A Concise Review of Linear Algebra

Tarang Srivastava

June 7, 2017

Contents

1 Vectors

1.1 The Geometry and Algebra of Vectors

Vectors is a directed line segment thatpro corresponds from one point A to another point B. The vector from point A to B is denoted by $\vec{AB}$. The **initial point** is is A and the **terminal point** is B. A vector from the origin to the origin, the **zero vector** is denoted by $\mathbf{0}$. A vector with its initial point in the origin is said to be in **standard position**. **vector addition** is by following one vector by another. Geometrically you move the second vector's initial point to the first vector's terminal point.

Algebraic Properties of Vectors

1. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
2. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$
3. $\mathbf{u}+\mathbf{0}=\mathbf{u}$
4. $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$
5. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$
6. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$
7. $c(d\mathbf{u})=(cd)\mathbf{u}$
8. $1\mathbf{u}=\mathbf{u}$

Linear Combinations A vector $\mathbf{v}$ is a **linear combination** of vectors $v_1, v_2, ..., v_k$ if there are scalars $c_1, c_2, ..., c_k$ such that $\mathbf{v}=c_1\mathbf{v}_1+c_2\mathbf{v}_2+...+c_k\mathbf{v}_k$. The scalars $c_1, c_2, ..., c_k$ are called the **coefficients** of the linear combination.

1.2 Length and Angle: The Dot Product

if

$$\mathbf{u}=\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix} \text{ and } \mathbf{v}=\begin{bmatrix}v_1\\v_2\\\vdots\\v_n\end{bmatrix}$$

then the **dot product** is defined by

$$\mathbf{u}\cdot\mathbf{v}=u_1v_1+u_2v_2+...+u_nv_n$$

Properties of the Dot Product

1. $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$
2. $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$

3. $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$

4. $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ iff $\mathbf{u} = 0$

The length or norm of a vector $\mathbf{v}$ is a nonnegative scalar $||\mathbf{v}||$ defined by

$$||v|| = \sqrt{(\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + ... + v_n^2}$$

The Cauchy-Shwarz Inequality: $|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}||\mathbf{v}|$
The Triangle Inequality: $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$
The distance $\mathrm{d}(\mathbf{u}, \mathbf{v}) = |\mathbf{u} - \mathbf{v}|$

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}$$

$$proj_u(\mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}$$

1.3 Lines and Planes

Equations of lines in R^2

$NormalForm$	$GeneralForm$	$VectorForm$	$ParametricForm$
$\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$	$ax + by = c$	$\mathbf{x} = \mathbf{p} + t\mathbf{d}$	$\begin{cases} x = p_1 + td_1 \\ y = p_2 + td_2 \end{cases}$

Equations of lines in R^3

$Type$	$NormalForm$	$GeneralForm$	$VectorForm$	$ParametricForm$
Lines	$\begin{cases} \mathbf{n}_1 \cdot \mathbf{x} = \mathbf{n}_1 \cdot \mathbf{p}_1 \\ \mathbf{n}_2 \cdot \mathbf{x} = \mathbf{n}_2 \cdot \mathbf{p}_2 \end{cases}$	$\begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \end{cases}$	$\mathbf{x} = \mathbf{p} + t\mathbf{d}$	$\begin{cases} x = p_1 + td_1 \\ y = p_2 + td_2 \\ z = p_3 + td_3 \end{cases}$
Planes	$\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$	$ax + by + cz = d$	$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$	$\begin{cases} x = p_1 + su_1 + tv_1 \\ y = p_2 + su_2 + tv_2 \\ z = p_3 + su_3 + tv_3 \end{cases}$

1.4 Code Vectors and Modular Arithmetic

A **binary code** is a set of binary vectors called **code vectors**. Converting a message into code vectors is called **encoding**, the reverse is **decoding**.
An error-detecting code can be achieved by using **parity check code** where and extra component called the **check digit** is added.

2 System of Linear Equations

2.1 Introduction to Systems of Linear Equations

Do row-echelon form don't be stupid

2.2 Direct Methods for Solving Linear Systems

The **coefficient matrix** contains the coefficients of the variables, and the **augmented matrix** is the coefficient matrix with another column for the constant terms.
The **rank** of a matrix is the number of nonzero rows in its row echelon form.
The Rank Therom states, given A the coefficient matrix of a system of linear equations with n variables. If the system is consistent, then

$$\text{number of free variables} = n - \text{rank}(A)$$

A system of linear equations is called **homogeneous** if the constant term in each equation is zero. If $[A|\mathbf{0}]$ is a homogeneous system of m linear equations with n variables, where $m < n$, then the system has infinitely many solutions.

2.3 Spanning Sets and Linear Independence

$[A|\mathbf{b}]$ is consistent iff $\mathbf{b}$ is a linear combination of the columns of A.
A set of vector $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$ is **linearly dependent** if there are scalars $c_1, c_2, ..., c_3$ at least one of which is not zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_k\mathbf{v}_k = \mathbf{0}$$

Let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ be (column) vectors in R^n and let A be the $n \times m$ matrix. $[\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m]$ with these vectors as its columns. Then $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ are linearly dependent iff the homogeneous linear system with augmented matrix $[A|\mathbf{0}]$ has nontrivial situation.

Let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ be (row) vectors in R^n and let A be then $m \times n$ matrix $\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_m \end{bmatrix}$ with these vectors as its rows. Then $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ are linearly depedent iff $rank(A) < m$.

3 Matrices

3.1 Matrix Operations

A matrix all of whose entries are zero is called a **zero matrix** and denoted by O.

$$A + O = A = O + A$$

$$A - A = O = -A + A$$

A square matrix whose nondiagonal entries are all zero is called a **diagonal matrix**. A diagonal matrix all whose entries are the same is a **scalar matrix**. If the scalar on the diagonal is 1, the scalar matrix is called an **identity matrix**. Example:

$$\begin{cases} x_1 - 2x_2 + 3x_3 = 5 \\ -x_1 + 3x_2 + x_3 = 5 \\ 2x_1 - x_2 + 4x_3 = 5 \end{cases}$$

This system can be represented as a matrix product

$$\begin{bmatrix} 1 & -2 & 3 \\ -1 & 3 & 1 \\ 2 & -1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 14 \end{bmatrix}$$

Partitioned Matrices A matrix can be regarded as a larger matrix composed of smaller submatrices. Given two vectors A and B, where AB exists, AB can be represented by breaking B into its column vectors. This givs $AB = A[b_1|b_2|...|b_r] = [Ab_1|Ab_2|...|Ab_r]$. This is called **matrix column representation**. A similar thing can be achieved by partitioning A into its rows, giving **matrix row representation**.

The **transpose** of a an $m \times n$ matrix A is the $n \times m$ matrix A^T obtained by interchanging the rows and columns of A. Given a square matrix A if $A^T = A$ then it is **symmetric**.

3.2 Matrix Algebra

Span of a matrix is the set of linear combinations of the matrices.
Matrices $A_1, A_2, ..., A_k$ of the same size are **linearly independent** if the only solution of the equation

$$c_1 A_1, c_2 A_2, ..., c_k A_k = O$$

is the trivial one: $c_1 = c_2 = ... = 0$. If there are nontrivial coefficients that satisfy that then those matrices are called **linearly dependent**.
Multiplicative Identity: $I_m A = A = A I_n$ if A is $m \times n$
If A is a square matrix, $A + A^T$ is a symmetric matrix. Also, for any matrix A, AA^T, $A^T A$ are symmetric matrices.

3.3 The Inverse of a Matrix

$$AA' = I$$

$$A'A = I$$

A' is an **inverse** of A. If A' exists then A is said to be **invertible**. If A is invertible, as in A' exists, then A' is unique.

1. $(A^{-1})^{-1} = A$
2. $(cA)^{-1} = \frac{1}{c} A^{-1}$

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

An **elementary matrix** is any matrix that can be obtained by performing elementary row operations to an identity matrix.

3.4 The LU Factorization

If A is a square matrix that can be reduced to row echelon form without any row interchanges, then A has an LU factorization.
Okay, so here is how you do LU factorization. Remember NO row interchanges. Alright, lets say we have a matrix A, we first row reduce this. But see we have to be very careful how we row reduce. We first start with the first column, and the first row, and try and get that entry to 0. We note down what row operations we did to get the first space 0, the first space will be A_{12} as in first column second row. We keep doing this until the first column is a number followed by zeros underneath it. Now we move to the next column, again working the same way, from top to bottom in the row. The matrix you have left is U. In order to find L we look at the row operations we did to get the given spaces to 0 and fill that with the negative coefficient of that row operation. So lets say we wanted the last row in the first column to be 0, and we wrote down something like $R_4 - 2R_1$ in this case the number we put in L is 2.
Now lets say we must perform a row interchange, we use a **permutation matrix**. if we interchange rows 2 and 3 for a 3×3 matrix we will do this by using $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and the inverse of permutation matrices is just their inverse

$$P^{-1} = P^T$$

every square matrix has a $P^T LU$ Factorization

3.5 Subspaces, Basis, Dimension and Rank

A **subspace** of R^n is any collection S of vectors in R^n such that

1. The zero vector is in S
2. if $\mathbf{u}$ and $\mathbf{v}$ are in S, then $\mathbf{u} + \mathbf{v}$ is in S
3. if $\mathbf{u}$ is in S and c is a scalar, c$\mathbf{u}$ is in S

row and column subspace are spans of the rows and columns of A. denoted by row(A) and col(A). The **nullspace** of A is the subspace consisting of solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. denoted by null(A).
A basis for a subspace is a set of vectors that it spans the subspace, and all vectors in set are linearly independent
The Basis Theorem states for a subspace S, and two bases of S have the same number of vectors.

The number of vectors in a basis for S is called the **dimension** denoted by dim S. **Rank** of a matrix A is the dimension of its row and column spaces and is denoted by rank(A)

$$\text{rank}(A^T) = \text{rank}(A)$$

The **nullity** of matrix is the dimension of its null space, denoted nullity(A).
The Rank Theorem if A is a $m \times n$ matrix, then

$$\text{rank}(A) + \text{nullity}(A) = n$$

4 Eigenvalues and Eigenvectors

4.1 Introduction

Let A be an $n \times n$ matrix. A scalar λ is called an **eigenvalue** of A if there is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Such a vector $\mathbf{x}$ is called an **eigenvector** of A corresponding to λ.
The collection of all eigenvectors corresponding to λ, together with the zero vector, is called the **eigenspace** of λ and is denoted by E_λ

4.2 Determinants

Let $A = [a_{ij}]$ be an $n \times n$ matrix, where $n \geq 2$. Then the **determinant** of A is the scalar

$$\text{det}A = |A| = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} det A_{1j}$$

A square matrix is invertible iff det A not equal to 0 If A is a square matrix, $det(kA) = k^n det A$ Determinant of a triangular matrix is the product of its diagonal and the determinant of the last 2 by 2 matrix.
Cramer's rule Let A be an invertible $n \times n$ matrix and let $\mathbf{b}$ be a vector in R^n. Then the unique solution $\mathbf{x}$ of the system $A\mathbf{x} = \mathbf{b}$ is given by

$$x_i = \frac{\det(A_i(\mathbf{b}))}{\det A}$$

You use the cramer rule by replacing A_i which is the ith column of A, by $\mathbf{b}$ and finding the determinant of that matrix.

If A is a matrix, and B is the result of row interchanging of A then det(A) = -det(B)

4.3 Eigenvalues and Eigenvectors of $n \times n$ Matrices

Eigenvalues are found by $\det(A - \lambda I) = 0$. The eigenspace for each respective λ is by finding the null space of matrix $A - \lambda I$ The nonzero vectors of which are the eigenvectors corresponding to that eigenvalue. Eigenvalues of a triangular matrix are the values on its diagonal. A position vector can be represented as a linear combination of eigenvectors of a matrix. Eigenvectors are linearly independent.

4.4 Similarity and Diagonalization

A is similar to be if $P^{-1}AP = B$ this is written as $A \sim B$
If $A \sim B$

1. det(A) = det(B)
2. A is invertible iff B is invertible
3. A and B have same eigenvalues

If A is a $n \times n$ matrix with n eigenvalues it is diagonalizable

5 Orthogonality

5.1 Orthogonality in R^n

A set of vectors $[\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k]$ is called and **orthogonal set** if all pairs of distinct vectors are orthogonal. The vectors in an orthogonal set are linearly independent.
An **orthogonal basis** for a subspace W is a basis of W that is an orthogonal set.

$$\mathbf{w} = c_1\mathbf{v}_1, c_2\mathbf{v}_2, ..., c_k\mathbf{v}_k$$

$$c_i = \frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$$

An orthogonal matrix is one whose columns form an orthonormal set. A square matrix Q is orthogonal iff $Q^{-1} = Q^T$ Q is an $n \times n$ matrix

1. Q is orthogonal
2. $||Q\mathbf{x}|| = ||\mathbf{x}||$
3. $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$

5.2 Orthogonal Complements and Orthogonal Projections

$$W^{\perp} = \mathbf{v} : \mathbf{v} \cdot \mathbf{w} = 0 \text{for all } \mathbf{w} \text{ in W}$$

$\mathbf{v} = \mathbf{w} + \mathbf{w}^{\perp}$
$(W^{\perp})^{\perp} = W$
Let A be an $m \times n$ matrix. Now think about the following statement, it should make complete sense.

$$(\text{row}(A))^{\perp} = \text{null}(A) \text{ and } (\text{col}(A))^{\perp} = \text{null}(A^T)$$

The **orthogonal projection** of v onto W, where w is defined by the and orthogonal basis of **u**

$$\text{proj}_W(v) = (\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1})\mathbf{u}_1 + ... + (\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k})\mathbf{u}_k$$

$$\text{perp}_W(\mathbf{v}) = \mathbf{v} - \text{proj}_W(v)$$

5.3 The Gram-Schmidt Process and the QR Factorization

Let $\mathbf{x}_1, ..., \mathbf{x}_k$ be a basis for a subspace W of

$$v_1 = x_1$$

$$v_k = x_k - ... - (\frac{\mathbf{v}_{k-1} \cdot \mathbf{x}}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}})\mathbf{v}_{k-1}$$

The **QR Factorization** for an $m \times n$ matrix A with linearly independent columns. Then A can be factored as A = QR, where Q is an $m \times n$ matrix with orthonormal columns, found using Gram-Schmidt Process and normalizing the orthogonal vectors. Then R can be find by such $A = QR$ which gives $Q^T S = R$ since Q is orthogonal matrix its inverse is equal to its transpose.

5.4 Orthogonal Diagonalization of Symmetric Matrices

A square matrix A is **orthogonally diagonalizable** if there exists and orthogonal matrix Q and a diagonal matrix D such that $Q^T AQ = D$. If A is orthogonally diagonalizable it is symmetric.
The Spectral Theorem states Let A be an $n \times n$ real matrix. Then A is symmetric iff it is orthogonally diagonalizable
The **spectral decomposition** is given by

$$A = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots + \lambda_n q_n q_n^T$$

5.5 Applications and the Perron-Frobenius Theorem

$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is where A is the matrix associated with f and is a symmetric square matrix.

6 Vector Spaces

6.1 Vector Spaces and Subspaces

If W is a subspace of V, then W contains the zero vector **0** of V and is closed under addition and scalar multiplication.

6.2 Linear Independence, Basis, and Dimension

For a set of vectors, and a linear combination, if there is one combination such that there is atleast one scalar that is not 0, and the linear combination equals 0 then the set of vectors is **linearly dependent** If B is a basis for V. For every vector **v** in V, there is exactly one way to write **v** as a linear combination of the basis vectors in B. This is because in order for B to be a basis it must be linearly independent.

Let V be a vector space with dim V = n. Then

1. Any linearly independent set in V contains at most n vectors

2. Any spanning set for V contains at least n vectors

3. Any linearly independent set of exactly n vectors in V is a basis for V

4. Any spanning set for V consisting of exactly n vectors is a basis for V

5. Any linearly independent set in V can be extended to a basis for V

6. Any spanning set for V can be reduced to a basis for V

6.3 Change of Basis

Let B = $u_1, ..., u_n$ and C = $v_1, ..., v_n$ be bases for a vector space V and let $P_{C \leftarrow B}$ be the change of basis matrix from B to C. Then

1. $P_{C \leftarrow B}[\mathbf{x}]_B = [\mathbf{x}]_C$

2. $P_{C \leftarrow B}$ is a unique matrix, with the property above for all $\mathbf{x}$ in V

3. $P_{C \leftarrow B}$ is invertible and $(P_{C \leftarrow B})^{-1} = P_{B \leftarrow C}$

6.4 Linear Transformation

A **linear transformation** from a vector space V to a vector space W is a mapping $T : V \rightarrow W$ such that, for all $\mathbf{u}$ and $\mathbf{v}$ in V for all scalars c

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

2. $T(c\mathbf{u}) = cT(\mathbf{u}$

Let $T : V \leftarrow W$ be a linear transformation. Then

1. $T(\mathbf{0}) = \mathbf{0}$

2. $T(-\mathbf{v}) = -T(\mathbf{v})$

3. $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$

if $T : U \leftarrow V$ and $S : V \leftarrow W$ are linear transformations, then $S \circ T : U \leftarrow W$ is a linear transformation

$T : U \leftarrow V$ is invertible if there is a linear transformation $T' : V \leftarrow W$ such that

$$T' \circ T = I_V \text{ and } T \circ T' = I_W$$

T' is called the inverse of T. The inverse of T' is unique

6.5 The Kernel and Range of a Linear Transformation

kernel is the set of all vectors that are mapped by T to $\mathbf{0}$. **Range** is the set o of all vectors that are images of vectors in V under T.

$$rank(T) = nullity(T) = dimV$$

For a transformation $T : V \to W$
A linear transformation $T : V \to W$ is called **one to one** if T maps distinct vectors. if range(T) = W then it is called **onto**
A transformation is one to one for all **u** and **v** if

$$T(\mathbf{u}) = T(\mathbf{v}) \text{ implies that } \mathbf{u} = \mathbf{v}$$

A linear transformation is one to one iff ker(T) = $\mathbf{0}$. A linear transformation is invertible iff it is onetoone and onto. A linear transformation is called an **isomorphism** if it is one to one and onto

6.6 The Matrix of a Linear Transformation

Let V and W be two finite dimensional vector spaces with bases B and C respectfully, where B = v.... Then A is defined by

$$A = [[T(v_1)]_C ... [T(v_n)]_C]$$

$$[S \circ T]_{D \leftarrow B} = [S]_{D \leftarrow C}[T]_{C \leftarrow B}$$

For a transformation $T : V \to V$ be a linear transformation. Then T is called Diagonalizable if there is a basis for C for V such that the matrix $[T]_C$ is a diagonal matrix

7 Distance and Approximation

7.1 Inner Product Spaces

An **inner product** on a vector space V is an operation that assigns to every pair of vectors **u** and **v** in V a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ such that the following properties hold

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{v}, \mathbf{u} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ iff $\mathbf{u} = \mathbf{0}$

Let **u** and **v** be vectors in an inner product space V and let c be a scalar

1. the length or norm of v is $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$

2. the distance between u and v is $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$

3. orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$

The Cauchy-Schwar Inequality

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

7.2 Norms and Distance Functions

List of norms

1. the sum norm of a vectors is the sum of the absolute values of its components

2. the max norm is the highest component of among the absolute values of each component

3. Euclidean norm is the standard norm by squaring each component and taking square root of the sum of all of it

A matrix norm is mapping that associates each matrix A with a real number $\|A\|$ called the **norm**

7.3 Least Squares Approximation

If W is a subspace of a normed linear space V and if $\mathbf{v}$ is a vector in V, then the **best approximation to v in W** is the vectors $\bar{\mathbf{v}}$ in W such that

$$\|\mathbf{v} - \bar{\mathbf{v}}\| < \|\mathbf{v} - \mathbf{w}\|$$

We can say the $\bar{\mathbf{v}} = proj_W(v)$
Least squares can be found using QR given by $A = QR$

$$\bar{\mathbf{x}} = R^{-1} Q^T \mathbf{b}$$

$$\text{proj}_W(\mathbf{v}) = A(A^T A)^{-1} A^T \mathbf{v}$$

The pseudoinverse of A is denoted by A^+

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

A short hand with the pseudoinverse is

$$\bar{\mathbf{x}} = A^+ \mathbf{b}$$

7.4 The Singular Value Decomposition

singular values of A are the square root of the eigenvalues of $A^T A$ and are denoted by $\sigma_1, ..., \sigma_n$

$$A = U \Sigma V^T$$

Here is how you do the SVD. Remember that. A is an $m \times n$ matrix. U is an $m \times m$ orthogonal matrix. Σ is an $m \times n$ diagonal matrix and V is an $n \times n$ orthogonal matrix. Σ is a diagonal matrix where on the diagonal are singular values in descending order, all the other spaces are 0, in order to satisfy the size of the matrix. V is the eigenvectors of $A^T A$, as the columns for V. $U = [\mathbf{u}_1, ..., \mathbf{u}_m]$ where $\mathbf{u}_i = \frac{1}{\sigma_1} A \mathbf{v}_i$