1

Preliminaries

In this first chapter, we refresh basic concepts in matrix analysis. The purpose is two-fold: (i) to ensure readers are familiar with the concepts relevant to our later exposition and (ii) to standardize notation. We expect readers to be familiar with these concepts; hence we omit all but the most trivial proofs. Readers who desire a more detailed treatment should refer to Part I and Lecture 6 of Trefethen and Bau's excellent textbook [4].

1.1 Euclidean space

1.1.1 Vectors

The basic building blocks are complex numbers (*scalars*), which we denote by italicized Greek and Latin letters, *e.g.* $\alpha, a \in \mathbb{C}$ *etc.*

Using these scalars, we can build column *vectors* of length n, which we denote by lower-case bold-faced letters, *e.g.* $\mathbf{x} \in \mathbb{C}^n$. The jth entry of $\mathbf{x}$ is denoted by $x_j \in \mathbb{C}$. Sometimes, we shall emphasize that a scalar or vector has real entries: $\mathbf{x} \in \mathbb{R}^n, v_i \in \mathbb{R}$.

A set of vectors is *linearly independent* provided the zero vector cannot be expressed as a nontrivial linear combination of the vectors in the set: $\alpha_1\mathbf{x}_1+\cdots+\alpha_n\mathbf{x}_n = 0 \iff \alpha_1 = \cdots = \alpha_n = 0$. Equivalently, no vector in that set can be expressed as a linear combination of the other vectors in the set.

A set $\mathcal{S}$ is called a *subspace* if it is closed under vector addition and scalar multiplication; *i.e.* (i) if $\mathbf{x}, \mathbf{y} \in \mathcal{S}$, then $\mathbf{x} + \mathbf{y} \in \mathcal{S}$ and (ii) if $\mathbf{x} \in \mathcal{S}$, and $\alpha \in \mathbb{C}$, then $\alpha\mathbf{x} \in \mathcal{S}$.

The *span* of a set of vectors is the set of all possible linear combinations of the vectors in the set:

$$\mathrm{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} = \{\alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n \mid \alpha_1, \ldots, \alpha_n \in \mathbb{C}\}.$$

The span of a set of vectors is a subspace.

A basis for a subspace $\mathcal{S}$ is a set of linearly independent vectors that span $\mathcal{S}$. Bases are not unique but every basis for a subspace contains the same number of vectors. We call this number the *dimension* of the subspace $\mathcal{S}$, written $\dim(\mathcal{S})$. If $\mathcal{S} \in \mathbb{C}^n$, then $\dim(\mathcal{S}) \leq n$.

We seek to extend our intuition about geometry naturally to $\mathbb{C}^n$, which implies we seek generalizations of the usual notions of *angle* and *distance* or *length*. First, we define the *dot product* between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ to be

$$\langle \mathbf{x} \cdot \mathbf{y} \rangle := \sum_{i=1}^{n} \bar{x}_i y_i = \mathbf{x}^*\mathbf{y},$$

where $\mathbf{x}^*$ denotes the conjugate-transpose of $\mathbf{x}$:

$$\mathbf{x}^* = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \end{bmatrix} \in \mathbb{C}^{1\times n},$$

a row vector consisting of the complex conjugates of the entries of $\mathbf{x}$. Sometimes we turn a column vector into a row vector without conjugating the entries; we call this the *transpose*: $\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$. Note that $\mathbf{x}^* = \bar{\mathbf{x}}^T$, and if $\mathbf{x} \in \mathbb{R}^n$, then $\mathbf{x}^* = \mathbf{x}^T$.

The dot product provides a notion of magnitude, or *norm*, of a vector $\mathbf{x} \in \mathbb{C}^n$. This is the *Euclidean norm* of $\mathbf{x}$:

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^*\mathbf{x}} = \sqrt{\sum_{i=1}^{n} |x_i|^2}.$$

Here the subscript is used to distinguish this specific norm. More generally, a norm must satisfy $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$.

A norm also satisfies: (i) $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for all $\alpha \in \mathbb{C}$, and (ii) the triangle inequality: $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. We call a vector of norm 1 a *unit vector*.

The dot product satisfies the *Cauchy-Schwarz inequality*:

$$|\mathbf{x}^*\mathbf{y}| \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2, \tag{1.1}$$

a very useful inequality that appears in many mathematical subjects. Equality holds in (1.1) when $\mathbf{y}$ is a scalar multiple of $\mathbf{x}$.

The dot product also provides a notion of acute *angle* between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$:

$$\angle(\mathbf{x}, \mathbf{y}) = \arccos\left(\frac{|\mathbf{x}^*\mathbf{y}|}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}\right).$$

This notion of angle is, in essence, a measure of the sharpness of the Cauchy-Schwarz inequality. The argument to arccos is always between 0 and 1 so the angle is always between 0 and $\pi/2$.

We say the vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* if $\mathbf{x}^*\mathbf{y} = 0$ because $\arccos(0) = \pi/2$. We denote this case using $\mathbf{x} \perp \mathbf{y}$. This definition of angle and norm immediately yields a generalization of the *Pythagorean Theorem* to $\mathbb{C}^n$: If $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, then

$$\|\mathbf{x}+\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2.$$

Two sets $\mathcal{U}$ and $\mathcal{V}$ are *orthogonal* if $\mathbf{u} \perp \mathbf{v}$ for every $\mathbf{u} \in \mathcal{U}$ and $\mathbf{v} \in \mathcal{V}$. The *orthogonal complement* of a set $\mathcal{U} \subseteq \mathbb{C}^n$ is the set of vectors orthogonal to every $\mathbf{u} \in \mathcal{U}$, denoted by

$$\mathcal{U}^\perp := \{\mathbf{v} \in \mathbb{C}^n : \mathbf{u}^*\mathbf{v} = 0 \text{ for all } \mathbf{u} \in \mathcal{U}\}.$$

The *sum* of two subspaces $\mathcal{U}$ and $\mathcal{V}$ is denoted by

$$\mathcal{U} + \mathcal{V} := \{\mathbf{u}+\mathbf{v} : \mathbf{u} \in \mathcal{U}, \mathbf{v} \in \mathcal{V}\}.$$

If the two subspaces intersect trivially ($\mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}$), then we call the sum of $\mathcal{U}$ and $\mathcal{V}$ a *direct sum*, denoted by $\mathcal{U} \oplus \mathcal{V}$. We use this special notation because in the case of a direct sum, every vector $\mathbf{x} \in \mathcal{U} \oplus \mathcal{V}$ can be decomposed *uniquely* as $\mathbf{x} = \mathbf{u}+\mathbf{v}$ for some $\mathbf{u} \in \mathcal{U}$ and $\mathbf{v} \in \mathcal{V}$.

1.1.2 Matrices

Using scalars, we can also build matrices with m rows and n columns, which we denote by a bold capital letter, *e.g.* $\mathbf{A} \in \mathbb{C}^{m \times n}$. The entry in the jth row and kth column is denoted by a_{jk}. We often split a matrix into columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ or rows $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T$. For example, we can represent $\mathbf{A} \in \mathbb{C}^{3 \times 2}$ as

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \text{ or } \mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \mathbf{a}_3^T \end{bmatrix}.$$

The conjugate-transpose and transpose of a matrix are defined and denoted similarly to the respective operations for a vector: the j, k entry of $\mathbf{A}^*$ is $\bar{a}_{kj}$ and then j, k entry of $\mathbf{A}^T$ is a_{kj}. For our 3×2 example matrix $\mathbf{A}$,

$$\mathbf{A}^* = \begin{bmatrix} \bar{a}_{11} & \bar{a}_{21} & \bar{a}_{31} \\ \bar{a}_{12} & \bar{a}_{22} & \bar{a}_{32} \end{bmatrix}$$
$$\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{23} \end{bmatrix}.$$

Matrices with the same number of rows and columns are called *square* matrices.

We can verify using a direct calculation that

$$(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$$
$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T.$$

We often add and multiply matrices and vectors block-wise in the same way we carry out matrix multiplication entry-wise. For example, the Karush-Kuhn-Tucker (KKT) conditions for some classes of optimization problems can be expressed as

$$\mathbf{Hx} + \mathbf{A}^T\mathbf{y} = -\mathbf{g}$$
$$\mathbf{Ax} = \mathbf{b},$$

which can also be expressed as

$$\begin{bmatrix} \mathbf{H} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} -\mathbf{g} \\ \mathbf{b} \end{bmatrix}.$$

We can also generalize the dot product to handle $m \times n$ matrices:

$$\langle \mathbf{A} \cdot \mathbf{B} \rangle = \mathrm{Tr}(\mathbf{A}^*\mathbf{B})$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a square matrix; *i.e.* the sum of its diagonal entries. The trace is a linear operator: it satisfies (i) $\mathrm{Tr}(\mathbf{A} + \mathbf{B}) = \mathrm{Tr}(\mathbf{A}) + \mathrm{Tr}(\mathbf{B})$ and (ii) $\mathrm{Tr}(\alpha\mathbf{A}) = \alpha\mathrm{Tr}(\mathbf{A})$. The trace of a product is also invariant with respect to cyclic permutations of the product; *i.e.*

$$\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB}) = \mathrm{Tr}(\mathbf{BCA}).$$

Like the dot product, the trace product also provides a norm. This norm is called the Frobenius norm:

$$\|\mathbf{A}\|_F := \sqrt{\mathrm{Tr}(\mathbf{A}^*\mathbf{A})} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}.$$

The trace product and the Frobenius norm also satisfy the Cauchy-Schwarz inequality:

$$|\mathrm{Tr}(\mathbf{A}^*\mathbf{B})| \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F.$$

If $\mathbf{A}$ is a scalar multiple of $\mathbf{B}$, then equality is attained.

In fact, we can use any vector norm, for example the aforementioned euclidean norm, to measure magnitude of a matrix by the largest amount the matrix stretches vectors. We define the *induced matrix norm* of a matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ as

$$\|\mathbf{A}\| := \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|},$$

which may be equivalently expressed as

$$\|\mathbf{A}\| := \max_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\|.$$

There is no simple formula for this norm and we shall describe how to compute it after developing the requisite machinery.

These matrix norms (and many others) satisfy the same properties the Euclidean norm satisfies: (i) $\|\mathbf{A}\| \geq 0$ and $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = \mathbf{0}$, (ii) $\|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|$ for $\alpha \in \mathbb{C}$, and (iii) $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$.

Both these norms are also *submultiplicative*; *i.e.* if $\mathbf{A} \in \mathbb{C}^{m\times n}$ and $\mathbf{x} \in \mathbb{C}^n$, then

$$\|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\|\|\mathbf{x}\|.$$

The result also holds if we replace $\mathbf{x}$ by a matrix $\mathbf{B} \in \mathbb{C}^{n\times p}$.

We say a matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ is *diagonal* if all its off-diagonal entries are zero ($a_{ij} = 0$ if $i \neq j$). We say $\mathbf{A}$ is *upper triangular* if all its entries below the main diagonal are zero ($a_{ij} = 0$ if $i > j$) and *lower triangular* if all its entries above the main diagonal are zero.

A square matrix is *Hermitian* if $\mathbf{A}^* = \mathbf{A}$ and *symmetric* if $\mathbf{A}^T = \mathbf{A}$. In the case of real matrices, these notions are the same and we call such matrices symmetric. In the case of complex matrices, Hermitian matrices are more common and [thankfully] more useful than symmetric matrices. We can also define *skew-Hermitian* and *skew-symmetric* martices that satisfy $\mathbf{A}^* = -\mathbf{A}$ and $\mathbf{A}^T = -\mathbf{A}$ respectively.

A square matrix is *unitary* if $\mathbf{U}^*\mathbf{U} = \mathbf{I}$. We say the columns of $\mathbf{U}$ are *orthonormal*. Since $\mathbf{U}$ is a square matrix (it has n orthonormal columns, each in $\mathbb{C}^n$), the columns of $\mathbf{U}$ are an *orthonormal basis* for $\mathbb{C}^n$ and such bases are useful because we can express a vector $\mathbf{x}$ as

$$\mathbf{x} = (\mathbf{x}^*\mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{x}^*\mathbf{u}_n)\mathbf{u}_n.$$

We also often encounter matrices $\mathbf{V} \in \mathbb{C}^{n\times k}, n > k$ with orthonormal columns. Such matrices satisfy $\mathbf{V}^*\mathbf{V} = \mathbf{I} \in \mathbb{C}^{k\times k}$, however, $\mathbf{V}\mathbf{V}^* \neq \mathbf{I} \in \mathbb{C}^{n\times n}$. There is no universally accepted term for such matrices although a commonly used term is *subunitary*. Premultiplying a vector by such a matrix preserves the norm because

$$\|\mathbf{V}\mathbf{x}\|_2 = \sqrt{(\mathbf{V}\mathbf{x})^*(\mathbf{V}\mathbf{x})} = \sqrt{\mathbf{x}^*\mathbf{V}^*\mathbf{V}\mathbf{x}} = \sqrt{\mathbf{x}^*\mathbf{x}} = \|\mathbf{x}\|_2.$$

The result also holds if we replace $\mathbf{x}$ by a matrix $\mathbf{A}$ and the Euclidean norm with either the induced matrix norm or the Frobenius norm :

$$\|\mathbf{U}\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{U}\mathbf{A}\mathbf{x}\|_2 = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{A}\|_2.$$

The intuition behind these results is that the product $\mathbf{U}\mathbf{x}$ is just a representation of $\mathbf{x}$ in a new orthonormal basis and such a change of basis should not affect the length/magnitude of $\mathbf{x}/\mathbf{A}$. These results are known as the *invariance* of the Euclidean/induced matrix/Frobenius norm under unitary transformations.

1.2 The four fundamental subspaces

The *range* or *column space* of a matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$, denoted by $\mathcal{R}(\mathbf{A})$, is given by

$$\mathcal{R}(\mathbf{A}) := \{\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^n\} \subseteq \mathbb{C}^m.$$

The *null space* or *kernel* of $\mathbf{A}$, denoted by $\mathcal{N}(\mathbf{A})$, is given by

$$\mathcal{N}(\mathbf{A}) := \{\mathbf{z} \in \mathbb{C}^n \mid \mathbf{A}\mathbf{z} = \mathbf{0}\} \subseteq \mathbb{C}^n.$$

The range and nullspace of $\mathbf{A}$ and $\mathbf{A}^*$ are referred to collectively as the four *fundamental subspaces* of $\mathbf{A}$.

The *column rank* of $\mathbf{A}$ is the dimension of $\mathcal{R}(\mathbf{A})$; *i.e.* the number of linearly independent vectors in a basis for $\mathcal{R}(\mathbf{A})$. The *row rank* of $\mathbf{A}$ is the dimension of $\mathcal{R}(\mathbf{A}^*)$. The row rank of a matrix is the same as its column rank so we refer to this quantity as simply the *rank* of a matrix.

The range and null space of $\mathbf{A}$ are subspaces of $\mathbb{C}^m$ and $\mathbb{C}^n$ respectively, and the span of a set of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{C}^m$ is the same as the range of the matrix whose columns are $\mathbf{x}_1, \dots, \mathbf{x}_n$:

$$\operatorname{span}\{\mathbf{x}_1, \dots, \mathbf{x}_n\} = \mathcal{R}\left(\begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}\right).$$

The set of vectors are linearly independent if

$$\mathcal{N}\left(\begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}\right) = \{\mathbf{0}\}.$$

The four fundamental subspaces of $\mathbf{A}$ satisfy what Gilbert Strang calls *the Fundamental Theorem of Linear Algebra.*

Theorem 1.1. If $\mathbf{A} \in \mathbb{C}^{m\times n}$, its four fundamental subspaces satisfy

$$\mathcal{R}(\mathbf{A}) \perp \mathcal{N}(\mathbf{A}^*) \text{ and } \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^*) = \mathbb{C}^m,$$
$$\mathcal{R}(\mathbf{A}^*) \perp \mathcal{N}(\mathbf{A}) \text{ and } \mathcal{R}(\mathbf{A}^*) \oplus \mathcal{N}(\mathbf{A}) = \mathbb{C}^n.$$

Thus, a vector $\mathbf{x} \in \mathbb{C}^m$ can be expressed uniquely as $\mathbf{x} = \mathbf{x}_R + \mathbf{x}_N$, where $\mathbf{x}_R \in \mathcal{R}(\mathbf{A})$ and $\mathbf{x}_N \in \mathcal{N}(\mathbf{A}^*)$ are orthogonal.

A square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is *nonsingular* or *invertible* if and only if the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x}$ for any $\mathbf{b} \in$

$\mathbb{C}^n$. We express the unique solution to such linear systems as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. Equivalently, $\mathbf{A}$ is nonsingular if $\mathcal{R}(\mathbf{A}) = \mathbb{C}^n$ or $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$. A square matrix that is not nonsingular is called *singular* and rectangular matrices (matrices that are not square) are always singular.

If $\mathbf{A}$ is nonsingular, then its inverse is unique. Thus, for a nonsingular matrix $\mathbf{A}$, $(\mathbf{A}^*)^{-1} = (\mathbf{A}^{-1})^*$ and if both $\mathbf{A}$ and $\mathbf{B}$ are nonsingular, then their product is also nonsingular:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

1.3 Projectors

A square matrix $\mathbf{P}$ is called a *projector* if $\mathbf{P}^2 = \mathbf{P}$. The name projector hints at the action of these matrices. Imagine shining a lamp onto the subspace $\mathcal{R}(\mathbf{P})$ from some direction. The shadow cast by a vector $\mathbf{v}$ represents $\mathbf{Pv}$, the projection of $\mathbf{v}$ onto $\mathcal{R}(\mathbf{P})$.

Our intuition about projectors says if $\mathbf{v} \in \mathcal{R}(\mathbf{P})$, then it should lie exactly on its shadow so $\mathbf{Pv} = \mathbf{v}$. The mathematical notion of a projector agrees with this intuiton: if $\mathbf{v} \in \mathcal{R}(\mathbf{P})$, then there exists some $\mathbf{x}$ such that $\mathbf{v} = \mathbf{Px}$ and

$$\mathbf{Pv} = \mathbf{P}^2\mathbf{x} = \mathbf{Px} = \mathbf{v}.$$

If $\mathbf{P}$ is a projector, then $\mathbf{I} - \mathbf{P}$ is also a projector and $\mathcal{R}((\mathbf{I} - \mathbf{P})) = \mathcal{N}(\mathbf{P}))$. Further, $\mathcal{R}(\mathbf{P}) \cap \mathcal{N}(\mathbf{P}) = \{\mathbf{0}\}$, so a projector partitions $\mathbb{C}^n$ into a direct sum of two subspaces. We say such a pair of subspaces are *complementary subspaces* and the pair of projectors are *complementary projectors.*

If the projector $\mathbf{P}$ is Hermitian, the fundamental theorem of linear algebra implies $\mathcal{R}(\mathbf{P}) \perp \mathcal{N}(\mathbf{P})$. In this case we call $\mathbf{P}$ an *orthogonal projector* because the space $\mathbf{P}$ projects onto is orthogonal to the space it projects along. The orthogonal projection of a vector $\mathbf{v}$ satisfies

$$\min_{\mathbf{w} \in \mathcal{R}(\mathbf{P})} \|\mathbf{v} - \mathbf{w}\| = \|\mathbf{v} - \mathbf{Pv}\|.$$

In general, the induced norm of a projector satisfies $\|\mathbf{P}\| \geq 1$ because $\|\mathbf{Pv}\| = \|\mathbf{v}\|$ for every $\mathbf{v} \in \mathcal{R}(\mathbf{P})$. If $\mathbf{P}$ is an orthogonal projector, then $\|\mathbf{P}\| = 1$ because $\mathbf{P}$ decomposes $\mathbf{v}$ into $\mathbf{Pv} + (\mathbf{I} - \mathbf{P})\mathbf{v}$, where $\mathbf{Pv}$ and $(\mathbf{I} - \mathbf{P})\mathbf{v}$ are orthogonal. The converse is also true.

Theorem 1.2. A projector $\mathbf{P}$ is orthogonal if and only if $\|\mathbf{P}\| = 1$.

We can construct an orthogonal projector onto a subspace $\mathcal{S}$ using an orthonormal basis for $\mathcal{S}$. Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be an orthonormal basis for $\mathcal{S}$ and $\mathbf{U}$ be the matrix whose jth column is $\mathbf{u}_i$. Then the matrix

$$\mathbf{P}_{\mathcal{S}} = \mathbf{U}\mathbf{U}^*$$

is an orthogonal projector onto $\mathcal{S}$. Sometimes, $\mathcal{S}$ is one-dimensional; *i.e.* $\mathcal{S} = \operatorname{span}\{\mathbf{v}\}$. In this case, we use a special case of this formula:

$$\mathbf{P}_{\mathcal{S}} = \frac{\mathbf{v}\mathbf{v}^*}{\mathbf{v}^*\mathbf{v}}.$$

We can also construct orthogonal projectors onto a subspace from non-orthonormal bases. In practice, this is almost always done by first orthogonalizing the basis and then forming the projector by the preceding construction.