1 Notation

1.1 Sets and spaces

$C[a,b] = \{f : [a,b] \to \mathbb{R} | f \text{ is continuous } \}$. $a < b$

col A is the range or span of a matrix $A \in \mathbb{F}^{n \times m}$

$\mathrm{d}(A, M_k) = \inf\{||A - S||_2 : S \in M_k\}$
$\dim\mathcal{V}$ is the dimension of $\mathcal{V}$, provided $\mathcal{V}$ is finite dimensional.

$\{e_1, e_2, ..., e_n\}$ is the set of vectors where e_j has all entries zero, except the jth entry, which is one: $\begin{bmatrix} 0 \\ ... \\ 0 \\ 1 \\ 0 \\ ... \\ 0 \end{bmatrix}$

$\mathbb{F}^n$ is the set of all vectors $\begin{bmatrix} u_1 \\ u_2 \\ ... \\ u_n \end{bmatrix}$ where $u_i \in \mathbb{F}$ for all i in $1, ...n$.

$\mathbb{F}^{n \times m}$ is the set of all matrices with n rows and m columns with elements in $\mathbb{F}$.

$\mathcal{L}(\mathcal{U}, \mathcal{W})$ is the set of all linear transformations from $\mathcal{V}$ to $\mathcal{W}$.

$\mathcal{L}(\mathcal{V}) := \mathcal{L}(\mathcal{V}, \mathcal{V})$.

$\mathcal{L}(\mathcal{V}, \mathbb{F})$ is the dual vector space of $\mathcal{V}$. Its elements are called linear functionals.

nullA, kerA and N(A) are all notations for the nullity (or null space) of A.

kerT is the kernel of the linear transformation T

M_k is the set of all matrices of rank less than or equal to k. In the context of finding the distance between some matrix A and this set, we furthermore say that all matrices in this set have the same shape as A. M_k is not a subspace of $R^{m \times n}$.
ranT is the range of the linear transformation T

P_n is the set of all polynomials of degree at most n with coefficients in $\mathbb{F}$.

$\mathcal{S}^\perp$ is the orthogonal complement of $\mathcal{S}$.

$\text{span}(v_1, v_2, ..., v_r) := \{c_1v_1 + c_2v_2 + ... + c_rv_r \in \mathbb{F}\}$

T($\mathcal{V}$) is the image of T.

1.2 Operations

A is the left inverse of A.
$\oplus$ is vector addition. Often, the circle around the plus sign is left out.

$\odot$ is scalar multiplication. Often, the circle around the dot is left out.

$\mathcal{P}_{\mathcal{U}}(v)$ is the orthogonal projection of v onto $\mathcal{U}$.

$-u$ is the **additive inverse** of u

$\mathcal{U} \oplus \mathcal{W}$ is the direct sum of $\mathcal{U}$ and $\mathcal{W}$.

$[\cdot]_\beta : \mathcal{V} \to \mathbb{F}^n$ is the β-basis representation function for an n-dimensional $\mathcal{V}$.

$|| \cdot ||_2$ is the matrix 2-norm.
$< \cdot, \cdot >$ is an inner product

$v \perp w$: v and w are perpendicular

$g(u_1, ...u_n)$ is the Gram determinant of $u_1, ..., u_n$

$G(u_1, ..., u_n)$ is the Grammian of $u_1, ..., u_n$

$\geq, >, \leq, <$ stand for positive semidefiniteness, positive definitenes, negatie semidefiniteness and negative definiteness of a matrix respectively.

Λ is a real diagonal matrix with diagonal entries $\lambda_1, \lambda_2, ...\lambda_n$. $\Lambda \geq 0$.

σ_i is the ith singular value of a matrix.

2 Definitions

2-norm: $||M||_2 = \sup_{\text{nonzero vectors } x} \frac{||Mx||}{||x||}$ where $||\cdot||$ is the usual Euclidian norm.

Adjoint: Given finite dimensional $\mathbb{F}$-inner product spaces $\mathcal{V}$ and $\mathcal{W}$ and $T : \mathcal{V} \mapsto \mathcal{W}$, the adjoint of T is T^*, such that $< T(v), w >_{\mathcal{W}} =< v, T^*(w) >_{\mathcal{V}}$ for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$. The adjoint is unique to T and it is also a linear transformation.

Asymptotically stable: A system of differential equations $\frac{dx}{dt}(t) = Ax(t)$ is asymptotically stable if $x(t) \to 0$ as $t \to \infty$ for all initial vectors $x(0) = x_0 \in \mathbb{R}^n$. See Lyuapunov's theorem for more.

Best approximation: For $u \in \mathcal{V}$ and a finite-dimensional subspace $\mathcal{U}$ the best approximation for v is a vector $u \in \mathcal{U}$ such that $||v-u|| \leq ||v-w||$ for all $w \in \mathcal{U}$. In fact, $u = \mathcal{P}_{\mathcal{U}}(v)$ with equality only if it and w are equal.

Basis: Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. The ordered vectors $v_1, v_2, ..., v_n$ for a basis of $\mathcal{V}$ if

1. $v_1, v_2, ..., v_n$ are linearly independent, and
2. $\text{span}(v_1, v_2, ..., v_n) = \mathcal{V}$.

Bases are not unique.

Basis representations: Let $\beta = (v_1, v_2, ..., v_n)$ be an ordered basis of $\mathcal{V}$. For any $u \in \mathcal{V}$, write $u = c_1v_1 + c_2v_2 + ... + c_nv_n$, where $c_1, c_2, ..., c_n \in \mathbb{F}$. The function $[\cdot]_\beta : \mathcal{V} \to \mathbb{F}^n$, defined by $[u]_\beta = \begin{bmatrix} u_1 \\ u_2 \\ ... \\ u_n \end{bmatrix}$ is called the β-basis representation function.

Definiteness of matrices:
For an n by n Hermitian matrix A, we say it is...

- positive semidefinite if $< Ax, x > \geq 0$ for all vectors $x \in \mathbb{F}^n$.
- positive definite if $< Ax, x > 0$ for all nonzero vectors $x \in \mathbb{F}^n$.
- negative semidefinite if $< Ax, x > \leq 0$ for all vectors $x \in \mathbb{F}^n$
- negative definite if $< Ax, x >< 0$ for all nonzero vectors $x \in \mathbb{F}^n$
- indefinite if $< Ax, x >$ takes both positive and negative values.

Remember: for the standard inner product, $< Ax, x >= x^* Ax$

Dimension: The dimension of a vector space is the size of its bases.

Direct sum: The sum of two subspaces is called a direct sum if their intersection consists of only the zero vector.

Dual vector space of $\mathcal{V}$: $\mathcal{L}(\mathcal{V}, \mathbb{F})$

Finite dimensional: If $\mathcal{V}$ has dimension n for some integer n, we say $\mathcal{V}$ is finite dimensional. Otherwise, $\mathcal{V}$ is inifinte dimensional.

Full column rank: An m by n matrix is full column rank if its rank is n.

Full row rank: An m by n matrix is full row rank if its rank is m.

Gram determinant: of a list of vectors; the determinant of the Grammian

Gram matrix/Grammian: See methods $\rightarrow$ find the orthogonal projection.

Hermitian: A linear transformation is Hermitian, or self-adjoint, if it is equal to its own adjoint.
A matrix A is Hermitian if $A^* = A$.

Hessian: The Hessian of a function F at point $\overline{x}$ is $\begin{pmatrix} \frac{\partial^2 F}{\partial x_1^2}(\overline{x}) & \frac{\partial^2 F}{\partial x_1 x_2}(\overline{x}) & \dots & \frac{\partial^2 F}{\partial x_1 x_n}(\overline{x}) \\ \frac{\partial^2 F}{\partial x_2 x_1}(\overline{x}) & \frac{\partial^2 F}{\partial x_2^2}(\overline{x}) & \dots & \frac{\partial^2 F}{\partial x_2 x_n}(\overline{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n x_1}(\overline{x}) & \frac{\partial^2 F}{\partial x_n x_2}(\overline{x}) & \dots & \frac{\partial^2 F}{\partial x_n^2}(\overline{x}) \end{pmatrix}$
This matrix can tell us more about the stationary point $\overline{x}$: It is a...

- local minimum if the Hessian is positive definite
- local maximum if the Hessian is negative definite
- saddle point if the Hessian is indefinite

Inner product: a map $\mathcal{V} \times \mathcal{V} \to \mathbb{F}$ such that:

1. if we fix any $w \in \mathcal{V}$ then the map $v \mapsto < v, w >$ is a linear map.
2. "symmetry": $< v, w >= \overline{< w, v >}$.
   - If $\mathbb{F} = \mathbb{R}$ then this simply means $< v, w >=< w, v >$ and in this case $w \mapsto < v, w >$ is an $\mathbb{R}$-linear map for all vectors $v \in \mathcal{V}$
   - IF $\mathbb{F} = \mathbb{C}$, then $w \mapsto < v, w >$ need not be alinear map over $\mathbb{C}$ for a fixed v.
3. "positive defineteness": $< v, v > \geq 0$ for all $v \in \mathcal{V}$ and $< v, v >= 0 \leftrightarrow v = \mathbf{0} \in \mathcal{V}$.

Inner product space: A vector space equipped with an inner product

Kernel: The kernel of a linear transformation is the set of all vectors v in the domain such that $T(v) = 0$.

Least squares solution: A "solution" x_0 to an inconsistent system Ax = y such that $||y - Ax_0||$ is as small as possible. This is the "best approximate solution". Here, the norm is defined by $||u|| = (\sum_{i=1}^{n} |u_i|^2)^{1/2}$

Left inverse: A matrix B is the left inverse of A if BA = I.

Left singular vectors: The left singular vectors of A are the columns of V in the singular value decomposition.

Linear combination: Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. For given $c_1, c_2, ..., c_r \in \mathbb{F}$ and $v_1, v_2, ..., v_r \in \mathcal{V}$, we say that an expression of the form:
$c_1v_1 + c_2v_2 + ... + c_rv_r$
is a linear combination of $v_1, v_2, ..., v_r$.

Linear dependence: A set of vectors $v_1, v_2, ..., v_r \in \mathcal{V}$ is linearly dependent if there exist scalars $c_1, c_2, ..., c_r \in \mathbb{F}$, not all equal to zero, such that $c_1v_1 + c_2v_2 + ... + c_rv_r = 0$.

Linear functional: A linear transformation $\phi : \mathcal{V} \mapsto \mathbb{F}$

Linear independence: A set of vectors $v_1, v_2, ..., v_r \in \mathcal{V}$ is linearly independent if they are not linearly dependent.
Equivalently, if $c_1v_1 + c_2v_2 + ... + c_rv_r = 0$ for $c_1, c_2, ..., c_r \in \mathbb{F}$ implies that $c_1 = c_2 = ... = c_r = 0$.

Linear isomorphism: A bijective linear transformation

Linear transformation: Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the same field $\mathbb{F}$. A function $T : \mathcal{V} \to \mathcal{W}$ is called a linear transformation if $T(au + bv) = aT(u) + bT(v)$ for all $a, b \in \mathbb{F}$ and $u, v \in \mathcal{V}$. Also called a linear map or an $\mathbb{F}$-linear map.

Linear operator: If $\mathcal{V} = \mathcal{W}$ and $T : \mathcal{V} \to \mathcal{V}$ is a linear transformation, we say that T is a linear operator.

Matrix representation: See methods

Minimum norm solution: A solution to a consistent system (A,y) such that $||s|| \leq ||u||$ whenever $Au = y$ ($||s|| = \sqrt{s \cdot s^T}$

Norm: A map $|| \cdot || : \mathcal{V} \to \mathbb{F}$ that satisfies:

- $||v|| = 0 \leftrightarrow v = 0$
- $||av|| = |a|||v||$ where a is a scalar
- Triangle inequality: $||u + v|| \leq ||u|| + ||v||$

Normed vector space: A vector space equipped with a norm $|| \cdot || : \mathcal{V} \to \mathbb{R}_{\geq 0}$

The **nullity** of A is the set of all vectors x such that $Ax = 0$.

Orthogonal basis: a basis $\{u_1, u_2, ..., u_n\}$ such that $< u_1, u_j >= 0$ for all $i \neq j$

Orthogonal complement: $\mathcal{U}^\perp$ of $\mathcal{U}$, a subset of an inner product space $\mathcal{V}$, $= \{v \in \mathcal{V} :< u, v >= 0 \text{ for all } u \in \mathcal{U}\}$

Orthogonal projection: The orthogonal projection of a vector $v \in \mathcal{V}$ onto a finite-dimensional subspace $\mathcal{U}$ of inner product space $\mathcal{V}$ is the vector $u \in \mathcal{U}$ such that $v = u + (v - u)$ where $v - u \in \mathcal{U}^\perp$. In less rigorous terms, u is the vector in $\mathcal{U}$ that is "closest" to v.
$\mathcal{P}_\mathcal{U}(v) = \sum_{i=1}^{r} < v, u_i > u_1$, where $u_1, u_2, ..., u_n$ is an orthonormal basis for $\mathcal{U}$.
For the case where we don't have an orthonormal basis, see methods.
The function that maps v to its orthogonal projection is a linear transformation.

Orthogonal projector: the function $\mathcal{P}_\mathcal{U}(v) : \mathcal{V} \mapsto \mathcal{U}$. It is a linear operator.

An **orthonormal basis** is a basis that is orthonormal. i.e. $< u_i, u_j >= 0$ for $i \neq j$ and $< u_i, u_j >= 1$ for $i = j$.

Perpendicular: Two vectors v and w are perpendicular iff $< v, w >= 0$. We denote this as $v \perp w$

Range: The range of $T : \mathcal{V} \to \mathcal{W}$ is defined as
$\{w \in \mathcal{W} | \exists v \in \mathcal{V}$ such that $T(v) = w\}$

Rank: The rank of a matrix is the dimension of the span of its column vectors.

Rank deficient: An m by n matrix is rank deficient if its rank is less than the length of its shortest side.

Riesz vector: the Riesz vector for the linear functional ϕ is the vector w as defined in the Riesz representation theorem.

Right singular vector: The right singular vectors of A are the columns of W in the singular value decomposition.

Scalar multiplication. For a field $\mathbb{F}$ of scalars and a set $\mathcal{V}$ with elements called vectors, we define the operation $\odot$: $\mathbb{F} \times \mathcal{V} \to V$. This assigns to any scalar $c \in \mathbb{F}$ and vector $v \in \mathcal{V}$ a new vector $c \odot v \in \mathcal{V}$

Singular values: The values used in singular value decomposition (see methods).

Span: Let $v_1, v_2, ..., v_r \in \mathcal{V}$. The set of all linear combinations of $v_1, v_2, ..., v_r$ is called the span of $v_1, v_2, ..., v_r$. The span takes finite sums.

Spanning set: Let $v_1, v_2, ..., v_r \in \mathcal{V}$. We say that $\{v_1, v_2, ..., v_r\}$ is a spanning set for $\mathcal{V}$ if every vector in $\mathcal{V}$ can be written as a linear combination of $v_1, v_2, ..., v_r$.

Square root of a matrix: The square root of a matrix A is a matrix B such that $B^2 = B \cdot B = A$.
Some matrices have many square roots, and some have none.

Standard basis: The basis with ones in each spot. For example, the standard basis for $\mathbb{F}^2$ is the set $\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\}$.

Stationary point: Let F: $\mathbb{R}^n \to \mathbb{R}$ be a C^2-function (twice continuously differentiable). We say $\overline{x} \in \mathbb{R}^n$ is a stationary point if the gradient of $\overline{x} = \begin{pmatrix} \frac{\partial F}{\partial x_i}(\overline{x}) \\ \vdots \\ \frac{\partial F}{\partial x_n}(\overline{x}) \end{pmatrix} = 0$. **Subspace**: For a vector space $\mathcal{V}$ over $\mathbb{F}$ and a subset $\mathcal{U} \subset \mathcal{V}$, $\mathcal{U}$ is a subspace of $\mathcal{V}$ if $\mathcal{U}$ is a vector space over $\mathbb{F}$, with the same vector addition and scalar multiplication operations as in $\mathcal{V}$.

Sum of two subspaces: $\mathcal{U} + \mathcal{W} = \{u + w | u \in \mathcal{U}$ and $w \in \mathcal{W}\}$

Unitary matrix: A matrix U such that $U^2 = I$.

Vector addition. For a field $\mathbb{F}$ of scalars and a set $\mathcal{V}$ with elements called vectors, we define the operation $\oplus$: $\mathcal{V} \times \mathcal{V} \to \mathcal{V}$. This assigns to any pair $u, v \in \mathcal{V}$ a new vector $u \oplus v \in \mathcal{V}$

Vector space: The set $\mathcal{V}$ is called a vector space over $\mathbb{F}$ if $\mathcal{V}$, $\mathbb{F}$ and the operations $\oplus$, $\odot$ satisfy the following axioms:

1. There exists an element $0 \in \mathcal{V}$ such that $0 \oplus u = u$ for all $u \in \mathcal{V}$. 0 is called the **zero vector**.
2. For all $u, v \in \mathcal{V}$, $u \oplus v = v \oplus u$
3. For all $u, v, w \in \mathcal{V}$, $u \oplus (v \oplus w) = (u \oplus v) \oplus w$.
4. For each $u \in \mathcal{V}$, there exists $x \in \mathcal{V}$ such that $u \oplus x = 0$.
5. For all $u \in \mathcal{V}$, $1 \odot u = u$.
6. For all $a, b \in \mathbb{F}$ and all $u \in \mathcal{V}$, $a \odot (b \odot u) = (ab) \odot u$.
7. For all $c \in \mathbb{F}$ and all $u, v \in \mathcal{V}$, $c \odot (u \oplus v) = (c \odot u) \oplus (c \odot v)$.
8. For all $a, b \in \mathbb{F}$ and all $u \in \mathcal{V}$, $(a + b) \odot u = (a \odot u) \oplus (b \odot u)$.

3 Theorems

3.1 Vector spaces

Properties of all vector spaces:
For a vector space $\mathcal{V}$ and $u \in \mathcal{V}$:

1. the zero vector is unique
2. $0 \odot u = 0$
3. the additive inverse z of u is unique
4. $(-1) \odot u = z$

3.2 Subspaces

Definition of subspace:
For $\mathcal{V}$ a vector space over $\mathbb{F}$ and $\mathcal{U} \subset \mathcal{V}$, $\mathcal{U}$ is a subspace of $\mathcal{V}$ iff the following three properties hold:

1. $\mathcal{U}$ is nonempty
2. $u + v \in \mathcal{U}$ for all $u, v \in \mathcal{U}$ (closed under vector addition)
3. $cu \in \mathcal{U}$ for all $c \in \mathbb{F}$ and $u \in \mathcal{U}$ (closed under scalar multiplication)

Contents of a subspace:
A subspace $\mathcal{U}$ of $\mathcal{V}$ always contains the zero vector of $\mathcal{V}$ because $0 \cdot u = 0 \in \mathcal{U}$.

3.3 Bases, linear (in)dependence and span

Narrowing down a spanning set
Let $v_1, v_2, ..., v_r$ be vectors in $\mathcal{V}$.
If $v_1, v_2, ..., v_r$ span $\mathbb{V}$ and v_i is a linear combination of the other vectors, then $\{v_1, v_2, ..., v_r\} \setminus v_1$ spans $\mathcal{V}$.

Singular matrix
The vectors $v_1, v_2, ..., v_r \in \mathbb{F}^n$ are linearly dependent iff the matrix $V := [v_1\ v_2\ ...\ v_n]$ is singular.

Unique expressions of vectors

- Every vector in the span of a set of linearly independent vectors can be written uniquely as a linear combination of said vectors.
- Every vector in a direct sum $\mathcal{U} \oplus \mathcal{W}$ can be uniquely expressed as a sum of an element of $\mathcal{U}$ and an element from $\mathcal{W}$.

Replacement lemma: Let $\mathcal{V}$ be a nonzero vector space over $\mathbb{F}$ and let r be a positive integer. Suppose that $u_1, u_2, ..., u_r$ span $\mathcal{V}$. Let $v = \sum_{i=1}^{r} c_i u_i$ be a nonzero vector in $\mathcal{V}$.

1. $c_j \neq 0$ for some $j \in \{1, 2, ..., r\}$.
2. if $c_j \neq 0$ then $v, u_1, ..., u_{j-1}, u_{j+1}, ..., u_r$ span $\mathcal{V}$. (Note we replaced u_j by v)
3. If $\{u_1, u_2, ..., u_r\}$ is a basis of $\mathcal{V}$ then the vectors $v, u_1, ..., u_{j-1}, u_{j+1}, ..., u_r$ also form a basis of $\mathcal{V}$.

Size of spanning sets

- Any two bases for a vector space have the same size.
- If we can find a spanning set for a vector space of size r then the dimension of that vector space is equal to or less than r.
- Suppose that $\mathcal{V}$ is finite dimensional with $\dim\mathcal{V} = n > r \geq 1$. If $v_1, v_2, ...v_r$ are linearly independent then there exist $w_{r+1}, ..., w_n$ such that $\{v_1, v_2, ..., v_r, w_{r+1}, ..., w_n\}$ is a basis of $\mathcal{V}$.

Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. If there is a basis for $\mathcal{V}$ of size n, then

1. Any set of linearly independent vectors in $\mathcal{V}$ cannot be larger than this basis.
2. If a set of linearly independent vectors in $\mathcal{V}$ has size n, it is also a basis of $\mathcal{V}$.

Dimensions of various spaces

- If $\mathcal{V} = \{0\}$ then $\mathcal{V}$ has dimension zero.
- $\dim\mathbb{F}^n = n$
- P_n and $C[a, b]$ are infinite dimensional.
- If $\mathcal{U}$ is a subspace of a finite dimensional vector space $\mathcal{V}$ then $\mathcal{U}$ is finite dimensional and $\dim\mathcal{U} \leq \dim\mathcal{V}$.
- Let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of a finite-dimensional vector space $\mathcal{V}$. Then $\dim(\mathcal{U} \cap \mathcal{W}) + \dim(\mathcal{U} + \mathcal{W}) = \dim(\mathcal{U}) + \dim(\mathcal{W})$.

3.4 Linear transformations

Properties of all linear transformations
Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the same field $\mathbb{F}$. Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then:

1. T(cv) = cT(v) for all $c \in \mathbb{F}$ and $v \in \mathcal{V}$.
2. T(0) = 0
3. $T(a_1v_1 + a_2v_2 + ... + a_nv_n) = a_1T(v_1) + a_2T(v_2) + ... + a_nT(v_n)$ for all $a_1, a_2, ..., a_n \in \mathbb{F}$ and $v_1, v_2, ..., v_n \in \mathcal{V}$.

Properties of basis representation functions
$[\cdot]_\beta : \mathcal{V} \to \mathbb{F}^n$ is a linear isomorphism and $\mathcal{V}$ and $\mathbb{F}^n$ are isomorphic.

Dimension Theorem: Given $\mathcal{V}$, $\mathcal{W}$ and T, we know ker(T) and T($\mathcal{V}$) are vector spaces.

- dim$\mathcal{V} < \infty$ iff ker(T) and T($\mathcal{V}$) are finite dimensional
- If those three are finite dimensional, then

$$\text{dim}\mathcal{V} = \text{dimker(T)} + \text{dimT}(\mathcal{V})$$

Note that the rank-nullity theorem is the Dimension Theorem applied to matrices.

Injectivity of T:
For T: $\mathcal{V} \to \mathcal{W}$

$$\text{T is injective} \leftrightarrow \text{ker(T)} = \{\mathbf{0}\} \leftrightarrow \text{dimkerT} = 0.$$

For T: $\mathcal{V} \to \mathcal{W}$, where $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional,

$$\text{T is injective} \leftrightarrow \text{T is surjective} \leftrightarrow \text{T is bijective}$$

Note that both last iff statements follow from the definitions of dimker and bijectivity. We include them out of completeness.

Properties of the adjoint of T:
If $\mathcal{V}, \mathcal{W}$ are finite-dimensional inner product spaces and $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ then:
ker T = (ranT*)$^\perp$ and ran T = (kerT*)$^\perp$.

3.5 Inner products and norms

Inner product for one vector space is also one for its subspaces
If $\mathcal{V}$ is an $\mathbb{F}$-inner product space with inner product $< \cdot, \cdot >$ and $\mathcal{W} \subset \mathcal{V}$ is an $\mathbb{F}$-subspace, then it is an inner product space as well, with the same inner product as $\mathcal{V}$ has.

Perpendicular vectors

- $v \perp w \leftrightarrow w \perp v$
- $0 \perp v$ for all v in the inner product space.

Injectiveness of inner product
$< v, u >=< v, w > \rightarrow u = w$
Fancy way of saying this: the map from $\mathcal{V}$ to the set of all linear maps $\mathcal{V} \rightarrow \mathbb{F}$ is injective. This is a map $u \in \mathcal{V} \mapsto [v \mapsto < v, u >]$.

Theorem of Pythagoras
If $v \perp w$, then $f : v \mapsto \sqrt{< v, v >}$ satisfies $f(v+w)^2 = f(v)^2 + f(w)^2$.

Cauchy-Schwarz-Bunyakowski
$| < v, w > | < f(v)f(w)$ with $f : v \mapsto < \sqrt{< v, v >}$ So: $| < v, w >< ||v||||w||$ with the norm derived from the inner product.

Norms arising from inner products, parallelogram rule
Any inner product space $(\mathcal{V}, < \cdot, \cdot >)$ is also a normed vector space, namely by taking $||v|| := \sqrt{< v, v >}$.
A specific norm on vector space $\mathcal{V}$ arises from some inner product on $\mathcal{V}$ iff $|| \cdot ||$ satisfies the **parallelogram law**:

$$||v + w||^2 - ||v - w||^2 = 2||v||^2 + 2||w||^2$$

This also means that the function $f : v \mapsto \sqrt{< v, v >}$ satisfies $f(v+w)^2 + f(v-w)^2 = 2f(v)^2 + 2f(w)^2$.

Triangle inequality for a derived norm
For a norm arising from an inner vector,

$$||u + v|| \leq ||u|| + ||v||$$

with equality iff one of the vectors is a real, nonnegative scalar multiple of the other. (Note the inequality holds for all norms.)

Benefit of orthonormal basis: expressing vectors in terms of an orthonormal basis
Given vector space $\mathcal{V}$ with $\dim\mathcal{V} = n < \infty$, orthonormal basis β and $v \in \mathcal{V}$, we have $[v]_\beta = \begin{bmatrix} < v, e_1 > \\ < v, e_2 > \\ ... \\ < v, e_n > \end{bmatrix}$

3.6 Minimum norm solutions and least squares solutions

Suppose $y \in \mathbb{F}^m, A \in \mathbb{M}_{m\times n}(\mathbb{F})$ and Ax = y is consistent. Then:

- The equation Ax = y has a unique minimum solution $s \in$colA$*$
- s is the only solution to Ax = y that lies in colA*
- AA*u = y is consistent. If $u_0 \in \mathbb{F}^m$ and AA*u_0 = y, then s = A*$u0$.

3.7 Singular value decomposition

Singular values and Σ:

- The singular values of a matrix are unique.
- The singular values of a matrix is greater than or equal to 0, so Σ is unique..
- If r = rankA, then A has r nonzero singular values.

AA^T and A^TA:

- nullA^TA = null A

V, W and A: We can decompose V and W in the following way: $W = \begin{pmatrix} W_1 & W_2 \end{pmatrix}$ and $V = \begin{pmatrix} V_1 & V_2 \end{pmatrix}$

- V and W are not unique in general.
- V and W are unitary ($VV^T = I$)
- $AW_2 = 0$
- $AW_1W_1^T = A$
- ranA = ranU_1
- ranA^T = ranV_1
- nullA = null V_1^T = ran V_2
- nullA^T = null U_1^T = ran U_2

SVD Theorem:
Let $A \in \mathbb{R}^{m \times n}$ be a matrix of rank r. There exist $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r 0$ and orthogonal $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ such that

$$A = U \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & 0^{r \times n-r} \\ & & & \sigma_r & \\ & 0^{m-r \times r} & & & 0^{m-r \times n-r} \end{bmatrix} \tag{1}$$

If rankA = n (full column rank), then the two zero matrices on the right are absent.
If rankA = m (full row rank) then the bottom two zero matrices are absent.
If A is square and nonsingular, all zero matrices are absent.

3.8 Other

Properties of the orthogonal complement, assuming nonemptiness:
For $\mathcal{U}$ and $\mathcal{W}$ (nonempty) subsets of inner product space $\mathcal{V}$:

- $\mathcal{U}^\perp$ is a subspace of $\mathcal{V}$, even if $\mathcal{U}$ isn't.
- $\mathcal{U} \subseteq \mathcal{W} \implies \mathcal{W}^\perp \subseteq \mathcal{U}^\perp$
- $\mathcal{U}^\perp = \text{span}(\mathcal{U})^\perp$
- If $\mathbf{0} \in \mathcal{W}$ then $\mathcal{W} \cap \mathcal{W}^\perp = \{\mathbf{0}\}$
- $\{\mathbf{0}\}^\perp = \mathcal{V}$
- $\{\mathbf{0}\} = \mathcal{V}^\perp$
- $\mathcal{U} \subseteq \text{span } \mathcal{U} \subseteq (\mathcal{U}^\perp)^\perp$, with equality between the latter two if span($\mathcal{U}$) is finite dimensional.

Not assuming nonemptiness:

- If $\mathcal{U}$ is a finite-dimensional subspace, then $\mathcal{U} = (\mathcal{U}^\perp)^\perp$
- $\mathcal{W} + \mathcal{W}^\perp = \mathcal{V}$, provided $\mathcal{W}$ is a finite-dimensional subspace of $\mathcal{V}$. Furthermore, then for all $v \in \mathcal{V}$ $\exists$ $u \in \mathcal{V}$ such that $v - u \in \mathcal{U}^\perp$. In other words, every vector in $\mathcal{V}$ can be expressed uniquely as a sum of two vectors, one from $\mathcal{U}$ and one from $\mathcal{U}^\perp$ where the vector in $\mathcal{U}$ is the orthogonal projection of v onto $\mathcal{U}$.

Riesz Representation Theorem:
Let $\mathcal{V}$ be a finite dimensional $\mathbb{F}$-inner product space and let $\phi : \mathcal{V} \mapsto \mathbb{F}$ be a linear functional. Then:

- There is a unique $w \in \mathcal{V}$ such that $\phi(v) =< v, w >$ for all $v \in \mathcal{V}$.
- Let $\{u_1, u_2, ..., u_n\}$ be an orthonormal basis for $\mathcal{V}$. The vector w from the above bullet point is
  $w = \overline{\phi(u_1)}u_1 + \overline{\phi(u_2)}u_2 + ... + \overline{\phi(u_3)}u_3$

w does not depend on the choice of orthonormal basis.

Lyuapunov's theorem:
A system of differential equations $\frac{dx}{dt}(t) = Ax(t)$ is asympototically stable if and only if there exists a positive definite matrix $P \in \mathbb{R}^{n\times n}$ such that $A^TP+PA < 0$.

Properties of ... definite matrices
Let $A \in \mathbb{F}^{n\times n}$ be a Hermitian matrix with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{R}$. We have:

- $A \geq 0$ iff all eigenvalues are greater than or equal to zero.
- $A > 0$ iff all eigenvalues are greater than zero.
- $A \leq 0$ iff all eigenvalues are less than or equal to zero.
- $A < 0$ iff all eigenvalues are less than zero.
- A is indefinite iff there is an eigenvalue greater than zero and one less than zero.
- $A \geq 0$ iff there exists a matrix $B \in \mathbb{F}^{n\times n}$ such that $A = B^*B$
- $A > 0$ iff there is a B such as above that is also nonsingular.
- $A > 0$ means A is nonsingular.
- $A \leq 0$ iff $-A \geq 0$ and $A < 0$ iff $-A > 0$.
- If $A \geq 0$ then the following holds: $< Ax, x >$ iff $Ax = 0$.
- If $A \geq 0$ then it has a unique positive semidefinite square root.

4 Methods

Finding norms:

- Euclidian norm for a vector (vector 2-norm): $\sqrt{\sum_i x_i^2}$
- Frobenius norm for a matrix: $\sqrt{\sum_{i,j} |A_{ij}|^2}$
- Matrix 2-norm: $||M||_2 = \sigma_1$, where σ_1 is M's largest singular value.

Matrix representation of a linear transformation
Let $\mathcal{V}$ and $\mathcal{W}$ be finite dimensional vector spaces over the same field $\mathbb{F}$ with bases β and γ respectively, where $|\beta| = n$ and $|\gamma| = m$. Let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$.
Now we have three different mappings: $[\cdot]_\beta$ from $\mathcal{V}$ to $\mathbb{F}^n$; $[\cdot]_\gamma$ from $\mathcal{W}$ to $\mathbb{F}^m$; and T from $\mathcal{V}$ to $\mathcal{W}$.
There exists a mapping from $\mathbb{F}^n$ to $\mathbb{F}^m$ as well, in the form of an $m \times n$ matrix. This mapping $M : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation, given by $M(c) = Ac$ where $c \in \mathbb{F}^n$. A is called the matrix representation of T with representation with respect to the bases β and γ.

In easier words: M, takes a vector in one "language" (basis), looks what T will do to change it, and expresses the result in another language.

To find the matrix representation, perform T on the first vector in the start basis. Next, express the result of that as a vector in the end basis. For example, if you have the start basis ($\begin{bmatrix}1\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\end{bmatrix}$), and end basis ($\begin{bmatrix}2\\0\end{bmatrix}$, $\begin{bmatrix}0\\-4\end{bmatrix}$), and $T(v) = \begin{bmatrix}6v_1\\2v_2\end{bmatrix}$, then $T(\begin{bmatrix}1\\0\end{bmatrix}) = \begin{bmatrix}6\\0\end{bmatrix}$.
In the end basis, this would be expressed as $\begin{bmatrix}3\\0\end{bmatrix}$. $T(\begin{bmatrix}0\\1\end{bmatrix}) = \begin{bmatrix}0\\2\end{bmatrix}$, which in our end basis is represented by $\begin{bmatrix}0\\-0.5\end{bmatrix}$. So our matrix representation is $\begin{bmatrix}3 & 0\\0 & -0.5\end{bmatrix}$.
The matrix ${}_\gamma[\cdot]_\beta$ is the matrix representation of the identity transformation: basically, it sends vectors from basis β to basis γ without changing anything else about them.

Finding dimensions: Often, the dimension theorem can help you find the dimension of a vector space given a convenient linear operator T.

Another way: if we know the dimension of $\mathcal{V}$, we can select a basis for it. Then ${}_\beta[T]_\beta$ will be a 4 by 4 matrix. The rank of this matrix is the dimension of T($\mathcal{V}$). ${}_\gamma[T]_\gamma =_\gamma [id]_\beta \, {}_\beta[T]_\beta \, {}_\gamma[id]_\gamma$. This means ${}_\gamma[T]_\gamma$ and ${}_\beta[T]_\beta$ are similar since ${}_\beta[id]_\gamma$ and ${}_\gamma[id]_\beta$ are each other's inverse.

Gram Schmidt: The Gram-Schmidt process turns a basis for an inner product space into an orthonormal basis.
Take basis

$$\{v_1, v_2, \ldots, v_n\}$$

and construct a new one

$$\{u_1, u_2, \ldots, u_n\}$$

by setting

$$u_1 = \frac{x_1}{||x_{1||}}$$

. Then for each k = 2,3,...,n compute:

$$x_k = v_k - < v_k, u_1 > u_1 - \cdots - < v_k, u_{k-1} > u_{k-1}$$

Normalise it:

$$u_k = \frac{x_k}{||x_{k||}}$$

Finding the best approximation/orthogonal projection:
By the Best Approximation Theorem, the orthogonal projection onto a subspace is the best approximation in that subspace.
Let $\mathcal{U}$ be a finite-dimensional subspace of an inner product space $\mathcal{V}$.
With an orthonormal basis: If we have an orthonormal basis $u_1, ..., u_r$ for $\mathcal{U}$, $\mathcal{P}_{\mathcal{U}}(v) = \sum_{i=1}^{r} < v, u_i > u_i$.
Without an orthonormal basis: If we have as spanning set $u_1, ..., u_n$ which is not an orthonormal basis, we can compute $\mathcal{P}_{\mathcal{U}}(v) = \sum_{j=1}^{n} c_j u_j$ where the c_j solve

$$\begin{bmatrix} < u_1, u_1 > & < u_2, u_1 > & ... & < u_n.u_1 > \\ < u_1, u_2 > & < u_2, u_2 > & ... & < u_n, u_2 > \\ \vdots & \vdots & \ddots & \vdots \\ < u_1, u_n > & < u_2, u_n > & ... & < u_n, u_n > \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} < v, u_1 > \\ \vdots \\ < v, u_n > \end{bmatrix}$$

This system is always consistent. If $u_1, u_2, ..., u_n$ are linearly independent it has a unique solution. The first matrix is called the Gram matrix or Grammian of $u_1, ..., u_n$. The Gram determinant of $u_1, ..., u_n$ is the determinant of the Gram matrix.

Finding a least squares solution to an inconsistent linear system:
For the system Ax=y, x_0 is a least squares solution iff $A^*Ax_0 = A^*y$.
This system is always consistent and has a unique solution if rankA = n.
This is basically a best approximation problem where $\mathcal{U}$ is the range or column space of the matrix A.

Solving a system of linear equations with a positive definite coefficient matrix, Cholesky factorisation: Solving a system with a lower or an upper triangular matrix is easily done via forward/back substitution. Cholesky factorization allows us to write any positive semidefinte matrix as the product of a lower and upper triangular matrix.
For such a matrix P, we can find matrices such that $P = LL^*$.
To find L, first we apply QR-factorization, which gives us $P = QR$ where Q is unitary and R is upper triangular. Then $L = R^*$.

Finding the singular value decomposition:
For an m by n matrix with, we can find matrices such that:
$A = V\Sigma W^*$,
where $V^{m\times m}$ and $W^{n\times n}$ are orthogonal and Σ is the same shape as A. Σ is a diagonal matrix with A's singular values along its diagonal.
Case 1: Tall or square matrix

1. First, find the eigenvalues of A^TA. Order these by size, such that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$. Then find the eigenvectors of A^TA. Preferably, these should be orthogonal.
2. If the eigenvectors are not orthogonal, perform Gram-Schmidt to get orthonormal vectors.
3. The singular values of A are the square roots of the eigenvalues we just found. There are min(m,n) - r singular values.
4. If we performed Gram-Schmidt: those are the columns of W. If we didn't: normalise the eigenvectors of A^TA and those will be the columns of W.
5. Each column of V is $v_i = \frac{1}{\sigma_i}Aw_i$. These will be orthonormal vectors
6. **Alternative method to find V:** Find the eigenvalues and eigenvectors of AA^T. Order the eigenvalues by size again. Orthonormalise these vectors to get the columns of V.

Case 2: wide matrix:
To find the SVT of A with rank r, we find it for A^T. Then $A^T = \overline{U}\begin{pmatrix}\overline{\Sigma_1} & 0\\ 0 & 0\end{pmatrix}\overline{V}^T$
where $\overline{\Sigma_1} = \begin{pmatrix}\overline{\sigma_1} & & & \\ & \overline{\sigma_2} & & \\ & & \ddots & \\ & & & \overline{\sigma_r}\end{pmatrix}$.

The compact form of an SVD, outer product expansion:
We can partition V, W and Σ as $\begin{pmatrix} U_1 & U_2 \end{pmatrix}$, $\begin{pmatrix} V_1 & V_2 \end{pmatrix}$ and Σ_1 (which is the square part of Σ with nonzero diagonal entries).
Then the compact SVD of A is $U_1 \Sigma \, {}_{V_1^T}$.
Alternatively, if $U_1 = \begin{pmatrix} u_1 & u_2 & ... & u_r \end{pmatrix}$ and $V_1 = \begin{pmatrix} v_1 & v_2 & ... & v_r \end{pmatrix}$ then A = $\sum_{i=1}^{r} \sigma_i u_i v_i^T$ is called the outer product expansion of A. It is a rank-1 matrix.

Finding the left inverse of a matrix using the SVD:
If A is full column rank, it has a left inverse.

If we have made an SVD of $A = U\Sigma V^T$, then $A^{=V} \begin{bmatrix} \Sigma_1^{-1} & 0 \end{bmatrix} U^T$.
In general, A is not unique.

Checking whether something is an SVD:
Given matrices $U\Sigma V^T$, we need to check the following to see whether it is an SVD of A:

- When you multiply the matrices out, you actually get A.
- U and V are orthogonal ($U^T U = I$)
- the columns of U are eigenvectors for AA^T and the columns of V are eigenvectors for $A^T A$
- The diagonal elements of Σ are singular values for A.

Finding a low rank approximation of a matrix:
For a matrix of rank r, we may want to know how "far" away it is from having a lower rank.
In other words, we want to find $d(A, M_k)$. Moreover, we would like to find a matrix $X \in M_k$ such that $||A - X||_2 = d(A, M_k)$.

How far away is a matrix from being singular? and finding the best approximation of A in M_k
S is singular $\leftrightarrow$ $rankS \leq n - 1$. So the distance of a matrix A to singular matrices is $d(A, M_{n-1}$.
If we sort the singular values of A by size ($\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r \geq 0$, $d(A, M_k) = \sigma_{k+1}$.
If we have a singular value decomposition of A $U\Sigma V^T, X = U\Sigma_k V^T$ (own notation for Σ_k), where Σ_k is Σ, with all singular values up to σ_k present, and the rest of the elements equal to zero.

Using the outer product expansion, $X = \sum_{i=1}^{k} \sigma_i u_i v_i^T$

---

5 Notable vector spaces

- $\mathbb{F}^n$ with $u+v$ and $c \cdot v$ defined in the usual way.. If $c \in \mathbb{F}$, this is a vector space over $\mathbb{F}$.
- The set $\mathbb{F}^{n \times m}$ with $A+B$ and $c \cdot A$ defined as the usual matrix addition and scalar multiplication. If $c \in \mathbb{F}$, this is a vector space over $\mathbb{F}$.
- P_n with $(p+q)(x) = p(x) + q(x)$ and $(c \cdot p)(x) = cp(x)$. If the coefficients are complex and $c \in \mathbb{C}$, this is a vector space over $\mathbb{C}$.
- $C[a,b]$ with $(f+g)(x) = f(x) + g(x)$ and $(c \cdot f)(x) = cf(x)$. If $a, b \in \mathbb{R}$ and $c \in \mathbb{R}$, this is a vector space over $\mathbb{R}$.
- $\mathcal{L}(\mathcal{V}, \mathcal{W})$

6 Notable subspaces

- For a matrix $A \in \mathbb{F}^{n \times m}$, nullA is a subspace of $\mathbb{F}^m$
- For a matrix $A \in \mathbb{F}^{n \times m}$, colA is a subspace of $\mathbb{F}^n$.
- $\{p_0 + p_1 | p_0, p_1 \in \mathbb{C}\} \subseteq P_n$ is a subspace of P_n.
- $\{f \in C[a,b] | f(a) = 0\} \subseteq C[a,b]$ is a subspace of $C[a,b]$.
- The span of a collection of vectors is always a subspace.
- For a vector space $\mathcal{V}$ over $\mathbb{F}$ and subspaces $\mathcal{U}$ and $\mathcal{W}$ of $\mathcal{V}$, the intersection $\mathcal{U} \cap \mathcal{W}$ is a subspace of $\mathcal{V}$.
- If $\mathcal{U}$ and $\mathcal{W}$ are subspaces of $\mathcal{V}$ then $\mathcal{U} + \mathcal{W}$ is a subspace of $\mathcal{V}$.
- kerT is a subspace of the domain
- ranT is a subspace of the codomain
- $\{0\}$ is a subspace
- Any vector space is a subspace of itself.
- Let $\mathcal{V}$ be a vector space and $v_1, v_2, ..., v_r \in \mathcal{V}$. Then span$(v_1, v_2, ..., v_r)$ is a subspace of $\mathcal{V}$.
- For any subset $S \in \mathcal{V}$, $S^\perp$ is a subspace of $\mathcal{V}$ regardless of whether S is.

7 Notable inner products and inner product spaces

- $\mathbb{R}^n$ is an inner product space with the usual dot product as inner product.
- $\mathbb{C}^n$ with the Hermitian form $a \cdot b = \sum_{j=1}^{n} a_j \overline{b_j}$
- $\mathrm{P}_n(\mathbb{R})$ with $< p(t), q(t) >= \int_a^b p(t)q(t)dt$
- $\mathrm{P}_n(\mathbb{C})$ with $< p(t), q(t) >= \int_a^b p(t)\overline{q(t)}dt$
- $\mathrm{P}_n(\mathbb{R})$ with $< p(t), q(t) >= \int_0^\infty p(t)q(t)e^{-t}dt$
- $\mathrm{P}_n(\mathbb{R})$ with $< p(t), q(t) >= \int_0^\pi p(cos(\alpha))q((\alpha))d\alpha$
- C[a,b] with inner product $\int_0^1 f(t)g(t)dt$

8 Notable norms and normed vector spaces

- $\mathbb{R}^n|$ with $||a|| := |a_1| + ... + |a_n|$, or similarly over $\mathbb{C}$.
- $P_n(\mathbb{R}) \to \mathbb{R}$ where $||a_0 + a_1 t + ... + a_n t^n|| := max|f(t)|$ with $0 \leq t \leq 1$.

9 Orthonormal bases

- $\{e_1, e_2, ...e_n\}$ is an orthonormal basis for $\mathbb{R}^n$ with the usual dot product.