

Linear Algebra Reference Sheet

By Danny Liu

1 Basic Knowledge

A vector space is an Abelian group under vector addition with defined scalar multiplication over \mathbb{F}
 v_1, v_2, \dots, v_n is linearly independent when $c_1v_1 + c_2v_2 + \dots + c_nv_n = 0$ has only the trivial solution
If E is an elementary matrix, then EA performs a row operation and AE a column operation
 $A \sim B$ if they represent the same linear operator under possibly different bases, written $B = PAP^{-1}$

1.1 Positive Definiteness

A symmetric matrix A is positive semi-definite if $x^T Ax \geq 0 \forall x \neq \mathbf{0} \iff Av = \lambda v \implies \lambda \geq 0$
Positive definite matrices $x^T Ax > 0$ are nonsingular and have positive diagonal elements
 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is convex $\iff \mathbf{H}_f$ is positive semi-definite $\iff f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
Taylor expansion of f at $\mathbf{z} \in [\mathbf{x}, \mathbf{y}]$ is $f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \mathbf{H}_f(\mathbf{z})(\mathbf{y} - \mathbf{x})$

1.2 Vectorization

$\nabla_x x^T Ax = \nabla_x (\sum_{i,j} x_i x_j A_{ij}) = (A + A^T)x$, $\nabla_x Ax = A$, $\nabla_x x^T y = y$, $\nabla_x x^T x = 2x$
 $\sum_i (\sigma(\theta^T x^{(i)}) - y^{(i)}) x^{(i)} = X^T (\sigma(X\theta) - y)$, $\nabla_\theta \sigma(X\theta) = \text{diag}(\sigma'(X\theta))X$
 $\nabla_x f = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})^T$, $\mathbf{J}_f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_n} \end{bmatrix}$, $\mathbf{H}_f = \mathbf{J}(\nabla_x f)^T = \nabla_x^2 f(x)^T$

2 Change of Basis

Let $B = \{b_1, b_2, \dots, b_n\}$ a basis for \mathcal{V} over \mathbb{F} and $T \in \mathcal{L}(\mathcal{V})$. Then
 $\mathcal{P}_B = [[e_1]_B \quad [e_2]_B \dots [e_n]_B]$ where $\mathcal{P}_B(v) = [v]_B$ and $\mathcal{P}_B^{-1} = [b_1 \quad b_2 \dots b_n]$
 $\mathcal{M}(T) = [Te_1 \quad Te_2 \dots Te_n]$ and $\mathcal{M}(T)^{-1} = [T^{-1}e_1 \quad T^{-1}e_2 \dots T^{-1}e_n]$
 $[\mathcal{M}(T)]_B = \mathcal{P}_B \mathcal{M}(T) \mathcal{P}_B^{-1} = \mathcal{P}_{BE} [\mathcal{M}(T)]_{EE} \mathcal{P}_{EB} = [[Te_1]_B \quad [Te_2]_B \dots [Te_n]_B]$

3 Orthonormal Bases

A list of vectors (v_1, \dots, v_m) is orthonormal if $\langle v_i, v_j \rangle = 0$ for $i \neq j$ and $\|v_i\| = 1$
If $\{e_1, \dots, e_n\}$ is an orthonormal basis for \mathcal{V} and $v \in \mathcal{V}$, then $v = \langle v, e_1 \rangle e_1 + \dots + \langle v, e_n \rangle e_n$
A matrix Q is orthogonal if its columns and rows form orthonormal bases or if $Q^{-1} = Q^T$
An orthogonal matrix can be thought of as a change of basis or a rotation to a new coordinate axes
Orthogonal operators preserve inner product and norm. $\langle Tx, Ty \rangle = \langle x, y \rangle$ and $\|Tv\| = \|v\|$

3.1 Gram-Schmidt

If v_i are independent, then there exist orthonormal e_i such that $\text{span}(v_1, \dots, v_m) = \text{span}(e_1, \dots, e_m)$
 $e_1 = \frac{v_1}{\|v_1\|}$, $e_2 = \frac{v_2 - \langle v_2, e_1 \rangle e_1}{\|v_2 - \langle v_2, e_1 \rangle e_1\|}$, $e_3 = \frac{v_3 - \langle v_3, e_2 \rangle e_2 - \langle v_3, e_1 \rangle e_1}{\|v_3 - \langle v_3, e_2 \rangle e_2 - \langle v_3, e_1 \rangle e_1\|} \dots$

4 Invertible Matrix Theorem

Let A be an $n \times n$ matrix that represents $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. Note this implies $\dim \mathcal{V} = \dim \mathcal{W}$. Then
 A is invertible $\iff A^T$ is invertible $\iff \det(A) \neq 0 \iff \text{rank}(A) = n \iff 0$ is not an eigenvalue of A
 $\iff \text{range } T = \mathcal{W} \iff \ker T = \{\mathbf{0}\} \iff T$ is injective $\iff T$ is surjective

5 Determinant and Trace

The determinant is a function from $n \times n$ matrices to \mathbb{F} defined recursively by $\det A = a$ if $A = [a]$ and $\det A = \sum_j (-1)^{1+j} a_{1j} \det A_{1j}$ otherwise where A_{ij} denotes deleting the i th row and j th column
 $|\det A|$ scales volume of vector transformed by A and is 0 when it is collapsed to a lower dimension
If A is triangular, then $\det A = \prod_i a_{ii}$ where a_{ii} are the diagonal elements
Interchanging rows or columns, $\det A = -\det A$. Adding rows or columns $\det A = \det A$
 $\det A = \prod_i \lambda_i$, $\det(AB) = \det A \det B$, $\det(cA) = c^n \det A$, $\det A^{-1} = \frac{1}{\det A}$, $\det A = \det A^T$

5.1 Characteristic Polynomial

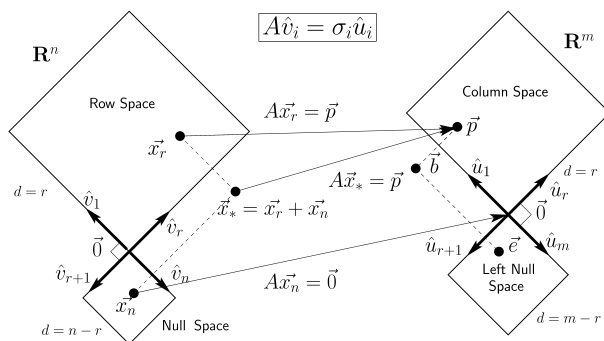
$\det(\lambda I - T) = p(\lambda) = (\lambda - \lambda_1)^{d_1} \dots (\lambda - \lambda_m)^{d_m}$ where $d_i = \dim G(\lambda_i, T) = \dim(\text{null}(T - \lambda I)^{\dim \mathcal{V}})$
If $p(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_1\lambda + c_0$, then $c_{n-1} = -\text{tr } A$ and $c_0 = (-1)^n \det A$
 $p(T) = O$ and the zeros of $p(\lambda)$ are the eigenvalues of T

5.2 Trace

The trace is a function from $n \times n$ matrices to \mathbb{F} defined by $\text{tr } A = \sum_i a_{ii}$
Trace and determinant are both invariant under similarity, so $\text{tr } A = \text{tr}(PJP^{-1}) = \sum_i \lambda_i$
 $\text{tr}(X^T Y) = \sum_{i,j} X_{ij} Y_{ij} = \text{vec}(X)^T \text{vec}(Y) = \langle X, Y \rangle$, $\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$

6 Fundamental Theorem of Linear Algebra

Every $m \times n$ real matrix A contains four fundamental subspaces described by $A = U\Sigma V^T$



This implies if $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, then $\mathcal{V} = \text{row } T \oplus \ker T$ and $\mathcal{W} = \text{row } T^* \oplus \ker T^*$
 $x \mapsto Ax$ is one-to-one $\iff \ker A = \{\mathbf{0}\}$. $\text{col } A = \text{Im } A = \text{range } A = \text{row } A^T$
 $A[v_1 \dots v_r] = [u_1 \dots u_r] \text{diag}(\sigma_1, \dots, \sigma_r) \implies A = U_r \Sigma_r V_r^T = u_1 \sigma_1 v_1^T + \dots + u_r \sigma_r v_r^T$

7 Rank-Nullity Theorem

Let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, then $\text{rank}(T) + \text{nullity}(T) = \dim \mathcal{V}$. Equivalently, $\dim(\text{Im } T) + \dim(\ker T) = \dim \mathcal{V}$.

8 Diagonalization

Let $T \in \mathcal{L}(\mathcal{V})$ and $\{\lambda_1, \dots, \lambda_m\}$ denote the distinct eigenvalues of T . Then

T is diagonalizable $\iff \mathcal{V}$ has a basis consisting of eigenvectors of $T \iff \dim G(\lambda_i, T) = \dim E(\lambda_i, T)$

$\iff \mathcal{V} = E(\lambda_1, T) \oplus \dots \oplus E(\lambda_m, T) \iff \mathcal{V} = U_1 \oplus \dots \oplus U_n$ where each U_i is invariant under T

$A = PDP^{-1}$ where $P = [v_1 \ v_2 \ \dots \ v_n]$ and $D = \text{diag}(\lambda_1, \dots, \lambda_n)$

$D = [A]_B$ where B is the basis consisting of eigenvectors of A and $P^{-1} = P_{BE}$

8.1 Spectral Theorem

If \mathcal{V} is a real inner-product space and $T \in \mathcal{L}(\mathcal{V})$, then \mathcal{V} has an orthonormal basis consisting of eigenvectors of T if and only if T is self-adjoint (symmetric).

If A is orthogonally diagonalizable, then $A^T = (PDP^{-1})^T = (P^{-1})^T DP^T = PDP^{-1} = A$

If A is symmetric, then $(A - \lambda I)^2 v = 0 \implies v^T (A - \lambda I)^2 v = 0 \implies \|(A - \lambda I)v\|^2 = 0 \implies Av = \lambda v$

9 Singular Value Decomposition

Any real $m \times n$ matrix A can be factored into $A = U\Sigma V^T$ where U, V are orthogonal matrices whose columns are the orthonormal eigenvectors of AA^T and $A^T A$ respectively and Σ is the $m \times n$ diagonal matrix of the square roots of the nonzero eigenvalues values of $A^T A$.

9.1 Key Insights

Add rest of basis vectors $v_{r+1} \dots v_n$ and $u_{r+1} \dots u_m$ to V_r and U_r to make them orthogonal matrices

Let $A = U_r \Sigma_r V_r^T = U\Sigma V^T$, then $A^T A = V\Sigma^2 V^T$ so the v_i 's diagonalize $A^T A \succeq 0$

Note that if $A \succeq 0$, then $A^T A = PD^2 P^T$ implies that svd of A is equivalent to eigendecomposition

When A is square, the transformation can be viewed as a change of basis (rotation 1) V^T , a scaling in that intermediate basis Σ and then another change of basis (rotation 2) U

$A^T = (U\Sigma V^T)^T = V\Sigma^T U^T$ viewed as inverse rotation (2), same scaling and inverse rotation (1)

$A^{-1} = (U\Sigma V^T)^{-1} = (V^T)^{-1} \Sigma^{-1} U^{-1} = V\Sigma^{-1} U^T$. Inverse scaling, assuming nonzero scaling terms.

10 QR Decomposition

Any $m \times n$ matrix A with linearly independent columns can be factored into $A = QR$, where Q is an $m \times n$ matrix with orthonormal columns and R is a nonsingular upper triangular matrix.

$$A = [u_1 \ u_2 \ \dots \ u_n] = [q_1 \ q_2 \ \dots \ q_n] \begin{bmatrix} \langle q_1, u_1 \rangle & \langle q_1, u_2 \rangle & \dots \\ 0 & \langle q_2, u_2 \rangle & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

If A does not have linearly independent columns, then R will be singular

11 Triangular Matrices

The inverse, product and sum of an upper (lower) triangular matrix is upper (lower) triangular

A triangular matrix is invertible if and only if the entries on its main diagonal are nonzero

$\mathcal{M}(T - \lambda I)$ is not invertible $\iff \lambda = d_i$ for some diagonal element $d_i \implies$ eigenvalues on main diagonal

If $Ax = b$ where A is lower triangular and nonsingular, then $x_1 = b_1/a_{11}$, ..., $x_n = \frac{b_n - \sum_{k=1}^{n-1} a_{nk}x_k}{a_{nn}}$

Inverse can be computed by solving $A[x_1 \ x_2 \ \dots \ x_n] = [e_1 \ e_2 \ \dots \ e_n]$ column by column

If $A = LU$, then let $Ux = y$ and solve $Ly = b$ via forward-substitution and $Ux = y$ by back-substitution

11.1 LU Decomposition

If A is square and nonsingular, then $A = LU$ for unit lower triangular L and upper triangular U . Defining E_{ij} to remove the i th row in the j th column, we have that $U = E_{n,n-1} \dots E_{32} E_{n,1} \dots E_{21} A$. $A = E_{21}^{-1} \dots E_{n,1}^{-1} E_{32}^{-1} \dots E_{n,n-1}^{-1} U = LU$, where E_{ij} has $-\frac{a_{ij}}{a_{jj}}$ at index (i, j) .
If A is singular, then there exists P such that the algorithm $PA = LU$ avoids dividing by zero.

11.2 LDU Decomposition

If A admits an LU decomposition, then $A = LDU$ for unit triangular L , U and diagonal D . Let $D = \text{diag}(u_{11}, \dots, u_{nn})$ and $U_1 = \frac{U}{\text{vec } D}$, then $A = LDU_1$ as L is already unit triangular.

11.3 Cholesky Decomposition

If A is real positive definite, then $A = LL^T$ for a lower triangular L with positive diagonal entries. $A \succ 0 \implies A = LDU = A^T = (LDU)^T = U^T D L^T = L D L^T = (L D^{\frac{1}{2}})(D^{\frac{1}{2}} L^T) = L_1 L_1^T$.
Since L is nonsingular, let $y = L^T x$. Then $y^T D y = x^T L D L x = x^T A x > 0 \implies D \succ 0$.
Once existence proved, find L by $l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2}$ and $l_{ij} = \frac{1}{l_{jj}}(a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk})$ for $i > j$.

12 Inner Products

$z = a + bi$, $|z|^2 = z\bar{z} = (a + bi)(a - bi) = a^2 + b^2$, $\langle u, v \rangle = \overline{\langle v, u \rangle}$, $\|v\|^2 = \langle v, v \rangle$.
Euclidean inner product over \mathbb{C} becomes $\langle (u_1, \dots, u_n), (v_1, \dots, v_n) \rangle = u_1 \bar{v}_1 + \dots + u_n \bar{v}_n$.
If $\langle u, v \rangle = 0$, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$ (Pythagorean).
 $|\langle u, v \rangle| \leq \|u\| \|v\|$ (Cauchy-Schwartz) and $\|u + v\| \leq \|u\| + \|v\|$ (Triangle).

12.1 Adjoints

The adjoint of $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is $T^* \in \mathcal{L}(\mathcal{W}, \mathcal{V})$ such that $\langle Tv, w \rangle = \langle v, T^*w \rangle$ for all $v \in \mathcal{V}, w \in \mathcal{W}$.
If B is an orthonormal basis for \mathcal{V} , then $[T^*]_B = [T]_B^*$ where $[T]_B^*$ is the conjugate transpose of $[T]_B$.

12.2 Orthogonal Projections

$\text{proj}_{\mathcal{V}} \mathbf{u} = \frac{\langle u, v \rangle}{\langle v, v \rangle} v = \frac{\langle u, v \rangle}{\|v\|} \hat{v} = \frac{\|u\| \|v\| \cos \theta}{\|v\|} \hat{v} = \|u\| \cos \theta \hat{v}$.
The orthogonal complement of $\mathcal{U} \subseteq \mathcal{V}$ is $\mathcal{U}^\perp = \{v \in \mathcal{V} \mid \langle v, u \rangle = 0 \ \forall u \in \mathcal{U}\}$.
 $\mathcal{V}^\perp = \{\mathbf{0}\}$ and \mathcal{U}^\perp is always a subspace of \mathcal{V} . If $\mathcal{U} \subseteq \mathcal{V}$ is a subspace, then $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$.
Let $\mathcal{U} \subseteq \mathcal{V}$ be a subspace. The orthogonal projection of \mathcal{V} onto \mathcal{U} is the operator $P_{\mathcal{U}}$ where if $v = u + w$ where $u \in \mathcal{U}$ and $w \in \mathcal{U}^\perp$, then $P_{\mathcal{U}}(v) = u$.
If $T \in \mathcal{L}(\mathcal{V})$, then \mathcal{U} is invariant under $T \iff P_{\mathcal{U}} T P_{\mathcal{U}} = T P_{\mathcal{U}} \iff \mathcal{U}^\perp$ is invariant under T^* .
Given a subspace \mathcal{U} of \mathcal{V} and a vector $v \in \mathcal{V}$, then $P_{\mathcal{U}}(v) := \underset{u \in \mathcal{U}}{\text{argmin}} \|u - v\|$.

12.3 Riesz Representation Theorem

If $\varphi : \mathcal{V} \rightarrow \mathbb{F}$ is a linear form, then there exists a unique $u \in \mathcal{V}$ such that $\varphi(v) = \langle v, u \rangle$ for all $v \in \mathcal{V}$.