Lesson Plan 3 for Uses of the Row Echelon Form

Frank Uhlig

Department of Mathematics and Statistics,
Auburn University, Auburn, AL 36849-5310, USA
uhligfd@auburn.edu

In Lesson Plan 3 we apply row reduction to square matrix inversion and then study how to factor matrices $A_{m,n}$ as $L_{m,m} \cdot R_{m,n}$ with L lower triangular and R upper triangular.
Row echelon form reductions help us to find the column vector space of a matrix and its nullspace leading to the Dimension Theorem and to a practical definition of linear (in-)dependance for sets of vectors.
The lecture takes about 6 hours of class time, i.e., 4 hours for lectures and 2 for student discussions, explorations and questions. This slow speed is essential for engaging students in the learning process.

(version August 3, 2023) iii + 26 p.; $\sim$ 6 hours of class time

Concepts, Notions and Definitions in Lesson Plan 3

For [1] p. 2

Matrix Invertibility

For [2] p. 4

Matrix Composition

Matrix Linearity Condition

Vector Dot Product and Matrix Products

For [1p] p.8

Practical Matrix Inversion

Multi-augmented Matrix

For [1a] p. 10

Commuting Matrices

(continued)

For [3] p. 12
LR Factorization
Triangular Matrices

For [4] p. 17
Vector Spaces
Span of Vectors
Invariants of a RREF
Pivot and Free Columns
Spanning Set
Subspace and Dimension
Matrix Subspaces
Nullspace or Kernel
Dimensions Theorem

For [5] p. 22
Linear Independence
Midterm Test

In this lesson we develop useful information and how to extract it from a Row Echelon Form of any matrix $A_{m,n}$.

Specifically we explain matrix inversion and discuss the existence of inverses for linear maps and matrix mappings in an abstract and proving way.

This is followed by the mechanics of matrix multiplication $A \cdot B$*, or more abstractly of the composition* $g(f(x))$ *of linear transformations* f *and* g*.*

Then we introduce the LR matrix decomposition of arbitrary matrices $A_{m,n}$ *as the product of a lower triangular elimination matrix* L *and an upper triangular matrix* R *and learn that matrices generally do not commute, i.e., that* $A \cdot B \neq B \cdot A$ *in general.*

[1] Matrix Inversion

When can a function $f : \mathbb{R}^n \to \mathbb{R}^m$ *be inverted, i.e., when is every* $b \in \mathbb{R}^m$ *the unique image under* f *of some* $x \in \mathbb{R}^n$*?*

To be invertible, any function $f : \mathbb{R}^n \to \mathbb{R}^m$ must satisfy two conditions :

(A) If some images $b \in \mathbb{R}^m$ have multiple origins in $\mathbb{R}^n$ under f, then f cannot be inverted since invertible functions need to map consistently one-to-one.

(B) If $f(\mathbb{R}^n) \subsetneqq \mathbb{R}^m$ then the inverse cannot exist as a function with the necessary domain $\mathbb{R}^m$.

How do (A) and (B) apply to linear transformations $\mathcal{L}$ *or to matrix* $\times$ *vector products* $A_{m,n} \cdot x$ *with* $x \in \mathbb{R}^n$?

The REF of A can help us :

For matrices $A_{m,n}$ there are three cases to look at.

($m > n$) If $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ has more rows than columns, then the row echelon of the REF of the augmented matrix $(A|b)_{m,n+1}$ contains at most n pivots and $m-n \geq 1$ zero rows at the bottom. Thus there are many vectors $b \in \mathbb{R}^m$ for which the linear system $Ax = b$ has no solution and according to (B) an inverse function for A does not exist.

($m < n$) If $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ has more columns (n) than rows (m), then the row echelon of the REF of $A_{m,n}$ can contain at most m pivots and must have at least $n-m \geq 1$ free columns. Thus for any vector $b \in \mathbb{R}^m$ there are many vectors in $\mathbb{R}^n$ that A maps to b. And according to (A) the matrix A cannot be inverted.

Thus only square matrices $A_{n,n} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ can be inverted – if A satisfies both conditions (A) and (B).

For $A_{n,n}$ to be invertible, there must be n pivots in its REF. If there are fewer than n pivots, then not all linear systems $Ax = b$ can be solved for x.

Theorem : A matrix $A_{m,n}$ is invertible if and only if $m = n$ and A's Row Echelon Form contains n pivots.

[2] Function and matrix concatenations or compositions

We compose or concatenate two functions $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$ in $g \circ f : \mathbb{R}^n \to \mathbb{R}^k$ by defining $(g \circ f)(x) = g(f(x))$ for all $x \in \mathbb{R}^n$. *(In reverse order, inner f to outer g)*

How do we concatenate two linear transformations $T = f$ and $S = g$ that are given by their standard matrix representations $A_{m,n}$ for $T = f$ and $B_{k,m} = g$ for S?

$$(B \circ A)x = B(A(x)) \left[\overset{(*)}{=} (B_{k,m} \cdot A_{m,n})x \in \mathbb{R}^k \text{ for all } x \in \mathbb{R}^n \right].$$

A question for **students** :

Is the composition of linear transformations itself a linear transformation?

How can one insure, i.e. prove that the concatenation of two compatible matrices $B \circ A$ is linear, i.e., can $B \circ A$ be represented by a standard matrix?

Who can prove the linearity condition for $B \circ A$, namely that

$$
\begin{aligned}
(B \circ A)(\alpha u + \beta v) &= \ldots \ldots \\
\ldots \ldots &= \alpha(B \circ A)(u) + \beta(B \circ A)(v)
\end{aligned}
$$

for all scalars α and β and all vectors $u, v \in \mathbb{R}^n$ if A is m by n and B is k by m so that $B \circ A$ maps n vectors to k vectors.

Now the teacher best leave the classroom – and lets the students solve this question internally, *from learner to learner*.

Back to the matrix product $B \cdot A$ defining equation

$$(B_{k,m} \cdot A_{m,n})x \overset{(*)}{=} B(A(x)) = (B \circ A)(x) \;.$$

Matrix multiplication is easily established by remembering the construction of standard matrix representations of linear transformations such as $B \circ A$.

The standard matrix representation of $B \circ A : \mathbb{R}^n \to \mathbb{R}^k$ contains the images of the unit vectors e_i of $\mathbb{R}^n$ in its columns.

Starting from A, what are the column vectors $A(e_i)$ of the matrix representation of A?

They are the columns a_i of $A = \begin{pmatrix} \vdots & & \vdots \\ a_1 & \cdots & a_n \\ \vdots & & \vdots \end{pmatrix}_{m,n}$; $A(e_i) = a_i$.

How does $B_{k,m}$ then map the vectors $A(e_i) = a_i \in \mathbb{R}^m$?

By evaluating the dot product of each of B's k rows in $\mathbb{R}^m$ with each of the columns $a_i \in \mathbb{R}^m$ of A.

Forming the matrix product $B \cdot A$ of $B_{k,m}$ and $A_{m,n}$ requires $k \cdot n$ dot product evaluations between all rows (k) of the first factor B and all columns (n) of the second matrix factor A. Each of these dot products in turn requires m multiplications. If $m = n = k$ and both A and B are square, then multiplying $B \cdot A$ requires n^3 multiplications. Therefore numericalists say that n by n matrix multiplication is an $O(n^3)$ process.

Expressed graphically, matrix multiplication works in this way:

$$B_{k,m} \cdot A_{m,n} = \underset{\textit{row } i}{\begin{pmatrix} & \vdots & \\ (& \cdots b_i \cdots &) \\ & \vdots & \end{pmatrix}} \underset{\times}{\cdot} \underset{\textit{column } j}{\begin{pmatrix} \cdots & \begin{pmatrix} \vdots \\ a_j \\ \vdots \end{pmatrix} & \cdots \end{pmatrix}} = \underset{\textit{column} \uparrow j}{\begin{pmatrix} & \vdots & \\ & (\cdots b_i \cdots) \begin{pmatrix} \vdots \\ a_j \\ \vdots \end{pmatrix} & \\ & \vdots & \end{pmatrix}_{k,n}} \leftarrow i\ .$$

[1p] Practical Matrix Inversion

Here we construct the inverse X of an invertible matrix $A_{n,n}$ via the REF of A.

Clearly the matrix product $A \cdot X$ acts like the identity matrix

$$I_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

would since each column x_i of the right inverse X of A satisfies $A(x_i) = e_i$ for $i = 1, ..., n$.

If we solve $A(x_i) = e_i$ for each i, we would perform the identical row reduction process of A n times. More economically we instead work on the multi-augmented matrix $(A|I_n)_{n,2n}$. The REF of this n by $2n$ matrix would have n pivots $\boxed{1}$ on its diagonal and all zeros in its lower triangular part if we row reduced it and scaled all pivots to 1 as explained in part [2].

But the upper triangle of the leading n by n matrix would still be dense.

To assemble the columns of A^{-1} in the trailing n by n matrix in the $(A|I)$ scheme, we must reduce the leading n by n matrix part above the diagonal to create I_n in its leading square.

Therefore we eliminate the entries of the nth column in the reduced multi-augmented matrix next by subtracting appropriate multiples of its last full row from each row above. This gives us the last column x_n of the inverse of A.

Then we repeat this upper triangle zero-out process with columns $n - 1$ and go backwards and up through column 2.

Once completed, we have computed the multi-augmented matrix $(I_n|A^{-1})$ which is the Reduced Row Echelon Form or the RREF of $(A|I)$ with A and A^{-1} interchanging their positions.

[1a] Abstract Linear Algebraic Results and Commuting Matrices

We have computed the matrix inverse $X = A^{-1}$ on the right side of $A_{n,n}$ if A is invertible.

What about left inverses Y of A with $YA = I_n$? If $YA = I$ and $AX = I$, then

$$Y = YI = Y(AX) = (YA)X = IX = X\ .$$

Thus right and left inverses Y and X of an invertible matrix A are identical. Since the right inverse is uniquely determined column-wise by the solutions of $A(x_i) = e_i$, A's inverse $X = A^{-1} = Y$ is unique and commutes with A, i.e., $A^{-1}A = AA^{-1}$.

Commuting matrices are a rare breed and this area is subject of much abstract matrix research.

Students should now look for commuting matrix pairs and non-commuting ones.

The zero matrix O_n commutes with some (?) or all (?) matrices $A_{n,n}$?

How about the identity matrix I_n?

Or polynomials in A?

Construct two by two matrix pairs that commute and that do not commute.

Do $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ commute or not?

Try to construct 2 by 2 and 3 by 3 matrix pairs with no zero entries that commute and some that do not commute?

These exercises are meant to practice and perfect matrix multiplication and to learn to reason mathematically.

[3] Row Reduction and the LR factorization of Matrices

The LR factorization of a matrix $A_{n,n}$ expresses $A = L \cdot R$ with

$$L = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ * & & 1 \end{pmatrix} \text{ lower triangular and } R = \begin{pmatrix} * & & * \\ & \ddots & \\ 0 & & * \end{pmatrix}$$

upper triangular n by n.

Warning : Not all square matrices have an LR factorization.

Show that $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *cannot be LR factored. Why?*

Assume that $A_{n,n}$ can be row reduced to a RREF without the need to switch rows in the process. Then our row reduction algorithm uses two actions repeatedly: scale the current pivot candidate to become $\boxed{1}$, and then subtract certain multiples of the pivot row from lower rows so that all entries in the pivot column below the nonzero pivot itself become zero.

Scaling the ith pivot row can be effected by left multiplying the row vector $s = (0, ..., 0, s_i = 1/\hat{a}_{i,i}, 0, ..., 0)$ with the partially reduced matrix $\hat{A}$ on the right.

In matrix notation, the scaling matrix

$$S = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & 0 & \\ & & 1 & & & & \\ & & & s_i & & & \\ & & & & \ddots & & \\ & 0 & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}_{n,n} \leftarrow i$$

effects the same scaling of row i and changes the ith pivot in $S \cdot \hat{A}$ to $\boxed{1}$.

Subsequent row reduction of the entries $b_{j,i}$ with $j > i$ in column i of $\hat{\hat{A}} = S \cdot \hat{A}$ is achieved again by left matrix multiplication of $\hat{\hat{A}}$ with

$$E_i = \begin{pmatrix} 1 & & \overset{\downarrow i}{0} & & & \\ & \ddots & & & 0 & \\ & 0 & 1 & 0 & & \\ & & -b_{i+1,i} & 1 & & \\ & 0 & \vdots & 0 & \ddots & 0 \\ & & -b_{n,i} & & & 1 \end{pmatrix}_{n,n} \begin{matrix} \\ \\ \leftarrow i \\ \\ \\ \\ \end{matrix}$$

for $i = 1, \ldots, n-1$ and $b_{j,i}$ denoting the entry of $\hat{\hat{A}}$ in row j and column i for $j > i$.

Note that the scaling matrices S and the elimination matrices E_i all have zeros in their upper triangle above and to the right of the diagonal. Such matrices are called lower triangular.

Theorem : Matrix products of lower triangular n by n matrices L and M remain lower triangular.

Proof : *What is the dot product of row* $l_j = (*, .., *_j, 0, ..., 0)$ *of* L *and column* $m_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ *_k \\ \vdots \\ * \end{pmatrix}$ *of* M *when* $j < k$ *and* $l_y \times m_k$ *evaluates the* (j, k) *entry of* $L \cdot M$ *in its upper triangle?*

Can the students figure this out? How do the $*$ *entries and the zero entries in* l_j *and* m_k*, respectively, match up if* $j < k$*?*

Theorem : The inverse of a lower triangular matrix L is lower triangular if L can be inverted. *Proof by students ... Hint: the REF of* L *is lower triangular.*

Combining all our insights that pertain to the row reduction of square matrices A via lower triangular matrix multiplication by M from the left, we understand that M transforms A to $M \cdot A = R$ in upper triangular form.

Thus $A = M^{-1}R = LR$ where $L = M^{-1}$ is lower triangular as claimed in the LR factorization for matrices A that allow such.

Historically, Heinz Rutishauser in 1957 was the first to propose matrix factorizations and reverse order multiplies in the ill-fated LR matrix factorization when trying to find intrinsic properties of matrices. This was quickly improved by John Francis - who dropped out of Cambridge after the first year - and by Vera Kublanovskaya in Leningrad around 1960. Both used orthogonal matrix factorizations instead of LR and invented the QR eigenvalue algorithm, our workhorse today.

[4] Sets of Vectors, Vector Spaces, Spanning Sets, Bases, and Subspaces Associated with Matrices

Here we study sets of $k \geq 1$ vectors in $u_i \in \mathbb{R}^n$ where $k < n$, $k = n$ or $k > n$.

The vectors $\{u_i\}$ and all their linear combinations (also called the span$(\{u_i\})$) form a subspace U of $\mathbb{R}^n$.

If $k = 1$ and $u_1 \neq 0$, then span(u_1) is the line through u_1 and the origin 0 of $\mathbb{R}^n$.

If $k = 2$ and $u_1 = u_2 = 0$, then span$(\{u_1, u_2\})$ consists of the origin $0 \in \mathbb{R}^n$.

It is the line through u_1 and the origin 0 if $u_1 = \alpha u_2 \neq 0 \in \mathbb{R}^n$.

span$(\{u_1, u_2\})$ is the plane spanned by $0, u_1$ and u_2 if u_1 and u_2 are both nonzero and do not lie on a line through 0.

Elementary geometry does not help us much further here, we need to use matrix theory to sort vector spans and vector spaces out.

For any set of k vectors $\{u_i\} \subset \mathbb{R}^n$ we now study the n by k column vector matrix $U = \begin{pmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_k \\ \vdots & & \vdots \end{pmatrix}_{n,k}$ and its row echelon form $R_{n,k}$ instead.

For students, this is a different 'abstract' matrix challenge than 'theorems' and 'proving'. We need to mentally visualize what a REF of any matrix $U_{n,k}$ signifies, what its invariants are; by whomever and however it was computed.

What are the invariants of the REF R of any specific matrix U?

The most significant information of a REF $R_{n,k}$ are the number p of pivot columns and their position and complementarily, the position and number f of free columns in $R_{n,k}$.

How many pivots p_{max} can $R_{n,k}$ maximally have?

$p_{max} = min(n, k)$.

How many free columns f_{min} can $R_{n,k}$ minimally have?

$f_{min} = k - p$.

The minimal number p_{min} of pivots in a REF $R_{n,k}$ is 0.

The maximal number of free columns f_{max} is $f_{max} = min(n, k)$.

Thus $0 \leq p \leq min(n, k)$ and $f = k - p$ because U's REF $R_{n,k}$ has k columns and each column is either pivot or free.

If $p = 0$, then the column vector matrix $U = O_{n,k}$ and all vectors u_i are the zero vector. And their span is the singleton vector $\{o\} \in \mathbb{R}^n$ which contains all of its linear combinations.

Each free column associated vector u_i can be expressed as a linear combination of the preceding pivot associated vectors. The pivot associated column vectors u_j of U suffice to generate every vector in span($\{u_i\}$). And each pivot associated column vector is necessary to span their generated subspace.

Therefore the pivot associated vectors u_j of U are a minimal spanning set for span($\{u_i\}$); no fewer will do and no more are needed to reach each vector in span($\{u_i\}$).

Definition : A minimal spanning set for a subspace $U \subset \mathbb{R}^n$ is called a basis for U. The number of basis vectors for a subspace $U \subset \mathbb{R}^n$ is called its dimension, $0 \leq \dim(U) \leq n$.

Definition : Given an n by k matrix A, the set of vectors $\{b = Ax \mid x \in \mathbb{R}^k\}$ is called the image $\text{im}(A)$ of A or the range space of A.

Definition : Given an n by k matrix A, the set of vectors $\{x \in \mathbb{R}^k \mid Ax = 0 \in \mathbb{R}^n\}$ is called the nullspace $\mathrm{null}(A)$ of A or the kernel of A.

Theorem : Given an n by k matrix A, the dimensions of $\mathrm{im}(A)$ and $\mathrm{null}(A)$ add up to the number of columns k of $A_{n,k}$, i.e.,

$$\dim(\mathrm{im}(A)) + \dim(\mathrm{null}(A)) = k \ .$$

This should become quickly obvious to **students***:*
what are the dimensions of the image space and of the nullspace of A as indicated by the pivot columns and free columns in A's REF.

And an open question for teacher and students:
How can we find a basis for the nullspace of a matrix from its REF?

How to solve $Ax = 0$?

[5] Linear (In-)dependence of Vectors

The logic based definition of linear independence for a set of vectors that is given in one form or another in every classical abstract Linear Algebra class reads as follows :

Classical abstract Definition : A set of vectors u_i is linearly independent if every linear combination of the u_i that equals the zero vector has zero coefficients.

Then students in an abstract Linear Algebra course are usually given several sets of vectors in $\mathbb{R}^n$ *and asked to reason and decide which sets are linearly independent and which not.*

How would a matrix based class approach this problem?

A linear combination of vectors u_i is the matrix $\times$ vector product of the column vector matrix U and the coefficient vector x.

When is $U \cdot x = 0$*?*

$U \cdot x = 0$ whenever $x \in \text{null}(U)$. When is $\text{null}(U) = 0$?
This happens when the Ref of U has only pivot columns and no free variables. Thus we have :

Matrix Theory based Definition of Linear Independence :

A set of vectors u_i is linearly independent if the REF of the column vector matrix U has as many pivots as the number of vectors u_i.

Now add some simple 'classical' homework problems ...

(1) If $o \in \{u_i\}$ then the u_i are linearly ...

(2) If $u_1 - 4u_3 \in \{u_i\}$ then the u_i are linearly ...

(3) , (4), ...

Classical abstract Definition of Linear Dependence :

A set of vectors u_i is linearly dependent if one of the vectors u_j is a linear combination of the others.

From Matrix Theory we now know that a set of vectors $\{u_i\}$ is **linearly dependent** if the **REF** of the column vector matrix U of the u_i has **a free variable**. *And we are done; in theory.*

Numerically :

As before, here are the steps to let software find bases for the nullspace and the image space of a set of vectors $\{u_i\}$ and $i = 1, .., m$ that are arranged column-wise in the matrix $U_{n,m}$.

For finding a basis for the image space of U, transpose U, i.e., write the vectors u_i row-wise and form $U^T_{m,n}$, the transpose of U and compute the RREF of U^T.

The RREF's pivot rows, transposed, are a basis of im(U).

To find a basis of the nullspace of U solve `U.'\zeros(m,n)` in Matlab.

Early Midterm Test

What are the invariants of a REF of a matrix ?

Assume that an instructor asks her/his 214 or more or fewer students to row reduce a given matrix $A_{m,n}$ to row echelon form $R_{m,n}$ in whatever sequence of operations on A they choose.

What will be common and the same in all 214 handed in tests?

The *'invariants'*.

What in the REFs can differ?

Next, Lesson Plan 4 will deal with matrix representations with respect to arbitrary bases, extending the standard matrix representation idea to allow and account for the intrinsic properties of each specific matrix such as its eigenvectors, eigenvalues and eigenspaces.

Lesson 4 will be the **turning point** *from simple linear properties in the behavior of matrices and the REF to nonlinear properties, actions and classifications of matrices.*

Matrix eigenvectors and eigenvalues will then be introduced and found via Krylov subspaces and vector iteration in subsequent lessons 5 through 7.