Matrix Theory and Norms

- In this chapter, we will review some important fundamentals of matrix algebra necessary to the development of multivariable frequency response techniques
- Importantly, we'll focus on the geometric construct of the:
  - *eigenvalue-eigenvector decomposition*
  - *singular value decomposition*
- Additionally, we'll discuss the notion of *norms*

Basic Concepts

Complex numbers

- Define the complex scalar quantity

$$c = \alpha + j\beta$$

  where $\alpha = \mathrm{Re}\{c\}$ and $\beta = \mathrm{Im}\{c\}$

- We may express c in polar form, $c = |c| \angle c$, where $|c|$ denotes the magnitude (or modulus) of c and $\angle c$ its phase angle:

$$|c| = \sqrt{\bar{c}c} = \sqrt{\alpha^2 + \beta^2}$$

$$\angle c = \arctan\left(\frac{\beta}{\alpha}\right)$$

  Here, the $\bar{c}$ notation indicates complex conjugation.

Complex vectors and matrices

- A complex column vector $\boldsymbol{a}$ with m components may be written,

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

  where the a_i are complex scalars

- We define the following transpose operations:

$$\boldsymbol{a}^T = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \end{bmatrix}$$

$$\boldsymbol{a}^* = \begin{bmatrix} \bar{a}_1 & \bar{a}_2 & \cdots & \bar{a}_m \end{bmatrix}$$

  – The latter of these is often called the *complex conjugate* or *Hermitian* transpose

- Similarly, we may define a complex $\ell \times m$ matrix A,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{\ell 1} & a_{\ell 2} & \cdots & a_{\ell m} \end{bmatrix}$$

  with elements

$$a_{ij} = \mathrm{Re}\{a_{ij}\} + j\,\mathrm{Im}\{a_{ij}\}$$

- Here, ℓ denotes the number of rows (outputs), and m the number of columns (inputs)

- Similar to the vector, we define the following operations:

$$A^T = \text{transpose of } A$$

$$\bar{A} = \text{conjugate of } A$$
$$A^* = \text{conjugate transpose of } A$$

- MATRIX INVERSE

$$A^{-1} = \frac{\text{adj } A}{\det\ A}$$

where $\text{adj } A$ is the classical adjoint ("adjugate") of A which is computed as the transpose of the matrix of cofactors c_{ij} of A,

$$c_{ij} = [\text{adj } A]_{ji} \triangleq (-1)^{i+j}\ \det\ A^{ij}$$

Here A^{ij} is the submatrix formed by deleting row i and column j of A.

Example 3.1

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\det\ A = a_{11}a_{22} - a_{12}a_{21}$$

$$A^{-1} = \frac{1}{\det\ A}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Some Determinant Identities

- $\det\ A = \sum_{i=1}^{n} a_{ij}\ c_{ij}$ (expansion along column j)
- $\det\ A = \sum_{j=1}^{n} a_{ij}\ c_{ij}$ (expansion along row i)
- Let c be a complex scalar and A be an $n \times n$ matrix,

$$\det\,(cA) = c^n\ \det\ A$$

- Let A be a nonsingular square matrix,

$$\det\ A^{-1} = \frac{1}{\det\ A}$$

- Let A and B be matrices of compatible dimension such that AB and BA are square,

$$\det\left(I + AB\right) = \det\left(I + BA\right)$$

- For block triangular matrices,

$$\begin{aligned}\det\begin{bmatrix} A & B \\ 0 & D \end{bmatrix} &= \det\begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \\ &= \det\left(A\right)\cdot\det\left(D\right)\end{aligned}$$

- *Schur's Formula* for the determinant of a partitioned matrix:

$$\begin{aligned}\det\begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \det\left(A\right)\cdot\det\left(D - CA^{-1}B\right) \\ &= \det\left(D\right)\cdot\det\left(A - BD^{-1}C\right)\end{aligned}$$

  where it is assumed that A and/or D are non-singular.

Vector Spaces

- We'll start with the simplest example of a vector space: the *Euclidian* n-dimensional space, $\mathbb{R}^n$, e.g.,

$$\boldsymbol{u} = [u_1, u_2, ..., u_n] \text{ with } u_i\ \in \mathbb{R} \text{ for } i = 1, 2, \ldots, n$$

- $\mathbb{R}^n$ 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 $k\boldsymbol{u}$ is also in $\mathbb{R}^n$

  - Given any $k \in \mathbb{R}$ and $\boldsymbol{u} \in \mathbb{R}^n$ we can obtain by scalar multiplication the unique vector $k\boldsymbol{u}$

> *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.*
> [Jacobson, Nathan, Basic Algebra I, 2nd ed., 2009]

- Some further results:
  - $k\left(\boldsymbol{u}+\boldsymbol{v}\right) = k\boldsymbol{u} + k\boldsymbol{v}$ (i.e., scalar multiplication is distributive under addition)
  - $\left(k+l\right)\boldsymbol{u} = k\boldsymbol{u} + l\boldsymbol{u}$
  - $\left(kl\right)\boldsymbol{u} = k\left(l\boldsymbol{u}\right)$
  - $1 \cdot \boldsymbol{u} = \boldsymbol{u}$
- 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 $\boldsymbol{u} \in V$ to generate a unique $k\boldsymbol{u} \in V$
  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 of vectors, S=$\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_m}\} \in V$ where V is a vector space of some F
- Define $L(S)$ =$\{\boldsymbol{v} \in V : \boldsymbol{v} = a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \ldots + a_m\boldsymbol{u_m}$ with $a_i \in F\}$
- Thus $L(S)$ is the set of all *linear combinations* of the vectors $\boldsymbol{u_i}$
- 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 to 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

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

$\Rightarrow$ *Linear Tranform*

Matrix Respresentation of Linear Mappings

- Linear mappings can be made concrete and conveniently 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:
  - $\boldsymbol{y} = T(\boldsymbol{x})$ is equivalent to $\boldsymbol{y} = A\boldsymbol{x}$
- 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 "1" appearing in the i^{th} position

Change of Basis

- Let $\boldsymbol{e}_1', \boldsymbol{e}_2', \ldots, \boldsymbol{e}'_m$ be a new basis for X
- Then each $\boldsymbol{e}_i'$ is in X and can be written as a linear combination of the $\boldsymbol{e}_i$:

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

$$E' = EP$$

  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

$$T' = P^{-1} T P$$

Image, Kernal, Rank & Nullity

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

Image. The image of a linear transformation T is,

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

Kernel. The kernel of a linear transformation T is,

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

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

Example 3.2

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

- It is clear to see that matrix A transforms a 3×1 input vector into a 2×1 output vector, which is the dimension of the image of A
- The vector $\boldsymbol{n}$,

$$\boldsymbol{n} = \alpha \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

  denotes the null space of A

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
- Note that in the case of $T \in L(X, Y)$ defined by $\boldsymbol{x} \rightarrow \boldsymbol{y} = A\boldsymbol{x}$, the dimension of $\mathrm{Image}(T)$ is the dimension of the linear span of the columns of A, which equals the number of linearly independent column vectors
- An alternative but equivalent 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 3.3

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

- 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
- All four 3×3 minors of A are equal to 0, but A has non-zero minors of dimension 2×2, indicating A has rank 2

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

Some additional properties of determinants:

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

1. $\det(A) = 0$ if and only if the columns (rows) of A are linearly dependent
   (a) If a column (row) of A is zero, then $\det(A) = 0$
   (b) $\det(A) = \det(A^T)$
   (c) $\det(AB) = \det(A) \cdot \det(B)$
   (d) Swapping two rows (columns) changes the sign of $\det(A)$.
      i. Scaling a row (column) by k, scales the determinant by k.
      ii. Adding a multiple of one row (column) onto another does not affect the determinant.

(e) If $A = \text{diagonal}\{a_{ii}\}$ or lower triangular, or $A =$ upper triangular with diagonal elements a_{ii}, then $\det(A) = a_{11} \cdot a_{22} \cdot \cdots \cdot a_{nn}$

Eigenvalues and Eigenvectors

What is an Eigenvalue?

- Often in engineering we deal with matrix functions, e.g.,

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

- Here, A is a square real matrix, and $G(s)$ is the transfer function matrix of a dynamic system with many inputs and many outputs
- Furthermore,

$$\boldsymbol{y} = e^{At}\boldsymbol{x}_0$$

  is the form of solution of n first-order ordinary differential equations
- For scalar systems, we know how to handle terms like $\dfrac{1}{s-a}$, whose inverse Laplace transform is just e^{at}
- But how do we handle the matrix $(sI - A)^{-1}$? And how do we evaluate the matrix exponential 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 $A' = P^{-1}AP$
- Suppose that it were possible to choose our new basis such that $P^{-1}AP$ were a *diagonal* matrix D

  - This transforms a general matrix problem into a diagonal (scalar) one – it now consists of a collection of simple elements, $1/(s-d_i)$ 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-Eigenvector Relationship

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

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

- Such A-invariant directions are called *eigenvectors* of A and the corresponding λ's are called *eigenvalues* of A; *eigen* in German means "own" or "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 kernal 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 3.4:

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

- Solving $\det(sI - A) = 0$ gives,

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

and thus,

$$\lambda_1 = 1 \quad \lambda_2 = 2 \quad \lambda_3 = 3$$

- Solving for the eigenvectors we obtain

$$\boldsymbol{w}_1 = \alpha_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{w}_2 = \alpha_2 \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} \quad \boldsymbol{w}_3 = \alpha_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

where the α's are arbitrary scaling factors.

IMPORTANT EIGENVALUE/EIGENVECTOR 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:
$$\lambda_i\left(A + kI\right) = \lambda_i\left(A\right) + k$$
7. The eigenvalues of the inverse of a matrix are given by:
$$\lambda_i\left(A^{-1}\right) = \frac{1}{\lambda_i\left(A\right)}$$
8. The eigenvectors of a matrix and its transpose are the same.
9. The left eigenvectors of a matrix A are the right eigenvectors of A^*
10. The eigenvalues/eigenvectors of a real matrix are real or appear in complex conjugate pairs.
11. $\rho\left(A\right) \triangleq \max_i \left|\lambda_i\left(A\right)\right|$ is called the *spectral radius* of A
12. Eigenvalues are invariant under similarity transformations.

Gershgorin's Theorem

- The eigenvalues of the $n \times n$ matrix A lie in the union of n circles in the complex plane, each with center a_{ii} and radius

$$r_i = \sum_{j \neq i} |a_{ij}|$$

  - Radius equals the sum of off-diagonal elements in row i

- They also lie in the union of n circles, each with center a_{ii} and radius

$$r_i = \sum_{j \neq i} |a_{ji}|$$

  - Radius equals the sum of off-diagonal elements in column i

Example 3.5

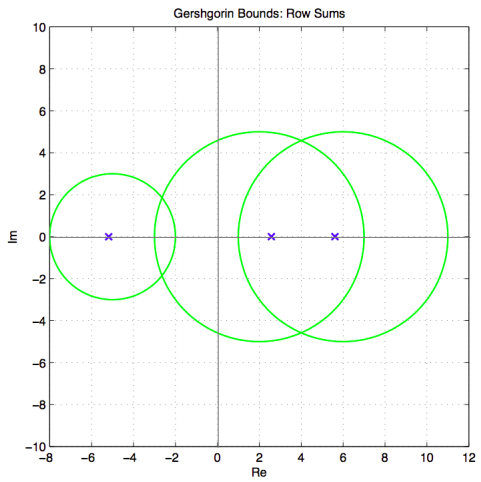
- Consider the matrix,

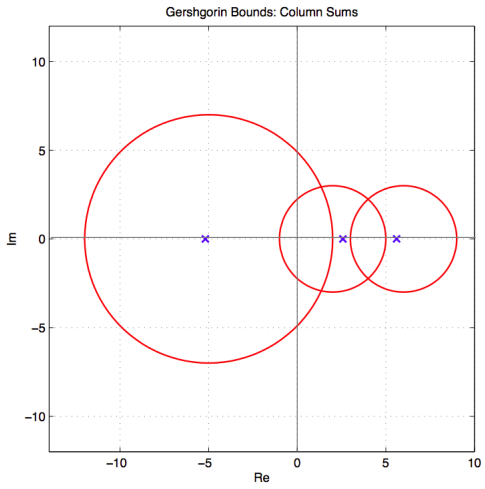
$$A = \begin{bmatrix} -5 & 1 & 2 \\ 4 & 2 & -1 \\ -3 & -2 & 6 \end{bmatrix}$$

- The computed eigenvalues are

$$\boldsymbol{\lambda} = \begin{bmatrix} -5.1866 \\ 5.6081 \\ 2.5785 \end{bmatrix}$$

- Gershgorin bands for both row and column sums are depicted in the plots below (overlaid with actual eigenvalues)





Eigen-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*, $V = W^{-1}$:

$$V = W^{-1} = \begin{bmatrix} \boldsymbol{v_1^T} \\ \boldsymbol{v_2^T} \\ \vdots \\ \boldsymbol{v_n^T} \end{bmatrix}$$

- The left eigenvectors $\boldsymbol{v}_i^T$ satisfy

$$\boldsymbol{v}_i^T (A - \lambda_i I) = 0 \quad \Leftrightarrow \quad \boldsymbol{v}_i^T A = \lambda_i \boldsymbol{v}_i^T$$

- 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 3.6

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

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 $\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

- 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 $\boldsymbol{w}_2 = \begin{bmatrix} 1000\epsilon \\ 1 \end{bmatrix}$
- The $\boldsymbol{w}_2$ 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., $\boldsymbol{w_j^T \cdot w_i = 0}$, for all $i \neq j$) . Therefore its eigenvalue/eigenvector decomposition becomes:

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

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

Corollary: 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: $N^*N = NN^*$
  - 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: $U^*U = UU^* = I$
  - 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,

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

  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_i'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).
- Matrices U and V are unitary matrices, i.e., they satisfy

$$U^* = U^{-1}$$

and

$$\|\lambda_i(U)\| = 1 \quad \forall i$$

- The largest singular value of M, σ_1, (also denoted, $\bar{\sigma}$) is an induced norm of the matrix M:

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

- It is important to note that,

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

and

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

- 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

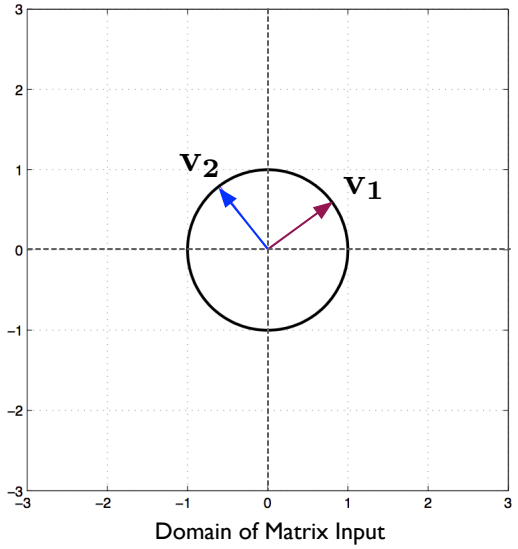
Example 3.7 (Golub & Van Loan):

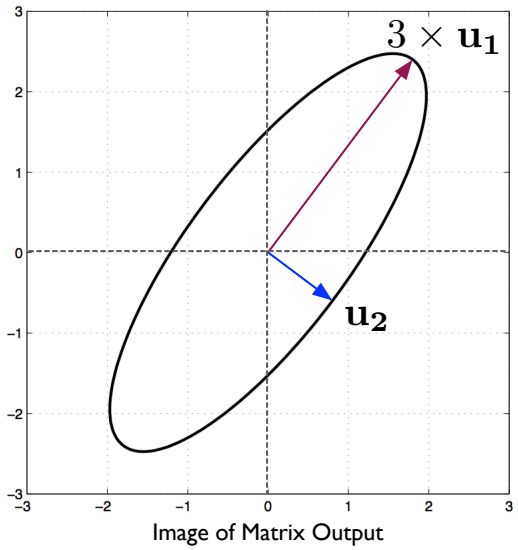
$$A = \begin{bmatrix} 0.96 & 1.72 \\ 2.28 & 0.96 \end{bmatrix} = U\Sigma V^T$$

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

- 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.

Singular value inequalities

- $\underline{\sigma}(A) \leq |\lambda_i(A)| \leq \bar{\sigma}(A)$
- $\bar{\sigma}(A^*) = \bar{\sigma}(A) \quad \text{and} \quad \bar{\sigma}\left(A^T\right) = \bar{\sigma}(A)$
- $\bar{\sigma}(AB) \leq \bar{\sigma}(A)\,\bar{\sigma}(B)$
- $\underline{\sigma}(A)\,\bar{\sigma}(B) \leq \bar{\sigma}(AB) \quad \text{or} \quad \bar{\sigma}(A)\,\underline{\sigma}(B) \leq \bar{\sigma}(AB)$
- $\underline{\sigma}(A)\,\underline{\sigma}(B) \leq \underline{\sigma}(AB)$
- $\max\left\{\bar{\sigma}(A),\ \bar{\sigma}(B)\right\} \leq \bar{\sigma}\begin{bmatrix} A \\ B \end{bmatrix} \leq \sqrt{2}\max\left\{\bar{\sigma}(A),\ \bar{\sigma}(B)\right\}$

- $\bar{\sigma}\begin{bmatrix} A \\ B \end{bmatrix} \leq \bar{\sigma}(A) + \bar{\sigma}(B)$
- $\bar{\sigma}\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} = \max\left\{ \bar{\sigma}(A),\ \bar{\sigma}(B) \right\}$
- $\sigma_i(A) - \bar{\sigma}(B) \leq \sigma_i(A+B) \leq \sigma_i(A) + \bar{\sigma}(B)$

Condition number

- The condition number of a matrix A is defined as the ratio of the maximum to the minimum singular values,

$$\kappa(A) = \frac{\bar{\sigma}(A)}{\underline{\sigma}(A)}$$

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

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

  where $p \geq 1$
- Three norms of particular interest to control theory are:
  - VECTOR 1-NORM

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

  - VECTOR 2-NORM (EUCLIDIAN NORM)

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

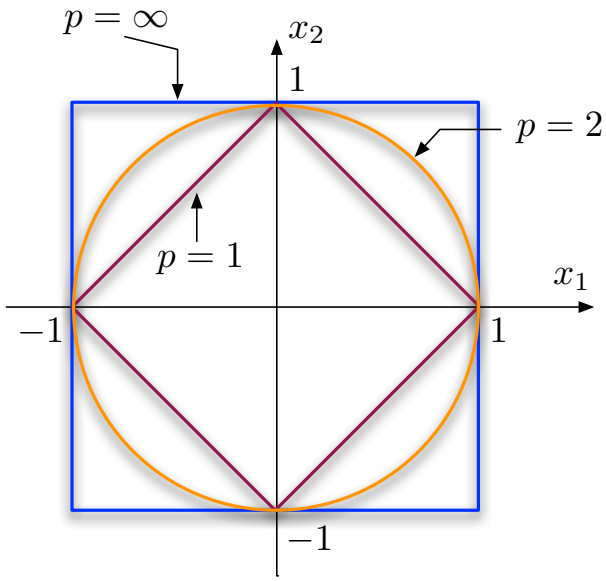
  - VECTOR ∞-NORM

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

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

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

- The following plot shows contours for the vector p-norms defined above ($n = 2$)



Matrix Norms

- Matrix norms satisfy the same properties outlined above for vector norms
- 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:

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

  - Norms defined in this manner are known as *induced* norms
  - NOTE: the *supremum* is used here instead of the maximum since the maximum value may not be achieved

- The matrix norms are computed by the following expressions:

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

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

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

SOME USEFUL MATRIX NORM RELATIONSHIPS:

- $\|AB\|_p \leq \|A\|_p \|B\|_p$ (multiplicative property)
- $\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2$
- $\max_{i,j} |a_{ij}| \leq \|A\|_2 \leq \sqrt{mn} \max_{i,j} |a_{ij}|$

- $\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$
- $\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{m} \|A\|_\infty$
- $\frac{1}{\sqrt{m}} \|A_1\| \leq \|A\|_2 \leq \sqrt{n} \|A\|_1$

Spectral Radius

- The spectral radius was defined earlier as

$$\rho(A) = \max_i |\lambda_i(A)|$$

- Note: the spectral radius is *not* a norm!

Example 3.8

- Consider the matrices

$$A_1 = \begin{bmatrix} 1 & 0 \\ 10 & 1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 10 \\ 0 & 1 \end{bmatrix}$$

- It is obvious that

$$\rho(A_1) = 1, \quad \rho(A_2) = 1$$

- But,

$$\rho(A_1 + A_2) = 12$$
$$\rho(A_1 A_2) = 101.99$$

- Hence the spectral radius does not satisfy the multiplicative property, i.e.,

$$\rho(A_1 A_2) \not\leq \rho(A_1)\rho(A_2)$$

System Norms

- Consider the input-output system below

  - Given information about allowable input signals $\boldsymbol{w}(t)$, how large can the outputs $\boldsymbol{z}(t)$ become?
  - To answer this question, we must examine relevant system norms.

  We shall evaluate the output signals in terms of the signal 2-norm, defined by

$$\|\boldsymbol{z}(t)\|_2 = \sqrt{\sum_i \int_{-\infty}^{\infty} |z_i(\tau)|^2 \, d\tau}$$

  and consider three inputs:

  1. $\boldsymbol{w}(t)$ is a series of unit impulses
  2. $\boldsymbol{w}(t)$ is any signal such that $\|w(t)\|_2 = 1$
  3. $\boldsymbol{w}(t)$ is any signal such that $\|\boldsymbol{w}(t)\|_2 = 1$, but $\boldsymbol{w}(t) = \mathbf{0}$ for $t \geq 0$ and we only measure $\boldsymbol{z}(t)$ for $t \geq 0$

- The relevant system norms are: $\mathcal{H}_2$, $\mathcal{H}_\infty$, and Hankel norms, respectively

$\mathcal{H}_2$ norm

- Consider a stricly proper system $G(s)$ (i.e., $D = 0$)
- For the $\mathcal{H}_2$ norm we use the Frobenius norm spatially (for the matrix) and integrate over frequency, i.e.,

$$\|G(s)\|_2 \triangleq \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{tr}\left(G(j\omega)^* G(j\omega)\right) d\omega}$$

  - $G(s)$ must be strictly proper, or else the $\mathcal{H}_2$ norm is infinite

- By Parseval's theorem we have

$$\|G(s)\|_2 = \|g(t)\|_2 \triangleq \sqrt{\int_0^{\infty} \mathrm{tr}\left(g^T(\tau)\, g(\tau)\right) d\tau}$$

  where $g(t)$ is a matrix of impulse responses.

- Note we can change the order of integration and summation to get

$$\|G(s)\|_2 = \|g(t)\|_2 = \sqrt{\sum_{ij} \int_0^{\infty} \left|g_{ij}(\tau)\right|^2 d\tau}$$

  where $g_{ij}(t)$is the ij^{th} element of the impulse response matrix.

- Thus, the $\mathcal{H}_2$ norm is the 2-norm output resulting from applying unit impulses $\delta_j(t)$ to each input

Numerical computation of the $\mathcal{H}_2$ norm

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

- Then,

$$\|G(s)\|_2 = \sqrt{\mathrm{tr}\left(B^T Q B\right)} \quad or \quad \|G(s)\|_2 = \sqrt{\mathrm{tr}\left(CPC^T\right)}$$

  where Q is the observability Gramian and P is the controllability Gramian.

H_∞ norm

- Consider a proper, stable system $G(s)$ (for this case, nonzero D is allowed)
- For the $\mathcal{H}_\infty$ norm, we use the singular value (induced 2-norm) spatially for the matrix and pick out the peak value as a function of frequency

$$\|G(s)\|_\infty \triangleq \max_\omega \bar{\sigma}\left(G(j\omega)\right)$$

  - The $\mathcal{H}_\infty$ norm is interpreted as the peak transfer function "magnitude"
- Time Domain Interpretation:
  - Worst-case steady-state gain for sinusoidal inputs at any frequency
  - Induced (worst-case) 2-norm in the time domain:

$$\|G(s)\|_\infty = \max_{\boldsymbol{w}(t)\neq 0} \frac{\|\boldsymbol{z}(t)\|_2}{\|\boldsymbol{w}(t)\|_2} = \max_{\|\boldsymbol{w}(t)\|_2=1} \|\boldsymbol{z}(t)\|_2$$

  - NOTE: at a given frequency ω, the gain depends on the *direction* of $\boldsymbol{w}(\omega)$

Numerical computation of the $\mathcal{H}_\infty$ norm

- Computation of the H_∞ norm is non-trivial
  - It can be approximated via the defnition $\|G(s)\|_\infty \triangleq \max_\omega \bar{\sigma}\left(G(j\omega)\right)$
  - But in practice, it's normally computed via its state-space representation as shown below
- Consider the system represented by

$$G(s) = C(sI - A)B + D$$

- The $\mathcal{H}_\infty$ norm is the smallest value of γ such that the Hamiltonian matrix H has no eigenvalues on the imaginary axis

$$H = \begin{bmatrix} A + BR^{-1}D^T C & BR^{-1}B^T \\ -C^T\left(I + DR^{-1}D^T\right)C & -\left(A + BR^{-1}D^T C\right)^T \end{bmatrix}$$

where $R = \gamma^2 I - D^T D$

  - The derivation of this expression will not be addressed here

Comparison of $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms

- We first recognize that we can write the Frobenius norm in terms of singular values,

$$\|G(s)\|_2 = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_i \sigma_i^2\left(G(j\omega)\right)d\omega}$$

- Minimizing the $\mathcal{H}_\infty$ norm essentially "pushes down the peak" of the largest singular value
- Minimizing the $\mathcal{H}_2$ norm "pushes down the entire function" of singular values over all frequencies

Example 3.9

- Consider the plant

$$G(s) = \frac{1}{s+a}$$

- Computing the H_2 norm

$$\|G(s)\|_2 = \left(\frac{1}{2\pi}\int_{-\infty}^{\infty}|G(j\omega)|^2 d\omega\right)^{\frac{1}{2}}$$

$$= \left(\frac{1}{2\pi a} \left[\tan^{-1} \left(\frac{\omega}{a} \right) \right]_{-\infty}^{\infty} \right)^{\frac{1}{2}}$$

$$= \sqrt{\frac{1}{2a}}$$

- Alternatively, consider the impulse response

$$g(t) = \mathcal{L}^{-1} \left(\frac{1}{s+a} \right) = e^{-at}, \quad t \geq 0$$

  – which gives

$$\|g(t)\|_2 = \sqrt{\int_0^{\infty} (e^{-at})^2 \, dt} = \sqrt{\frac{1}{2a}}$$

- Computing the $\mathcal{H}_\infty$ norm,

$$\|G(s)\|_\infty = \max_\omega |G(j\omega)| = \max_\omega \frac{1}{(\omega^2 + a^2)^{\frac{1}{2}}} = \frac{1}{a}$$

- Note that there is no general relationship between the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms

Hankel norm

- A special norm in systems theory is the Hankel norm, which relates the *future* output of a system to its *past* input:

$$\|G(s)\|_H \triangleq \max_{\boldsymbol{w}(t)} \frac{\sqrt{\int_0^{\infty} \|\boldsymbol{z}(\tau)\|_2^2 \, d\tau}}{\sqrt{\int_{-\infty}^{0} \|\boldsymbol{w}(\tau)\|_2^2 \, d\tau}}$$

- It can be shown that the Hankel norm is equal to

$$\|G(s)\|_H = \sqrt{\rho(PQ)}$$

where ρ is the spectral radius and P and Q are the controllability and observability Gramians, respectively

- The name "Hankel norm" derives from the fact that the matrix PQ is in the form of a Hankel matrix
- The Hankel singular values are given by

$$\sigma_{H,i} = \sqrt{\lambda_i (PQ)}$$

- Hankel and $\mathcal{H}_\infty$ norms can be related by

$$\|G(s)\|_H = \bar{\sigma}_H \leq \|G(s)\|_\infty \leq 2 \sum_{i=1}^{n} \sigma_{H,i}$$

Example 3.10

- Let's compute the various norms for the system of Example 3.9 using state-space formulations.
- For the plant $G(s) = {}^{1}/_{(s+a)}$ we have the following state-space realization:

$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right] = \left[\begin{array}{cc} -a & 1 \\ 1 & 0 \end{array}\right]$$

- The controllability Gramian P is the solution of the Lyapunov equation

$$\begin{aligned} AP + PA^T &= -BB^T \\ -aP - aP &= -1 \\ P &= \frac{1}{2a} \end{aligned}$$

- Similarly, for the observability Gramian,

$$Q = \frac{1}{2a}$$

- So for the $\mathcal{H}_2$ norm we have

$$\|G(s)\|_2 = \sqrt{\mathrm{tr}\left(B^T Q B\right)} = \sqrt{\frac{1}{2a}}$$

- Computing the H_∞ norm, we first find the eigenvalues of the Hamiltonian matrix,

$$\lambda(H) = \lambda\begin{bmatrix} -a & \frac{1}{\lambda^2} \\ -1 & a \end{bmatrix} = \pm\sqrt{a^2 - \frac{1}{\gamma^2}}$$

  - Clearly, H has no imaginary eigenvalues for $\gamma > \frac{1}{a}$, so

$$\|G(s)\|_\infty = \frac{1}{a}$$

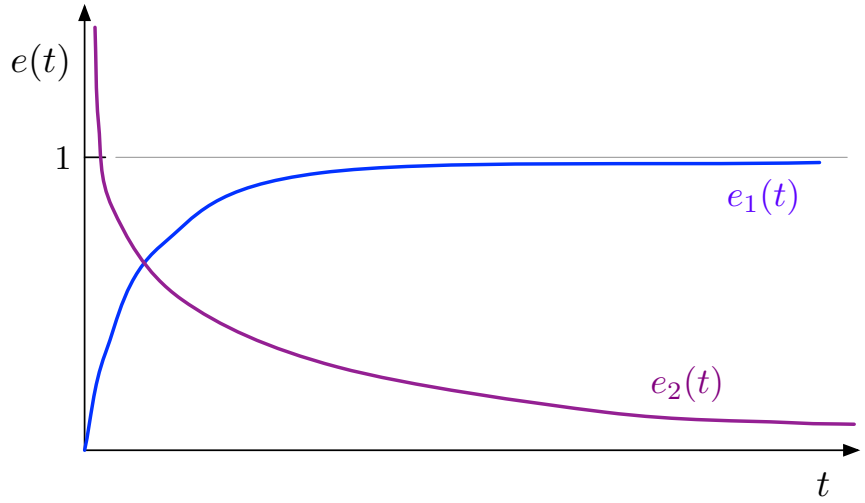
- Finally, the Hankel matrix is $PQ = \frac{1}{4a^2}$, so the Hankel norm is

$$\|G(s)\|_H = \sqrt{\rho(PQ)} = \frac{1}{2a}$$

Signal Norms

Example 3.11

- Consider the following two time signals



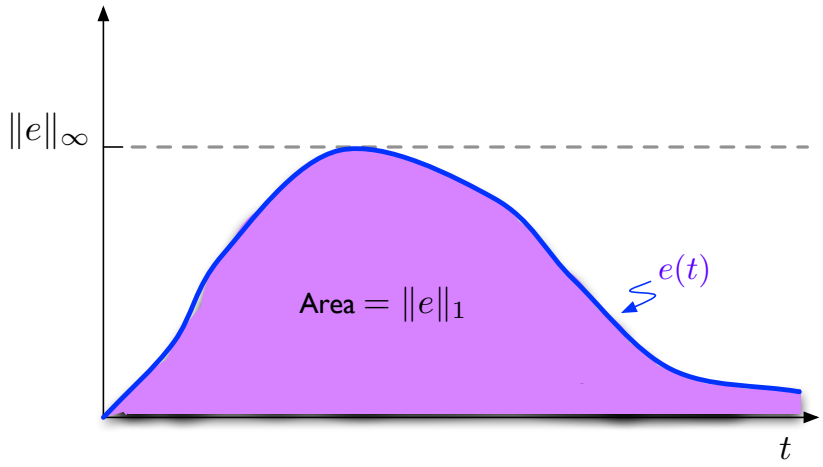
- For these two signals, we compute the following values for their respective norms:

$$\|e_1(t)\|_\infty = 1$$
$$\|e_1(t)\|_2 = \infty$$

$$\|e_2(t)\|_\infty = \infty$$
$$\|e_2(t)\|_2 = 1$$

Computation of signal norms

1. "Sum up" the channels at a given time or frequency using a vector norm (for a scalar signal, take the absolute value)
2. "Sum up" in time or frequency using a temporal norm



- We define the temporal p-norm of a time-varying vector as

$$\|e(t)\|_p = \left(\int_{-\infty}^{\infty} \sum_i |e_i(\tau)|^p \, d\tau \right)^{1/p}$$

- The following temporal norms are commonly used:
  - 1-norm in time

$$\|e(t)\|_1 = \int_{-\infty}^{\infty} \sum_i |e_i(\tau)| \, d\tau$$

  - 2-norm in time

$$\|e(t)\|_2 = \sqrt{\int_{-\infty}^{\infty} \sum_i |e_i(\tau)|^2 \, d\tau}$$

  - ∞-norm in time

$$\|e(t)\|_\infty = \sup_\tau \left(\max_i |e_i(\tau)| \right)$$

Additional Topics

Toeplitz & Hankel Matrix Forms

- Systems theory makes extensive use of Toeplitz and Hankel matrices to simplify much of the algebra.
- Consider the polynomial $n(z)$:

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

- 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 often used for changing polynomial convolution into a matrix-vector multiplication

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 3.12

- The function $e^{\alpha x}$ is analytic for all values of x; therefore it has a 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 3.13

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

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 3.14

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

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

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

(mostly blank)

(mostly blank)