Fall 2019 Math 201-003 at Drexel: blackboard notes of 2019-10-21

Darij Grinberg

February 12, 2021

1. Block matrices (aka partitioned matrices)

This little chapter follows [Lay, §2.4].

We have already seen constructions like $\begin{bmatrix} A \\ B \end{bmatrix}$ or $[A \mid B]$: these were two ways of gluing two matrices A and B together.

This can be generalized. For example, you can glue four matrices A, B, C, D together to $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ if their dimensions match appropriately. You can do this with more than four matrices, too. Essentially, you can start with any "matrix of matrices", and glue them together into a single big matrix.

Definition 1.0.1. Let n and m be two nonnegative integers. Assume you have nm many matrices $A_{1,1}, A_{1,2}, \ldots, A_{n,m}$ (that is, you have a matrix $A_{i,j}$ for each $i \in \{1, 2, \ldots, n\}$ and each $j \in \{1, 2, \ldots, m\}$). Then,

$$\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,m} \end{bmatrix}$$

denotes the matrix obtained by gluing all these nm many matrices together into a single big matrix. This is only defined if their dimensions match: that is,

- For any given $j \in \{1, 2, \ldots, m\}$, the n matrices $A_{1,j}, A_{2,j}, \ldots, A_{n,j}$ must have the same number of columns.
- For any given $i \in \{1, 2, \ldots, n\}$, the m matrices $A_{i,1}, A_{i,2}, \ldots, A_{i,m}$ must have the same number of rows.

Example 1.0.2. Let

$$A = \begin{pmatrix} a & a' \\ a'' & a''' \end{pmatrix}, \qquad B = \begin{pmatrix} b & b' \\ b'' & b''' \end{pmatrix},$$
$$C = \begin{pmatrix} c & c' \\ c'' & c''' \end{pmatrix}, \qquad D = \begin{pmatrix} d & d' \\ d'' & d''' \end{pmatrix}.$$

Then,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{pmatrix} a & a' & b & b' \\ a'' & a''' & b'' & b''' \\ c & c' & d & d' \\ c'' & c''' & d'' & d''' \end{pmatrix}.$$

Example 1.0.3. Let

$$A = \begin{pmatrix} a & a' \\ a'' & a''' \end{pmatrix}, \qquad B = \begin{pmatrix} b \\ b'' \end{pmatrix},$$
$$C = \begin{pmatrix} c & c' \end{pmatrix}, \qquad D = \begin{pmatrix} d \end{pmatrix}.$$

Then,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{pmatrix} a & a' & b \\ a'' & a''' & b'' \\ c & c' & d \end{pmatrix}.$$

Convention 1.0.4. The matrix

$$\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,m} \end{bmatrix}$$

in Definition 1.0.1 is called the **block matrix** (or **partitioned matrix**) built from the nm matrices $A_{1,1}, A_{1,2}, \ldots, A_{n,m}$. Its construction is called the **block-matrix construction**. We will always use **square brackets** [] to signalize that we are gluing together a bunch of matrices, whereas **parentheses** () simply mean that we are forming a matrix out of numbers. So, the things inside square brackets are matrices to be glued together, rather than just numbers to be put each in a single cell.

(This notation is not standard; texts like [Strickland] use square brackets for **any** matrix.)

The little matrices $A_{i,j}$ are called the **blocks** in the block matrix $\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,m} \end{bmatrix}$.

You can imagine the blocks being separated by horizontal and vertical lines as follows:

$$\begin{array}{|c|c|c|c|} \hline A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ \hline A_{2,1} & A_{2,2} & \cdots & A_{2,m} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline A_{n,1} & A_{n,2} & \cdots & A_{n,m} \\ \hline \end{array}.$$

These lines are called the **dividers** of our block matrix. (Of course, they are not a property of the resulting matrix, but of the way it has been glued together.)

So, on the surface, block matrices are a way to just pigeonhole the contents of a matrix into some sort of boxes – an organizational tool. However, it turns out to have extra advantages in the sense that **sometimes**, you can manipulate block matrices on the level of blocks.

The most important such situation is when you want to multiply two block matrices: For example, it would be great if the equality

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} U & V \\ W & T \end{bmatrix} = \begin{bmatrix} AU+BW & AV+BT \\ CU+DW & CV+DT \end{bmatrix}$$

was true, meaning that we could multiply two block matrices "as if their blocks were entries".

In full generality, this is not true, because depending on the dimensions of the blocks, products like AU might not even be well-defined. But let us assume that they are defined. For example:

Example 1.0.5. Let

$$A = \begin{pmatrix} a \end{pmatrix}, \qquad B = \begin{pmatrix} b & b' \end{pmatrix}, \qquad C = \begin{pmatrix} c \\ c' \end{pmatrix}, \qquad D = \begin{pmatrix} d & d' \\ d'' & d''' \end{pmatrix},$$

$$U = \begin{pmatrix} u \end{pmatrix}, \qquad V = \begin{pmatrix} v & v' \end{pmatrix}, \qquad W = \begin{pmatrix} w \\ w' \end{pmatrix}, \qquad T = \begin{pmatrix} t & t' \\ t'' & t''' \end{pmatrix}.$$

Then,

$$\begin{aligned} & \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} U & V \\ W & T \end{bmatrix} \\ &= \begin{pmatrix} a & b & b' \\ c & d & d' \\ c' & d'' & d''' \end{pmatrix} \begin{pmatrix} u & v & v' \\ w & t & t' \\ w' & t'' & t''' \end{pmatrix} \\ &= \begin{pmatrix} au+bw+b'w' & av+bt+b't'' & av'+bt'+b't''' \\ cu+dw+d'w' & cv+dt+d't'' & cv'+dt'+d't''' \\ d''w+c'u+d'''w' & d''t+c'v+d'''t'' & d''t'+d'''t'''+v'c' \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned}
&\left[\begin{array}{cc} AU+BW & AV+BT \\ CU+DW & CV+DT \end{array}\right] \\
&= \begin{pmatrix} au+bw+b'w' & av+bt+b't'' & av'+bt'+b't''' \\ cu+dw+d'w' & cv+dt+d't'' & cv'+dt'+d't''' \\ d''w+c'u+d'''w' & d''t+c'v+d'''t'' & d''t'+d'''t'''+v'c' \end{pmatrix}.
\end{aligned}$$

Thus,

$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]\left[\begin{array}{cc} U & V \\ W & T \end{array}\right] = \left[\begin{array}{cc} AU+BW & AV+BT \\ CU+DW & CV+DT \end{array}\right]$$

is true in this case!

This holds more generally:

Theorem 1.0.6. Let A, B, C, D, U, V, W and T be matrices that can be glued together to block matrices $\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]$ and $\left[\begin{array}{cc} U & V \\ W & T \end{array}\right]$. Then,

$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]\left[\begin{array}{cc} U & V \\ W & T \end{array}\right] = \left[\begin{array}{cc} AU+BW & AV+BT \\ CU+DW & CV+DT \end{array}\right]$$

if

$$\begin{aligned}
&(\text{the number of columns of } A) = (\text{the number of rows of } U) \qquad \text{and} \\
&(\text{the number of columns of } B) = (\text{the number of rows of } W).
\end{aligned}$$

Proof idea. The product

$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]\left[\begin{array}{cc} U & V \\ W & T \end{array}\right]$$

has as many rows as $\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]$ and as many columns as $\left[\begin{array}{cc} U & V \\ W & T \end{array}\right]$. Thus, we can write this product as a block matrix $\left[\begin{array}{cc} P & Q \\ R & S \end{array}\right]$ whose horizontal divider is at the same position as the horizontal divider of $\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]$ (that is, P has as many rows as A, and R has as many rows as C) and whose vertical divider is at the same position as the vertical divider of $\left[\begin{array}{cc} U & V \\ W & T \end{array}\right]$ (that is, P has as many columns as U, and Q has as many columns as V).

Now, consider the (i,j)-th entry of S for some integers i and j. This is an entry of the block matrix

$$\left[\begin{array}{cc} P & Q \\ R & S \end{array}\right] = \left[\begin{array}{cc} A & B \\ C & D \end{array}\right]\left[\begin{array}{cc} U & V \\ W & T \end{array}\right]$$

that lies i rows below the horizontal divider and j columns east of the vertical divider. Hence, it must equal

$$\begin{aligned}
& \operatorname{row}_i \left[\begin{array}{cc} C & D \end{array}\right] \cdot \operatorname{col}_j \left[\begin{array}{c} V \\ T \end{array}\right] \qquad \text{(by [lina, Proposition 2.19 }\textbf{(b)}\text{])} \\
& = \left[\begin{array}{cc} \operatorname{row}_i C & \operatorname{row}_i D \end{array}\right] \cdot \left[\begin{array}{c} \operatorname{col}_j V \\ \operatorname{col}_j T \end{array}\right] \\
& \qquad \left(\text{since } \operatorname{row}_i \left[\begin{array}{cc} C & D \end{array}\right] = \left[\begin{array}{cc} \operatorname{row}_i C & \operatorname{row}_i D \end{array}\right] \text{ and } \operatorname{col}_j \left[\begin{array}{c} V \\ T \end{array}\right] = \left[\begin{array}{c} \operatorname{col}_j V \\ \operatorname{col}_j T \end{array}\right]\right) \\
& = \underbrace{(\operatorname{row}_i C) \cdot (\operatorname{col}_j V)}_{=(CV)_{i,j}} + \underbrace{(\operatorname{row}_i D) \cdot (\operatorname{col}_j T)}_{=(DT)_{i,j}} \\
& \qquad \left(\begin{array}{c} \text{because when you multiply the row vector } \left[\begin{array}{cc} \operatorname{row}_i C & \operatorname{row}_i D \end{array}\right] \\ \text{with the column vector } \left[\begin{array}{c} \operatorname{col}_j V \\ \operatorname{col}_j T \end{array}\right], \\ \text{you get a sum where the first few addends are products} \\ \text{of an entry of } \operatorname{row}_i C \text{ with an entry of } \operatorname{col}_j V, \\ \text{while the remaining addends are products} \\ \text{of an entry of } \operatorname{row}_i D \text{ with an entry of } \operatorname{col}_j T \end{array}\right) \\
& = (CV)_{i,j} + (DT)_{i,j} = (CV + DT)_{i,j}.
\end{aligned}$$

Thus, $S = CV + DT$. This explains why a part of the matrix

$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right] \left[\begin{array}{cc} U & V \\ W & T \end{array}\right]$$

(namely, the bottom-right part) is a copy of $CV + DT$. Similarly, you can find copies of $AU + BW$ and $AV + BT$ and $CU + DW$ in the appropriate positions. Thus, the matrix equals $\left[\begin{array}{cc} AU + BW & AV + BT \\ CU + DW & CV + DT \end{array}\right]$. □

Even more generally, you can always multiply two block matrices "as if the blocks were entries", provided that the products are well-defined (i.e., that the sizes of the parts into which the vertical dividers subdivide the left matrix are exactly the sizes of the parts into which the horizontal dividers subdivide the right matrix). Here is a more rigorous (and probably clearer, too) way to state this:

Theorem 1.0.7. Consider two block matrices

$$\left[\begin{array}{cccc} A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,m} \end{array}\right] \qquad \text{and} \qquad \left[\begin{array}{cccc} B_{1,1} & B_{1,2} & \cdots & B_{1,p} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m,1} & B_{m,2} & \cdots & B_{m,p} \end{array}\right]$$

such that for each $j \in \{1, 2, \ldots, m\}$, we have

$$\begin{aligned} &\left(\text{the number of columns of } A_{i,j} \text{ for any } i\right) \\ &= \left(\text{the number of rows of } B_{j,k} \text{ for any } k\right). \end{aligned}$$

Then,

$$\begin{aligned} &\left[\begin{array}{cccc} A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,m} \end{array}\right] \cdot \left[\begin{array}{cccc} B_{1,1} & B_{1,2} & \cdots & B_{1,p} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m,1} & B_{m,2} & \cdots & B_{m,p} \end{array}\right] \\ &= \left[\begin{array}{ccc} A_{1,1}B_{1,1} + A_{1,2}B_{2,1} + \cdots + A_{1,m}B_{m,1} & \cdots & A_{1,1}B_{1,p} + A_{1,2}B_{2,p} + \cdots + A_{1,m}B_{m,p} \\ A_{2,1}B_{1,1} + A_{2,2}B_{2,1} + \cdots + A_{2,m}B_{m,1} & \cdots & A_{2,1}B_{1,p} + A_{2,2}B_{2,p} + \cdots + A_{2,m}B_{m,p} \\ \vdots & \ddots & \vdots \\ A_{n,1}B_{1,1} + A_{n,2}B_{2,1} + \cdots + A_{n,m}B_{m,1} & \cdots & A_{n,1}B_{1,p} + A_{n,2}B_{2,p} + \cdots + A_{n,m}B_{m,p} \end{array}\right]. \end{aligned}$$

(The right hand side here is a block matrix, whose (i,j)-th block is $A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j}$.)

Block matrices have even nicer properties when some of the blocks are zero matrices:

Definition 1.0.8. Let $D_1, D_2, \ldots, D_n$ be any n square matrices (not necessarily of the same size). Then, the block matrix

$$\left[\begin{array}{cccc} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n \end{array}\right]$$

(where the 0s stand for zero matrices of the appropriate sizes) is called a **block-diagonal matrix** with diagonal blocks $D_1, D_2, \ldots, D_n$.

Example 1.0.9. For example,

$$\begin{pmatrix} a & a' & 0 \\ a'' & a''' & 0 \\ 0 & 0 & b \end{pmatrix}$$

is a block-diagonal matrix with blocks $D_1 = \begin{pmatrix} a & a' \\ a'' & a''' \end{pmatrix}$ and $D_2 = \begin{pmatrix} b \end{pmatrix}$.

Of course, any square matrix is a block-diagonal matrix, with a single diagonal block. But you don't gain much from this point of view. Block-diagonality is at its most useful when the blocks are small.

Proposition 1.0.10. Consider two block-diagonal matrices

$$\begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} E_1 & 0 & \cdots & 0 \\ 0 & E_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_n \end{bmatrix},$$

where the size of each block D_i equals the size of the corresponding E_i. Then, the product of these matrices can be computed blockwise:

$$\begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n \end{bmatrix} \cdot \begin{bmatrix} E_1 & 0 & \cdots & 0 \\ 0 & E_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_n \end{bmatrix} = \begin{bmatrix} D_1E_1 & 0 & \cdots & 0 \\ 0 & D_2E_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_nE_n \end{bmatrix}.$$

Example 1.0.11. Let us see what this says in the case when $n = 2$ and all blocks D_1, D_2, E_1, E_2 are 2×2-matrices:

$$\begin{pmatrix} a_1 & a_2 & 0 & 0 \\ a_1' & a_2' & 0 & 0 \\ 0 & 0 & b_1 & b_2 \\ 0 & 0 & b_1' & b_2' \end{pmatrix} \begin{pmatrix} x_1 & x_2 & 0 & 0 \\ x_1' & x_2' & 0 & 0 \\ 0 & 0 & y_1 & y_2 \\ 0 & 0 & y_1' & y_2' \end{pmatrix}$$
$$= \begin{pmatrix} a_1x_1 + a_2x_1' & a_1x_2 + a_2x_2' & 0 & 0 \\ a_1'x_1 + a_2'x_1' & a_1'x_2 + a_2'x_2' & 0 & 0 \\ 0 & 0 & b_1y_1 + b_2y_1' & b_1y_2 + b_2y_2' \\ 0 & 0 & b_1'y_1 + b_2'y_1' & b_1'y_2 + b_2'y_2' \end{pmatrix}.$$

Definition 1.0.12. Let $D_1, D_2, \ldots, D_n$ be any n square matrices (not necessarily of the same size). Then, the block matrix

$$\begin{bmatrix} D_1 & * & \cdots & * \\ 0 & D_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n \end{bmatrix}$$

(where the $*$s are arbitrary matrices) is called a **block-upper-triangular matrix** with diagonal blocks $D_1, D_2, \ldots, D_n$.

Example 1.0.13. Consider a block-upper-triangular matrix $\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$ with A and D being square matrices. Assume that A and D are invertible. Is the whole matrix $\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$ invertible, too? What is its inverse?

Let A be of size $n \times n$, and let D be of size $m \times m$.

Let us first assume that $\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$ is invertible, and try to find its inverse. (Once we have a formula for the inverse, we will then verify directly that it is indeed an inverse, and thus our invertibility assumption will not be necessary anymore.)

Let the inverse of $\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$ be $\begin{bmatrix} X & Y \\ Z & W \end{bmatrix}$ (with X being of size $n \times n$ and W of size $m \times m$). Thus,

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \cdot \begin{bmatrix} X & Y \\ Z & W \end{bmatrix} = I_{n+m} = \begin{bmatrix} I_n & 0 \\ 0 & I_m \end{bmatrix}.$$

Comparing this with

$$\begin{aligned} \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \cdot \begin{bmatrix} X & Y \\ Z & W \end{bmatrix} &= \begin{bmatrix} AX + BZ & AY + BW \\ 0X + DZ & 0Y + DW \end{bmatrix} \qquad \text{(by Theorem 1.0.6)} \\ &= \begin{bmatrix} AX + BZ & AY + BW \\ DZ & DW \end{bmatrix}, \end{aligned}$$

we obtain

$$\begin{bmatrix} AX + BZ & AY + BW \\ DZ & DW \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_m \end{bmatrix}.$$

Since the block matrices on both sides of this equality have the same dimensions, we can rewrite this equality as the following system of equations:

$$\begin{aligned} AX + BZ &= I_n; \\ AY + BW &= 0; \\ DZ &= 0; \\ DW &= I_m. \end{aligned}$$

The fourth equation ($DW = I_m$) shows that W is a right inverse of D, and thus must be the inverse of D (since a right inverse of a square matrix must always be its inverse). That is, $W = D^{-1}$. Thus, from $DZ = 0$, we can obtain $Z = 0$ (just compare $\underbrace{D^{-1}D}_{=I_m} Z = I_m Z = Z$ with $D^{-1} \underbrace{DZ}_{=0} = D^{-1}0 = 0$). Thus, the first two of our above four equations become

$$\begin{aligned} AX &= I_n; \\ AY + BD^{-1} &= 0. \end{aligned}$$

Now, $AX = I_n$ entails that X is a right inverse of A, and thus is the inverse of A (since a right inverse of a square matrix must always be its inverse). Thus, $X = A^{-1}$. Hence, the equation $AY + BD^{-1} = 0$ can be solved for Y (by multiplying with A^{-1} on the left and bringing $A^{-1}BD^{-1}$ on the right hand side): It gives $Y = -A^{-1}BD^{-1}$.

(Be careful when solving these linear equations: Matrix multiplication is not commutative, so $A^{-1}B$ is not the same as BA^{-1}, and you cannot write $\dfrac{B}{A}$ for it.)

So now we know that $X = A^{-1}$, $Y = -A^{-1}BD^{-1}$, $Z = 0$ and $W = D^{-1}$. Hence, $\begin{bmatrix} X & Y \\ Z & W \end{bmatrix} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}$. That is, the putative inverse of $\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$ is $\begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}$.

Let us now check that this is indeed the case:

$$\begin{aligned}
\begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \cdot \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix} &= \begin{bmatrix} AA^{-1} + B0 & A\left(-A^{-1}BD^{-1}\right) + BD^{-1} \\ 0A^{-1} + D0 & 0\left(-A^{-1}BD^{-1}\right) + DD^{-1} \end{bmatrix} \\
&\qquad \text{(by Theorem 1.0.6)} \\
&= \begin{bmatrix} I_n & 0 \\ 0 & I_m \end{bmatrix} = I_{n+m}
\end{aligned}$$

and (by a similar computation)

$$\begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix} \cdot \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = I_{n+m}.$$

Thus, the matrix $\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$ is indeed invertible and its inverse is $\begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}$.

Caution: When multiplying two block matrices "as if the blocks were numbers", make sure to keep the order of the factors intact!

$$\begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = AX + BY \text{ is not the same as } XA + BY$$

(and XA might not even be well-defined). In every product of blocks, the factor that comes from the left matrix must come first (because matrix multiplication is not commutative).

2. Determinants

Now, we will study determinants of (square) matrices. I will follow [Strickland, Appendix B and §12], which is the best introduction I know to this subject.

2.1. Motivation

Back in the 2019-10-16 class, we have learned how to find the inverse of a square matrix (if it exists). We can use this tactic to compute the inverse of an arbitrary 2×2-matrix $\begin{pmatrix} a & b \\ a' & b' \end{pmatrix}$, with the caveat that we sometimes have to divide by certain numbers that may be 0. But if we assume that all these numbers are nonzero (unless they simplify to 0 for all a, b, a', b'), we get the result $\begin{pmatrix} \dfrac{b'}{ab'-ba'} & \dfrac{-b}{ab'-ba'} \\ \dfrac{-a'}{ab'-ba'} & \dfrac{a}{ab'-ba'} \end{pmatrix}$.

So the inverse of $\begin{pmatrix} a & b \\ a' & b' \end{pmatrix}$ is the matrix $\begin{pmatrix} \dfrac{b'}{ab'-ba'} & \dfrac{-b}{ab'-ba'} \\ \dfrac{-a'}{ab'-ba'} & \dfrac{a}{ab'-ba'} \end{pmatrix}$, as long as $ab' - ba' \neq 0$. Each entry in this matrix contains the denominator $ab' - ba'$. Moreover, (as you can check – it is not obvious) this denominator determines whether the matrix has an inverse or not:

- If $ab' - ba' = 0$, then $\begin{pmatrix} a & b \\ a' & b' \end{pmatrix}$ has no inverse.
- If $ab' - ba' \neq 0$, then $\begin{pmatrix} a & b \\ a' & b' \end{pmatrix}$ has an inverse, which is $\begin{pmatrix} \dfrac{b'}{ab'-ba'} & \dfrac{-b}{ab'-ba'} \\ \dfrac{-a'}{ab'-ba'} & \dfrac{a}{ab'-ba'} \end{pmatrix}$.

This is nicer than one might expect.

Similarly, we can find the inverse of a 3×3-matrix:

$$\begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}^{-1} = \begin{pmatrix} \dfrac{b'c''-b''c'}{D} & \dfrac{cb''-bc''}{D} & \dfrac{bc'-cb'}{D} \\ \dfrac{a''c'-a'c''}{D} & \dfrac{ac''-ca''}{D} & \dfrac{ca'-ac'}{D} \\ \dfrac{a'b''-a''b'}{D} & \dfrac{ba''-ab''}{D} & \dfrac{ab'-ba'}{D} \end{pmatrix}$$

where $D = ab'c'' + bc'a'' + ca'b'' - ac'b'' - ba'c'' - cb'a''$. Again, all entries have the same denominator, D. Again, D determines whether the inverse exists: If $D = 0$, then $\begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}$ has no inverse; otherwise, D has an inverse, given by the above formula.

What is this number D ? It is a sum of 6 products, each with a $+$ or $-$ sign. Three of the products have a $+$ sign, and the other three have a $-$ sign. Each product contains exactly one entry from each row of the matrix, and exactly one entry from each column of the matrix.

Wouldn't it be nice to have such a D for all sizes of square matrices?

So here is our **wishlist:** We want to assign to each $n \times n$-matrix A a number $\det A$ (which would generalize the D for 3×3-matrices, and the $ab' - ba'$ for 2×2-matrices) with the following properties:

- The number $\det A$ is a sum of products of entries of A, some with $-$ signs.
- Each of these products has n factors, with one from each row and one from each column of the matrix.
- The matrix A is invertible if and only if $\det A \neq 0$.

We have found such numbers for 2×2- and 3×3-matrices:

$$\det \begin{pmatrix} a & b \\ a' & b' \end{pmatrix} = ab' - ba';$$

$$\det \begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix} = ab'c'' + bc'a'' + ca'b'' - ac'b'' - ba'c'' - cb'a''.$$

Also, it is easy to check that

$$\det \begin{pmatrix} a \end{pmatrix} = a.$$

How do we define $\det A$ in general, when A is an arbitrary $n \times n$-matrix?

- Ideally, $\det A$ should be a sum over all possible "generalized diagonals" of A, where a "generalized diagonal" is a selection of n entries with one taken from each row and one taken from each column. For example, the generalized diagonals of $\begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}$ are

  $$\left(a, b', c''\right), \qquad \left(a, c', b''\right), \qquad \left(b, a', c''\right), \qquad \left(b, c', a''\right),$$
  $$\left(c, a', b''\right), \qquad \left(c, b', a''\right).$$

  (The order doesn't matter: i.e., we don't distinguish between two "generalized diagonals" that select the same entries in a different order. So (a, b', c'') is considered the same generalized diagonal as (b', a, c''). Thus, we can always list a "generalized diagonal" in the top-to-bottom order, starting with the entry in row 1, and moving down.)

- Each of these "generalized diagonals" is multiplied through and gets a sign ($+$ or $-$). What is the rule for these signs? For example, here they are for our 3×3-matrix:

$$+\left(a, b', c''\right), \qquad -\left(a, c', b''\right), \qquad -\left(b, a', c''\right), \qquad +\left(b, c', a''\right),$$
$$+\left(c, a', b''\right), \qquad -\left(c, b', a''\right).$$

2.2. Permutations

In order to formalize the "generalized diagonals" without even speaking of matrices, and subsequently to define their signs, we will use **permutations**:

Definition 2.2.1. Let n be a nonnegative integer. Let $[n]$ be the n-element set $\{1, 2, \ldots, n\}$.

A **permutation** of $[n]$ means a bijective map from $[n]$ to $[n]$.

Here, a map f from a set X to a set Y is said to be

- **injective** (aka 1-to-1) if it sends distinct elements of X to distinct elements of Y (that is, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$);
- **surjective** (aka onto) if every element of Y is taken as a value of f (that is, for each $y \in Y$, there exists some $x \in X$ such that $f(x) = y$);
- **bijective** if it is both injective and surjective.

You can catch up on basic properties of maps (also known as functions) from sources such as [LeLeMe16, §4.4–4.5], [Hammac15, §12.2] or [Day16, §3.F].

We can write a map $f : X \to Y$ in **two-line notation**, meaning that we list the elements of X as $x_1, x_2, \ldots, x_k$ (in some order) and write $f = \begin{pmatrix} x_1 & x_2 & \cdots & x_k \\ f(x_1) & f(x_2) & \cdots & f(x_k) \end{pmatrix}$ (not a matrix, just a table of values of f). Using this notation, the permutations of $[3]$ are

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

The permutations of $[n]$ will be used to index the products that appear in $\det A$ when A is an $n \times n$-matrix.

If A is any $n \times n$-matrix, and σ is a permutation of $[n]$, then we can obtain a "generalized diagonal" of A by picking

- the $\sigma(1)$-st entry from the 1-st row of A;

- the $\sigma(2)$-nd entry from the 2-nd row of A;
- the $\sigma(3)$-rd entry from the 3-rd row of A;
- and so on

(i.e., we pick the $\sigma(i)$-th entry from row i of A for each $i \in \{1, 2, \ldots, n\}$). The product of these entries is

$$A_{1,\sigma(1)} A_{2,\sigma(2)} \cdots A_{n,\sigma(n)}.$$

This product contains exactly one entry from each row (clearly), and furthermore contains exactly one entry from each column (indeed, the injectivity of σ shows that it contains no two entries from the same column, whereas the surjectivity of σ shows that it contains at least one entry from each column). Thus, the entries in this product form a "generalized diagonal".

Thus, we have a guess which products should occur in $\det A$: namely, the products

$$A_{1,\sigma(1)} A_{2,\sigma(2)} \cdots A_{n,\sigma(n)} \qquad \text{for } \sigma \text{ being a permutation of } [n].$$

But how do we decide what signs ($+$ or $-$) they get?

So we want to assign a sign to each permutation σ. Here is how we do this:

Definition 2.2.2. Let σ be a permutation of $[n]$.

(a) An **inversion** of σ means a pair (i, j) of integers such that $1 \leq i < j \leq n$ but $\sigma(i) > \sigma(j)$. (This is called a **reversal** of σ in [Strickland, Appendix B].)

(b) We define $\ell(\sigma)$ as the number of inversions of σ. (This is called the **Coxeter length** of σ. It is denoted by $l(\sigma)$ in [Strickland, Appendix B].)

(c) We define the **sign** $\operatorname{sign}\sigma$ of σ to be $(-1)^{\ell(\sigma)}$. Thus,

- we have $\operatorname{sign}\sigma = 1$ if σ has an even number of inversions;
- we have $\operatorname{sign}\sigma = -1$ if σ has an odd number of inversions.

(Note that $\operatorname{sign}\sigma$ is called $\operatorname{sgn}(\sigma)$ in [Strickland, Appendix B].)

Example 2.2.3. Let us compute the inversions of the 6 permutations of [3]:

permutation σ	inversions	$\ell(\sigma)$	sign $\operatorname{sign}\sigma$
$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$	none	0	$(-1)^0 = 1$
$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$	$(2,3)$	1	$(-1)^1 = -1$
$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$	$(1,2)$	1	$(-1)^1 = -1$
$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$	$(2,3)$ and $(1,3)$	2	$(-1)^2 = 1$
$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$	$(1,2)$ and $(1,3)$	2	$(-1)^2 = 1$
$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$	$(1,2)$ and $(1,3)$ and $(2,3)$	3	$(-1)^3 = -1$

Now, we can try to define $\det A$ as a sum of products of generalized diagonals, one for each permutation σ of $[n]$. These products get multiplied by $\operatorname{sign}\sigma$, and then get all added together:

Definition 2.2.4. Let A be an $n \times n$-matrix. Then, the **determinant** $\det A$ of A is defined to be the sum

$$\sum_{\sigma \text{ is a permutation of } [n]} \operatorname{sign}\sigma \cdot A_{1,\sigma(1)} A_{2,\sigma(2)} \cdots A_{n,\sigma(n)}.$$

Here, the $\sum$ sign is the **summation sign**; thus, the determinant is defined as a big sum, with one addend for each permutation of $[n]$.

Example 2.2.5. If $n = 3$, then this becomes

$$\begin{aligned}
\det A &= 1 A_{1,1} A_{2,2} A_{3,3} + (-1) A_{1,1} A_{2,3} A_{3,2} + (-1) A_{1,2} A_{2,1} A_{3,3} \\
&\qquad + 1 A_{1,2} A_{2,3} A_{3,1} + 1 A_{1,3} A_{2,1} A_{3,2} + (-1) A_{1,3} A_{2,2} A_{3,1} \\
&= A_{1,1} A_{2,2} A_{3,3} + A_{1,2} A_{2,3} A_{3,1} + A_{1,3} A_{2,1} A_{3,2} \\
&\qquad - A_{1,1} A_{2,3} A_{3,2} - A_{1,2} A_{2,1} A_{3,3} - A_{1,3} A_{2,2} A_{3,1}.
\end{aligned}$$

This is exactly the formula $D = ab'c'' + bc'a'' + ca'b'' - ac'b'' - ba'c'' - cb'a''$ when $A = \begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}$.

Example 2.2.6. If $n = 2$, then the definition of $\det A$ becomes

$$\det A = 1A_{1,1}A_{2,2} + (-1)\, A_{1,2}A_{2,1} = A_{1,1}A_{2,2} - A_{1,2}A_{2,1}.$$

This is exactly $ab' - ba'$ if $A = \begin{pmatrix} a & b \\ a' & b' \end{pmatrix}$.

References

[Day16] Martin V. Day, *An Introduction to Proofs and the Mathematical Vernacular*, 7 December 2016.
`https://web.archive.org/web/20180712152432/https://www.math.vt.edu/people/day/ProofsBook/IPaMV.pdf` .

[Hammac15] Richard Hammack, *Book of Proof*, 3rd edition.
`https://www.people.vcu.edu/~rhammack/BookOfProof/`

[Lay] David C. Lay, Steven R. Lay, Judi J. McDonald, *Linear Algebra and its Applications*, 5th edition.

[LeLeMe16] Eric Lehman, F. Thomson Leighton, Albert R. Meyer, *Mathematics for Computer Science*, revised Tuesday 6th June 2018,
`https://courses.csail.mit.edu/6.042/spring18/mcs.pdf` .

[lina] Darij Grinberg, *Notes on linear algebra*, version of 13 December 2016.
`https://github.com/darijgr/lina`

[Strickland] Neil Strickland, *MAS201 Linear Mathematics for Applications*, lecture notes, 28 September 2013.
`http://neil-strickland.staff.shef.ac.uk/courses/MAS201/`