

8.1 Partitions and Block Multiplication

A large matrix often has structure that its individual entries hide: a corner of zeros, a copy of the identity, a smaller matrix repeated. Chapter 7 already met one such shape, the block triangular matrix, and computed its determinant from two smaller determinants. This section makes the idea systematic. We cut matrices into rectangular pieces, prove that products can be computed piece by piece, and use the pieces to read off structure: direct sums become block diagonal matrices, and a subspace that an operator maps into itself becomes a zero block in the corner.

8.1.1 Cutting a matrix into blocks

Take two 4×4 matrices over $\mathbb{Q}$ whose top-right 2×2 corners are zero:

$$\mathbf{M} = \left(\begin{array}{cc|cc} 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 2 & 1 \\ 2 & 1 & 1 & 1 \end{array} \right), \quad \mathbf{N} = \left(\begin{array}{cc|cc} 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 1 & 0 \\ 1 & 1 & 3 & 1 \end{array} \right).$$

The lines cut each matrix into four 2×2 pieces. Name them: $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$ and $\mathbf{N} = \begin{pmatrix} \mathbf{E} & \mathbf{O} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}$,

with $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, $\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, and so on. If we pretend that the pieces are numbers and multiply as for 2×2 matrices, keeping the order of every product, we get

$$\mathbf{MN} \stackrel{?}{=} \begin{pmatrix} \mathbf{AE} + \mathbf{OG} & \mathbf{AO} + \mathbf{OH} \\ \mathbf{CE} + \mathbf{DG} & \mathbf{CO} + \mathbf{DH} \end{pmatrix} = \begin{pmatrix} \mathbf{AE} & \mathbf{O} \\ \mathbf{CE} + \mathbf{DG} & \mathbf{DH} \end{pmatrix}.$$

With $\mathbf{AE} = \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix}$, $\mathbf{CE} + \mathbf{DG} = \begin{pmatrix} 2 & 2 \\ 4 & 0 \end{pmatrix}$, and $\mathbf{DH} = \begin{pmatrix} 5 & 1 \\ 4 & 1 \end{pmatrix}$, this predicts

$$\mathbf{MN} = \begin{pmatrix} 3 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 2 & 2 & 5 & 1 \\ 4 & 0 & 4 & 1 \end{pmatrix},$$

and multiplying out the sixteen entries directly gives exactly this matrix. Two things happened. The zero corner of the product came for free, with no arithmetic. And the product of two 4×4 matrices became four products of 2×2 matrices, $\mathbf{AE}$, $\mathbf{CE}$, $\mathbf{DG}$, and $\mathbf{DH}$; the other four products in the block formula have a zero factor and cost nothing. To use this safely we need to say precisely what “cutting” means and prove that the pretend multiplication is always correct.

Draw horizontal lines between rows and vertical lines between columns; the pieces are the blocks.

Definition 8.1.1. Block Partition

Let $\mathbf{A} = (a_{ij}) \in M_{m \times n}(F)$. A **row partition** of $\mathbf{A}$ is a list of **positive** integers $m_1, \dots, m_a$ with $m_1 + \dots + m_a = m$, and a **column partition** is a list of positive integers $n_1, \dots, n_b$ with $n_1 + \dots + n_b = n$. Put $\mu_r = m_1 + \dots + m_r$ and $\nu_s = n_1 + \dots + n_s$, with $\mu_0 = \nu_0 = 0$. For $1 \leq r \leq a$ and $1 \leq s \leq b$, the (r, s) **block** of $\mathbf{A}$ is the matrix $\mathbf{A}_{rs} \in M_{m_r \times n_s}(F)$ with entries

$$(\mathbf{A}_{rs})_{ij} = a_{\mu_{r-1}+i, \nu_{s-1}+j} \quad (1 \leq i \leq m_r, 1 \leq j \leq n_s).$$

We then write $\mathbf{A} = (\mathbf{A}_{rs})$, displayed as an $a \times b$ array of blocks, and call $\mathbf{A}$ a **block matrix**.

Let $\mathbf{B} \in M_{n \times p}(F)$. Partitions of $\mathbf{A}$ and $\mathbf{B}$ are **conformable** if the column partition of $\mathbf{A}$ and the row partition of $\mathbf{B}$ are **the same list** $n_1, \dots, n_b$, in the same order.

In words: the row partition says where the horizontal lines go. Block row r consists of the rows with indices in $I_r = \{\mu_{r-1} + 1, \dots, \mu_r\}$, a run of m_r consecutive rows, and similarly block column s consists of the columns in $J_s = \{\nu_{s-1} + 1, \dots, \nu_s\}$. The block $\mathbf{A}_{rs}$ is what survives when we keep only the rows in I_r and the columns in J_s , renumbered from 1. Conformability says that the vertical lines through $\mathbf{A}$ sit at the same places as the horizontal lines through $\mathbf{B}$. That is exactly what will make each product “block of $\mathbf{A}$ times block of $\mathbf{B}$ ” defined.

Well-definedness. The runs $I_1, \dots, I_a$ are disjoint and cover $\{1, \dots, m\}$, and the same holds for $J_1, \dots, J_b$. So every entry of $\mathbf{A}$ lies in exactly one block. Conversely, if we are given matrices $\mathbf{A}_{rs} \in M_{m_r \times n_s}(F)$, there is exactly one $\mathbf{A} \in M_{m \times n}(F)$ with these blocks. Block sizes must fit: all blocks in block row r have m_r rows, and all blocks in block column s have n_s columns. We allow only positive parts, so no block is empty.

Examples.

- **Every entry a block.** With $m_r = 1$ and $n_s = 1$ for all r, s , the block $\mathbf{A}_{rs}$ is the 1×1 matrix (a_{rs}) . A matrix is its own finest block matrix.
- **One block.** With $a = b = 1$, the single block is $\mathbf{A}$ itself. This degenerate partition matters because a statement about block matrices then says something about ordinary matrices, and the statement must not break.
- **Columns as blocks.** Row partition m , column partition $1, 1, \dots, 1$. The blocks are the n columns of $\mathbf{A}$: $\mathbf{A} = (\mathbf{a}_1 \ \cdots \ \mathbf{a}_n)$, as in Chapter 0.
- **The hook.** For $\mathbf{M}$ and $\mathbf{N}$ above, both partitions are $2, 2$, so the column partition of $\mathbf{M}$ equals the row partition of $\mathbf{N}$: they are conformable.

Nonexample by minimal change. Keep $\mathbf{M}$ with column partition $2, 2$, but cut $\mathbf{N}$ with row partition $1, 3$. The product $\mathbf{MN}$ is still defined, and both partitions are legitimate. What fails is conformability: the first block column of $\mathbf{M}$ has 2 columns, while the first block row of $\mathbf{N}$ has 1 row, so the “product” $\mathbf{A} \cdot \mathbf{N}_{11}$ of a 2×2 matrix with a 1×2 matrix is undefined.

Quick check. Let $\mathbf{A} \in M_{3 \times 5}(F)$ have row partition $1, 2$ and column partition $3, 2$, and let $\mathbf{B} \in M_{5 \times 4}(F)$ have row partition $2, 3$. Are the partitions conformable? What are the sizes of the blocks $\mathbf{A}_{21}$ and $\mathbf{B}_{2t}$?

Solution. No. The column partition of $\mathbf{A}$ is $3, 2$ and the row partition of $\mathbf{B}$ is $2, 3$. Both lists sum to 5, but they differ in order, and conformability asks for the same list. The block $\mathbf{A}_{21}$ is 2×3 (block row 2 has 2 rows, block column 1 has 3 columns), and every $\mathbf{B}_{2t}$ has 3 rows.

8.1.2 Block multiplication

Here is the theorem that justifies the computation in the hook. The proof is nothing but the entry formula for a product, with the sum over the middle index cut into the runs $J_1, \dots, J_b$.

Theorem 8.1.1. Block Multiplication

Let $\mathbf{A} \in M_{m \times n}(F)$ and $\mathbf{B} \in M_{n \times p}(F)$ be conformably partitioned: $\mathbf{A}$ with row partition $m_1, \dots, m_a$ and column partition $n_1, \dots, n_b$, and $\mathbf{B}$ with row partition $n_1, \dots, n_b$ and column partition $p_1, \dots, p_c$. Partition $\mathbf{AB} \in M_{m \times p}(F)$ with row partition $m_1, \dots, m_a$ and column partition $p_1, \dots, p_c$. Then for all $1 \leq r \leq a$ and $1 \leq t \leq c$,

$$(\mathbf{AB})_{rt} = \sum_{s=1}^b \mathbf{A}_{rs} \mathbf{B}_{st} = \mathbf{A}_{r1} \mathbf{B}_{1t} + \mathbf{A}_{r2} \mathbf{B}_{2t} + \cdots + \mathbf{A}_{rb} \mathbf{B}_{bt}.$$

Each product $\mathbf{A}_{rs} \mathbf{B}_{st}$ is defined and lies in $M_{m_r \times p_t}(F)$.

Idea. Look at one entry of $\mathbf{AB}$, in row i^* and column k^* . It is a sum over the middle index $j = 1, \dots, n$. Cut that sum into the runs $J_1, \dots, J_b$ given by the common partition. The part over J_s pairs row i^* of $\mathbf{A}$, restricted to the columns of block column s , with column k^* of $\mathbf{B}$, restricted to the rows of block row s . That is one entry of $\mathbf{A}_{rs} \mathbf{B}_{st}$. The only real work is translating indices between the big matrix and the blocks.

Proof. Let μ_r, ν_s, ρ_t be the partial sums of the three partitions, as in Definition 8.1.1, and $J_s = \{\nu_{s-1} + 1, \dots, \nu_s\}$. The block $\mathbf{A}_{rs}$ is $m_r \times n_s$ and $\mathbf{B}_{st}$ is $n_s \times p_t$, so their product is defined and is $m_r \times p_t$, as is the block $(\mathbf{AB})_{rt}$.

Fix r, t , and let $1 \leq i \leq m_r$ and $1 \leq k \leq p_t$. Put $i^* = \mu_{r-1} + i$ and $k^* = \rho_{t-1} + k$. By Definition 8.1.1 and Definition 0.9.9,

$$((\mathbf{AB})_{rt})_{ik} = (\mathbf{AB})_{i^*k^*} = \sum_{j=1}^n a_{i^*j} b_{jk^*}.$$

The sets $J_1, \dots, J_b$ are disjoint with union $\{1, \dots, n\}$, so the finite sum may be grouped according to them. In the group for J_s , substitute $j = \nu_{s-1} + l$; as l runs through $1, \dots, n_s$, j runs through J_s exactly once. Hence

$$\begin{aligned} ((\mathbf{AB})_{rt})_{ik} &= \sum_{s=1}^b \sum_{j \in J_s} a_{i^*j} b_{jk^*} \\ &= \sum_{s=1}^b \sum_{l=1}^{n_s} a_{\mu_{r-1}+i, \nu_{s-1}+l} b_{\nu_{s-1}+l, \rho_{t-1}+k} \\ &= \sum_{s=1}^b \sum_{l=1}^{n_s} (\mathbf{A}_{rs})_{il} (\mathbf{B}_{st})_{lk}, \end{aligned}$$

where the last step is the definition of the blocks $\mathbf{A}_{rs}$ and $\mathbf{B}_{st}$. By Definition 0.9.9 the inner sum is $(\mathbf{A}_{rs} \mathbf{B}_{st})_{ik}$, and by Definition 0.9.6 the outer sum is the (i, k) -entry of $\sum_s \mathbf{A}_{rs} \mathbf{B}_{st}$. Since i and k were arbitrary, $(\mathbf{AB})_{rt} = \sum_s \mathbf{A}_{rs} \mathbf{B}_{st}$. This proves the theorem. ■

So conformably partitioned matrices multiply like matrices whose entries are matrices. The

commonest case is a 2×2 block matrix:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{A}' & \mathbf{B}' \\ \mathbf{C}' & \mathbf{D}' \end{pmatrix} = \begin{pmatrix} \mathbf{AA}' + \mathbf{BC}' & \mathbf{AB}' + \mathbf{BD}' \\ \mathbf{CA}' + \mathbf{DC}' & \mathbf{CB}' + \mathbf{DD}' \end{pmatrix},$$

whenever the partitions are conformable. Sums and scalar multiples of matrices with the same partition are also computed blockwise, straight from Definition 0.9.6 and Definition 0.9.7, since each entry lies in exactly one block.

Over a commutative ring. The proof uses only sums and products of entries and the re-grouping of a finite sum. So block multiplication holds verbatim for matrices with entries in any commutative ring R in the sense of Chapter 7 §4, for instance $R = F[x]$. Section 3 uses this.

Warning. Keep the order inside every block product. The top-left block of the product above is $\mathbf{AA}' + \mathbf{BC}'$, not $\mathbf{A}'\mathbf{A} + \mathbf{C}'\mathbf{B}$: blocks are matrices, and matrices do not commute. In the hook, $\mathbf{CE} + \mathbf{DG} = \begin{pmatrix} 2 & 2 \\ 4 & 0 \end{pmatrix}$, while $\mathbf{EC} + \mathbf{GD} = \begin{pmatrix} 0 & 0 \\ 4 & 2 \end{pmatrix}$. For the same reason, the formula $ad - bc$ for a 2×2 determinant does **not** survive with blocks: in general $\det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \neq \det(\mathbf{AD} - \mathbf{BC})$. Take $\mathbf{A} = \mathbf{I}_2$, $\mathbf{B} = \mathbf{E}_{12}$, $\mathbf{C} = \mathbf{E}_{21}$, $\mathbf{D} = \mathbf{E}_{11}$ in $M_2(\mathbb{Q})$. Then $\mathbf{AD} - \mathbf{BC} = \mathbf{E}_{11} - \mathbf{E}_{12}\mathbf{E}_{21} = \mathbf{E}_{11} - \mathbf{E}_{11} = \mathbf{O}$, with determinant 0. But

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

and subtracting row 4 from row 1, which does not change the determinant (Theorem 7.5.1), gives the permutation matrix that swaps $\mathbf{e}_1$ and $\mathbf{e}_4$, whose determinant is -1 . So $\det \mathbf{M} = -1 \neq 0$. Section 3 shows that the formula does hold when $\mathbf{C}$ and $\mathbf{D}$ commute; here $\mathbf{CD} = \mathbf{E}_{21} \neq \mathbf{O} = \mathbf{DC}$.

Transposing a block matrix transposes each block and reflects the array of blocks.

Proposition 8.1.2. Transpose of a Block Matrix

Let $\mathbf{A} \in M_{m \times n}(F)$ have row partition $m_1, \dots, m_a$, column partition $n_1, \dots, n_b$, and blocks $\mathbf{A}_{rs}$. Give $\mathbf{A}^\top$ the row partition $n_1, \dots, n_b$ and the column partition $m_1, \dots, m_a$. Then $(\mathbf{A}^\top)_{sr} = (\mathbf{A}_{rs})^\top$. In particular $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^\top = \begin{pmatrix} \mathbf{A}^\top & \mathbf{C}^\top \\ \mathbf{B}^\top & \mathbf{D}^\top \end{pmatrix}$.

Proof. For $1 \leq j \leq n_s$ and $1 \leq i \leq m_r$, by Definition 8.1.1 and Definition 0.9.10, $((\mathbf{A}^\top)_{sr})_{ji} = (\mathbf{A}^\top)_{\nu_{s-1}+j, \mu_{r-1}+i} = a_{\mu_{r-1}+i, \nu_{s-1}+j} = (\mathbf{A}_{rs})_{ij} = ((\mathbf{A}_{rs})^\top)_{ji}$. ■

8.1.3 Rows, columns, and outer products

The finest and coarsest partitions turn [Theorem 8.1.1](#) into facts we already know, and one mixed choice gives something new.

Partition $\mathbf{A} \in M_{m \times n}(F)$ trivially (one block) and $\mathbf{B} \in M_{n \times p}(F)$ into columns (row partition n , column partition $1, \dots, 1$). These are conformable, and the theorem says that the k -th block of $\mathbf{AB}$, its k -th column, is $\mathbf{A}\mathbf{b}_k$. That is the column view of Theorem 0.9.2. Partitioning $\mathbf{A}$ into rows and $\mathbf{B}$ trivially gives the row view in the same way.

Now cut $\mathbf{A}$ into columns and $\mathbf{B}$ into rows. Both middle partitions are $1, 1, \dots, 1$ (with n ones), so they are conformable, and the product has a single block.

Corollary 8.1.3. Outer Product Expansion

Let $\mathbf{A} \in M_{m \times n}(F)$ have columns $\mathbf{a}_1, \dots, \mathbf{a}_n \in F^m$, and let $\mathbf{B} \in M_{n \times p}(F)$ have rows $\mathbf{r}_1, \dots, \mathbf{r}_n \in M_{1 \times p}(F)$. Then

$$\mathbf{AB} = \mathbf{a}_1\mathbf{r}_1 + \mathbf{a}_2\mathbf{r}_2 + \cdots + \mathbf{a}_n\mathbf{r}_n = \sum_{k=1}^n \mathbf{a}_k\mathbf{r}_k,$$

where each $\mathbf{a}_k\mathbf{r}_k \in M_{m \times p}(F)$ is column k of $\mathbf{A}$ times row k of $\mathbf{B}$.

Proof. Give $\mathbf{A}$ the row partition m and the column partition $1, \dots, 1$, so that its blocks are $\mathbf{A}_{1k} = \mathbf{a}_k$; give $\mathbf{B}$ the row partition $1, \dots, 1$ and the column partition p , so that $\mathbf{B}_{k1} = \mathbf{r}_k$. The partitions are conformable, and by [Theorem 8.1.1](#) the single block of $\mathbf{AB}$ is $\sum_k \mathbf{A}_{1k}\mathbf{B}_{k1} = \sum_k \mathbf{a}_k\mathbf{r}_k$. ■

Each term $\mathbf{a}_k\mathbf{r}_k$ is a column times a row, an **outer product**, and it has rank at most 1: its columns are the multiples $r_{kj}\mathbf{a}_k$ of the single vector $\mathbf{a}_k$. So the corollary writes every product of an $m \times n$ and an $n \times p$ matrix as a sum of n matrices of rank at most one, in the spirit of Exercise (B3: Rank-one pieces), which writes a matrix of rank r as a sum of r rank-one matrices. For example,

$$\begin{aligned} & \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ -1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 2 \\ 3 & 0 & 6 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} -2 & 2 & 2 \\ -4 & 4 & 4 \\ -1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 2 & 4 \\ -1 & 4 & 10 \\ -1 & 1 & 1 \end{pmatrix}, \end{aligned}$$

which is the product computed entry by entry.

8.1.4 Block diagonal matrices

Diagonal matrices are the easiest matrices to compute with: they multiply, power, and invert entry by entry along the diagonal. Replace the diagonal entries by square blocks and all of this survives, one block at a time.

A block diagonal matrix is several square matrices placed corner to corner along the diagonal, with zeros everywhere else.

Definition 8.1.2. Block Diagonal Matrix

Let $k_1, \dots, k_r$ be positive integers and $\mathbf{A}_i \in M_{k_i}(F)$ **square** for each i . Their **direct sum** $\mathbf{A}_1 \oplus \dots \oplus \mathbf{A}_r$ is the matrix in $M_{k_1 + \dots + k_r}(F)$ which, with row partition and column partition both equal to $k_1, \dots, k_r$, has (i, i) block $\mathbf{A}_i$ and (i, j) block $\mathbf{O}$ for $i \neq j$. A square matrix is **block diagonal with respect to** the partition $k_1, \dots, k_r$ if it has this form. For two blocks,

$$\mathbf{A} \oplus \mathbf{B} = \begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{B} \end{pmatrix}.$$

In words: the same partition is used for rows and columns, so the diagonal blocks are square; the blocks on the diagonal may be anything; every block off the diagonal is zero. Whether a matrix is block diagonal depends on the partition, which is part of the statement.

Examples.

- **Diagonal matrices.** $\text{diag}(d_1, \dots, d_n) = (d_1) \oplus \dots \oplus (d_n)$, with all $k_i = 1$.
- **The identity.** $\mathbf{I}_k \oplus \mathbf{I}_l = \mathbf{I}_{k+l}$, and more generally $\mathbf{I}_k \oplus \mathbf{O}_l$ is the matrix of a projection (Corollary 4.5.2).
- **A genuine block.** $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \oplus (2) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, with partition 2, 1.
- **One block.** With $r = 1$, $\mathbf{A}_1$ itself. Every square matrix is block diagonal with respect to the trivial partition, which is why the partition must always be named.

Nonexample by minimal change. Change the $(1, 3)$ -entry of the third example from 0 to 1.

The partition 2, 1 still has square diagonal blocks $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and (2) , and the $(2, 1)$ block $\begin{pmatrix} 0 & 0 \end{pmatrix}$

is still zero. What fails is the clause that the $(1, 2)$ block is zero: it is now $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The matrix is block upper triangular, which is treated below, but not block diagonal for this partition.

Why the symbol $\oplus$. The matrix $\mathbf{A} \oplus \mathbf{B}$ acts on $F^{k+l} = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_k) \oplus \text{span}(\mathbf{e}_{k+1}, \dots, \mathbf{e}_{k+l})$ by sending each of the two summands into itself, acting as $\mathbf{A}$ on the first and as $\mathbf{B}$ on the second. The theorem on adapted bases below makes this precise for any operator and any direct sum, and it is the reason the same symbol is used.

Warning. $\mathbf{A} \oplus \mathbf{B}$ is a matrix, not a subspace, and the order matters. For subspaces, $U \oplus W$ and $W \oplus U$ are the same set. For matrices, $(1) \oplus (2) = \text{diag}(1, 2)$ and $(2) \oplus (1) = \text{diag}(2, 1)$ are different matrices. They are similar (swap the two basis vectors), but they are not equal.

Theorem 8.1.4. Arithmetic of Block Diagonal Matrices

Let $\mathbf{A}, \mathbf{A}' \in M_k(F)$, $\mathbf{B}, \mathbf{B}' \in M_l(F)$, and $c \in F$. Then:

- (a) $(\mathbf{A} \oplus \mathbf{B}) + (\mathbf{A}' \oplus \mathbf{B}') = (\mathbf{A} + \mathbf{A}') \oplus (\mathbf{B} + \mathbf{B}')$, $c(\mathbf{A} \oplus \mathbf{B}) = c\mathbf{A} \oplus c\mathbf{B}$, and $(\mathbf{A} \oplus \mathbf{B})(\mathbf{A}' \oplus \mathbf{B}') = \mathbf{A}\mathbf{A}' \oplus \mathbf{B}\mathbf{B}'$;
- (b) $(\mathbf{A} \oplus \mathbf{B})^j = \mathbf{A}^j \oplus \mathbf{B}^j$ for every $j \in \mathbb{N}$, and $p(\mathbf{A} \oplus \mathbf{B}) = p(\mathbf{A}) \oplus p(\mathbf{B})$ for every $p \in F[x]$;
- (c) $\mathbf{A} \oplus \mathbf{B}$ is invertible if and only if $\mathbf{A}$ and $\mathbf{B}$ are both invertible, and then $(\mathbf{A} \oplus \mathbf{B})^{-1} = \mathbf{A}^{-1} \oplus \mathbf{B}^{-1}$;
- (d) $\det(\mathbf{A} \oplus \mathbf{B}) = \det \mathbf{A} \det \mathbf{B}$ and $\text{tr}(\mathbf{A} \oplus \mathbf{B}) = \text{tr} \mathbf{A} + \text{tr} \mathbf{B}$.

Proof.

- (a) Sums and scalar multiples are computed blockwise, and by Theorem 8.1.1 the product has blocks $\mathbf{A}\mathbf{A}' + \mathbf{O} \cdot \mathbf{O} = \mathbf{A}\mathbf{A}'$, $\mathbf{A}\mathbf{O} + \mathbf{O}\mathbf{B}' = \mathbf{O}$, $\mathbf{O}\mathbf{A}' + \mathbf{B}\mathbf{O} = \mathbf{O}$, and $\mathbf{O} \cdot \mathbf{O} + \mathbf{B}\mathbf{B}' = \mathbf{B}\mathbf{B}'$.
- (b) By induction on j . For $j = 0$, $(\mathbf{A} \oplus \mathbf{B})^0 = \mathbf{I}_{k+l} = \mathbf{I}_k \oplus \mathbf{I}_l$ by Definition 0.9.14. If $(\mathbf{A} \oplus \mathbf{B})^j = \mathbf{A}^j \oplus \mathbf{B}^j$, then $(\mathbf{A} \oplus \mathbf{B})^{j+1} = (\mathbf{A}^j \oplus \mathbf{B}^j)(\mathbf{A} \oplus \mathbf{B}) = \mathbf{A}^{j+1} \oplus \mathbf{B}^{j+1}$ by (a). For $p = \sum_j c_j x^j$, (a) gives $p(\mathbf{A} \oplus \mathbf{B}) = \sum_j c_j (\mathbf{A}^j \oplus \mathbf{B}^j) = (\sum_j c_j \mathbf{A}^j) \oplus (\sum_j c_j \mathbf{B}^j) = p(\mathbf{A}) \oplus p(\mathbf{B})$.
- (c) The matrix $\mathbf{A} \oplus \mathbf{B}$ is block upper triangular with square diagonal blocks, so $\det(\mathbf{A} \oplus \mathbf{B}) = \det \mathbf{A} \det \mathbf{B}$ by Theorem 7.8.4. Its diagonal entries are those of $\mathbf{A}$ followed by those of $\mathbf{B}$, so the traces add (Definition 0.9.12).
- (d) $(\Leftarrow)$ By (a), $(\mathbf{A} \oplus \mathbf{B})(\mathbf{A}^{-1} \oplus \mathbf{B}^{-1}) = \mathbf{I}_k \oplus \mathbf{I}_l = \mathbf{I}_{k+l}$, so by Theorem 3.4.5 $\mathbf{A} \oplus \mathbf{B}$ is invertible with inverse $\mathbf{A}^{-1} \oplus \mathbf{B}^{-1}$. $(\Rightarrow)$ If $\mathbf{A} \oplus \mathbf{B}$ is invertible, then $\det \mathbf{A} \det \mathbf{B} \neq 0$ by (d) and Theorem 7.5.2. A product in a field is nonzero only if both factors are (Theorem 0.7.1 (d)), so $\det \mathbf{A} \neq 0 \neq \det \mathbf{B}$, and both are invertible by Theorem 7.5.2. ■

By induction on the number of blocks, the same rules hold for $\mathbf{A}_1 \oplus \cdots \oplus \mathbf{A}_r$: everything is done block by block. For instance $(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \oplus (2))^5 = \begin{pmatrix} 1 & 5 \\ 0 & 1 \end{pmatrix} \oplus (32)$, since $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^j = \begin{pmatrix} 1 & j \\ 0 & 1 \end{pmatrix}$ by induction.

8.1.5 Block triangular matrices

Allow one more nonzero block: $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{pmatrix}$ with $\mathbf{A} \in M_k(F)$, $\mathbf{D} \in M_l(F)$ square, and $\mathbf{B} \in M_{k \times l}(F)$. Such a matrix is **block upper triangular**; with the zero block above the diagonal instead, **block lower triangular**. Chapter 7 computed $\det \mathbf{M} = \det \mathbf{A} \det \mathbf{D}$ (Theorem 7.8.4). Products and inverses keep the shape.

Theorem 8.1.5. Block Triangular Products and Inverses

Let $k, l \geq 1$, and let $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{pmatrix}$ and $\mathbf{M}' = \begin{pmatrix} \mathbf{A}' & \mathbf{B}' \\ \mathbf{O} & \mathbf{D}' \end{pmatrix}$, with $\mathbf{A}, \mathbf{A}' \in M_k(F)$, $\mathbf{D}, \mathbf{D}' \in M_l(F)$, and $\mathbf{B}, \mathbf{B}' \in M_{k \times l}(F)$.

- (a) $\mathbf{M}\mathbf{M}' = \begin{pmatrix} \mathbf{A}\mathbf{A}' & \mathbf{A}\mathbf{B}' + \mathbf{B}\mathbf{D}' \\ \mathbf{O} & \mathbf{D}\mathbf{D}' \end{pmatrix}$; in particular it is block upper triangular.
- (b) $\mathbf{M}$ is invertible if and only if $\mathbf{A}$ and $\mathbf{D}$ are both invertible.
- (c) In that case

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{O} & \mathbf{D}^{-1} \end{pmatrix}.$$

Idea. Where does the formula come from? Solve $\mathbf{M} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix}$ by back substitution with blocks. The second block row says $\mathbf{D}\mathbf{y} = \mathbf{g}$, so $\mathbf{y} = \mathbf{D}^{-1}\mathbf{g}$. The first says $\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \mathbf{f}$, so $\mathbf{x} = \mathbf{A}^{-1}\mathbf{f} - \mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1}\mathbf{g}$. Reading off the coefficients of $\mathbf{f}$ and $\mathbf{g}$ gives the four blocks of $\mathbf{M}^{-1}$. The proof only has to check the answer.

Proof.

- (a) By [Theorem 8.1.1](#), the blocks of $\mathbf{M}\mathbf{M}'$ are $\mathbf{A}\mathbf{A}' + \mathbf{B}\mathbf{O} = \mathbf{A}\mathbf{A}'$, $\mathbf{A}\mathbf{B}' + \mathbf{B}\mathbf{D}'$, $\mathbf{O}\mathbf{A}' + \mathbf{D}\mathbf{O} = \mathbf{O}$, and $\mathbf{O}\mathbf{B}' + \mathbf{D}\mathbf{D}' = \mathbf{D}\mathbf{D}'$.
- (b) By [Theorem 7.8.4](#), $\det \mathbf{M} = \det \mathbf{A} \det \mathbf{D}$. By [Theorem 7.5.2](#), $\mathbf{M}$ is invertible if and only if $\det \mathbf{A} \det \mathbf{D} \neq 0$, which in a field happens exactly when $\det \mathbf{A} \neq 0$ and $\det \mathbf{D} \neq 0$ ([Theorem 0.7.1](#) (f) and (d)), that is, when $\mathbf{A}$ and $\mathbf{D}$ are both invertible.
- (c) Let $\mathbf{X}$ be the displayed matrix. By (a) with $\mathbf{A}' = \mathbf{A}^{-1}$, $\mathbf{B}' = -\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1}$, and $\mathbf{D}' = \mathbf{D}^{-1}$,

$$\mathbf{M}\mathbf{X} = \begin{pmatrix} \mathbf{A}\mathbf{A}^{-1} & -\mathbf{A}\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} + \mathbf{B}\mathbf{D}^{-1} \\ \mathbf{O} & \mathbf{D}\mathbf{D}^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_k & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_l \end{pmatrix} = \mathbf{I}_{k+l}.$$

By [Theorem 3.4.5](#), $\mathbf{X} = \mathbf{M}^{-1}$. ■

Transposing with [Proposition 8.1.2](#) turns this into the lower triangular version: if $\mathbf{A}$ and $\mathbf{D}$ are invertible, then $\begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{O} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{D}^{-1} \end{pmatrix}$. A special case is used constantly: $\begin{pmatrix} \mathbf{I}_k & \mathbf{X} \\ \mathbf{O} & \mathbf{I}_l \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I}_k & -\mathbf{X} \\ \mathbf{O} & \mathbf{I}_l \end{pmatrix}$ for every $\mathbf{X} \in M_{k \times l}(F)$. Adding $\mathbf{X}$ times the second block of coordinates to the first is undone by subtracting it.

Quick check. The matrix $\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, with row partition 2, 1 and column partition 1, 2, has a zero lower-left block and is invertible. Does this contradict [Theorem 8.1.5](#)?

Solution. No. The theorem needs the diagonal blocks square, which means equal row and

column partitions. Here the top-left block $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is 2×1 and the bottom-right block $\begin{pmatrix} 1 & 0 \end{pmatrix}$ is 1×2 , so “**A** is invertible” does not even make sense. (The lower-left block is the 1×1 entry in position $(3, 1)$, which is 0.)

8.1.6 Matrices of operators in adapted bases

Now the shapes pay off. For an operator $T \in \mathcal{L}(V)$ and a subspace $U \subseteq V$, we say that **T maps U into U** if $Tu \in U$ for every $u \in U$. Then the **restriction** $T|_U: U \rightarrow U$, $u \mapsto Tu$, is an operator on U , linear because its values are those of T . Such subspaces are the subject of Chapter 9; here we only ask what they do to the matrix of T .

When $V = U \oplus W$ and a basis of U is followed by a basis of W , the combined list is a basis of V (Theorem 1.7.5), which we call **adapted** to the decomposition. If T maps both pieces into themselves, the matrix splits.

Theorem 8.1.6. Direct Sums Give Block Diagonal Matrices

Let $V = U \oplus W$ be finite-dimensional with $U \neq \{0\} \neq W$, let $\mathcal{B}_U = (u_1, \dots, u_k)$ and $\mathcal{B}_W = (w_1, \dots, w_l)$ be bases of U and W , and let $\mathcal{B} = (u_1, \dots, u_k, w_1, \dots, w_l)$. For $T \in \mathcal{L}(V)$, the following are equivalent:

- (a) T maps U into U and W into W ;
- (b) $[T]_{\mathcal{B}}^{\mathcal{B}}$ is block diagonal with respect to the partition k, l .

In that case $[T]_{\mathcal{B}}^{\mathcal{B}} = [T|_U]_{\mathcal{B}_U}^{\mathcal{B}_U} \oplus [T|_W]_{\mathcal{B}_W}^{\mathcal{B}_W}$.

Proof. By Theorem 1.7.5, $\mathcal{B}$ is a basis of V . By Definition 4.1.1, column $j \leq k$ of $[T]_{\mathcal{B}}^{\mathcal{B}}$ holds the coordinates of Tu_j , and column $k + q$ those of Tw_q .

- (a) $\Rightarrow$ (b). Since $Tu_j \in U = \text{span}(\mathcal{B}_U)$, write $Tu_j = \sum_i a_{ij}u_i$. By uniqueness of coordinates (Theorem 1.5.1), the last l entries of column j are 0 and the first k are $a_{1j}, \dots, a_{kj}$, which is column j of $[T|_U]_{\mathcal{B}_U}^{\mathcal{B}_U}$. In the same way $Tw_q = \sum_p d_{pq}w_p$, so column $k + q$ is 0 in its first k entries and equals column q of $[T|_W]_{\mathcal{B}_W}^{\mathcal{B}_W}$ in its last l . This shows $[T]_{\mathcal{B}}^{\mathcal{B}} = [T|_U]_{\mathcal{B}_U}^{\mathcal{B}_U} \oplus [T|_W]_{\mathcal{B}_W}^{\mathcal{B}_W}$.
- (b) $\Rightarrow$ (a). If the $(2, 1)$ block is zero, then the coordinates of each Tu_j vanish in the last l places, so $Tu_j \in \text{span}(\mathcal{B}_U) = U$. For $u = \sum_j c_j u_j \in U$, linearity gives $Tu = \sum_j c_j Tu_j \in U$, since U is a subspace. The $(1, 2)$ block being zero gives $Tw_q \in W$ in the same way, and hence T maps W into W . ■

For example, let $T = \frac{d^2}{dx^2}$ on $\mathbb{R}[x]_{\leq 3} = U \oplus W$ with $U = \text{span}(1, x^2)$ (even polynomials) and $W = \text{span}(x, x^3)$ (odd polynomials). Since $T(1) = 0$, $T(x^2) = 2$, $T(x) = 0$, and $T(x^3) = 6x$, T maps each piece into itself, and in the basis $\mathcal{B} = (1, x^2, x, x^3)$

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 6 \\ 0 & 0 \end{pmatrix}.$$

In the standard order $(1, x, x^2, x^3)$ the same operator has the matrix with 2 in position $(1, 3)$

and 6 in position $(2, 4)$, which is not block diagonal for the partition $2, 2$. **Ordering the basis to match the decomposition is part of the method.**

If T maps only one subspace U into itself, we can still take a basis of U first, but the remaining basis vectors need not be sent anywhere special. What survives is a zero block in the lower left, and the lower-right block turns out to describe the operator that T induces on the quotient V/U . This was Exercise (C2: The operator induced on a quotient) in Chapter 2; we now record it as a theorem, with its converse.

Theorem 8.1.7. Subspaces Mapped into Themselves Give Block Triangular Matrices

Let V be finite-dimensional, $T \in \mathcal{L}(V)$, and U a subspace with $\{0\} \neq U \neq V$. Let $\mathcal{B}_U = (u_1, \dots, u_k)$ be a basis of U , extended to a basis $\mathcal{B} = (u_1, \dots, u_k, w_1, \dots, w_l)$ of V .

- (a) T maps U into U if and only if the lower-left $l \times k$ block of $[T]_{\mathcal{B}}^{\mathcal{B}}$ is zero.
- (b) In that case the rule $\bar{T}(v + U) = Tv + U$ defines an operator $\bar{T}$ on V/U , and

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} [T|_U]_{\mathcal{B}_U}^{\mathcal{B}_U} & \mathbf{B} \\ \mathbf{O} & [\bar{T}]_{\bar{\mathcal{B}}}^{\bar{\mathcal{B}}} \end{pmatrix}$$

for some $\mathbf{B} \in M_{k \times l}(F)$, where $\bar{\mathcal{B}} = (w_1 + U, \dots, w_l + U)$ is a basis of V/U .

Idea. Read the matrix one column at a time. The first k columns describe Tu_j , which lie in U exactly when their w -coordinates vanish. The last l columns describe Tw_q , which may have any u -part; passing to V/U erases that part and leaves precisely the w -coordinates, which are the lower-right block.

Proof.

- (a) Column $j \leq k$ of $[T]_{\mathcal{B}}^{\mathcal{B}}$ is the coordinate vector of Tu_j (Definition 4.1.1), and by Theorem 1.5.1 its last l entries vanish if and only if $Tu_j \in \text{span}(u_1, \dots, u_k) = U$. If T maps U into U , this holds for every j . Conversely, if it holds for every j , then for $u = \sum_j c_j u_j \in U$ we get $Tu = \sum_j c_j Tu_j \in U$, because U is a subspace.
- (b) Let $\pi: V \rightarrow V/U$ be the quotient map, linear with kernel U (Theorem 4.4.4). The composite πT is linear (Theorem 2.4.2), and for $u \in U$, $\pi(Tu) = 0$ because $Tu \in U$. So $U \subseteq \ker(\pi T)$, and by Theorem 4.4.10 (a) there is exactly one linear $\bar{T}: V/U \rightarrow V/U$ with $\bar{T}(v + U) = Tv + U$. By Theorem 4.4.5, $\bar{\mathcal{B}}$ is a basis of V/U .

By (a), the lower-left block is zero, and as in the proof of Theorem 8.1.6, the upper-left block is $[T|_U]_{\mathcal{B}_U}^{\mathcal{B}_U}$. For $1 \leq q \leq l$, write $Tw_q = \sum_{i=1}^k b_{iq}u_i + \sum_{p=1}^l d_{pq}w_p$, so that column $k + q$ of $[T]_{\mathcal{B}}^{\mathcal{B}}$ is $(b_{1q}, \dots, b_{kq}, d_{1q}, \dots, d_{lq})$. Applying the linear map π , and using $\pi(u_i) = 0$,

$$\bar{T}(w_q + U) = \pi(Tw_q) = \sum_{p=1}^l d_{pq}(w_p + U).$$

By Definition 4.1.1, column q of $[\bar{T}]_{\bar{\mathcal{B}}}^{\bar{\mathcal{B}}}$ is $(d_{1q}, \dots, d_{lq})$, which is the lower-right part of column $k + q$ of $[T]_{\mathcal{B}}^{\mathcal{B}}$. This proves (b), with $\mathbf{B} = (b_{iq})$. ■

The upper-left block depends only on the basis $\mathcal{B}_U$, and the lower-right block only on the cosets $w_1 + U, \dots, w_l + U$. So replacing each w_q by $w_q + u'_q$ with $u'_q \in U$ can change the top-right block $\mathbf{B}$, but leaves both diagonal blocks unchanged. The theorem splits T into two smaller operators, and every quantity that sees only diagonal blocks splits with it. By Theorem 7.8.4, $\det T = \det(T|_U) \det \bar{T}$. Applied over $F[x]$ (Chapter 7 §8 notes that the block triangular determinant holds over any commutative ring), $x\mathbf{I} - [T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} x\mathbf{I}_k - [T|_U]_{\mathcal{B}_U}^{\mathcal{B}_U} & -\mathbf{B} \\ \mathbf{O} & x\mathbf{I}_l - [\bar{T}]_{\bar{\mathcal{B}}}^{\bar{\mathcal{B}}} \end{pmatrix}$ gives $p_T = p_{T|_U} p_{\bar{T}}$. Chapter 9 builds on exactly this factorization.

Example. A Plane Mapped into Itself

Let $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$, $T(x, y, z) = (y, z, x)$, and $U = \{(x, y, z) : x + y + z = 0\}$. Show that T maps U into U , find $[T]_{\mathcal{B}}^{\mathcal{B}}$ for the basis $\mathcal{B} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{e}_3)$ with $\mathbf{u}_1 = (1, -1, 0)$, $\mathbf{u}_2 = (0, 1, -1)$, and interpret the lower-right block.

Solution. If $x + y + z = 0$, then the entries of $T(x, y, z) = (y, z, x)$ also sum to 0, so T maps U into U . The vectors $\mathbf{u}_1, \mathbf{u}_2$ lie in U and are independent (look at the first entries, then the last), and $\dim U = 2$ since U is the kernel of the nonzero functional $(x, y, z) \mapsto x + y + z$ (Theorem 2.3.1). So $(\mathbf{u}_1, \mathbf{u}_2)$ is a basis of U (Theorem 1.6.4), and $\mathcal{B}$ is a basis of $\mathbb{R}^3$: in a relation $a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{e}_3 = \mathbf{0}$, the entries sum to $c = 0$, and then $a = b = 0$, so $\mathcal{B}$ is an independent list of 3 vectors (Theorem 1.6.4). Now

$$T\mathbf{u}_1 = (-1, 0, 1) = -\mathbf{u}_1 - \mathbf{u}_2,$$

$$T\mathbf{u}_2 = (1, -1, 0) = \mathbf{u}_1,$$

$$T\mathbf{e}_3 = (0, 1, 0) = \mathbf{u}_2 + \mathbf{e}_3,$$

so

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \left(\begin{array}{cc|c} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 1 \end{array} \right).$$

The lower-left block is zero, as Theorem 8.1.7 predicts. The lower-right block (1) says $\bar{T}(\mathbf{e}_3 + U) = \mathbf{e}_3 + U$, so $\bar{T}$ is the identity on the line $\mathbb{R}^3/U$. Indeed $T\mathbf{v} - \mathbf{v} = (y - x, z - y, x - z)$ has entries summing to 0, so $T\mathbf{v} + U = \mathbf{v} + U$ for every $\mathbf{v}$: T moves vectors, but never changes the sum of their coordinates, and that sum is all that $\mathbb{R}^3/U$ remembers.

8.1.7 Exercises

8.1.7.1 A. Check your understanding

Exercise. A1

- Define what it means for partitions of $\mathbf{A} \in M_{m \times n}(F)$ and $\mathbf{B} \in M_{n \times p}(F)$ to be conformable.
- State the block multiplication formula for $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{A}' & \mathbf{B}' \\ \mathbf{C}' & \mathbf{D}' \end{pmatrix}$.
- True or false: the top-left block of that product is $\mathbf{A}'\mathbf{A} + \mathbf{C}'\mathbf{B}$. Justify your answer.
- True or false: if $\mathbf{A} \in M_k(F)$ and $\mathbf{B} \in M_l(F)$, then $\mathbf{A} \oplus \mathbf{B} = \mathbf{B} \oplus \mathbf{A}$. Justify your answer.

- (e) True or false: if $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{pmatrix}$ with $\mathbf{A}$, $\mathbf{D}$ square is invertible, then $\mathbf{M}^{-1}$ is block upper triangular. Justify your answer.
- (f) In a basis whose first k vectors form a basis of U , what does a zero lower-left block of $[T]_{\mathcal{B}}^{\mathcal{B}}$ say about T ?

Solution.

- (a) The column partition of $\mathbf{A}$ and the row partition of $\mathbf{B}$ are the same list $n_1, \dots, n_b$, in the same order (Definition 8.1.1).
- (b) It is $\begin{pmatrix} \mathbf{AA}' + \mathbf{BC}' & \mathbf{AB}' + \mathbf{BD}' \\ \mathbf{CA}' + \mathbf{DC}' & \mathbf{CB}' + \mathbf{DD}' \end{pmatrix}$, provided the partitions are conformable (Theorem 8.1.1).
- (c) False in general. The block is $\mathbf{AA}' + \mathbf{BC}'$, with each block of the left factor on the left. For 4×4 matrices with partition 2, 2, take $\mathbf{A} = \mathbf{E}_{12}$, $\mathbf{A}' = \mathbf{E}_{21}$ in $M_2(F)$, and all other blocks zero. Then the top-left block of the product is $\mathbf{AA}' = \mathbf{E}_{11}$, while $\mathbf{A}'\mathbf{A} + \mathbf{C}'\mathbf{B} = \mathbf{E}_{22}$.
- (d) False. For $k = l = 1$, $(1) \oplus (2) = \text{diag}(1, 2) \neq \text{diag}(2, 1) = (2) \oplus (1)$.
- (e) True. By Theorem 8.1.5 (b), $\mathbf{A}$ and $\mathbf{D}$ are invertible, and (c) gives $\mathbf{M}^{-1}$ with lower-left block $\mathbf{O}$.
- (f) That T maps U into U (Theorem 8.1.7 (a)).

8.1.7.2 B. Practice

Exercise. B1: Multiplying block matrices

Let

$$\mathbf{M} = \left(\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & 1 & -1 \\ \hline 2 & 0 & 3 \end{array} \right), \quad \mathbf{N} = \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 1 & 1 & 2 \\ \hline 1 & -1 & 1 \end{array} \right) \in M_3(\mathbb{Q}),$$

both with row and column partition 2, 1.

- (a) Compute $\mathbf{MN}$ blockwise, and check one block against the entry formula.
- (b) Explain why the partitions would **not** be conformable if $\mathbf{N}$ were given the row partition 1, 2 instead.

Solution.

- (a) Write $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{c} & d \end{pmatrix}$ and $\mathbf{N} = \begin{pmatrix} \mathbf{E} & \mathbf{f} \\ \mathbf{g} & h \end{pmatrix}$ with $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\mathbf{c} = \begin{pmatrix} 2 & 0 \end{pmatrix}$, $d = (3)$, $\mathbf{E} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $\mathbf{f} = \begin{pmatrix} 0 \end{pmatrix}$, $\mathbf{g} = \begin{pmatrix} 1 & -1 \end{pmatrix}$, $h = (1)$. The partitions are conformable

(both middle partitions are $2, 1$), so by [Theorem 8.1.1](#)

$$\mathbf{A}\mathbf{E} + \mathbf{b}\mathbf{g} = \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 0 & 2 \end{pmatrix},$$

$$\mathbf{A}\mathbf{f} + \mathbf{b}h = \begin{pmatrix} 4 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix},$$

$$\mathbf{c}\mathbf{E} + d\mathbf{g} = \begin{pmatrix} 2 & 0 \\ 3 & -3 \end{pmatrix} + \begin{pmatrix} 3 & -3 \end{pmatrix} = \begin{pmatrix} 5 & -3 \end{pmatrix}, \quad \mathbf{c}\mathbf{f} + dh = \mathbf{O}_1 + (3) = (3).$$

Hence

$$\mathbf{M}\mathbf{N} = \begin{pmatrix} 4 & 1 & 5 \\ 0 & 2 & 1 \\ 5 & -3 & 3 \end{pmatrix}.$$

Check the $(3, 3)$ -entry directly: row 3 of $\mathbf{M}$ is $(2, 0, 3)$ and column 3 of $\mathbf{N}$ is $(0, 2, 1)$, giving $0 + 0 + 3 = 3$.

- (b) The column partition of $\mathbf{M}$ is $2, 1$, and the row partition $1, 2$ is a different list. Concretely, the top-left block $\mathbf{A}$ of $\mathbf{M}$ is 2×2 , while the top-left block of $\mathbf{N}$ would be 1×2 , and a 2×2 matrix cannot multiply a 1×2 matrix.

Exercise. B2: Inverting a block triangular matrix

Use [Theorem 8.1.5](#) to find the inverse of

$$\mathbf{M} = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \in M_4(\mathbb{Q}),$$

and check one block of $\mathbf{M}\mathbf{M}^{-1}$.

Solution. With the partition $2, 2$, $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{pmatrix}$ with $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, $\mathbf{D} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$. Both $\det \mathbf{A} = 1$ and $\det \mathbf{D} = 1$ are nonzero, so by [Theorem 0.9.8](#)

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \quad \mathbf{D}^{-1} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}.$$

Then $\mathbf{A}^{-1}\mathbf{B} = \begin{pmatrix} -1 & -1 \\ 3 & 2 \end{pmatrix}$ and $-\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} = -\begin{pmatrix} -1 & 1 \\ 3 & -4 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -3 & 4 \end{pmatrix}$. By [Theorem 8.1.5](#) (c),

$$\mathbf{M}^{-1} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 2 & -3 & 4 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Check the top-right block of $\mathbf{M}\mathbf{M}^{-1}$:

$$\begin{aligned}\mathbf{A}(-\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1}) + \mathbf{B}\mathbf{D}^{-1} &= \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -3 & 4 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix} + \begin{pmatrix} 1 & -2 \\ 2 & -3 \end{pmatrix} = \mathbf{O},\end{aligned}$$

as it must be.

Exercise. B3: A projection in an adapted basis

Let $\mathbf{P} = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{R})$, $U = \text{span}((1, 1, 0), (0, 1, 1))$, and $W = \text{span}(\mathbf{e}_1)$.

- (a) Show that $\mathbf{P}^2 = \mathbf{P}$, that $\mathbf{P}\mathbf{u} = \mathbf{u}$ for $\mathbf{u} \in U$ and $\mathbf{P}\mathbf{e}_1 = \mathbf{0}$, and that $\mathbb{R}^3 = U \oplus W$.
- (b) Let $\mathcal{B} = ((1, 1, 0), (0, 1, 1), \mathbf{e}_1)$. Use [Theorem 8.1.6](#) to write down $[T_{\mathbf{P}}]_{\mathcal{B}}^{\mathcal{B}}$ without computing, and confirm it as $\mathbf{S}^{-1}\mathbf{P}\mathbf{S}$, where $\mathbf{S}$ has the vectors of $\mathcal{B}$ as columns.
- (c) Hence find $\mathbf{P}^{100}$ and $\text{tr } \mathbf{P}$.

Solution.

- (a) Multiplying, $\mathbf{P}^2$ has rows $(0, 1, -1)$, $(0, 1, 0)$, $(0, 0, 1)$, so $\mathbf{P}^2 = \mathbf{P}$. Also $\mathbf{P}(1, 1, 0) = (1, 1, 0)$, $\mathbf{P}(0, 1, 1) = (0, 1, 1)$, and $\mathbf{P}\mathbf{e}_1 = \mathbf{0}$ (the first column is zero), so $\mathbf{P}$ fixes every vector of U by linearity. The three vectors of $\mathcal{B}$ are the columns of $\mathbf{S} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, and expanding along the last column, $\det \mathbf{S} = 1 \cdot (1 - 0) = 1 \neq 0$. So $\mathbf{S}$ is invertible (Theorem 7.5.2), its columns form a basis of $\mathbb{R}^3$ (Theorem 3.4.4), and $\mathbb{R}^3 = U \oplus W$ by Theorem 1.7.5.
- (b) $T_{\mathbf{P}}$ maps U into U (as the identity) and W into W (as zero). By [Theorem 8.1.6](#), $[T_{\mathbf{P}}]_{\mathcal{B}}^{\mathcal{B}} = \mathbf{I}_2 \oplus \mathbf{O}_1 = \text{diag}(1, 1, 0)$. By Theorem 4.2.3, $[T_{\mathbf{P}}]_{\mathcal{B}}^{\mathcal{B}} = \mathbf{S}^{-1}\mathbf{P}\mathbf{S}$. Here $\mathbf{S}^{-1} = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{pmatrix}$ (check $\mathbf{S}\mathbf{S}^{-1} = \mathbf{I}_3$), and $\mathbf{P}\mathbf{S}$ has columns $\mathbf{P}(1, 1, 0) = (1, 1, 0)$, $\mathbf{P}(0, 1, 1) = (0, 1, 1)$, $\mathbf{P}\mathbf{e}_1 = \mathbf{0}$, so $\mathbf{S}^{-1}\mathbf{P}\mathbf{S} = \mathbf{S}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \text{diag}(1, 1, 0)$.
- (c) $\mathbf{P} = \mathbf{S} \text{diag}(1, 1, 0) \mathbf{S}^{-1}$, so $\mathbf{P}^{100} = \mathbf{S} \text{diag}(1, 1, 0)^{100} \mathbf{S}^{-1} = \mathbf{S} \text{diag}(1, 1, 0) \mathbf{S}^{-1} = \mathbf{P}$ by [Theorem 8.1.4](#) (b) (or directly from $\mathbf{P}^2 = \mathbf{P}$). By Theorem 4.2.5, $\text{tr } \mathbf{P} = \text{tr } \text{diag}(1, 1, 0) = 2$, which matches $0 + 1 + 1$.

8.1.7.3 C. Going deeper

Exercise. C1: Square-zero matrices of half rank

Let $n \geq 1$.

- (a) For $\mathbf{X} \in M_n(F)$, let $\mathbf{N}_{\mathbf{X}} = \begin{pmatrix} \mathbf{O} & \mathbf{X} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \in M_{2n}(F)$. Show that $\mathbf{N}_{\mathbf{X}}^2 = \mathbf{O}$ and that $\text{rank } \mathbf{N}_{\mathbf{X}} = \text{rank } \mathbf{X}$.
- (b) Let V have dimension $2n$, and let $T \in \mathcal{L}(V)$ satisfy $T^2 = 0$ and $\text{rank } T = n$. Show that $\text{im } T = \ker T$. Choose $w_1, \dots, w_n$ with $(Tw_1, \dots, Tw_n)$ a basis of $\text{im } T$, and put $u_i = Tw_i$. Prove that $\mathcal{B} = (u_1, \dots, u_n, w_1, \dots, w_n)$ is a basis of V and that $[T]_{\mathcal{B}}^{\mathcal{B}} = \mathbf{N}_{\mathbf{I}_n}$.
- (c) Deduce that for every invertible $\mathbf{X} \in M_n(F)$, the matrix $\mathbf{N}_{\mathbf{X}}$ is similar to $\mathbf{N}_{\mathbf{I}_n}$.

Hint: for the independence in (b), apply T .

Solution.

- (a) By Theorem 8.1.1, $\mathbf{N}_{\mathbf{X}}^2 = \begin{pmatrix} \mathbf{O} \cdot \mathbf{O} + \mathbf{X} \cdot \mathbf{O} & \mathbf{O} \cdot \mathbf{X} + \mathbf{X} \cdot \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} = \mathbf{O}$. The first n columns of $\mathbf{N}_{\mathbf{X}}$ are zero, and column $n+j$ is $\begin{pmatrix} \mathbf{x}_j \\ \mathbf{0} \end{pmatrix}$, where $\mathbf{x}_j$ is column j of $\mathbf{X}$. A combination $\sum_j c_j \begin{pmatrix} \mathbf{x}_j \\ \mathbf{0} \end{pmatrix}$ is zero exactly when $\sum_j c_j \mathbf{x}_j = \mathbf{0}$, so a list of these columns is independent exactly when the corresponding columns of $\mathbf{X}$ are. Hence $\dim \text{col}(\mathbf{N}_{\mathbf{X}}) = \dim \text{col}(\mathbf{X})$, that is, $\text{rank } \mathbf{N}_{\mathbf{X}} = \text{rank } \mathbf{X}$.
- (b) If $v = Tx \in \text{im } T$, then $Tv = T^2x = 0$, so $\text{im } T \subseteq \ker T$. By Theorem 2.3.1, $\dim \ker T = 2n - n = n = \dim \text{im } T$, so $\text{im } T = \ker T$ by Theorem 1.6.8. The vectors w_i exist because $(Tw_1, \dots, Tw_n)$ can be any basis of $\text{im } T$, each of whose vectors is T of something. Let $\sum_i a_i u_i + \sum_i b_i w_i = 0$. Applying T , and using $Tu_i = T^2w_i = 0$, gives $\sum_i b_i u_i = 0$. The u_i form a basis of $\text{im } T$, so all $b_i = 0$, and then $\sum_i a_i u_i = 0$ gives all $a_i = 0$. So $\mathcal{B}$ is a linearly independent list of $2n = \dim V$ vectors, hence a basis (Theorem 1.6.4). Since $Tu_i = 0$ and $Tw_i = u_i$, column $i \leq n$ of $[T]_{\mathcal{B}}^{\mathcal{B}}$ is zero and column $n+i$ is $\mathbf{e}_i$. So $[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} \mathbf{O} & \mathbf{I}_n \\ \mathbf{O} & \mathbf{O} \end{pmatrix} = \mathbf{N}_{\mathbf{I}_n}$.
- (c) If $\mathbf{X}$ is invertible, then by (a) the operator $T = T_{\mathbf{N}_{\mathbf{X}}}$ on F^{2n} satisfies $T^2 = 0$ and $\text{rank } T = \text{rank } \mathbf{N}_{\mathbf{X}} = \text{rank } \mathbf{X} = n$ (Theorem 4.1.5, and Theorem 3.4.4 for $\text{rank } \mathbf{X} = n$). By (b) there is a basis $\mathcal{B}$ with $[T]_{\mathcal{B}}^{\mathcal{B}} = \mathbf{N}_{\mathbf{I}_n}$. The matrix of T in the standard basis is $\mathbf{N}_{\mathbf{X}}$, so $\mathbf{N}_{\mathbf{X}}$ and $\mathbf{N}_{\mathbf{I}_n}$ are similar by Theorem 4.2.4.

Exercise. C2: Powers of block triangular matrices

Let $\mathbf{A} \in M_k(F)$, $\mathbf{D} \in M_l(F)$, and $\mathbf{B} \in M_{k \times l}(F)$, and let $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{pmatrix}$.

- (a) Prove that for every $j \geq 1$,

$$\mathbf{M}^j = \begin{pmatrix} \mathbf{A}^j & \mathbf{S}_j \\ \mathbf{O} & \mathbf{D}^j \end{pmatrix}, \quad \mathbf{S}_j = \sum_{i=0}^{j-1} \mathbf{A}^i \mathbf{B} \mathbf{D}^{j-1-i}.$$

- (b) Deduce that if $\mathbf{D} = \mathbf{O}$, then $\mathbf{M}^j = \begin{pmatrix} \mathbf{A}^j & \mathbf{A}^{j-1} \mathbf{B} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}$ for $j \geq 1$, and that if moreover

$\mathbf{A}^r = \mathbf{O}$, then $\mathbf{M}^{r+1} = \mathbf{O}$.

- (c) For $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and $\mathbf{D} = \mathbf{O}_1$, find the least j with $\mathbf{M}^j = \mathbf{O}$.

Solution.

- (a) By induction on j . For $j = 1$, $\mathbf{S}_1 = \mathbf{A}^0 \mathbf{B} \mathbf{D}^0 = \mathbf{B}$. Suppose the formula holds for j . By [Theorem 8.1.5 \(a\)](#),

$$\mathbf{M}^{j+1} = \mathbf{M}^j \mathbf{M} = \begin{pmatrix} \mathbf{A}^j \mathbf{A} & \mathbf{A}^j \mathbf{B} + \mathbf{S}_j \mathbf{D} \\ \mathbf{O} & \mathbf{D}^j \mathbf{D} \end{pmatrix},$$

and $\mathbf{A}^j \mathbf{B} + \mathbf{S}_j \mathbf{D} = \mathbf{A}^j \mathbf{B} \mathbf{D}^0 + \sum_{i=0}^{j-1} \mathbf{A}^i \mathbf{B} \mathbf{D}^{j-i} = \sum_{i=0}^j \mathbf{A}^i \mathbf{B} \mathbf{D}^{j-i} = \mathbf{S}_{j+1}$. This completes the induction.

- (b) If $\mathbf{D} = \mathbf{O}$, every term of $\mathbf{S}_j$ with $j - 1 - i \geq 1$ contains a factor $\mathbf{D} = \mathbf{O}$, and only $i = j - 1$ survives, with $\mathbf{D}^0 = \mathbf{I}_l$. So $\mathbf{S}_j = \mathbf{A}^{j-1} \mathbf{B}$ and $\mathbf{D}^j = \mathbf{O}$. If also $\mathbf{A}^r = \mathbf{O}$, then $\mathbf{M}^{r+1}$ has blocks $\mathbf{A}^{r+1} = \mathbf{O}$ and $\mathbf{A}^r \mathbf{B} = \mathbf{O}$, so $\mathbf{M}^{r+1} = \mathbf{O}$.

- (c) Here $\mathbf{A}^2 = \mathbf{O}$, so $\mathbf{M}^3 = \mathbf{O}$ by (b). But $\mathbf{M}^2 = \begin{pmatrix} \mathbf{A}^2 & \mathbf{A} \mathbf{B} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}$ with $\mathbf{A} \mathbf{B} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq \mathbf{O}$,

so $\mathbf{M}^2 \neq \mathbf{O}$; and $\mathbf{M} \neq \mathbf{O}$. The least such j is 3. Explicitly, $\mathbf{M} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ and

$$\mathbf{M}^2 = \mathbf{E}_{13}.$$