This primer is designed to provide an overview of many basic concepts and properties of linear matrix algebra as briefly as possible; i.e. proofs are left to the reader.

Definitions and algebra

A matrix is a rectangular array of numbers (real or complex) arranged in horizontal rows and vertical columns. We shall denote matrices by boldface capital letters, such as $A$, and the entries (elements) of the matrix by lowercase letter with subscripts, such as $a_{ij}$, where the first index ($i$) increments down rows, and the second index ($j$) increments out columns; we may also refer to $A$ as $[a_{ij}]$. The entries for $i = j$ are called the main diagonal, and matrices having only non-zero entries on the main diagonal are called diagonal. Summing the diagonal entries of a matrix gives the trace: $tr(A) = \sum_{i=1}^{n} a_{ii}$.

The order (or size) of the matrix is $m \times n$ (“$m$ by $n$”), where $m$ is the number of rows, and $n$ is the number of columns; the shorthand notation $A \sim m \times n$ will be used. Matrices with the same number of rows and columns are square. A special square matrix called the identity matrix ($I$) is a square, diagonal, matrix having only ones along the main diagonal. Whenever “0” appears it is understood that this is a matrix of the appropriate size having all entries equal to zero.

Vectors are matrices with only one row or column; here vectors will be denoted by bold lowercase letters (e.g. $x$). A column vector with $m$ elements has order $m \times 1$ and a row vector with $n$ elements has order $1 \times n$. The transpose of $A$, $A^T$, has rows and columns interchanged; i.e., the first row becomes the first column, the second row becomes the second column, etc. For complex matrices, the Hermitian transpose of $A$, $A^H$ is obtained by taking the transpose and complex conjugate of each matrix element. In the case where only complex conjugation is required (no transpose), an overbar will be used (e.g. $\bar{A}$)
Addition and subtraction

Matrix addition and subtraction proceed element-wise: $A + B = [a_{ij} + b_{ij}]$.

These operations are both associative: $(A + B) + C = A + (B + C)$
and commutative: $A + B = B + A$

Matrix multiplication

Given matrices $A \sim m \times n$ and $B \sim n \times p$, their product $AB$ gives a matrix $C \sim m \times p$, whose elements are defined by

$$c_{ij} = \sum_k a_{ik} b_{kj},$$

that is, a sum over element-by-element multiplication of the $i^{th}$ row of $A$ by the $j^{th}$ column of $B$.

Matrix multiplication is associative: $(AB)C = A(BC)$

and distributive: $A(B + C) = AB + AC$

but, in general, not commutative: $AB \neq BA$

transposition: $(AB)^T = B^T A^T$

Multiplication of a matrix by a scalar simply multiplies every element of the matrix by the scalar, and is commutative ($kA = Ak$).

Matrix inverse

$B$ is an inverse of a square matrix $A$ if $AB = BA = I$. A matrix that has no inverse is called singular, and one that does is called nonsingular. The inverse of a non-singular matrix is unique.

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

$$(A^T)^{-1} = (A^{-1})^T$$

Determinants
The determinant of a square matrix is a scalar. For a $2 \times 2$ matrix, \( \det(A) = |A| \), is
\[
\begin{vmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{4}
\]

In general, determinants of $n \times n$ square matrices larger than $2 \times 2$ may be obtained by reduction to the $2 \times 2$ form through
\[
\det(A) = \sum_{j=1}^{n} (-1)^{j+1}a_{1j} \det(A_{1j}). \tag{5}
\]
Here $A_{1j}$ is the $(n-1) \times (n-1)$ matrix that results from removing the first row and $j^{th}$ column from $A$.

Properties:
- If $\det(A) \neq 0$, then $A$ is nonsingular.
- If any row or column of $A$ is zero, then $\det(A) = 0$.
- If two rows of $A$ are equal, then $\det(A) = 0$.
- $\det(AB) = \det(A) \det(B)$.
- $\det(A^T) = \det(A)$.
- $\det(A^{-1}) = \det(A)^{-1}$, provided $A^{-1}$ exists.
- $\det(cA) = c^n \det(A)$ for scalar $c$.

Eigenvalues and eigenvalues

A nonzero column vector, $x$, is a right eigenvector of $A$, with eigenvalue, $\lambda$, if
\[
Ax = \lambda x. \tag{6}
\]
Rearranging (6) gives $(A - \lambda I)x = 0$, which has a unique non-trivial solution for $x$ provided that $\det(A - \lambda I) = 0$. Expanding the determinant gives the characteristic polynomial of $A$,
\[
\det(A - \lambda I) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_1 \lambda + a_0 = 0. \tag{7}
\]

The **Cayley-Hamilton Theorem** says that every matrix satisfies its own characteristic equation:
\[
a_n A^n + a_{n-1} A^{n-1} + \cdots + a_1 A + a_0 = 0. \tag{8}
\]
Left eigenvectors are defined by $y^T A = \lambda y^T$. Note that the left eigenvectors of $A$ are the right eigenvectors of $A^T$; therefore these vectors are also called the adjoint eigenvectors.

Properties:
- $tr(A) = \sum_{i=1}^n \lambda_i$.
- $det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$.
- Eigenvectors associated with unique eigenvalues are linearly independent.
- A matrix is singular if it has a zero eigenvalue.
- If $x$ and $\lambda$ are an eigenvector–eigenvalue pair for $A$, then $x$ and $\lambda^{-1}$ are an eigenvector–eigenvalue pair for $A^{-1}$.
- The eigenvalues of $A^T$ are the same as for $A$.
  - Therefore there are left and right eigenvectors of $A$ that share the same eigenvalue.
- Left and right eigenvectors are orthogonal: $y_i^T x_j = 0$, for $i \neq j$.

Matrices $A$ and $B$ are similar if there is an invertible matrix, $X$ so that

$$A = X B X^{-1}. \quad (9)$$

An important special case involves diagonal $B$, which means that $X$ diagonalizes $A$. In this case, $X$ is the eigenvector matrix of $A$, which has as column vectors the right eigenvectors of $A$. By the properties outlined above, this means that $X^{-1}$ is the transpose of the matrix of adjoint eigenvectors, $Y^T$, where the columns of $Y$ are ordered so that the $i^{th}$ column of $X$ and $Y$ share the same eigenvalue. Then $B$ has the eigenvalues of $A$ along the main diagonal, ordered so that they correspond to the appropriate columns of $X$. This shows that if $A$ has a zero eigenvalue, then it must be singular.

**Functions of matrices** It turns out that the Cayley-Hamilton Theorem can be used to evaluate arbitrary functions of matrices. The result is

$$f(A) = a_n A^n + a_{n-1} A^{n-1} + \cdots + a_1 A + a_0 = 0. \quad (10)$$

The coefficients $a_i$ may be determined by:

(i) In (10), replace the lefthand side by $r(\lambda)$ and on the righthand side $A^j$ by $\lambda^j$. There is one such equation for each eigenvalue.
(ii) Solve the $n$ equations for the $n$ unknowns ($a_j$).
(iii) Substitute the $a_j$ into (10).
Note that if $A$ has repeat eigenvalues, additional steps must be taken to obtain a unique solution.

A function that comes up frequently is $\exp(A \cdot t)$, which may be evaluated from the similarity transformation of $A$ using its eigenvectors, $E$, and diagonal matrix of eigenvalues, $\Lambda$:

$$A t = E \Lambda E^{-1} t. \quad (11)$$

The matrix exponential may then be evaluated from

$$\exp(A t) = E T E^{-1}, \quad (12)$$

where $T$ is a diagonal matrix with diagonal entries given by $t_{ii} = e^\lambda_i t$.

A **matrix square root** is a matrix, $X$, such that

$$XX = A. \quad (13)$$

If $A$ and $X$ are positive semidefinite (see below) then $X$ is unique. Further insight may be gained by noting that any positive definite matrix may be expressed as $A = LL^H$ (Cholesky factorization), and any matrix may be expanded in a singular value decomposition (SVD; see below); i.e. $L = USV^T$. Direct substitution gives the square root of $A$: $X = USU^T$.

**Differentiation** and **integration** of matrices proceeds by operating on each element of the matrix individually.

**Vector Spaces**

Vector spaces are defined by sets of vectors that obey the basic algebraic operations described in the first few pages of this primer (addition, subtraction, multiplication, etc.). A **subspace** is some subset of a larger set of vectors that is also a vector space. A set of $n$ vectors $x_i = \{x_1, \ldots, x_n\}$ is **linearly independent** if the only coefficients that satisfy the equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \quad (14)$$

are $a_1 = \cdots a_n = 0$; i.e. there is no way to scale and add the vectors to get zero. Note that if a set of vectors is linearly independent, then any subset of those vectors is too.
The **span** of a set of vectors is the set of all possible linear combinations of these vectors, and this is denoted by \( \text{span}\{x_1, \ldots, x_n\} \); we often refer to this set of linear combinations as “the space spanned by” the vectors \( \{x_i\} \). For example, in a three-dimensional Cartesian coordinate system, the unit vectors along each coordinate axis span the entire space—every point in the space is “reachable” by scaling and adding these vectors. Such **basis vectors** may be defined for any set of vectors space by the smallest subset of vectors that are linearly independent and span the space of the full set. The **dimension** of the space is then given by the number of basis vectors.

The **range** of an \( m \times n \) matrix \( A \), \( \text{ran}(A) \), is defined by all vectors \( y \) such that

\[
A \mathbf{x} = \mathbf{y},
\]

for every vector \( \mathbf{x} \) in a vector space. This **mapping** is **one-to-one** if for every \( \mathbf{x} \) there is a unique \( \mathbf{y} \). If the column vectors of \( A \) are written as the set \( \{x_1, \ldots, x_n\} \), then \( \text{ran}(A) = \text{span}\{x_1, \ldots, x_n\} \).

Similarly, the **null space** of \( A \), \( \text{null}(A) \), is defined by

\[
A \mathbf{x} = \mathbf{0}.
\]

The **rank** of a matrix is defined by

\[
\text{rank}(A) = \text{dim} \left( \text{ran}(A) \right).
\]

Note that \( \text{rank}(A^T) = \text{rank}(A) \), and \( \text{dim} \left( \text{null}(A) \right) + \text{dim} \left( \text{ran}(A) \right) = n \). \( A \) is **rank deficient** if \( \text{rank}(A) < \min\{m, n\} \).

**Inner products and norms**

The **inner product** between \( n \times 1 \) column vectors \( \mathbf{x} \) and \( \mathbf{y} \) is

\[
r = \langle \mathbf{x}, \mathbf{y} \rangle_W = \mathbf{x}^H W \mathbf{y}.
\]

Here, \( W \) is the \( n \times n \) Hermitian matrix that defines the inner product; the result, \( r \), is scalar. This is the vector dot product, which gives the fraction of \( \mathbf{x} \) in the direction of \( \mathbf{y} \). If \( r = 0 \) the vectors are **orthogonal**. Note that **orthogonality depends on the chosen inner**
product, as defined by $W$. When $W = I$, the norm is called the Euclidean (or Frobenius) norm, which is the sum of the squared values; if the mean value of the entries in the column vectors is zero, then $r$ gives the variance in the chosen coordinates. Note that we may always re-express the inner product in term of the Euclidean norm by changing coordinates through the linear transformation

$$ x' = W^{1/2}x, \quad y' = W^{1/2}y. $$

(19)

This point is crucial: a choice of inner product is equivalent to a choice of coordinates.

Properties:

• $\langle x, x \rangle_W > 0$ for $x \neq 0$. The inner product is only zero for $x = 0$.
• $\langle x, y \rangle_W = \overline{\langle y, x \rangle_W}$.
• Distributive rule: $\langle x + y, z \rangle_W = \langle x, z \rangle_W + \langle y, z \rangle_W$.
• **Schwarz inequality**: $|\langle x, y \rangle_W|^2 \leq |\langle x, x \rangle_W|^2 |\langle y, y \rangle_W|^2$.

Norms are used to measure the “size” of vectors and matrices, and are denoted by $\| : \|$.

They must satisfy the following properties:

(i) $\|X\| \geq 0$.
(ii) $\|X\| = 0$ iff $X = 0$.
(iii) $\|cX\| = |c|\|X\|$ for any scalar $c$.
(iv) $\|X + Y\| \leq \|X\| + \|Y\|$
(v) $\|XY\| \leq \|X\|\|Y\|$

Inner products are often used to define norms, and the most common is the Euclidean norm, based on the Euclidean inner product:

$$ \|x\| = [(x, x)]^{1/2} = [x^T x]^{1/2}. $$

(20)

Special matrices

The adjoint of an $m \times n$ matrix $A$ is an $n \times m$ matrix $A^*$ defined by

$$ \langle x, Ay \rangle_W = \langle A^* x, y \rangle_W, $$

(21)

for all $m \times 1$ vectors $x$ and all $n \times 1$ vectors $y$. The adjoint matrix is given by

$$ A^* = W^{-1} A^H W. $$

(22)
In the limiting case of $W = I$, $A^* = A^H$. A matrix is self adjoint if $A^* = A$; $A$ must be square.

A Hermitian matrix satisfies: $A = A^H$; real matrices that satisfy $A = A^T$ are symmetric. These matrices have real eigenvalues and orthogonal eigenvectors.

A normal matrix satisfies: $AA^H = A^H A$. These matrices also have orthogonal eigenvectors.

A unitary matrix satisfies: $A^{-1} = A^H$. These matrices are also normal. The columns of a unitary matrix form an orthonormal basis. The eigenvalues of a unitary matrix satisfy $|\lambda| > 1$.

An orthogonal matrix is a real matrix that satisfies: $A^{-1} = A^T$. Orthogonal matrices are also unitary.

A positive semidefinite matrix is a Hermitian matrix that satisfies: $\langle Ax, y \rangle \geq 0$; if the inequality is greater than zero, the matrix is positive definite.

Singular value decomposition (SVD) and the generalized inverse

Any real $m \times n$ matrix may be decomposed as (SVD):

$$ A = U S V^T, \quad (23) $$

where $U(\sim m \times n)$ and $V(\sim n \times n)$ are orthogonal matrices, and $S(\sim n \times n)$ is a diagonal matrix, with values $\sigma_i$ along the main diagonal; the $\sigma_i$ are called the singular values of $A$.

Let $u_i$ and $v_i$ be the column vectors of $U$ and $V$, respectively. $u_i$ and $v_i$ are called the left and right singular vectors of $A$, respectively, and they satisfy

$$ Av_i = \sigma_i u_i \quad A^T u_i = \sigma_i v_i, \quad (24) $$

If $A$ has $r$ non-zero singular values ($\sigma$) then:

- $\text{rank}(A) = r$.
- $\text{null}(A) = \text{span}\{v_{r+1}, \ldots, v_n\}$.
- $\text{ran}(A) = \text{span}\{u_1, \ldots, u_r\}$.
\[ \mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \]
\[ \|\mathbf{A}\| = \sum_{i=1}^{r} \sigma_i^2. \]

So if \( \mathbf{A} \) is \( n \times n \) square with \( n \) non-zero singular values, then \( \mathbf{A} \) has no null space.

Note also that (23) leads directly to the following eigenvector problems:

\[ (\mathbf{A}^T \mathbf{A}) \mathbf{V} = \mathbf{V} \mathbf{S}^2. \]
\[ (\mathbf{A} \mathbf{A}^T) \mathbf{U} = \mathbf{U} \mathbf{S}^2. \]

Note that in both cases the matrix in parenthesis is symmetric, and therefore has real eigenvalues and orthogonal eigenvectors.

One very important application of SVD involves the inversion of a matrix. In particular, it generalizes the concept of a matrix inverse beyond nonsingular square matrices to non-square matrices, and even singular square matrices! (23) leads directly to the generalized inverse (pseudoinverse) for any matrix:

\[ \mathbf{A}^{-1} = \mathbf{V} \tilde{\mathbf{S}} \mathbf{U}^T, \quad (25) \]

where \( \tilde{\mathbf{S}} \) is a diagonal \( m \times n \) matrix with the first \( r \) values along the main diagonal given by the inverse of the non-zero singular values \( \sigma_i^{-1} \); the other values zero.