{rref}(A) = \begin{bmatrix} 0 & \mathbf{1} & 0 & 1 & 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \underline{\mathbf{1}} & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \operatorname{rrref}(A) = \begin{bmatrix} 0 & \mathbf{1} & 0 & 1 & 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \underline{\mathbf{1}} & -1 \end{bmatrix}.$$

Recall (or observe directly) that, for this example, $\texttt{bound} = (2,5,7)$ and $\texttt{free} = (1,3,4,6,8)$. □

Finally, for most purposes, it is sufficient to have a **b-form** for A, i.e., a matrix R that satisfies the following two conditions:

(3.8)(i) $\operatorname{null} R = \operatorname{null} A$;

(3.8)(ii) $R(:,\texttt{b}) = \operatorname{id}$ for some sequence `b`.

Certainly, in these terms, the $\operatorname{rrref}(A)$ is a `bound`-form for A, but a matrix A may have many `b`-forms, and, as we shall see, only the two conditions (3.8)(i-ii) really matter.

3.5 For each of the matrices A in HP(3.1), determine its rrref.

A complete description for $\operatorname{null} A$ obtained from a b-form

If R is a `b`-form for A, then it is easy to determine all solutions of the homogeneous linear system $A? = 0$, i.e., all the elements of $\operatorname{null} A$.

In recognition of the special case $R = \operatorname{rrref}(A)$, I'll use `f` for a sequence **complementary to** `b` in the sense that it contains all the indices in $\underline{n}$ but not in `b`.

In `MATLAB`, one would obtain `f` from n and `b` by the commands `f = 1:n; f(b) = [];` □

We now obtain from any `b`-form R for A a 1-1 matrix C with the property that $\operatorname{null} A = \operatorname{ran} C$, thus getting a description both as a range and as a nullspace. Since such a C is 1-1 onto $\operatorname{null} A$, this implies that every $x \in \operatorname{null} A$ can be written *in exactly one way* in the form $x = Ca$. We will soon learn to call such a C a 'basis' for the vector space $\operatorname{null} A$.

In the discussion, we use the following notation introduced earlier: If x is an n-vector and p is a list of length r with range in $\underline{n}$, then x_p is the r-vector

$$x_p = (x_{p(i)} : i = 1{:}r).$$

With this, by property (3.8)(i),

$$x \in \operatorname{null} A \iff 0 = Rx = \sum_j R(:,j)x_j = R(:,\mathtt{b})x_{\mathtt{b}} \ + \ R(:,\mathtt{f})x_{\mathtt{f}} \ .$$

Since $R(:,\mathtt{b}) =$ id by property (3.8)(ii), we conclude that

$$x \in \operatorname{null} A \iff x_{\mathtt{b}} = -R(:,\mathtt{f})x_{\mathtt{f}}.$$

We can write this even more succinctly in matrix form as follows:

$$\operatorname{null} A = \operatorname{ran} C,$$

with C the $(n \times \#\mathtt{f})$-matrix whose ‘$\mathtt{f}$-rows’ form an identity matrix, and whose ‘$\mathtt{b}$-rows’ are formed by the ‘$\mathtt{f}$-columns’ of $-R$:

$$C(\mathtt{f},:) = \text{id}, \qquad C(\mathtt{b},:) = -R(:,\mathtt{f}). \tag{3.9}$$

E.g., for the earlier numerical example and with $R = \operatorname{rrref}(A)$,

$$C = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -2 & -3 \\ 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -0 & -0 & -1 & -2 & -3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -0 & -0 & -0 & -0 & -0 \\ 0 & 0 & 0 & 0 & 0 \\ -0 & -0 & -0 & -0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that C is 1-1, since $x := Ca = 0$ implies that $0 = x_{\mathtt{f}} = C(\mathtt{f},:)a = a$. Therefore, C is (or, the columns of C form) a ‘basis’ for $\operatorname{null} A$, in the sense that C is a 1-1 onto column map to $\operatorname{null} A$.

Finally, when $R = \operatorname{rrref}(A)$, then the resulting C is ‘upper triangular’ in the sense that then

$$i > \mathtt{free}(j) \implies C(i,j) = 0. \tag{3.10}$$

3.6 Determine a ‘basis’ for the nullspace of $A := \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ and use it to describe the solution set of the system $A? = (1,2)$. Draw a picture indicating both the solution set and $\operatorname{null} A$.

3.7 For each of the matrices A in HP(3.1), give a ‘basis’ for $\operatorname{null} A$.

The factorization $A = A(:,\mathtt{bound})\operatorname{rrref}(A)$

Continuing with our $\mathtt{b}$-form R for A, we claim that

$$A(:,\mathtt{b})R = A.$$

For the proof, we compare $A(:,\mathtt{b})R =: M$ and A column by column. First, $M(:,\mathtt{b}) = A(:,\mathtt{b})R(:,\mathtt{b}) = A(:,\mathtt{b})$, by property (3.8)(ii). As to $M(:,\mathtt{f}) = A(:,\mathtt{b})R(:,\mathtt{f})$, we observe that, for any c (of length $\#\mathtt{f}$), the vector x with

$$x_{\mathtt{b}} := R(:,\mathtt{f})c, \qquad x_{\mathtt{f}} := -c,$$

is in $\operatorname{null} R = \operatorname{null} A$, hence

$$0 = Ax = A(:,\mathtt{b})x_{\mathtt{b}} + A(:,\mathtt{f})x_{\mathtt{f}} = A(:,\mathtt{b})R(:,\mathtt{f})c + A(:,\mathtt{f})(-c).$$

In other words,

$$M(:\mathtt{f})c = A(:,\mathtt{b})R(:,\mathtt{f})c = A(:,\mathtt{f})c, \quad \forall c \in \mathbb{F}^{\#\mathtt{f}},$$

showing that also $M(:,\mathtt{f}) = A(:,\mathtt{f})$. This proves our claim that $A(:,\mathtt{b})R = A$, hence, in particular,

$$\boxed{\textbf{(3.11)} \qquad A = A(:,\mathtt{bound})\operatorname{rrref}(A).}$$

3.8 Prove: *If M is such that $MA = \operatorname{rrref}(A) =: R$, and $\mathtt{bound}$ is the increasing sequence of indices of bound columns of A, then M is a left inverse for $A(:,\mathtt{bound})$.*

A 'basis' for $\operatorname{ran} A$

Here is a first consequence of the factorization $A = A(:,\mathtt{b})R$ (with R satisfying (3.8)(i–ii)): The factorization implies that $\operatorname{ran} A \subset \operatorname{ran} A(:,\mathtt{b})$, while certainly $\operatorname{ran} A(:,\mathtt{b}) \subset \operatorname{ran} A$. Hence

$$\operatorname{ran} A = \operatorname{ran} A(:,\mathtt{b}).$$

Also, $A(:,\mathtt{b})$ is 1-1: For, if $A(:,\mathtt{b})a = 0$, then the n-vector x with $x_{\mathtt{b}} = a$ and with $x_{\mathtt{f}} = 0$ is in $\operatorname{null} A = \operatorname{null} R$, hence $a = x_{\mathtt{b}} = -R(:,\mathtt{f})x_{\mathtt{f}} = -R(:,\mathtt{f})0 = 0$. Consequently, $A(:,\mathtt{b})$ is (or, the columns of $A(:,\mathtt{b})$ form) a 'basis' for $\operatorname{ran} A$.

3.9 For each of the matrices A in HP(3.1), give a 'basis' for $\operatorname{ran} A$.

3.10 Let A be the $n \times n$ matrix $[0, e_1, \ldots, e_{n-1}]$ (with e_j denoting the jth unit vector, of the appropriate length). (a) What is its rref? (b) In the equation $A? = 0$, which unknowns are bound, which are free? (c) Give a 'basis' for $\operatorname{null} A$ and a 'basis' for $\operatorname{ran} A$.

3.11 Let M be the 6×3-matrix $[e_3, e_2, e_1]$. (a) What is its rref? (b) Use (a) to prove that M is 1-1. (c) Construct a left inverse for M. (d) (off the wall:) Give a matrix P for which $\operatorname{null} P = \operatorname{ran} M$.

3.12 Let $N := M^{\mathrm{t}}$, with M the matrix in the previous problem. (a) What is its rref? (b) Use (a) to prove that N is onto. (c) Construct a right inverse for N.

3.13 Use the rref to prove that $\operatorname{ran} U = \operatorname{ran} V$, with

$$U := \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -1 & 1 & 3 \end{bmatrix}, \qquad V := \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ -4 & -5 \end{bmatrix}.$$

(Hints: Proving two sets to be equal usually involves showing that each is a subset of the other. In this case, applying elimination to $[V, U]$ as well as to $[U, V]$ should provide all the information you need.)

Uniqueness of the $\operatorname{rrref}(A)$

If R is a $\mathtt{b}$-form for A, then, as we just proved, $A = A(:,\mathtt{b})R$ and $A(:,\mathtt{b})$ is 1-1. Hence, if also S is a $\mathtt{b}$-form for A, then we have $A(:,\mathtt{b})R = A = A(:,\mathtt{b})S$ and, since $A(:,\mathtt{b})$ is 1-1, this implies that $R = S$. In other words, the matrix R is uniquely determined by the condition that $A(:,\mathtt{b})R = A$. In particular, $\operatorname{rrref}(A)$ is uniquely determined, since we already observed that, by (3.4), the sequence $\mathtt{bound}$ only depends on $\operatorname{null} A$.

Further, since $\operatorname{rref}(A)$ differs from $\operatorname{rrref}(A)$ only by those additional $m - \#\mathtt{bound}$ zero rows, it follows that each A also has a *unique* rref.

This finishes the proof of the following summarizing theorem.

(3.12) Theorem: For given $A \in \mathbb{F}^{m\times n}$, there is exactly one matrix R having the properties 1. and 2. (listed in (3.7)) of a rrref for A. Further, with $\mathtt{bound}$ and $\mathtt{free}$ the indices of bound and free unknowns, $A(:,\mathtt{bound})$ is 1-1 onto $\operatorname{ran} A$, and $C \in \mathbb{F}^{n\times\#\mathtt{free}}$, given by $C(\mathtt{free},:) = \mathrm{id}$, $C(\mathtt{bound},:) = -R(:,\mathtt{free})$, is 1-1 onto $\operatorname{null} A$, and C is 'upper triangular' in the sense that $C(i,j) = 0$ for $i > \mathtt{free}(j)$.

$\operatorname{rrref}(A)$ and the solving of $A? = y$

(3.5)Corollary(i) is exactly what we need when considering the linear system

$$A? = y \tag{3.13}$$

for given $A \in \mathbb{F}^{m\times n}$ and given $y \in \mathbb{F}^m$. For, here we are hoping to write y as a linear combination of the columns of A, and (3.5) tells us that this is possible exactly when the last unknown in the *homogeneous* system

$$[A, y]? = 0 \tag{3.14}$$

is free. Further, the factorization (3.11), applied to the **augmented** matrix $[A, y]$, tells us how to write y as a linear combination of the columns of A in case that can be done. For, with $R = \mathrm{rrref}([A, y])$, it tells us that

$$y = [A, y](:\texttt{bound})R(:, n + 1),$$

and this gives us y in terms of the columns of A precisely when $n + 1 \notin \texttt{bound}$, i.e., when the $(n + 1)$st unknown is free.

(3.15) Proposition: For $A \in \mathbb{F}^{m\times n}$ and $y \in \mathbb{F}^m$, the equation

$$A? = y$$

has a solution if and only if the last column of $[A, y]$ is free, in which case the last column of $\mathrm{rrref}([A, y])$ provides the unique solution to

$$A(:\texttt{bound})? = y.$$

More generally, if $R = \mathrm{rrref}([A, B])$ for some arbitrary matrix $B \in \mathbb{F}^{m\times s}$ and all the unknowns corresponding to columns of B are free, then, by (3.11), applied to $[A, B]$ rather than A, we have

$$B = A(:, \texttt{bound})R(:, n + [1:s]).$$

A numerical example, continued: Recall our earlier example in which we used elimination to convert a given matrix to its rrref, as follows:

$$\begin{bmatrix} 0 & 2 & 0 & 2 & 5 & 4 & 0 & 6 \\ 0 & 1 & 0 & 1 & 2 & 2 & 0 & 3 \\ 0 & 2 & 0 & 2 & 5 & 4 & -1 & 7 \\ 0 & 1 & 0 & 1 & 3 & 2 & -1 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 1 & 2 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{-1} & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & \mathbf{1} & 0 & 1 & 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & -1 \end{bmatrix},$$

hence $\texttt{bound} = (2, 5, 7)$, $\texttt{free} = (1, 3, 4, 6, 8)$. Now, the elimination algorithm is entirely unaware of how we got the initial matrix. In particular, we are free to interpret in various ways the array on the left as being of the form $[A, B]$. As soon as we specify the number of columns, in A or B, we know A and B exactly.

First, choose B to be a one-column matrix. Then, since the last unknown is free, we conclude that

$$(6, 3, 7, 4) = A(:, \texttt{bound})R(:, 8) = \begin{bmatrix} 2 & 5 & 0 \\ 1 & 2 & 0 \\ 2 & 5 & -1 \\ 1 & 3 & -1 \end{bmatrix} (3, 0, -1).$$

If we choose B to be a three-column matrix instead, then the linear system $A? = B$ is unsolvable since now one of the columns of B (the second one) corresponds to a bound unknown. What about the other two columns of this B? The first one corresponds to a free unknown, hence is a weighted sum of the columns to the left of it, hence is in $\operatorname{ran} A$. But the last one fails to be in $\operatorname{ran} A$ since its unknown is free only because of the presence of the seventh column, and this seventh column is *not* in the span of the columns to the left of it, hence neither is the eighth column. Indeed, the corresponding column of R has its last entry nonzero, showing that $A(:, \texttt{bound}(3))$ is needed to write the last column of A as a weighted sum of columns to the left of it. □

3.14 Use elimination to show that $\begin{bmatrix} 2 & -1 & 0 \\ 1 & 2 & 1 \\ 0 & 2 & -1 \end{bmatrix}$ is 1-1 and onto.

3.15 Use elimination to settle the following assertions, concerning the linear system $A? = y$, with the (square) matrix A and the right side y given by

$$[A, y] := \begin{bmatrix} 1 & -2 & 3 & 1 \\ 2 & k & 6 & 6 \\ -1 & 3 & k-3 & 0 \end{bmatrix}.$$

(a) If $k = 0$, then the system has an infinite number of solutions. (b) For another specific value of k, which you must find, the system has no solutions. (c) For all other values of k, the system has a unique solution.

(To be sure, there probably is some preliminary work to do, after which it is straightforward to answer all three questions.)

3.16 Here are three questions that can be settled **without doing any arithmetic**. Please do so.

(i) Can both of the following equalities be right?

$$\begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} = \mathrm{id}_2 = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} -4 & 2 \\ 3 & 5 \end{bmatrix}$$

(ii) How does one find the coordinates of $e_1 \in \mathbb{R}^2$ with respect to the vector sequence $(1,3), (2,5)$ (i.e., numbers α, β for which $e_1 = (1,3)\alpha + (2,5)\beta$), given that

$$AV := \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} = \mathrm{id}_2\,?$$

(iii) How does one conclude at a glance that the following equation must be wrong?

$$\begin{bmatrix} -5 & 2 \\ 3 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 3 & 5 & 0 \end{bmatrix} = \mathrm{id}_3\,?$$

The pigeonhole principle for square matrices

We are ready for a discussion of our basic problem, namely solving $A? = y$, in case $A \in \mathbb{F}^{m\times n}$, hence $y \in \mathbb{F}^m$. When is A 1-1, onto, invertible? We answer all these questions by applying elimination to the augmented matrix $[A, y]$.

If A is 1-1, then, by (3.5)Corollary, all its columns must be *bound*. In particular, there must be enough rows to bind them, i.e., $m \ge n$. Further, if $m = n$, then, by the time we reach the last column of $[A, y]$, there is no row left to bind it. Therefore, the last column must be free regardless of the choice of y, hence, by (3.5)Corollary, $y \in \operatorname{ran} A$ for every $y \in \mathbb{F}^m = \operatorname{tar} A$, i.e., A is onto.

If A is onto, then, for $i = 1{:}m$, there is $b_i \in \mathbb{F}^n$ so that $Ab_i = e_i \in \mathbb{F}^m$. Hence, with $B := [b_1, \ldots, b_m] \in \mathbb{F}^{n\times m}$, we have $AB = A[b_1, \ldots, b_m] = [Ab_1, \ldots, Ab_m] = [e_1, \ldots, e_m] = \mathrm{id}_m$. It follows that B is 1-1, hence B has at least as many rows as columns, i.e., $n \ge m$, and A is a left inverse for B. Further, if $n = m$, then, by the previous paragraph, B is also onto, hence invertible, hence any left inverse must be its inverse. In particular $A = B^{-1}$ and therefore, in particular, A is 1-1.

Note that the argument just given provides the proof of the 'Amazing Fact' (2.18), since it concludes from $AB = \mathrm{id}$ (with A, B square) that A must be the inverse of B, and this implies, in particular, that also $BA = \mathrm{id}$.

But we have proved much more, namely the following basic Theorem.

(3.16) Theorem (pigeonhole principle for square matrices): A square matrix is 1-1 if and only if it is onto.

In other words, when dealing with a *square* matrix, 1-1 *or* onto is already enough to have 1-1 *and* onto, i.e., to have invertibility.

We also now know that only square matrices are invertible.

(3.17) Proposition: An invertible matrix is necessarily square. More precisely, if $A \in \mathbb{F}^{m\times n}$, then (i) A 1-1 implies that $m \ge n$; and (ii) A onto implies that $m \le n$.

If $A \in \mathbb{F}^{n\times n}$ is invertible, then the first n columns of $[A,\ \mathrm{id}_n]$ are necessarily bound and the remaining n columns are necessarily free. Therefore, if $R := \mathrm{rrref}([A,\ \mathrm{id}_n])$, then $R = [\,\mathrm{id}_n, ?]$ and, with (3.11), necessarily $[A,\ \mathrm{id}_n] = AR = [A\,\mathrm{id}_n, A?]$, hence $? = A^{-1}$, i.e., $R = [\,\mathrm{id}_n, A^{-1}]$.

practical note: Although `MATLAB` provides the function `inv(A)` to generate the inverse of `A`, there is usually no reason to compute the inverse of a matrix, nor would you solve the linear system $A? = y$ in practice by computing $\mathrm{rref}([A, y])$ or by computing `inv(A)*y`. Rather, in `MATLAB` you would compute the solution of `A`$?$ =`y` as `A\y`. For this, `MATLAB` also uses elimination, but in a more sophisticated form, to keep rounding error effects as small as possible. In effect, the choice of pivot rows is more elaborate than we discussed above.

□

(3.18) Example: Triangular matrices There is essentially only one class of square matrices whose invertibility can be settled by inspection, namely the class of triangular matrices.

Assume that the square matrix A is **upper triangular**, meaning that $i > j \Longrightarrow A(i,j) = 0$. If all its diagonal elements are nonzero, then each of its unknowns has a pivot row, hence is bound and, consequently, A is 1-1, hence, by (3.16)Theorem, it is invertible. Conversely, if some of its diagonal elements are zero, then there must be a first zero diagonal entry, say $A(i,i) = 0 \ne A(k,k)$ for $k < i$. Then, for $k < i$, row k is a pivot row for x_k, hence, when it comes time to decide whether x_i is free or bound, all rows not yet used as pivot rows do not involve x_i explicitly, and so x_i is free. Consequently, $\mathrm{null}\,A$ is nontrivial and A fails to be 1-1.

Exactly the same argument can be made in case A is **lower triangular**, meaning that $i < j \Longrightarrow A(i,j) = 0$, provided you are now willing to carry out the elimination process from right to left, i.e., in the order x_n, x_{n-1}, etc., and, correspondingly, recognize a row as pivot row for x_k in case x_k is the last unknown that appears explicitly (i.e., with a nonzero coefficient) in that row.

(3.19) Proposition: A square triangular matrix is invertible if and only if all its diagonal entries are nonzero.

□

(3.20) Example: Interpolation If $V \in L(\mathbb{F}^n, X)$ and $Q \in L(X, \mathbb{F}^n)$, then QV is a linear map from $\mathbb{F}^n$ to $\mathbb{F}^n$, i.e., a square matrix, of order n. If QV is 1-1 or onto, then (3.16)Theorem tells us that QV is invertible. In particular, V is 1-1 and Q is onto, and so, for every $y \in \mathbb{F}^n$, there exists exactly one $p \in \operatorname{ran} V$ for which $Qp = y$. This is the essence of *interpolation*.

For example, take $X = \mathbb{R}^{\mathbb{R}}$, $V = [()^0, ()^1, \ldots, ()^{k-1}]$, hence $\operatorname{ran} V$ equals $\Pi_{<k}$, the collection of all polynomials of degree $< k$. Further, take $Q : X \to \mathbb{R}^k : f \mapsto (f(\tau_1), \ldots, f(\tau_k))$ for some fixed sequence $\tau_1 < \cdots < \tau_k$ of points. Then the equation

$$QV? = Qf$$

asks for the (power) coefficients of a polynomial of degree $< k$ that agrees with the function f at the k distinct points τ_j.

We investigate whether QV is 1-1 or onto, hence invertible. For this, consider the matrix QW, with the columns of $W := [w_1, \ldots, w_k]$ the so-called **Newton polynomial**s

$$w_j : t \mapsto \prod_{h<j} (t - \tau_h).$$

Observe that $(QW)(i,j) = (Qw_j)(\tau_i) = \prod_{h<j}(\tau_i - \tau_h) = 0$ if and only if $i < j$. Therefore, QW is square and lower triangular with nonzero diagonal entries, hence invertible by (3.19)Proposition, while w_j is a polynomial of exact degree $j - 1 < k$, hence $w_j = Vc_j$ for some k-vector c_j. It follows that the invertible matrix QW equals

$$QW = [Qw_1, \ldots, Qw_k] = [QVc_1, \ldots, QVc_k] = (QV)[c_1, \ldots, c_k].$$

In particular, QV is onto, hence invertible, hence also V is 1-1, therefore invertible as a linear map from $\mathbb{R}^k$ to its range, $\Pi_{<k}$. We have proved:

(3.21) Proposition: For every $f : \mathbb{R} \to \mathbb{R}$ and every k distinct points $\tau_1, \ldots, \tau_k$ in $\mathbb{R}$, there is exactly one choice of coefficient vector a for which the polynomial $[()^0, \ldots, ()^{k-1}]a$ of degree $< k$ agrees with f at these τ_j.

In particular, (i) the column map $[()^0, \ldots, ()^{k-1}] : \mathbb{R}^k \to \Pi_{<k}$ is invertible, and (ii) any polynomial of degree $< k$ with more than $k - 1$ distinct zeros must be 0.

□

3.17 For each of the following matrices A, use elimination (to the extent necessary) to (a) determine whether it is invertible and, if it is, to (b) construct the inverse (see the remark following (3.17)Proposition).

(a) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}$; (b) $\begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}$; (c) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$; (d) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 4 \end{bmatrix}$; (e) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 8 \end{bmatrix}$; (f) $[e_1 - e_3, e_2, e_3 + e_4, e_4] \in \mathbb{R}^{4\times 4}$.

3.18 (a) Construct the unique element of $\operatorname{ran}[()^0, ()^2, ()^4]$ that agrees with $()^1$ at the three points 0, 1, 2.

(b) Could (a) have been carried out if the pointset had been -1, 0, 1 (instead of 0, 1, 2)?

3.19 T/F

(a) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is in row echelon form.

(b) If all unknowns in the linear system $A? = 0$ are free, then $A = 0$;

(c) If all unknowns in the linear system $A? = 0$ are bound, then A is invertible.

(d) If some unknowns in the linear system $A? = 0$ are free, then A cannot be invertible.

(e) The inverse of an upper triangular matrix is lower triangular.

(f) A linear system of n equations in $n + 1$ unknowns always has solutions.

(g) Any square matrix in row echelon form is upper triangular.

(h) If A and B are square matrices of the same order, then $AB? = 0$ has the same number of bound unknowns as does $BA? = 0$.

(i) If A and B are square matrices of the same order, and AB is invertible, then also BA is invertible.

(j) If $\operatorname{null} A = \operatorname{null} B$, then $A? = 0$ and $B? = 0$ have the same free and bound unknowns.

