Mathematical Fundamentals

Vector Spaces

- Simplest example of a vector space is the *Euclidian* n-dimensional space, $\mathbb{R}^n$, e.g.,

$$\mathbf{u} = (u_1, u_2, ..., u_n) \text{ with } u_i \in \mathbb{R} \text{ for } i = 1, 2, \ldots, n$$

- $\mathbb{R}^n$ exhibits some special properties that lead to a general definition of a vector space:
  - $\mathbb{R}^n$ is an *additive abeliean group*
  - Any n-tuple $k\mathbf{u}$ is also in $\mathbb{R}^n$
  - Given any $k \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^n$ we can obtain by scalar multiplication the unique vector $k\mathbf{u}$

> In abstract algebra, an *abelian group*, also called a commutative group, is a group in which the result of applying the group operation to two group elements does not depend on the order in which they are written (the axiom of commutativity). Abelian groups generalize the arithmetic of addition of integers. They are named after Niels Henrik Abel. [Jacobson, Nathan, *Basic Algebra I*, 2nd ed., 2009]

- Some further results:

- $k(\mathbf{u}+\mathbf{v}) = k\mathbf{u} + k\mathbf{v}$ (i.e., scalar multiplication is distributive under addition)
- $(k+l)\,\mathbf{u} = k\mathbf{u} + l\mathbf{u}$
- $(kl)\,\mathbf{u} = k(l\mathbf{u})$
- $1 \cdot \mathbf{u} = \mathbf{u}$

- To define a vector space in general we need the following four conditions:

1. A set V which forms an additive abeliean group
2. A set F which forms a field
3. A scalar multiplication of $k \in F$ with $\mathbf{u} \in V$ to generate a unique $k\mathbf{u} \in V$
4. Four scalar multiplication axioms from above.

Then we say that V is a *vector space* of F.

Examples:

- All ordered n-tuples, $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ with $u_i \in \mathbb{C}$ forms a vector space $\mathbb{C}^n$ over $\mathbb{C}$
- The set of all $n \times m$ real matrices $\mathbb{R}^{n\times m}$ is a vector space over $\mathbb{R}$

Linear Spans, Spanning Sets and Bases

- Let S be the set S=$\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_m}\} \in V$ where V is a vector space of some F
- Define $L(S)$ =$\{\boldsymbol{v} \in V : \boldsymbol{v} = a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \ldots + a_m\boldsymbol{u_m} \text{ with } a_i \in F\}$
- Thus $L(S)$ is the set of all *linear combinations* of the $\boldsymbol{u_i}$
- Then,

- $L(S)$ is a subspace of V (note, it is closed under vector addition and scalar multiplication)
- $L(S)$ is the smallest subspace containing S

- We say that $L(S)$ is the *linear span* of S, or that $L(S)$ is *spanned* by S
- Conversely, S is called a *spanning set* for $L(S)$

IMPORTANT: A spanning set $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}$ of V is a *basis* if the $\boldsymbol{v_i}$ are linearly independent

NOTE: If $dim(V) = m$, then any set $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n} \in V$ with $n > m$ must be linearly dependent

Linear Transformations, Matrix Representation & Matrix Algebra

Linear Mappings

- Let X, Y be vector spaces over F and let T be a mapping of X into Y,

$$T : X \to Y$$

- Then $T \in L(X, Y)$, the set of linear mappings (or transformations) of X into Y, if

$$T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$$

and

$$T(k\boldsymbol{u}) = k \cdot T(\boldsymbol{u}), \text{ for all } \boldsymbol{u}, \boldsymbol{v} \in X \text{ and } k \in F$$

$\Rightarrow$ This defines a *Linear Tranform*

Matrix Respresentation of Linear Mappings

- Linear mappings can be made concrete and represented by matrix multiplications...

– once $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$ are both expressed in terms of their coordinates relative to appropriate basis sets

- We can now show that A is a matrix representation of the linear transform T:
  – $\mathrm{y} = T(\mathrm{x})$ is equivalent to $\mathrm{y} = A\mathrm{x}$
- Note that when dealing with actual vectors in $\mathbb{R}^{\mathbf{k}}$ or $\mathbb{C}^{\boldsymbol{k}}$, an all-time favorite basis set is the *standard basis set*, which consists of vectors $\boldsymbol{e_i}$ which are zero everywhere except for a "$\mathbf{1}$" appearing in the i^{th} position

<u>*Change of Basis*</u>

- Let $\boldsymbol{e}'_{\mathbf{1}}, \boldsymbol{e}'_{\mathbf{2}}, \ldots, \boldsymbol{e}'_{\boldsymbol{m}}$ be a new basis for X
- Then each $\boldsymbol{e}'_{\boldsymbol{i}}$ is in X and can be written as a linear combination of the $\boldsymbol{e_i}$:

$$\boldsymbol{e}'_{\boldsymbol{i}} = \sum_{j=1}^{n} p_{ij} \boldsymbol{e}_j$$

$$E' = EP$$

where P is a square and invertible matrix with elements p_{ji}

- Thus we can write,

$$E = E' P^{-1}$$

meaning that P^{-1} transforms the old coordinates into the new coordinates.

- We define a *similarity transformation* as

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

Isomorphism. A linear transformation is an isomorphism if the mapping is "one-to-one"

Image. The image of a linear transformation T is,

$\mathrm{Image}(T) = \{y \in Y : T(x) = y,\ x \in X\}$

Kernel. The kernel of a linear transformation T is,

$\ker(T) = \{x \in X : T(x) = 0\}$

- $\mathrm{Image}(T)$ is a subspace of Y (the co-domain) and $\ker(T)$ is a subspace of X (the domain)

Example 1.1

$$A : \mathbb{R}^3 \to \mathbb{R}^2 \text{ with } A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

- Rank of a Matrix
- The *rank* of a matrix is defined as the number of linearly independent columns (or equivalently the number of linearly independent rows) contained in the matrix
- Our definition of the rank of T is consistent with this because in the case of $T \in L(X, Y)$ defined by $\boldsymbol{x} \to A\boldsymbol{x}$, the dimension of $\mathrm{Image}(T)$ is nothing other than the dimension of the linear span of the columns of A, which obviously is equal to the number of linearly independent column vectors
- Note that an alternative but equivalent and perhaps more convenient way to define the rank of a square matrix is by way of *minors:*
  - If an $n \times n$ matrix A has a non-zero $r \times r$ minor, while every minor of order higher than r is zero, then A has rank r.

Example 1.2

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}$$

- Since the dimension of A is 4×3, it's rank can be no larger than 3
- But since column 2 plus column 3 equals column 1, the rank reduces to 2

Corrollary: A square $n \times n$ matrix is full rank (non-singular) if its determinant is non-zero.

- Some helpful properties of determinants:

  Let $A, B \in \mathbb{R}^{\mathrm{n \times n}}$

  1. $\det(A) = 0$ if and only if the columns (rows) of A are linearly dependent
  2. If a column (row) of A is zero, then $\det(A) = 0$
  3. $\det(A) = \det(A^T)$
  4. $\det(AB) = \det(A) \cdot \det(B)$
  5. Swapping two rows (columns) changes the sign of $\det(A)$.
     (a) Scaling a row (column) by k, scales the determinant by k.
     (b) Adding a multiple of one row (column) onto another does not affect the determinant.
  6. If $A = \mathrm{diagonal}\{a_{ii}\}$ or lower triangular, or $A =$ upper triangular with diagonal elements a_{ii}, then $\det(A) = a_{11}, a_{22}, \ldots, a_{nn}$

Define the *adjoint* of a square matirx A as follows:

- Let $\gamma_{i,j} = (-1)^{i+j} det(M_{i,})$ where $M_{i,j}$ is the matrix A with the i^{th}row and j^{th}column removed.
- Then,

$$\mathrm{adj}(A) = \gamma_{i,j}^T$$

- Then the *matrix inverse* is defined as:

$$A^{-1} = \frac{1}{\det(A)} \cdot \mathrm{adj}(A)$$

Eigenvalue & Singular Value Decompositions

- Often in engineering we have to deal with scalar functions of a matrix
- For example

$$G(s) = C(sI - A)^{-1}B$$

where A is a square real matrix, represents the transfer function matrix of a dynamic system with many inputs and many outputs.

- Equally,

$$y = e^{At}\boldsymbol{x_o}$$

is the form of solution of n 1^{st} order ordinary differential equations.

- We know how to handle terms like $\dfrac{1}{s-a}$, whose inverse Laplace Transform is just e^{at}.
- But how do we handle $(sI - A)^{-1}$? And how do we evaluate e^{At}?
  - Remember that an $n \times n$ matrix A can be thought of as a matrix representation of a linear tranformation on $\mathbb{R}^n$, relative to some basis (e.g., the standard basis).

- Yet we know that a change of basis, described in matrix form by $E^{'} = EP$, will transform A to $P^{-1}AP$.

- Suppose then that it were possible to choose our new basis such that $P^{-1}AP$ were a diagonal matrix D.
- This transforms general matrix problems into "diagonal" problems – they consist now of a collection of scalar problems, $1/(s-d_i)$ and $e^{d_i t}$, which we can easily solve.
- Such a basis exists for almost all $A's$ and is defined by the set of *eigenvectors* of A.
- The corresponding diagonal elements d_i of D turn out to be the *eigenvalues* of A.

Eigenvalue Decomposition

- A linear transformation T maps $\boldsymbol{x} \to \boldsymbol{y} = A\boldsymbol{x}$ where in general $\boldsymbol{x}$, $\boldsymbol{y}$ have different directions
- There exist directions $\boldsymbol{w}$ however which are invariant under the mapping

$$A\boldsymbol{w} = \lambda\boldsymbol{w} \text{ where } \lambda \text{ is a scalar.}$$

- Such A-invariant directions are called *eigenvectors* of A and the corresponding λ's are called *eigenvalues* of A; *eigen* in German means "self".
- The above equation can be re-arranged as:

$$(\lambda I - A)\,\boldsymbol{w} = \mathbf{0}$$

and thus implies that $\boldsymbol{w}$ lies in the kernel of $(\lambda I - A)$.

- A non-trivial solution for $\boldsymbol{w}$ can be obtained if and only if, $\ker(\lambda I - A) \neq \{\mathbf{0}\}$, which is only possible if $(\lambda I - A)$ is singular, i.e.,

$$\det(\lambda I - A) = 0$$

- This equation is called the *characteristic equation* of A, and defines all possible values for the eigenvalues λ
- The λ may assume one of the n roots, λ_i of the characteristic equation, and are thus called *characteristic roots* or simply the *eigenvalues* of A.
- Interesting and important properties of the eigenvalues of a matrix A:
  - $\mathrm{Trace}(A) \equiv$ sum of the diagonal elements of $A = \sum_{i=1}^{n} \lambda_i =$ sum of the eigenvalues
  - $\det(A) \equiv$ determinant of $A = \prod_{i=i}^{n} \lambda_i =$ product of the eigenvalues
- Once the values of λ have been determined, the corresponding eigenvectors may be calculated by solving the set of homogeneous equations:

$$(\lambda I - A)\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = 0 \quad \text{where } w_i \text{ denotes the } i^{th} \text{ element of } \boldsymbol{w}$$

NOTE: If $A\boldsymbol{w} = \boldsymbol{w}$, then $A(k\boldsymbol{w}) = kA\boldsymbol{w} = k\lambda\boldsymbol{w} = \lambda(k\boldsymbol{w})$, so that if $\boldsymbol{w}$ is an eigenvector of A, then so also will be any scalar multiple of $\boldsymbol{w}$ – the important property of $\boldsymbol{w}$ is its *direction*, not its scaling or length.

Example:

$$A = \begin{bmatrix} 1 & -3 & 5 \\ 0 & 1 & 2 \\ 0 & -1 & 4 \end{bmatrix}$$

so that,

$$\det(\lambda I - A) = \lambda^3 - 6\lambda^2 + 11\lambda - 6 = 0$$

IMPORTANT EIGENVECTOR/EIGENVALUE PROPERTIES:

1. The eigenvalues of a diagonal or triangular matrix are equal to the diagonal elements of the matrix.
2. The eigenvectors of a diagonal matrix (with distinct diagonal elements) are the standard basis vectors.
3. The eigenvalues of the identity matrix are all "1" and the eigenvectors are arbitrary.
4. The eigenvalues of a scalar matrix kI are all k and the eigenvectors are arbitrary.
5. Multiplying a matrix by a scalar k has the effect of multiplying all its eigenvalues by k , but leaves the eigenvectors unaltered.
6. (Shift Theorem) Adding onto a matrix a scalar matrix, kI, has the effect of adding k to all its eigenvalues, but leaves the eigenvectors unaltered.
7. The eigenvectors of a matrix and its transpose are the same.
8. The eigenvalues/eigenvectors of a real matrix are real or appear in complex conjugate pairs.

Characteristic Decomposition

- By convention, we collect all the eigenvectors of an $n \times n$ square matrix A as the column vectors of the *eigenvector matrix*, W:

$$W = [\boldsymbol{w_1}, \boldsymbol{w_2}, \cdots, \boldsymbol{w_n}]$$

- Upon inversion we can obtain the dual set of *left eigenvectors* as the row vectors of the *dual eigenvector matrix*, $V = W^{-1}$:

$$V = W^{-1} = \begin{bmatrix} \boldsymbol{v}_1^t \\ \boldsymbol{v}_2^t \\ \vdots \\ \boldsymbol{v}_n^t \end{bmatrix}$$

- Combining the eigenvector and dual eigenvector matrices together we get the eigenvalue/eigenvector decomposition (or characteristic decomposition):

$$AW = W\Lambda$$

  where Λ is a diagonal matrix containing the eigenvalues, λ_i, of A. Post-mulitplying the above by W^{-1} we obtain:

$$A = W\Lambda W^{-1} = W\Lambda V$$

$\Rightarrow$ This is also known as the *spectral decomposition* of A.

- The eigen-decomposition can be ill-conditioned with respect to its computation as shown in the following example.

Example

$$A = \begin{bmatrix} 1 & \epsilon \\ 0 & 1.001 \end{bmatrix}$$

where by inspection, $\lambda_1 = 1$ and $\lambda_2 = 1.001$.

- It is easy to see that,

$$\begin{bmatrix} 1 & \epsilon \\ 0 & 1.001 \end{bmatrix} \begin{matrix} 1 \\ 0 \end{matrix} = 1 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- Hence the first eigenvector is $\mathrm{w}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

- From,

$$\begin{bmatrix} 1 & \epsilon \\ 0 & 1.001 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a + \epsilon b \\ 1.001b \end{bmatrix} = (1.001) \begin{bmatrix} a \\ b \end{bmatrix}$$

we get,

$$\begin{aligned} \epsilon b &= .001a \\ 1.001b &= 1.001b \end{aligned}$$

which gives,

$$a = 1000\epsilon b$$

- Letting $b = 1$, then the second eigenvector is $\mathrm{w}_2 = \begin{bmatrix} 1000\epsilon \\ 1 \end{bmatrix}$.

- The w_2 eigenvector has a component which varies 1000 times faster than ϵ in the A matrix.

- In general, eigenvectors can be extremely sensitive to small changes in matrix elements when the eignvalues are clustered closely together.

Some Special Matrices

REAL SYMMETRIC MATRIX: $S^T = S$

- Here S has real eigenvalues, λ_i and hence real eigenvectors, $\boldsymbol{w_i}$, where the eigenvectors are orthonormal (i.e., $\boldsymbol{w_j^t} \cdot \boldsymbol{w_i} = \boldsymbol{0}$, for all

$i \neq j$) . Therefore its eigenvalue/eigenvector decomposition becomes:

$$S = R\Lambda R^T,$$

where $R^T R = RR^T = I$.

Corrollary: Any quadratic form, $\sum a_{ij} x_i x_j$ can be written as $\boldsymbol{x}^t S \boldsymbol{x}$ and is positive definite (i.e., is positive for $\boldsymbol{x} \neq 0$) if and only if $\lambda_i(S) > 0$ for all i.

NORMAL MATRIX: $N^* N = N N^*$

If $(\lambda_i,\ \boldsymbol{w}_i)$ is an eigen-pair of N, then $\left(\bar{\lambda}_i,\ \boldsymbol{w}_i\right)$ is an eigen-pair of N^*

The eigenvectors of N are orthogonal.

HERMITIAN MATRIX: $H^* = H$

H has real eigenvalues and orthogonal eigenvectors.

UNITARY MATRIX: $U^* U = U U^* = I$

The eigenvalues of U have unit modulus and the eigenvectors are orthogonal.

Jordan Form

- If the $n \times n$ matrix A cannot be diagonalized, then it can always be brought into *Jordan* form via a similarity transformation,

$$T^{-1} A T = J = \begin{bmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_q \end{bmatrix}$$

where

$$J_i = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix} \in \mathbb{R}^{n_i \times n_i}$$

is called a Jordan block of size n_i with eigenvalue λ_i. Note that J is block-diagonal and upper bi-diagonal.

Singular Value Decomposition

- Despite its extreme usefulness, the eignenvalue/eigenvector decomposition suffers from two main drawbacks:
  1. It can only handle square matrices
  2. Its computation may be sensitive to even small errors in the elements of a matrix
- To overcome both of these we can instead use the Singular Value Decomposition (when appropriate).
- Let M be any matrix of dimension $p \times m$. It can then be factorized as:

$$M = U \Sigma V^*$$

where $UU^* = I_p \, and \, VV^* = I_m$. Here, Σ is a rectangular matrix of dimension $p \times m$ which has non-zero entries only on its leading diagonal:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N & 0 & \cdots & 0 \end{bmatrix}$$

and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N \geq 0$.

- This factorization is called the *Singular Value Decomposition* (SVD).

- The σ's are called the singular values of M.
- The number of positive (non-zero) singular values is equal to the rank of the matrix M.
- The SVD can be computed very reliably, even for very large matrices, although it is computationally heavy.
  - In MATLAB it can be obtained by the function svd: [U,S,V] = svd(M).
- The largest singular value of M, σ_1, (also denoted, $\bar{\sigma}$) is an induced norm of the matrix M:

$$\sigma_1 = \sup_{\mathbf{x}} \frac{\|M\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$$

- It is important to note that,

$$MM^* = U\Sigma^2U^*$$

and

$$M^*M = V\Sigma^2V^*$$

- These are eigen-decompositions since $U^*U = I$ and $V^*V = I$ with Σ diagonal.

SOME PROPERTIES

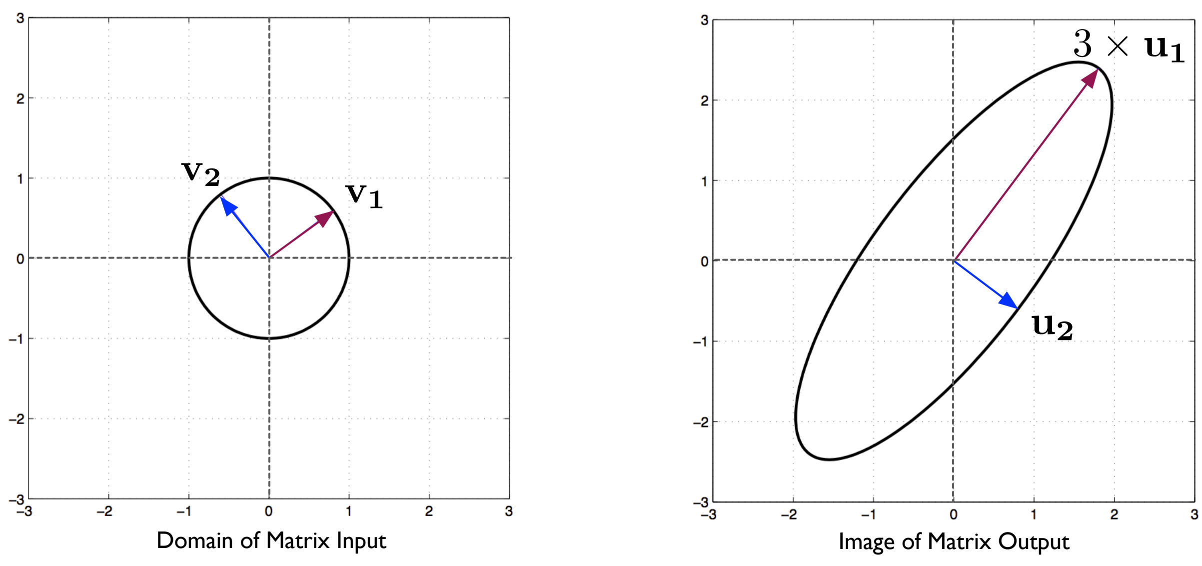
- $\sigma_i = \sqrt{\lambda_i(M^*M)} = \sqrt{\lambda_i(MM^*)}$= i^{th} *singular value* of M
- $\boldsymbol{v}_i = \boldsymbol{w}_i(M^*M)$ = i^{th} *input principal direction* of M
- $\boldsymbol{u}_i = \boldsymbol{w}_i(MM^*)$ = i^{th} *output principal direction* of M

Example (Golub & Van Loan):

$$A = \begin{bmatrix} 0.96 & 1.72 \\ 2.28 & 0.96 \end{bmatrix} = U\Sigma V^T$$

$$= \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{bmatrix}^T$$

- Any real matrix, viewed geometrically, maps a unit-radius (hyper)sphere into a (hyper)ellipsoid.
- The singular values σ_i give the lengths of the major semi-axes of the ellipsoid.
- The output principal directions $\boldsymbol{u}_i$ give the mutually orthogonal directions of these major axes.
- The input principal directions $\boldsymbol{v}_i$ are mapped into the $\boldsymbol{u}_i$ vectors with gain σ_i such that, $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$.



- Numerically, the SVD can be computed much more accurately than the eigenvectors, since only orthogonal transformations are used in the computation.

Vector & Matrix Norms

Vector Norms

A vector norm on $\mathbb{C}$ is a function $f : \mathbb{C} \to \mathbb{R}$ with the following properties:

- $f(\boldsymbol{x}) \geq 0 \quad \forall \quad \boldsymbol{x} \in \mathbb{C}^n$
- $f(\boldsymbol{x}) = 0 \Leftrightarrow \boldsymbol{x} = 0$
- $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y}) \quad \forall \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$
- $f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x}) \quad \forall \quad \alpha \in \boldsymbol{C},\ \boldsymbol{x} \in \mathbb{C}^n$

The *p-norms* are defined by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Three norms of particular interest to control theory are:

- 1-NORM

$$\|\boldsymbol{x}\|_1 = \sum_{i-1}^{n} |x_i|$$

- 2-NORM

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i-1}^{n} |x_i|^2} = \sqrt{\boldsymbol{x}^* \boldsymbol{x}}$$

- ∞-NORM

$$\|\boldsymbol{x}\|_\infty = \max_i |x_i|$$

The following relationships hold for 1, 2, and ∞ norms for vectors:

- $\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_1 \leq \sqrt{n}\, \|\boldsymbol{x}\|_2$

- $\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_2 \le \sqrt{n}\,\|\boldsymbol{x}\|_\infty$
- $\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_1 \le n\,\|\boldsymbol{x}\|_\infty$

Matrix Norms

Matrix norms satisfy the same properties outlined above for vector norms. The matrix norms normally used in control theory are also the 1-norm, 2-norm and ∞-norm.

- Another useful norm is the *Frobenius* norm, defined as follows:

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} \qquad \forall \quad A \in \mathbb{C}^{m\times n}$$

Here a_{ij} denotes the element of A in the i^{th} row, j^{th} column.

- The p-norms for matrices are defined in terms of *induced* norms, that is they are induced by the p-norms on vectors.
- One way to think of the norm $\|A\|_p$ is as the maximum gain of the matrix A as measured by the p-norms of vectors acting as inputs to the matrix A:

$$\|A\|_p = \sup_{\boldsymbol{x}\neq 0}\frac{\|A\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p} \quad \forall \quad A \in \mathbb{C}^m$$

- The matrix norms are computed by the following expressions:

$$\|A\|_1 = \max_j \sum_{i=1}^{m}|a_{ij}|$$

$$\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

$$\|A\|_2 = \bar{\sigma}(A)$$

SOME USEFUL RELATIONSHIPS:

- $\|AB\|_p \leq \|A\|_p \|B\|_p$
- $\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2$
- $\max_{i,j} |a_{ij}| \leq \|A\|_2 \leq \sqrt{mn} \max_{i,j} |a_{ij}|$
- $\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$
- $\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\,\|A\|_\infty$
- $\frac{1}{\sqrt{m}} \|A_1\| \leq \|A\|_2 \leq \sqrt{n}\,\|A\|_1$

Toeplitz & Hankel Matrix Forms

Model-based Predictive Control formulations make extensive use of Toeplitz and Hankel matrices to simplify much of the algebra.

Consider the polynomial $n(z)$:

$$n(z) = n_0 + n_1 z^{-1} + \cdots + n_m z^{-m}$$

Then we define the Toeplitz matrices, Γ_n, C_n for $n(z)$ from the following matrix form:

$$\Gamma_n = \left[\frac{C_n}{M_n} \right]$$

where,

$$C_n = \begin{bmatrix} n_0 & 0 & 0 & \cdots & 0 \\ n_1 & n_0 & 0 & \cdots & 0 \\ n_2 & n_1 & n_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n_m & n_{m-1} & n_{m-2} & \vdots & \end{bmatrix}$$

and

$$M_n = \begin{bmatrix} 0 & n_m & n_{m-1} & \vdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & n_0 \end{bmatrix}$$

Define the Hankel matrix H_n as

$$H_n = \begin{bmatrix} n_1 & n_2 & n_3 & \cdots & n_{m-1} & n_m \\ n_2 & n_3 & n_4 & \cdots & n_m & 0 \\ n_3 & n_4 & n_5 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n_{m-1} & n_m & 0 & \vdots & 0 & 0 \\ n_m & 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Note: The dimension of matrices Γ_n, C_n, H_n are not defined here as they are implicit in the context in which they are used.

Toeplitz matrices defined in this manner are used for changing polynomial convolution into a matrix-vector multiplication. As might be guessed, the relationship between matrix/vector multiplication and polynomial operations is very useful for analysis of predictions.

Analytic Functions of Matrices

Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, and let $p(x)$ be an infinite series in a scalar variable x,

$$p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_k x^k + \ldots$$

Such an infinite series may converge or diverge depending on the value of x.

EXAMPLES

$$1 + x + x^2 + x^3 + \ldots + x^k + \ldots$$

- This geometric series converges for all $|x| < 1$.

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} + \cdots$$

- This series converges for all values of x and is in fact, e^x.

These results have analogies for matrices.

Let A be an $n \times n$ matrix with eigenvalues λ_i. If the infinite series $p(x)$ defined above is convergent for each of the n values of λ_i, then the corresponding matrix inifinite series

$$p(A) = a_0 I + a_1 A + a_2 A^2 + \cdots + a_k A^k + \cdots = \sum_{k=0}^{\infty} a_k A^k$$

converges.

DEFINITION: A single-valued function $f(z)$, with z a complex scalar, is analytic at a point z_0 if and only if its derivative exists at every point in some neighborhood of z_0. Points at which the function is not analytic are called s*ingular points*.

RESULT. If a function $f(z)$ is analytic at every point in some circle Ω in the complex plane, then $f(z)$ can be represented as a comnvergent power series (Taylor series) at every point z inside Ω.

IMPORTANT RESULT. If $f(z)$ is any function which is analytic withing a circle in the complex plane which contains all the eigenvalues λ_i of A, then a corresponding matrix function $f(A)$ can be defined by a convergent power series.

EXAMPLE:

The function $e^{\alpha x}$ is analytic for all values of x. Therefore it has convergent series representation:

$$e^{\alpha x} = 1 + \alpha x + \frac{\alpha^2 x^2}{2!} + \frac{\alpha^3 x^3}{3!} + \cdots + \frac{\alpha^k x^k}{k!} + \cdots$$

The corresponding matrix function is

$$e^{\alpha A} = I + \alpha A + \frac{\alpha^2 A^2}{2!} + \frac{\alpha^3 A^3}{3!} + \cdots + \frac{\alpha^k A^k}{k!} + \cdots$$

EXAMPLE:

$$e^{At} \cdot e^{Bt} = e^{(A+B)t}$$

if and only if, $AB = BA$.

OTHER USEFUL REPRESENTATIONS:

$$\sin A = A - \frac{A^3}{3!} + \frac{A^5}{5!} - \cdots$$

$$\cos A = I - \frac{A^2}{2!} + \frac{A^4}{4!} - \cdots$$

$$\sinh = A + \frac{A^3}{3!} + \frac{A^5}{5!} + \cdots$$

$$\cosh = I + \frac{A^2}{2!} + \frac{A^4}{4!} + \cdots$$

$$\sin^2 A + \cos^2 A = I$$

$$\sin A = \frac{e^{jA} - e^{-jA}}{2j} \qquad \text{and} \qquad \cos A = \frac{e^{jA} + e^{-jA}}{2}$$

Cayley-Hamilton Theorem

If the characteristic polynomial of a matrix A is

$$|\lambda I - A| = (-\lambda)^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + \cdots + c_1\lambda + c_0 = \triangle(\lambda)$$

then the corresponding matrix polynomial is

$$\triangle(A) = (-1)^n A^n + c_{n-1}A^{n-1} + c_{n-2}A^{n-2} + \cdots + c_1 A + c_0 I$$

CAYLEY-HAMILTON THEOREM. Every matrix satisfies its own characteristic equation.

$$\triangle(A) = 0$$

The proof of this follows by applying a diagonalizing similarity transformation to A and substituing into $\triangle(A)$.

EXAMPLE:

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

$$|\lambda I - A| = (3 - \lambda)(2 - \lambda) - 1 = \lambda^2 - 5\lambda + 5$$

$$\triangle(A) = A^2 - 5A + 5I = \begin{bmatrix} 10 & 5 \\ 5 & 5 \end{bmatrix} - 5 \cdot \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} + 5 \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$