Linear Algebra Primer
Gregory J. Hakim
University of Washington
7 January 2005 v1.1

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 $\mathbf{A}$, 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 $\mathbf{A}$ 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**: $tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$.

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 $\mathbf{A} \sim m \times n$ will be used. Matrices with the same number of rows and columns are **square**. A special square matrix called the **identity matrix** ($\mathbf{I}$) 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. $\mathbf{x}$). 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 $\mathbf{A}$, $\mathbf{A}^T$, 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 $\mathbf{A}$, $\mathbf{A}^H$ 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. $\bar{\mathbf{A}}$)

Addition and subtraction

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

These operations are both associative: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$

and commutative: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

Matrix multiplication

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

$$c_{ij} = \sum_{k}^{n} a_{ik}\, b_{kj}, \tag{1}$$

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

Matrix multiplication is associative: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$

and distributive: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

but, in general, **not** commutative: $\mathbf{AB} \neq \mathbf{BA}$

transposition: $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

Multiplication of a matrix by a scalar simply multiplies every element of the matrix by the scalar, and is commutative ($k\mathbf{A} = \mathbf{A}k$).

Matrix inverse

$\mathbf{B}$ is an inverse of a square matrix $\mathbf{A}$ if $\mathbf{AB} = \mathbf{BA} = I$. 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.

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{2}$$

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T \tag{3}$$

Determinants

The determinant of a square matrix is a scalar. For a 2×2 matrix, $det(\mathbf{A}) = \mid \mathbf{A} \mid$, 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 2×2 may be obtained by reduction to the 2×2 form through

$$det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j}\, det(\mathbf{A}_{1j}). \tag{5}$$

Here $\mathbf{A}_{1j}$ is the $(n-1) \times (n-1)$ matrix that results from removing the first row and j^{th} column from $\mathbf{A}$.

Properties:

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

Eigenvectors and eigenvalues

A nonzero column vector, $\mathbf{x}$, is a right eigenvector of $\mathbf{A}$, with eigenvalue, λ, if

$$\mathbf{Ax} = \lambda \mathbf{x}. \tag{6}$$

Rearranging (6) gives $(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = 0$, which has a unique non-trivial solution for $\mathbf{x}$ provided that $det(\mathbf{A} - \lambda \mathbf{I}) = 0$. Expanding the determinant gives the **characteristic polynomial** of $\mathbf{A}$,

$$det(\mathbf{A} - \lambda \mathbf{I}) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_1 \lambda + a_0 = 0. \tag{7}$$

The **Cayley-Hamilton Theorem** says that every matrix satisfies its own characteristic equation:

$$a_n \mathbf{A}^n + a_{n-1} \mathbf{A}^{n-1} + \cdots + a_1 \mathbf{A} + a_0 = 0. \tag{8}$$

Left eigenvectors are defined by $\mathbf{y}^T\mathbf{A} = \lambda\mathbf{y}^T$. Note that the left eigenvectors of $\mathbf{A}$ are the right eigenvectors of $\mathbf{A}^T$; therefore these vectors are also called the **adjoint** eigenvectors.

Properties:

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

Matrices $\mathbf{A}$ and $\mathbf{B}$ are **similar** if there is an invertible matrix, $\mathbf{X}$ so that

$$\mathbf{A} = \mathbf{X}\,\mathbf{B}\,\mathbf{X}^{-1}. \tag{9}$$

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

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

The coefficients a_i may be determined by:

(i) In (10), replace the lefthand side by $r(\lambda)$ and on the righthand side $\mathbf{A}^j$ by λ^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 $\mathbf{A}$ has repeat eigenvalues, additional steps must be taken to obtain a unique solution.

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

$$\mathbf{A}t \;=\; \mathbf{E}\,\mathbf{\Lambda}\,\mathbf{E}^{-1}\,t. \tag{11}$$

The matrix exponential may then be evaluated from

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

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

A **matrix square root** is a matrix, $\mathbf{X}$, such that

$$\mathbf{X}\,\mathbf{X} \;=\; \mathbf{A}. \tag{13}$$

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

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 $\mathbf{x}_i = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is **linearly independent** if the only coefficients that satisfy the equation

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n \;=\; 0 \tag{14}$$

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 $span\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$; we often refer to this set of linear combinations as "the space spanned by" the vectors $\{\mathbf{x}_i\}$. 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 $\mathbf{A}$, $ran(\mathbf{A})$, is defined by all vectors $\mathbf{y}$ such that

$$\mathbf{A}\,\mathbf{x} \,=\, \mathbf{y}, \tag{15}$$

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 $\mathbf{A}$ are written as the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, then $ran(\mathbf{A}) = span\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$.

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

$$\mathbf{A}\,\mathbf{x} \,=\, 0. \tag{16}$$

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

$$rank(A) \,=\, dim(ran(\mathbf{A})). \tag{17}$$

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

Inner products and norms

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

$$r = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{W}} \,=\, \mathbf{x}^H \mathbf{W} \mathbf{y}. \tag{18}$$

Here, $\mathbf{W}$ 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 $\mathbf{W}$! When $\mathbf{W} = I$, 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**

$$\mathbf{x}' \;=\; \mathbf{W}^{1/2}\mathbf{x}, \qquad\qquad \mathbf{y}' \;=\; \mathbf{W}^{1/2}\mathbf{y}. \tag{19}$$

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

Properties:

- $\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{W}} > 0$ for $\mathbf{x} \neq 0$. The inner product is only zero for $\mathbf{x} = 0$.
- $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{W}} \;=\; \overline{\langle \mathbf{y}, \mathbf{x} \rangle}_{\mathbf{W}}$.
- Distributive rule: $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle_{\mathbf{W}} \;=\; \langle \mathbf{x}, \mathbf{z} \rangle_{\mathbf{W}} \;+\; \langle \mathbf{y}, \mathbf{z} \rangle_{\mathbf{W}}$.
- **Schwarz inequality**: $|\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{W}}|^2 \;\leq |\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{W}}|^2 \, |\langle \mathbf{y}, \mathbf{y} \rangle_{\mathbf{W}}|^2$.

Norms are used to measure the "size" of vectors and matrices, and are denoted by $\| \cdot \|$. They must satisfy the the following properties:
(i) $\|\mathbf{X}\| \geq 0$.
(ii) $\|\mathbf{X}\| = 0$ iff $\mathbf{X} = 0$.
(iii) $\|c\mathbf{X}\| \;=\; |c| \|\mathbf{X}\|$ for any scalar c.
(iv) $\|\mathbf{X} + \mathbf{Y}\| \leq \|\mathbf{X}\| + \|\mathbf{Y}\|$
(v) $\|\mathbf{X}\mathbf{Y}\| \leq \|\mathbf{X}\| \|\mathbf{Y}\|$

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

$$\|\mathbf{x}\| \;=\; [\langle \mathbf{x}, \mathbf{x} \rangle]^{1/2} \;=\; \left[\mathbf{x}^T \mathbf{x}\right]^{1/2}. \tag{20}$$

Special matrices

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

$$\langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle_{\mathbf{W}} \;=\; \langle \mathbf{A}^* \mathbf{x}, \mathbf{y} \rangle_{\mathbf{W}}, \tag{21}$$

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

$$\mathbf{A}^* \;=\; \mathbf{W}^{-1} \, \mathbf{A}^H \, \mathbf{W}. \tag{22}$$

In the limiting case of $\mathbf{W} = \mathbf{I}, \mathbf{A}^* = \mathbf{A}^H$. A matrix is **self adjoint** if $\mathbf{A}^* = \mathbf{A}$; $\mathbf{A}$ must be square.

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

A **normal** matrix satisfies: $\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A}$. These matrices also have orthogonal eigenvectors.

A **unitary** matrix satisfies: $\mathbf{A}^{-1} = \mathbf{A}^H$. 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: $\mathbf{A}^{-1} = \mathbf{A}^T$. Orthogonal matrices are also unitary.

A **positive semidefinite** matrix is a Hermitian matrix that satisfies: $\langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle \geq 0$; 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):

$$\mathbf{A} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T, \tag{23}$$

where $\mathbf{U}(\sim m \times n)$ and $\mathbf{V}(\sim n \times n)$ are orthogonal matrices, and $\mathbf{S}(\sim n \times n)$ is a diagonal matrix, with values σ_i along the main diagonal; the σ_i are called the **singular values** of $\mathbf{A}$. Let $\mathbf{u}_i$ and $\mathbf{v}_i$ be the column vectors of $\mathbf{U}$ and $\mathbf{V}$, respectively. $\mathbf{u}_i$ and $\mathbf{v}_i$ are called the left and right singular vectors of $\mathbf{A}$, respectively, and they satisfy

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \qquad \mathbf{A}^T\mathbf{u}_i = \sigma_i \mathbf{v}_i. \tag{24}$$

If $\mathbf{A}$ has r non-zero singular values (σ) then:

- $rank(\mathbf{A}) = r$.
- $null(\mathbf{A}) = span\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$.
- $ran(\mathbf{A}) = span\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$.

- $\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$.
- $\|\mathbf{A}\| = \sum_{i=1}^{r} \sigma_i^2$

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:

$$\mathbf{A}^{-1} = \mathbf{V}\,\tilde{\mathbf{S}}\,\mathbf{U}^T, \tag{25}$$

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 σ_i^{-1}; the other values zero.