MATRIX MECHANICS

1. Introduction

These notes study the action of linear transformations on vector spaces, which amounts to analyzing matrix multiplication. We will emphasize computations over proofs. Most results will be stated either without proof, or with only an outline of the proof. The aim is to explain how to use the results in practice.

Many results will be stated in the setting of the real numbers, $\mathbb{R}$. Most of them are true also in the setting of the complex numbers, $\mathbb{C}$, and some are in fact only true in general if you allow complex numbers (notably, eigenspace results). Nevertheless, you may consider that all numbers are real, except in cases in which non-real numbers arise explicitly.

A word on the various different terms for mathematical results: lemma, theorem, proposition, and corollary. There is no precise formula determining when to use which term, and different authors use them differently. Lemmas tend to be small results, often simple to prove, not really of independent interest, but used as tools in proving larger results. Theorems tend to be major results of fundamental interest. Propositions might be said to be between lemmas and theorems: important in their own right, but not landmarks. Corollaries are results which follow more or less immediately from the result preceding them. All that said, you are perfectly welcome to simply treat all four words as synonymous.

A few very small words on notation: "$:=$" means an equation which defines its left hand side, and "$\square$" means the end of a proof (or, more likely, its outline).

2. Vector spaces

2.1. **Subspaces, linear combinations, and spans.**

Definition 2.1.

- An $m \times n$ *matrix* is a rectangular array of numbers, m high and n wide.
- An *n-vector* is a $n \times 1$ matrix, i.e., a column vector of height n. We write $\mathbb{R}^n$ for the set of all n-vectors.

$\mathbb{R}^n$ is a *vector space*. This means that vectors can be added to each other and multiplied by scalars:

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} v_1' \\ \vdots \\ v_n' \end{pmatrix} := \begin{pmatrix} v_1 + v_1' \\ \vdots \\ v_n + v_n' \end{pmatrix}, \qquad c \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} := \begin{pmatrix} cv_1 \\ \vdots \\ cv_n \end{pmatrix}.$$

Definition 2.2. A subset V of $\mathbb{R}^n$ is a *subspace* if it is *closed* under addition and scalar multiplication: $V + V = V$ and $\mathbb{R}V = V$. In other words:

- For v_1 and v_2 in V and c in $\mathbb{R}$, both $v_1 + v_2$ and cv_1 are in V.

Notation. When we use the term *vector space*, we will always mean a subspace of $\mathbb{R}^n$ (or, when it is necessary to allow complex numbers, $\mathbb{C}^n$). If V and U are vector spaces such that V contains U, we say that "U is a subspace of V".

Definition 2.3. Let $v_1, \dots, v_r$ be n-vectors, and let $c_1, \dots, c_r$ be scalars.

- A *linear combination* of $v_1, \dots, v_r$ is any vector of the form $c_1v_1 + \cdots c_rv_r$.
- $\operatorname{Span}\{v_1, \dots, v_r\}$ is the set of all linear combinations of $v_1, \dots, v_r$.

Proposition 2.1. $\operatorname{Span}\{v_1, \dots, v_r\}$ *is a subspace of* $\mathbb{R}^n$.

If $V = \operatorname{Span}\{v_1, \dots, v_r\}$, one says that

- "V is the span of $v_1, \dots, v_r$", or
- "the vectors $v_1, \dots, v_r$ span V".

2.2. Linear dependence and independence.

Definition 2.4. Let $v_1, \dots, v_r$ be n-vectors.

- They are *linearly dependent* if there exist scalars $c_1, \dots, c_r$, not all zero, such that $c_1v_1 + \cdots c_rv_r = 0$.
- Conversely, they are *linearly independent* if $c_1v_1 + \cdots c_rv_r = 0$ is only true when $c_1, \dots, c_r$ are all zero.

Proposition 2.2. *Let V be a vector space. Any collection of vectors which spans V is at least as large as any collection of linearly independent vectors in V.*

This proposition is very important. Let us restate it in a different way:

Proposition 2.2′. *Let $V = \operatorname{Span}\{v_1, \dots, v_s\}$.*

(i) *If $w_1, \dots, w_r$ are linearly independent vectors in V, then $r \leq s$.*
(ii) *If $w_1, \dots, w_r$ are vectors in V with $r > s$, then they are linearly dependent.*

Idea of proof. Suppose that $w_1, \dots, w_r$ are in $\operatorname{Span}\{v_1, \dots, v_s\}$. Then for $1 \leq i \leq r$ and $1 \leq j \leq s$, there are scalars A_{ij} such that

$$\begin{aligned} A_{11}v_1 + \cdots + A_{1s}v_s &= w_1, \\ A_{21}v_1 + \cdots + A_{2s}v_s &= w_2, \\ &\vdots \\ A_{r1}v_1 + \cdots + A_{rs}v_s &= w_r. \end{aligned}$$

If $r > s$, it is always possible to find a non-trivial linear combination of the left sides of these equations which is zero. Then the corresponding linear combination of the right sides is also zero, showing that the w_i's are linearly dependent. The procedure one uses to carry this out is row reduction of the $r \times s$ matrix A of coefficients. □

Definition 2.5. The *standard basis* of $\mathbb{R}^n$ is the set of vectors $e_1, \dots, e_n$, where e_j is the vector with a 1 in the i^{th} entry and 0's elsewhere.

Lemma 2.3. *The vectors $e_1, \dots, e_n$ are linearly independent and span $\mathbb{R}^n$.*

Corollary 2.4. (i) *Any $n+1$ vectors in $\mathbb{R}^n$ are linearly dependent.*
(ii) *Any set of linearly independent vectors in $\mathbb{R}^n$ is of size at most n.*

The next two lemmas are also very important.

Lemma 2.5. *Suppose that $v_1, \dots, v_r$ are linearly independent, and v_{r+1} is any vector not in* $\operatorname{Span}\{v_1, \dots, v_r\}$. *Then $v_1, \dots, v_{r+1}$ are also linearly independent.*

Lemma 2.6. *Suppose that $v_1, \dots, v_r$ are n-vectors.*

(i) *If they are linearly dependent, then we can delete one of them without changing their span.*

(ii) *Conversely, if they are linearly independent, then deleting any of them reduces their span.*

2.3. **Bases and dimension.** Let V be a vector space, i.e., subspace of $\mathbb{R}^n$.

Definition 2.6. A *basis* of V is a set of linearly independent vectors spanning V.

Theorem 2.7. (i) *Bases of V do exist, and they are of size $\leq n$.*

(ii) *Any set of linearly independent vectors in V is contained in a basis of V.*

(iii) *Any set of vectors which span V contains a basis of V.*

Idea of proof. For (ii), use Lemma 2.5 to repeatedly add vectors to the given linearly independent set, while maintaining linear independence. This process stops only when the set spans V.

For (iii), use Lemma 2.6 to repeatedly remove vectors from the given spanning set, while maintaining full span. This process stops only when the set becomes linearly independent.

For (i), apply (ii) to the empty set of vectors, and use Corollary 2.4. $\square$

Theorem 2.8. *All bases of V have the same size.*

Proof. By Proposition 2.2, any one basis is no bigger than any other. $\square$

Definition 2.7. The *dimension* of V, written as $\dim(V)$, is the size of its bases.

2.4. **Sums of vector spaces.** Let $V_1, V_2, \ldots, V_p$ be subspaces of $\mathbb{R}^n$.

Definition 2.8. The *vector space sum* $V_1 + \cdots + V_p$ is

$$V_1 + \cdots + V_p := \big\{ v_1 + \cdots + v_p : v_i \in V_i \text{ for } 1 \leq i \leq p \big\}.$$

Lemma 2.9. *$V_1 + \cdots + V_p$ and $V_1 \cap \cdots \cap V_p$ are vector spaces.*

Definition 2.9. Suppose that $V_1, \ldots, V_p$ have the following property: the only way to write the zero vector as a sum of p vectors, one from each subspace V_i, is to take all the summand vectors to be zero. To put it another way:

- If $v_1 + \cdots + v_p = 0$, where $v_i \in V_i$ for $1 \leq i \leq p$, then $v_i = 0$ for all i.

In this case we say that the sum of the V_i is *direct*, and we write it as

$$V_1 \oplus \cdots \oplus V_p.$$

Proposition 2.10. *Dimension "distributes" over direct sums:*

$$\dim(V_1 \oplus \cdots \oplus V_p) = \dim(V_1) + \cdots + \dim(V_p).$$

Idea of proof. Fix a basis of each the subspaces V_i, and use the direct sum property to show that the union of all these bases is a basis of the direct sum space. It is helpful to use "descriptive notation" in writing out the details. For example, you could write $v_1^i, \ldots, v_{d_i}^i$ for your basis of V_i, where d_i is $\dim(V_i)$. $\square$

Proposition 2.11. *The sum of two vector spaces is direct if and only if their intersection is zero: $V_1 \cap V_2 = 0 \iff V_1 + V_2$ is a direct sum.*

Theorem 2.12. *Suppose that U is a subspace of a vector space V. Then it is possible to find a subspace U' of V such that $U + U' = V$ and the sum is direct:*

$$V = U \oplus U'.$$

Idea of proof. Let $u_1, \ldots, u_s$ be a basis of U. Use Lemma 2.5 to pick vectors $u'_1, \ldots, u'_{s'}$ extending your basis of U to a basis of V. Let U' be $\mathrm{Span}\{u'_1, \ldots, u'_{s'}\}$, and argue that the sum of U and U' is direct. $\square$

Warning. The subspace U' in Theorem 2.12 is not unique. The misconception that it is unique is the "diagonal fallacy". It is an error that arises in many forms. Another of its forms is the misconception that if three vector spaces have the property that any two of them intersect in zero, then the sum of all three is direct. It is a good exercise to find a counterexample to this statement.

Theorem 2.13. $\dim(V_1 + V_2) + \dim(V_1 \cap V_2) = \dim(V_1) + \dim(V_2)$.

Idea of proof. Use Theorem 2.12 to choose vector spaces U'_1 and U'_2 such that

$$V_1 = U'_1 \oplus (V_1 \cap V_2), \qquad V_2 = U'_2 \oplus (V_1 \cap V_2).$$

Then argue that $V_1 + V_2 = U'_1 \oplus U'_2 \oplus (V_1 \cap V_2)$. $\square$

3. Matrices

3.1. **Matrix multiplication.** Let A be an $m \times n$ matrix. Here are some observations and notation:

- A has n columns. Each of them is an $m \times 1$ matrix, i.e., an m-vector. We write $\mathrm{Col}_j(A)$ for the j^{th} column.
- A has m rows. Each of them is an $1 \times n$ matrix, i.e., an n-vector "turned on its side". We write $\mathrm{Row}_i(A)$ for its i^{th} row.
- We write A_{ij} for the entry in the i^{th} row and the j^{th} column of A. Thus

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix}.$$

It is often useful to picture A in terms of its constituent rows or columns:

$$A = \begin{pmatrix} \mathrm{Row}_1(A) \\ \vdots \\ \mathrm{Row}_m(A) \end{pmatrix} = \begin{pmatrix} \mathrm{C} & & \mathrm{C} \\ \mathrm{o} & \cdots & \mathrm{o} \\ \mathrm{l}_1 & & \mathrm{l}_n \\ (A) & & (A) \end{pmatrix}.$$

Definition 3.1. Let A and B be matrices, and let c be a scalar.

- Matrices can be multiplied by scalars: the entries of cA are the corresponding entries of A, all multiplied by c:

$$(cA)_{ij} := cA_{ij}.$$

- Matrices of the same size can be added: if A and B are both $m \times n$, then the entries of $A + B$ are the sums of the corresponding entries of A and B:

$$(A + B)_{ij} := A_{ij} + B_{ij}.$$

- Matrices of equal "contacting" sizes can be multiplied: if B is $\ell \times m$ and A is $m \times n$, then the product BA is the $\ell \times n$ matrix with ij-entry

$$(BA)_{ij} = B_{i1}A_{1j} + \cdots + B_{im}A_{mj}.$$

We may think of the matrix product BA in terms of the dot products of the rows of B with the columns of A. Recall that the dot product of two n-vectors is the sum of the products of their entries:

$$v \cdot v' := v_1 v_1' + \cdots + v_n v_n'.$$

Thus the ij-entry of BA is $(\mathrm{Row}_i\, B) \cdot (\mathrm{Col}_j\, A)$. It is helpful to visualize this as

$$BA = \begin{pmatrix} \mathrm{Row}_1(B) \\ \vdots \\ \mathrm{Row}_\ell(B) \end{pmatrix} \begin{pmatrix} \mathrm{C} & & \mathrm{C} \\ \mathrm{o} & \cdots & \mathrm{o} \\ \mathrm{l}_1 & & \mathrm{l}_n \\ (A) & & (A) \end{pmatrix}$$

$$= \begin{pmatrix} \mathrm{Row}_1(B) \cdot \mathrm{Col}_1(A) & \cdots & \mathrm{Row}_1(B) \cdot \mathrm{Col}_n(A) \\ \vdots & \ddots & \vdots \\ \mathrm{Row}_\ell(B) \cdot \mathrm{Col}_1(A) & \cdots & \mathrm{Row}_\ell(B) \cdot \mathrm{Col}_n(A) \end{pmatrix}.$$

Remark. In working with matrices it is necessary to be comfortable with *summation notation*. For example, we may write the ij-entry of the product BA as

$$(BA)_{ij} = \sum_{r=1}^{m} B_{ir} A_{rj}. \tag{3.1}$$

Lemma 3.1. *Matrix multiplication has the following properties:*

(i) *It distributes over matrix addition.*

(ii) *It commutes with scalar multiplication.*

Lemma 3.2. *Matrix multiplication is associative: if C, B, and A are of sizes such that the products $(CB)A$ and $C(BA)$ can be formed, then they are equal.*

To say this another way, suppose that C is $k \times \ell$, B is $\ell \times m$, and A is $m \times n$. Then $(CB)A$ and $C(BA)$ are equal: their ij-entries are

$$\big((CB)A\big)_{ij} = \big(C(BA)\big)_{ij} = \sum_{r=1}^{\ell} \sum_{s=1}^{m} C_{ir} B_{rs} A_{sj}. \tag{3.2}$$

Idea of proof. Derive (3.2) from (3.1). □

Warning. Matrix multiplication is not commutative. Even when A and B are both $n \times n$ matrices, AB and BA are usually different.

We now define the *transpose*, an important concept. The *main diagonal* of a matrix runs diagonally "southeast" from its upper left corner. Transposition is "reflection across the main diagonal":

Definition 3.2. If A is $m \times n$, its transpose A^T is the $n \times m$ matrix with entries

$$(A^T)_{ij} := A_{ji}.$$

The transpose can be visualized as follows:

$$A^T = \begin{pmatrix} \mathrm{Col}_1(A) \\ \vdots \\ \mathrm{Col}_n(A) \end{pmatrix} = \begin{pmatrix} \mathrm{R} & & \mathrm{R} \\ \mathrm{o} & \cdots & \mathrm{o} \\ \mathrm{w}_1 & & \mathrm{w}_m \\ (A) & & (A) \end{pmatrix}. \tag{3.3}$$

Proposition 3.3. *The transpose "anti-distributes" over matrix multiplication:*

- *If B is $\ell \times m$ and A is $m \times n$, then $(BA)^T = A^T B^T$.*

Proof. Use (3.3): $(A^T B^T)_{ij} = (\mathrm{Col}_i\, A) \cdot (\mathrm{Row}_j\, B) = (BA)_{ji} = ((BA)^T)_{ij}$. $\square$

3.2. **Matrices as linear functions.**

Definition 3.3. An $m \times n$ matrix A defines a function from $\mathbb{R}^n$ to $\mathbb{R}^m$:

$$A : \mathbb{R}^n \to \mathbb{R}^m, \quad A(v) = Av.$$

The point here is that if v is an n-vector, i.e., an $n \times 1$ matrix, then the matrix product Av is an $m \times 1$ matrix, i.e., an m-vector.

Lemma 3.4. *The function A is "linear": for v and v' in $\mathbb{R}^n$ and $c \in \mathbb{R}$,*

$$A(v + v') = Av + Av', \qquad A(cv) = cAv$$

Proof. This follows from Lemma 3.1. $\square$

Proposition 3.5. *Composition of linear functions corresponds to matrix multiplication: for B an $\ell \times m$ matrix, A an $m \times n$ matrix, and $v \in \mathbb{R}^n$, we have*

$$A(Bv) = (AB)v.$$

Proof. This follows from Lemma 3.2. $\square$

Lemma 3.6. *Suppose that A is an $m \times n$ matrix and v is an n-vector. Then we may write the m-vector Av in terms of either the rows or the columns of A:*

$$A = \begin{pmatrix} \mathrm{Row}_1(A) \cdot v \\ \vdots \\ \mathrm{Row}_m(A) \cdot v \end{pmatrix}, \qquad A = v_1 \begin{pmatrix} \mathrm{C} \\ \mathrm{o} \\ \mathrm{l}_1 \\ (A) \end{pmatrix} + \cdots + v_n \begin{pmatrix} \mathrm{C} \\ \mathrm{o} \\ \mathrm{l}_n \\ (A) \end{pmatrix}.$$

Definition 3.4. Let A be an $m \times n$ matrix, regarded as a function from $\mathbb{R}^n$ to $\mathbb{R}^m$.

- The *range* of A is $\mathcal{R}(A) := \{Av : v \in \mathbb{R}^n\}$.
- The *null space* of A is $\mathcal{N}(A) := \{v \in \mathbb{R}^n : Av = 0\}$.

Lemma 3.7. *Let A be an $m \times n$ matrix.*

(i) *$\mathcal{R}(A)$ is the subspace of $\mathbb{R}^m$ spanned by the columns of A.*
(ii) *$\mathcal{N}(A)$ is a subspace of $\mathbb{R}^n$. It consists of all n-vectors whose dot products with all the rows of A are zero.*

Proof. This follows from Lemma 3.6. $\square$

Theorem 3.8. *For any matrix, the dimension of the domain is the sum of the dimensions of the range and null space. In other words, for any $m \times n$ matrix A,*

$$\dim \mathcal{R}(A) + \dim \mathcal{N}(A) = n.$$

Idea of proof. Choose a basis $w_1, \ldots, w_r$ of $\mathcal{R}(A)$, and then choose vectors $v_1, \ldots, v_r$ such that $Av_i = w_i$ for $1 \le i \le r$. Define

$$V := \mathrm{Span}\,\{v_1, \ldots, v_r\}.$$

Use the linear independence of the w_i's to verify that the v_i's are also linearly independent. Hence V and $\mathcal{R}(A)$ both have dimension r.

To conclude the argument, deduce that $\mathbb{R}^n$ is the direct sum of V and $\mathcal{N}(A)$, and apply Proposition 2.10. $\square$

Definition 3.5. The *rank* of a matrix is the dimension of its range:

$$\operatorname{rank}(A) := \dim \mathcal{R}(A).$$

4. Invertible matrices

Recall from Definition 2.5 the standard basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$.

Proposition 4.1. *Let $w_1, \ldots, w_n$ be arbitrary vectors in $\mathbb{R}^m$. Then there exists a unique $m \times n$ matrix A such that $Ae_i = w_i$ for $1 \le i \le n$, namely, the matrix whose i^{th} column is w_i:*

$$A = \begin{pmatrix} | & & | \\ w_1 & \cdots & w_n \\ | & & | \end{pmatrix}.$$

5. Square matrices

Throughout this chapter, A denotes an $n \times n$ matrix.

Notation.
- If c is a scalar, "$A + c$" means $A + cI$.
- Blank regions in matrices are understood to contain only zeroes.
- Starred regions in matrices are understood to contain arbitrary entries.

5.1. Eigenvectors, eigenvalues, and eigenspaces.

Definition 5.1. Suppose that v is a non-zero n-vector, λ is a scalar, and $Av = \lambda v$. Then v is an *eigenvector* of A, of *eigenvalue* λ.

- Eigenvectors of eigenvalue λ may be referred to as λ-*eigenvectors*.
- If A has λ-eigenvectors, we say that λ is an *eigenvalue of* A.

Definition 5.2. The λ-*eigenspace* of A is the set of all λ-eigenvectors of A, together with the zero vector. We denote it by V_λ:

$$V_\lambda := \big\{v \in \mathbb{R}^n : Av = \lambda v\big\}.$$

Lemma 5.1. *The λ-eigenspace of A is the null space of $A - \lambda$:*

$$V_\lambda = \mathcal{N}(A - \lambda).$$

In particular, λ is an eigenvalue of A if and only if $A - \lambda$ is not invertible.

Definition 5.3. The *characteristic polynomial* of A is the determinant of $t - A$:

$$\operatorname{char}_A(t) := \det(t - A).$$

We will usually write the factorization of $\operatorname{char}_A(t)$ like this:

$$\operatorname{char}_A(t) := (t - \lambda_1)^{d_1} \cdots (t - \lambda_r)^{d_r}.$$

Here $\lambda_1, \ldots, \lambda_r$ are the distinct roots of $\operatorname{char}_A(t)$, and $d_1, \ldots, d_r$ are their multiplicities. The degree of char_A is n, so $d_1 + \cdots + d_r = n$.

Proposition 5.2.
(i) *The eigenvalues of A are the roots of* $\operatorname{char}_A(t)$.
(ii) *The sum of the eigenspaces of A is direct:*

$$V_{\lambda_1} + \cdots + V_{\lambda_r} = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_r}.$$

(ii) *For $1 \le s \le r$, $\dim(V_{\lambda_s}) \le d_s$. Therefore*

$$\dim(V_{\lambda_1}) + \cdots + \dim(V_{\lambda_r}) \le n.$$

Definition 5.4. A is *diagonalizable* if there is a diagonal matrix Λ and an invertible matrix P such that $A = P\Lambda P^{-1}$. In this case, we say that

- "*P diagonalizes A*", or
- "$P\Lambda P^{-1}$ is a *diagonalization* of A".

Proposition 5.3. (i) *Suppose that $P\Lambda P^{-1}$ is a diagonalization of A. Then* $\mathrm{Col}_i(P)$ *is an eigenvector of A, of eigenvalue Λ_{ii}.*

(ii) *Conversely, suppose that $v_1, \ldots, v_n$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A. Let λ_i be the eigenvalue of v_i. Then $A = P\Lambda P^{-1}$, where*

$$P = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

Theorem 5.4. *The following conditions are equivalent:*

(i) *A is diagonalizable.*

(ii) $\dim(V_{\lambda_s}) = d_s$ *for* $1 \leq s \leq r$.

(iii) *The direct sum of the eigenspaces of A is all of $\mathbb{R}^n$, i.e.,*

$$V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_r} = \mathbb{R}^n.$$

The next proposition gives one of the many uses of diagonalizations. In fact, (5.1) is true for any function $f(x)$ with a power series expansion. A particularly important example is the exponential function e^x.

Proposition 5.5. *If $A = P\Lambda P^{-1}$ and $f(x)$ is any polynomial, then*

$$f(A) = Pf(\Lambda)P^{-1} = P \begin{pmatrix} f(\Lambda_{11}) & & \\ & \ddots & \\ & & f(\Lambda_{nn}) \end{pmatrix} P^{-1}. \tag{5.1}$$

Diagonalization. Here is a summary of the procedure:

Step 1. Compute the characteristic polynomial $\mathrm{char}_A(t) = \det(t - A)$. Factor it:

$$\mathrm{char}_A(t) := (t - \lambda_1)^{d_1} \cdots (t - \lambda_r)^{d_r}, \qquad d_1 + \cdots + d_r = n. \tag{5.2}$$

Here $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues, and $d_1, \ldots, d_r$ are their multiplicities.

Step 2. The λ_s-eigenspace V_{λ_s} is $\mathcal{N}(A - \lambda_s)$. Use row operations to find bases of all the eigenspaces. A is diagonalizable if and only if $\dim(V_{\lambda_s}) = d_s$ for all s.

Step 3. Assume that A is diagonalizable. Let $v_1, \ldots, v_n$ be a list of the basis vectors of all the eigenspaces, ordered as follows: the basis of V_{λ_1} first, then the basis of V_{λ_2}, and so on, ending with the basis of V_{λ_r}. Define P by $\mathrm{Col}_i(P) = v_i$. Let Λ_s be the $d_s \times d_s$ matrix $\lambda_s I$. Then

$$A = P\,\Lambda P^{-1}, \quad \text{where} \quad \Lambda = \begin{pmatrix} \Lambda_1 & & \\ & \ddots & \\ & & \Lambda_r \end{pmatrix}.$$

5.2. **Generalized eigenspaces.** For most matrices, the characteristic polynomial has no repeated roots. In this case there are n distinct eigenvalues, each eigenspace is 1-dimensional, and the matrix is diagonalizable.

When the characteristic polynomial has repeated roots, the matrix may be diagonalizable, but usually it is not. In this section we describe an analog of diagonalization for non-diagonalizable matrices. As usual, write the factorization of the characteristic polynomial as in (5.2).

Definition 5.5. The λ_s*-generalized eigenspace* of A is

$$V^{\mathrm{g}}_{\lambda_s} := \mathcal{N}(A-\lambda_s)^{d_s}.$$

Theorem 5.6. (i) *$\dim(V^{\mathrm{g}}_{\lambda_s}) = d_s$ for all s.*

(ii) *The sum of the $V^{\mathrm{g}}_{\lambda_s}$ is direct and is all of $\mathbb{R}^n$:*

$$\mathbb{R}^n = V^{\mathrm{g}}_{\lambda_1} \oplus \cdots \oplus V^{\mathrm{g}}_{\lambda_r}.$$

We now explain how to construct a certain type of basis of $V^{\mathrm{g}}_{\lambda_s}$.

Proposition 5.7. $\mathcal{N}(A-\lambda_s) \subseteq \mathcal{N}(A-\lambda_s)^2 \subseteq \cdots \subseteq \mathcal{N}(A-\lambda_s)^{d_s}$.

Definition 5.6. Choose a basis $v^s_1, \ldots, v^s_{d_s}$ of $V^{\mathrm{g}}_{\lambda_s}$ in the following order:

- First, choose a basis of $\mathcal{N}(A-\lambda_s)$.
- Second, extend it to a basis of $\mathcal{N}(A-\lambda_s)^2$.
- Continue extending it to the null spaces of successively higher powers of $A-\lambda_s$ until you have reached a basis of $\mathcal{N}(A-\lambda_s)^{d_s}$.

Proposition 5.8. *Let P be the matrix whose columns are the following basis of $\mathbb{R}^n$, in the following order:*

$$v^1_1, \ldots, v^1_{d_1},\ v^2_1, \ldots, v^2_{d_2},\ \ldots,\ v^r_1, \ldots, v^r_{d_r}. \tag{5.3}$$

Then $A = P\Gamma P^{-1}$, for a "block-diagonal" matrix Γ of the following form:

$$\Gamma = \begin{pmatrix} \Gamma_1 & & \\ & \ddots & \\ & & \Gamma_r \end{pmatrix},$$

where Γ_s is an upper triangular $d_s \times d_s$ matrix whose diagonal entries are all λ_s:

$$\Gamma_s = \begin{pmatrix} \lambda_s & & & & \\ & \lambda_s & & * & \\ & & \ddots & & \\ & & & \lambda_s & \\ & & & & \lambda_s \end{pmatrix}.$$

We will refer $P\Gamma P^{-1}$ as a "generalized diagonalization" of A.

Generalized diagonalization. Here is a summary of the procedure:

Step 1. Exactly as for diagonalization, factor $\mathrm{char}_A(t)$ as $(t-\lambda_1)^{d_1} \cdots (t-\lambda_r)^{d_r}$.

Step 2. The generalized eigenspace $V^{\mathrm{g}}_{\lambda_s}$ is $\mathcal{N}(A-\lambda_s)^{d_s}$. It is of dimension d_s. For each s, find a basis of $\mathcal{N}(A-\lambda_s)$. Extend it to a basis of $\mathcal{N}(A-\lambda_s)^2$. Continue until you have a basis of $\mathcal{N}(A-\lambda_s)^{d_s}$ as in Definition 5.6:

$$v_1^s, \dots, v_{d_s}^s. \tag{5.4}$$

Step 3. Let $v_1, \dots, v_n$ be the basis (5.3) of $\mathbb{R}^n$ given by combining all the bases (5.4) of the generalized eigenspaces, in order of increasing s. Define P by $\mathrm{Col}_i(P) = v_i$.

At this point you could obtain Γ as $P^{-1}AP$. However, it can be difficult to find P^{-1} and compute the triple product. There is another more conceptual way to obtain the $d_s \times d_s$ matrices Γ_s which make up Γ:

Step 4. Check that $(A-\lambda_s)v_1^s = 0$. Then compute $(A-\lambda_s)v_d^s$ for $1 < d \leq d_s$. It will be a linear combination of the preceding basis vectors:

$$(A-\lambda_s)v_d^s = g_{1,d}v_1^s + g_{2,d}v_2^s + \cdots + g_{d-1,d}v_{d-1}^s.$$

The scalars $g_{c,d}$ are the entries of Γ_s above the diagonal: for $c < d$,

$$(\Gamma_s)_{c,d} = g_{c,d}.$$

Because Γ_s is upper triangular with all diagonal entries equal to λ_s, we now know all of its entries. Combining all the Γ_s into Γ completes the process.

Here is a brief justification of Step 4. Because P maps the standard basis of $\mathbb{R}^n$ to the basis in (5.3) and $A = P\Gamma P^{-1}$, we have the following statement:

- Γ acts on the standard basis as A acts on the basis in (5.3).

6. Orthogonal matrices

Definition 6.1. Let v and v' be vectors in $\mathbb{R}^n$.

- The *length* of v is $\|v\| := \sqrt{v^T v}$.
- v is said to be a *unit vector* if its length is 1.
- v and v' are said to be *orthogonal* if their dot product $v^T v'$ is 0.
- For $v \neq 0$, we write $\hat{v}$ for $v/\|v\|$, the unit vector in the direction of v.

Definition 6.2. A collection of mutually orthogonal unit vectors in $\mathbb{R}^n$ is said to be *orthonormal* (sometimes abbreviated to ON).

Lemma 6.1. *Any collection of orthonormal vectors in $\mathbb{R}^n$ is linearly independent and may be extended to an orthonormal basis of $\mathbb{R}^n$.*

More generally, if V is a subspace of $\mathbb{R}^n$, any collection of orthonormal vectors in V may be extended to an orthonormal basis of V.

Idea of proof. Suppose that $u_1, \dots, u_r$ are ON vectors in V. To see that they are linearly independent, suppose that $c_1u_1 + \cdots + c_ru_r = 0$. Check that taking the dot product of this equation with u_s gives $c_s = 0$.

The extension property if proven using the *Gram-Schmidt process.* Suppose that $u_1, \dots, u_r$ do not span V. Choose $v \in V$ not in their span. Let

$$w = v - (u_1^T v)u_1 - \cdots - (u_r^T v)u_r. \tag{6.1}$$

Check that $u_s^T w = 0$ for $1 \leq s \leq r$, but $w \neq 0$. Define $u_{r+1} := \hat{w}$. If $u_1, \dots, u_{r+1}$ span V, we are finished. If not, repeat the process. $\square$

Definition 6.3. A real square matrix M is *orthogonal* if its transpose is its inverse:

$$M^T = M^{-1}.$$

Lemma 6.2. (i) *The determinant of an orthogonal matrix is* ± 1.

(ii) *Transposes of orthogonal matrices are orthogonal:*
- M *orthogonal* $\implies M^T$ *orthogonal.*

(iii) *Products of orthogonal matrices are orthogonal:*
- M, M' *both* $n \times n$ *orthogonal* $\implies MM'$ *orthogonal.*

(iv) M *is orthogonal if and only if its columns are orthonormal.*

Proposition 6.3. *Let* M *be a* 2×2 *orthogonal matrix.*

(i) *If* $\det(M) = 1$, *then* M *is rotation by some angle* θ:

$$M = R_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

(ii) *If* $\det(M) = -1$, *then* M *is reflection over some angle* θ:

$$M = F_\theta := \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}.$$

Definition 6.4. The *trace* of an $n \times n$ matrix A is

$$\operatorname{tr}(A) := A_{11} + \cdots + A_{nn}.$$

Proposition 6.4. *Let* M *be a* 3×3 *orthogonal matrix.*

(i) *If* $\det(M) = 1$, *then* M *has an eigenvector* v *of eigenvalue* 1, *and* M *rotates* $\mathbb{R}^3$ *by an angle* θ *around the axis* $\mathbb{R}v$, *where*

$$1 + 2\cos\theta = \operatorname{tr}(M).$$

(ii) *If* $\det(M) = -1$, *then* M *has an eigenvector* v *of eigenvalue* -1, *and* M *both reflects* $\mathbb{R}^3$ *across the plane* $v^\perp$ *and rotates it by an angle* θ *around the axis* $\mathbb{R}v$, *where*

$$-1 + 2\cos\theta = \operatorname{tr}(M).$$

7. Symmetric matrices

Definition 7.1. If V is any subspace of $\mathbb{R}^n$, then $V^\perp$ is the set of all vectors in $\mathbb{R}^n$ orthogonal to every vector in V:

$$V^\perp := \{ w \in \mathbb{R}^n : w^T v = 0 \ \forall\ v \in V \}.$$

Lemma 7.1. (i) $V^\perp$ *is a subspace of* $\mathbb{R}^n$.

(ii) *The sum* $V + V^\perp$ *is direct and is all of* $\mathbb{R}^n$.

(iii) $(V^\perp)^\perp = V$.

Definition 7.2. Two subspaces V and V' of $\mathbb{R}^n$ are said to be *orthogonal* if every vector in V is orthogonal to every vector in V'. This is equivalent to the condition

$$V' \subseteq V^\perp.$$

Lemma 7.2. *Suppose that* $V_1, \ldots, V_r$ *are "mutually orthogonal" subspaces of* $\mathbb{R}^n$: *for all* $i \neq j$, V_i *and* V_j *are orthogonal to one another. Then the sum* $V_1 + \cdots + V_r$ *is direct. Such direct sums are said to be* orthogonal.

Definition 7.3. A real square matrix S is *symmetric* if it is its own transpose:

$$S = S^T.$$

Theorem 7.3 (the spectral theorem)**.** *Let S be a symmetric $n \times n$ matrix.*

(i) *The eigenvalues of S are real.*

(ii) *S "admits orthonormal bases of eigenvectors": it is possible to find an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of S.*

(iii) *S "may be diagonalized by an orthogonal matrix": it is possible to find an orthogonal matrix M and a real diagonal matrix Λ such that*

$$S = M\Lambda M^T.$$

Theorem 7.3′ (a reformulation of the spectral theorem)**.** *If S is a symmetric $n \times n$ matrix, then $\mathbb{R}^n$ is the orthogonal direct sum of its eigenspaces.*

Orthogonal diagonalization. Let S be a symmetric matrix.

Step 1. As for ordinary diagonalization, factor $\mathrm{char}_S(t)$ as $(t-\lambda_1)^{d_1}\cdots(t-\lambda_r)^{d_r}$.

Step 2. The λ_s-eigenspace V_{λ_s} is $\mathcal{N}(S-\lambda_s)$. Use row reduction and the Gram-Schmidt process to find an ON basis of each eigenspace.

Step 3. Let $u_1, \ldots, u_n$ be a list of the ON basis vectors of all the eigenspaces, ordered according to the order $V_{\lambda_1}, \ldots, V_{\lambda_r}$. Define M by $\mathrm{Col}_i(M) = u_i$. Let Λ_s be the $d_s \times d_s$ matrix $\lambda_s I$. Then M will be orthogonal, and

$$S = M\Lambda M^T, \quad \text{where} \quad \Lambda = \begin{pmatrix} \Lambda_1 & & \\ & \ddots & \\ & & \Lambda_r \end{pmatrix}.$$

8. Projection matrices

Definition 8.1. Let V be a subspace of $\mathbb{R}^n$. The *orthogonal projection matrix P_V* is the (unique) matrix with 1-eigenspace V and 0-eigenspace $V^\perp$.

Lemma 8.1. (i) *P_V is symmetric, and "idempotent": $P_V^2 = P_V$.*

(ii) *The sum $P_V + P_{V^\perp}$ is the identity I, and the product $P_V P_{V^\perp}$ is zero.*

(iii) *If $u_1, \ldots, u_r$ is an orthonormal basis of V, then $P_V = u_1 u_1^T + \cdots + u_r u_r^T$.*

Corollary 8.2. *If v is any non-zero vector, then projection to the line $\mathbb{R}v$ is*

$$P_{\mathbb{R}v} = \hat{v}\hat{v}^T = \frac{vv^T}{v^T v}.$$

Warning. Suppose that V is a 2-dimensional subspace spanned by unit vectors u_1 and u_2, which are *not* orthogonal. Then P_V is *not* $u_1 u_1^T + u_2 u_2^T$.

Theorem 7.3″ (another reformulation of the spectral theorem)**.** *Let S be a symmetric matrix, with distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, and corresponding eigenspaces $V_{\lambda_1}, \ldots, V_{\lambda_r}$. Then*

$$S = \lambda_1 P_{V_{\lambda_1}} + \cdots + \lambda_r P_{V_{\lambda_r}}$$

Remark. Projections give a better way to write the Gram-Schmidt process (6.1):

$$w = v - P_{\mathrm{Span}\{u_1,\ldots,u_r\}} v.$$

This gives a concise way to understand a factorization of invertible matrices known as the *QR-decomposition*, or, in a more general context, the *Iwasawa decomposition*:

Theorem 8.3. *Every invertible matrix A has a unique factorization*

$$A = KDU,$$

where K is orthogonal, D is diagonal with positive diagonal entries, and U is upper triangular with 1's on the diagonal.

Idea of proof. Write $v_1, \dots, v_n$ for the columns of A. Define $w_1, \dots, w_n$ as follows. First, set $w_1 := v_1$. Define $w_2, w_3, \dots$ via the Gram-Schmidt process:

$$w_2 := v_2 - P_{\mathbb{R}w_1} v_2, \quad w_3 := v_3 - P_{\mathrm{Span}\{w_1, w_2\}} v_3, \quad w_4 := v_4 - P_{\mathrm{Span}\{w_1, w_2, w_3\}} v_4,$$

and so on. Then $w_1, \dots, w_n$ are orthogonal, so $\hat{w}_1, \dots, \hat{w}_n$ are ON. Set

$$K := \begin{pmatrix} | & & | \\ \hat{w}_1 & \cdots & \hat{w}_n \\ | & & | \end{pmatrix}, \qquad D := \begin{pmatrix} \|w_1\| & & \\ & \ddots & \\ & & \|w_n\| \end{pmatrix},$$

and define U to be the upper triangular matrix with 1's on the diagonal and super-diagonal entries $U_{ij} := w_i^T v_j / w_i^T w_i$. Check that KD has columns $w_1, \dots, w_n$, and U encodes the column operations necessary to reverse the Gram-Schmidt process and transform KD back into A. The point is that because the w_i's are orthogonal,

$$P_{\mathrm{Span}\{w_1, \dots, w_r\}} = P_{\mathbb{R}w_1} + \cdots + P_{\mathbb{R}w_r}. \qquad \square$$

9. Positive matrices

Definition 9.1. Let S be a symmetric matrix.

- S is *positive definite symmetric* (PDS) if $v^T S v > 0$ for all vectors $v \neq 0$.
- S is *positive semidefinite symmetric* (PSDS) if $v^T S v \geq 0$ for all v.

Proposition 9.1. *Let S be a symmetric matrix.*

(i) *S is PDS if and only if its eigenvalues are all positive.*

(ii) *S is PSDS if and only if its eigenvalues are all non-negative.*

Idea of proof. We will only discuss (i); the reasoning for (ii) is similar. If v is an eigenvector of eigenvalue λ, then

$$0 < v^T S v = \lambda v^T v = \lambda \|v\|^2 \implies \lambda > 0.$$

Conversely, suppose that the eigenvalues $\lambda_1, \dots, \lambda_n$ of S are all positive, and let $u_1, \dots, u_n$ be a corresponding ON eigenbasis. Any non-zero vector v may be written as $c_1 u_1 + \cdots + c_n u_n$, where not all of the c_i are 0. Check that

$$v^T S v = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2 > 0. \qquad \square$$

Corollary 9.2. *A PSDS matrix is PDS if and only if it is invertible.*

Proposition 9.3. *Every PSDS matrix has a unique PSDS square root.*

Idea of proof. Let S be PSDS, and let $M \Lambda M^T$ be an orthogonal diagonalization. By Proposition 9.1, the diagonal entries of Λ are non-negative. Define $\Lambda^{1/2}$ to be the diagonal matrix such that for $1 \leq i \leq n$, the diagonal entry $\Lambda_{ii}^{1/2}$ is the *non-negative* square root of Λ_{ii}. Then the PSDS square root of S is

$$S^{1/2} := M \Lambda^{1/2} M^T.$$

The uniqueness follows from the fact that S and $S^{1/2}$ have the same eigenspaces: for every eigenvalue λ of S, the λ-eigenspace of S and the $\lambda^{1/2}$-eigenspace of $S^{1/2}$ are equal. (It is instructive to work out the details!) $\square$

Lemma 9.4. *Let S be PSDS. Then $S^{1/2}$ and S have the same nullspaces:*

$$\mathcal{N}(S^{1/2}) = \mathcal{N}(S).$$

Proof. As noted in the last proof, S and $S^{1/2}$ have the same 0-eigenspaces. □

The next theorem gives computationally effective criteria for positivity. We will omit the proof, but it is easier than you might think: try it for small matrices. It leads to a factorization of PDS matrices called the *Cholesky decomposition*:

Theorem 9.5. *Let S be symmetric. The following conditions are equivalent:*

(i) *S is PDS.*

(ii) *For all r, the upper left $r \times r$ submatrix of S has positive determinant.*

(iii) *Performing lower triangular row operations on S reduces it to an upper triangular matrix whose pivots are all positive. In other words, there is a lower triangular matrix L with 1's on the diagonal, and an upper triangular matrix U with positive diagonal entries, such that*

$$S = LU.$$

Corollary 9.6. *Let S be PDS.*

(i) *There is a unique lower triangular matrix L with 1's on the diagonal, and a unique diagonal matrix D with positive diagonal entries, such that*

$$S = LDL^T.$$

(ii) *The Cholesky decomposition: there is a unique lower triangular matrix T with positive diagonal entries such that*

$$S = TT^T.$$

Proof. Begin by performing lower triangular row operations on S to obtain the factorization LU in Theorem 9.5(iii). Then let D be the diagonal matrix with the same diagonal entries as U. Because S is symmetric, U will be DL^T, giving (i). For (ii), define T to be $LD^{1/2}$. □

10. The polar decomposition

Lemma 10.1. *Let A be any $m \times n$ matrix.*

(i) *$A^T A$ is $n \times n$ and PSDS.*

(ii) *A and $A^T A$ have the same null spaces: $\mathcal{N}(A) = \mathcal{N}(A^T A)$.*

(ii) *A and $A^T A$ have the same rank.*

Proof. For (i), note that $A^T A$ is symmetric, and for any vector v,

$$v^T (A^T A) v = (Av)^T (Av) = \|Av\|^2 \geq 0.$$

For (ii), note that if $Av = 0$, then $A^T Av = 0$. Conversely, if $A^T Av = 0$, then multiplying on the left by v^T yields $\|Av\| = 0$, proving $Av = 0$. □

Definition 10.1. The *modulus* of an arbitrary $m \times n$ A is the PSDS $n \times n$ matrix

$$|A| := (A^T A)^{1/2}.$$

Lemma 10.2. *$|A|$ is invertible if and only if the null space $\mathcal{N}(A)$ of A is 0.*

Proof. Being square, $|A|$ invertible if and only if $\mathcal{N}(|A|) = 0$. By Lemmas 9.4 and 10.1(ii), $\mathcal{N}(|A|) = \mathcal{N}(A^T A) = \mathcal{N}(A)$. □

Definition 10.2. Let A be a square matrix. A *polar decomposition* of A is a factorization YS, where Y is orthogonal and S is PSDS.

Theorem 10.3. (i) *If $A = YS$ is a polar decomposition, then $S = |A|$.*

(ii) *Invertible matrices have unique polar deompositions*

(iii) *Every square matrix has at least one polar decomposition.*

Proof. For (i), note that $A^TA = S^TY^TYS = S^2$, as $Y^T = Y^{-1}$. This means that S is a PSDS square root of A^TA, so it must be $|A|$.

For (ii), if A is invertible, then $|A|$ is invertible, so the only choice for Y is $A|A|^{-1}$. We must check that this matrix is orthogonal, i.e., its transpose is its inverse. Keeping in mind that $|A|$ and hence also $|A|^{-1}$ are symmetric, compute:

$$(A|A|^{-1})^T(A|A|^{-1}) = |A|^{-1}(A^TA)|A|^{-1} = |A|^{-1}|A|^2|A|^{-1} = I.$$

For (iii), use A's singular value decomposition $K\Sigma M^T$, given in the next section: it has the property that $|A| = M\Sigma M^T$, so $Y = KM^T$ gives $A = Y|A|$. □

11. The singular value and Schmidt decompositions

Let A be an arbitrary $m \times n$ matrix of rank r.

Definition 11.1. A *singular value decomposition* (SVD) of A is a factorization

$$A = K\Sigma M^T, \quad \text{where}$$

- K is $m \times m$ orthogonal;
- M is $n \times n$ orthogonal;
- Σ is "$m \times n$ diagonal": its only non-zero entries are $\Sigma_{11}, \Sigma_{22}, \dots, \Sigma_{rr}$, which are assumed to be positive and non-increasing.

Definition 11.2. A *reduced SVD* (rSVD) of A is a factorization

$$A = \tilde{K}\tilde{\Sigma}\tilde{M}^T, \quad \text{where}$$

- $\tilde{K}$ is $m \times r$ with ON columns;
- $\tilde{M}$ is $n \times r$ with ON columns;
- $\tilde{\Sigma}$ is diagonal with non-increasing positive diagonal entries.

Definition 11.3. A *Schmidt decomposition* of A is a sum:

$$A = \sigma_1 w_1 u_1^T + \cdots + \sigma_r w_r u_r^T, \quad \text{where} \tag{11.1}$$

- $w_1, \dots, w_r$ are ON vectors in $\mathbb{R}^m$;
- $u_1, \dots, u_r$ are ON vectors in $\mathbb{R}^n$;
- $\sigma_1 \geq \cdots \geq \sigma_r$ are positive scalars.

Before explaining how to compute these decompositions, we give some propositions showing that they are all equivalent, and although none of them is unique, the scalars appearing in them are.

Proposition 11.1. *The non-zero entries of Σ in any SVD, the diagonal entries of $\tilde{\Sigma}$ in any rSVD, and the scalars σ_i in any Schmidt decomposition are all the same: they are the non-zero eigenvalues of the modulus $|A|$, listed in non-increasing order:*

$$\sigma_i = \tilde{\Sigma}_{ii} = \Sigma_{ii} \textit{ for } 1 \leq i \leq r.$$

Idea of proof. Suppose we have all three types of decompositions of A:

$$A = K\Sigma M^T = \tilde{K}\tilde{\Sigma}\tilde{M}^T = \sigma_1 w_1 u_1^T + \cdots + \sigma_r w_r u_r^T.$$

Transposing everything gives all three types of decompositions of A^T:

$$A^T = M\Sigma^T K^T = \tilde{M}\tilde{\Sigma}\tilde{K}^T = \sigma_1 u_1 w_1^T + \cdots + \sigma_r u_r w_r^T.$$

Because $\tilde{K}$ has ON columns, $\tilde{K}^T\tilde{K} = I_{r\times r}$. Hence

$$A^T A = M\Sigma^T \Sigma M^T = \tilde{M}\tilde{\Sigma}^2\tilde{M}^T = \sigma_1^2 u_1 u_1^T + \cdots + \sigma_r^2 u_r u_r^T.$$

It follows that as i runs from 1 to r, the σ_i^2, the $\tilde{\Sigma}_{ii}^2$, and the Σ_{ii}^2 all run through the positive eigenvalues of $A^T A$. Hence they are all the same. □

Corollary 11.2. *SVDs, rSVDs, and Schmidt decompositions all have the following property: the transpose of the decomposition is the decomposition of the transpose.*

In other words, the transpose of an SVD, rSVD, or Schmidt decomposition of A is an SVD, rSVD, or Schmidt decomposition of A^T.

Proposition 11.3. *To convert an rSVD to a Schmidt decomposition and vice versa, use the following equations:*

$$\tilde{K} := \begin{pmatrix} | & & | \\ w_1 & \cdots & w_r \\ | & & | \end{pmatrix}, \quad \tilde{\Sigma} := \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}, \quad \tilde{M} := \begin{pmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{pmatrix}.$$

Proof. Just check that for this $\tilde{K}$, $\tilde{\Sigma}$, and $\tilde{M}$, $\tilde{K}\tilde{\Sigma}\tilde{M}$ is nothing but (11.1). □

Proposition 11.4. *To convert an SVD $K\Sigma M^T$ to an rSVD $\tilde{K}\tilde{\Sigma}\tilde{M}^T$,*

- *let $\tilde{K}$ be the first r columns of K;*
- *let $\tilde{M}$ be the first r columns of M;*
- *let $\tilde{\Sigma}$ be the upper left $r \times r$ block of Σ.*

Conversely, to convert an rSVD $\tilde{K}\tilde{\Sigma}\tilde{M}^T$ to an rSVD $K\Sigma M^T$,

- *let the first r columns of K be $\tilde{K}$, and let the remaining columns be any extension of the first r to an ON basis of $\mathbb{R}^m$;*
- *let the first r columns of M be $\tilde{M}$, and let the remaining columns be any extension of the first r to an ON basis of $\mathbb{R}^n$;*
- *let Σ be the $m \times n$ matrix with Σ as its upper left block and 0's elsewhere.*

Proof. Note that Σ and $\tilde{\Sigma}$ have the following shapes:

$$\Sigma = \begin{pmatrix} \tilde{\Sigma} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix}, \quad \text{where} \quad \tilde{\Sigma} := \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}.$$

This implies that only the first r columns of K and M affect the SVD:

$$K\Sigma M^T = \Sigma_{11}\operatorname{Col}_1(K)\operatorname{Col}_1(M)^T + \cdots + \Sigma_{rr}\operatorname{Col}_r(K)\operatorname{Col}_r(M)^T.$$ □

Definition 11.4.

- The scalars $\sigma_1 \geq \cdots \geq \sigma_r$ are called the *singular values.*
- The ON vectors $u_1, \ldots, u_r$ are called the *right singular vectors.*
- The ON vectors $w_1, \ldots, w_r$ are called the *left singular vectors.*

So far we have seen that SVDs, rSVDs, and Schmidt decompositions are essentially identical. We summarize this in one equation, with matrix sizes as subscripts:

$$A_{m\times n} = K_{m\times m}\,\Sigma_{m\times n}\,M^T_{n\times n} = \tilde{K}_{m\times r}\,\tilde{\Sigma}_{r\times r}\,\tilde{M}^T_{n\times r} = \sum_{i=1}^{r}\sigma_i\,(w_i)_{m\times 1}\,(u_i)^T_{n\times 1}.$$

- K is orthogonal, and $\tilde{K}$ is its first r columns: $w_1,\ldots,w_r$;
- M is orthogonal, and $\tilde{M}$ is its first r columns: $u_1,\ldots,u_r$;
- $\tilde{\Sigma}$ is invertible and diagonal, with the singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$ on the diagonal, and Σ has $\tilde{\Sigma}$ as its upper left block and 0's elsewhere.

Notation. For $p, q \geq r$, it will be convenient to write $\Sigma_{p\times q}$ for the $p \times q$ matrix with $\tilde{\Sigma}$ as its upper left block and 0's elsewhere.

Proposition 11.5. *Suppose that $K\Sigma M^T$ is an SVD of A.*

(i) $A^TA = M\Sigma^2_{n\times n}M^T$ *is an orthogonal diagonalization.*

(ii) $AA^T = K\Sigma^2_{m\times m}K^T$ *is an orthogonal diagonalization.*

(iii) $|A| = M\Sigma_{n\times n}M^T$ *and* $|A^T| = K\Sigma_{m\times m}K^T$ *are orthogonal diagonalizations.*

Proof. First, (iii) follows from (i) and (ii). For (i) and (ii), use $\Sigma = \Sigma_{m\times n}$ to verify

$$\Sigma^T\Sigma = \Sigma^2_{n\times n}, \qquad \Sigma\Sigma^T = \Sigma^2_{m\times m}.$$

Since K and M are orthogonal, this gives

$$A^TA = M\Sigma^T\Sigma M^T = M\Sigma^2_{n\times n}M^T, \qquad AA^T = K\Sigma\Sigma^TK^T = K\Sigma^2_{m\times m}K^T. \quad \square$$

This proposition gives a sense of how to construct SVDs: the columns of M are eigenvectors of A^TA, the columns of K are eigenvectors of AA^T, and the singular values are the square roots of their eigenvalues. Here are the details.

Theorem 11.6. *Every matrix A has SVDs, rSVDs, and Schmidt decompositions.*

Proof. By Propositions 11.3 and 11.4, it is enough to construct any one of the three. The following procedure constructs a Schmidt decomposition.

Step 1. Recall that A is an arbitrary $m \times n$ matrix of rank r. Its modulus $|A|$ is PSDS, and also of rank r. Therefore it admits an ON basis $u_1,\ldots,u_n$ of eigenvectors with non-negative eigenvalues, precisely r of which are positive. Order this basis so its eigenvalues are non-increasing. Then the first r will be positive: write them as

$$\sigma_1 \geq \cdots \geq \sigma_r.$$

Step 2. Define vectors $w_1,\ldots,w_r$ in $\mathbb{R}^m$ by

$$w_i := \tfrac{1}{\sigma_i}Au_i.$$

We claim that they are ON. To see this, use $A^TA = |A|^2$ and $|A|u_i = \sigma_iu_i$:

$$w_i^Tw_i = \tfrac{1}{\sigma_i^2}u_i^T(A^TA)u_i = \tfrac{1}{\sigma_i^2}u_i^T|A|^2u_i = u_i^Tu_i = 1,$$

$$w_j^Tw_i = \tfrac{1}{\sigma_j\sigma_i}u_j^T(A^TA)u_i = \tfrac{1}{\sigma_j\sigma_i}u_j^T|A|^2u_i = \tfrac{\sigma_i}{\sigma_j}u_j^Tu_i = 0 \quad \text{for } i \neq j.$$

Step 3. Define $A' := \sigma_1w_1u_1^T + \cdots + \sigma_rw_ru_r^T$. We need to show that $A' = A$. It will suffice to show that they have the same action on the ON basis $u_1,\ldots,u_n$.

We ordered the u_i so that for $i > r$, $|A|u_i = 0$. Since A has the same null space as $|A|$, $Au_{i>r}$ is also 0. And because the u_i are ON, $A'u_{i>r}$ is 0 too.

For $i \leq r$, every summand of A' except for the i^{th} kills u_i, and because $u_i^T u_i = 1$, the i^{th} summand sends u_i to $\sigma_i w_i$. By the definition of w_i, this is Au_i. □

Remark. SVDs, rSVDs, and Schmidt decompositions are not unique. For example, u_i and w_i can be replaced by $-u_i$ and $-w_i$. We saw an analogous phenomenon when studying diagonalization: eigenvectors are not intrinsic, but eigenspaces are. This points the way to a form of the SVD which does possess uniqueness, which we refer to as the *intrinsic SVD* (iSVD). In it, right and left singular vectors are replaced by right and left "singular spaces". We now briefly describe it, but you may skip this material.

Definition 11.5. Let $\sigma_1 > \cdots > \sigma_d > 0$ be the *distinct* singular values of A. (Warning: this indexing is different from the indexing $\sigma_1 \geq \cdots \geq \sigma_r$ used earlier.)

- Let U_δ be the σ_δ-eigenspace of $|A|$.
- Let W_δ be the σ_δ-eigenspace of $|A^T|$.
- Let ι_δ be $\frac{1}{\sigma_\delta} A|_{U_\delta}$, the restriction of $\frac{1}{\sigma_\delta} A$ to U_δ.

The spaces U_δ and W_δ are the right and left *singular spaces*. The iSVD is the following statement. Recall that P_{U_δ} denotes orthogonal projection from $\mathbb{R}^n$ to U_δ.

Proposition 11.7. (i) $A = \sigma_1 \iota_1 P_{U_1} + \cdots + \sigma_d \iota_d P_{U_d}$.

(ii) *ι_δ is a bijection from U_δ to W_δ, and its inverse is* $\frac{1}{\sigma_\delta} A^T|_{W_\delta}$.

(iii) *ι_δ is an isometry: for all $u, u' \in U_\delta$, $(\iota_\delta u) \cdot (\iota_\delta u') = u \cdot u'$.*

12. The pseudoinverse

Let A be an arbitrary $m \times n$ matrix of rank r.

Definition 12.1. The *pseudo-inverse* A^+ of A may be defined in terms of any rSVD of A, or equivalently, any Schmidt decomposition of A:

$$A^+ := \tilde{M} \tilde{\Sigma}^{-1} \tilde{K}^T = \frac{1}{\sigma_1} u_1 w_1^T + \cdots + \frac{1}{\sigma_r} u_r w_r^T. \tag{12.1}$$

This definition shows you how to compute A^+, and it even gives you the Schmidt decomposition of A^+, but it has two flaws: it does not tell you what A^+ really is, and it does not explain why you get the same A^+ no matter which rSVD you choose. We now rectify this by giving an intrinsic definition of A^+.

Theorem 12.1. *Consider the restriction of A to $\mathcal{R}(A^T)$, written $A|_{\mathcal{R}(A^T)}$.*

(i) $\mathbb{R}^n = \mathcal{R}(A^T) \oplus^\perp \mathcal{N}(A)$.

(ii) $\mathbb{R}^m = \mathcal{R}(A) \oplus^\perp \mathcal{N}(A^T)$.

(iii) *$A|_{\mathcal{R}(A^T)}$ is a bijection from $\mathcal{R}(A^T)$ to $\mathcal{R}(A)$.*

(iv) *$A^T|_{\mathcal{R}(A)}$ is a bijection from $\mathcal{R}(A)$ to $\mathcal{R}(A^T)$.*

Proposition 12.2. *A^+ is the unique $n \times m$ matrix with the following properties:*

(i) $\mathcal{N}(A^+) = \mathcal{N}(A^T)$.

(ii) $\mathcal{R}(A^+) = \mathcal{R}(A^T)$.

(iii) *The restriction $A^+|_{\mathcal{R}(A)}$ is the inverse of $A|_{\mathcal{R}(A^T)}$.*

Corollary 12.3. (i) *A^+A is the projection $P_{\mathcal{R}(A^T)}$.*

(ii) *AA^+ is the projection $P_{\mathcal{R}(A)}$.*

Remark. The following variation of (12.1) can be easier for computations, because it often contains no irrational square roots. Write A_i for the i^{th} summand in the Schmidt decomposition of A, the rank 1 matrix $\sigma_i w_i u_i^T$. Then

$$A^+ = \frac{1}{\sigma_1^2} A_1 + \cdots + \frac{1}{\sigma_r^2} A_r.$$