510: **Fall 2019**

Lecture 1: Introduction and Overview

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1 Today

Objective: This course is a bootcamp course in linear algebra. We will largely focus on finite dimensional linear algebra with short forays into infinitely dimensional concepts where and when necessary.

The motivation for the material in this course is to ensure that graduate students are all 'on the same page' in terms of their linear algebra background as needed for courses such as 547 (Linear Systems Theory) and 578 (Convex optimization).

Linear algebra is the foundation for these courses and many more.

Topics to be covered include:

- linear algebra review, QR decomposition
- Solutions to linear equations
- regularized least squares, least norm problems
- Eigen-decompositions and Jordan canonical form
- positive semidefiniteness, matrix norms
- Singular value decompositions
- applied operator theory & functional analysis (e.g., adjoint operators)
- Introduction to autonomous linear systems, including finding solutions using linear algebra
- Stability via Lyapunov equations

Linear Algebra Background: *This course is not meant to be a first course in linear algebra!* You should already have a strong background in linear algebra. In this course we will review some topics in a more rigorous manner including learning to prove linear algebra results.

2 Course Organization

Canvas will be used as the "course website". You will upload your homeworks (and self-assessments there). I will post any lecture materials I provide on canvas. Information about office hours will also be posted there.

Course prerequisites.

- knowledge of linear algebra, matrix analysis, differential equations & Laplace transform
- comfort with mathematical proofs
- familiarity with computational tools (e.g., Python)
- exposure to numerical linear algebra

The course is suitable for students interested in PhD research in systems theory.

questions? check out course website, reference texts, come to office hours to consult

Grading. Grading will be based on four components:

1. homework sets: 25%
2. exam 1 (midterm): 30%
3. exam 2 (final, comprehensive): 40%
4. course participation: 5%

Homework: There will be (roughly) weekly homeworks. You will use a 'self-assessment' to self-grade your homeworks. The purpose of self-assessment is so that you are forced to go through and review your work and think critically about the solution.

Important: There is most often not *one* right answer. The self-assessment process lets you go through and point myself and the TA to the problems in which you believe your solution is correct but you are not sure because it is different.

Self-Assessment Process: You can give yourself the following point designations per problem.

- **2**: Your solution is exactly correct on all parts.
- **1**: You have any doubt about the correctness of any part of your solution, or it is incomplete.
- **0**: If you have made no attempt at any part of the problem.

If you give yourself a **1**, we will look at your solution and mark it up to a **2**. Essentially, you earn back the extra point by simply going through and critically assessing your solution.

In giving your self each of the point designations, you have to justify why. In particular, if you give your self **1**pt, then say what you did wrong. **Do not provide a corrected solution because you will have the solutions during the assessment so it makes no sense to provide a corrected solution.** Instead, provide an assessment of what you did wrong. We will look and give you back the extra point. **Do not give yourself the extra point; we will do this!** If you deviate from this process at all, we reserve the right to give you zero points for that problem.

No late homeworks will be accepted!: even a couple mins or seconds after the deadline, your homework will be considered not submitted. Plan accordingly.

Exams: the exams will be in-class. The final will be comprehensive.

2.1 Books

There is no book for the course. There are several reference books you can consider looking at including the following:

- Linear algebra done right, Sheldon Axler
- Linear Functional Analysis, Rynne and Youngson
- Optimization by vector space methods, Luenberger
- Principles of Mathematical Analysis, Rudin
- Linear Systems Theory, Callier and Desoer (Appendix A, B primarily)

3 Why study linear algebra so intensely?

Linear algebra is the foundation for almost everything we do in engineering and computer science, and most of the pure sciences!

- automatic control systems

- signal processing
- communications
- economics, finance
- operations research and logistics
- circuit analysis, simulation, design
- mechanical and civil engineering
- aeronautics
- navigation, guidance

Let us discuss some concrete examples.

3.1 Optimization and Machine Learning

Relevant secondary courses: CSE/EE578: Convex Optimization, CSE546: Machine Learning, MATH/ AMATH515A: Optimization: Fundamentals and Applications, CSE535: Theory of optimization and continuous algorithms. . . so many more!

Setting statistics aside[1], basically everything in optimization and machine learning boils down to some linear algebra concept or depends on some core linear algebra concept.

The following is one of many examples.

Note: You do not have to know all of the mathematics behind this example; it is simply to demonstrate the connections between linear algebra and optimization/ML problems.

Example. Principle Component Analysis.

Cool visualization: `http://setosa.io/ev/principal-component-analysis/`

PCA is a powerful statistical tool for analyzing data sets and is formulated in the language of linear algebra. Largely, it used as a tool in exploratory data analysis or for making predictive models. Indeed, it can be used to address all kinds of questions of relevance to data analysis including:

- *Is there a simpler way of visualizing data?* Data is a priori a collection of points in $\mathbb{R}^n$, where n might be very large. Might want to project onto lower dimensional subspaces of important explanatory variables.
- *Which variables are correlated?*
- *Which variables are the most significant in describing the full data set?*

PCA is an *orthogonal linear transformation* that transforms the data to a new coordinate system such that the greatest *variance* by some projection of the data lies on the first coordinate (i.e., first principal component), the second greatest variance on the second coordinate, and so on. . .

Connecting ML+Opt via Linear Algebra: Let's see it in action.

Consider a data matrix, $X \in \mathbb{R}^{n\times p}$, with column-wise zero empirical mean (the sample mean of each column has been shifted to zero), where

- each of the n rows, denoted $x_i \in \mathbb{R}^{1\times p}$, represents a different repetition of the experiment
- each of the p columns gives a particular kind of feature (e.g., the results from a particular sensor)

Zero mean: $\sum_{i=1}^n x_i = 0_{1\times d}$ which can be enforced by subtracting the mean $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

The output of PCA is ℓ orthonormal vectors w_j, $j \in \{1, \dots, \ell\}$ (i.e. the top ℓ principle components) that

[1] Another very foundational area worth becoming familiar with; ask and I will happily provide you with references or course suggestions.

maximize

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{\ell}(x_i \cdot v_j)^2$$

Mathematically, the transformation is defined by a set, with cardinality ℓ, of p-dimensional vectors of weights or coefficients $w_k = (w_{k1}, \dots, w_{kp})$, $k = 1, \dots, \ell$, that map each row vector x_i of X to a new vector of principal component scores $t_i = (t_{i1}, \dots, t_{il})$, given by

$$t_{ik} = x_i \cdot w_k, \ i = 1, \dots, n, \ k = 1, \dots, \ell$$

in such a way that the individual variables $t_{i1}, \dots, t_{i\ell}$ of t_i considered over the data set successively inherit the maximum possible variance from X, with each coefficient vector w constrained to be a unit vector (where ℓ is usually selected to be less than p to reduce dimensionality).

e.g., consider computing the first principle component: in order to maximize variance, the first weight vector w_1 thus has to satisfy

$$w_1 = \arg\max_{\|w\|=1}\left\{\sum_i t_{i1}^2\right\} = \arg\max_{\|w\|=1}\left\{\sum_i (x_i \cdot w)^2\right\}$$

Equivalently, writing this in matrix form gives

$$w_1 = \arg\max_{\|w\|=1}\left\{\|Xw\|^2\right\} = \arg\max_{\|w\|=1} w^T X^T X w$$

and since $\|w\| = 1$, this is equivalent to

$$w_1 = \arg\max\left\{\frac{w^T X^T X w}{w^T w}\right\}$$

where

$$X = \begin{bmatrix} — & x_1 & — \\ & \vdots & \\ — & x_n & — \end{bmatrix}$$

and

$$Xw = \begin{bmatrix} x_1 \cdot w \\ \vdots \\ x_n \cdot w \end{bmatrix}$$

The quantity to be maximised can be recognised as a *Rayleigh quotient*. A standard result for a positive semidefinite matrix such as $X^T X$ is that the quotient's maximum possible value is the *largest eigenvalue of the matrix*, which occurs when w is the corresponding eigenvector.

The matrix $X^T X$ has a natural interpretation. The (i, j) entry of this matrix is the inner product of the i–th row of X and the j–th column of X—i.e., of the i–th and j–th columns of X. So $X^T X$ just collects the inner products of columns of X, and is a symmetric matrix.

Other examples:

- *Solutions to systems of linear equations*. Solving linear equations is very basic problem that shows up in all kinds of application areas.

  e.g., $Ax = b$. This shows up in a lot of ML problems where you need to invert a matrix. In a lot of ML applications, you need to compute vector-Jacobian products, inverse Hessians, etc.
- *Gradient-based Optimization*. Using a local linearization of a cost function to determine the direction of steepest descent towards finding a local minima.

3.2 Dynamical Systems and Control Theory

Relevant secondary courses: AA/ME/EE547 (linear systems theory, I)-548 (linear systems theory, II)-549(Estimation and System Identification); AMATH575 A/B/C: Dynamical Systems; AMATH 502 A: Introduction to Dynamical Systems Theory and Chaos; EE550: Nonlinear Optimal Control; AMATH 518 Theory of Optimal Control, etc.

Linear dynamical systems are central to control theory. They represent a class of systems that we understand very well and can be used to approximate nonlinear phenomena, particularly in the design of controllers or observers for real world systems.

Example. Controllability/Observability. Consider the linear system

$$
\begin{aligned}
x_{k+1} &= Ax_k + Bu_k \\
y_k &= Cx_k
\end{aligned}
$$

Questions about the controllability or observability of such a system boil down to characterizing the rank of a particular operator.

Controllability: informally, a system—in particular, the pair (A, B)—is controllable on $[k_0, k_1]$ if and only if (iff) $\forall$ (x_0, k_0) and $\forall$ (x_1, k_1), there exists a control input $(u_k)_{k\in[k_0,k_1-1]}$ that transfers x_0 at time k_0 to x_1 at time k_1.

Observability: informally, a system—in particular, the pair (C, A)—is observable on $[k_0, k_1]$ iff for all in inputs $(u_k)_{k\in[k_0,k_1-1]}$ and for all corresponding outputs $(y_k)_{k\in[k_0,k_1]}$, the state x_0 at time k_0 is uniquely determined.

Controllability is tantamount to assessing the surjectivity of a map L_c which maps the input sequence to the state space. Observability is tantamount to assessing the injectivity of a map L_o that maps initial states to the output space. Eventually (if you take 547) you will see that this reduces to checking rank conditions on the matrices

$$
C = \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}
$$

and

$$
O = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}
$$

for controllability and observability, respectively.

Other examples.

- *Stability of a linear system.* Towards the end of this course, we will actually cover this topic. Stability can be assessed based on characterizing the eigen structure of a dynamical system or solving a Lyapunov equation which is defined by a special linear operator which we will study in this course.
- *Optimal control.* e.g., finding the least norm control input to a linear system.
- *Characterizing invariant spaces.* Invariant spaces are important for all kinds of aspects of control theory including providing safety guarantees or understanding reachability or robustness.

3.3 Numerical Methods for Differential Equations

Relevant secondary courses: AA/EE/ME 547-548 (linear systems theory); AMATH 516: Numerical Optimization; MATH 554-556 Linear Analysis, etc.

Example. Solving differential equations numerically. In solving differential equations such as

$$
\dot{x} = f(x),\ x \in \mathbb{R}^n,
$$

one approach is to discretize the differential equation. Indeed,

$$x(t+h) - x(t) = h\dot{x} + O(h^2) = hf(x(t)) + O(h^2) \implies x_{k+1} = x_k + hf(x_k)$$

where we define $x_{k+1} = x(t+h)$[2]. Whether or not this process converges depends on the stability of the discretization which, in turn, depends on the choice of *step-size* h relative to the behavior of the dynamics. For instance, in the case of a linear system

$$\dot{x} = Ax, \ x \in \mathbb{R}^n, \ A \in \mathbb{R}^{n \times n}$$

we have a discretized system

$$x_{k+1} = x_k + hAx_k = (I + hA)x_k$$

Discrete time systems, as we will learn, are stable when their eigenvalues are inside the unit circle. Hence, h would need to be chosen such that $\mathrm{spec}(I + hA) \in D_1(0)$. To further illustrate this, consider the scalar case:

$$\dot{x}(t) = -ax(t) \implies x_{k+1} = (1 - ha)x_k$$

Q: for what values of a does the continuous time system decay to zero?

A: $0 < a \in \mathbb{R}$. Why is this the case? Well, the solution is $x(t) = x(t_0)e^{-a(t-t_0)}, \ \forall t \geq t_0$ and we know an exponential function will decay to zero if the exponent is negative.

Hence, if $a > 0$, $|1 - ha| < 1$ if $h < h_0$ where h_0 is the largest h such that $|1 - ha| = 1$.

$$h = a \implies |1 - ha| = 1$$

Hence, for any $0 < h < a$, the discretized linear system will be stable.

4 Cool Visualizations

http://setosa.io/ev/

[2]In particular, the continuous time t at iteration k is $t = t_0 + kh$.

b

EE547: **Fall 2018**

Lecture 1b: Functions, Linear Functions, Examples, and More...

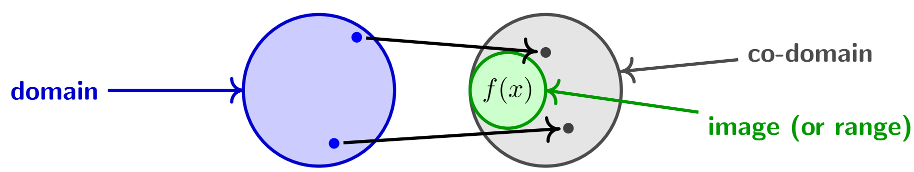
Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1.1 Functions

First, what is a function? And, what kind of notation will we use?

Given two sets $\mathcal{X}$ and $\mathcal{Y}$, by $f : \mathcal{X} \to \mathcal{Y}$ we mean that $\forall x \in \mathcal{X}$, f assigns a *unique* $f(x) \in \mathcal{Y}$.



$$f : \mathcal{X} \to \mathcal{Y}$$

- f maps $\mathcal{X}$ into $\mathcal{Y}$ (alternative notation: $f : x \mapsto y$.
- image of f (range):
$$f(\mathcal{X}) = \{f(x) |\ x \in \mathcal{X}\}$$

1.1.1 Characterization of functions

Injectivity (one-to-one):

$$f \text{ is injective } \iff [f(x_1) = f(x_2) \Rightarrow x_1 = x_2] \iff [x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)]$$

Surjectivity (onto):

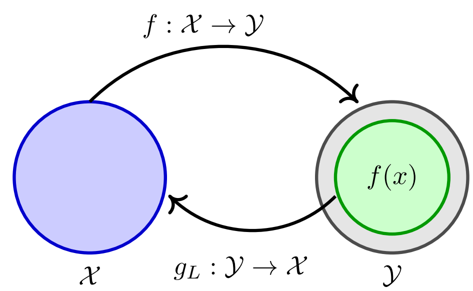
$$f \text{ is surjective} \iff \forall\ y \in \mathcal{Y},\ \exists\ x \in \mathcal{X},\ \text{such that } y = f(x)$$

Bijective (injective and surjective):

$$\forall\ y \in \mathcal{Y},\ \exists!\ x \in \mathcal{X} \text{ s.t. } y = f(x)$$

Example. Consider $f : \mathcal{X} \to \mathcal{Y}$ and let $\mathbf{1}_{\mathcal{X}}$ be the identity map on $\mathcal{X}$. We define the *left inverse* of f as the map $g_L : \mathcal{Y} \to \mathcal{X}$ such that $g_L \circ f \equiv \mathbf{1}_{\mathcal{X}}$. (composition: $g_L \circ f : \mathcal{X} \to \mathcal{X}$ such that $g_L \circ f : x \mapsto g_L(f(x))$)

0



Exercise. Prove that

$$[f \text{ has a left inverse } g_L] \iff [f \text{ is injective}]$$

Before we do the proof, some remarks on constructing proofs. Consider proving

$$\texttt{Expression A} \Longrightarrow \texttt{Expression B}$$

One can provide things with a direct argument (e.g., suppose `Expression A` is true, and argue that `Expression B` must hold), finding a contradiction (i.e., by supposing `Expression A` holds and `Expression B` does not and arguing a contradiction in the logic), or via the contrapositive argument. The statement

$$\texttt{Expression A} \Longrightarrow \texttt{Expression B}$$

is equivalent to

$$\texttt{Expression B} \Longrightarrow \texttt{Expression A}$$

(i.e. not B implies not A). One can also disprove a statement by coming up with a counter-example.

Proof: ($\Leftarrow$): Assume f is injective. Then, by definition,

$$f(x_1) = f(x_2) \Rightarrow x_1 = x_2$$

Construct $g_L : \mathcal{Y} \to \mathcal{X}$ such that, on $f(\mathcal{X})$, $g_L(f(x)) = x$. This is a well-defined function due to injectivity of f. Hence, $g_L \circ f \equiv \mathbf{1}_{\mathcal{X}}$.

($\Rightarrow$): Assume f has a left inverse g_L. Then, by definition, $g_L \circ f \equiv \mathbf{1}_{\mathcal{X}}$. Which implies that $\forall\, x \in \mathcal{X}$, $g_L(f(x)) = x$. We can use these facts to find a contradiction. Indeed, suppose f is not injective. Then, $\exists$ some $x_1 \neq x_2$ such that $f(x_1) = f(x_2)$. Hence, $g_L(f(x_1)) = x_1$ AND $g_L(f(x_2)) = x_2$. Yet, since g_L is a function $x_1 = x_2$ which contradicts our assumption ($\rightarrow\leftarrow$). Thus f has to be injective. ■

Remark. Some of the homeworks (particularly early on) will require you to derive arguments such as the above. So I will try to derive some of these in the class. Please ask questions if you feel uncomfortable at all.

Analogously, we define the *right inverse* of f as the map $g_R : \mathcal{Y} \to \mathcal{X}$ such that $f \circ g_R \equiv \mathbf{1}_{\mathcal{Y}}$.

DIY Exercise. prove that

$$[f \text{ has a right inverse } g_R] \iff [f \text{ is surjective}]$$

Remark. DIY exercises you find in the notes will sometimes appear in the official homeworks but not always. They are called out in the notes to draw your attention to practice problems that will help you grasp the material.

Finally, g is called a *two-sided inverse* or simply, an *inverse* of f (and is denoted f^{-1})

- $\iff g \circ f \equiv \mathbf{1}_{\mathcal{X}}$ and $f \circ g \equiv \mathbf{1}_{\mathcal{Y}}$ (that is, g is both a left and right inverse of f)
- $\iff f$ is invertible, f^{-1} exists

1.2 Linear equations and functions

Another characterization of functions is linearity.

Probably the most common representation of a linear function you have seen is a system of linear equations. Consider a system of linear equations:

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \; &\vdots \; \vdots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{aligned}$$

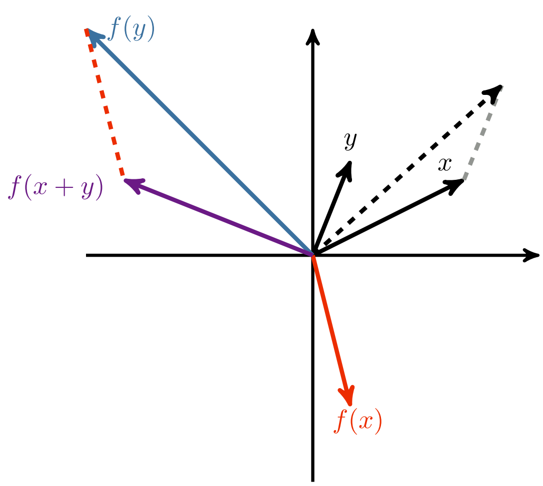
We can write this in matrix form as $y = Ax$ where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Formally, a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is linear if

- $f(x + y) = f(x) + f(y)$, $\forall\, x, y \in \mathbb{R}^n$
- $f(\alpha x) = \alpha f(x)$, $\forall\, x \in \mathbb{R}^n$, $\forall\, \alpha \in \mathbb{R}$

That is, superposition holds!



1.2.1 Linear Maps

Let (U, F) and (V, F) be linear spaces over the same field F. Let $\mathcal{A}$ be a map from U to V.

Definition 1.1 (Linear operator/map) *$\mathcal{A} : U \to V$ is said to be a linear map (equiv. linear operator) iff for each $v_1, v_2 \in V$*

$$\mathcal{A}(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 \mathcal{A}(v_1) + \alpha_2 \mathcal{A}(v_2), \quad \forall\, \alpha_1, \alpha_2 \in F$$

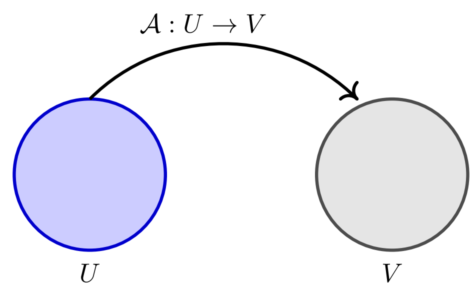


Figure 1.1: $\mathcal{A} : V \to W$ s.t. $\mathcal{A}(v) = w$, $v \in V$, $w \in W$.

Remark. We will show that the operator notation $\mathcal{A}$ operates on $v \in V$ is equivalent to pre-multiplication of v by a matrix if $\mathcal{A}$ is linear.

Example. Consider the following mapping on the set of polynomials of degree 2:

$$\mathcal{A} : as^2 + bs + c \mapsto cs^2 + bs + a$$

is this map linear?

Proof: Let

$$\begin{aligned} v_1 &= a_1 s^2 + b_1 s + c_1 \\ v_2 &= a_2 s^2 + b_2 s + c_2 \end{aligned}$$

Then

$$\begin{aligned} \mathcal{A}(\alpha_1 v_1 + \alpha_2 v_2) &= \mathcal{A}(\alpha_1(a_1 s^2 + b_1 s + c_1) + \alpha_2(a_2 s^2 + b_2 s + c_2)) \\ &= (\alpha_1 c_1 + \alpha_2 c_2)s^2 + (\alpha_1 b_1 + \alpha_2 b_2)s + (\alpha_1 a_1 + \alpha_2 a_2) \\ &= \alpha_1(c_1 s^2 + b_1 s + a_1) + \alpha_2(c_2 s^2 + b_2 s + a_2) \\ &= \alpha_1 \mathcal{A}(v_1) + \alpha_2 \mathcal{A}(v_2) \end{aligned}$$

Hence, $\mathcal{A}$ is a linear map. ■

DIY Exercises: Are the following linear maps?

1.
$$\mathcal{A} : as^2 + bs + c \mapsto \int_0^s (bt + a)\ dt$$

Let $v_1 = a_1 s^2 + b_1 s + c_1$, $v_2 = a_2 s^2 + b_2 s + c_2$

$$\begin{aligned} \mathcal{A}(\alpha_1 v_1 + \alpha_2 v_2) &= \mathcal{A}(\alpha_1(a_1 s^2 + b_1 s + c_1) + \alpha_2(a_2 s^2 + b_2 s + c_2)) \\ &= \mathcal{A}((\alpha_1 a_1 + \alpha_2 a_2)s^2 + (\alpha_1 b_1 + \alpha_2 b_2)s + (\alpha_1 c_1 + \alpha_2 c_2)) \\ &= \int_0^s ((\alpha_1 b_1 + \alpha_2 b_2)t + (\alpha_1 a_1 + \alpha_2 a_2))\ dt \\ &= \int_0^s (\alpha_1 b_1 t + \alpha_1 a_1)\ dt + \int_0^s (\alpha_2 b_2 t + \alpha_2 a_2)\ dt \end{aligned}$$

2.
$$\mathcal{A} : v(t) \mapsto \int_0^1 v(t)\ dt + k$$

for $v(\cdot) \in C([0,1], \mathcal{R})$.

3. $\mathcal{A} : \mathbb{R}^3 \to \mathbb{R}^3$ such that $\mathcal{A}(v) = Av$ with

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 5 \\ 7 & 0 & 16 \end{bmatrix}$$

   Aside: what happens to the zero vector under this (or any) linear map?

Given a linear map $\mathcal{A} : U \to V$ we define the follow two spaces.

Definition 1.2 *The range space of $\mathcal{A}$ to be the subspace*

$$\mathcal{R}(\mathcal{A}) = \{v |\ v = \mathcal{A}(u), u \in U\}$$

The range is also called the image of $\mathcal{A}$.

Definition 1.3 *The null space of $\mathcal{A}$ to be the subspace*

$$\mathcal{N}(\mathcal{A}) = \{u |\ \mathcal{A}(u) = 0\} \subset U$$

The null space is also called the kernel of $\mathcal{A}$.

DIY Exercise. Prove that $\mathcal{R}(\mathcal{A})$ and $\mathcal{N}(\mathcal{A})$ are linear subspaces.

1.2.2 Matrix multiplication as a linear function

Consider $f : \mathbb{R}^n \to \mathbb{R}^m$ given by $f(x) = Ax$ where $A \in \mathbb{R}^{m\times n}$. Matrix multiplication as a function is linear!

Turns out the converse is also true: i.e. **any** linear function $f : \mathbb{R}^n \to \mathbb{R}^m$ can be written as $f(x) = Ax$ for some $A \in \mathbb{R}^{m\times n}$.

representation via matrix multiplication is **unique**: for any linear function f there is only one matrix A for which $f(x) = Ax$ for all x

$y = Ax$ is a concrete representation of a generic linear function

Interpretation of $y = Ax$:

- y is a measurement or observation and x is unknown to be determined
- x is 'input' or 'action'; y is 'output' or 'result'
- $y = Ax$ defines a function or transformation that maps $x \in \mathbb{R}^n$ into $y \in \mathbb{R}^m$

Interpretation of a_{ij}:

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

a_{ij} is a gain factor from j-th input (x_j) to i-th output (y_i). The *sparsity pattern* of A—i.e. list of zero/non-zero elements of A—shows which x_j affect which y_i.

1.2.3 Applications for Linear Maps $y = Ax$

Linear models or functions of the form $y = Ax$ show up in a large number of applications that can be broadly categorized as follows:

estimation or inversion.

- y_i is the i–th measurement or sensor reading (we see this)
- x_j is the j–th parameter to be estimated or determined
- a_{ij} is the sensitivity of the i–th sensor to the j–th parameter

e.g.,

- find x, given y
- find all x's that result in y (i.e., all x's consistent with measurements)
- if there is no x such that $y = Ax$,find x s.t. $y \approx Ax$ (i.e. if the sensor readings are inconsistent, find x which is almost consistent)

control or design.

- x is vector of design parameters or inputs (which we can choose)
- y is vector of results, or outcomes
- A describes how input choices affect results

e.g.,

- find x so that $y = y_{\text{des}}$
- find all x's that result in $y = y_{\text{des}}$ (i.e., find all designs that meet specifications)
- among the x's that satisfy $y = y_{\text{des}}$, find a small one (i.e. find a small or efficient x that meets the specifications)

mapping or transformation.

- x is mapped or transformed to y by a linear function A

e.g.,

- determine if there is an x that maps to a given y
- (if possible) find an x that maps to y
- find all x's that map to a given y
- if there is only one x that maps to y, find it (i.e. decode or undo the mapping)

1.2.4 Other views of matrix multiplication

Matrix Multiplication as

mixture of columns. write $A \in \mathbb{R}^{m\times n}$ in terms of its columns

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

where $a_j \in \mathbb{R}^m$. Then, $y = Ax$ can be written as

$$y = x_1a_1 + x_2a_2 + \cdots + x_na_n$$

x_j's are scalars and a_j's are m dimensional vectors. the x's given the *mixture* or weights for the mixing of columns of A.

One very important example is when $x = e_j$, which is the j–th unit vector.

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ \cdots, \ e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

In this case, $Ae_j = a_j$, the j–th column of A.

inner product with rows. write A in terms of its rows

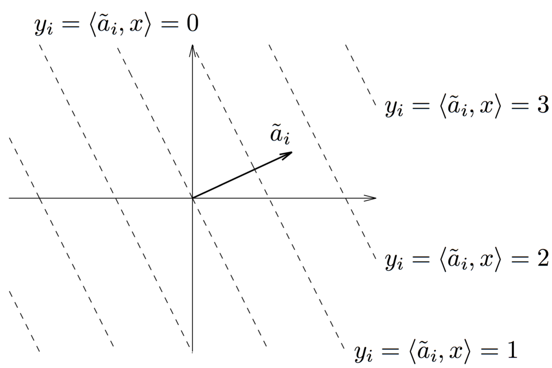
$$A = \begin{bmatrix} \tilde{a}_1^T \\ \tilde{a}_2^T \\ \vdots \\ \tilde{a}_n^T \end{bmatrix}$$

where $\tilde{a}_i \in \mathbb{R}^n$. Then $y = Ax$ can be written as

$$y = \begin{bmatrix} \tilde{a}_1^T x \\ \tilde{a}_2^T x \\ \vdots \\ \tilde{a}_n^T x \end{bmatrix}$$

Thus, $y_i = \langle \tilde{a}_i, x\rangle$, i.e. y_i is the inner product of the i–th row of A with x.

This has a nice geometric interpretation. $y_i = \tilde{a}_i^T x = \alpha$ is a hyperplane in $\mathbb{R}^n$ (normal to $\tilde{a}_i$

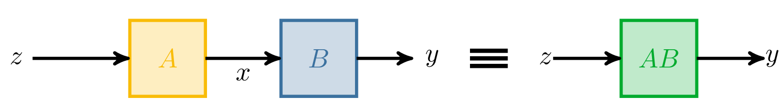


1.2.4.1 Composition of matrices

For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, $C = AB \in \mathbb{R}^{m\times p}$ where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

composition interpretation: $y = Cz$ represents composition of $y = Ax$ and $x = Bz$



We can also interpret this via columns and rows. That is, $C = AB$ can be written as

$$C = \begin{bmatrix} c_1 & \cdots & c_p \end{bmatrix} = AB = \begin{bmatrix} Ab_1 & \cdots & Ab_p \end{bmatrix}$$

i.e. the i–th column of C is A acting on the i–th column of B.

similarly we can write

$$C = \begin{bmatrix} \tilde{c}_1^T & \vdots & \tilde{c}_m^T \end{bmatrix} = AB = \begin{bmatrix} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{bmatrix}$$

i.e. the i–th row of C is the i–th row of A acting (on the left) on B.

as above, we can write entries of C as inner products:

$$c_{ij} = \tilde{a}_i^T b_j = \langle \tilde{a}_i, b_j \rangle$$

i.e. the entries of C are inner products of rows of A and columns of B

- $c_{ij} = 0$ means the i–th row of A is *orthogonal* to the j–th column of B
- (Gram matrices are defined this way): **Gram matrix** of vectors $v_1, \dots, v_n$ defined as $G_{ij} = v_i^T v_j$ (gives inner product of each vector with the others)

$$G = \begin{bmatrix} v_1 & \cdots v_n \end{bmatrix}^T \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

510: **Fall 2019**

Lecture 2: Review of Mathematical Preliminaries & Functions

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

The material in this lecture should be a review.

1 Review of Mathematical Preliminaries

It is assumed that you are familiar with mathematical notation; if not, please take the time to review. Some of the main notion that will be used in this course is summarized below.

Quantifiers.

- $\forall$: for all
- $\exists$: there exists; $\exists!$: there exists a unique; $\nexists$: there does not exist
- $a \implies b$: statement or condition a implies statement or condition b; $\impliedby$ is the opposite.
- $\iff$: if and only if (i.e. $a \iff b$ means $a \implies b$ and $b \implies a$).

Further, $a \implies b$ is equivalent to b is **necessary** for a and a is **sufficient** for b, hence when proving an iff ($\iff$) statement, we often say $\implies$ is the sufficient condition and $\impliedby$ is the necessary condition.

Familiar spaces.

- $\mathbb{R}$: real numbers
- $\mathbb{C}$: complex numbers
- $\mathbb{C}_+$, $\mathbb{R}_+$: right (positive) complex plane or positive reals
- $\mathbb{C}_+^\circ$: open right half complex plane
- $\mathbb{C}_-$, $\mathbb{R}_-$: left (negative) complex plane or negative reals
- $\mathbb{Z}$: integers
- $\mathbb{N}$: natural numbers

Set notation.

- $\in$: "in" e.g., $a \in \mathbb{R}$ means a lives in the real numbers $\mathbb{R}$
- $\notin$: "not in"
- $\subset$: subset e.g., $A \subset B$: the set A is a subset (contained in) the set B
- $\not\subset$: "not contained in"
- $\cup$: union e.g., $A \cup B$ is the union of the sets A and B
- $\cap$: intersection

2 Review of Basic Proof Techniques

It is assumed that you have some basic knowledge of proof techniques. If you do not, please review.

Do not get scared off by the work "prove". Basically, proving something is as simple as constructing a logical argument by invoking a series of facts which enable you to deduce a particular statement or outcome.

There are several basic proof techniques that you can apply including the following. For each of the proof techniques below, consider the following statement

$$\texttt{ExpressionA} \implies \texttt{ExpressionB}$$

- **direct.** The direct method of proof is as its sounds. You start by assuming `Expression A` is true and you try to argue based on other facts that are known to be true and using logical deduction that `Expression B` must hold as a result.
- **contradiction.** There are a number of ways to argue by contradiction, but the most straightforward approach is to assume that `Expression A` holds but `Expression B` does not. Then reason why this assumption therefore leads to a contradicting state of the world.
- **contrapositive.** To construct a proof by this approach, you prove the contrapositive statement by direct proof. That is, you use the direct proof method to prove

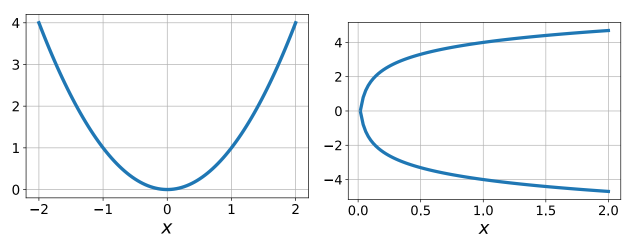
$$\sim \texttt{ExpressionB} \implies \sim \texttt{ExpressionA}$$

  where $\sim$ `ExpressionB` means "not `Expression B`" (i.e. `Expression B` does not hold.
- **disprove by counterexample.** a statement such as the one above can also be shown to be false by construction of a counterexample. This is an specific example which shows that the statement cannot be true.

3 Functions

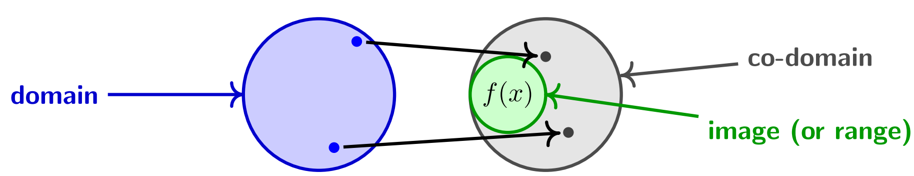
First, what is a function? And, what kind of notation will we use?

Given two sets X and Y, by $f : X \to Y$ we mean that $\forall x \in X$, f assigns a *unique* $f(x) \in Y$.



(a) function (b) not function

Figure 1: Example: a function and something that is not a function.



$$f : X \to Y$$

- f maps X into Y (alternative notation: $f : x \mapsto y$.
- image of f (range):

$$f(X) = \{f(x)|\ x \in X\}$$

3.1 Characterization of functions

Injectivity (one-to-one):

$$f \text{ is injective } \iff [f(x_1) = f(x_2) \implies x_1 = x_2] \iff [x_1 \neq x_2 \implies f(x_1) \neq f(x_2)]$$

Surjectivity (onto):

$$f \text{ is surjective} \iff \forall\, y \in Y,\ \exists\, x \in X, \text{ such that } y = f(x)$$

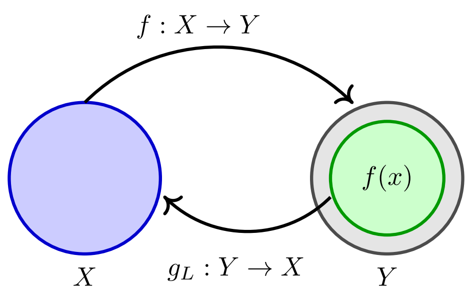
Bijective (injective and surjective):

$$\forall\, y \in Y,\ \exists!\, x \in X \text{ s.t. } y = f(x)$$

Example. Consider $f : X \to Y$ and let $\mathbf{1}_X$ be the identity map on X.

Left inverse. the *left inverse* of f is defined as the map $g_L : Y \to X$ such that $g_L \circ f \equiv \mathbf{1}_X$. (composition: $g_L \circ f : X \to X$ such that $g_L \circ f : x \mapsto g_L(f(x))$)

Right inverse. the *right inverse* of f is defined as the map $g_R : X \to Y$ such that $f \circ g_R \equiv \mathbf{1}_Y$.



Exercise. Prove that

$$[f \text{ has a left inverse } g_L] \iff [f \text{ is injective}]$$

Proof. Direct Proof of $\Longleftarrow$: Assume f is injective. Then, by definition, for any x_1, x_2,

$$f(x_1) = f(x_2) \quad \implies \quad x_1 = x_2$$

Construct a mapping $g_L : Y \to X$ such that, on $f(X) = \{f(x), \forall x \in X\}$, $g_L(f(x)) = x$. This mapping is in fact a well-defined function due to injectivity of f. That is, since f is injective, for each x there is a unique $f(x)$. Hence, this mapping satisfies the definition of a function. Hence, $g_L \circ f \equiv \mathbf{1}_X$.

Contradiction proof of $\Longrightarrow$: Assume f has a left inverse g_L but f is not injective. Then, by definition

$$g_L \circ f \equiv \mathbf{1}_X \implies \forall\, x \in X,\ g_L(f(x)) = x.$$

Since we have assumed that f is not injective, $\exists$ some $x_1 \neq x_2$ such that $f(x_1) = f(x_2)$. Using the definition of left inverse $g_L(f(x_1)) = x_1$ and $g_L(f(x_2)) = x_2$. Yet, since g_L is a function $x_1 = x_2$ (since $f(x_1) = f(x_2)$) which contradicts our assumption ($\rightarrow\leftarrow$). Thus f has to be injective. $\square$

Remark. Many of the homework problems will require you to derive arguments such as the above. So I will try to derive some of these in the class. Please ask questions if you feel uncomfortable at all.

DIY Exercise. prove that

$$[f \text{ has a right inverse } g_R] \Longleftrightarrow [f \text{ is surjective}]$$

Remark. DIY exercises you find in the notes will sometimes appear in the official homeworks but not always. They are called out in the notes to draw your attention to practice problems that will help you grasp the material.

Two-sided inverse. A function g is called a *two-sided inverse* or simply, an *inverse* of f (and is denoted f^{-1})

- $\Longleftrightarrow g \circ f \equiv \mathbf{1}_X$ and $f \circ g \equiv \mathbf{1}_Y$ (that is, g is both a left and right inverse of f)
- $\Longleftrightarrow f$ is invertible, f^{-1} exists

The following is a commutative diagram which we can use to illustrate these functional properties:

$$\begin{tikzcd} X \arrow[rr, "f \circ g"] \arrow[dr, "g"'] & & Z \\ & Y \arrow[ur, "f"'] & \end{tikzcd}$$

4 Linear functions

A map (or function) $f : X \to Y$ is linear if

- $f(x_1 + x_2) = f(x_1) + f(x_2),\ \forall x_1, x_2 \in X$
- $f(\alpha x) = \alpha f(x)$, $\forall x \in X$ and $\forall\ \alpha \in \mathbb{R}$.

Note. We can also check this definition by combining the two properties. Indeed, f is linear iff

$$f(\alpha x_1 + \beta x_2) = \alpha f(x_1) + \beta f(x_2),\ \forall\ x_1, x_2 \in X,\ \forall \alpha, \beta \in \mathbb{R}$$

Example. Consider the following mapping on the set of polynomials of degree 2:

$$f : as^2 + bs + c \mapsto cs^2 + bs + a$$

is this map linear?

Proof. Let

$$\begin{aligned} v_1 &= a_1 s^2 + b_1 s + c_1 \\ v_2 &= a_2 s^2 + b_2 s + c_2 \end{aligned}$$

Then

$$\begin{aligned} f(\alpha_1 v_1 + \alpha_2 v_2) &= f(\alpha_1(a_1 s^2 + b_1 s + c_1) + \alpha_2(a_2 s^2 + b_2 s + c_2)) \\ &= (\alpha_1 c_1 + \alpha_2 c_2)s^2 + (\alpha_1 b_1 + \alpha_2 b_2)s + (\alpha_1 a_1 + \alpha_2 a_2) \\ &= \alpha_1(c_1 s^2 + b_1 s + a_1) + \alpha_2(c_2 s^2 + b_2 s + a_2) \\ &= \alpha_1 f(v_1) + \alpha_2 f(v_2) \end{aligned}$$

Hence, f is a linear map. $\square$

DIY Exercise. Is the following map linear map?

$$f : as^2 + bs + c \mapsto \int_0^s (bt + a)\ dt$$

Let $v_1 = a_1 s^2 + b_1 s + c_1$, $v_2 = a_2 s^2 + b_2 s + c_2$

$$\begin{aligned}
f(\alpha_1 v_1 + \alpha_2 v_2) &= f(\alpha_1(a_1 s^2 + b_1 s + c_1) + \alpha_2(a_2 s^2 + b_2 s + c_2)) \\
&= f((\alpha_1 a_1 + \alpha_2 a_2)s^2 + (\alpha_1 b_1 + \alpha_2 b_2)s + (\alpha_1 c_1 + \alpha_2 c_2)) \\
&= \int_0^s ((\alpha_1 b_1 + \alpha_2 b_2)t + (\alpha_1 a_1 + \alpha_2 a_2))\ dt \\
&= \int_0^s (\alpha_1 b_1 t + \alpha_1 a_1)\ dt + \int_0^s (\alpha_2 b_2 t + \alpha_2 a_2)\ dt
\end{aligned}$$

4.1 Matrix multiplication as a linear function

Consider $f : \mathbb{R}^n \to \mathbb{R}^m$ given by $f(x) = Ax$ where $A \in \mathbb{R}^{m \times n}$. Matrix multiplication as a function is linear!

Turns out the converse is also true!

Matrix representation theorem. any linear function $f : \mathbb{R}^n \to \mathbb{R}^m$ can be written as $f(x) = Ax$ for some $A \in \mathbb{R}^{m \times n}$.

e.g.,

$$\mathcal{L}(X) = A^T X + XA$$

$$\begin{aligned}
\mathcal{L}(a_1 X_1 + a_2 X_2) &= A^T(a_1 X_1 + a_2 X_2) + (a_1 X_1 + a_2 X_2)A \\
&= a_1(A^T X_1 + X_1 A) + a_2(A^T X_2 + X_2 A) \\
&= a_1 \mathcal{L}(X_1) + a_2 \mathcal{L}(X_2)
\end{aligned}$$

Suppose that $X, A \in \mathbb{R}^{2\times 2}$. We can find the matrix representation by viewing $\mathcal{L}$ as a map from $\mathbb{R}^4$ to $\mathbb{R}^4$ by first expressing

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},\ X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$$

Then, if you actually expressed $\mathcal{L}(X)$ in terms of this decomposition and took into consideration the mapping

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

You would find that

$$\mathcal{L}(X) = LX$$

where

$$L = \begin{bmatrix} 2A_{11} & A_{21} & A_{21} & 0 \\ A_{12} & A_{11} + A_{22} & 0 & A_{21} \\ A_{12} & 0 & A_{22} + A_{11} & A_{21} \\ 0 & A_{12} & A_{21} & 2A_{22} \end{bmatrix}$$

e.g., **derivatives of nonlinear maps are linear operators** Suppose we are given a *non-linear* map (or function) $f : \mathbb{R}^n \to \mathbb{R}^m$ which is differentiable at a point $x_0 \in \mathbb{R}^n$. Then, for any x "near" x_0, $f(x)$ is "very near" $f(x_0) + Df(x_0)(x - x_0)$ where

$$(Df(x_0))_{ij} = \frac{\partial f_i}{\partial x_j}\Big|_{x_0}$$

is the derivative (Jacobian) matrix such that $Df(x_0) \in \mathbb{R}^{m\times n}$.

With $y = f(x)$, $y_0 = f(x_0)$, we define the input deviation (or variation) as $\delta x = x - x_0$ and output deviation as $\delta y = y - y_0$ so that $\delta y \approx Df(x_0)\delta x$.

When deviations are small, they are approximately related by a linear function.

Such a representation via matrix multiplication is **unique**: for any linear function f there is only one matrix A for which $f(x) = Ax$ for all x

The expression $y = Ax$ is a concrete representation of a generic linear function
Interpretation of $y = Ax$:

- y is a measurement or observation and x is unknown to be determined
- x is 'input' or 'action'; y is 'output' or 'result'
- $y = Ax$ defines a function or transformation that maps $x \in \mathbb{R}^n$ into $y \in \mathbb{R}^m$

Interpretation of a_{ij}:

$$y_i = \sum_{j=1}^{n} a_{ij}x_j$$

a_{ij} is a gain factor from j-th input (x_j) to i-th output (y_i). The *sparsity pattern* of A—i.e. list of zero/non-zero elements of A—shows which x_j affect which y_i.

4.2 Applications for Linear Maps $y = Ax$

Linear models or functions of the form $y = Ax$ show up in a large number of applications that can be broadly categorized as follows:

estimation or inversion.

- y_i is the i–th measurement or sensor reading (we see this)
- x_j is the j–th parameter to be estimated or determined
- a_{ij} is the sensitivity of the i–th sensor to the j–th parameter

e.g.,

- find x, given y
- find all x's that result in y (i.e., all x's consistent with measurements)
- if there is no x such that $y = Ax$,find x s.t. $y \approx Ax$ (i.e. if the sensor readings are inconsistent, find x which is almost consistent)

control or design.

- x is vector of design parameters or inputs (which we can choose)
- y is vector of results, or outcomes
- A describes how input choices affect results

e.g.,

- find x so that $y = y_{\text{des}}$
- find all x's that result in $y = y_{\text{des}}$ (i.e., find all designs that meet specifications)
- among the x's that satisfy $y = y_{\text{des}}$, find a small one (i.e. find a small or efficient x that meets the specifications)

mapping or transformation.

- x is mapped or transformed to y by a linear function A

e.g.,

- determine if there is an x that maps to a given y
- (if possible) find an x that maps to y
- find all x's that map to a given y
- if there is only one x that maps to y, find it (i.e. decode or undo the mapping)

4.3 Other views of matrix multiplication

Matrix Multiplication as

mixture of columns. write $A \in \mathbb{R}^{m\times n}$ in terms of its columns

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

where $a_j \in \mathbb{R}^m$. Then, $y = Ax$ can be written as

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

x_j's are scalars and a_j's are m dimensional vectors. the x's given the *mixture* or weights for the mixing of columns of A.

One very important example is when $x = e_j$, which is the j–th unit vector.

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},\ e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix},\ \cdots,\ e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

In this case, $Ae_j = a_j$, the j–th column of A.

inner product with rows. write A in terms of its rows

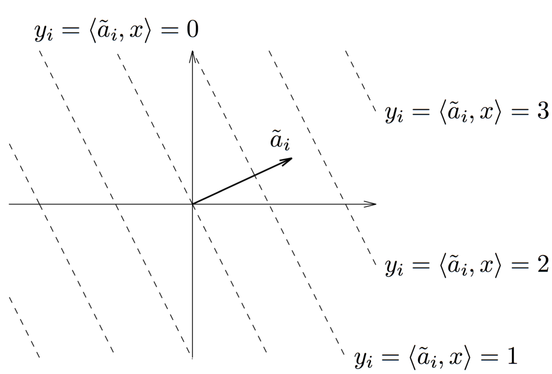
$$A = \begin{bmatrix} — & \tilde{a}_1^T & — \\ — & \tilde{a}_2^T & — \\ \vdots & & \\ — & \tilde{a}_n^T & — \end{bmatrix} = \begin{bmatrix} \langle \tilde{a}_1, x\rangle \\ \vdots \\ \langle \tilde{a}_n, x\rangle \end{bmatrix}$$

where $\tilde{a}_i \in \mathbb{R}^n$. Then $y = Ax$ can be written as

$$y = \begin{bmatrix} \tilde{a}_1^T x \\ \tilde{a}_2^T x \\ \vdots \\ \tilde{a}_n^T x \end{bmatrix}$$

Thus, $y_i = \langle \tilde{a}_i, x\rangle$, i.e. y_i is the inner product of the i–th row of A with x.

This has a nice geometric interpretation. $y_i = \tilde{a}_i^T x = \alpha$ is a hyperplane in $\mathbb{R}^n$ (normal to $\tilde{a}_i$

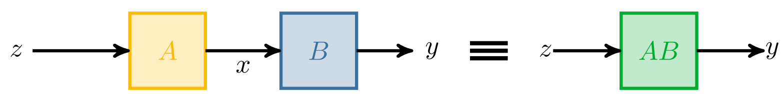


4.4 Composition of matrices

For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, $C = AB \in \mathbb{R}^{m\times p}$ where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

composition interpretation: $y = Cz$ represents composition of $y = Ax$ and $x = Bz$



We can also interpret this via columns and rows. That is, $C = AB$ can be written as

$$C = \begin{bmatrix} c_1 & \cdots & c_p \end{bmatrix} = AB = \begin{bmatrix} Ab_1 & \cdots & Ab_p \end{bmatrix}$$

i.e. the i–th column of C is A acting on the i–th column of B.

Similarly we can write

$$C = \begin{bmatrix} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_m^T \end{bmatrix} = AB = \begin{bmatrix} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{bmatrix}$$

i.e. the i–th row of C is the i–th row of A acting (on the left) on B.

as above, we can write entries of C as inner products:

$$c_{ij} = \tilde{a}_i^T b_j = \langle \tilde{a}_i, b_j \rangle$$

i.e. the entries of C are inner products of rows of A and columns of B

- $c_{ij} = 0$ means the i–th row of A is *orthogonal* to the j–th column of B
- (Gram matrices are defined this way): **Gram matrix** of vectors $v_1, \ldots, v_n$ defined as $G_{ij} = v_i^T v_j$ (gives inner product of each vector with the others)

$$G = \begin{bmatrix} v_1 & \cdots v_n \end{bmatrix}^T \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

510: **Fall 2019**

Lecture 3: Vector Spaces and Linear Spaces

Lecturer: L.J. Ratliff

References: Appendix A.2 & A.3 of **[C&D]**; Chapter 1 and 2 of **[Ax]**

1 A Review of Fields and Rings

Definition 1. A field $\mathbb{F}$ is an object consisting of a set of elements and two binary operations

1. addition $(+)$
2. multiplication $(\cdot)$

such that the following axioms are obeyed:

- *Addition $(+)$ satisfies the following*: for any $\alpha,\ \beta,\ \gamma \in \mathbb{F}$,
  - associative: $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$
  - commutative: $\alpha+\beta = \beta+\alpha$
  - $\exists$ identity element $0 \in \mathbb{F}$ such that $\alpha + 0 = \alpha$
  - $\exists$ inverse: $\forall \alpha \in \mathbb{F}$, $\exists(-\alpha) \in \mathbb{F}$ such that $\alpha + (-\alpha) = 0$
- *Multiplication $(\cdot)$ satisfies the following*: for any $\alpha,\ \beta,\ \gamma \in \mathbb{F}$,
  - associative: $(\alpha\cdot\beta)\cdot\gamma = \alpha\cdot(\beta\cdot\gamma)$
  - commutative: $\alpha\cdot\beta = \beta\cdot\alpha$
  - $\exists$ identity element 1 such that $\alpha \cdot 1 = \alpha \ \ \forall \alpha \in \mathbb{F}$
  - $\exists$ inverse: $\forall \alpha \neq 0$, $\exists \alpha^{-1}$ such that $\alpha\cdot\alpha^{-1} = \alpha^{-1}\cdot\alpha = 1$
  - distributive (over $(+)$):

$$\begin{aligned} \alpha\cdot(\beta+\gamma) &= \alpha\cdot\beta + \alpha\cdot\gamma \\ (\beta+\gamma)\cdot\alpha &= \beta\cdot\alpha + \gamma\cdot\alpha \end{aligned}$$

Note. Throughout this course, a field $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. If not clear from context, assume $\mathbb{C}$ as it is the more general case.

Examples. (see if you can prove that each of the following are fields)

- $\mathbb{R}$: reals
- $\mathbb{C}$: complex numbers
- $\mathbb{R}(s)$: the field of rational functions in s with coefficients in $\mathbb{R}$—i.e.

$$\frac{s^2+3s+1}{s+1}$$

  (This is important for frequency domain analysis of controllability, observability, stability, etc. You will see this, e.g., in 547)
- $\mathbb{C}(s)$: field of rational functions in s with coefficients in $\mathbb{C}$

DIY Exercise. Show that
- the polynomials in s w/ coefficients in $\mathbb{R}$ (denoted by $\mathbb{R}[s]$) is *not a field*
- strictly proper rational functions, denoted by $\mathbb{R}_{p,o}(s)$ is *not a field*

Answer:

- $sf(s) = 1 \in \mathbb{R}[s]$ implies $0 = 1$ in $\mathbb{R}$ by taking $s = 0$. Or you can see this via the fact that $1/x$ is not a polynomial. More generally, we know that for polynomials $\deg(fg) = \deg(f) + \deg(g)$ for any polynomials f, g. Any polynomial of degree $n \geq 1$ would then need a polynomial of degree $-n$ as an inverse, which does not exist in the ring of polynomials.
- A similar argument can be made for strictly proper rational functions.

Definition 2 (Ring)**.** A ring is the same as a field *except* the multiplication operator is not commutative and there is no inverse for non-zero elements under the multiplication operator.

Examples.

- *Commutative Rings*: $\mathbb{Z}, \mathbb{R}[s], \mathbb{C}[s], \mathbb{R}_{p,o}(s)$
- *Non-commutative rings*: $\mathbb{R}^{n\times n}, \mathbb{C}^{n\times n}$, $\mathbb{R}^{n\times n}[s]$, $\mathbb{C}^{n\times n}[s]$, $\mathbb{R}^{n\times n}(s)$, $\mathbb{C}^{n\times n}(s)$, $\mathbb{R}_p^{n\times n}(s)$, ...

DIY Exercise. Show that the set $\{0, 1\}$ with
- $(\cdot)$ = binary AND
- $(+)$ = binary XOR

is a field.

2 Vector Spaces (or Linear Spaces)

Before formally defining a vector space, let's consider some examples.

- A plane (set of all pairs of real numbers): $\mathbb{R}^2 = \{(x, y)|\ x, y \in \mathbb{R}\}$
- More generally, $\mathbb{R}^n = \{(x_1, \ldots, x_n)|\ x_i \in \mathbb{R}, i \in \{1, \ldots, n\}\}$

Definition. A vector space $(V, \mathbb{F})$ is a set of vectors V, and a field of scalars $\mathbb{F}$, and two binary operations:

- vector addition '+'
- scalar multiplication '$\cdot$' (multiplication of vectors by scalars)

such that

- *Addition* $(+ : V \times V \to V,\ + : (x, y) \mapsto x + y)$ satisfies the following: for any $x,\ y,\ z \in \mathbb{F}$,
  - associative: $(x + y) + z = x + (y + z)$
  - commutative: $x + y = y + x$
  - $\exists$ identity $\mathbf{0}$ ("zero vector") such that $x + \mathbf{0} = \mathbf{0} + x = x$
  - $\exists$ inverse: $\forall x \in V$, $\exists(-x) \in \mathbb{F}$ such that $x + (-x) = \mathbf{0}$
- *Scalar Multiplication* $(\cdot : \mathbb{F} \times V \to V,\ \cdot : (\alpha, x) \mapsto \alpha \cdot x)$ satisfies the following: for any $x,\ y \in V$ and for any $\alpha,\ \beta \in \mathbb{F}$,
  - $(\alpha \cdot \beta) \cdot x = \alpha(\beta \cdot x)$
  - $1 \cdot x = x$
  - $0 \cdot x = \mathbf{0}$

  where 1 and 0 are the multiplicative and additive identities of the field $\mathbb{F}$.

- distributive laws:

$$\forall x \in V,\ \forall \alpha, \beta \in F \quad (\alpha+\beta)x = \alpha \cdot x + \beta \cdot x$$
$$\forall x, y \in V,\ \forall \alpha \in F \quad \alpha(x+y) = \alpha \cdot x + \alpha \cdot y$$

Examples.

1. $(\mathbb{F}^n, \mathbb{F})$ the space of n-tuples in $\mathbb{F}$ over the field $\mathbb{F}$ is a vector space. (i.e. $(\mathbb{R}^n, \mathbb{R})$, $(\mathbb{C}^n, \mathbb{C})$)
2. The function space $\mathcal{F}(U, V)$ defined by: $(V, \mathbb{F})$ is a vector space and U is a set (i.e. $\mathbb{R}, \mathbb{R}^n, \ldots$). Hence $\mathcal{F}(U, V)$ is the class of all functions which map U to V. Then, $(\mathcal{F}(U, V), \mathbb{F})$ with
   - addition: $(f+g)(u) = f(u) + g(u)$, $f, g \in \mathcal{F}(U, V)$, $u \in U$
   - scalar multiplication: $(\alpha f)u = \alpha f(u)$, $\forall u \in U$.

   is a vector space. e.g.,
   - Continuous functions $f : [t_0, t_1] \to \mathbb{R}^n$—notation

   $$(C([t_0, t_1], \mathbb{R}), \mathbb{R})$$

   - k–times differentiable functions $f : [t_0, t_1] \to \mathbb{R}^n$—notation

   $$(C^k([t_0, t_1], \mathbb{R}^n), \mathbb{R})$$

3. Another example of vector spaces of functions:

   $$V = \{\xi : \mathbb{R}_+ \to \mathbb{R}^n |\ \xi \text{ is differentiable}\}$$

   where vector sum the sum of functions, i.e.

   $$(\xi + \phi)(t) = \xi(t) + \phi(t)$$

   and scalar multiplication is defined by

   $$(\alpha\xi)(t) = \alpha\xi(t)$$

   e.g., a point in V is a trajectory in $\mathbb{R}^n$ or the flow of a differential equation for example.

   Now, consider

   $$W = \{\xi \in V |\ \dot{\xi} = A\xi\}$$

   points in W are trajectories of the linear system $\dot{\xi} = A\xi$

   DIY Exercise. Show that W is a subspace of V.

3 Vector subspaces (or linear subspaces)

Let $(V, \mathbb{F})$ be a linear space and W a subset of V (denoted $W \subset V$). Then $(W, \mathbb{F})$ is called a *subspace* of $(V, \mathbb{F})$ if $(W, \mathbb{F})$ is itself a vector space.

How to check if W is a subspace of V?

step 1. verify that W is a subset of V (thus W inherits the vector space axioms of V)

step 2. verify *closure* under vector addition and scalar multiplication—i.e. $\forall w_1, w_2 \in W$, $\forall \alpha \in \mathbb{F}$,

$$\alpha w_1 + \alpha w_2 \in W$$

This is equivalent to...

Defn. A subspace W of V is a subset of V such that

1. $0 \in W$
2. $u, v \in W \implies u + v \in W$
3. $u \in W$, $\alpha \in \mathbb{F} \implies \alpha u \in W$

Examples.

- Is $\{(x_1, x_2, 0) |\ x_1, x_2 \in \mathbb{F}\}$ is a subspace of $\mathbb{F}^3$? (e.g., is a plane in $\mathbb{R}^3$ is a subspace?)
- Prove that if W_1, W_2 are subspaces of V, then
  i) $W_1 \cap W_2$ is a subspace
  ii) $W_1 \cup W_2$ is <u>*not*</u> necessarily a subspace

Proof. 1. simple; **DIY Exercise**
2. for (i), consider two arbitrary subspaces of V. Then to show that $W_1 \cap W_2$ is a subspace, we need to show that $W_1 \cap W_2 \subset V$, which is obvious since $W_i \subset V$ for each i and $W_1 \cap W_2 \subset W_1$ and $W_1 \cap W_2 \subset W_2$. Now we need to check its closed under vector addition and scalar multiplication: let $v_1, v_2 \in W_1 \cap W_2$ and $\alpha_1, \alpha_2 \in \mathbb{F}$. Then, we know that $\alpha_1 v_1 + \alpha_2 v_2 \in W_1$ and $\alpha_1 v_1 + \alpha_2 v_2 \in W_2$ since W_1 and W_2 are subspaces and $v_i \in W_1 \cap W_2$ (i.e. its in both spaces) for each $i = 1, 2$. Hence,

$$\alpha_1 v_1 + \alpha_2 v_2 \in W_1 \cap W_2$$

and we are done.
for (ii) a counterexample is sufficient: Let

$$W_1 = \text{span}\left\{\begin{bmatrix}1\\0\end{bmatrix}\right\}$$

$$W_2 = \text{span}\left\{\begin{bmatrix}0\\1\end{bmatrix}\right\}$$

and $V = \mathbb{R}^2$. Consider

$$w = \begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix}$$

but $w \notin W_1 \cup W_2$. Hence, $W_1 \cup W_2$ is not closed under vector addition and thus is not a subspace.

□

4 Sums and Direct Sums

Suppose $W_1, \ldots, W_m$ are subspaces of V. The sum of the W_i's, denoted by $W_1 + \cdots + W_m$ is defined to be the set of all possible sums of elements in $W_1, \ldots, W_m$—i.e.,

$$W_1 + \cdots + W_m = \{w_1 + \cdots + w_m |\ w_i \in W_i, \forall i \in \{1, \ldots, m\}\}$$

DIY Exercise. Show the following:

$$W_1, \ldots, W_m \text{ subspaces of } V \implies \sum_{i=1}^{m} W_i \text{ subspace of } V$$

Example. Consider $U = \{(x, 0, 0) \in \mathbb{F}^3 |\ x \in \mathbb{F}\}$ and $W = \{(0, y, 0) \in \mathbb{F}^3 |\ y \in \mathbb{F}\}$. Show that $U + W = \{(x, y, 0) |\ x, y \in \mathbb{F}\}$ is a subspace of $\mathbb{F}^3$.

Suppose that $W_1, \ldots, W_m$ are subspaces of V such that $V = W_1 + \cdots + W_m$. We say that V is the <u>direct sum</u> of subspaces $W_1, \ldots, W_m$, written as

$$V = W_1 \oplus \cdots \oplus W_m$$

if each element of V can be written uniquely as a sum $w_1 + \cdots + w_m$ where each $w_i \in W_i$.

Example. Consider $U = \{(x, y, 0) \in \mathbb{F}^3 |\ x, y \in \mathbb{F}\}$ and $W = \{(0, 0, z) \in \mathbb{F}^3 |\ z \in \mathbb{F}\}$. Then $\mathbb{F}^3 = U \oplus W$.

Proposition 1. Suppose that $U_1, \dots, U_m$ are subspaces of V. Then, $V = U_1 \oplus \cdots \oplus U_m$ if and only if both the following conditions hold:

1. $V = U_1 + \cdots + U_m$
2. the only way to write zero as a sum $u_1 + \cdots + u_m$, where each $u_i \in U_i$ is by taking all the $u_i = 0$, $i \in \{1, \dots, m\}$.

Proof. ($\Longrightarrow$): Suppose that $V = U_1 \oplus \cdots \oplus U_m$. By definition, 1. holds. Now, we need to show 2. Suppose $u_j \in U_j$ for each $j \in \{1, \dots, m\}$ and that

$$0 = u_1 + \cdots + u_m$$

Then, each $u_j = 0$ due to the uniqueness part of the definition of the direct sum—i.e., recall that a space is a direct sum of U_j, $j \in \{1, \dots, m\}$ if each element in that space can be written uniquely as a sum $u_1 + \cdots + u_m$. Hence, $0 = 0 + \cdots + 0$ where $0 \in U_j$.

($\Longleftarrow$): Suppose that 1. and 2. hold. Let $v \in V$. By 1.,

$$v = u_1 + \cdots + u_m$$

for some $u_j \in U_j$, $j \in \{1, \dots, m\}$. To show that this representation is unique, suppose that we also have

$$v = v_1 + \cdots + v_m$$

where $v_j \in U_j$, $j \in \{1, \dots, m\}$. Subtracting these two expressions for v we have that

$$0 = (u_1 - v_1) + \cdots + (u_m - v_m)$$

so that $u_j - v_j \in U_j$ so that the equation above and 2. imply that each $u_j - v_j = 0$. Thus $u_j = v_j$, $j \in \{1, \dots, m\}$. ☐

Proposition 2. Suppose that U and W are subspaces of V. Then $V = U \oplus W$ if and only if $V = U + W$ and $U \cap W = \{0\}$.

Proof. **DIY Exercise.** ☐

5 Linear Independence and Dependence

Suppose $(V, \mathbb{F})$ is a linear space. A linear combination of vectors $(v_1, \dots, v_p)$, $v_i \in V$ is a vector of the form

$$\alpha_1 v_1 + \cdots + \alpha_p v_p, \ \alpha_i \in \mathbb{F}$$

- The set of vectors $\{v_i \in V, \ i \in \{1, \dots, p\}\}$ is said to be linearly independent if and only if (iff)

  $$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_p v_p = 0, \ \alpha_i \in \mathbb{F} \implies \alpha_i = 0, \ \forall i = 1, \dots, p$$

  where $\alpha_i \in F$.
- The set of vectors is said to be linearly dependent iff $\exists$ scalars $\alpha_i \in \mathbb{F}$, $i \in \{1, \dots, p\}$ *not all zero* such that

  $$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_p v_p = 0$$

Example. Let $\mathbb{F} = \mathbb{R}$, $k \in \{0, 1, 2, \dots, n\}$, $f_k : [-1, 1] \to \mathbb{R}$ such that $f_k(t) = t^k$. Show that the set of vectors $(f_k)_{k=0}^n$ is linearly independent in $(\mathcal{F}_n([-1, 1], \mathbb{R}), \mathbb{R})$.

proof sketch.

$$(f_k)_0^n = \{f_0, f_1, \dots, f_n\}$$

Hence we need to show that, $\forall \ t \in [-1, 1]$,

$$\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n = 0 \implies \alpha_i = 0, \alpha_i \in \mathbb{R}$$

DIY Exercise. DIY Exercise. Finish proof.

Proof. To show this is the case, take derivatives of

$$\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n = 0$$

up to the n–th derivative to get a system of equations. Indeed,

$$\begin{aligned} \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n &= 0 \\ \alpha_1 + 2\alpha_2 t + 3\alpha_2 t^2 + \cdots + n\alpha_n t^{n-1} &= 0 \\ \vdots &= \vdots \\ n!\alpha_n &= 0 \end{aligned}$$

so that

$$\begin{bmatrix} 1 & t & t^2 & \cdots & t^{n-1} & t^n \\ 0 & 1 & 2t & \cdots & (n-1)t^{n-2} & nt^{n-1} \\ 0 & 0 & 2 & \cdots & (n-1)(n-2)t^{n-3} & n(n-1)t^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & n! \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The determinant of this matrix is $2!3!\cdots(n-1)!n!$ which is the product of the diagonal (since its upper triangular). This is clearly non-zero for any t. Thus there is a unique solution and the only solution is when all the α's are zero.

□

Span. The set of all linear combinations of $(v_1, \ldots, v_p)$ is called the span of $(v_1, \ldots, v_p)$. We use the notation

$$\operatorname{span}(v_1, \ldots, v_p) = \left\{ \sum_{i=1}^{p} \alpha_i v_i \,\middle|\, \alpha_i \in \mathbb{F},\ i \in \{1, \ldots, p\} \right\}$$

If $\operatorname{span}(v_1, \ldots, v_p) = V$, then we say $(v_1, \ldots, v_p)$ **spans** V.

Note. If some vectors are removed from a linearly independent list, the remaining list is also linearly independent, as you should verify.

Dimension. The notion of a spanning set let us define **dimension**. If we can find a set of spanning vectors with finite cardinality for a space V, then V is said to be finite dimensional, and otherwise it is said to be infinite dimensional. We denote by $\dim(V)$ the dimension of the space V.

Example.
1. $(\mathbb{R}^{n\times n}, \mathbb{R})$, the space of $n \times n$ real valued matrices, has dimension n^2.
2. $\mathrm{PC}([0,1], \mathbb{R})$ is infinite dimensional. Indeed, this can be shown by showing it contains a subspace of infinite dimension: $\operatorname{span}\{(t \to t^n)_0^\infty\}$.

The following is an alternative representation of linear dependence. Indeed, it states that given a linearly dependent list of vectors, with the first vector not zero, one of the vectors is in the span of the previous ones and, furthermore, we can throw out that vector without changing the span of the original list.

Lemma 1. If $(v_1, \ldots, v_p)$ is a linearly dependent set of vectors in V with $v_1 \neq 0$, then there exists $j \in \{2, \ldots, p\}$ such that the following hold:

(a) $v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$

(b) if the j–th term is removed from $(v_1, \ldots, v_p)$, then the span of the remaining list equals span$(v_1, \ldots, v_p)$

This lemma let's us prove a stronger result: the cardinality of linearly independent sets is always smaller than the cardinality of spanning sets.

Theorem 1. In a finite-dimensional vector space, the length of every linearly independent set of vectors is less than or equal to the length of every spanning set of vectors.

DIY Exercise. Construct a proof using the above lemma.

Proposition 3. Every subspace of a finite-dimensional vector space is finite dimensional.

You cannot embed an infinite dimensional space in a finite dimensional one.

6 Basis and Coordinate Representation

With these concepts, we can introduce the notion of a basis for a space V.

Definition 3. Suppose $(V, \mathbb{F})$ is a linear space. Then a set of vectors $\{v_1, v_2, \ldots, v_n\}$ is called a **basis** of V if

(a) $\{v_1, v_2, \ldots, v_n\}$ spans V
(b) $\{v_1, v_2, \ldots, v_n\}$ is a linearly independent set.

Example. The canonical example is the standard basis in $\mathbb{R}^n$—i.e.,

$$\{e_1, \ldots, e_n\}$$

where each vector e_i is all zeros with the exception of the i–th entry which is has value one.

Why do we care about bases? Any vector $x \in V$ may be written as a *linear combination* of the basis vectors. Indeed, consider a basis $\{v_i\}_{i=1}^n$ of V and a vector $x \in V$. Then,

$$x = \xi_1 v_1 + \xi_2 v_2 + \cdots \xi_n v_n = \sum_{i=1}^n \xi_i v_i$$

The vector

$$\xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} \in \mathbb{F}^n$$

is called the <u>coordinate vector</u> of x w.r.t. $\{v_i\}_{i=1}^n$. The ξ_i's are called the <u>coordinates</u> of x w.r.t. $\{v_i\}_{i=1}^n$.

Generating Subspaces w/ Linear Combinations. Using linear combinations we can generate subspaces, as follows. Recall the notion of **span**.

Proposition 4. For a non-empty subset S of $\mathbb{R}^n$, show that span(S) is always a subspace of $\mathbb{R}^n$.

Proof. **DIY Exercise**.

Let $V = \text{span}(S)$ be the subspace of $\mathbb{R}^n$ spanned by some $S \subseteq \mathbb{R}^n$. Then S is said to **generate** or **span** V, and to be a **generating** or **spanning set** for V.

Note. If V is already known to be a subspace, then finding a spanning set S for V can be useful, because it is often easier to work with the smaller spanning set than with the entire subspace V, e.g., if we are trying to understand the behavior of linear transformations on V.

Example.

1. Let S be the unit circle in $\mathbb{R}^3$ which lies in the xy plane. Then span(S) is the entire xy plane.
2. Let
$$S = \{(x, y, z) \in \mathbb{R}^3 |\ x = y = 0,\ 1 < z < 3\}$$
Then span(S) is the z-axis.

If turns out the coordinate representation with respect to a basis is unique.

Proposition 5. A set $\{v_1, \ldots, v_p\}$ of vectors in V is a basis of V iff every $v \in V$ can be written uniquely in the form
$$v = \xi_1 v_1 + \cdots + \xi_p v_p,\ \xi_i \in \mathbb{F}$$

Proof. ($\Longrightarrow$): Suppose not. Then $\exists\ \xi, \xi'$ such that
$$x = \xi_1 v_1 + \xi_2 v_2 + \cdots \xi_n v_n \tag{1}$$
and
$$x = \xi_1' v_1 + \xi_2' v_2 + \cdots \xi_n' v_n \tag{2}$$
Subtracting (2) from (1) gives
$$0 = (\xi_1 - \xi_1')v_1 + (\xi_2 - \xi_2')v_2 + \cdots (\xi_n - \xi_n')v_n$$
which, for $\xi_i \neq \xi_i'$ implies that the $\{v_i\}_{i=1}^n$ are linearly dependent, contradicting the assumption that $\{v_i\}_{i=1}^n$ is a basis.

($\Longrightarrow$): Suppose that every $x \in V$ can be written uniquely as in (1). Then, clearly $\{v_i\}_{i=1}^n$ spans V. Now, we need to argue that that $\{v_i\}_{i=1}^n$ is a linearly independent set. Again since *any* $x \in V$ can be expressed uniquely in the form (1) for some ξ_i, then it is certainly the case for $0 \in V$ so that
$$0 = \xi_1 v_1 + \cdots + \xi_n v_n,\ \xi_i \in \mathbb{F}$$
Since this representation is unique[1], it has to be the case that $\xi_i = 0$ for each $i \in \{1, \ldots, n\}$. Thus $\{v_i\}_{i=1}^n$ is linearly independent and hence a basis for V.

□

Fact 1. If $\{v_1, v_2, \ldots, v_n\}$ is a basis for $(V, \mathbb{F})$, then any other basis also has n elements. The number of elements in the basis is called the **dimension** of the vector space.

Fact 2. A basis of a vector space is *NOT* unique.

Can you come up with an example in $\mathbb{R}^3$? The following are both bases for $\mathbb{R}^3$:
$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

[1] Suppose its unique, but v_i are not linearly independent. Then, there exists some $\xi_i \neq 0$ such that $\sum_j \xi_j v_j = 0$. WLOG, suppose v_1 can be written as a linear combination of the other vectors:
$$\begin{aligned} \xi_1(\alpha_2 v_2 + \cdots \alpha_n v_n) + \xi_2 v_2 + \cdots + \xi_n v_n &= 0 \\ 0 \cdot v_1 + (\xi_1 \alpha_2 + \xi_2)v_2 + \cdots + (\xi_1 \alpha_n + \xi_n)v_n &= 0 \end{aligned}$$
which implies that we have another representation of 0 in terms of the basis vectors and hence, this contradicts uniqueness.

DIY Exercise. what is a basis for $\mathbb{R}^{2\times 2}$?
DIY Exercise. the linear space of polynomials with real coefficients defined over the field of reals, denoted $(\mathbb{R}[s], \mathbb{R})$ is an example of an infinite dimensional vector space—i.e.,

$$B = \{1, s, s^2, s^3, \ldots, s^k, \ldots\}$$

show this infinite set of vectors is linearly independent over $\mathbb{R}$.

6.1 Simple Examples

Let's do some simple examples.

Sometimes the coefficients of a vector with respect to a basis are call *coordinates*. That is, coordinates of a vector

$$v = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{R}^n$$

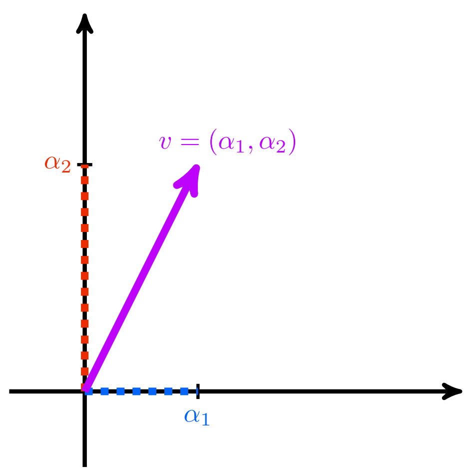
relative to the standard basis

$$e_1 = (1, 0, \ldots, 0, 0),\ e_2 = (0, 1, \ldots, 0, 0),\ \ldots,\ e_n = (0, 0, \ldots, 0, 1)$$

are

$$(\alpha_1, \alpha_2, \ldots, \alpha_n)$$

e.g., in $\mathbb{R}^2$, the coefficients w.r.t. the standard basis literally tell you the coordinates



Key: Just like writing v wrt the standard basis, we can represent a vector v in any of basis by finding its coordinates via a transformation.

Example. Consider vectors $u_1 = (2, 1)$ and $u_2 = (3, 1)$. These vectors form a basis for $\mathbb{R}^2$.

1. Can you show that these vectors form a basis?
   soln. you need to check that they are linearly independent. That means the only α's such that

$$\alpha_1 u_1 + \alpha_2 u_2 = \alpha_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 0$$

   are zero.

$$2\alpha_1 = -3\alpha_2,\quad \alpha_1 = -\alpha_2 \iff 2(-\alpha_2) = -3(\alpha_2) \iff (2-3)\alpha_2 = -\alpha_2 = 0 \iff \alpha_1 = \alpha_2 = 0$$

2. Find coordinates of the vector $v = (7,4)$ with respect to the basis $\{u_1, u_2\}$.
   soln. i.e. we need to find α_1, α_2 such that
   $$v = \begin{bmatrix} 7 \\ 4 \end{bmatrix} = \alpha_1 u_1 + \alpha_2 u_2 = \alpha_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2\alpha_1 + 3\alpha_2 \\ \alpha_1 + \alpha_2 \end{bmatrix}$$
   hence, we need to solve the set of linear equations
   $$\begin{aligned} 7 &= 2\alpha_1 + 3\alpha_2 \\ 4 &= \alpha_1 + \alpha_2 \end{aligned}$$
   from the second, $\alpha_1 = 4 - \alpha_2$ so that
   $$7 = 2(4 - \alpha_2) + 3\alpha_2 \iff -1 = \alpha_2 \iff \alpha_2 = -1,\ \alpha_1 = 5$$

3. Find the vector w whose coordinates with respect to the basis u_1, u_2 are $(7,4)$.
   sol. That is, find w_1, w_2 such that
   $$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 7u_1 + 4u_2 = 7\begin{bmatrix} 2 \\ 1 \end{bmatrix} + 4\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 14 + 12 \\ 7 + 4 \end{bmatrix} = \begin{bmatrix} 26 \\ 11 \end{bmatrix}$$

4. (**change of coordinates**). Given a vector $v \in \mathbb{R}^2$, let (x, y) be its standard coordinates—i.e., coordinates with respect to the standard basis $e_1 = (1,0)$, $e_2 = (0,1)$, and let (α, β) be its coordinates with respect to the basis $u_1 = (2,1)$, $u_2 = (3,1)$. Find a relation between (x, y) and (α, β).
   sol. By definition,
   $$v = x\begin{bmatrix} 1 \\ 0 \end{bmatrix} + y\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad v = \alpha\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \beta\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$
   so that
   $$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \iff \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}^{-1}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

What is the generalization of this? Given a vector $x \in \mathbb{R}^n$ with coordinates $(\alpha_i)_{i=1}^n$ w.r.t. basis $\{v_i\}_{i=1}^n$ and coordinates $(\beta_i)_{i=1}^n$ w.r.t. basis $\{w_i\}_{i=1}^n$,

$$\underbrace{\begin{bmatrix} | & \cdots & | \\ v_1 & \cdots & v_n \\ | & \cdots & | \end{bmatrix}}_{V}\underbrace{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}}_{\alpha} = \underbrace{\begin{bmatrix} | & \cdots & | \\ w_1 & \cdots & w_n \\ | & \cdots & | \end{bmatrix}}_{W}\underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}}_{\beta}$$

so that

$$\alpha = V^{-1}W\beta$$

510: **Fall 2019**

Lecture 4: Linear Maps & Matrix Representation Revisited

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

References: Appendix A.4 of **[C&D]**; Chapter 3 of **[Ax]**

Last time we concluded with some examples which were leading up to representation of operations (transformations) on linear spaces with respect to a basis. Before we dive into this further, let us revisit the linear map concept from lecture 2.

1 Linear Maps

Recall the definition of a linear map: A **linear map** from V to W is a function $f : V \to W$ with the following properties:

additivity. $f(x + z) = f(x) + f(z)$ for all $x, z \in V$.
homogeneity. $f(ax) = af(x)$ for all $a \in \mathbb{F}$ and $x \in V$.

We denote by $\mathcal{L}(V, W)$ the set of all linear maps from V to W.

Suppose $(v_1, \ldots, v_n)$ is a basis of V and $\mathcal{A} : V \to W$ is linear. If $v \in V$, then we can write v in the form

$$v = a_1v_1 + \cdots + a_nv_n$$

The linearity of $\mathcal{A}$ implies that

$$\mathcal{A}(v) = a_1\mathcal{A}(v_1) + \cdots + a_n\mathcal{A}(v_n)$$

That is, the values of $\mathcal{A}(v_i)$, $i \in \{1, \ldots, n\}$ determine the values of $\mathcal{A}$ on arbitrary vectors in V.

Converse: Construction of Linear Map. Linear maps can be constructed that take on arbitrary values on a basis. Indeed, given a basis $(v_1, \ldots, v_n)$ of V and any choice of vectors $w_1, \ldots, w_n \in W$, we can construct a linear map $\mathcal{A} : V \to W$ such that $\mathcal{A}(v_j) = w_j$ for $j \in \{1, \ldots, n\}$. To do this, we define $\mathcal{A}$ as follows:

$$\mathcal{A}\left(\sum_{i=1}^n a_iv_i\right) = \sum_{i=1}^n a_iw_i$$

where $a_i \in \mathbb{F}$, $i = 1, \ldots, n$ are arbitrary elements of $\mathbb{F}$.

Fact. The set of all linear maps $\mathcal{L}(V, W)$ is a vector space.

2 Null and Range Spaces

There are several important spaces associated with linear maps.

Null Space. For $\mathcal{A} \in \mathcal{L}(V, W)$, we define the **null space** of $\mathcal{A}$ to be the subset of V consisting of those vectors that map to zero under $\mathcal{A}$—i.e.,

$$\ker(\mathcal{A}) = \mathcal{N}(\mathcal{A}) = \{v \in V |\ \mathcal{A}(v) = 0\}$$

Examples.

1. Consider the map $\mathcal{A} \in \mathcal{L}(P(\mathbb{R}), P(\mathbb{R}))$, where $P(\mathbb{R})$ is the set of polynomials over the field $\mathbb{R}$, defined by
$$(\mathcal{A}p)(x) = x^2 p(x),\ \ p(x) \in P(\mathbb{R})$$
Hence,
$$\ker(\mathcal{A}) = \{0\}$$
The reason for this is that the only polynomial $p(x)$ such that $p(x)x^2 = 0$ for all $x \in \mathbb{R}$ is the zero polynomial.
2. Recall that derivatives are linear maps. Indeed, consider $\mathcal{A} \in \mathcal{L}(P(\mathbb{R}), P(\mathbb{R}))$ defined by
$$\mathcal{A}(p) = Dp$$
The only functions whose derivative equals the zero function are the constant functions, so in this case the null space of $\mathcal{A}$ equals the set of constant functions.

Range Space. For $\mathcal{A} \in \mathcal{L}(V, W)$ the **range** of $\mathcal{A}$ is the subset of W consisting of those vectors that are of the form $\mathcal{A}(v)$ for some $v \in V$—i.e.
$$\text{range}(\mathcal{A}) = \mathcal{R}(\mathcal{A}) = \{\mathcal{A}(v) |\ v \in V\}$$

Examples.

1. Consider the map $\mathcal{A} \in \mathcal{L}(P(\mathbb{R}), P(\mathbb{R}))$, defined by
$$(\mathcal{A}p)(x) = x^2 p(x),\ \ p(x) \in P(\mathbb{R})$$
Then the range of $\mathcal{A}$ is the set of polynomials of the form
$$a_2x^2 + \cdots + a_m x^m$$
where $a_2, \ldots, a_m \in \mathbb{R}$.
2. Consider $\mathcal{A} \in \mathcal{L}(P(\mathbb{R}), P(\mathbb{R}))$ defined by
$$\mathcal{A}(p) = Dp$$
Then
$$\text{range}(\mathcal{A}) = P(\mathbb{R})$$
since every polynomial $q \in P(\mathbb{R})$ is differentiable.

2.1 Some results on Null and Range Spaces

Proposition 1. If $\mathcal{A} \in \mathcal{L}(V, W)$, then $\mathcal{N}(\mathcal{A})$ is a subspace of V.

Proof. **Q**: What do we need to check? **A**: (1) $0 \in V$, (2) $u, v \in V \implies u + v \in V$, (3) $u \in V,\ a \in \mathbb{F} \implies au \in V$. OR (1) check $\mathcal{N}(\mathcal{A}) \subset V$ and (2) $\forall\ u, v \in V,\ \forall a \in \mathbb{F},\ av_1 + av_2 \in V$.

So, let's try the former:

(1). Suppose $\mathcal{A} \in \mathcal{L}(V, W)$. By additivity,
$$\mathcal{A}(0) = \mathcal{A}(0 + 0) = \mathcal{A}(0) + \mathcal{A}(0) \implies \mathcal{A}(0) = 0 \implies 0 \in \mathcal{N}(\mathcal{A})$$

(2). Suppose $u, v \in \mathcal{N}(\mathcal{A})$, then
$$\mathcal{A}(u + v) = \mathcal{A}(u) + \mathcal{A}(v) = 0 + 0 = 0 \implies u + v \in \mathcal{N}(\mathcal{A})$$

(3). Suppose $u \in \mathcal{N}(\mathcal{A})$, $a \in \mathbb{F}$, then

$$\mathcal{A}(au) = a\mathcal{A}(u) = a \cdot 0 = 0 \implies au \in \mathcal{N}(\mathcal{A})$$

$\square$

Definition 1. A linear map $\mathcal{A} : V \to W$ is **injective** if for any $u, v \in V$

$$\mathcal{A}(u) = \mathcal{A}(v) \quad \implies \quad u = v$$

Because of the next proposition, only Example 1—i.e. $\mathcal{A}(p) = x^2 p(x)$—above is injective.

Proposition 2. Suppose $\mathcal{A} \in \mathcal{L}(V, W)$.

$$\mathcal{A} \text{ injective} \iff \mathcal{N}(\mathcal{A}) = \{0\}$$

DIY Exercise.

Proof. ($\Longrightarrow$) Suppose that $\mathcal{A}$ is injective.
WTS. $\mathcal{N}(\mathcal{A}) = \{0\}$.

$$\text{Prop 1} \quad \implies \quad \{0\} \subset \mathcal{N}(\mathcal{A})$$

It is sufficient to show that $\mathcal{N}(\mathcal{A}) \subset \{0\}$:
Suppose $v \in \mathcal{N}(\mathcal{A})$.

$$[\mathcal{A}(v) = 0 = \mathcal{A}(0)] \text{ and } \mathcal{A} \text{ injective} \quad \implies \quad v = 0$$

which shows the claim.

($\Longleftarrow$). Suppose that $\mathcal{N}(\mathcal{A}) = \{0\}$.
WTS. $\mathcal{A}$ is injective.
Suppose $u, v \in V$ and $\mathcal{A}(u) = \mathcal{A}(v)$. Then,

$$0 = \mathcal{A}(u) - \mathcal{A}(v) = \mathcal{A}(u - v) \implies u - v \in \mathcal{N}(\mathcal{A}) = \{0\} \implies u - v = 0 \implies u = v$$

which shows the claim. $\square$

Proposition 3. For any $\mathcal{A} \in \mathcal{L}(V, W)$, $\mathcal{R}(\mathcal{A})$ is a subspace of W.

DIY Exercise. proof.

Definition 2. A linear map $\mathcal{A} : V \to W$ is **surjective** if its range is all of the co-domain W.

Recall Example 2—i.e., $\mathcal{A}(p) = Dp$. This map is surjective since $\mathcal{R}(\mathcal{A}) = P(\mathbb{R})$.

Combining the above propositions and definitions leads us to a key result in linear algebra which is worth internalizing.

Theorem 1. If V is finite dimensional and $\mathcal{A} \in \mathcal{L}(V, W)$, then $\mathcal{R}(\mathcal{A})$ is a finite-dimensional subspace of W and

$$\dim(V) = \dim(\mathcal{N}(\mathcal{A})) + \dim(\mathcal{R}(\mathcal{A}))$$

Lemma 1. Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of the vector space.

Proof. Suppose V is a finite-dimensional vector space of dimension n. Let $\mathcal{A} \in \mathcal{L}(V, W)$ and let $(u_1, \dots, u_m)$ be a basis of $\mathcal{N}(\mathcal{A})$ so that $\dim(\mathcal{N}(\mathcal{A})) = m$.

We can extend $(u_1, \dots, u_m)$ to a basis of V—that is, we can find $n - m$ vectors $u_{m+1}, \dots, u_n$ such that

$$(u_1, \dots, u_m, u_{m+1}, \dots, u_n)$$

is a basis for V.

<u>WTS.</u> $\mathcal{R}(\mathcal{A})$ is finite-dimensional with dimension $n - m$.

Claim. $(\mathcal{A}(u_{m+1}), \dots, \mathcal{A}(u_n))$ is a basis for $\mathcal{R}(\mathcal{A})$.

Need to show: (1) the set spans $\mathcal{R}(\mathcal{A})$ and (2) it is linearly independent.

Let $v \in V$. Since $\text{span}(u_1, \dots, u_m, u_{m+1}, \dots, u_n) = V$,

$$v = \sum_{i=1}^{n} a_i u_i,\ a_i \in \mathbb{F},\ \forall i$$

so that

$$\mathcal{A}(v) = \sum_{i=1}^{n} a_i \mathcal{A}(u_i) = \sum_{i=m+1}^{n} a_i \mathcal{A}(u_i) \tag{1}$$

since $u_1, \dots, u_m \in \mathcal{N}(\mathcal{A})$. Now,

$$(1) \implies \text{span}(\mathcal{A}(u_j))_{j=m+1}^{n} = \mathcal{R}(\mathcal{A}).$$

Let's show (2). Suppose that $c_{m+1}, \dots, c_n \in \mathbb{F}$ and

$$c_{m+1}\mathcal{A}(u_{m+1}) + \cdots + c_n \mathcal{A}(u_n) = 0 \implies \mathcal{A}(c_{m+1}u_{m+1} + \cdots + c_n u_n) = 0 \implies \sum_{j=m+1}^{n} c_j u_j \in \mathcal{N}(\mathcal{A})$$

Now, since $\text{span}(u_1, \dots, u_m) = \mathcal{N}(\mathcal{A})$ and $(u_j)_{j=1}^{n}$ are linearly independent,

$$\sum_{j=m+1}^{n} c_j u_j = \sum_{i=1}^{m} d_i u_i,\ d_i \in \mathbb{F} \implies c_j, d_i = 0\ \forall i, j$$

This shows the claim and also that $\dim(\mathcal{R}(\mathcal{A})) = n - m$.

□

There are two corollaries to this theorem which are important and can be used as sanity checks.

Corollary 1. If V and W are finite-dimensional vector spaces such that $\dim(V) > \dim(W)$, then <u>no</u> linear map from V to W is injective.

Proof.

$$\dim(V) > \dim(W) \implies \dim(\mathcal{N}(\mathcal{A})) = \dim(V) - \dim(\mathcal{R}(\mathcal{A})) \geq \dim(V) - \dim(W) > 0$$

□

Corollary 2. If V and W are finite-dimensional vector spaces such that $\dim(V) < \dim(W)$, then <u>no</u> linear map from V to W is surjective.

Proof.

$$\dim(V) < \dim(W) \implies \dim(\mathcal{R}(\mathcal{A})) = \dim(V) - \dim(\mathcal{N}(\mathcal{A})) \leq \dim(V) < \dim(W)$$

□

Implications for Systems of Linear Equations. Consider

$$a_{ji} \in \mathbb{F},\ j \in \{1, \dots, m\},\ i \in \{1, \dots, n\}$$

Define $\mathcal{A} : \mathbb{F}^n \to \mathbb{F}^m$ by

$$\mathcal{A}(x_1, \dots, x_n) = \left(\sum_{i=1}^{n} a_{1i} x_i, \dots, \sum_{i=1}^{n} a_{mi} x_i \right)$$

Consider $\mathcal{A}(x) = 0$ so that

$$\sum_{i=1}^{n} a_{ji} x_i = 0,\ j \in \{1, \dots, m\}$$

Suppose we know the a_{ji}'s and want to find x's that satisfy these equations. There are clearly m equations and n unknowns. Trivially,

$$x_1 = \cdots = x_n = 0$$

is a solution. Are there others? i.e.,

$$\dim(\mathcal{N}(\mathcal{A})) > \dim(\{0\}) = 0$$

This happens exactly when $\mathcal{A}$ is not injective which is equivalent to $n > m$.

Take away: a homogeneous system of linear equations in which there are more variables than equations must have nonzero solutions.

Examining $\mathcal{A}(x) = c$ for some $c \in \mathbb{F}^m$, we can draw an analogous conclusion via similar reasoning.

Take away: an inhomogeneous system of linear equations in which there are more equations than variables has no solution for some choice of the constant terms.

3 More on the Matrix Representation Theorem

Key Idea: Any linear map between finite dimensional linear spaces can be represented as a matrix multiplication.

$$\begin{tikzcd}
x \in U \arrow[r, "\mathcal{A}"] \arrow[d, "Q"'] & y = \mathcal{A}(x) \in V \arrow[d, "P"] \\
Qx \arrow[r, "A"] & AQx = Py
\end{tikzcd}$$

Let $\mathcal{A} : U \to V$ be a linear map from (U, F) to (V, F) where $\dim(U) = n$ and $\dim(V) = m$. Let $\mathcal{U} = \{u_j\}_{j=1}^{n}$ be a basis for U and let $\mathcal{V} = \{v_j\}_{j=1}^{m}$ be a basis for V.

We will get a matrix representation of A wrt to $\mathcal{U}$ and $\mathcal{V}$ in a few steps:

(step 1) represent x in terms of a basis for U;

(step 2) use linearity to get the map applied to each basis element for U

(step 3) use the uniqueness of a representation wrt a basis to get a representation in terms of basis vectors for V for each map of the basis vectors of U

(step 4) compile these representations to get a matrix

Now in detail...

(step 1) We showed last time that given a basis, the coordinate representation of each x is unique—that is, for any $x \in U$, $\exists!\ \xi = (\xi_1, \dots, \xi_n) \in F^n$ such that

$$x = \sum_{j=1}^{n} \xi_j u_j$$

where $\xi \in F^n$ is the component vector.

(step 2) By linearity,

$$\mathcal{A}(x) = \mathcal{A}\left(\sum_{j=1}^{n} \xi_j u_j\right) = \sum_{j=1}^{n} \xi_j \mathcal{A}(u_j)$$

(step 3) Now each $\mathcal{A}(u_j) \in V$, thus each $\mathcal{A}(u_j)$ has a *unique* representation in terms of the $\{v_j\}_{j=1}^m$ such that

$$\mathcal{A}(u_j) = \sum_{i=1}^{m} a_{ij} v_i \quad \forall\ j \in \{1, \dots, n\}$$

i.e.

$$\mathcal{A}(u_1) = \sum_{i=1}^{m} a_{i,1} v_i, \quad \mathcal{A}(u_2) = \sum_{i=1}^{m} a_{i,2} v_i, \quad \dots$$

remark: $(a_{ij})_i$ is the j–th column of A

Thus $\{a_{i,j}\}_{i=1}^m$ is the representation of $\mathcal{A}(u_j)$ in terms of $\{v_1, v_2, \dots, v_m\}$. In fact, each $\mathcal{A}(u_j)$ forms a column of the matrix representation

$$\mathcal{A}(u_j) \longleftrightarrow \begin{bmatrix} a_{1,j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

Therefore,

$$\mathcal{A}(x) = \sum_{j=1}^{n} \xi_j \mathcal{A}(u_j) = \sum_{j=1}^{n} \xi_j \sum_{i=1}^{m} a_{ij} v_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} \xi_j\right) v_i = \sum_{i=1}^{m} \eta_i v_i$$

Thus, the representation of $\mathcal{A}(x)$ with respect to $\{v_1, v_2, \dots, v_m\}$ is

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_m \end{bmatrix} \in F^m$$

where

$$\eta_i = \sum_{j=1}^{n} a_{ij}\xi_j = \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} \quad \text{for each } i \in \{1, \dots, m\}$$

so that

$$\eta = A\xi, \ a \in F^{m\times n}$$

(step 4) recall that the unique representation of x with respect to $\{u_1, u_2, \dots, u_n\}$ is

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} \in F^n$$

And, we just argued that η was the unique representation of

$$(\mathcal{A}(u_1), \dots, \mathcal{A}(u_n))$$

wrt $\mathcal{V}$. Hence, A is the matrix representation of the linear operator $\mathcal{A}$ from U to V.

In most applications we replace vectors and linear maps by their representative, viz. component vectors and matrices, e.g., we write $\mathcal{N}(A)$ instead of $\mathcal{N}(\mathcal{A})$.

Theorem 2 (Matrix Representation)**.** Let (U, F) have basis $\{u_j\}_{j=1}^n$ and let (V, F) have basis $\{v_i\}_{i=1}^m$. Let $\mathcal{A} : U \to V$ be a linear map. Then, w.r.t. these bases, the linear map A is represented by the $m \times n$ matrix

$$A = (a_{ij})_{i=1,j=1}^{m,n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \dots & \ddots & \dots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in F^{m\times n}$$

where the j–th column of A is the component vector of $\mathcal{A}(u_j)$ w.r.t. the basis $\{v_i\}_{i=1}^m$.

Example. (Observable Canonical Form). Let $\mathcal{A} : (\mathbb{R}^n, \mathbb{R}) \to (\mathbb{R}^n, \mathbb{R})$ be a linear map and suppose that

$$\mathcal{A}^n = -\alpha_1\mathcal{A}^{n-1} - \alpha_2\mathcal{A}^{n-2} - \cdots - \alpha_{n-1}\mathcal{A} - \alpha_n I, \quad \alpha_i \in \mathbb{R}$$

where I is the identity map from $\mathbb{R}^n$ to itself. Let $b \in \mathbb{R}^n$. Suppose

$$(b, \mathcal{A}(b), \dots, \mathcal{A}^{n-1}(b))$$

is a basis for $\mathbb{R}^n$. Show that, w.r.t. this basis, the vector b and the linear map $\mathcal{A}$ are represented by

$$\bar{b} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\alpha_n \\ 1 & 0 & \cdots & 0 & -\alpha_{n-1} \\ 0 & 1 & \cdots & 0 & -\alpha_{n-2} \\ \vdots & & \vdots & \dots & \vdots \\ 0 & 0 & \cdots & 1 & -\alpha_1 \end{bmatrix}$$

Proof.

$$\bar{b} = 1 \cdot b + 0 \cdot \mathcal{A}(b) + \cdots + 0 \cdot \mathcal{A}^{n-1}(b)) = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{(1st column)}$$

$$\mathcal{A}(b) = 0 \cdot b + 1 \cdot \mathcal{A}(b) + 0 \cdot \mathcal{A}^2(b) + \cdots + 0 \cdot \mathcal{A}^{n-1}(b) \ \text{(2nd column)}$$

$$\mathcal{A}(\mathcal{A}^{n-1}(b)) = \mathcal{A}^n(b) = -\alpha_n b - \alpha_{n-1}\mathcal{A}(b) + \cdots - \alpha_1 \mathcal{A}^{n-1}(b) \text{ (n-th column)}$$

hence

$$A = \begin{bmatrix} 0 & 0 & \cdots \\ 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \cdots \\ 0 & 0 & \cdots \end{bmatrix}$$

as expected. □

Remark. We will see later that this is the observable canonical form.

Fact. The composition of two linear operators is just matrix multiplication.

4 Change of Basis for Matrix Representation

We can apply this result to institute a change of basis.

Setup:

- Let $\mathcal{A} : U \to V$, where U, V are vector spaces over the same field F, be a linear map with $\dim U = m$ and $\dim V = n$.
- Let U with elements x have bases $\{u_j\}_{j=1}^m$ and $\{\bar{u}_j\}_{j=1}^m$ generating component vectors ξ and $\bar{\xi}$, resp., in F^m.
- Let V with elements y have bases $\{v_j\}_{j=1}^n$ and $\{\bar{v}_j\}_{j=1}^n$ generating component vectors η and $\bar{\eta}$, resp., in F^n.
- Let A be the matrix representation of $\mathcal{A} : U \to V$ w.r.t. the bases $\{u_j\}_{j=1}^m$ and $\{v_j\}_{j=1}^n$.
- Let $\bar{A}$ be the matrix representation of $\mathcal{A} : U \to V$ w.r.t. the bases $\{\bar{u}_j\}_{j=1}^m$ and $\{\bar{v}_j\}_{j=1}^n$.

Note. A and $\bar{A}$ are said to be equivalent, and $\bar{A} = QAP$ is said to be a similarity transform.

Moreover, if $U = V$ and $\{\bar{v}_j\}_{j=1}^n = \{\bar{u}_j\}_{j=1}^n$, then

$$\bar{A} = P^{-1}AP$$

Then, by the above proposition, $x \in U$ has generating component vectors ξ and $\bar{\xi}$ for bases $\{u_j\}_{j=1}^m$ and $\{\bar{u}_j\}_{j=1}^m$, resp. Moreover,

$$\xi = P\bar{\xi}$$

with non-singular $P \in F^{m\times m}$ given by

$$P = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}^{-1} \begin{bmatrix} \bar{u}_1 & \cdots & \bar{u}_m \end{bmatrix}$$

Similarly, $y \in V$ has generating component vectors η and $\bar{\eta}$ for bases $\{v_i\}_{i=1}^n$ and $\{\bar{v}_i\}_{i=1}^n$, resp., and

$$\bar{\eta} = Q\eta$$

with non-singular $Q \in F^{n\times n}$ given by

$$Q = \begin{bmatrix} \bar{v}_1 & \cdots & \bar{v}_n \end{bmatrix}^{-1} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

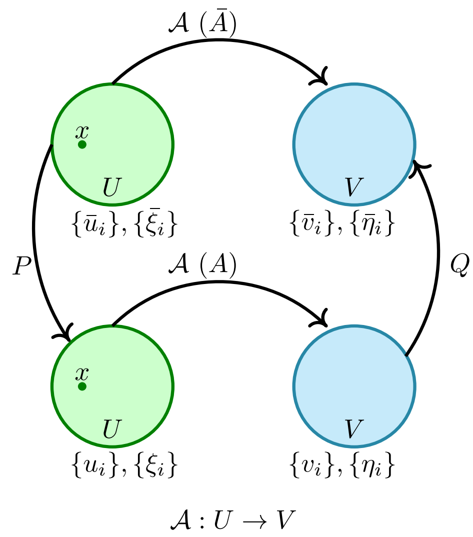


Figure 1: Change of Basis

By the matrix representation theorem (Theorem 2), we know that $\eta = A\xi$. Hence,

$$\bar{\eta} = Q\eta = QA\xi = QAP\bar{\xi} = \bar{A}\bar{\xi}$$

where $\bar{A} = QAP$.

So, if A is the matrix representation of $\mathcal{A}$ with respect to $\{u_i\}, \{v_j\}$, then $\bar{A} = QAP$ is the matrix representation of $\mathcal{A}$ with respect to $\{\bar{u}_i\}, \{\bar{v}_j\}$.

DIY Exercise. Let $\mathcal{A} : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear map. Consider

$$B = \{b_1, b_2, b_3\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

and

$$C = \{c_1, c_2, c_3\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Clearly, B and C are bases for $\mathbb{R}^3$. Suppose $\mathcal{A}$ maps vectors in B in the following way:

$$\mathcal{A}(b_1) = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \; \mathcal{A}(b_2) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \; \mathcal{A}(b_3) = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}$$

Write down the matrix representation of $\mathcal{A}$ with respect to B and then C.

510:	**Fall 2019**

Lecture 6: More on Linear Maps

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

References: Appendix A **[C&D]**; Chapter 3 & 5 of **[Ax]**

1 Change of Basis for Matrix Representation

We can apply this result to institute a change of basis.

Setup:

- Let $\mathcal{A} : U \to V$, where U, V are vector spaces over the same field F, be a linear map with $\dim U = m$ and $\dim V = n$.
- Let U with elements x have bases $\{u_j\}_{j=1}^m$ and $\{\bar{u}_j\}_{j=1}^m$ generating component vectors ξ and $\bar{\xi}$, resp., in F^m.
- Let V with elements y have bases $\{v_j\}_{j=1}^n$ and $\{\bar{v}_j\}_{j=1}^n$ generating component vectors η and $\bar{\eta}$, resp., in F^n.
- Let A be the matrix representation of $\mathcal{A} : U \to V$ w.r.t. the bases $\{u_j\}_{j=1}^m$ and $\{v_j\}_{j=1}^n$.
- Let $\bar{A}$ be the matrix representation of $\mathcal{A} : U \to V$ w.r.t. the bases $\{\bar{u}_j\}_{j=1}^m$ and $\{\bar{v}_j\}_{j=1}^n$.

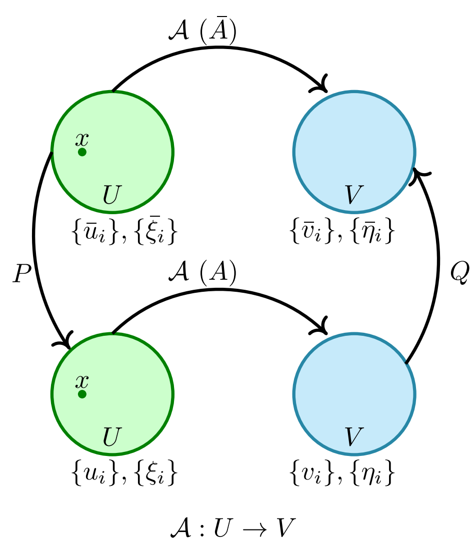


Figure 1: Change of Basis

Note. A and $\bar{A}$ are said to be equivalent, and $\bar{A} = QAP$ is said to be a similarity transform.

Moreover, if $U = V$ and $\{\bar{v}_j\}_{j=1}^n = \{\bar{u}_j\}_{j=1}^n$, then

$$\bar{A} = P^{-1}AP$$

Then, by the above proposition, $x \in U$ has generating component vectors ξ and $\bar{\xi}$ for bases $\{u_j\}_{j=1}^m$ and $\{\bar{u}_j\}_{j=1}^m$, resp. Moreover,

$$\xi = P\bar{\xi}$$

with non-singular $P \in F^{m\times m}$ given by

$$P = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}^{-1} \begin{bmatrix} \bar{u}_1 & \cdots & \bar{u}_m \end{bmatrix}$$

Similarly, $y \in V$ has generating component vectors η and $\bar{\eta}$ for bases $\{v_i\}_{i=1}^n$ and $\{\bar{v}_i\}_{i=1}^n$, resp., and

$$\bar{\eta} = Q\eta$$

with non-singular $Q \in F^{n\times n}$ given by

$$Q = \begin{bmatrix} \bar{v}_1 & \cdots & \bar{v}_n \end{bmatrix}^{-1} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

By the matrix representation theorem, we know that $\eta = A\xi$. Hence,

$$\bar{\eta} = Q\eta = QA\xi = QAP\bar{\xi} = \bar{A}\bar{\xi}$$

where $\bar{A} = QAP$.

So, if A is the matrix representation of $\mathcal{A}$ with respect to $\{u_i\}, \{v_j\}$, then $\bar{A} = QAP$ is the matrix representation of $\mathcal{A}$ with respect to $\{\bar{u}_i\}, \{\bar{v}_j\}$.

1.1 Invariant Subspaces

There are some nice connections between invariant subspaces of an operator and the existence of matrix representations with particular forms.

definition. An invariant subspace of a linear mapping $A \in \mathcal{L}(V)$ on a vector space V to itself is a subspace W of V such that $A(W) \subset W$. Sometimes we call W an A–invariant subspace.

If W is A invariant, then we can restrict A to W, giving rise to a new linear operator which we denote by

$$A|_W : W \to W$$

The invariant subspaces of a given linear transformation A expose the structure of A. In the next few weeks, we will see the following interesting fact: when V is a finite-dimensional vector space over an (algebraically) closed field[1], linear transformations acting on V are characterized (up to similarity) by the *Jordan canonical form*, which decomposes V into invariant subspaces of A.

Indeed, many fundamental questions regarding A can be translated to questions about invariant subspaces of A.

Suppose that W is an invariant subspace for a liner operator $A : V \to V$. Choose a basis $\{w_1, \dots, w_k\}$ for W. Complete it to a basis for V. Then, with respect to this basis, the matrix representation of A takes the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

where

$$A_{11} = A|_W$$

[1] An algebraically closed field is a field $\mathbb{F}$ which contains a root for every non-constant polynomial in $\mathbb{F}[s]$, the ring of polynomials in s with coefficients in $\mathbb{F}$. For example, the real numbers are not algebraically closed because $x^2 + 1 = 0$ does not contain a root in the real numbers. On the other hand, the complex numbers are algebraically closed.

This means that we can write $V = W \oplus U$ where U is spanned by the linearly independent vectors chosen to complete the basis $\{w_1, \ldots, w_k\}$ for W. Hence, if we write

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} : \begin{matrix} W \\ \oplus \\ U \end{matrix} \to \begin{matrix} W \\ \oplus \\ U \end{matrix}$$

it has to be the case that $A_{21} : W \to U$ is the zero map.

remark. Determining whether a given subspace W is invariant under A is ostensibly a problem of geometric nature. *Matrix representation* allows one to phrase this problem algebraically.

A *projection operator* P onto the A–invariant subspace W is defined by $P(w + u) = w$ with $w \in W$ and $u \in U$. Projection operators are formally defined as operators $P : V \to V$ such that $P^2 = P$; we will extend this definition to *orthogonal projections* after we introduce Hilbert spaces in the next few lectures.

The projection operator has matrix representation

$$P = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} : \begin{matrix} W \\ \oplus \\ U \end{matrix} \to \begin{matrix} W \\ \oplus \\ U \end{matrix}.$$

Example. Show that $W = \mathcal{R}(P)$ is invariant under A if and only if $PAP = AP$.

Proof. ($\Longleftarrow$): Suppose $AP = PAP$ and $x \in W$. Then

$$Ax = APx = PAPx \in W$$

The first equality is due to the fact that

$$P = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

and $x \in W$. The second equality is by assumption. And the inclusion at the end is again because of the structure of P—i.e. because P has the above structure and $AP = PAP$ it means that $APx \in W$.
($\Longrightarrow$): Suppose that W is invariant under A. Then $APx \in W$ (since $W = \mathcal{R}(P)$) and thus, $APx = PAPx$ for all $x \in V$. □

DIY Exercise. Show that if P is a projection, so is $I - P$.

Hence,

$$AP = PA \iff \mathcal{R}(P), \mathcal{R}(I - P) \text{ are } A\text{–invariant}$$

Moreover, A has matrix representation

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} : \begin{matrix} \mathcal{R}(P) \\ \oplus \\ \mathcal{R}(I - P) \end{matrix} \to \begin{matrix} \mathcal{R}(P) \\ \oplus \\ \mathcal{R}(I - P) \end{matrix}$$

Example. Show $\mathcal{N}(A)$ is A–invariant.

Proof. Indeed, consider $v \in \mathcal{N}(A)$. We claim $Av \in \mathcal{N}(A)$. Now

$$v \in \mathcal{N}(A) \implies Av = 0 \implies A(Av) = A(0) = 0 \in \mathcal{N}(A)$$

since A is linear and we know that $\mathcal{N}(A)$ is a subspace so that it contains 0. □

2 Rank and Nullity

Recall that we showed in Lec 4 that for a linear map $\mathcal{A} \in \mathcal{L}(U,V)$ with $\dim(U) = n$ and $\dim(V) = m$,

$$\dim(\mathcal{R}(\mathcal{A})) + \dim(\mathcal{N}(\mathcal{A})) = n = \dim(\underbrace{\text{Domain}(\mathcal{A})}_{U})$$

where $\mathcal{R}(\mathcal{A}) \subset V$ is a subspace and $\mathcal{N}(\mathcal{A}) \subset U$ is a subspace.

Moreover, we showed that any linear map $\mathcal{A} \in \mathcal{L}(U,V)$ on finite dimensional spaces can be represented by a matrix $A \in \mathbb{F}^{m\times n}$ where $\dim(U) = n$ and $\dim(V) = m$ as above.

Definition. The rank of a matrix $A \in \mathbb{F}^{m\times n}$, denoted by $\text{rank}(A)$, is the dimension of the range space. That is,

$$\text{rank}(A) = \dim(\mathcal{R}(A))$$

The nullity of a matrix $A \in \mathbb{F}^{m\times n}$, denoted by $\text{null}(A)$, is the dimension of the null space. That is,

$$\text{null}(A) = \dim(\mathcal{N}(A))$$

Theorem 1 (Rank-Nullity)**.** Consider a linear map $A \in \mathbb{R}^{m\times n}$. Then,

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

Proof. **DIY Exercise** □

Proof Sketch. Note that $\mathcal{N}(A) \subset F^n$ while $\mathcal{R}(A) \subset F^m$. Consider a basis for $\mathcal{N}(A)$ where $\dim \mathcal{N}(A) = k$. Complete that basis (i.e. generate $n-k$ linearly independent vectors independent of the original basis for $\mathcal{N}(A)$). Consider the representation of x w.r.t. this completed basis. To prove the Rank-Nullity theorem, you need to know that you can extend a basis of a subspace to a basis for the whole space.

Lemma 1. Let (U,F) be a finite-dimensional linear space with $\dim U = n$. Suppose $\{u_1, u_2, \dots, u_m\}$ is a set of m linearly independent vectors. There exists $n-m$ additional vectors $\{\tilde{u}_{m+1}, \dots, \tilde{u}_n\}$ such that

$$\{u_1, u_2, \dots, u_m, \tilde{u}_{m+1}, \dots, \tilde{u}_n\}$$

forms a basis for U—i.e. any set of linearly independent vectors can be extended to a basis.

Proof. If $m = n$, no extension is needed. Hence we assume $m < n$. It follows that $\{u_1, \dots, u_m\}$ is not a basis for U so there must exists a $v \in U$ such that

$$\sum_{i=1}^{m} \alpha_i u_i \neq v, \quad \forall \ \{\alpha_i\}_{i=1}^m \subset F$$

Then,

$$\{u_1, \dots, u_m, v\}$$

is linearly independent so that we can take $\tilde{u}_{m+1} = v$. The lemma follows by induction. □

Fact 1. Let $A \in \mathbb{F}^{m\times n}$. Then,

$$0 \leq \text{rank}(A) \leq \min(m,n)$$

and $\text{rank}(A)$ is equivalent to

a. maximum number of linearly independent column vectors of A
b. maximum number of linearly independent row vectors of A
c. largest integer r such that at least one minor of order r is non-zero.

DIY Exercise. Let $A \in \mathbb{F}^{n\times n}$. Show that A has an inverse if and only if $\text{rank}(A) = n$.

Recall:

$$A \in \mathbb{F}^{n\times n} \text{ invertible} \iff A \text{ non-singular (i.e. } \det(A) \neq 0)$$

Note that $\text{rank}(A) = n$ is equivalent to $\det(A) \neq 0$.

Definition. Let $A \in \mathbb{F}^{m\times n}$. The row rank of A, denoted by $\text{rowrk}(A)$, is the maximum number of linearly independent row vectors. The column rank of A, denoted by $\text{colrk}(A)$, is the maximum number of linearly independent column vectors.

By **Fact 1**,

$$\text{rank}(A) = \text{rowrk}(A) = \text{colrk}(A)$$

The column rank of A is the dimension of the column space of A, while the row rank of A is the dimension of the row space of A.

The matrix $A \in \mathbb{F}^{m\times n}$ is said to be *full-row rank* or *full-column rank* if and only if $\text{rank}(A) = m$ or $\text{rank}(A) = n$, respectively.

DIY Exercise. Let $A \in \mathbb{F}^{m\times n}$ and I_m be the $m \times m$ identity matrix. Show that

a. A has a right inverse (equiv. A surjective) iff

$$\text{rank}(A) = \text{rank}(\begin{bmatrix} A & I_m \end{bmatrix})$$

(equiv. A has *full-row rank*).

b. A has a left inverse (equiv. A injective) iff

$$\mathcal{N}(A) = \{0\}$$

(equiv. A has *full-column rank*).

Theorem 2 (Sylvester's Inequality.)**.** Let $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times p}$ be two matrices. Then, $AB \in \mathbb{F}^{m\times p}$ and

$$\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$$

DIY Exercise. Prove Thm 2. Hints: Let W be the co-domain of A and let $A|_{\mathcal{R}(B)} : \mathcal{R}(B) \to W$ be the restriction of A to $\mathcal{R}(B)$. Note that that the following are true:

$$\mathcal{R}(AB) = \mathcal{R}(A|_{\mathcal{R}(B)}) \subset \mathcal{R}(A)$$

$$\mathcal{N}(A|_{\mathcal{R}(B)}) \subset \mathcal{N}(A)$$

$$\mathcal{R}(B) = \text{domain}(A|_{\mathcal{R}(B)})$$

Applying the rank-nullity theorem to $A|_{\mathcal{R}(B)}$, we get

$$\dim(\mathcal{R}(B)) = \dim(\mathcal{R}(A|_{\mathcal{R}(B)})) + \dim(\mathcal{N}(A|_{\mathcal{R}(B)}))$$

Then, argue from here to get the result.

Combining Sylvester's inequality with the rank-nullity theorem, we get the following result.

Theorem 3 (Rank and Nullity are invariant under equivalence.)**.** Let $A \in \mathbb{F}^{m\times n}$ and $P \in \mathbb{F}^{n\times n}$, $Q \in \mathbb{F}^{m\times m}$ be nonsingular. Then,

$$\text{rank}(A) = \text{rank}(AP) = \text{rank}(QA) = \text{rank}(QAP)$$

and

$$\text{null}(A) = \text{null}(AP) = \text{null}(QA) = \text{null}(QAP)$$

This theorem is an algebraic consequence of the obvious geometric fact that the range and null space of a linear map A does not change under a change of basis in its domain or co-domain or both.

Why important? Sylvester's inequality and the rank-nullity theorem are the main tools to prove the following fact, which is very useful for our analysis since we will see that transformation, e.g., to the controllable canonical form does not change key facts about the system which allow us to analyze its properties.

Fact. Rank and nullity are invariant under equivalence (i.e. if two matrices are equivalent via similarity transformation, then they have the same rank and nullity).

3 Eigenvalues and Eigenvectors

Why do we care? Eigenvalues and eigenvectors are important for

- assessing stability; recall the example
$$\dot{x} = -\lambda x$$
whose solution is $x(t) = x_0 e^{-\lambda t}$ and if $\lambda > 0$ solution decays to zero, otherwise it blows up. λ is an 'eigenvalue' for this scalar system.
- as we will see, it is also important for assessing controllability and observability

Let $(V, \mathbb{C})$ be a vector space and $A \in \mathcal{L}(V, V)$. If there exists $\alpha \in \mathbb{C}$ and $v \in V$, $v \neq 0$, such that

$$Av = \lambda v$$

then λ is an *eigenvalue* of A and v is an *eigenvector* of A.

Note that if v is an eigenvector, then any nonzero multiple of v is an eigenvector.

The subspace $\ker(A - \lambda I) \subset V$ is called the *eigenspace* (corresponding to λ) and the *geometric multiplicity* of λ is the number $m_\lambda = \text{null}(A - \lambda I)$.

Definition. *Spectrum*. The set of all eigenvalues of A is known as the *spectrum* of A, denoted by spec(A). The *spectral radius* of A is defined as

$$\rho(A) = \max\{|\lambda| : \lambda \in \text{spec}(A)\}$$

The definition of *eigenpairs* (x, λ)—i.e. $Ax = \lambda x$—is equivalent to

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

Thus, λ is an eigenvalue of A iff the linear system $(A - \lambda I)x = 0$ has a nontrivial—i.e. $x \neq 0$—solution.

This, in turn, is equivalent to
$$\det(A - \lambda I) = 0.$$

Therefore we define the so-called characteristic polynomial of A.

Definition. *Characteristic Polynomial*.

$$\chi_A(\lambda) = \det(A - \lambda I_n) = \alpha_0 + \alpha_1 \lambda + \alpha_2 \lambda^2 + \cdots + \alpha_n \lambda^n$$

The eigenvalues of A are the roots of the *characteristic polynomial*:

$$\chi_A(\lambda) = \det(A - \lambda I_n) = \alpha_0 + \alpha_1 \lambda + \alpha_2 \lambda^2 + \cdots + \alpha_n \lambda^n$$

where I_n is the $n \times n$ matrix identity. The characteristic polynomial is an n-degree polynomial.

3.1 Examples

Ex 1. Consider $A \in \mathcal{L}(\mathbb{F}^3)$ defined by

$$A(x, y, z) = (2y, 0, 5z)$$

Find the eigenvalues and eigenvectors.

sol. suppose λ is an eigenvalue of A. taking heed of the eigenpair equation, we get

$$\begin{aligned} 2y &= \lambda x \\ 0 &= \lambda y \\ 5z &= \lambda z \end{aligned}$$

If $\lambda \neq 0$, then the second equation implies $y = 0$ so that $x = 0$. Since an eigenvalue must have a nonzero eigenvector, this means the solution is such that $z \neq 0$ which implies that $\lambda = 5$. That is, $\lambda = 5$ is the only non-zero eigenvalue of A. The set of eigenvectors corresponding to $\lambda = 5$ is

$$\{(0, 0, z) : \ x \in \mathbb{F}\}$$

If $\lambda = 0$, then the first and third equations imply that $y = 0$ and $z = 0$. With these values for y, z, the equations above are satisfied for all values of x. Thus zero is an eigenvalue of A. The corresponding eigenvectors are

$$\{(x, 0, 0) : \ x \in \mathbb{F}\}$$

Ex 2. Consider the map $A \in \mathcal{L}(\mathbb{F}^n)$, where n is a positive integer, defined by

$$A(x_1 + \cdots + x_n) = (x_1 + \cdots + x_n, \ldots, x_1 + \cdots + x_n)$$

i.e. A is an operator whose matrix (wrt standard basis) consists of all ones. What are the eigenpairs?

sol. Suppose $\lambda \in \operatorname{spec}(A)$. Then,

$$\begin{aligned} x_1 + \cdots + x_n &= \lambda x_1 \\ &\vdots \\ x_1 + \cdots + x_n &= \lambda x_n \end{aligned}$$

so that

$$\lambda x_1 = \cdots = \lambda x_n$$

Hence, either $\lambda = 0$ or $x_1 = \cdots = x_n$.

Suppose first that $\lambda = 0$. Then we get the homogeneous system of equations

$$\begin{aligned} x_1 + \cdots + x_n &= 0 \\ &\vdots \\ x_1 + \cdots + x_n &= 0 \end{aligned}$$

so that $\lambda = 0$ is an eigenvalue with corresponding eigenvectors

$$\{(x_1, \ldots, x_n) \in \mathbb{F}^n : \ x_1 + \cdots + x_n = 0\}$$

Now consider the case that $x_1 = \cdots = x_n$. Let $x_1 = \cdots = x_n = \xi$. Then, the system we get is

$$n\xi = \lambda\xi$$

so that $\lambda = n$ has to be the eigenvalue since an eigenvalue must have a non-zero eigenvector. The corresponding eigenvector is

$$\{(x_1, \ldots, x_n) \in \mathbb{F}^n | \ x_1 = \cdots = x_n\}$$

3.2 Polynomials Applied to Operators and Cayley-Hamilton

The main reason that a richer theory exists for operators (which map a vector space into itself) than for linear maps is that operators can be raised to powers.

When the domain and co-domain of a map are the same, we will denote the set of linear operators as follows:

$$\mathcal{L}(V) = \{A : V \to V \mid A \text{ is linear}\}$$

Consider $A \in \mathcal{L}(V)$. Then $A \circ A = A^2 \in \mathcal{L}(V)$. More generally, for $m \in \mathbb{Z}_+$, then

$$A^m = \underbrace{A \circ \cdots \circ A}_{m \text{ times}}$$

Consider $A \in \mathcal{L}(V)$ and $p \in P(\mathbb{F})$ such that

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m, \; z \in \mathbb{F}$$

Then, $p(A)$ is the operator defined by

$$p(A) = a_0 I + a_1 A + a_2 A^2 + \cdots + a_m A^m$$

DIY Exercise. Verify that $p \mapsto p(A)$ is linear.

Given $\lambda_1, \ldots, \lambda_p$, the coefficients α_i are determined by solving the above equation.

Suppose we define

$$\chi_A(X) = \alpha_0 I + \alpha_1 X + \alpha_2 X^2 + \cdots + \alpha_n X^n$$

Theorem 4. Cayley-Hamilton. For any square matrix A with elements in a field F,

$$\chi_A(A) = 0$$

i.e.

$$0 = \alpha_0 I + \alpha_1 A + \alpha_2 A^2 + \cdots + \alpha_n A^n$$

Fact. λ is an eigenvalue of A iff $\chi_A(\lambda) = 0$.

Example. It is well known that even real matrices can have complex eigenvalues. For instance,

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

has a characteristic polynomial

$$\chi_A(z) = \det\left(\begin{bmatrix} z & -1 \\ 1 & z \end{bmatrix}\right) = z^2 + 1$$

so that its eigenvalues are $\lambda_{1,2} = \pm i$ with associated eigenvalues

$$x_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}, \; x_2 = \begin{bmatrix} i \\ 1 \end{bmatrix}$$

3.3 Multiplicities

The characteristic polynomial also leads to the following definition.

Definition. *Algebraic Multiplicity*. Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the distinct eigenvalues of A. Then

$$\det(A - \lambda I_n) = (\lambda_1 - \lambda)^{m_1}(\lambda_2 - \lambda)^{m_2} \cdots (\lambda_p - \lambda)^{m_p}$$

where $\sum_{i=1}^{p} m_i = n$ and m_i is the *algebraic multiplicity* of λ_i.

This result can also be seen from the fact that for a matrix A,

$$\det(A) = \prod_{i=1}^{n} \lambda_i$$

Definition. *Geometric Multiplicity*. Let (v, λ) be an eigenvalue/eigenvector pair for $A \in \mathbb{C}^{n\times n}$. Then

$$\text{null}(A - \lambda I)$$

is the geometric multiplicity of λ.

In particular, the null space of $A - \lambda I$ (or the eigenspace of A for λ) is the space of all eigenvectors of A for λ and the zero vector. Its dimension is the *geometric multiplicity* of λ.
Note: It is called *geometric* because it refers to the dimension of a particular space. On the other hand, the algebraic multiplicity of λ is the number of times λ is a root of the characteristic polynomial for A— i.e. $\det(A - \lambda I) = 0$.

3.4 Examples: Multiplicities and Eigenpairs

Example.[of Multiplicities] Consider

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

so that $\chi_A(z) = (z - 1)^3$. Thus, $\lambda = 1$ is an eigenvalue (in fact, the only one) of A with algebraic multiplicity 3. To determine its geometric multiplicity we need to find the associated eigenvectors. Thus we solve $(A - \lambda I)x = 0$ for the special case of $\lambda = 1$. That is

$$(A - I) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} x = 0$$

Hence $x_1 = -x_3$ from the second row or

$$x = \begin{bmatrix} \alpha \\ \beta \\ -\alpha \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

and thus the geometric multiplicity of $\lambda = 1$ is only 2 since all vectors in the null space can be written in terms of the two basis vectors given.

Example. Consider $A \in \mathcal{L}(\mathbb{F}^2)$ defined such that

$$A(w, v) = (-v, w)$$

If $\mathbb{F} = \mathbb{R}$, then this operator has a nice geometric interpretation: A is just a counterclockwise rotation by $90°$ about the origin in $\mathbb{R}^2$.

An operator has an eigenvalue if and only if there exists a nonzero vector in its domain that gets sent by the operator to a scalar multiple of itself.

The rotation of a nonzero vector in $\mathbb{R}^2$ obviously never equals a scalar multiple of itself. Hence, if $\mathbb{F} = \mathbb{R}$, the operator A has no eigenvalues. However, this is no longer the case if $\mathbb{F} = \mathbb{C}$. To find eigenvalues of A, we must find the scalars λ such that

$$A(w, v) = \lambda(w, v)$$

has some solution other than $w = v = 0$. For A defined above, this is equivalent to

$$\begin{aligned} -v &= \lambda w \\ w &= \lambda v \end{aligned}$$

Substituting the value for w given by the second equation into the first gives

$$-v = \lambda^2 v \implies -1 = \lambda^2$$

since $v \neq 0$ (otherwise this would imply that $w = 0$ and not lead to an eigenpair). The solutions are thus

$$\lambda = \pm i$$

so that

$$\{v \in \mathbb{C}^2 |\ v = (w, \pm iw),\ w \in \mathbb{C}\}$$

are the eigenvectors.

3.5 Eigenbasis

Given the notion of 'null space', we can define an eigenbasis for a linear map A.

Definition. *Eigenbasis*. An eigenbasis corresponding to A is a basis for V consisting entirely of eigenvectors for A.

To find an eigenbasis, you find a basis for each eigenspace of A where the null space of $A - \lambda I$ is called the eigenspace associated with λ. That is, the vector space E_λ of all eigenvectors corresponding to λ:

$$E_\lambda = \text{span}\{x |\ Ax = \lambda x,\ \lambda \in \mathbb{C}\}$$

The collection of all these basis vectors for an eigenbasis for A.

Theorem 5. Let V be a vector space and let $A \in \mathcal{L}(V, V)$. Let $\{\lambda_1, \ldots, \lambda_k\}$ be a set of distinct eigenvalues of A, and for each $1 \leq j \leq k$ let x_j be an eigenvector corresponding to λ_j. Then the set $\{x_1, \ldots, x_k\}$ is linearly independent.

Corollary 1. Each operator on V has at most $\dim(V)$ distinct eigenvalues.

Fact. Two similar matrices A and $\bar{A} = P^{-1}AP$ have the same spectrum (i.e. same eigenvalues).

This is important because we can transform our system and preserve key properties like controllability, observability, stability.

Fact. Similar matrices have the same characteristic polynomial, eigenvalues, algebraic and geometric multiplicities. The eigenvectors, however, are in general different.

Theorem 6. A matrix $A \in \mathbb{C}^{m \times m}$ is *nondefective*—i.e., has no defective eigenvalues meaning the geometric and algebraic multiplicites are equal—if and only if A is similar to a diagonal matrix, i.e.,

$$A = X \Lambda X^{-1}$$

where

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}$$

is the matrix formed with the eigenvectors of A as its columns, and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m)$.

We will talk more about this when we discuss Jordan form, but essentially there is a *full set* of eigenvectors.

Consequences of Distinct Eigenvalues. An operator $A \in \mathcal{L}(V)$ has a diagonal matrix

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

with respect to a basis $(v_1, \ldots, v_n)$ of V if and only if

$$\begin{aligned} Av_1 &= \lambda_1 v_1 \\ &\cdots \\ Av_n &= \lambda_n v_n \end{aligned}$$

An operator $A \in \mathcal{L}(V)$ has a diagonal matrix with respect to some basis of V if and only if V has a basis consisting of eigenvectors of A.

Example. *Q*. Does $A \in \mathcal{L}(\mathbb{C}^2)$ defined by

$$A(w, v) = (v, 0)$$

have a diagonal matrix with respect to some basis?

A. No. The only eigenvalue is 0 and the set of eigenvectors is the 1d subspace

$$\{(w, 0) \in \mathbb{C}^2 | \; w \in \mathbb{C}\}$$

so that there are not enough linearly independent eigenvectors of A to form a basis of the 2d space $\mathbb{C}^2$.

Proposition 1. Suppose $A \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of A. Then the following are equivalent:

a. A has a diagonal matrix with respect to some basis of V.
b. V has a basis consisting of eigenvectors of A
c. there exists 1d subspaces $U_1, \ldots, U_n$ of V, each invariant under A, such that

$$V = U_1 \oplus \cdots \oplus U_n$$

d. $V = \mathcal{N}(A - \lambda_1 I) \oplus \cdots \oplus \mathcal{N}(A - \lambda_m I)$
e. $\dim(V) = \dim(\mathcal{N}(A - \lambda_1 I)) + \cdots + \dim(\mathcal{N}(A - \lambda_m I))$

Fact. If matrix $A \in \mathbb{R}^{n \times n}$ (or $\in \mathbb{C}^{n \times n}$) has m distinct eigenvalues ($\lambda_i \neq \lambda_j$, $i \neq j$) then it has (at least) m linearly independent eigenvectors.

Proposition 2. If all eigenvalues of A are distinct then A is diagonalizable.

Definition. *Diagonalizable.* An $n \times n$ matrix A is diagonalizable iff the sum of the dimensions of its eigenspaces is n—aka there exists a matrix P such that

$$A = P \Lambda P^{-1}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$,

$$P = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

with

$$Av_i = \lambda_i v_i$$

(i.e. col vectors of P are right eigenvectors of A)

Proof. Proof of Prop. 2 (By contradiction) Assume λ_i, $i \in \{1, \ldots, m\}$ are distinct and v_i, $i = 1, \ldots, m$ are linearly dependent. That is, there exists α_i such that

$$\sum_{i=1}^{m} \alpha_i v_i = 0$$

where all α_i are not zero. We can assume w.l.o.g that $\alpha_1 \neq 0$. Multiplying on the left by $(\lambda_m I - A)$,

$$0 = (\lambda_m I - A) \sum_{i=1}^{m} \alpha_i v_i = (\lambda_m I - A) \sum_{i=1}^{m-1} \alpha_i v_i + \alpha_m (\lambda_m I - A) v_m = \sum_{i=1}^{m-1} \alpha_i (\lambda_m - \lambda_i) v_i$$

since $Av_i = \lambda_i v_i$. Then multiply by $(\lambda_{m-1} I - A)$ to get that

$$0 = (\lambda_{m-1} I - A) \sum_{i=1}^{m-1} \alpha_i (\lambda_m - \lambda_i) v_i = \sum_{i=1}^{m-2} \alpha_i (\lambda_{m-1} - \lambda_i)(\lambda_m - \lambda_i) v_i = 0$$

Repeatedly multiply by $(\lambda_{m-k} I - A)$, $k = 2, \ldots, m-2$ to obtain

$$\alpha \prod_{i=2}^{m} (\lambda_i - \lambda_1) v_i = 0$$

As $\lambda_1 \neq \lambda_i$, $i = 2, \ldots, m$, the above implies that $\alpha_1 = 0$ which is a contradiction. □

3.6 Connections with Invariant Subspaces

The notion of a subspace that gets mapped into itself is sufficiently important to deserve a name.

Definition. For $A \in \mathcal{L}(V)$ and $U \subset V$ a subspace, we say U is invariant under A if $u \in U$ implies $Au \in U$.

That is, U is invariant under A if $A|_U$ is an operator on U.

Example. If A is the differentiation operator on $P_7(\mathbb{R})$ (i.e., polynomials of dimension seven over $\mathbb{R}$), then $P_4(\mathbb{R})$ is invariant under A because the derivative of any polynomial o f degree at most four is also a polynomial with degree at most four.

Example. Consider $A \in \mathcal{L}(V)$.

- $\{0\}$ is clearly invariant under A.
- V is clearly invariant under A
- $\mathcal{N}(A)$ is invariant under A:

$$u \in \mathcal{N}(A) \implies Au = 0 \implies A(Au) = 0 \implies Au \in \mathcal{N}(A)$$

We can describe a subspace of dimension one by the following: for $v \in V$,

$$U = \{au |\ a \in \mathbb{F}\}$$

is a subspace of V of dimension one. Indeed, all 1d subspaces are of this form. To see this, consider the following line of reasoning. If $A \in \mathcal{L}(V)$ and U is invariant under A, then $Au \in U$ and hence, there exists a scalar $\lambda \in \mathbb{F}$ such that $Au = \lambda u$. Conversely, if $u \in V$ such that $u \neq 0$ and $Au = \lambda u$ for some $\lambda \in \mathbb{F}$, then the subspace $U = \{au|\ a \in \mathbb{F}\}$ is a 1d subspace of V invariant under A.

The notion of an eigenvalue and eigenvector is closely related to 1d invariant subspaces. Recall that (λ, v) such that $Av = \lambda v$ is an eigenpair. Note also that

$$Av = \lambda v \iff (A - \lambda I)u = 0$$

so that λ is an eigenvalue if and only if $A - \lambda I$ is injective.

510: **Fall 2019**

Lecture 7: Normed Linear Spaces and Inner Products

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

References: Appendix A.7 **[C&D]**; Chapter 6 of **[Ax]**

1 Normed Linear Spaces

Why? We need to be able to measure distances between things and hence, we need to formally introduce a metric onto the linear spaces we consider. This is going to be important for classical stability notions.

Intuitively the norm of a vector is a measure of its *length*. Of course, since elements of a linear space can be vectors in $\mathbb{R}^n$, matrices, functions, etc. Hence, the concept of norm must be defined in general terms. Let the field F be $\mathbb{R}$ or $\mathbb{C}$.

Definition (*Normed Linear Space.*) A linear space (V, F) is said to be a *normed linear space* if there exists a map $\|\cdot\| : V \to \mathbb{R}_+$ satisfying

1. the 'triangle inequality' holds:
$$\|v_1 + v_2\| \le \|v_1\| + \|v_2\|, \quad \forall\ v_1, v_2 \in V$$
2. length under scalar multiplication satisfies
$$\|\alpha v\| = |\alpha|\|v\|, \quad \forall\ v \in V, \alpha \in F$$
3. the only element with zero length is the zero vector:
$$\|v\| = 0 \iff v = 0$$

The expression $\|v\|$ is read 'the norm of v' and the function $\|\cdot\|$ is said to be the *norm* on V. It is often convenient to label the space $(V, F, \|\cdot\|)$.

Example.

1. ℓ_p–norms:
$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$$
   - ℓ_1–norm (sum norm)
$$\|x\|_1 = \sum_{i=1}^n |x_i|$$
   - ℓ_2–norm (Euclidean norm)
$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$$
   - ℓ_∞–norm (sup norm)
$$\|x\|_\infty = \max_i |x_i|$$

Recall: sup ='supremum' is the least upper bound and inf ='infimum' is the greatest lower bound. The concepts of infimum and supfctum are similar to minimum and maximum, but are more useful in analysis because they better characterize special sets which may have no minimum or maximum.

2. Function norms: $\mathcal{C}([t_0, t_1], \mathbb{R}^n)$
   - L_1–norm
   $$\|f\|_1 = \int_{t_0}^{t_1} \|f(t)\| \, dt$$
   where $\|f(t)\|$ is any of the ℓ_p norms from above
   - L_2–norm
   $$\|f\|_2 = \left(\int_{t_0}^{t_1} \|f(t)\|^2 \, dt \right)^{1/2}$$

Note. since f maps into $\mathbb{R}^n$, this is equivalent to each component function of f being in $\mathcal{C}([t_0, t_1], F)$ and since (as we will see) all norms on $\mathbb{R}^n$ are equivalent, it does not matter which norm we choose on $\mathbb{R}^n$.

What do we do with norms? A typical engineering use of norms is in deciding whether or not an iterative technique converges or not—this will be very important in our analysis of stability.

Aside on Convergence (Recall from Calculus).

Definition 1. Convergence. Given a normed linear space $(V, F, \|\cdot\|)$ and a sequence of vectors $\{v_i\}_{i=1}^{\infty} \subset V$, we say the sequence $\{v_i\}_{i=1}^{\infty}$ *converges* to the vector $v \in V$ iff the sequence of non-negative real numbers $\|v_i - v\|$ tends to zero as $i \to \infty$.

We denote convergence by $v_i \to v$ or $\lim_{i\to\infty} v_i = v$.

1.1 Equivalence of Norms

Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on (V, F) are said to be equivalent if $\exists\ m_\ell,\ m_u \in \mathbb{R}_+$ such that for all $v \in V$

$$m_\ell \|v\|_a \le \|v\|_b \le m_u \|v\|_a$$

It is crucial to note that the same m_ℓ and m_u must work for all $v \in V$.

How can we use equivalence of norms? Consider (V, F) and two equivalent norms $\|\cdot\|_a$ and $\|\cdot\|_b$. Then, convergence is equivalent—i.e.

$$v_i \overset{i\to\infty}{\longrightarrow} v \ \text{ in } \ \|\cdot\|_a \iff v_i \overset{i\to\infty}{\longrightarrow} v \ \text{ in } \ \|\cdot\|_b$$

Example.

1. $$\|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2$$
2. $$\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty$$
3. $$\|x\|_\infty \le \|x\|_1 \le n\|x\|_\infty$$

Aside: Many other things are equivalent too.

- Cauchy nature of sequences is equivalent—i.e.

$$(v_i)_{i=1}^{\infty} \text{ Cauchy in } \ \|\cdot\|_a \iff (v_i)_{i=1}^{\infty} \text{ Cauchy in } \ \|\cdot\|_b$$

- Completeness is equivalent—i.e.

$$(V, F) \text{ complete in } \|\cdot\|_a \iff (V, F) \text{ complete in } \|\cdot\|_b$$

- density is equivalent—i.e.

$$X \text{ a dense subset of } V \text{ in } \|\cdot\|_a \iff X \text{ a dense subset of } V \text{ in } \|\cdot\|_b$$

What is the point of this? As far as convergence is concerned, $\|\cdot\|_a$ and $\|\cdot\|_b$ lead to the same answer, yet in practice the analysis may be more easy in one than the other.

1.2 Continuity

You have probably seen the notion of continuity in your calculus class from undergrad.

Intuition. Continuous maps have the desirable property that a small perturbation in x results in a small perturbation in the operator at that point. They are paramount in the study of robustness and the effect of perturbations of data.

Definition. (*Continuity.*) Let $F = \mathbb{R}$ (or $\mathbb{C}$) and consider two normed linear spaces $(U, G, \|\cdot\|_U)$ and $(V, F, \|\cdot\|_V)$. Let f be a map (or operator) s.t. $f : U \to V$.

a. (Local continuity). We say f is *continuous at* $u \in U$ iff for every $\varepsilon > 0$, there exxists $\delta > 0$ (possibly depending on ε at u) such that considering points $u' \in U$

$$\|u' - u\|_U < \delta \implies \|f(u') - f(u)\|_V < \varepsilon$$

b. (Global continuity). We say f is *continuous on* U iff it is continuous at every $u \in U$.

1.3 Induced Norms

Definition. (*Induced Operator Norm.*) The *induced (operator) norm* of f is defined to be

$$\|f\| = \sup_{u \neq 0} \left\{ \frac{\|f(u)\|_V}{\|u\|_U} \right\}$$

Let $A : U \to V$ be a (continuous) linear operator. Let U and V be endowed with the norms $\|\cdot\|_U$ and $\|\cdot\|_V$, resp. Then, the induced norm of A is defined by

$$\|A\|_i = \sup_{u \neq 0} \frac{\|Au\|_V}{\|u\|_U}$$

Interpretation. The induced norm $\|A\|$ is the maximum gain of the map A over all directions; moreover, $\|A\|$ depends on the choice of the norms $\|\cdot\|_U$ and $\|\cdot\|_V$ in the domain and co-domain, resp.

Definition. (*Bounded Operator.*) Let V, W be normed vector spaces over field F (=$\mathcal{R}$ or $\mathbb{C}$). A linear transformation or linear operator $A : V \to W$ is *bounded* if $\exists\ 0 < K < \infty$ such that

$$\|Ax\|_W \leq K\|x\|_V, \quad \forall x \in V$$

Theorem 1. Connections between bounded and continuous linear operators. Let V, W be normed vector spaces and let $A : V \to W$ be a linear transformation. The following statements are equivalent:

a. A is a bounded linear transformation
b. A is continuous everywhere in V

c. A is continuous at $0 \in V$.

Proof. [a. $\implies$ b.] Suppose a. Let K as in the definition of bounded linear transformation. By linearity of A we have that

$$\|A(x) - A(\tilde{x})\|_W = \|A(x - \tilde{x})\|_W \le K\|x - \tilde{x}\|_V$$

which implies b.

[b. $\implies$ c.] trivial

[c. $\implies$ a.] If A is continuous at $0 \in V$, $\exists\ \delta > 0$ such that for $v \in V$ with $\|v\| < \delta$ we have that $\|Av\| < 1$ since continuity at a point x implies that for every $\varepsilon > 0$, $\exists\ \delta > 0$ s.t.

$$\|x - x'\|_V < \delta \implies \|A(x) - A(x')\|_Q < \varepsilon$$

so, in particular, it is true for $\varepsilon = 1$.

Now, let $x \in V$ with $x \ne 0$. Then

$$\left\| \frac{\delta}{2\|x\|_V} x \right\|_V = \frac{\delta}{2} < \delta \implies \left\| A\left(\frac{\delta}{2\|x\|_V} x \right) \right\|_W < 1$$

But by linearity of A,

$$\|Ax\|_W = \left\| \frac{2\|x\|_V}{\delta} A\left(\delta \frac{x}{2\|x\|_V} \right) \right\|_W = \frac{2\|x\|_V}{\delta} \left\| A\left(\delta \frac{x}{2\|x\|_V} \right) \right\|_W \le \frac{\|x\|_V}{\delta} \cdot 1 = \frac{2}{\delta}\|x\|_V$$

And, since $\|Ax\|_W \le (2/\delta)\|x\|_V$ for $x = 0$ too, then $K = 2/\delta$.

□

Facts about induced norms. Let $(U, \mathbb{F}, \|\cdot\|_U)$, $(V, \mathbb{F}, \|\cdot\|_V)$, and $(W, \mathbb{F}, \|\cdot\|_W)$ be normed linear spaces and let $A : V \to W$, $\tilde{A} : V \to W$ and $B : U \to V$ be linear maps. Note that $\forall \alpha \in \mathbb{F}$, $\alpha A : V \to W$, $A + \tilde{A} : V \to W$ and $AB : U \to W$ are also linear maps. Then we have that

1. $\|Av\|_W \le \|A\|\|v\|_V$
2. $\|\alpha A\| = |\alpha|\|A\|, \quad \forall\ \alpha \in F$
3. $\|A + \tilde{A}\| \le \|A\| + \|\tilde{A}\|$
4. $\|A\| = 0 \iff A = 0$ (zero map)
5. $\|AB\| \le \|A\|\|B\|$

Matrix norms are *induced norms* or *operator norms*. Consider $A \in \mathbb{F}^{m\times n}$, $A : \mathbb{F}^n \to \mathbb{F}^m$. The induced matrix norm is defined by

$$\|A\| = \sup\{\|Ax\| : \ x \in \mathbb{F}^n,\ \|x\| = 1\} = \sup\left\{ \frac{\|Ax\|}{\|x\|} : \ x \in \mathbb{F}^n,\ x \ne 0 \right\}$$

With this definition, we can show the following:

- 1–norm (column sums)

$$\|A\|_{i,1} = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|$$

- 2–norm (largest singular value—we will discuss singular values later)

$$\|A\|_{i,2} = \sqrt{\lambda_{\max}(A^*A)} = \sigma_{\max}(A)$$

- (not induced) Forbenius norm

$$\|A\|_{i,\mathbb{F}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} = (\mathrm{tr}(A^TA))^{1/2}$$

  Note: $\|A\|_2 \le \|A\|_{\mathbb{F}}$
- ∞–norm (row sums)

$$\|A\|_{i,\infty} = \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|$$

Suppose A is an $m \times n$ matrix and we have the induced norm

$$\|A\|_p = \sup_{x\neq 0}\frac{\|Ax\|_p}{\|x\|_p}$$

Since $\|x\|_p$ is a scalar, we have

$$\|A\|_p = \sup_{x\neq 0}\frac{\|Ax\|_p}{\|x\|_p} = \sup_{x\neq 0}\left\|\frac{Ax}{\|x\|_p}\right\|_p$$

Since $x/\|x\|_p$ has unit length, we can express the induced norm as follows:

$$\|A\|_p = \sup_{\|x\|_p=1}\|Ax\|_p$$

That is, $\|A\|_p$ is the supremum of $\|Ax\|_p$ on the unit ball $\|x\|_p = 1$. Note that $\|Ax\|_p$ is a continuous function of x and the unit ball $B_1(0) = \{x|\ \|x\|_p = 1\}$ is closed and bounded (*compact*).

Fact. On a compact set, a continuous function always achieves its maximum and minimum values.

The above fact enables us to replace the 'sup' with a 'max'. Indeed,

$$\|A\|_p = \max_{x\neq 0}\frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|_p=1}\|Ax\|_p$$

When computing the norm of an operator A, the definition is the starting point and the remainder of the process is the following:
step 1. First, find a *candidate* for the norm, call it K for now, that satisfies

$$\|Ax\|_p \le K\|x\|_p, \quad \forall\ x$$

step 2. Then, find at least one non-zero x_0 for which

$$\|Ax_0\|_p = K\|x_0\|_p$$

step 3. Finally, set $\|A\|_p = K$. That is,

$$\|Ax_0\| = K\|x_0\| = \|A\|\|x_0\| \implies \frac{\|Ax_0\|}{\|x_0\|} = \|A\|$$

which is the definition

We can do this because

$$\|Ax\|_p \le \|A\|_p\|x\|_p$$

which is clearly the case by inspecting the definition of $\|A\|_p$.

Recall that

$$\|x\|_\infty = \max_i |x_i|$$

The sup norm is the max of row sums:

$$\|A\|_\infty = \max_i \left(\sum_{j=1}^{n} |a_{ij}| \right) \qquad \text{(max of row sum)}$$

Proof. We prove this using the above steps. Let $A \in \mathbb{R}^{m\times n}$. Then, $Ax \in \mathbb{R}^m$ is a vector so that

$$\begin{aligned}
\|Ax\|_\infty &= \left\| \begin{matrix} \sum_{k=1}^n a_{1k}x_k \\ \vdots \\ \sum_{k=1}^n a_{mk}x_k \end{matrix} \right\|_\infty \\
&= \max_{i\in\{1,\dots,m\}} \left| \sum_{k=1}^{n} a_{ik}x_k \right| \\
&\leq \max_{i\in\{1,\dots,m\}} \sum_{k=1}^{n} |a_{ik}x_k| \\
&\leq \left(\max_{i\in\{1,\dots,m\}} \sum_{k=1}^{n} |a_{ik}| \right) \max_{k\in\{1,\dots,n\}} |x_k| \\
&= \left(\max_{i\in\{1,\dots,m\}} \sum_{k=1}^{n} |a_{ik}| \right) \|x\|_\infty
\end{aligned}$$

so that we have completed step 1—that is,

$$K = \max_{i\in\{1,\dots,m\}} \sum_{k=1}^{n} |a_{ik}|$$

so that

$$\|Ax\|_p \leq K\|x\|_\infty, \quad \forall\ x$$

Step 2 above requires that we find non-zero x_0 for which equality holds in the above inequality. Examination reveals that we have equality if x_0 is defined to have the components

$$x_k = \begin{cases} \frac{a_{i^*k}}{|a_{i^*k}|}, & a_{ik} \neq 0 \\ 1, & a_{i^*k} = 0 \end{cases} \qquad 1 \leq k \leq n$$

where i^* above is

$$i^* = \arg \max_{i\in\{1,\dots,m\}} \sum_{j=1}^{n} |a_{ij}|$$

You can check this by starting with the definition for $\|Ax\|_\infty$ and plugging in this choice of x_0 to see that

$$\|Ax_0\| = K\|x_0\|$$

$$
\begin{aligned}
\|A\|_\infty = \sup_{\|x\|\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} &= \sup_{\|x\|\neq 0} \frac{\max_{i\in\{1,\ldots,m\}} \left|\sum_{j=1}^n a_{ij}x_j\right|}{\max_{i\in\{1,\ldots,m\}} |x_i|} \\
&= \sup_{\|x\|=1} \max_{i\in\{1,\ldots,m\}} \left|\sum_{j=1}^n a_{ij}x_j\right| \\
&\leq \sup_{\|x\|=1} \max_{i\in\{1,\ldots,m\}} \sum_{j=1}^n |a_{ij}||x_j| \quad \text{(triangle inequality)} \\
&\leq \sup_{\|x\|=1} \max_{i\in\{1,\ldots,m\}} \sum_{j=1}^n |a_{ij}| \max_j |x_j| \\
&= \max_{i\in\{1,\ldots,m\}} \sum_{j=1}^n |a_{ij}|
\end{aligned}
$$

where equality holds when we let x be the vector with $\|x\|_\infty = 1$ such that it *extracts* the max row sum, i.e. for $1 \leq j \leq n$, let

$$
x_j = \begin{cases} \frac{a_{i^*j}}{|a_{i^*j}|}, & a_{i^*j} \neq 0 \\ 1 & , a_{i^*j} = 0 \end{cases} \tag{1}
$$

where

$$
i^* = \arg \max_{i\in\{1,\ldots,m\}} \sum_{j=1}^n |a_{ij}|
$$

Allowing for the vector x to have the above property, we have $\|A\|_\infty = \max_{i\in\{1,\ldots,m\}} \sum_{j=1}^n |a_{ij}|$

□

Definition. (*Positive definite matrix.*) A matrix $P \in \mathbb{R}^{n\times n}$ such that for all $x \in \mathbb{R}^n$

$$
x^T P x \geq 0, \text{ whenever } x \neq 0
$$

is said to be positive definite.

A positive definite matrix P has only positive eigenvaules. Indeed, for all (non-trivial) eigenvectors x we have that

$$
x^T P x = \lambda x^T x = \lambda \|x\|^2 > 0
$$

and thus, $\lambda > 0$.

Fact. Every positive definite matrix induces a norm via

$$
\|x\|_P^2 = x^T P x
$$

This is because $P = SS^{1/2}$

2 Inner Product Space (Hilbert Spaces)

Let the field F be $\mathbb{R}$ or $\mathbb{C}$ and consider the linear space (H, F). The function

$$
\langle \cdot, \cdot \rangle : H \times H \to F \quad [(x,y) \mapsto \langle x, y\rangle]
$$

is called *inner product* iff

1. $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle, \quad \forall x, y, z \in H$
2. $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle, \quad \forall \alpha \in F, \ \forall x, y \in H$

3. $\|x\|^2 = \langle x, x\rangle > 0 \iff x \in H$ s.t. $x \neq 0$
4. $\langle x, y\rangle = \overline{\langle y, x\rangle}$

where the overbar denotes the complex conjugate of $\langle y, x\rangle$. Moreover, in (c) $\|\cdot\|$ is called the norm induced by the inner product.

definition.(*Hilbert Space.*) A space equipped with an inner product is called a *Hilbert space* and denoted by $(H, F, \langle\cdot,\cdot\rangle)$.

Example. The following are examples:

1. $(F^n, F, \langle\cdot,\cdot\rangle)$ is a Hilbert space under the inner product

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i \bar{y}_i = y^* x$$

2. $L_2([t_0, t_1], F^n)$ (space of square, integrable F^n-valued functions on $[t_0, t_1]$) defined by the inner product by

$$\langle f, g\rangle = \int_{t_0}^{t_1} f(t) g(t)^* \; dt$$

Theorem 2. Schwawrz's Inequality. Let $(H, F, \langle\cdot,\cdot\rangle)$ be an inner product space.

$$|\langle x, y\rangle| \leq \|x\|\|y\| \;\; \forall\; x, y \in H$$

This inequality implies the triangle inequality for the inner product norm $\|x\|^2 = \langle x, x\rangle$.

Proof. Choose $\alpha \in \mathbb{F}$ with $|\alpha| = 1$ such that

$$\alpha\langle x, y\rangle = |\langle x, y\rangle|$$

(basically choose α to have the sign which makes the expression on the left non-negative). Then, for all $\lambda \in \mathbb{R}$, we have

$$0 \leq \|\lambda x + \alpha y\|^2 = \lambda^2\|x\|^2 + 2\lambda|\langle x, y\rangle| + \|y\|^2$$

Since this last polynomial in λ is non-negative, we have

$$|\langle x, y\rangle|^2 - \|x\|^2\|y\|^2 \leq 0$$

□

2.1 Orthogonality

Definition. (*Orthogonality.*) Two vectors $x, y \in H$ are said to be *orthogonal* $\iff \langle x, y\rangle = 0$.

If two vectors are orthogonal, they satisfy

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 \;\; \text{(Pythogoras' Theorem)}$$

Definition. (*Orthogonal Complement.*) If M is a subspace of a Hilbert space, the subset

$$M^\perp = \{y \in H |\; \langle x, y\rangle = 0, \;\; \forall x \in M\}$$

is called the *orthogonal complement* of M.

$M \cap M^\perp = \{0\}$.

Proof. Suppose not. Then there exists $x \neq 0$ with $x \in M \cap M^{\perp}$. Hence,

$$\langle x, y \rangle = 0 \quad \forall y \in M$$

but $x \in M$ so

$$\langle x, x \rangle = 0$$

This in turn implies that $x = 0 \rightarrow\leftarrow$. □

2.2 Adjoints

Let F be $\mathbb{R}$ or $\mathbb{C}$ and let $(U, F, \langle \cdot, \cdot \rangle_U)$ and $(V, F, \langle \cdot, \cdot \rangle_V)$ be Hilbert spaces. Let $\mathcal{A} : U \to V$ be continuous and linear.

Definition. (*Adjoint.*) The *adjoint* of $\mathcal{A}$, denoted $\mathcal{A}^*$, is the map $\mathcal{A}^* : V \to U$ such that

$$\langle v, Au \rangle_V = \langle A^* v, u \rangle_U$$

Consider $A : V \to W$.

matrix type	notation	Adjoint property
transpose	A^T	$\langle Af, g \rangle = \langle f, A^* g \rangle, \forall\ f \in V, g \in W$ (adjoint)
symmetric	$A = A^T$	$\langle Af, g \rangle = \langle A^* f, g \rangle, \forall\ f \in V, g \in W$ (self-adjoint)
orthogonal	$A^{-1} = A^T$	$\langle f, g \rangle = \langle Af, Ag \rangle, \forall\ f \in V, g \in W$ (isometry)

Example. Let $f(\cdot), g(\cdot) \in C([t_0, t_1], \mathbb{R}^n)$ and define $\mathcal{A} : C([t_0, t_1], \mathbb{R}^n) \to \mathbb{R}$ by

$$\mathcal{A} : f(\cdot) \mapsto \langle g(\cdot), f(\cdot) \rangle$$

Find the adjoint map of $\mathcal{A}$.

soln.

$$\mathcal{A}^* : \mathbb{R} \to C([t_0, t_1], \mathbb{R}^n)$$

such that

$$\langle v, \mathcal{A} f(\cdot) \rangle_{\mathbb{R}} = \langle \mathcal{A}^* v, f(\cdot) \rangle_{C([t_0, t_1], \mathbb{R}^n)}$$

where $v \in \mathbb{R}$ and $f(\cdot) \in C([t_0, t_1], \mathbb{R}^n)$.

c.f. with

$$\langle \mathcal{A}^* v, f(\cdot) \rangle \implies \mathcal{A}^* : v \mapsto v^* g(\cdot)$$

(multiplication by $g(\cdot)$). Indeed,

$$\begin{aligned} \langle v, \mathcal{A} f(\cdot) \rangle_{\mathbb{R}} &= \langle v, \langle g(\cdot), f(\cdot) \rangle_{C([t_0, t_1], \mathbb{R}^n)} \rangle_{\mathbb{R}} \\ &= v^* \langle g(\cdot), f(\cdot) \rangle_{C([t_0, t_1], \mathbb{R}^n)} \\ &= \langle v^* g(\cdot), f(\cdot) \rangle_{C([t_0, t_1], \mathbb{R}^n)} \end{aligned}$$

We can also compute the matrix of an adjoint of a linear operator.

Definition. (*Adjoint of a matrix.*) The adjoint of matrix A is

$$A^* = \overline{A^T}$$

Regarding a matrix representation of a linear operator T, if the matrix A is its representation with respect to bases α and β of V and W, respectively, then we would like for the adjoint of A to coincide with the matrix of the adjoint of T with respect to β and α.

We need the following definition.

Definition. (*Orthonormal Basis.*) An *orthonormal basis* for an inner product space V with finite dimension is a basis for V whose vectors are *orthonormal*—that is, they are all unit vectors and orthogonal to each other.

For example, the standard basis for $\mathbb{R}^n$ is an orthonormal basis.

Theorem 3. Adjoint of Matrix Representation. Let V, W be finite dimensional inner product spaces with α, β as orthonormal bases for V and W, respectively. If $T : V \to W$ is a linear operator, then

$$[T^*]_\alpha^\beta = ([T]_\beta^\alpha)^*$$

where $[T]_\alpha^\beta$ denotes the matrix representation of T wrt the bases α and β for V and W, respectively.

Proof. If (a_{ij}) is the matrix representation of T wrt α, β, then

$$T\alpha_i = \sum_j a_{ij}\beta_j, \quad 1 \le i \le m$$

Since α is orthonormal,

$$\langle T\alpha_i, \beta_j\rangle = a_{ij}, \quad 1 \le i \le m, \ 1 \le j \le n$$

Indeed, plugging in $T\alpha_i$ from above we have

$$\left\langle \sum_\ell a_{i\ell}\beta_\ell, \beta_j \right\rangle = \sum_\ell a_{i\ell}\langle \beta_\ell, \beta_j\rangle = a_{ij}$$

since $\langle \beta_\ell, \beta_j\rangle = 1$ only when $\ell = j$. Similarly, if $B = (b_{ij})$ is the representation of $([T]_\beta^\alpha)^*$, then

$$\langle T^*\beta_i, \alpha_j\rangle = b_{ij}$$

Then

$$b_{ij} = \langle T^*\beta_i, \alpha_j\rangle = \langle \beta_i, T\alpha_j\rangle = \overline{\langle T\alpha_j, \beta_i\rangle} = a_{ji}^*$$

so that

$$B = \overline{A^T}$$

□

Fact. ("Necessity") This is NOT true in general—i.e. it is true when both α and β are orthonormal.

We can still write $[T^*]_\beta^\alpha$ in terms of $[T]_\alpha^\beta$ without assuming α and β are orthonormal. Indeed, let $\alpha = \{\alpha_i\}_{i=1}^m$ and $\beta = \{\beta_j\}_{j=1}^n$. Then

$$[T^*]_\beta^\alpha = C^{-1}([T]_\alpha^\beta)^* B$$

where

$$C = \begin{bmatrix} \langle\alpha_1, \alpha_1\rangle & \langle\alpha_2, \alpha_1\rangle & \cdots & \langle\alpha_m, \alpha_1\rangle \\ \langle\alpha_1, \alpha_2\rangle & \langle\alpha_2, \alpha_2\rangle & \cdots & \langle\alpha_m, \alpha_2\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\alpha_1, \alpha_m\rangle & \langle\alpha_2, \alpha_1\rangle & \cdots & \langle\alpha_m, \alpha_m\rangle \end{bmatrix}$$

and

$$B = \begin{bmatrix} \langle\beta_1, \beta_1\rangle & \langle\beta_2, \beta_1\rangle & \cdots & \langle\beta_n, \beta_1\rangle \\ \langle\beta_1, \beta_2\rangle & \langle\beta_2, \beta_2\rangle & \cdots & \langle\beta_n, \beta_2\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\beta_1, \beta_n\rangle & \langle\beta_2, \beta_1\rangle & \cdots & \langle\beta_n, \beta_n\rangle \end{bmatrix}$$

2.3 Self-Adjoint Maps

Let $(H, F, \langle\cdot,\cdot\rangle_H)$ be a Hilbert space and let $\mathcal{A} : H \to H$ be a continuous linear map with adjoint $\mathcal{A}^* : H \to H$.

Definition. (*Self-Adjoint.*) The map $\mathcal{A}$ is ***self-adjoint*** iff $\mathcal{A} = \mathcal{A}^*$, or equivalently

$$\langle x, \mathcal{A}y \rangle_H = \langle \mathcal{A}x, y \rangle_H, \quad \forall x, y \in H$$

Definition. (*Hermitian matrices.*) Let $H = F^n$ and let $\mathcal{A}$ be represented by a matrix $A : F^n \to F^n$, $A \in F^{n\times n}$. Then A is self-adjoint iff the matrix A is Hermitian (i.e. $A = A^*$, viz. $a_{ij} = \bar{a}_{ij}$, $\forall i, j \in \{1, \ldots, n\}$.

Fact. Let $\mathcal{A} : H \to H$ be a self-adjoint, continuous linear map. Then

a. all eigenvalues of A are real

b. if λ_i and λ_k are distinct eigenvalues of A, then any eigenvector v_i associated with λ_i is orthogonal to any eigenvector v_k associated with λ_k.

Proof of part (a). From $Av_i = \lambda_i v_i$ and $A^* = A$, we get that

$$\langle v_i, Av_i \rangle = \lambda_i \|v_i\|^2 = \langle Av_i, v_i \rangle = \overline{\langle v_i, Av_i \rangle}$$

It follows that λ_i is real. $\square$

Proof of part (b). From $Av_i = \lambda_i v_i$ and $Av_k = \lambda_k v_k$ we obtain

$$\lambda_i \langle v_k, v_i \rangle = \langle v_k, Av_i \rangle$$

and

$$\lambda_k \langle v_k, v_i \rangle = \langle Av_k, v_i \rangle = \langle v_k, Av_i \rangle$$

Subtracting the last two equations, we get

$$(\lambda_i - \lambda_k)\langle v_i, v_k \rangle = 0$$

Since $\lambda_i \neq \lambda_k$,

$$\langle v_i, v_k \rangle = 0$$

$\square$

2.4 Orthogonal Projection

Recall that we already defined a direct sum of spaces:

$$V = U \oplus W \iff \forall x \in V, \ \exists! \ x = u + v : \ u \in U, \ v \in V$$

Fact. Let U be a closed subspace of $(H, \mathbb{F}, \langle \cdot, \cdot \rangle)$. Then, we have the direct sum decomposition

$$H = U \oplus U^\perp$$

Equivalently, $\forall x \in H$, $\exists! \ u \in U$ called the orthogonal projection of x onto the subspace U such that $x - u \in U^\perp$.

The operator that does this operation, called the orthogonal projection operator, is such that its range space is U and its null space is $U^\perp$.

DIY Exercise. Let $(H, \mathbb{C}, \langle \cdot, \cdot \rangle)$ be a Hilbert space and U be a closed subspace defined by an orthonormal basis $(u_i)_1^k$. For any $x \in H$, let x_p be the orthogonal projection of x onto U. By direct calculation, show that

$$x_p = \sum_{i=1}^{k} \langle u_i, x \rangle u_i$$

and

$$\|x - x_p\| \leq \|x - y\|, \ \forall y \in U, \ y \neq x_p$$

Hint: write $y = \sum_{i=1}^k \eta_i u_i$ and minimize $\|x - y\|$ w.r.t. η_i's. Further show that if $H = \mathbb{C}^n$ and the basis $(u_i)_{i=1}^k$ is not orthonormal, then $x_p = U(U^*U)^{-1}U^*x$ where U is $n \times k$ with columns from the u_i's.

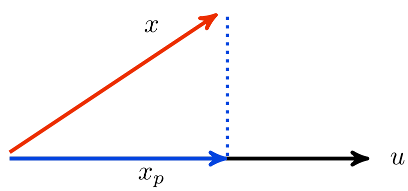
In fact, in the orthonormal case, if we define $U = \begin{bmatrix} u_1 & \cdots & u_k \end{bmatrix}$, then

$$P_U = UU^\top$$

This does exactly this operation

$$\sum_{i=1}^k \langle u_i, x \rangle u_i$$

Why is this the case? Well consider a $U = [u]$ as a line so that we have just a single vector u spanning U. Then, consider a vector x we want to project onto U.



Then the scalar projection of x onto u can be found using basic cosine properties. Indeed,

$$|x_p| = \|x\| \cos(\theta) = u \cdot x$$

Hence,

$$x_p = \langle u, x \rangle u$$

2.5 Properties of the Adjoint

Let A be a linear continuous map from U to V (both Hilbert spaces). Then, A^* is linear and continuous with induced norm $\|A^*\| = \|A\|$. Moreover, $A^{**} = A$.

Theorem 4. Finite Rank Operator Lemma (FROL). Let $\mathbb{F}$ be $\mathbb{R}$ or $\mathbb{C}$. Let $(H, \mathbb{F}, \langle \cdot, \cdot \rangle_H)$ be a Hilbert space and consider $\mathbb{F}^m$ as the Hilbert space $(\mathbb{F}^m, \mathbb{F}, \langle \cdot, \cdot \rangle_{F^m})$. Let $A : H \to \mathbb{F}^m$ be a continuous linear map with adjoint $A^* : \mathbb{F}^m \to H$. Then

$$A^* : \mathbb{F}^m \to H, \ AA^* : \mathbb{F}^m \to \mathbb{F}^m, \ A^*A : H \to H$$

are continuous linear maps with AA^* and A^*A self-adjoint. Furthermore,

a. $H = \mathcal{R}(A^*) \overset{\perp}{\oplus} \mathcal{N}(A)$, $F^m = \mathcal{R}(A) \overset{\perp}{\oplus} \mathcal{N}(A^*)$

b. The restriction $A|_{\mathcal{R}(A^*)}$ is a bijection of $\mathcal{R}(A^*)$ onto $\mathcal{R}(A)$ and

$$\mathcal{N}(AA^*) = \mathcal{N}(A^*), \quad \mathcal{R}(AA^*) = \mathcal{R}(A)$$

c. The restriction $A^*|_{\mathcal{R}(A)}$ is a bijection of $\mathcal{R}(A)$ onto $\mathcal{R}(A^*)$ and

$$\mathcal{N}(A^*A) = \mathcal{N}(A), \quad \mathcal{R}(A^*A) = \mathcal{R}(A^*)$$

The following result proves that $\mathbb{F}^m = \mathcal{R}(A) \overset{\perp}{\oplus} \mathcal{N}(A^*)$.

Proposition 1. Let V and W be finite-dimensional nonzero inner product spaces. Let $T \in \mathcal{L}(V, W)$. Then $\ker(T^*) = (\text{im}(T))^\perp$

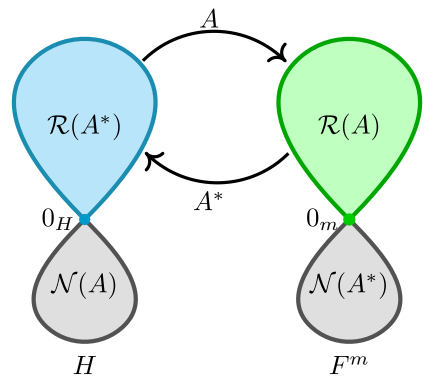


Figure 1: The orthogonal decomposition of the domain and the co-domain of a finite rank operator $A : H \to F^m$ and its associated bijections.

Proof. We must show that $\ker(T^*) \subset (\mathrm{im}(T))^\perp$ and $\ker(T^*) \supset (\mathrm{im}(T))^\perp$.

Let $w \in \ker(T^*)$. Then $T^*(w) = 0$. Hence, for all $v \in V$, we have $0 = \langle v, 0\rangle = \langle v, T^*(w)\rangle$ which implies that $\langle T(v), w\rangle = 0$ for all $v \in V$. Therefore w is orthogonal to every vector in $\mathrm{im}(T)$ so $w \in (\mathrm{im}(T))^\perp$.

Now, let $w \in (\mathrm{im}(T))^\perp$. Then every vector in $\mathrm{im}(T)$ is orthogonal to w, that is $\langle T(v), w\rangle = 0$ for all $v \in V$. Thus $\langle v, T^*(w)\rangle = 0$ for all $v \in V$. But if $\langle v, T^*(w)\rangle = 0$ for all $v \in V$ then this must imply that $T^*(w) = 0$ so $w \in \ker(T^*)$. $\square$

You can argue $H = \mathcal{R}(A^*) \overset{\perp}{\oplus} \mathcal{N}(A)$ similarly.

Proposition 2. Let V and W be finite dimensional inner product spaces and let $T \in \mathcal{L}(V, W)$. Then $\mathrm{im}(T^*) = (\ker(T))^\perp$.

The following are important comments on this result:

- It is crucial to notice that $AA^* : \mathbb{F}^m \to \mathbb{F}^m$ and hence, by

$$\mathcal{N}(AA^*) = \mathcal{N}(A^*), \quad \mathcal{R}(AA^*) = \mathcal{R}(A),$$

  the study of the range of A and the null space of A^* is equivalent to study of the range and null space, resp., of any (Hermitain) matrix representation M of AA^* (cf. controllability. . .)
- This result also leads to singular value decomposition which is often very useful in studying the modes of a system—i.e. given a matrix A we can decompose it into

$$A = U\Sigma V \in \mathbb{F}^{m\times n}$$

  where Σ contains the singular values of A (or the square roots of the common positive eigenvalues of A^*A and AA^*) and $U = [U_1 : U_2] \in \mathbb{F}^{m\times m}$ where U_1 cols form an orthonormal basis of $R(A)$ and $R(U_2) = N(A^*)$ (cols of U_2 form an orthonormal basis for $N(A^*)$). Cols of U form a complete orthonormal basis of eigenvectors of AA^*. And, V is also unitary $V^*V = I_n$ and $R(V_1) = R(A^*)$ and $R(V_2) = N(A)$. cols of V_1 form ortho basis of $R(A^*)$ and cols of V_2 form ortho basis of $N(A)$.

510: **Fall 2019**

Lecture 8: Singular Value Decomposition

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

References: Appendix A.7 **[C&D]**; Chapter 7 of **[Ax]**

1 Finite Rank Operator Lemma

Theorem 1. Finite Rank Operator Lemma (FROL). Let $\mathbb{F}$ be $\mathbb{R}$ or $\mathbb{C}$. Let $(H, \mathbb{F}, \langle\cdot,\cdot\rangle_H)$ be a Hilbert space and consider $\mathbb{F}^m$ as the Hilbert space $(\mathbb{F}^m, \mathbb{F}, \langle\cdot,\cdot\rangle_{F^m})$. Let $A : H \to \mathbb{F}^m$ be a continuous linear map with adjoint $A^* : \mathbb{F}^m \to H$. Then

$$A^* : \mathbb{F}^m \to H,\ AA^* : \mathbb{F}^m \to \mathbb{F}^m,\ A^*A : H \to H$$

are continuous linear maps with AA^* and A^*A self-adjoint. Furthermore,

a. $H = \mathcal{R}(A^*) \overset{\perp}{\oplus} \mathcal{N}(A)$, $\mathbb{F}^m = \mathcal{R}(A) \overset{\perp}{\oplus} \mathcal{N}(A^*)$

b. The restriction $A|_{\mathcal{R}(A^*)}$ is a bijection of $\mathcal{R}(A^*)$ onto $\mathcal{R}(A)$ and

$$\mathcal{N}(AA^*) = \mathcal{N}(A^*),\quad \mathcal{R}(AA^*) = \mathcal{R}(A)$$

c. The restriction $A^*|_{\mathcal{R}(A)}$ is a bijection of $\mathcal{R}(A)$ onto $\mathcal{R}(A^*)$ and

$$\mathcal{N}(A^*A) = \mathcal{N}(A),\quad \mathcal{R}(A^*A) = \mathcal{R}(A^*)$$

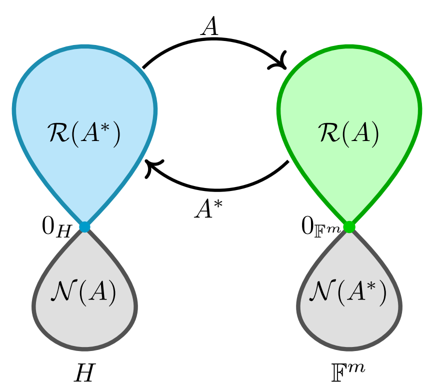


Figure 1: The orthogonal decomposition of the domain and the co-domain of a finite rank operator $A : H \to F^m$ and its associated bijections.

2 Singular Value Decomposition

FROL leads to singular value decomposition which is often very useful in studying the modes of a system or for solving systems of equations or for understanding the contribution that different features have in terms of explaining things like correlations/variance.

Note: this is not the only decomposition one can perform on a matrix which is useful, but it is one that is widely used in a number of applications.

2.1 Overview of SVD

$$A = \underset{\text{left singular vecs}}{U}\;\underset{\text{singular values}}{\Sigma}\;\underset{\text{right singular vecs}}{V^*}$$

Given a matrix $A \in \mathbb{F}^{m\times n}$ we can decompose it into

$$A = U\Sigma V^* \in \mathbb{F}^{m\times n}$$

where

- **Singular Values**: Σ contains the *singular values* of A (or the square roots of the common positive eigenvalues of A^*A and AA^*)
- **Orthonormal basis of $\mathcal{R}(A)$:** $U \in \mathbb{F}^{m\times m}$ is such that $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ where columns $\mathcal{R}(U_1) = \mathcal{R}(A)$ and $\mathcal{R}(U_2) = \mathcal{N}(A^*)$
- **Orthonormal basis for $\mathcal{N}(A)$**: $V \in \mathbb{F}^{n\times n}$ is such that $V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ where $\mathcal{R}(V_1) = \mathcal{R}(A^*)$ and $\mathcal{R}(V_2) = \mathcal{N}(A)$.

2.2 Reminder of some basic classes matrices/operators

Definition. (*positive operator/positive semi-definite (PSD)*) An operator $A \in \mathcal{L}(V)$ is called positive if A is self-adjoint ($A = A^*$, i.e. Hermitian) and

$$\langle Av, v\rangle \geq 0,\ \forall v \in V$$

Example: if U is a subspace of V, then the orthogonal projection P_U is a positive operator.
Recall that a symmetric matrix has real eigenvalues:

Suppose $A = A^\top$, then λ is real.

Proof. suppose $Av = \lambda v$, $v \neq 0$, $v \in \mathbb{C}^n$. Then,

$$\bar{v}^\top Av = \bar{v}^\top(Av) = \lambda\bar{v}^\top v = \lambda\sum_i |v_i|^2$$

But, we also have that

$$\bar{v}^\top Av = \overline{(Av)}^\top v = \overline{(\lambda v)}^\top v = \bar{\lambda}\sum_i |v_i|^2$$

Hence, $\bar{\lambda} = \lambda$. □

Definition. (*square root*) An operator T is called the square root of an operator A if $T^2 = A$. Sometimes we write $T = A^{1/2} = \sqrt{A}$.

Fact. Square roots in general are not unique, however for PSD matrices, they are unique. That is, if A is a positive operator, then $\sqrt{A}$ denotes the unique positive square root of A.

How to compute? if $A \in \mathbb{R}^{n\times n}$ and it is diagonalizable, we can compute it via diagonalization. Then,

$$\sqrt{A} = VD^{1/2}V^{-1},\ \text{for}\ \ A = VDV^{-1}$$

Other ways include Jordan decomposition (which we will get to after the midterm).

Definition. (*unitary*) A matrix $U \in \mathbb{F}^{n\times n}$ is said to be unitary if and only if $U^*U = UU^* = I$. if $\mathbb{F} = \mathbb{R}$, then U is called orthogonal.

Unitary matrices preserve inner product and hence, length:

$$U \in \mathbb{F}^{n\times n} \implies \forall x, y \in \mathbb{F}^n,\ \langle Ux, Uy\rangle_{\mathbb{F}^n} = \langle x, y\rangle$$

Hence, $\|Ux\|_2 = \|x\|_2$.

Fact. Let $A \in \mathbb{F}^{m\times n}$ have rank r. Then, the Hermitian PSD matrices A^*A and AA^* are such that

$$r = \text{rank}(AA^*) = \text{rank}(A) = \text{rank}(A^*) = \text{rank}(A^*A)$$

Further, AA^* and A^*A has exactly r identical (strictly) positive eigenvalues $\sigma_i^2 > 0$, $i \in \{1, \ldots, r\}$.

DIY Exercise. Prove the above fact.

2.3 The Nitty Gritty

SVD is like an eigendecomposition. We have a rotation, a scaling and another rotation. Just the coordinates are changed. Indeed, if we have a diagonalizable matrix A, then

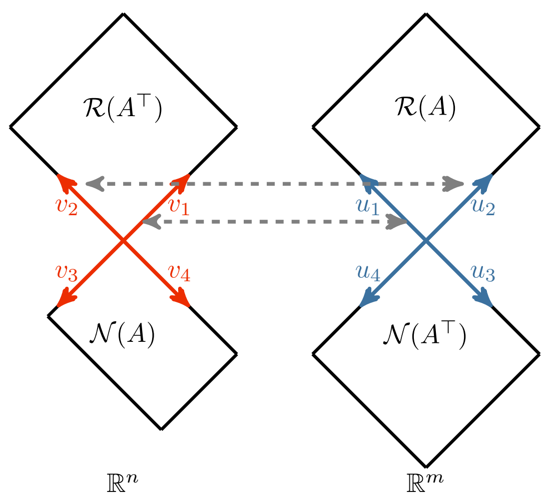
$$A = Q\lambda Q^\top$$

where $Q^\top = Q^{-1}$ for an orthonormal basis. (Recall from undergrad linear algebra that if λ_i are not distinct you can take any basis and Gram schmidt it to get an orthonormal basis). Then when we have

$$y = Ax$$

what we are doing is resolving x into q_i coordinates (rotate), then scaling by the λ_i's (dialate), then reconstituting with the basis q_i (rotate back). Note that you can write this as

$$A = \sum_{i=1}^{n} \lambda_i q_i q_i^\top$$



$$A^\top u_i = \frac{1}{\sigma_i} v_i \quad Av_i = \sigma_i u_i,\ i = 1, 2$$

$$Av_j = 0,\ A^\top u_j = 0,\ j = 3, 4$$

subspace	definition	dimension	basis
$\mathcal{R}(A) \subseteq \mathbb{R}^m$	column space (image) of A	r	first r columns of U
$\mathcal{N}(A^\top) \subseteq \mathbb{R}^m$	nullspace of $A^\top$	$m-r$	last $m-r$ columns of U
$\mathcal{R}(A^\top) \subseteq \mathbb{R}^n$	row space (image) of A	r	first r columns of V
$\mathcal{N}(A) \subseteq \mathbb{R}^n$	nullspace of A	$n-r$ last $n-r$	columns of V
$\mathcal{N}(A) \overset{\perp}{\oplus} \mathcal{R}(A^\top)$	$\mathbb{R}^n$	n	all n columns of V
$\mathcal{N}(A^\top) \overset{\perp}{\oplus} \mathcal{R}(A)$	$\mathbb{R}^m$	m	all m columns of U

Definition. (*singular values*) Suppose $A \in \mathcal{L}(V)$. The singular values of A are the eigenvalues of $\sqrt{A^*A}$, with each eigenvalue λ repeated $\dim(\mathcal{N}(\sqrt{A^*A} - \lambda I))$ times.

A coupled relevant facts:

- Recall: $\mathcal{N}(\sqrt{A^*A} - \lambda I)$ is the set of all eigenvectors of A corresponding to λ and the zero vector.
- A^*A is a positive operator for every $A \in \mathcal{L}(V)$, so that $\sqrt{A^*A}$ is well-defined.
- singular values of A are all non-negative because they are the eigenvalues of the positive operator $\sqrt{A^*A}$.

Theorem 2. Singular Value Decomposition. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Let $A \in \mathbb{F}^{m\times n}$ be a matrix of rank r. Then there exists matrices $U \in \mathbb{F}^{m\times m}$ and $V \in \mathbb{F}^{n\times n}$ and $\Sigma_1 \in \mathbb{R}^{r\times r}$ such that

a. $V \in \mathbb{F}^{n\times n}$ is a unitary matrix ($V^*V = I$) of the form

$$V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}, \ V_1 \in \mathbb{F}^{n\times r}$$

where the columns of V_1 form an orthonormal basis of $\mathcal{R}(A^*)$ (i.e., $\mathcal{R}(V_1) = \mathcal{R}(A^*)$), the columns of V_2 form an orthonormal basis of $\mathcal{N}(A)$, and the columns of V form a complete orthonormal basis of eigenvectors of A^*A.

b. $U \in \mathbb{F}^{m\times m}$ is a unitary matrix ($U^*U = I$) of the form

$$U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}, \ U_1 \in \mathbb{F}^{m\times r}$$

where the columns of U_1 form an orthonormal basis of $\mathcal{R}(A)$ (i.e., $\mathcal{R}(U_1) = \mathcal{R}(A)$), the columns of U_2 form an orthonormal basis of $\mathcal{N}(A^*)$, and the columns of U form a complete orthonormal basis of eigenvectors of AA^*.

c. Let $x \in \mathcal{R}(A^*)$ and $y \in \mathcal{R}(A)$ be represented by component vectors ξ and η respectively, according to the following scheme:

$$\mathbb{F}^r \to \mathcal{R}(A^*) : \xi \mapsto x = V_1\xi, \ \mathbb{F}^r \to \mathcal{R}(A) : \eta \mapsto y = U_1\eta$$

Then, the bijections induced by the orthogonal decomposition, viz.

$$A|_{\mathcal{R}(A^*)} : \mathcal{R}(A^*) \to \mathcal{R}(A), \ A^*|_{\mathcal{R}(A)} : \mathcal{R}(A) \to \mathcal{R}(A^*)$$

have representations

$$\Sigma_1 : \xi \mapsto \eta = \Sigma_1\xi, \ \Sigma_1 : \eta \mapsto \xi = \Sigma_1\eta$$

with

$$\Sigma_1 = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$$

such that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$$

where σ_i (i.e., the positive singular values of A) are the square roots of the common positive eigenvalues of A^*A and AA^*.

d. A has a dyadic expansion

$$A = U_1 \Sigma_1 V_1^* = \sum_{i=1}^{r} \sigma_i u_i v_i^*$$

where u_i, v_i are the columsn of U_1 and V_1, respecitvely.

e. $A \in \mathbb{F}^{m\times n}$ has singular value decomposition (SVD) given by

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

where

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

Recall our discussion of orthogonal projections. For simplicitly, suppose we have a subspace W of $\mathbb{F}^n$. Then we can find an orthogonal projection on to this space by doing the following:

$$x_p = \sum_{i=1}^{m} \langle x, w_i\rangle w_i$$

where $\{w_i\}$ is an orthonormal basis for W which is of dimension m.

Hence, we can rewrite the dyadic expansion to look like a projection:

$$Av = \sum_{i=1}^{r} \sigma_i u_i v_i^* v = \sum_{i=1}^{r} \sigma_i \langle v, v_i\rangle u_i$$

where we can think of writing the coordinate representation of v in terms of the vectors that span $\mathcal{R}(A^*)$:

$$\begin{bmatrix} \langle v, v_n\rangle \\ \vdots \\ \langle v, v_n\rangle \end{bmatrix} = \xi$$

Then, $\Sigma_1 : \xi \mapsto \eta$, so that the terms $\sigma_i\langle v, v_i\rangle = \eta_i$. Then we are left with

$$Av = \sum_{i=1}^{r} \eta_i u_i$$

where $\{u_i\}$ forms a basis for $\mathcal{R}(A)$.

Fact. Suppose $A \in \mathcal{L}(V)$. Then the singular values of A are the non-negative square roots of the eigenvalues of A^*A, with each eigenvalue λ repeated $\dim(\mathcal{N}(A^*A - \lambda I))$ times.

Note also that

$$A^\top A = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma^2 V^*$$

Hence, v_i are eigenvectors of $A^\top A$ (nonzero eigenvalues) since this is an eigendecomposition, $\sigma_i = \sqrt{\lambda_i(A^\top A)}$ and $\lambda_i(A^\top A) = 0$ for $i > r$), and $\|A\| = \sigma_1$.

Same for $AA^\top$:

$$AA^\top = (U\Sigma V^*)(U\Sigma V^*)^* = U\Sigma^2 U^*$$

so that u_i are eigenvectors of $AA^\top$, $\sigma_i = \sqrt{\lambda_i(AA^\top)}$.

2.4 Some geometric intuition

The matrices U and V are isometries so that they preserve length in what ever coordinate system. So the way they operate on their domain is simply to rotate. The operator Σ on the other hand is diagonal and hence only performs scaling of each of the coordinates. So basically the SVD breaks up any matrix operation into three peices: rotation, scaling, then rotation.

What does A do to the unit sphere? Consider for simplicity that $A \in \mathbb{F}^{n\times n}$ to be non-singular. Apply the unitary transformation $x = V\xi$ and $y = U\eta$ in the domain and codomain of A (here $\mathbb{F}^n = \mathcal{R}(V) = \mathcal{R}(A) = \mathcal{R}(U) = \mathcal{R}(A^*)$). Let

$$S_n = \{x \in \mathbb{F}^n |\ \|x\|_2 = 1\}$$

(i.e., S_n is the unit sphere in $\mathbb{F}^n$). Since V is unitary (i.e., it preserves length), we have that

$$S_n = \{\xi \in \mathbb{F}^n |\ \|\xi\|_2^2 = \sum_{i=1}^n |\xi_i|^2 = 1\}$$

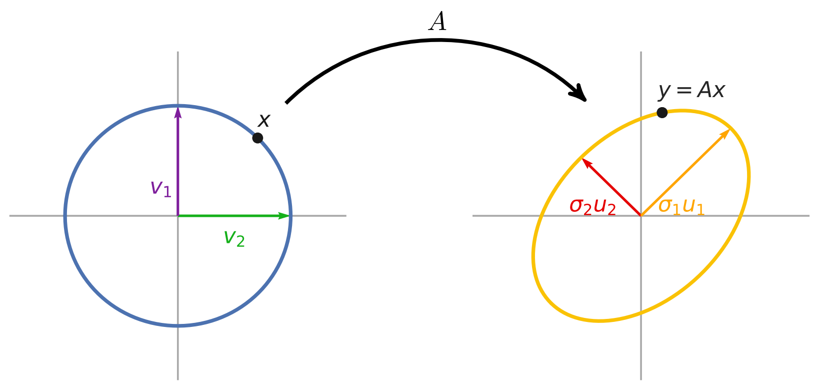
Let $A(S_n)$ denote the image of S_n under A. That is,

$$A(S_n) = \{y \in \mathbb{F}^n |]y = Ax,\ \|x\|_2 = 1\}$$

Hence, since U is unitary and by the SVD, i.e., $AV = U\Sigma$, we have that

$$A(S_n) = \left\{ y = U\eta \middle| \sum_{i=1}^n \left(\frac{|\eta_i|}{\sigma_i}\right)^2 = 1 \right\}$$

Note that column-wise, $Av_i = \sigma_i u_i$. Thus, $A(S_n)$, the image of the unit sphere S_n under A, is an *ellipsoid* whose principal axes are along the U_j's and are of length σ_i, the positive singular values of A. The action of A consists of, first, a "rotation" ($v_i \mapsto u_i$), then a scaling ($u_i \mapsto \sigma_i u_i$), i.e. the size of the action of A is dictated by its singular values.



3 Facts for me

- T is normal if $TT^* = T^*T$.
- Spectral theorem: T normal $\iff$ V has an orthonormal basis consisting of evecs of T $\iff$ T has a diangal with respect to some orthonormal basis of V

4 Steps for computing

1. Computer $AA^\top$. The eigenvectors of this operator make up the columns of U. So, find the eigepairs of $AA^\top$ in the usual way: first find the eigenvalues via solving the characteristic polynomial for its roots. Then find x in the null space of $AA^\top - \lambda_i I$ for each i:

$$(AA^\top - \lambda_i I)x = 0$$

   Orthonormalize them via Gram-Schmidt, e.g.
2. Eigenvectors of $A^\top A$ make up the columns of V.
3. Construct Σ by taking the common eigenvlaues of $AA^\top$ and $A^\top A$ and take the square root.

Fact. $AA^\top$ and $A^\top A$ have the same non-zero eigenvalues. Indeed, consider $\lambda \neq 0$, $\lambda \in \text{spec}(A^\top A)$. Then,

$$A^\top Ax = \lambda x \implies AA^\top(Ax) = \lambda(Ax) \implies AA^\top y = \lambda y$$

($x \notin \mathcal{N}(A)$ because otherwise $\lambda = 0$ since eigenvectors have to be non-trivial). Hence, $\lambda \in \text{spec}(AA^\top)$. Suppose $\lambda \in \text{spec}(AA^\top)$. Then,

$$AA^\top z = \lambda z \implies z^\top A^\top A = \lambda z^\top$$

or

$$AA^\top z = \lambda z \implies A^\top AA^\top z = \lambda A^\top z \implies A^\top Aw = \lambda w$$

Hence, $\lambda \in \text{spec}(A^\top A)$.

5 Connections to PCA

Recall the PCA discussion we had on lecture 1. The goal was to perform dimensionality reduction of a data set, either to ease interpretation or as a way to avoid overfitting and to prepare for subsequent analysis. Consider a sample covariance matrix from data matrix $X \in \mathbb{R}^{N\times p}$ which is defined by

$$S = \frac{X^\top X}{N}$$

since X has zero mean. Rows of X represent cases or observations and columns represent the variables of each observed unit.

If you do the eigendecomposition of $X^\top X$ you get

$$X^\top X = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma^* U^* U\Sigma V^* = V\Sigma^2 V^*$$

since $U^*U = I$ and the elements of Σ are real. Thus, if you have done the SVD then you already have the eigendecomposition for $X^\top X$.

The eigenvectors of $X^\top X$ (i.e., v_j, $j = 1, \ldots, p$) are called the principle component directions of X. The first principle component direction v_1 has the following properties:

1. v_1 is the eigenvector associated with the larges eigenvalue σ_1^2 of $X^\top X$.
2. $z_1 = Xv_1$ has the largest sample variance amongst all normalized linear combinations of the columns of X.
3. z_1 is called the first principle component of X and $\text{var}(x_1) = \sigma_1^2/N$.

The second principal component direction v_2 (the direction orthogonal to the first component that has the largest projected variance) is the eigenvector corresponding to the second largest eigenvalue, σ_2^2 of $X^\top X$ and so on. PCA projects the data along the directions where the data varies the most. The variance of the data along the principal component directions is associated with the magnitude of the eigenvalues.

6 Linear Systems of Equations and SVD

Solution of linear equations numerically difficult for matrices with 'bad condition'. Regular matrices in numeric approximation can be singular. **SVD helps with this issue.**

Generally in solving $Ax = b$, there are two possible sources of error:

1. intrinsic errors in A and b.
2. numerical errors due to rounding.

Solution $Ax = b$. Suppose that $A \in \mathbb{R}^{n\times n}$ for simplicity. Then,

$$b = Ax = U_1\Sigma_1 V_1^* x \implies U_1^* b = \Sigma_1 V_1^* x \implies x = V_1\Sigma_1^{-1}U_1^* b$$

if $\sigma_1 \geq \cdots \geq \sigma_r > 0$. If $b \in \mathcal{R}(A)$, then we get the exact solution. If $b \notin \mathcal{R}(A)$, then the above gives the closest possible solution in the sense that it minimizes $\|Ax - b\|$. Recall that The columns of U correspond to non-zero singular values of A are an orthonormal set of basis vectors for $\mathcal{R}(A)$. And, the columns of V that correspond to *zero* singular values form an orthonormal basis for $\mathcal{N}(A)$. So we are transforming b into its coordinate representation for the basis for $\mathcal{R}(A)$, then appropriately scaling it, then rotating it to a coordinate representation for the domain in terms of a orthonormal basis for $\mathcal{R}(A^*)$.

One of the important aspects of SVD is it allows us to assess the sensitivity of solutions of $Ax = b$ to sources of noise or error. Let $A \in \mathbb{F}^{n\times n}$ and suppose that $\det(A) \neq 0$, $b \neq 0$, and $A = U\Sigma V^*$. Then, $x_0 = A^{-1}b$ is the solution. Suppose that as a result of noisy measurements, round-off errors, etc. we only have an approximate value $A + \delta A$ for A and $b + \delta b$ for b. Then we have a perturbed system of equations

$$(A + \delta A)(x + \delta x_0) = b + \delta b$$

To better understand the error, we need to relate δx to that of δA and δb.

Recall our recent definitions of norms. Choose some norm $\|\cdot\|$ (all finite dimensional norms are equivalent in that what we can prove with one norm, we can prove with another). Suppose that

$$\|\delta A\| << \|A\|, \ \|\delta b\| << \|b\|$$

To better understand the error, lets compute an upper bound on

$$\frac{\|\delta x\|}{\|x\|}$$

Dropping higher order terms, we have the approximate equation

$$\delta A x_0 + A\delta x = \delta b$$

Compute δx, taking norms of both sides, and using properties of the induced norm, we have

$$\|\delta x\| \leq \|A^{-1}\|(\|\delta b\| + \|\delta A\|\|x_0\|)$$

Dividing by $\|x_0\| > 0$, and noting that $\|b\| \leq \|A\|\|x_0\|$, we have that

$$\frac{\|\delta x\|}{\|x_0\|} \leq \|A\|\|A^{-1}\| \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|} \right)$$

The number

$$\kappa(A) = \|A\|\|A^{-1}\|$$

is the condition number of A and it depends on the norm chosen. It is always greater than or equal to one.

Matrices for which $\kappa(A)$ is large are called ill-conditioned; for such matrices, the solution is very sensitive to some small changes in A and in b. Since all norms on $\mathbb{F}^n$ are equivalent, it can be shown that if a matrix is very ill-conditioned in some norm, it will also be ill-conditioned in any other norm.

So what does this mean for SVD? We know that

$$\sigma_1 = \max_i \sigma_i = \max\{\|Ax\|_2 |\ \|x\|_2 = 1\} = \|A\|_2$$

and

$$\sigma_n = \min_i \sigma_i = \min\{\|Ax\|_2 |\ \|x\|_2 = 1\}$$

so that

$$\|A^{-1}\|_2 = 1/\sigma_n$$

since $A^{-1} = V\Sigma^{-1}U^*$. We define the condition number (for the 2-norm) to be

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}$$

Hence if $\sigma_1 >> \sigma_n > 0$, for some δA and δb the resulting

$$\frac{\|\delta x\|_2}{\|x_0\|_2}$$

may be very large.

The smallest singular value σ_n of A is a measure of how far the nonsingular matrix A is from being singular.

7 Some Cool Visualization Tools/Tutorials with Applications

- Very simple introduction to svd with python:
  `https://lambein.xyz/blog/2018/12/20/svd-visualization.html`
- Nice interactive deep learning book with a chapter on SVD (may want to check out the other chapters, I really enjoyed them.): `https://github.com/hadrienj/deepLearningBook-Notes`

510: **Fall 2019**

Lecture 9: Least Squares and Regularized Least Squares

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

References: Chapter 6 of Calafiore and El Ghaoui; Chapter 12 of Boyd (linked on course site).

1 Review

So far we have already in some sense discussed solutions to linear equations. In particular, we discussed

- when solutions to $Ax = b$ exist and when they are unique in terms of surjectivity and injectivity of the map A
- least norm solutions via projections
- SVD, which is a solution technique and a way to assess sensitivity in terms of the singular values of a matrix (i.e., condition number).

Least squares is widely used as a method to construct a mathematical model from some data, experiments, or observations. Suppose we have an n-vector x, and a scalar y, and we believe that they are related, perhaps approximately, by some function

2 The Least Squares Problem

Consider the problem of finding $x \in \mathbb{R}^n$ such that

$$Ax = b, \quad A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m$$

There are several scenarios one can be in:

a. **under-determined**: $n < m$, i.e., there are more variables than equations. in general, here there are infinitely many solutions. it is also possible that there are no solutions because the equations are inconsistent.
b. **"square"**: $m = n$, i.e., the number of equations equals the number of variables.
c. **over-determined**: $m > n$, i.e., there are more equations than variables. The equations will be inconsistent in general.

In the latter case, we seek to find the solution x that solves

$$\min_x \|Ax - b\|_2^2$$

In the case of a., we can use least squares to assure that a solution exists, and to decide which of the many possible solutions to choose. Here we want to find the x that minimizes

$$\min_x \|Ax - b\|_2^2 + \mu\|x\|_2^2$$

This is a regularized least squares problem. We will come back to this. Alternatively, we can solve this problem via the method of Lagrange multipliers by mapping a constrained problem to minimization of a Lagrangian:

$$\left.\begin{array}{ll} \min\limits_x & \|x\|_2^2 \\ \text{s.t.} & Ax - b = 0 \end{array}\right\} \implies \min_x \|x\|_2^2 + \lambda^\top(Ax - b)$$

If the system is "square", then there is exactly one solution if A is full rank.

3 Solution Via Optimization/Calculcus

Consider

$$f(x) = \|Ax - b\|_2^2 = (Ax - b)^\top (Ax - b)$$

We know that the minimizer occurs at critical points where the derivative is zero:

$$\frac{\partial f}{\partial x_i}(x) = 0$$

The vector notation for this is

$$Df(x) = 0$$

3.1 A primer on derivatives

This should be review but in case it isnt I want to make sure you guys have seen it.

if you have a quadratic form $f(x) = \langle Ax, x\rangle = x^\top Ax$ over the real numbers, there are some simple rules for computing derivatives. The usual definition of $f'(x)$ when $f : \mathbb{R} \to \mathbb{R}$ is

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

Intuitively, f' represents the rate of change of f at the point x. In particular, starting at x, if you move a distance of h and want to figure out how much f changed, you would multiply the derivative $f'(x)$ by the amount of change, h, and add it to $f(x)$ which is where you started:

$$f(x+h) \simeq f(x) + \underbrace{f'(x)h}_{\text{linear approximation of } f \text{ at } x}$$

This can be generalized to functions $f : \mathbb{R}^m \to \mathbb{R}^n$. The idea behind a generalized derivative is that we still want a linear approximation to f at a point x, although now we allow x to be a vector and not just a real number. In this case, the derivative is the so called Jacobian. The approximation form still holds:

$$f(x+h) \simeq f(x) + \underbrace{J(x)}_{Df(x)} h$$

where

$$J(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

Suppose $f : \mathbb{R}^m \to \mathbb{R}$, then the Jacobian of f is simply one row:

$$J(x) = Df(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

so that

$$f(x+h) \simeq f(x) + \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_m \end{bmatrix}$$

The transpose of the *row vector* $J(x) = Df(x)$ is usually called the gradient of f at x, and is denoted by $\nabla f(x)$. This causes great confusion, since the matrix multiplication above no longer works if we transpose the Jacobian, so I am not an advocate.

Starting with the linear case,

$$f(x) = \langle x, b\rangle = b^\top x = \sum_{k=1}^{m} b_k x_k$$

the derivative is

$$Df(x) = b^\top$$

The reasoning behind this is as follows:

Now, generalizing a bit, if $f(x) = Ax$ where $A \in \mathbb{R}^{m\times m}$. Recall the classical notion of a derivative

$$f(x+h) \simeq f(x) + Df(x)h$$

If $f(x) = Ax$, then

$$f(x+h) = A(x+h) = Ax + Ah$$

Hence, setting this equal to our expanded version of f, we get

$$Ax + Ah = f(x) + Df(x)h$$

Subtracting $f(x) = Ax$, we get that $Ah = Df(x)h$ so that $Df(x) = A$.

Now, consider $f(x, y(x))$. How do we differentiate this? Well this is the total derivative. First, expand y:

$$y(x+h) \simeq y(x) + \frac{\partial y}{\partial x}h$$

We already know an expansion for f, hence we get

$$f(x+h, y(x+h)) \simeq f(x, y(x)) + \frac{\partial f}{\partial x}h + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x}h = f(x, y(x)) + \underbrace{\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x}\right)}_{Df(x)} h$$

Now, consider

$$f(x) = \langle x, Ax \rightarrow= x^\top Ax$$

Then, we can think of this as follows:

$$f(x, y(x)) = x^\top y(x),\ \ y(x) = Ax$$

We know that

$$Df(x, y(x)) = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}y'(x)$$

We know that

$$\frac{\partial f}{\partial x} = y(x)^\top$$

by above (i.e., its a linear function of x directly. Similarly,

$$\frac{\partial f}{\partial y} = y(x)^\top$$

Further,

$$y'(x) = A$$

Combining this we get

$$Df(x) = y(x)^\top + x^\top A = x^\top A^\top + x^\top A = x^\top(A^\top + A)$$

If $A = A^\top$, then

$$Df(x) = 2x^\top A^\top$$

So people often write this in its transposed form

$$\nabla f(x) = 2Ax$$

This is what is in the Boyd book for example. I am not fond of this notion of reasons noted above (its not covariant—i.e., its components do not change in the same way as changes to scale of the reference axes, very problematic in geometry).

3.2 Back to Least Squares

To minimize this problem, we set the derivative to zero:

$$Df(x) = 2(Ax-b)^\top A = 0 \implies x^\top A^\top A = b^\top A \implies x = \underbrace{(A^\top A)^{-1}A^\top}_{\text{pseudo-inverse}} b$$

Now, consider

$$L(x,\lambda) = \|x\|_2^2 + \lambda^\top(Ax-b) = \lambda^\top(Ax-b) + \lambda x^\top x$$

so that

$$\frac{\partial L}{\partial x}(x) = \lambda^\top A + 2x^\top = 0,\ \frac{\partial L}{\partial \lambda}L = x^\top A - b^\top = 0 \implies x = -A^\top\lambda/2,\ \lambda = -2(AA^\top)^{-1}b \implies x = A^\top(AA^\top)^{-1}b$$

Coming back to the regularized problem, we can similarly minimize

$$f(x) = \|Ax-b\|_2^2 + \mu\|x\|_2^2$$

to get

$$x_\mu = (A^\top A + \mu I)^{-1}A^\top b$$

Hence, the regularized solution converges to the least-norm solution as $\mu \to 0$. That is

$$(A^\top A + \mu I)^{-1}A^\top \to A^\top(AA^\top)^{-1}$$

for "fat" A as $\mu \to 0$. You will show this in your homework.

4 Residuals

Let's consider the overdetermined case and the corresponding problem

$$\min_x \|Ax-b\|_2^2$$

We call

$$r = Ax - b$$

the residual of the problem and

$$\|Ax-b\|_2^2 = r_1^2 + \cdots + r_m^2$$

This is the sum of squared residuals. Note that the least squares problem is often written as $Ax \simeq b$. That is, this is equivalent in some notations to

$$\min_x \|Ax-b\|_2^2$$

5 Let's talk about fat and skinny matrices

A key property of the two scenarios in `a.` and `c.` is the shape of the "data" matrix A. Indeed, if the system is under-determined, A is a skinny matrix. If it is full rank, then A has a left inverse given by

$$(A^\top A)^{-1}A^\top$$

If the system is over-determined then A is a fat matrix. If it is full rank, then A has a right inverse given by

$$A^\top(AA^\top)^{-1}$$

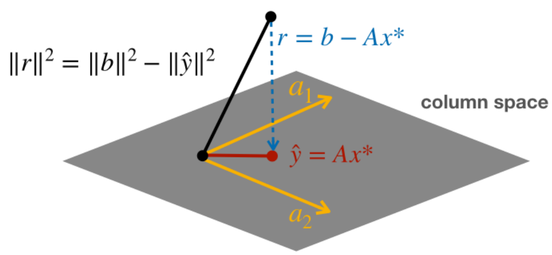
This operator should look familiar to you as you already showed on your exam that when A is surjective, $AA^\top$ is invertible, that $A^\top(AA^\top)^{-1}A$ is a projection onto $\mathcal{R}(A^\top)$, and finally that $I - A^\top(AA^\top)^{-1}A$ is a projection onto $\mathcal{N}(A)$. It turns out in the "skinny" case, $A(A^\top A)^{-1}A^\top$ is a projection onto $\mathcal{R}(A)$ and $I - A(A^\top A)^{-1}A^\top$ is a projection onto $\mathcal{N}(A^\top)$ (you should be able to prove this on your own now).

6 Intuition

Solving $Ax = b$ is an attempt to represent b in m-dimensional space with a linear combination of the n columns of A But those columns only give an n-dimensional plane inside the much larger m-dimensional space if $m > n$ as in the overdetermined case. The vector b is unlikely to lie in that plane, so $Ax = b$ is unlikely to be solvable. The vector Ax^* is the "nearest" to the point b in the plane. The error vector or residual vector r is orthogonal to the plane (i.e. the column space of A). That is

$$a_i^\top r = 0,\ \ i = 1, \dots, n \iff A^\top r = 0$$

where $r = b - Ax^*$.



The error vector $r = b - Ax^*$ is perpendicular to the column space:

$$\begin{bmatrix} a_1^\top \\ \vdots \\ a_n^\top \end{bmatrix} r = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

This geometric equation $A^\top r = 0$ finds x^*: the projection $\hat{y} = Ax^*$ (the combination of columns that is closest to b) It also gives us the so called normal equation for x^*:

$$A^\top r = A^\top (b - Ax^*) = 0 \implies A^\top A x^* = A^\top b$$

where

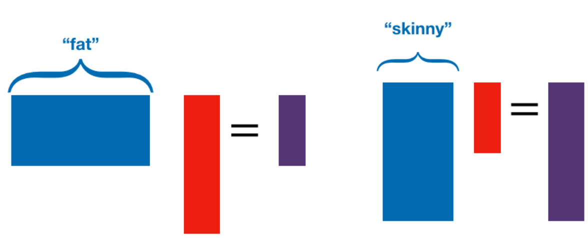
$$x^* = (A^\top A)^{-1} A^\top b$$

and

$$\hat{y} = Ax^* = \underbrace{A(A^\top A)^{-1}A^\top}_{P\text{: projection}} b = Pb$$

Here, $Ax = b$ has no solution but $Ax = \hat{y}$ has one solution x^*. That is measurements are inconsistent in $Ax = b$ but consistent in $Ax^* = \hat{y}$.

Now, $P = A(A^\top A)^{-1}A^\top$ is symmetric, $P^2 = P$ and $P \in \mathbb{R}^{m \times m}$ but has rank of N.



7 Example

Consider

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}, \; b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

The over-determined set of three equations in two variables $Ax = b$ has no solution. Indeed, from the first equation $2x_1 = 1$, we get that $x_1 = 1/2$ and from the last equation $2x_2 = -1$ we get that $x_2 = -1/2$ but this means the second equation $-x_1 + x_2 = 0$ doesn't hold.

Thus, to find an approximate solution we formulate the least squares problem:

$$\min(2x_1 - 1)^2 + (-x_1 + x_2)^2 + (2x_2 + 1)^2$$

If you find the solution, you will see that $x^* = (1/3, -1/3)$ and the residuals are

$$r = Ax^* - b = (-1/3, -2/3, 1/3)$$

This solution is the linear combination of the columns of A that is closest to b:

$$\frac{1}{3}\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -2/3 \\ -2/3 \end{bmatrix}$$

8 Least Squares Via SVD

Note: the singular vectors are real if A is real. SVD is a common tool for solving least squares. Indeed,

$$\begin{aligned} \|Ax - b\|_2^2 &= \|U\Sigma V^\top x - b\| \\ &= \|U^*(U\Sigma V^\top x - b)\|_2^2 \text{ since } U \text{ is orthogonal} \\ &= \|\Sigma V^\top x - U^* b\|_2^2 \\ &= \|\Sigma y - U^\top b\|_2^2 \end{aligned}$$

Now the goal is to find y that minimizes

$$\|\Sigma y - U^\top b\|_2^2 = \|\mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)y - z\|_2^2$$

So, pick

$$y_i = \begin{cases} \frac{z_i}{\sigma_i}, & \sigma_i \neq 0 \\ 0, & \sigma_i = 0 \end{cases}$$

Then, set $x = Vy$.

The SVD solution is the most stable computation of x^*. This is in part because we reduce the problem to only taking into consideration the singular values in the form of a singular matrix.

One of the important aspects of SVD is it allows us to assess the sensitivity of solutions of $Ax = b$ to sources of noise or error. Let $A \in \mathbb{F}^{n\times n}$ and suppose that $\det(A) \neq 0$, $b \neq 0$, and $A = U\Sigma V^*$. Then, $x_0 = A^{-1}b$ is the solution. Suppose that as a result of noisy measurements, round-off errors, etc. we only have an approximate value $A + \delta A$ for A and $b + \delta b$ for b. Then we have a perturbed system of equations

$$(A + \delta A)(x + \delta x_0) = b + \delta b$$

To better understand the error, we need to relate δx to that of δA and δb.

Recall our recent definitions of norms. Choose some norm $\|\cdot\|$ (all finite dimensional norms are equivalent in that what we can prove with one norm, we can prove with another). Suppose that

$$\|\delta A\| << \|A\|, \; \|\delta b\| << \|b\|$$

To better understand the error, lets compute an upper bound on

$$\frac{\|\delta x\|}{\|x\|}$$

Dropping higher order terms, we have the approximate equation

$$\delta A x_0 + A\delta x = \delta b$$

Compute δx, taking norms of both sides, and using properties of the induced norm, we have

$$\|\delta x\| \leq \|A^{-1}\|(\|\delta b\| + \|\delta A\|\|x_0\|)$$

Dividing by $\|x_0\| > 0$, and noting that $\|b\| \leq \|A\|\|x_0\|$, we have that

$$\frac{\|\delta x\|}{\|x_0\|} \leq \|A\|\|A^{-1}\| \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|}\right)$$

The number

$$\kappa(A) = \|A\|\|A^{-1}\|$$

is the condition number of A and it depends on the norm chosen. It is always greater than or equal to one.

Matrices for which $\kappa(A)$ is large are called ill-conditioned; for such matrices, the solution is very sensitive to some small changes in A and in b. Since all norms on $\mathbb{F}^n$ are equivalent, it can be shown that if a matrix is very ill-conditioned in some norm, it will also be ill-conditioned in any other norm.

So what does this mean for SVD? We know that

$$\sigma_1 = \max_i \sigma_i = \max\{\|Ax\|_2 | \ \|x\|_2 = 1\} = \|A\|_2$$

and

$$\sigma_n = \min_i \sigma_i = \min\{\|Ax\|_2 | \ \|x\|_2 = 1\}$$

so that

$$\|A^{-1}\|_2 = 1/\sigma_n$$

since $A^{-1} = V\Sigma^{-1}U^*$. We define the condition number (for the 2-norm) to be

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}$$

Hence if $\sigma_1 >> \sigma_n > 0$, for some δA and δb the resulting

$$\frac{\|\delta x\|_2}{\|x_0\|_2}$$

may be very large.

The smallest singular value σ_n of A is a measure of how far the nonsingular matrix A is from being singular.

9 Recursive Least Squares: A dynamical perspective

In off-line or batch identification, data up to some $t = N$ is first collected, then the model parameter vector $\hat{x}$ is computed.

In on-line or recursive identification, the model parameter vector $\hat{x}_t$ is required for every $t = n < N$.

Consider data $\{\phi_0, \ldots, \phi_{n-1}, y_1, \ldots, y_n\}$ and suppose that we believe

$$y_{k+1} \simeq \phi_k^\top \theta, \ k = 0, \ldots, n-1$$

Here ϕ_i and y_i are scalars. The least squares method as we have seen is to choose θ to minimize

$$V_n(\theta) = \sum_{k=0}^{n-1} (y_{k+1} - \phi_k^\top \theta)^2$$

One can show that by setting $\partial V_n(\theta)/\partial\theta = 0$, V_n is minimized by

$$\hat{\theta}_n = \left(\sum_{k=0}^{n-1} \phi_k \phi_k^\top \right)^{-1} \sum_{k=0}^{n-1} \phi_k y_{k+1}$$

assuming the inverse exists.

Let $Y_n = (y_1, \ldots, y_n)^\top$ and $X_n^\top = [\phi_0, \phi_1, \ldots, \phi_{n-1}]$. With this notation we can write

$$(X_n^\top X_n)\hat{\theta}_n = X_n^\top Y_n$$

and $\hat{\theta}_n$ above is the solution to this equation (check your self).

Moreover, every solution minimizes

$$\|Y_n - X_n\theta\|^2 = \sum_{k=0}^{n-1} (y_{k+1} - \phi_k^\top \theta)^2$$

This $\hat{\theta}_n$ is the LSE.

Note that the LSE does not require any probability structure; it is just the result of fitting a linear model to data in a certain way.

Let's see the recursive version. Suppose one more datum (y_{n+1}, ϕ_n) becomes available so that we now have $Y_{n+1} = (Y_n^\top, y_{n+1})^\top$ and $X_{n+1}^\top = [X_n^\top, \phi_n]$. Then we have a new LSE

$$\hat{\theta}_{n+1} = (X_{n+1}^\top X_{n+1})^{-1} X_{n+1}^\top Y_{n+1}$$

assuming again that the inverse exists. We can also obtain $\hat{\theta}_{n+1}$ from $\hat{\theta}_n$ as follows.

First, verify that the LSE can be written as

$$\hat{\theta}_{n+1} = \hat{\theta}_n + R_n^{-1}\phi_n(y_{n+1} - \phi_n^\top \hat{\theta}_n), \;\; R_{n+1} = R_n + \phi_n \phi_n^\top$$

510: **Fall 2019**

Lecture 10: Solutions to Autonomous ODEs

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

References: **Solutions to ODEs**: Chapter 3 **[C&D]**; **Jordan Form**: Chapter 4 **[C&D]**; Chapter 8.D **[Ax]**

1 Solutions to Autonomous LTI Systems

Recall from your basic ODE class that scalar ODE problems have solutions of the form:

$$\dot{x} = \frac{dx}{dt} = ax(t),\ x(t_0) = x_0 \implies x(t) = x_0 e^{a(t-t_0)}$$

1.1 Fundamental Theorem for ODEs

Consider the general ODE

$$\dot{x} = f(x,t)$$

The function f must satisfy two assumptions:

(A1) Let $\mathcal{D}$ be a set in $\mathbb{R}_+$ which contains at most a finite number of points per unit interval. $\mathcal{D}$ is the set of possible discontinuity points; it may be empty. For each fixed, $x \in \mathbb{R}^n$, the function $t \in \mathbb{R}_+ \backslash \mathcal{D} \to f(x,t) \in \mathbb{R}^n$ is *continuous* and for any $\tau \in \mathcal{D}$, the left-hand and right-hand limits $f(x, \tau_-)$ and $f(x, \tau_+)$, respectively, are finite vectors in $\mathbb{R}^n$.

(A2) There is a piecewise continuous function $k(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\|f(\xi,t) - f(\xi',t)\| \le k(t)\|\xi - \xi'\| \ \ \forall\, t \in \mathbb{R}_+,\ \forall \xi, \xi' \in \mathbb{R}^n$$

This is called a *global Lipschitz condition* because it must hold for all ξ and ξ'.

Theorem. (Fundamental Theorem of Existence and Uniqueness.) Consider

$$\dot{x}(t) = f(x,t)$$

where initial condition (t_0, x_0) is such that $x(t_0) = x_0$. Suppose f satisfies (A1) and (A2). Then,

1. For each $(t_0, x_0) \in \mathbb{R}_+ \times \mathbb{R}^n$ there exists a continuous function $\phi : \mathbb{R}_+ \to \mathbb{R}^n$ such that

$$\phi(t_0) = x_0$$

   and

$$\dot{\phi}(t) = f(\phi(t), t), \quad \forall t \in \mathbb{R}_+ \backslash \mathcal{D}$$

2. This function is unique. The function ϕ is called the solution through (t_0, x_0) of the differential equation.

1.2 Applying it to Linear Time Varying Systems

Recall

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) \quad \text{(state DE)} \\ y(t) &= C(t)x(t) + D(t)u(t) \quad \text{(read-out eqn.)} \end{aligned}$$

with initial data (t_0, x_0) and the assumptions on $A(\cdot), B(\cdot), C(\cdot), D(\cdot), u(\cdot)$ all being PC:

- $A(t) \in \mathbb{R}^{n \times n}$
- $B(t) \in \mathbb{R}^{n \times m}$
- $C(t) \in \mathbb{R}^{p \times n}$
- $D(t) \in \mathbb{R}^{p \times m}$

The input function $u(\cdot) \in \mathcal{U}$, where $\mathcal{U}$ is the set of piecewise continuous functions from $\mathbb{R}_+ \to \mathbb{R}^m$.
This system satisfies the assumptions of our existence and uniqueness theorem. Indeed,

1. For all fixed $x \in \mathbb{R}^n$, the function $t \in \mathbb{R}_+ \backslash \mathcal{D} \to f(x,t) \in \mathbb{R}^n$ is continuous where $\mathcal{D}$ contains all the points of discontinuity of $A(\cdot), B(\cdot), C(\cdot), D(\cdot), u(\cdot)$
2. There is a PC function $k(\cdot) = \|A(\cdot)\|$ such that

$$\|f(\xi,t) - f(\xi',t)\| = \|A(t)(\xi - \xi')\| \le k(t)\|\xi - \xi'\| \quad \forall t \in \mathbb{R}_+,\ \forall \xi, \xi' \in \mathbb{R}^n$$

Hence, by the above theorem, the differential equation has a unique continuous solution $x : \mathbb{R}_+ \to \mathbb{R}^n$ which is clearly defined by the parameters $(t_0, x_0, u) \in \mathbb{R}_+ \times \mathbb{R}^n \times U$.

Theorem. (Existence of the state transition map/flow.) Under the assumptions and notation above, for every triple $(t_0, x_0, u) \in \mathbb{R}_+ \times \mathbb{R}^n \times U$, the state transition map

$$x(\cdot) = \xi(\cdot, t_0, x_0, u) : \mathbb{R}_+ \to \mathbb{R}^n$$

is a continuous map well-defined as the unique solution of the state differential equation

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)$$

with (t_0, x_0) such that $x(t_0) = x_0$ and $u(\cdot) \in U$.

1.3 Zero-State and Zero-Input Maps

The state transition function of a linear system is equal to its *zero-input* state transition function and its *zero-state* state transition map:

$$\xi(t, t_0, x_0, u) \quad = \quad \underbrace{\xi(t, t_0, x_0, 0)}_{\text{zero-input state trans. func.}} \quad + \quad \underbrace{\xi(t, t_0, 0, u)}_{\text{zero-state state trans. func.}}$$

Due to the fact that the state transition map and the response map are *linear*, they have the property that for fixed $(t, t_0) \in \mathbb{R}_+ \times \mathbb{R}_+$ the maps

$$\xi(t, t_0, \cdot, 0) : \mathbb{R}^n \to \mathbb{R}^n : x_0 \mapsto \xi(t, t_0, x_0, 0)$$

Hence by the *Matrix Representation Theorem*, they are representable by matrices. Therefore there exists a matrix $\Phi(t, t_0) \in \mathbb{R}^{n \times n}$ such that

$$\xi(t, t_0, x_0, 0) = \Phi(t, t_0)x_0, \quad \forall x_0 \in \mathbb{R}^n$$

(State transition matrix.) $\Phi(t, t_0)$ is called the *state transition matrix*.

Consider the matrix differential equation

$$\dot{X} = A(t)X, \quad X(\cdot) \in \mathbb{R}^{n \times n}$$

Let $X(t_0) = X_0$.

The state transition matrix $\Phi(t, t_0)$ is defined to be the solution of the above matrix differential equation starting from $\Phi(t_0, t_0) = I$. That is,

$$\frac{\partial}{\partial t}\Phi(t, t_0) = A(t)\Phi(t, t_0)$$

and $\Phi(t_0, t_0) = I$.

1.4 State Transition Matrix

Definition. (LTI State transition matrix.) The state transition matrix for

$$\dot{x} = Ax, \ x \in \mathbb{R}^n, \ A \in \mathbb{R}^{n \times n}$$

is the *matrix exponential* e^{At} defined to be

$$e^{At} = I + \frac{At}{1!} + \frac{A^2t^2}{2!} + \cdots$$

where I is the $n \times n$ identity matrix.

Proof. It is easy to verify that

$$\Phi(t, 0) = e^{At} \ \text{ and } \ \Phi(t, t_0) = e^{A(t-t_0)}$$

by checking that

$$x(t) = e^{A(t-t_0)}x_0$$

satisfies the differential equation

$$\dot{x} = Ax, \quad x(t_0) = x_0$$

Indeed, by the fact that $\Phi(t, t_0)$ satisfies the ODE (by definition) we know that

$$\frac{\partial}{\partial t}\Phi(t, t_0) = A\Phi(t, t_0)$$

and

$$\frac{\partial}{\partial t}\exp(A(t-t_0))x_0 = A\exp(A(t-t_0))x_0$$

In addition, $\Phi(t_0, t_0) = I$ and $\exp(A(t_0 - t_0)) = I$. Hence, $\Phi(t, t_0)$ and $\exp(A(t - t_0))$ satisfy the same ODE so they are equal. □

Hence, the matrix exponential is our friend and we need computational approaches for expressing this little monster.

2 The Matrix Exponential

First, we note that the matrix exponential has several important properties.

- $e^0 = I$
- $e^{A(t+s)} = e^{At}e^{As}$
- $e^{(A+B)t} = e^{At}e^{Bt} \iff AB = BA$
- $(e^{At})^{-1} = e^{-At}$
- $\frac{d}{dt}e^{At} = Ae^{At} = e^{At} \cdot A$
- Let $z(t) \in \mathbb{R}^{n \times n}$. Then the solution to

  $$\dot{z}(t) = Az(t)$$

  with $z(0) = I$ is

  $$z(t) = e^{At}$$

Recall that

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

This is also true for the matrix exponential—i.e.

$$\exp(At) = \sum_{k=0}^{\infty} \frac{(At)^k}{k!}$$

Fact. Note also that Cayley-Hamilton implies that the matrix exponential is expressible as a polynomial of order $n-1$!

Using the series representation of e^{At} to compute e^{At} is difficult unless, e.g., the matrix A is ***nilpotent*** in which case the series yields a closed form solution.

Definition. (Nilpotent) A nilpotent matrix is such that $A^k = 0$ for some k.

Example. Consider

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

Then

$$A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

so that

$$e^{At} = I + At = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

Hence we need an alternative method to compute it.

2.1 Review of Laplace

Definition. (Laplace Transform)

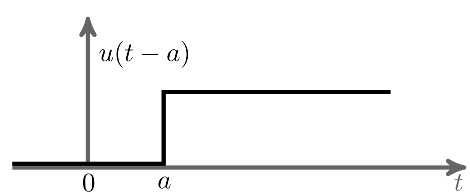
$$\mathcal{L}f(t) = \int_0^\infty f(t)e^{-st}\ dt$$

The Laplace transform has the following properties:

- Linearity:

$$\mathcal{L}(af(t) + bg(t)) = a\underbrace{F(s)}_{\mathcal{L}f(t)} + b\underbrace{G(s)}_{\mathcal{L}g(t)}$$

- Time Delay: Let u be a step function.



$$f(t-a)u(t-a) \xrightarrow{\mathcal{L}} e^{-as}F(s)$$

- First derivative (technically should be $t = 0^-$):

$$\mathcal{L}\dot{f}(t) = sF(s) - f(0)$$

- Integration:

$$\mathcal{L}\left(\int_{0^-}^\infty f(\tau)\ d\tau\right) = \frac{F(s)}{s}$$

2.2 Computation of e^{At} via Laplace

Use the Laplace transform of $\dot{X} = AX$, $X \in \mathbb{R}^{n\times n}$, $X(0) = I$:

$$s\hat{X}(s) - X(0) = A\hat{X}(s)$$

so that

$$s\hat{X}(s) - A\hat{X}(s) = I \implies \hat{X}(s) = (sI - A)^{-1}$$

We know (from property 6) that $X(t) = e^{At}$ so that

$$e^{At} = X(t) = \mathcal{L}^{-1}(\hat{X}(s)) = \mathcal{L}^{-1}\left((sI - A)^{-1}\right)$$

Example. Consider the same example above with

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

so that

$$(sI - A) = \begin{bmatrix} s & -1 \\ 0 & s \end{bmatrix}$$

Recall that the inverse of a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Now,

$$(sI - A)^{-1} = \frac{1}{s^2}\begin{bmatrix} s & 1 \\ 0 & s \end{bmatrix} = \begin{bmatrix} \frac{1}{s} & \frac{1}{s^2} \\ 0 & \frac{1}{s} \end{bmatrix}$$

where we recall that $\mathcal{L}(f(t)) = \frac{F(s)}{s}$ so that $\mathcal{L}(1) = \frac{1}{s}$; it is also easy to show that the ramp function transforms to $\frac{1}{s^2}$. Hence,

$$e^{At} = \mathcal{L}^{-1}\begin{bmatrix} \frac{1}{s} & \frac{1}{s^2} \\ 0 & \frac{1}{s} \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

Example. Consider

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

Then

$$(sI - A)^{-1} = \begin{bmatrix} s-1 & -1 \\ 0 & s-1 \end{bmatrix}^{-1} = \frac{1}{(s-1)^2}\begin{bmatrix} s-1 & 1 \\ 0 & s-1 \end{bmatrix}$$

where we recall that $\mathcal{L}e^{at} = \frac{1}{s-a}$, $s > a$ which can be verified by direct integration. Hence,

$$e^{At} = \mathcal{L}^{-1}\begin{bmatrix} \frac{1}{s-1} & \frac{1}{(s-1)^2} \\ 0 & \frac{1}{s-1} \end{bmatrix} = \begin{bmatrix} e^t & te^t \\ 0 & e^t \end{bmatrix}$$

3 Computing the Matrix Exponential

Computation of the matrix exponential is important for expressing the solution of a autonomous or controlled linear time invariant dynamical system. So we need ways to compute it that are tractable.

4 Distinct Eigenvalues

If matrix $A \in \mathbb{R}^{n\times n}$ (or $\in \mathbb{C}^{n\times n}$) has m distinct eigenvalues ($\lambda_i \neq \lambda_j$, $i \neq j$) then it has (at least) m linearly independent eigenvectors.

If all eigenvalues of A are distinct then A is diagonalizable.

Q: do you remember what diagonalizable means?

Diagonalizable. An $n \times n$ matrix A is diagonalizable iff the sum of the dimensions of its eigenspaces is n—aka there exists a matrix P such that

$$A = P\Lambda P^{-1}$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$,

$$P = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

with

$$Av_i = \lambda_i v_i$$

(i.e. col vectors of P are right eigenvectors of A)

Proof. Proof of Prop. 4 (By contradiction) Assume λ_i, $i \in \{1, \dots, m\}$ are distinct and v_i, $i = 1, \dots, m$ are linearly dependent. That is, there exists α_i such that

$$\sum_{i=1}^{m} \alpha_i v_i = 0$$

where all α_i are not zero. We can assume w.l.o.g that $\alpha_1 \neq 0$. Multiplying on the left by $(\lambda_m I - A)$,

$$0 = (\lambda_m I - A) \sum_{i=1}^{m} \alpha_i v_i = (\lambda_m I - A) \sum_{i=1}^{m-1} \alpha_i v_i + \alpha_m(\lambda_m I - A)v_m = \sum_{i=1}^{m-1} \alpha_i(\lambda_m - \lambda_i)v_i$$

since $Av_i = \lambda_i v_i$. Then multiply by $(\lambda_{m-1} I - A)$ to get that

$$0 = (\lambda_{m-1} I - A) \sum_{i=1}^{m-1} \alpha_i(\lambda_m - \lambda_i)v_i = \sum_{i=1}^{m-2} \alpha_i(\lambda_{m-1} - \lambda_i)(\lambda_m - \lambda_i)v_i = 0$$

Repeatedly multiply by $(\lambda_{m-k} I - A)$, $k = 2, \dots, m-2$ to obtain

$$\alpha \prod_{i=2}^{m} (\lambda_i - \lambda_1)v_i = 0$$

As $\lambda_1 \neq \lambda_i$, $i = 2, \dots, m$, the above implies that $\alpha_1 = 0$ which is a contradiction. □

For each $n \times n$ complex matrix A, define the exponential of A to be the matrix

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

It is not difficult to show that this sum converges for all complex matrices A of any finite dimension. But we will not prove this here.

If A is a 1×1 matrix $[t]$, then $e^A = [e^t]$, by the Maclaurin series formula for the function $y = e^t$. More generally, if D is a diagonal matrix having diagonal entries $d_1, d_2, \dots, d_n$, then we have

$$e^D = I + D + \frac{1}{2!}D^2 + \cdots = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} + \text{diag}(d_1, \dots, d_n) + \text{diag}\left(\frac{d_1^2}{2!}, \frac{d_1^2}{2!}, \dots, \frac{d_1^2}{2!}\right) = \text{diag}\left(e^{d_1}, \dots, e^{d_n}\right)$$

The situation is more complicated for matrices that are not diagonal. However, if a matrix A happens to be diagonalizable, there is a simple algorithm for computing e^A, a consequence of the following lemma.

Let A and P be complex $n \times n$ matrices, and suppose that P is invertible. Then

$$\exp(P^{-1}AP) = P^{-1}\exp(A)P$$

Proof. Recall that, for all integers $m \geq 0$, we have $(P^{-1}AP)^m = P^{-1}A^mP$. The definition for exponential then yields

$$\begin{aligned}
\exp(P^{-1}AP) &= I + P^{-1}AP + \frac{1}{2!}(P^{-1}AP)^2 + \cdots \\
&= I + P^{-1}AP + \frac{1}{2!}P^{-1}A^2P + \cdots \\
&= P^{-1}\left(I + A + \frac{A^2}{2!} + \cdots\right)P \\
&= P^{-1}\exp(A)P
\end{aligned}$$

If a matrix A is diagonalizable, then there exists an invertible P so that $A = PDP^{-1}$, where D is a diagonal matrix of eigenvalues of A, and P is a matrix having eigenvectors of A as its columns. In this case, $e^A = Pe^DP^{-1}$. ◻

Let A denote the matrix

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

You can asily verify that 4 and 3 are eigenvalues of A, with corresponding eigenvectors

$$w_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{and} \quad w_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

It follows that

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

so that

$$\exp(A) = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} e^4 & 0 \\ 0 & e^3 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 2e^4 - e^3 & e^4 - e^3 \\ 2e^3 - 2e^4 & 2e^3 - e^4 \end{bmatrix}$$

The definition of the exponential as a sum immediately reveals many other familiar properties. The following proposition is easy to prove:

Let $A \in \mathbb{C}^{n\times n}$.

1. If 0 denotes the zero matrix, then $e^0 = I$.
2. $A^m e^A = e^A A^m$ for all integers m
3. $(e^A)^T = e^{(A^T)}$
4. If $AB = BA$ then $Ae^B = e^BA$ and $e^Ae^B = e^Be^A$.

Unfortunately not all familiar properties of the scalar exponential function $y = e^t$ carry over to the matrix exponential. For example, we know from calculus that $e^{s+t} = e^se^t$ when s and t are numbers. However this is often not true for exponentials of matrices. In other words, it is possible to have $n \times n$ matrices A and B such that $e^{A+B} \neq e^Ae^B$. Exactly when we have equality, $e^{A+B} = e^Ae^B$, depends on specific properties of the matrices A and B. What do you think they are?

Let A and B be complex $n \times n$ matrices. If $AB = BA$ then $e^{A+B} = e^Ae^B$.

Proof. **DIY exercise** ◻

5 Generalized Eigenvectors

Last time we talked about the case when A had distinct eigenvalues and we said you could simply diagonalize as $A = P\Lambda P^{-1}$ and then write

$$\exp(At) = P\text{diag}(\exp(\lambda_1 t), \dots, \exp(\lambda_n t))P^{-1}$$

Question: What about when A is not diagonalizable?
First, some preliminaries. Consider a vector space (V, F) and a linear map $\mathcal{A} : V \to V$.

Definition. (Invariant Subspaces.) A subspace $M \subset V$ is said to be A–invariant or invariant under A if given $x \in M$, $Ax \in M$. This is often written as $A[M] \subset M$ or even $AM \subset M$.

Example.
1. $\mathcal{N}(A)$ is A–invariant.
2. $\mathcal{R}(A)$ is A–invariant.
3. $\mathcal{N}(A - \lambda_i I)$ where $\lambda_i \in \sigma(A)$ is A–invariant.
4. If
$$p(A) = A^k + \alpha_1 A^{k-1} + \cdots + \alpha_{k-1}A + \alpha_k I$$
then, $\mathcal{N}(p(A))$ is A–invariant.
5. Let the subspaces M_1 and M_2 be A–invariant. Let
$$M_1 + M_2 = \{x \in V :\ x = x_1 + x_2,\ x_i \in M_i \text{ for } i = 1, 2\}$$
Then, $M_1 \cap M_2$ and $M_1 + M_2$ are A–invariant.

Definition. (Generalized Eigenvectors) Suppose λ is an eigenvalue of the square matrix A. We say that v is a generalised eigenvector of A with eigenvalue λ, if v is a nonzero element of the null space of $(A - \lambda I)^j$—i.e. $\mathcal{N}(A - \lambda I)^j$—for some positive integer j.

Fact. Null spaces eventually stabilize—that is, the null spaces $\mathcal{N}(A - \lambda I)^j$ are increasing with j and there is a unique positive integer k such that $\mathcal{N}(A - \lambda I)^j = \mathcal{N}(A - \lambda I)^k$ for all $j \geq k$.

Definition. (Generalized Eigenspace) Consider $A \in \mathbb{F}^{n\times n}$ with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$. Define the generalized eigenspace pertaining to λ_i by

$$E_{\lambda_i} = \{x \in \mathbb{C}^n |\ (A - \lambda_i I)^n x = 0\}$$

Intuition: Observe that all the eigenvectors pertaining to λ_i are in E_{λ_i}. For a given E_{λ_i}, we can interpret the spaces in a hierarchical viewpoint. We know that E_{λ_i} contains all the eigenvectors pertaining to λ_i. Call these eigenvectors the **first order generalized eigenvectors**. If the span of these is not equal to E_{λ_i}, then there must be a vector $x \in E_{\lambda_i}$ for which $y = (A - \lambda_i I)^2 x = 0$ but $(A - \lambda_i I)x \neq 0$. That is to say y is an eigenvector of A pertaining to λ_i. Call such vectors **second order generalized eigenvectors**. In general, we call an $x \in E_{\lambda_i}$ a **generalized eigenvector of order** p if $y = (A - \lambda_i I)^p x = 0$ but $(A - \lambda_i I)^{p-1} x \neq 0$. For this reason we will call E_{λ_i} the space of generalized eigenvectors.

Fact. Let $A \in \mathbb{C}^{n\times n}$ with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$ and invariant subspaces E_{λ_i}, $i \in \{1, \dots, k\}$.
1. Let $x \in E_{\lambda_i}$ be a generalized eigenvector of order p. Then the vectors
$$x, (A - \lambda_i I)x, (A - \lambda_i I)^2 x, \dots, (A - \lambda_i I)^{p-1} x \tag{1}$$
are linearly independent.
2. The subspace of $\mathbb{C}^n$ generated by the above vectors is an invariant subspace of A.

Example. Consider

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

- $$\chi_A(\lambda) = (\lambda - 3)(\lambda - 1)^2$$
- eigenvalues: $\lambda = 1, 3$
- eigenvectors:
$$\begin{aligned} \lambda_1 = 3: &\qquad v_1 = (1,2,2) \\ \lambda_2 = 1: &\qquad v_2 = (1,0,0) \end{aligned}$$
- The last generalized eigenvector will be a vector $v_3 \neq 0$ such that
$$(A - \lambda_2 I)^2 v_3 = 0$$
but
$$(A - \lambda_2 I)v_3 \neq 0$$
Pick $v_3 = (0,1,0)$. Note that $(A - \lambda_2 I)v_3 = v_2$.

Tip. How many powers of $(A-\lambda I)$ do we need to compute in order to find all of the generalized eigenvectors for λ?

If A is an $n \times n$ matrix and λ is an eigenvalue with algebraic multiplicity k, then the set of generalized eigenvectors for λ consists of the nonzero elements of $\mathcal{N}(A - \lambda I)^k$. In other words, we need to take at most k powers of $A - \lambda I$ to find all of the generalized eigenvectors for λ.

Yet another example. Determine generalized eigenvectors for the matrix
$$A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \\ 0 & -1 & 1 \end{bmatrix}$$

- single eigenvalue of $\lambda = 1$
- single eigenvector $v_1 = (-2, 0, 1)$
- now we look at
$$(A - I)^2 = \begin{bmatrix} 2 & 0 & 4 \\ 0 & 0 & 0 \\ -1 & 0 & -2 \end{bmatrix}$$
to get generalized eigenvector $v_2 = (0,1,0)$.
- Finally, $(A - I)^3 = 0$ so that $v_3 = (1,0,0)$.

6 Jordan Normal Form

To get some intuition for why we can find a form that looks like the Jordan form (i.e. a block diagonal decomposition) let us recall the following result.

First, recall the definition of the direct sum of subspaces:

Definition. (Direct Sum.) V is the direct sum of $M_1, M_2, \ldots, M_k$, denoted as
$$V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$$
if for all $x \in V$, $\exists!\ x_i \in M_i$, $i = 1, \ldots, k$ such that
$$x = x_1 + x_2 + \cdots + x_k$$

Fact. The direct sum is the generalization of linear independence; e.g., check that if $V = M_1 \oplus \cdots \oplus M_k$, then $M_i \cap M_j = \{0\}$.

Theorem. (Second Representation Theorem.) Let $A : V \to V$ be a linear map. Let $V = M_1 \oplus M_2$ where $\dim V = n$, $\dim M_1 = k$, and $\dim M_2 = n - k$ be a finite dimensional vector space. If M_1 is A–invariant, then A has a representation of the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

where $A_{11} \in F^{k\times k}$, $A_{12} \in F^{k\times(n-k)}$, $A_{22} \in F^{(n-k)\times(n-k)}$. Moreover, if both M_1 and M_2 are A–invariant then

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$$

Proof. Let $\{b_1, b_2, \ldots, b_k\}$ be a basis for M_1 and let $\{b_{k+1}, \ldots, b_n\}$ be a basis for M_2. By assumption $V = M_1 \oplus M_2$ so that $\{b_i\}_{i=1}^n$ is a basis for V and any $x \in V$ has a unique representation as

$$x = \sum_{i=1}^{n} \xi_i b_i$$

Moreover, A has a matrix representation $A = (a_{ij})$ dictated by

$$Ab_j = \sum_{i=1}^{n} a_{ij} b_i \quad \forall\ j \tag{2}$$

Now for all $j = 1, \ldots, k$, $b_j \in M_1$ which is A–invariant so that $Ab_j \in M_1$ with basis $\{b_i\}_{i=1}^k$. Thus by (2), for all $j \in \{1, \ldots, k\}$

$$Ab_j = \sum_{i=1}^{k} a_{ij} b_i$$

i.e. $A_{ij} = 0$ for all $i = k + 1, \ldots, n$, for all $j = 1, \ldots, k$.

□

Essentially what this is saying is that since M_1 is A–invariant, if I apply A to a basis vector in M_1 it has to stay in M_1 so any vector $x \in M_1$ written as $x = \sum_{i=1}^k \xi_i b_i$ is such that $Ax \in M_1$ with $Ax = \sum_{i=1}^k \eta_i b_i$ and no non-zero basis vectors are coming from the basis of M_2.

Why useful? We can use the second representation theorem applied to

$$\mathbb{C}^n = \mathbb{N}(A - \lambda_1 I)^{m_1} \oplus \mathbb{N}(A - \lambda_2 I)^{m_2} \oplus \cdots \oplus \mathbb{N}(A - \lambda_p I)^{m_p}$$

to write A via similarity transform into a matrix that has 'nice structure' (Jordan blocks) so that with respect to this structure e^{At} is easy to compute.

We are also going to use this result quite a bit in terms of decomposition of controllable and observable subspaces. So keep it in your pocket.

6.1 Minimal Polynomial

In order to show this decomposition, we need to revisit the characteristic polynomial and its cousin the minimal polynomial.

We know that

$$\det(sI - A) = \chi_A(s) \quad \text{(characteristic polynomial)}$$

We can write

$$\chi_A(s) = (s - \lambda_1)^{d_1}(s - \lambda_2)^{d_2} \cdots (s - \lambda_p)^{d_p}$$

where $d_1, \ldots, d_p$ are the *multiplicities* of $\lambda_1, \ldots, \lambda_p \in \mathbb{C}$, respectively and

$$\sum_{i=1}^{p} d_i = n$$

By Cayley-Hamilton, we know that

$$\chi_A(A) = 0_{n\times n}$$

Let $\psi_A(s)$ be the polynomial of least degree such that

$$\psi_A(A) = 0_{n\times n}$$

Definition. (Minimal Polynomial.) Given a matrix $A \in \mathcal{C}^{n\times n}$, we call *minimal polynomial* of A the annihilating polynomial ψ of least degree. The minimal polynomial is of the form

$$\psi_A(s) = (s-\lambda_1)^{m_1}\cdots(s-\lambda_p)^{m_p}$$

for some integers $m_i \le d_i$.

Proposition. $\psi_A(s)$ divides $\chi_A(s)$

That is,

$$\frac{\chi_A(s)}{\psi_A(s)} = q(s)$$

for some polynomial $q(s)$.

Example.

1. Consider

$$A_1 = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix}$$

   Then,

$$\chi_A(s) = (s-\lambda_1)^2(s-\lambda_2) \text{ and } \psi_A(s) = (s-\lambda_1)(s-\lambda_2)$$

2. Consider

$$A_2 = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{bmatrix}$$

   Then,

$$\chi_A(s) = (s-\lambda_1)^3 \text{ and } \psi_A(s) = (s-\lambda_1)^2$$
$$\psi_A(A) = (A-\lambda_1 I)(A-\lambda_1 I) = A^2 - 2\lambda_1 A + \lambda_1^2 I = 0$$

   where

$$A^2 = \begin{bmatrix} \lambda_1^2 & 2\lambda_1 & 0 \\ 0 & \lambda_1^2 & 0 \\ 0 & 0 & \lambda_1^2 \end{bmatrix}$$

3. Consider

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

   Then

$$\chi_A(s) = (s-3)^2(s-4)$$

   The minimal polynomial is either

$$(s-3)(s-4)$$

   or

$$(s-3)^2(s-4)$$

   It cannot be the former since $(A-3I)(A-4I) \ne 0$.

That is, $\psi_A(s)$ is the least degree polynomial such that $\psi_A(A) = 0$.

Theorem (Decomposition)

$$\mathbb{C}^n = \mathcal{N}(A - \lambda_1 I)^{m_1} \oplus \mathcal{N}(A - \lambda_2 I)^{m_2} \oplus \cdots \oplus \mathbb{N}(A - \lambda_p I)^{m_p}$$

Proof.

$$\frac{1}{\psi_A(s)} = \frac{1}{(s-\lambda_1)_1^m \cdots (s-\lambda_p)^{m_p}} = \frac{n_1(s)}{(s-\lambda_1)^{m_1}} + \cdots + \frac{n_p(s)}{(s-\lambda_p)^{m_p}}$$

so that

$$1 = n_1(s)q_1(s) + \cdots n_p(s)q_p(s)$$

where

$$q_i(s) = \frac{\psi_A(s)}{(s-\lambda_i)^{m_i}}$$

Thus,

$$I = n_1(A)q_1(A) + \cdots + n_p(A)q_p(A)$$

so that

$$x = \underbrace{n_1(A)q_1(A)}_{x_1} x + \cdots + \underbrace{n_p(A)q_p(A)}_{x_1} x$$

which in turn implies that

$$x_i = n_i(A)q_i(A) = n_i(A)\frac{\psi_A(A)}{(A-\lambda_i I)^{m_i}}$$

so that

$$(A - \lambda_i I)^{m_i} x_i = 0_n \quad \Longrightarrow \quad x_i \in \mathbb{N}(A - \lambda_i I)_i^m$$

To show the decomposition is unique, we argue by contradiction. Let

$$x_i \in \mathbb{N}(A - \lambda_i I)^{m_i}$$

so that

$$x_1 + \cdots + x_p = 0_n$$

and wlog assume $x_1 \neq 0$. Then

$$x_1 = -x_2 - x_3 - \cdots - x_p$$

so that

$$(A - \lambda_2 I)^{m_2} \cdots (A - \lambda_p I)^{m_p} x_1 = 0_n$$

But $q_1(s)$ and $(s - \lambda_1)^{m_1}$ are *co-prime* meaning that

$$h_1(s)q_1(s) + h_2(s)(s-\lambda_1)^{m_1} = 1$$

This implies that

$$h_1(A)\underbrace{q_1(A)x_1}_{0} + h_2(A)\underbrace{(A-\lambda_1 I)^{m_1} x_1}_{0} = x_1 \quad \Longrightarrow \quad x_1 = 0 \;\; \rightarrow\leftarrow$$

□

Definition. (Multiplicities)

1. The geometric multiplicity of an eigenvalue λ is the dimension of E_λ.
2. The algebraic multiplicity of an eigenvalue λ is the number of times λ appears as a root of $\chi_A(\lambda)$.

Note. In general, the algebraic multiplicity and geometric multiplicity of an eigenvalue can differ. However, the geometric multiplicity can never exceed the algebraic multiplicity.

Fact. If for every eigenvalue of A, the geometric multiplicity equals the algebraic multiplicity, then A is said to be diagonalizable.

If the minimal polynomial is

$$\psi_A(\lambda) = \prod_{i=1}^{k}(\lambda - \lambda_i)^{m_i}$$

with $1 \leq m_i \leq d_i$ and d_i the algebraic multiplicity, then

$$N_i = \mathcal{N}((A - \lambda_i I)^{m_i})$$

is the algebraic eigenspace and $\mathcal{N}(A-\lambda_i I)$ is the geometric eigenspace with $\dim(N_i)$ the algebraic multiplicity and $\dim(\mathcal{N}(A - \lambda_i I))$ the geometric multiplicity.

Proposition. $\dim \mathbb{N}(A - \lambda_i I)^{m_i} = d_i$

Proof. see C& D, p.115 □

6.2 Jordan Form Details

Definition. (Jordan Block.) Let $\lambda \in \mathbb{C}$. A Jordan block $J_k(\lambda)$ is a $k \times k$ upper triangular matrix of the form

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix}$$

A Jordan matrix is any matrix of the form

$$J = \operatorname{diag}(J_{n_1}(\lambda_1), \ldots, J_{n_k}(\lambda_k))$$

where the matrices J_{n_1} are Jordan blocks. If $J \in \mathbb{C}^{n\times n}$, then $n_1 + n_2 + \cdots + n_k = n$.

Recall that

$$\chi_A(s) = \det(sI - A) = (s - \lambda_1)^{n_1} \cdots (s - \lambda_k)^{n_k}$$

When eigenvalues are distinct, $n_i = 1$ so that A is diagonalizable.

Theorem. (semisimple system) A square complex $n \times n$ matrix is *semisimple* if and only if there exists a nonsingular complex $n \times n$ matrix T^{-1} and diagonal complex $n \times n$ matrix Λ for which

$$A = T^{-1}\Lambda T$$

or equivalently

$$\Lambda = TAT^{-1}$$

The columns $e_i \in \mathbb{C}^n$ of T^{-1} organized as

$$T^{-1} = \begin{bmatrix} | & | & & | \\ e_1 & e_2 & \cdots & e_n \\ | & | & & | \end{bmatrix} \in \mathbb{C}^{n\times n}$$

and the diagonal entries $\lambda_i \in \mathbb{C}$ of Λ organized as

$$\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{C}^{n\times n}$$

may be taken as n eigenvectors associated according to (1) with n eigenvalues λ_i of A that form a *spectral list*.

Note: We call *spectral list* of A any n–tuple $(\lambda_i)_{i=1}^n$ of eigenvalues that is complete as a list of roots of the characteristic polynomial χ_A
In other words, $A \in \mathbb{C}^{n\times n}$ is semisimple iff A is *diagonalizable by a similarity transformation.*

Example 1. Modal Decomposition. Consider

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned}$$

Define $z = Tx$. Then

$$\begin{aligned} \dot{z} &= TAT^{-1}z + TBu \\ y &= CT^{-1}z \end{aligned}$$

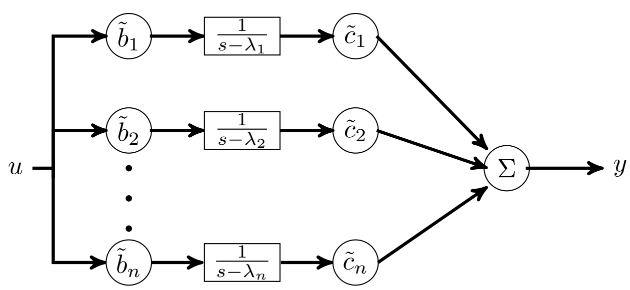
Now consider the case where we have a single input/single output (SISO, e.g., $m = 1 = p$). Define

$$TB = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \vdots \\ \tilde{b}_n \end{bmatrix}, \; CT^{-1} = \begin{bmatrix} \tilde{c}_1 & \tilde{c}_2 & \cdots & \tilde{c}_n \end{bmatrix}$$

Then

$$C(sI - A)^{-1}b = \frac{\tilde{c}_1\tilde{b}_1}{s - \lambda_1} + \cdots + \frac{\tilde{c}_n\tilde{b}_n}{s - \lambda_n} = \sum_{i=1}^{n} \frac{\tilde{c}_i\tilde{b}_i}{s - \lambda_i}$$

which is called the *modal decomposition.*



Note: $c(sI - A)^{-1}b$ is called the transfer function. If $\tilde{c}_i$ or $\tilde{b}_i$ is zero, then the transfer function does not contain the term $1/(s - \lambda_i)$.

Theorem. (Jordan Normal Form.) Let $A \in \mathbb{C}^{n\times n}$. Then there is a non-singular matrix P such that

$$A = P\text{diag}(J_{n_1}(\lambda_1), \ldots, J_{n_k}(\lambda_k))P^{-1} = PJP^{-1}$$

where $J_{n_i}(\lambda_i)$ is a Jordan block and $\sum_{i=1}^k n_i = n$. The Jordan form J is unique up to permutations of the blocks. The eigenvalues $\lambda_1, \ldots, \lambda_k$ are not necessarily distinct. If A is real with real eigenvalues, then P can be taken as real.

Suppose that

$$P^{-1}AP = J = \text{diag}(J_1, \ldots, J_k)$$

where $J_i = J_{n_i}(\lambda_i)$. Express P as

$$P = \begin{bmatrix} P_1 & \cdots & P_k \end{bmatrix}$$

where $P_i \in \mathbb{C}^{n\times n_i}$ are the columns of P associated with i–th Jordan block J_i. We have that

$$AP_i = P_iJ_i$$

Let $P_i = [v_{i1}\ v_{i2}\ \cdots\ v_{in_i}]$ so that

$$Av_{i1} = \lambda_i v_{i1}$$

that is, the first column of each P_i is an eigenvector associated with eigenvalue λ_i. For $j = 2, \ldots, n_i$,

$$Av_{ij} = v_{ij-1} + \lambda_i v_{ij}$$

These $v_{i1}, \ldots, v_{in_i}$ are the generalized eigenvectors.

Example. Let A be an n by n square matrix. If its characteristic equation $\chi_A(t) = 0$ has a repeated root then A may not be diagonalizable, so we need the Jordan Canonical Form. Suppose λ is an eigenvalue of A, with multiplicity r as a root of $\chi_A(t) = 0$. The vector v is an eigenvector with eigenvalue λ if $Av = \lambda v$ or equivalently

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

The trouble is that this equation may have fewer than r linearly independent solutions for v. So we generalize and say that v is a generalized eigenvector with eigenvalue λ if

$$(A - \lambda I)^k v = 0$$

for some positive k. Now one can prove that there are exactly r linearly independent generalized eigenvectors. Finding the Jordan form is now a matter of sorting these generalized eigenvectors into an appropriate order.

To find the Jordan form carry out the following procedure for each eigenvalue λ of A.

1. First solve $(A - \lambda I)v = 0$, counting the number r_1 of linearly independent solutions.
2. If $r_1 = r$ good, otherwise $r_1 < r$ and we must now solve
$$(A - \lambda I)^2 v = 0.$$
There will be r_2 linearly independent solutions where $r_2 > r_1$.
3. If $r_2 = r$ good, otherwise solve
$$(A - \lambda I)^3 v = 0$$
to get $r_3 > r_2$ linearly independent solutions, and so on.
4. Eventually one gets $r_1 < r_2 < \cdots < r_{N-1} < r_N = r$.

   Fact. The number N is the size of the largest Jordan block associated with λ, and r_1 is the total number of Jordan blocks associated to λ. If we define $s_1 = r_1$, $s_2 = r_2 - r_1$, $s_3 = r_3 - r_2$, $\ldots$, $s_N = r_N - r_{N-1}$ then s_k is the number of Jordan blocks of size at least k by k associated to λ.

5. Finally put $m_1 = s_1 - s_2$, $m_2 = s_2 - s_3$, $\ldots$, $m_{N-1} = s_{N-1} - s_N$ and $m_N = s_N$. Then m_k is the number of k by k Jordan blocks associated to λ. Once we've done this for all eigenvalues then we've got the Jordan form!

To find P such that $J = P^{-1}AP$ we need to do more work. We do the following for each eigenvalue λ:

1. First find the Jordan block sizes associated to λ by the above process. Put them in decreasing order
$$N_1 \geq N_2 \geq \cdots \geq N_k.$$
2. Now find a vector $v_{1,1}$ such that $(A - \lambda I)^{N_1} v_{1,1} = 0$ but $(A - \lambda I)^{N_1 - 1} v_{1,1} \neq 0$.
3. Define $v_{1,2} = (A - \lambda I)v_{1,1}$, $v_{1,3} = (A - \lambda I)v_{1,2}$ and so on until we get v_{1,N_1}. We can't go further because $(A - \lambda I)v_{1,N_1} = 0$.
4. If we only have one block we're OK, otherwise we can find $v_{2,1}$ such that $(A - \lambda I)^{N_2} v_{2,1} = 0$, and $(A - \lambda I)^{N_2 - 1} v_{2,1} \neq 0$ and AND $v_{2,1}$ is not linearly dependent on $v_{1,1}, \ldots, v_{1,N_1}$.
5. Define $v_{2,2} = (A - \lambda I)v_{2,1}$ etc.
6. keep going if you have more blocks, otherwise you will have r linearly independent vectors $v_{1,1}, \ldots, v_{k,N_k}$. Let
$$P_\lambda = \begin{bmatrix} v_{k,N_k} & \cdots & v_{1,1} \end{bmatrix}$$
be the n by r matrix.
7. do this for all λ's then stack the P_λ's up horizontally to get P

6.3 Functions of a matrix

Definition. (Functions of a matrix.) Let $\hat{f}(s)$ be any function of s analytic on the spectrum of A and $\hat{p}(s)$ be a polynomial such that

$$\hat{f}^k(\lambda_\ell) = \hat{p}^k(\lambda_\ell)$$

for $0 \leq k \leq m_\ell - 1$ and $1 \leq \ell \leq \sigma$. Then

$$\hat{f}(A) = \hat{p}(A)$$

In fact, if $m = \sum_{i=1}^{\sigma} m_i$ then

$$\hat{p}(s) = a_1 s^{m-1} + a_2 s^{m-2} + \cdots + a_m s^{\sigma}$$

where $a_1, a_2, \ldots, a_n$ are functions of

$$(\hat{f}(\lambda_1), \hat{f}^1(\lambda_1), \hat{f}^2(\lambda_1), \ldots, \hat{f}^{m_1}(\lambda_1), \hat{f}(\lambda_2), \ldots)$$

and hence

$$\hat{f}(A) = a_1 A^{m-1} + \cdots + a_m A^0 = \sum_{\ell=1}^{\sigma} \sum_{k=0}^{m_\ell - 1} p_{k\ell}(A) f^k(\lambda_\ell)$$

where $p_{k\ell}$'s are polynomials independent of f.

Example. Define

$$J_2(\lambda, \varepsilon) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda + \varepsilon \end{bmatrix}$$

with eigenvalues $\lambda_1 = \lambda$ and $\lambda_2 = \lambda + \varepsilon$. For any $\varepsilon \neq 0$, $J_2(\lambda, \varepsilon)$ is diagonalizable. Computing eigenvector

$$\begin{aligned} [\lambda_1 I - J_2(\lambda, \varepsilon)] v_1 &= \begin{bmatrix} 0 & -1 \\ 0 & -\varepsilon \end{bmatrix} v_1 = 0 \implies v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \big[\lambda_2 I - J_2(\lambda, \varepsilon)\big] v_2 &= \begin{bmatrix} \varepsilon & -1 \\ 0 & 0 \end{bmatrix} v_2 = 0 \implies v_1 = \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix} \end{aligned}$$

and

$$T = \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix}, \; T^{-1} = \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 1/\varepsilon \end{bmatrix}$$

we can evaluate

$$f(J_2(\lambda, \varepsilon)) = T f(\Lambda) T^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix} \begin{bmatrix} f(\lambda) & 0 \\ 0 & f(\lambda + \varepsilon) \end{bmatrix} \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 1/\varepsilon \end{bmatrix} = \begin{bmatrix} f(\lambda) & (f(\lambda + \varepsilon) - f(\lambda))/\varepsilon \\ 0 & f(\lambda + \varepsilon) \end{bmatrix}$$

As $J_2(\lambda, \varepsilon) \to J_2(\lambda)$ as $\varepsilon \to 0$ and f is continuous, if f is also differentiable at λ

$$f(J_2(\lambda, \varepsilon)) = \lim_{\varepsilon \to 0} f(J_2(\lambda, \varepsilon)) = \begin{bmatrix} f(\lambda) & f'(\lambda) \\ 0 & f(\lambda) \end{bmatrix}$$

6.4 Functions of a matrix (repeated eigenvalues)

Theorem. (General Form of $f(A)$) Let $A \in \mathbb{C}^{n \times n}$ have a minimal polynomial ψ_A given by

$$\psi_A(s) = \prod_{k=1}^{\sigma} (s - \lambda_k)^{m_k}$$

Let the domain Δ contain $\sigma(A)$, then for any analytic function $f : \Delta \to \mathbb{C}$. we have

$$f(A) = \sum_{k=1}^{\sigma} \sum_{\ell=0}^{m_k - 1} f^{(\ell)}(\lambda_k) p_{k\ell}(A)$$

where $p_{k\ell}$'s are polynomials independent of f.

Consider

$$J = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix} \in F^{n \times n}$$

Claim:

$$f(J) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ 0 & \ddots & \ddots & \cdots \\ \vdots & \ddots & \ddots & f^1(\lambda) \\ 0 & \cdots & \cdots & f(\lambda) \end{bmatrix}$$

Proof. the minimum polynomial is $(s-\lambda)^n$. Thus,

$$f(J) = \sum_{\ell=0}^{n-1} f^{(\ell)}(\lambda) p_\ell(J)$$

Choose

$$f_1(s) = 1 \implies f_1(J) = I = f_1^{(0)} p_0(J) \implies p_0(J) = I$$
$$f_2(s) = s - \lambda \implies f_2(J) = J - \lambda I = f_2^{(1)}(\lambda) p_1(J) \implies p_1(J) = J - \lambda I$$
$$f_3(s) = (s-\lambda)^2 \implies f_3(J) = (J-\lambda I)^2 = f_3^{(2)}(\lambda) p_2(J) \implies 2p_2(J) = (J - \lambda I)^2$$

Hence

$$\begin{aligned} p_0(J) &= I \\ p_1(J) &= J - \lambda I \\ p_2(J) &= \frac{1}{2}(J-\lambda I)^2 \end{aligned}$$

Thus,

$$f(J) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2} & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \ddots & \ddots & \frac{f''(\lambda)}{2} \\ \vdots & \cdots & \ddots & \ddots & f'(\lambda) \\ 0 & \cdots & \cdots & \cdots & f(\lambda) \end{bmatrix}$$

□

Hence we have

Theorem. (Spectral Mapping Theorem.)

$$\sigma(f(J)) = f(\sigma(J)) = \{f(\lambda), f(\lambda), \ldots, f(\lambda)\}$$

and more generally that

$$\sigma(f(A)) = f(\sigma(A))$$

Example.

$$e^{J_i(\lambda t)} = \begin{bmatrix} e^{\lambda t} & te^{\lambda t} & \frac{t^2}{2!}e^{\lambda t} & \cdots & \frac{t^{i-1}}{(i-1)!}e^{\lambda t} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{t^2}{2!}e^{\lambda t} \\ & & & \ddots & te^{\lambda t} \\ 0 & & & e^{\lambda t} & \end{bmatrix}$$

What does this mean? Well more generally if we had

$$J = \operatorname{diag}\left\{ \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix}, \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix}, \begin{bmatrix} \lambda_2 & 1 \\ 0 & \lambda_2 \end{bmatrix}, \lambda_2 \right\}$$

Recall that this Jordan form my be obtained from A by the similarity transform

$$J = TAT^{-1}$$

where

$$T^{-1} = \begin{bmatrix} e_1 & v_1 & w_1 & e_2 & v_2 & e_3 & v_3 & e_4 \end{bmatrix}$$

where $e_1, \dots, e_4$ are eigenvectors and the rest are generalized eigenvectors. Then

$$f(A) = f(T^{-1}JT) = T^{-1}f(J)T$$

where

$$f(J) = \operatorname{diag}\left\{ \begin{bmatrix} f(\lambda_1) & f'(\lambda_1) & \frac{f''(\lambda_1)}{2} \\ 0 & f(\lambda_1) & f'(\lambda_1) \\ 0 & 0 & f(\lambda_1) \end{bmatrix}, \begin{bmatrix} f(\lambda_1) & f'(\lambda_1) \\ 0 & f(\lambda_1) \end{bmatrix}, \begin{bmatrix} f(\lambda_2) & f'(\lambda_2) \\ 0 & f(\lambda_2) \end{bmatrix}, f(\lambda_2) \right\}$$

Example. Compute e^{At} for

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

Eigenvalues:

$$\det(\lambda I - A) = \left| \begin{bmatrix} \lambda & -1 \\ 1 & \lambda_2 \end{bmatrix} \right| = \lambda^2 + 2\lambda + 1 = (\lambda+1)^2 = 0 \implies \lambda_1 = \lambda_2 = -1$$

Eigenvector:

$$(1I - A)v_1 = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} v_1 = 0, \implies v_1 = \begin{bmatrix} 1 \\ 01 \end{bmatrix}$$

Generalized eigenvector:

$$v_{11} = v_1,\ (A - 1I)v_{12} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} v_{12} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} v_{11}, \implies v_{12} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Jordan form:

$$T = \begin{bmatrix} v_{11} & v_{12} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix},\ T^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix},\ J = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

Matrix exponential:

$$e^{At} = Te^{J_2(-t)}T^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} e^{-t} & te^{-t} \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} (t+1)e^{-t} & te^{-t} \\ -te^{-t} & (1-t)e^{-t} \end{bmatrix}$$

510: **Fall 2019**

Lecture 11: Jordan Form and Functions of Matrix

Lecturer: L.J. Ratliff

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

References: **Jordan Form**: Chapter 4 **[C&D]**; Chapter 8.D **[Ax]**

1 Generalized Eigenvectors

Last time we talked about the case when A had distinct eigenvalues and we said you could simply diagonalize as $A = P\Lambda P^{-1}$ and then write

$$\exp(At) = P\text{diag}(\exp(\lambda_1 t), \ldots, \exp(\lambda_n t))P^{-1}$$

Definition. (Multiplicities)

1. The *geometric multiplicity* of an eigenvalue λ is the dimension of E_λ.
2. The *algebraic multiplicity* of an eigenvalue λ is the number of times λ appears as a root of $\chi_A(\lambda)$.

In general, the algebraic multiplicity and geometric multiplicity of an eigenvalue can differ. However, the geometric multiplicity can never exceed the algebraic multiplicity.

Fact. If for every eigenvalue of A, the geometric multiplicity equals the algebraic multiplicity, then A is said to be diagonalizable.

Question: What about when A is not diagonalizable?

Definition. (Generalized Eigenvectors) Suppose λ is an eigenvalue of the square matrix A. We say that v is a generalised eigenvector of A with eigenvalue λ, if v is a nonzero element of the null space of $(A - \lambda I)^j$—i.e. $\mathcal{N}(A - \lambda I)^j$—for some positive integer j.

Fact. Null spaces eventually stabilize—that is, the null spaces $\mathcal{N}(A-\lambda I)^j$ are increasing with j and there is a unique positive integer k such that $\mathcal{N}(A - \lambda I)^j = \mathcal{N}(A - \lambda I)^k$ for all $j \geq k$.

Fact. The number of generalized eigenvectors associated to an eigenvalue λ_i is equal to the difference between the algebraic multiplicity d_i and the geometric multiplicity m_i; indeed,

$$\dim(\mathcal{N}(A - \lambda_i I)) = m_i \implies \exists\ m_i \text{ L.I. vectors spanning } \mathcal{N}(A - \lambda_i I)$$

and d_i is the number of times λ_i is a root of the characteristic polynomial. So, there is at least $d_i - m_i$ generalized eigenvectors. Further, if $d_i = m_i$, then there are no generalized eigenvectors.

Definition. (Generalized Eigenspace) Consider $A \in \mathbb{F}^{n\times n}$ with spectrum $\sigma(A) = \{\lambda_1, \ldots, \lambda_k\}$. Define the generalized eigenspace pertaining to λ_i by

$$E_{\lambda_i} = \{x \in \mathbb{C}^n | \ (A - \lambda_i I)^n x = 0\}$$

Intuition: Observe that all the eigenvectors pertaining to λ_i are in E_{λ_i}. For a given E_{λ_i}, we can interpret the spaces in a hierarchical viewpoint.

a. We know that E_{λ_i} contains all the eigenvectors pertaining to λ_i. Call these eigenvectors the **first order generalized eigenvectors**.

b. If the span of these is not equal to E_{λ_i}, then there must be a vector $x \in E_{\lambda_i}$ for which $y = (A - \lambda_i I)^2 x = 0$ but $(A - \lambda_i I)x \neq 0$. That is to say y is an eigenvector of A pertaining to λ_i. Call such vectors **second order generalized eigenvectors**.

c. In general, we call an $x \in E_{\lambda_i}$ a **generalized eigenvector of order** p if $y = (A - \lambda_i I)^p x = 0$ but $(A - \lambda_i I)^{p-1} x \neq 0$.

For this reason we will call E_{λ_i} the space of generalized eigenvectors.

Fact. Let $A \in \mathbb{C}^{n \times n}$ with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$ and invariant subspaces E_{λ_i}, $i \in \{1, \dots, k\}$.

1. Let $x \in E_{\lambda_i}$ be a generalized eigenvector of order p. Then the vectors

$$x, (A - \lambda_i I)x, (A - \lambda_i I)^2 x, \dots, (A - \lambda_i I)^{p-1} x \tag{1}$$

are linearly independent.

2. The subspace of $\mathbb{C}^n$ generated by the above vectors is an invariant subspace of A.

Example. Consider

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

- $$\chi_A(\lambda) = (\lambda - 3)(\lambda - 1)^2$$
- eigenvalues: $\lambda = 1, 3$
- eigenvectors:

$$\begin{aligned} \lambda_1 = 3: &\quad v_1 = (1, 2, 2) \\ \lambda_2 = 1: &\quad v_2 = (1, 0, 0) \end{aligned}$$

- The last generalized eigenvector will be a vector $v_3 \neq 0$ such that

$$(A - \lambda_2 I)^2 v_3 = 0$$

but

$$(A - \lambda_2 I) v_3 \neq 0$$

Pick $v_3 = (0, 1, 0)$. Note that $(A - \lambda_2 I) v_3 = v_2$.

Tip on computation. Q. How many powers of $(A - \lambda I)$ do we need to compute in order to find all of the generalized eigenvectors for λ?

A. At most d, where d is the algebraic multiplicity of λ. Indeed, if A is an $n \times n$ matrix and λ is an eigenvalue with algebraic multiplicity d, then the set of generalized eigenvectors for λ consists of the nonzero elements of $\mathcal{N}(A - \lambda I)^d$. In other words, we need to take at most d powers of $A - \lambda I$ to find all of the generalized eigenvectors for λ.

Example. Determine generalized eigenvectors for the matrix

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

- single eigenvalue of $\lambda = 1$
- single eigenvector $v_1 = (-2, 0, 1)$
- now we look at

$$(A - I)^2 = \begin{bmatrix} 2 & 0 & 4 \\ 0 & 0 & 0 \\ -1 & 0 & -2 \end{bmatrix}$$

to get generalized eigenvector $v_2 = (0, 1, 0)$.

- Finally, $(A - I)^3 = 0$ so that $v_3 = (1, 0, 0)$.

2 Jordan Normal Form

To get some intuition for why we can find a form that looks like the Jordan form (i.e. a block diagonal decomposition) let us recall the following result.

First, recall the definition of the direct sum of subspaces:

Definition. (Direct Sum.) V is the direct sum of $M_1, M_2, \ldots, M_k$, denoted as

$$V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$$

if for all $x \in V$, $\exists!\ x_i \in M_i$, $i = 1, \ldots, k$ such that

$$x = x_1 + x_2 + \cdots + x_k$$

Fact. The direct sum is the generalization of linear independence; e.g., check that if $V = M_1 \oplus \cdots \oplus M_k$, then $M_i \cap M_j = \{0\}$.

Theorem. (Second Representation Theorem.) Let $A : V \to V$ be a linear map. Let $V = M_1 \oplus M_2$ where $\dim V = n$, $\dim M_1 = k$, and $\dim M_2 = n - k$ be a finite dimensional vector space. If M_1 is A–invariant, then A has a representation of the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

where $A_{11} \in F^{k\times k}$, $A_{12} \in F^{k\times(n-k)}$, $A_{22} \in F^{(n-k)\times(n-k)}$. Moreover, if both M_1 and M_2 are A–invariant then

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$$

Proof. Let $\{b_1, b_2, \ldots, b_k\}$ be a basis for M_1 and let $\{b_{k+1}, \ldots, b_n\}$ be a basis for M_2. By assumption $V = M_1 \oplus M_2$ so that $\{b_i\}_{i=1}^n$ is a basis for V and any $x \in V$ has a unique representation as

$$x = \sum_{i=1}^{n} \xi_i b_i$$

Moreover, A has a matrix representation $A = (a_{ij})$ dictated by

$$Ab_j = \sum_{i=1}^{n} a_{ij} b_i \quad \forall\ j \tag{2}$$

Now for all $j = 1, \ldots, k$, $b_j \in M_1$ which is A–invariant so that $Ab_j \in M_1$ with basis $\{b_i\}_{i=1}^k$. Thus by (2), for all $j \in \{1, \ldots, k\}$

$$Ab_j = \sum_{i=1}^{k} a_{ij} b_i$$

i.e. $A_{ij} = 0$ for all $i = k + 1, \ldots, n$, for all $j = 1, \ldots, k$.

□

Essentially what this is saying is that since M_1 is A–invariant, if I apply A to a basis vector in M_1 it has to stay in M_1 so any vector $x \in M_1$ written as $x = \sum_{i=1}^k \xi_i b_i$ is such that $Ax \in M_1$ with $Ax = \sum_{i=1}^k \eta_i b_i$ and no non-zero basis vectors are coming from the basis of M_2.

Why useful? We can use the second representation theorem applied to

$$\mathbb{C}^n = \mathcal{N}(A-\lambda_1 I)^{m_1} \oplus \mathcal{N}(A-\lambda_2 I)^{m_2} \oplus \cdots \oplus \mathcal{N}(A-\lambda_p I)^{m_p}$$

to write A via similarity transform into a matrix that has 'nice structure' (Jordan blocks) so that with respect to this structure e^{At} is easy to compute.

We are also going to use this result quite a bit in terms of decomposition of controllable and observable subspaces. So keep it in your pocket.

2.1 Minimal Polynomial

In order to show the decomposition

$$\mathbb{C}^n = \mathcal{N}(A-\lambda_1 I)^{m_1} \oplus \mathcal{N}(A-\lambda_2 I)^{m_2} \oplus \cdots \oplus \mathcal{N}(A-\lambda_p I)^{m_p}$$

we need to revisit the characteristic polynomial and its cousin the minimal polynomial.

We know that

$$\det(sI - A) = \chi_A(s) \quad \text{(characteristic polynomial)}$$

We can write

$$\chi_A(s) = (s-\lambda_1)^{d_1}(s-\lambda_2)^{d_2}\cdots(s-\lambda_p)^{d_p}$$

where $d_1, \ldots, d_p$ are the ***algebraic multiplicities*** of $\lambda_1, \ldots, \lambda_p \in \mathbb{C}$, respectively and

$$\sum_{i=1}^{p} d_i = n$$

By Cayley-Hamilton, we know that

$$\chi_A(A) = 0_{n\times n}$$

Let $\psi_A(s)$ be the polynomial of least degree such that

$$\psi_A(A) = 0_{n\times n}$$

Definition. (Minimal Polynomial.) Given a matrix $A \in \mathcal{C}^{n\times n}$, we call *minimal polynomial* of A the annihilating polynomial ψ of least degree. The minimal polynomial is of the form

$$\psi_A(s) = (s-\lambda_1)^{m_1}\cdots(s-\lambda_p)^{m_p}$$

for some integers $m_i \leq d_i$.

Proposition. $\psi_A(s)$ divides $\chi_A(s)$. That is,

$$\frac{\chi_A(s)}{\psi_A(s)} = q(s)$$

for some polynomial $q(s)$.

Example.

1. Consider

$$A_1 = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix}$$

   Then,

$$\chi_A(s) = (s-\lambda_1)^2(s-\lambda_2) \quad \text{and} \quad \psi_A(s) = (s-\lambda_1)(s-\lambda_2)$$

2. Consider

$$A_2 = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{bmatrix}$$

Then,

$$\chi_A(s) = (s-\lambda_1)^3 \quad \text{and} \quad \psi_A(s) = (s-\lambda_1)^2$$

$$\psi_A(A) = (A-\lambda_1 I)(A-\lambda_1 I) = A^2 - 2\lambda_1 A + \lambda_1^2 I = 0$$

where

$$A^2 = \begin{bmatrix} \lambda_1^2 & 2\lambda_1 & 0 \\ 0 & \lambda_1^2 & 0 \\ 0 & 0 & \lambda_1^2 \end{bmatrix}$$

3. Consider

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

Then

$$\chi_A(s) = (s-3)^2(s-4)$$

The minimal polynomial is either

$$(s-3)(s-4)$$

or

$$(s-3)^2(s-4)$$

It cannot be the former since $(A-3I)(A-4I) \neq 0$.

That is, $\psi_A(s)$ is the least degree polynomial such that $\psi_A(A) = 0$.

Theorem (Decomposition)

$$\mathbb{C}^n = \mathcal{N}(A-\lambda_1 I)^{m_1} \oplus \mathcal{N}(A-\lambda_2 I)^{m_2} \oplus \cdots \oplus \mathcal{N}(A-\lambda_p I)^{m_p}$$

Proof.

$$\frac{1}{\psi_A(s)} = \frac{1}{(s-\lambda_1)_1^m \cdots (s-\lambda_p)^{m_p}} = \frac{n_1(s)}{(s-\lambda_1)^{m_1}} + \cdots + \frac{n_p(s)}{(s-\lambda_p)^{m_p}}$$

so that

$$1 = n_1(s)q_1(s) + \cdots n_p(s)q_p(s)$$

where

$$q_i(s) = \frac{\psi_A(s)}{(s-\lambda_i)^{m_i}}$$

Thus,

$$I = n_1(A)q_1(A) + \cdots + n_p(A)q_p(A)$$

so that

$$x = \underbrace{n_1(A)q_1(A)}_{x_1} x + \cdots + \underbrace{n_p(A)q_p(A)}_{x_1} x$$

which in turn implies that

$$x_i = n_i(A)q_i(A) = n_i(A)\frac{\psi_A(A)}{(A-\lambda_i I)^{m_i}}$$

so that

$$(A-\lambda_i I)^{m_i} x_i = 0_n \implies x_i \in \mathcal{N}(A-\lambda_i I)_i^m$$

To show the decomposition is unique, we argue by contradiction. Let

$$x_i \in \mathcal{N}(A - \lambda_i I)^{m_i}$$

so that

$$x_1 + \cdots + x_p = 0_n$$

and wlog assume $x_1 \neq 0$. Then

$$x_1 = -x_2 - x_3 - \cdots - x_p$$

so that

$$(A - \lambda_2 I)^{m_2} \cdots (A - \lambda_p I)^{m_p} x_1 = 0_n$$

But $q_1(s)$ and $(s - \lambda_1)^{m_1}$ are *co-prime* meaning that

$$h_1(s)q_1(s) + h_2(s)(s - \lambda_1)^{m_1} = 1$$

This implies that

$$h_1(A) \underbrace{q_1(A)x_1}_{0} + h_2(A) \underbrace{(A - \lambda_1 I)^{m_1} x_1}_{0} = x_1 \quad \Longrightarrow \quad x_1 = 0 \ \rightarrow\leftarrow$$

□

If the minimal polynomial is

$$\psi_A(\lambda) = \prod_{i=1}^{p} (\lambda - \lambda_i)^{m_i}$$

with $1 \leq m_i \leq d_i$ and d_i the algebraic multiplicity, then

$$\mathcal{N}((A - \lambda_i I)^{m_i})$$

is the algebraic eigenspace and $\mathcal{N}(A - \lambda_i I)$ is the geometric eigenspace with $\dim(\mathcal{N}((A - \lambda_i I)^{m_i}))$ equal to the algebraic multiplicity and $\dim(\mathcal{N}(A - \lambda_i I))$ equal to the geometric multiplicity.

Proposition.

$$\dim(\mathcal{N}((A - \lambda_i I)^{m_i})) = d_i.$$

Proof. see C& D, p.115 □

2.2 Jordan Form Details

Definition. (Jordan Block.) Let $\lambda \in \mathbb{C}$. A Jordan block $J_k(\lambda)$ is a $k \times k$ upper triangular matrix of the form

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix}$$

A Jordan matrix is any matrix of the form

$$J = \mathrm{diag}(J_{n_1}(\lambda_1), \ldots, J_{n_p}(\lambda_p))$$

where the matrices J_{n_1} are Jordan blocks. If $J \in \mathbb{C}^{n \times n}$, then $n_1 + n_2 + \cdots + n_p = n$.

Recall that

$$\chi_A(s) = \det(sI - A) = (s - \lambda_1)^{d_1} \cdots (s - \lambda_p)^{d_p}$$

When eigenvalues are distinct, $d_i = 1$ for each i so that A is diagonalizable.

Theorem. (semisimple system.) A square complex $n \times n$ matrix is *semisimple* if and only if there exists a nonsingular complex $n \times n$ matrix T^{-1} and diagonal complex $n \times n$ matrix Λ for which

$$A = T^{-1}\Lambda T$$

or equivalently

$$\Lambda = TAT^{-1}$$

The columns $e_i \in \mathbb{C}^n$ of T^{-1} organized as

$$T^{-1} = \begin{bmatrix} | & | & & | \\ e_1 & e_2 & \cdots & e_n \\ | & | & & | \end{bmatrix} \in \mathbb{C}^{n\times n}$$

and the diagonal entries $\lambda_i \in \mathbb{C}$ of Λ organized as

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{C}^{n\times n}$$

may be taken as n eigenvectors associated according to (1) with n eigenvalues λ_i of A that form a *spectral list*.

Note: We call *spectral list* of A any n–tuple $(\lambda_i)_{i=1}^n$ of eigenvalues that is complete as a list of roots of the characteristic polynomial χ_A
In other words, $A \in \mathbb{C}^{n\times n}$ is semisimple iff A is *diagonalizable by a similarity transformation.*

Example.Modal Decomposition. Consider

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned}$$

Define $z = Tx$. Then

$$\begin{aligned} \dot{z} &= TAT^{-1}z + TBu \\ y &= CT^{-1}z \end{aligned}$$

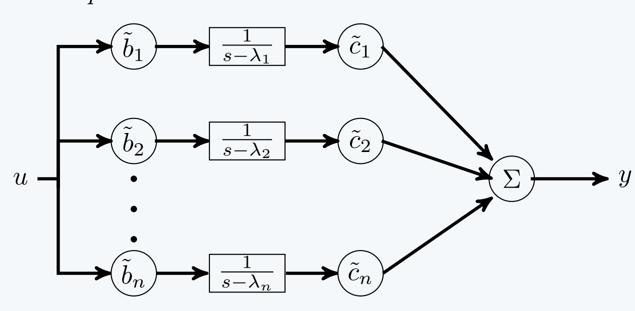
Now consider the case where we have a single input/single output (SISO, e.g., $m = 1 = p$). Define

$$TB = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \vdots \\ \tilde{b}_n \end{bmatrix}, \; CT^{-1} = \begin{bmatrix} \tilde{c}_1 & \tilde{c}_2 & \cdots & \tilde{c}_n \end{bmatrix}$$

Then

$$C(sI - A)^{-1}b = \frac{\tilde{c}_1\tilde{b}_1}{s-\lambda_1} + \cdots + \frac{\tilde{c}_n\tilde{b}_n}{s-\lambda_n} = \sum_{i=1}^{n} \frac{\tilde{c}_i\tilde{b}_i}{s-\lambda_i}$$

which is called the *modal decomposition.*



Note: $c(sI - A)^{-1}b$ is called the transfer function. If $\tilde{c}_i$ or $\tilde{b}_i$ is zero, then the transfer function does not contain the term $1/(s - \lambda_i)$.

Theorem. (Jordan Normal Form.) Let $A \in \mathbb{C}^{n\times n}$. Then there is a non-singular matrix P such that

$$A = P\mathrm{diag}(J_{n_1}(\lambda_1), \ldots, J_{n_k}(\lambda_p))P^{-1} = PJP^{-1}$$

where $J_{n_i}(\lambda_i)$ is a Jordan block and $\sum_{i=1}^{p} n_i = n$. The Jordan form J is unique up to permutations of the blocks. The eigenvalues $\lambda_1, \ldots, \lambda_p$ are not necessarily distinct. If A is real with real eigenvalues, then P can be taken as real.

Suppose that

$$P^{-1}AP = J = \mathrm{diag}(J_1, \ldots, J_p)$$

where $J_i = J_{n_i}(\lambda_i)$. Express P as

$$P = \begin{bmatrix} P_1 & \cdots & P_k \end{bmatrix}$$

where $P_i \in \mathbb{C}^{n\times n_i}$ are the columns of P associated with i–th Jordan block J_i. We have that

$$AP_i = P_i J_i$$

Let $P_i = [v_{i1}\ v_{i2}\ \cdots\ v_{in_i}]$ so that

$$Av_{i1} = \lambda_i v_{i1}$$

that is, the first column of each P_i is an eigenvector associated with eigenvalue λ_i. For $j = 2, \ldots, n_i$,

$$Av_{ij} = v_{ij-1} + \lambda_i v_{ij}$$

These $v_{i1}, \ldots, v_{in_i}$ are the generalized eigenvectors.

3 How to compute the Jordan Form

Let A be an $n \times n$ square matrix. If its characteristic equation $\chi_A(s) = 0$ has a repeated root then A may not be diagonalizable, so we need the *Jordan Canonical Form*. Suppose λ is an eigenvalue of A, with multiplicity r as a root of $\chi_A(s) = 0$. The vector v is an eigenvector with eigenvalue λ if

$$Av = \lambda v \iff (A - \lambda I)v = 0$$

The trouble is that this equation may have fewer than r linearly independent solutions for v. So we generalize and say that v is a generalized eigenvector with eigenvalue λ if

$$(A - \lambda I)^k v = 0$$

for some positive k.

Now one can prove that there are exactly r linearly independent generalized eigenvectors. Finding the Jordan form is now a matter of sorting these generalized eigenvectors into an appropriate order.

3.1 Jordan Form Procedure

To find the Jordan form carry out the following procedure.
Step 1: Compute Jordan Blocks for λ. For each λ_i an eigenvalue of A, let d_i be the algebraic multiplicity. Then, do the following:

1. First solve $(A - \lambda_i I)v = 0$, counting the number r_1 of linearly independent solutions.
2. If $r_1 = d_i$ good, otherwise $r_1 < d_i$ and we must now solve

   $$(A - \lambda_i I)^2 v = 0.$$

   There will be r_2 linearly independent solutions where $r_2 > r_1$.
3. If $r_2 = d_i$ good, otherwise solve

   $$(A - \lambda_i I)^3 v = 0$$

   to get $r_3 > r_2$ linearly independent solutions, and so on...

4. Eventually, one gets $r_1 < r_2 < \cdots < r_{N-1} < r_N = d_i$.

   The number N is the size of the largest Jordan block associated with λ_i, and r_1 is the total number of Jordan blocks associated to λ_i. Define

$$s_1 = r_1,\ s_2 = r_2 - r_1,\ s_3 = r_3 - r_2,\ \ldots, s_N = r_N - r_{N-1}$$

   so that s_k is the number of Jordan blocks of size at least $k \times k$ associated to λ_i.
5. Finally, put

$$m_1 = s_1 - s_2,\ m_2 = s_2 - s_3,\ \ldots, m_{N-1} = s_{N-1} - s_N,\ m_N = s_N.$$

   Then m_k is the number of $k \times k$ Jordan blocks associated to λ.

Once we've done this for all eigenvalues then we've got the Jordan form!

Step 2: Compute the similarity transform matrix P. To find P such that

$$J = P^{-1}AP,$$

for each eigenvalue λ do

1. First find the Jordan block sizes associated to λ by the above process. Put them in decreasing order

$$N_1 \geq N_2 \geq \cdots \geq N_p.$$

2. Now find a vector $v_{1,1}$ such that $(A - \lambda I)^{N_1} v_{1,1} = 0$ but $(A - \lambda I)^{N_1 - 1} v_{1,1} \neq 0$.
3. Define $v_{1,2} = (A - \lambda I)v_{1,1}$, $v_{1,3} = (A - \lambda I)v_{1,2}$ and so on until we get v_{1,N_1}. We can't go further because $(A - \lambda I)v_{1,N_1} = 0$.
4. If we only have one block we're OK, otherwise we can find $v_{2,1}$ such that $(A - \lambda I)^{N_2} v_{2,1} = 0$, and $(A - \lambda I)^{N_2 - 1} v_{2,1} \neq 0$ and AND $v_{2,1}$ is not linearly dependent on $v_{1,1}, \ldots, v_{1,N_1}$.
5. Define $v_{2,2} = (A - \lambda I)v_{2,1}$ etc.
6. keep going if you have more blocks, otherwise you will have r linearly independent vectors $v_{1,1}, \ldots, v_{p,N_p}$. Let

$$P_\lambda = \begin{bmatrix} v_{p,N_p} & \cdots & v_{1,1} \end{bmatrix}$$

   be the n by r matrix.

Do this procedure for all λ's and then stack the P_λ's up horizontally to get P. voila!

3.2 Examples

What easier way to learn something than through doing!

Example 1. Consider

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

Then,

$$\det(sI - A) = \det\left(\begin{bmatrix} s+1 & 1 & -2 \\ -1 & s-2 & 1 \\ 1 & 0 & s-2 \end{bmatrix}\right) \det\left(\begin{bmatrix} s-1 & -1 \\ 0 & s-1 \end{bmatrix}\right) = (s-1)^5 \implies \lambda = 1,\ d = 5$$

Step 1: Computing the J-form by way of eigenspaces.

a. Find $\mathcal{N}(A-I)$.

$$\begin{bmatrix} -2 & -1 & 2 & -1 & 2 \\ 1 & 1 & -1 & 0 & -1 \\ -1 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix} = 0$$

Row reduction makes this set of linear equations equivalent to

$$\begin{bmatrix} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix} = 0 \implies x = \begin{bmatrix} x_3 - x_4 \\ x_4 \\ x_3 \\ x_4 \\ 0 \end{bmatrix} \implies \mathcal{N}(A-I) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right)$$

Hence, $r_1 = 2 < d = 5$.

b. Find $\mathcal{N}((A-I)^2)$.

$$(A-I)^2 x = 0 \implies \begin{bmatrix} 1 & 1 & -1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = 0 \implies \begin{bmatrix} -x_2 + x_3 + x_5 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = 0$$

Hence,

$$\mathcal{N}(A-I)^2 = \operatorname{span}\left(\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right)$$

hence, $r_2 = 4 < d = 5$.

c. Find $\mathcal{N}((A-I)^3)$. Note that $(A-I)^3 = 0$. Hence,

$$\mathcal{N}((A-I)^3) = \operatorname{span}(e_1, e_2, e_3, e_4, e_5)$$

where e_i's are the standard basis vectors. Hence, $r_3 = d = 5$ so we are done.

d. $N = 3$ in the procedure so this is the size of the largest Jordan block, and $r_1 = 2$ is the total number of Jordan blocks. Define

$$s_1 = r_1 = 2, \; s_2 = r_2 - r_1 = 4 - 2 = 2, \; s_3 = r_3 - r_2 = 1$$

And,

$$m_1 = s_1 - s_2 = 0, \; m_2 = s_2 - s_3 = 1, \; m_3 = 1$$

so that m_k is the number of $k \times k$ J-blocks associated to $\lambda = 1$.

Step 2: Get P.

a. J-block sizes associated to $\lambda = 1$ are

$$N_1 = 3 \geq N_2 = 2$$

b. Find $v_{1,1}$ s.t.

$$(A-I)^3 v_{1,1} = 0, \; (A-I)^2 v_{1,1} \neq 0$$

Let $v_{1,1} = e_3$ since

$$(A-I)v_{1,1} = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

Then, define

$$v_{1,2} = (A - I)v_{1,1} = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \; v_{1,3} = (A - I)v_{1,2} = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

c. Find $v_{2,1}$ s.t.

$$(A - I)^2 v_{2,1} = 0, \; (A - I)v_{2,1} \neq 0, \; v_{2,1} \text{ L.I. from } \{v_{1,1}, v_{1,2}, v_{1,3}\}$$

There are many possible solutions. Try

$$v_{2,1} = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Indeed,

$$(A - I)^2 v_{2,1} = 0, \; (A - I)v_{2,1} = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}$$

and $v_{2,1}$ is linearly independent from $\{v_{1,1}, v_{1,2}, v_{1,3}\}$ which can easily be seen by the fact that none of the vectors in the first chain have an element in the fourth entry. Then,

$$v_{2,2} = (A - I)v_{2,1} = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}$$

d. Now

$$P = \begin{bmatrix} v_{1,3} & v_{1,2} & v_{1,1} & v_{2,2} & v_{2,1} \end{bmatrix} = \begin{bmatrix} -1 & 2 & 0 & -2 & 2 \\ 0 & -1 & 0 & 1 & 0 \\ -1 & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$J = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

4 Functions of a matrix

We need to understand how to compute functions of a matrix more generally so that we can compute

$$f(A) = \exp(A)$$

when A is not diagonalizable. We know that for any $A = PJP^{-1}$ where J is a Jordan form for A,

$$\exp(A) = \exp(PJP^{-1}) = P\exp(J)P^{-1}$$

But since J is not diagonal, we need to understand how to compute it.

4.1 Short cut for $f(\cdot) = \exp(\cdot)$

The exponential function has some nice properties in that it can be expressed as an infinite series

$$\exp(A) = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$$

We also know that we can write a Jordan form J as

$$J = D + N$$

where D is a diagonal matrix and N is a nilpotent matrix (cf. Theorem 14 page 118 **[C&D]**). Then, let m be the power of N such that $N^m = 0$. Using the property of the matrix exponential which says

$$\exp(A + B) = \exp(A)\exp(B)$$

we can then deduce that

$$\exp(J) = \exp(D + N) = \exp(D)\exp(N) = \exp(D)\left(\sum_{k=0}^{m} \frac{A^k}{k!}\right)$$

For nilpotent matrices with small m, this is a fine approach to computing the function $\exp(\cdot)$ where the argument is a non-diagonalizable matrix A.

What about more generally?

4.2 Analytic functions of a matrix

An *analytic function* is a function that is locally given by a convergent power series—i.e., f is real analytic on an open set U in the real line if for any $x_0 \in U$ one can write

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n$$

where $a_n \in \mathbb{R}$ for each n and the series is convergent to $f(x)$ for $x$in a neighborhood of x_0.

We can view $\exp(At)$ as the analytic function $f(\lambda t)$ in which A has been replaced by λ; this "substitution" requires clarification. However in the case of polynomials, the procedure is almost obvious and we have seen it before with Cayley Hamilton:

$$p(\lambda) = \lambda^3 + 3\lambda^2 + 5\lambda + 2 \iff p(A) = A^3 + 3A^2 + 5A + 2I$$

Let $A \in \mathbb{C}^{n\times n}$ and let $\sigma(A)$ denote the spectrum of A (containing distinct eigenvalues of A) with $p = |\sigma(A)|$. The minimal polynomial of A is given as above by

$$\psi_A(\lambda) = \prod_{i=1}^{p} (\lambda - \lambda_i)^{m_i}$$

Definition. (Functions of a matrix.) Let $f(s)$ be any function of s analytic on the spectrum of A and $q(s)$ be a polynomial such that

$$f^{(k)}(\lambda_\ell) = q^{(k)}(\lambda_\ell)$$

for $0 \le k \le m_\ell - 1$ and $1 \le \ell \le p$. Then

$$f(A) = q(A)$$

In fact, if $m = \sum_{i=1}^{p} m_i$ then

$$q(s) = a_1 s^{m-1} + a_2 s^{m-2} + \cdots + a_m s^p$$

where $a_1, a_2, \dots, a_n$ are functions of

$$(f(\lambda_1), f^{(1)}(\lambda_1), f^{(2)}(\lambda_1), \dots, f^{(m_1)}(\lambda_1), f(\lambda_2), \dots)$$

and hence

$$f(A) = a_1 A^{m-1} + \cdots + a_m A^0 = \sum_{\ell=1}^{p} \sum_{k=0}^{m_\ell - 1} q_{k,\ell}(A) f^{(k)}(\lambda_\ell)$$

where $q_{k,\ell}$'s are polynomials independent of f.

This leads to the following theorem.

Theorem. (General Form of $f(A)$.) Let $A \in \mathbb{C}^{n\times n}$ have a minimal polynomial ψ_A given by

$$\psi_A(s) = \prod_{k=1}^{p} (s - \lambda_k)^{m_k}$$

Let the domain Δ contain $\sigma(A)$, then for any analytic function $f : \Delta \to \mathbb{C}$. we have

$$f(A) = \sum_{k=1}^{p} \sum_{\ell=0}^{m_k - 1} f^{(\ell)}(\lambda_k) q_{k,\ell}(A)$$

where $q_{k,\ell}$'s are polynomials independent of f.

Example. *(Demonstrating the limit definition for derivatives in functions of matrix.)* Define

$$J_2(\lambda, \varepsilon) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda + \varepsilon \end{bmatrix}$$

with eigenvalues $\lambda_1 = \lambda$ and $\lambda_2 = \lambda + \varepsilon$. For any $\varepsilon \neq 0$, $J_2(\lambda, \varepsilon)$ is diagonalizable. Computing eigenvector

$$\begin{aligned} [\lambda_1 I - J_2(\lambda, \varepsilon)] v_1 &= \begin{bmatrix} 0 & -1 \\ 0 & -\varepsilon \end{bmatrix} v_1 = 0 \implies v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \left[\lambda_2 I - J_2(\lambda, \varepsilon)\right] v_2 &= \begin{bmatrix} \varepsilon & -1 \\ 0 & 0 \end{bmatrix} v_2 = 0 \implies v_1 = \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix} \end{aligned}$$

and

$$T = \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix}, \; T^{-1} = \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 1/\varepsilon \end{bmatrix}$$

we can evaluate

$$f(J_2(\lambda, \varepsilon)) = T f(\Lambda) T^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix} \begin{bmatrix} f(\lambda) & 0 \\ 0 & f(\lambda + \varepsilon) \end{bmatrix} \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 1/\varepsilon \end{bmatrix} = \begin{bmatrix} f(\lambda) & (f(\lambda + \varepsilon) - f(\lambda))/\varepsilon \\ 0 & f(\lambda + \varepsilon) \end{bmatrix}$$

Suppose $J_2(\lambda, \varepsilon) \to J_2(\lambda)$ as $\varepsilon \to 0$ and f is continuous, then if f is also differentiable at λ

$$f(J_2(\lambda, \varepsilon)) = \lim_{\varepsilon \to 0} f(J_2(\lambda, \varepsilon)) = \begin{bmatrix} f(\lambda) & f'(\lambda) \\ 0 & f(\lambda) \end{bmatrix}$$

Now more generally. . . Consider

$$J = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix} \in \mathbb{F}^{n\times n}$$

Claim:

$$f(J) = \begin{bmatrix} f(\lambda) & f^{(1)}(\lambda) & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ 0 & \ddots & \ddots & \cdots \\ \vdots & \ddots & \ddots & f^{(1)}(\lambda) \\ 0 & \cdots & \cdots & f(\lambda) \end{bmatrix}$$

Proof. The minimum polynomial is $(s-\lambda)^n$ so that

$$f(J) = \sum_{\ell=0}^{n-1} f^{(\ell)}(\lambda) q_\ell(J)$$

(that is, we can drop the inner sum since $m = 1$). Recall that the q_ℓ's are independent of the f. Hence, we can determine what they have to be by cleverly choosing and f and solving for the q's. Indeed, choose

$$\begin{array}{lll} f_1(s) = 1 & \Longrightarrow \quad f_1(J) = I = f_1^{(0)} q_0(J) & \Longrightarrow \quad q_0(J) = I \\ f_2(s) = s - \lambda & \Longrightarrow \quad f_2(J) = J - \lambda I = f_2^{(1)}(\lambda) q_1(J) & \Longrightarrow \quad q_1(J) = J - \lambda I \\ f_3(s) = (s-\lambda)^2 & \Longrightarrow \quad f_3(J) = (J - \lambda I)^2 = f_3^{(2)}(\lambda) q_2(J) & \Longrightarrow \quad 2q_2(J) = (J - \lambda I)^2 \end{array}$$

Hence

$$\begin{aligned} p_0(J) &= I \\ p_1(J) &= J - \lambda I \\ p_2(J) &= \frac{1}{2}(J - \lambda I)^2 \end{aligned}$$

so that

$$f(J) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2} & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \frac{f''(\lambda)}{2} \\ \vdots & \ddots & \ddots & \ddots & f'(\lambda) \\ 0 & \cdots & 0 & 0 & f(\lambda) \end{bmatrix}$$

□

Theorem. (Spectral Mapping Theorem.)

$$\sigma(f(J)) = f(\sigma(J)) = \{f(\lambda), f(\lambda), \ldots, f(\lambda)\}$$

and more generally that

$$\sigma(f(A)) = f(\sigma(A))$$

For some Jordan block J of size k,

$$e^{J(\lambda t)} = \begin{bmatrix} e^{\lambda t} & te^{\lambda t} & \frac{t^2}{2!}e^{\lambda t} & \cdots & \frac{t^{k-1}}{(k-1)!}e^{\lambda t} \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \frac{t^2}{2!}e^{\lambda t} \\ \vdots & \ddots & \ddots & \ddots & te^{\lambda t} \\ 0 & \cdots & 0 & 0 & e^{\lambda t} \end{bmatrix}$$

What does this mean? Well more generally if we had

$$J = \operatorname{diag}\left\{ \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix}, \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix}, \begin{bmatrix} \lambda_2 & 1 \\ 0 & \lambda_2 \end{bmatrix}, \lambda_2 \right\}$$

Recall that this Jordan form may be obtained from A by the similarity transform

$$J = P^{-1}AP$$

where

$$P = \begin{bmatrix} e_1 & v_1 & w_1 & e_2 & v_2 & e_3 & v_3 & e_4 \end{bmatrix}$$

where $e_1, \ldots, e_4$ are eigenvectors and the rest are generalized eigenvectors. Then

$$f(A) = f(PJP^{-1}) = Pf(J)P^{-1}$$

where

$$f(J) = \operatorname{diag}\left\{ \begin{bmatrix} f(\lambda_1) & f'(\lambda_1) & \frac{f''(\lambda_1)}{2} \\ 0 & f(\lambda_1) & f'(\lambda_1) \\ 0 & 0 & f(\lambda_1) \end{bmatrix}, \begin{bmatrix} f(\lambda_1) & f'(\lambda_1) \\ 0 & f(\lambda_1) \end{bmatrix}, \begin{bmatrix} f(\lambda_2) & f'(\lambda_2) \\ 0 & f(\lambda_2) \end{bmatrix}, f(\lambda_2) \right\}$$

Example. Compute e^{At} where

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

Eigenvalues:

$$\det(\lambda I - A) = \det\left(\begin{bmatrix} \lambda & -1 \\ 1 & \lambda+2 \end{bmatrix}\right) = \lambda^2 + 2\lambda + 1 = (\lambda+1)^2 = 0 \implies \lambda_1 = \lambda_2 = -1$$

Since we have repeated eigenvalues, we need to check the dimension of $\mathcal{N}(A+I)$. Indeed,

$$A + I = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \implies \operatorname{rank}(A+I) = 1$$

Since there is only one eigenvector, there must be a generalized eigenvector.
Generalized eigenvectors: Goal is to find $v_{1,2} \in \mathcal{N}((A+I)^2)$ where $(A+I)v_{1,2}! = 0$:

$$(A+I)^2 = 0 \implies \mathcal{N}((A+I)^2) = \operatorname{span}(e_1, e_2)$$

Let's try e_1. That is,

$$(A+I)^2 e_1 = 0,\ (A+I)e_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \implies v_{1,2} = e_1,\ v_{1,1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Jordan form:

$$P = \begin{bmatrix} v_{1,1} & v_{1,2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix},\ P^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix},\ J = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

Matrix exponential:

$$e^{At} = Pe^{J(-t)}P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} e^{-t} & te^{-t} \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} (t+1)e^{-t} & te^{-t} \\ -te^{-t} & (1-t)e^{-t} \end{bmatrix}$$

Lecture 12: Stability

Lecturer: L.J. Ratliff

References: Chapter 7 **[C&D]**

1 Stability

Recall that the solution to a linear time varying ODE $\dot{x}(t) = A(t)x(t),\ x(t_0) = x_0$ is given by

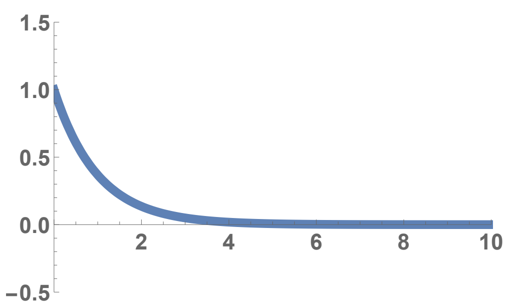
$$x(t) = \Phi(t, t_0)x_0$$

where $\Phi(t, t_0)$ is the state transition matrix.
Equilibrium point:

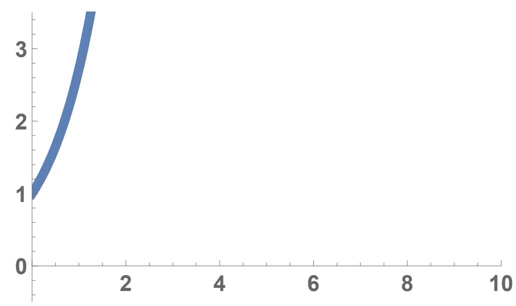
$$x_0 = \theta_n \implies x(t) = \theta_n\ \forall t$$

$x_e = \theta_n$ is called the *equilibrium point*.
Q: When is an equilibrium point stable? For some intuition, let us consider

$$\dot{x} = -\lambda x,\ x(0) = x_0,\ \lambda > 0$$



$\lambda = -1$



$\lambda = 1$

Definition 1. (Stable Equilibrium (Eq.)) $x_e = \theta_n$ is *stable* $\iff \forall\ x_0 \in \mathbb{R}^n$, $\forall t_0 \in \mathbb{R}^n$ the map $t \mapsto x(t) = \Phi(t, t_0)x_0$ is bounded for all $t \geq t_0$.

Definition 2. Asymptotic $x_e = \theta_n$ is *asymptotically stable* $\iff x_0 = \theta_n$ is stable and $t \mapsto x(t) = \Phi(t, t_0)x_0$ tends 0 as $t \to \infty$.

Definition 3. Exponential $x_e = \theta_n$ is *exponentially stable* $\iff \exists M, \alpha > 0$ such that

$$\|x(t_0)\| \leq M \exp(-\alpha(t - t_0))\|x_0\|$$

Theorem 1. (necessary and sufficient conditions.) $x_0 = 0$ is asymptotically stable $\iff \Phi(t, 0) \to 0$ as $t \to \infty$

Proof. $(\Longleftarrow)$

$$x(t) = \Phi(t, t_0)x_0 = \Phi(t, 0)\Phi(0, t_0)x_0$$

since $\Phi(t, 0) \to 0$ as $t \to \infty$ then $\|\Phi(t, 0)\| \to 0$ as $t \to \infty$ and

$$\|x(t)\| \leq \|\Phi(t, 0)\|\|\Phi(0, t_0)\|\|x_0\|$$

thus $x(t) \to 0$ as $t \to \infty$.
($\Longrightarrow$) By contradiction; assume that $t \to \Phi(t,0)$ does not tend to zero as $t \to \infty$, i.e. $\exists\ i, j$ such that

$$\Phi_{ij}(t,0) \not\to 0 \quad \text{as} \quad t \to \infty$$

choose

$$x_0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

with 1 in the j–th spot. Thus,

$$x_i(t) = \Phi_{ij}(t,0) \not\to 0 \quad \text{as} \quad t \to \infty$$

contradicting the asymptotic stability of 0. $\square$

1.1 Two stability results

From the above derivation of functions of matrices, we can show that

$$\exp(tA) = \sum_{k=1}^{\sigma} \sum_{\ell=0}^{m_k-1} t^\ell \exp(\lambda_k t) p_{k\ell}(A)$$

This gives rise to the following stability condition:

Proposition 1. Consider the differential equation $\dot{x} = Ax$, $x(0) = x_0$. From the above expression:

$$\{\exp(At) \to 0 \quad \text{as} \quad t \to \infty\} \Longleftrightarrow \{\forall \lambda_k \in \sigma(A),\ \operatorname{Re}(\lambda_k) < 0\}$$

and

$$\{t \mapsto \exp(At) \text{ is bounded on } \mathbb{R}_+\} \Longleftrightarrow \left\{ \begin{array}{ll} \forall \lambda_k \in \sigma(A), & \operatorname{Re}(\lambda_k) < 0\ \& \\ m_k = 1 \text{ when} & \operatorname{Re}(\lambda_k) = 0 \end{array} \right\}$$

We have a similar situation for discrete time systems:

$$\forall\ \nu \in \mathbb{N},\ A^\nu = \sum_{k=1}^{\sigma} \sum_{\ell=1}^{m_k-1} \nu(\nu-1)\cdots(\nu-\ell+1)\lambda_k^{\nu-\ell} p_{k\ell}(A)$$

The above gives rise to the following stability condition:

Proposition 2. Consider the discrete time system $x(k+1) = Ax(k)$, $k \in \mathbb{N}$, with $x(0) = x_0$. Then for $k \in \mathbb{N}$, $x(k) = A^k x_0$. From the above equation, we have that

$$\left\{A^k \to 0 \text{ as } k \to \infty\right\} \Longleftrightarrow \{\forall\ \lambda_i \in \sigma(A),\ |\lambda_i| < 1\}$$

and

$$\left\{k \to A^k \text{ is bounded on } \ \mathbb{N}_+\right\} \Longleftrightarrow \left\{ \begin{array}{ll} \forall \lambda_i \in \sigma(A), & |\lambda_i| \leq 1\ \& \\ m_i = 1 \ \text{ when} & |\lambda_i| = 1 \end{array} \right\}$$

1.2 LTI

remark: for the time-invariant case, asymptotic stability is equivalent to exponential stability.

Theorem 2. exponential stability the system $\dot{x} = Ax$ is exponentially stable iff all of the eigenvalues of A are in the open left half plane.

Proof. Follows from properties of matrix exponential. (proof pg 185 of C& D; proof not provided in hespanha) □

Theorem 3. Theorem 8.1 Hespanha The system $\dot{x} = Ax$ is

1. marginally stable if and only if all the eigenvalues of A have negative or zero real parts and all the Jordan blocks corresponding to eigenvalues with zero real parts are 1×1
2. asymptotically stable if and only if all the eigenvalues of A have strictly negative real parts
3. exponentially stable if and only if all the eigenvalues of A have strictly negative real parts, or
4. unstable if and only if at least one eigenvalue of A has a positive real part or zero real part, but the corresponding Jordan block is larger than 1×1.

Note: When all the eigenvalues of A have strictly negative real parts, all entries of e^{At} converge to zero exponentially fast, and therefore e^{At} converges to zero exponentially fast (for every matrix norm); i.e., there exist constants $c, \lambda > 0$ such that

$$\|e^{At}\| \leq ce^{-\lambda t}, \forall t \in \mathbb{R}$$

In this case, for a sub-multiplicative norm, we have

$$\|x(t)\| = \|e^{A(t-t_0)}x_0\| \leq \|e^{A(t-t_0)}\|\|x_0\| \leq ce^{-\lambda(t-t_0)}\|x\|_0,\ \forall t \in \mathbb{R}$$

This means asymptotic stability is equivalent to exponential stability for LTI systems.

2 Lyapunov Stability LTI Systems

Consider an LTI system

$$\dot{x} \quad = \quad Ax, \quad x(t_0) = x_0$$

The results in previous lectures about matrix exponentials provides us with simple conditions to classify the continuous-time homogeneous LTI system—since the state transition matrix is $e^{A(t-t_0)}$—in terms of its Lyapunov stability, without explicitly computing the solution to the system.

The following result (see, e.g., Theorem 8.1 Hespanha) summarizes how to characterize Lyapunov stability in terms of the eigenvalues of A.

Theorem 4. The system $\dot{x} = Ax$ is

1. *marginally stable* if and only if all the eigenvalues of A have negative or zero real parts and all the Jordan blocks corresponding to eigenvalues with zero real parts are 1×1
2. *asymptotically stable* if and only if all the eigenvalues of A have strictly negative real parts (i.e. in open left half plane),
3. *exponentially stable* if and only if all the eigenvalues of A have strictly negative real parts (i.e. in open left half plane), or
4. *unstable* if and only if at least one eigenvalue of A has a positive real part or zero real part, but the corresponding Jordan block is larger than 1×1.

The proof of the above results can be found in C&D page 185.

Note: When all the eigenvalues of A have strictly negative real parts, all entries of e^{At} converge to zero exponentially fast, and therefore e^{At} converges to zero exponentially fast (for every matrix norm); i.e., there exist constants $c, \lambda > 0$ such that

$$\|e^{At}\| \leq ce^{-\lambda t}, \forall t \in \mathbb{R}$$

In this case, for a sub-multiplicative norm, we have

$$\|x(t)\| = \|e^{A(t-t_0)}x_0\| \leq \|e^{A(t-t_0)}\|\|x_0\| \leq ce^{-\lambda(t-t_0)}\|x\|_0, \ \forall t \in \mathbb{R}$$

This implies the following fact.

For the time-invariant case, asymptotic stability is equivalent to exponential stability.

Why J-block 1x1? Recall from last time that

$$\exp(At) = \sum_{i=1}^{p} \sum_{\ell=0}^{m_k-1} t^\ell \exp(\lambda_k t) p_{k\ell}(A)$$

where

$$\mathbb{C}^n = \mathcal{N}(A-\lambda_1 I)^{m_1} \oplus \cdots \oplus \mathcal{N}(A-\lambda_p I)^{m_p}$$

Hence, if $\mathrm{Re}(\lambda_k) = 0$ and $m_k > 1$, then there are terms in the sum with t in them, which means the solution will be unbounded.

The conditions of the above theorem do not generalize to time-varying systems, even if the eigenvalues of $A(t)$ do not depend on t. One can find matrix-valued signals $A(t)$ that are stability matrices for every fixed $t \geq 0$, but the time-varying system $\dot{x} = A(t)x$ is not even stable.

2.1 Lyapunov Equation for Characterizing Stability

Recall the definition of a positive definite matrix.

Definition 4. Positive Definite Matrix A symmetric matrix $Q \in \mathbb{R}^{n\times n}$ is *positive definite* if

$$x^T Q x \geq 0, \quad \forall\ x \in \mathbb{R}^n/\{0\}$$

The Lyapunov stability theorem provides an alternative condition to check whether or not the continuous-time homogeneous LTI system

$$\dot{x} = Ax \tag{1}$$

is asymptotically stable.

Theorem 5. The following five conditions are equivalent:

1. The system (1) is asymptotically (equiv. exponentially) stable
2. All the eigenvalues of A have strictly negative real parts
3. For every symmetric positive definite matrix Q, there exists a unique solution P to the following *Lyapunov equation*

   $$A^T P + PA = -Q$$

   Moreover, P is symmetric and positive-definite.
4. There exists a symmetric positive-definite matrix P for which the following Lyapunov matrix inequality holds:

   $$A^* P + PA < 0$$

Proof. Proof Sketch. We have already seen that 1 and 2 are equivalent. To show that 2 implies 3, we need to show that

$$P = \int_0^\infty e^{A^T t} Q e^{At}\, dt$$

is the unique solution to $A^T P + PA = -Q$. This can be done by showing that i) the integral is well-defined (i.e. finite), ii) P as defined solves $A^T P + PA = -Q$, iii) P as defined is symmetric and positive definite, and lastly, iv) no other matrix solves the equation. Indeed,

i) This follows from exponential stability—i.e.

$$\|e^{A^Tt}Qe^{At}\| \to 0$$

exponentially fast as $t \to \infty$. Hence, the integral is absolutely convergent.

ii) We simply need to compute by direct verification

$$A^TP + PA = \int_0^\infty A^Te^{A^Tt}Qe^{At} + e^{A^Tt}Qe^{At}A\ dt$$

But,

$$\frac{d}{dt}\left(e^{A^Tt}Qe^{At}\right) = A^Te^{A^Tt}Qe^{At} + e^{A^Tt}Qe^{At}A$$

so that

$$\begin{aligned} A^TP + PA &= \int_0^\infty \frac{d}{dt}\left(e^{A^Tt}Qe^{At}\right)\ dt \\ &= \left(e^{A^Tt}Qe^{At}\right)\Big|_{t=0}^\infty \\ &= \left(\lim_{t\to\infty} e^{A^Tt}Qe^{At}\right) - e^{A^T0}Qe^{A0} \end{aligned}$$

And, the right-hand side is equal to $-Q$ since $\lim_{t\to\infty} e^{At} = 0$ (by asymptotic stability) and $e^{A0} = I$.

iii) this follows by direct computation. Symmetry easily follows:

$$P^T = \int_0^\infty (e^{A^Tt}Qe^{At})^T\ dt = \int_0^\infty \left(e^{At}\right)^T Q^T \left(e^{A^Tt}\right)^T\ dt = \int_0^\infty e^{A^Tt}Qe^{At}\ dt = P$$

To check positive definiteness, pick an arbitrary vector z and compute:

$$z^TPz = \int_0^\infty z^Te^{A^Tt}Qe^{At}z\ dt = \int_0^\infty w(t)^TQw(t)\ dt$$

where $w(t) = e^{At}z$. Since Q is positive definite, we get that $z^TPz \geq 0$. Moreover

$$z^TPz = 0 \implies \int_0^\infty w(t)^TQw(t)\ dt = 0$$

which only happens if $w(t) = e^{At}z = 0$ for all $t \geq 0$, from which one concludes that $z = 0$, because e^{At} is non-singular (recall all state transition matrices are!). Thus P is positive definite.

iv) prove by contradiction. assume there is some $\bar{P}$ that solves it:

$$A^TP + PA = -Q, \quad \text{and} \quad A^T\bar{P} + \bar{P}A = -Q$$

Then

$$A^T(P - \bar{P}) + (P - \bar{P})A = 0$$

Multiplying by e^{A^Tt} and e^{At} on the left and right, respectively, we conclude that

$$e^{A^Tt}A^T(P - \bar{P})e^{At} + e^{A^Tt}(P - \bar{P})Ae^{At} = 0, \quad \forall\ t \geq 0$$

Yet,

$$\frac{d}{dt}\left(e^{A^Tt}(P - \bar{P})e^{At}\right) = e^{A^Tt}A^T(P - \bar{P})e^{At} + e^{A^Tt}(P - \bar{P})Ae^{At} = 0$$

implying that $e^{A^Tt}(P - \bar{P})e^{At}$ must be constant. But,because of stability, this quantity must converge to zero as $t \to \infty$, so it must be always zero. Since e^{At} is nonsingular, this is possible only if $P = \bar{P}$.

The implication that condition 3 $\implies$ condition 4 follows immediately, because if we select $Q = -I$ in condition 3, then the matrix P that solves the Lyapunov equation also satisfies $A^T P + PA < 0$.

Now to complete the proof we need to show that condition 4 implies condition 2. Let P be a symmetric positive-definite matrix for which

$$A^T P + PA < 0 \tag{2}$$

holds and define

$$Q = -(A^T P + PA)$$

Consider an arbitrary solution to the LTI system and define the scalar time-dependent map

$$v(t) = x^T(t)Px(t) \geq 0$$

Taking derivatives, we have

$$\dot{v} = \dot{x}^T Px + x^T P\dot{x} = x^T(A^T P + PA)x = -x^T Qx \leq 0$$

Thus, $v(t)$ is nonincreasing and we conclude that

$$v(t) = x^T(t)Px(t) \leq v(0) = x^T(0)Px(0)$$

But since $v = x^T Px \geq \lambda_{\min}(P)\|x\|^2$ we have that

$$\|x\|^2 \leq \frac{x^T(t)Px(t)}{\lambda_{\min}(P)} = \frac{v(t)}{\lambda_{\min}(P)} \leq \frac{v(0)}{\lambda_{\min}(P)}$$

which means that the system is stable. To verify that it is actually exponentially stable, we go back to the derivative of v, and using the facts that $x^T Qx \geq \lambda_{\min}(Q)\|x\|^2$ and $v = x^T Px \leq \lambda_{\max}(P)\|x\|^2$, we get that

$$\dot{v} = -x^T Qx \leq -\lambda_{\min}(Q)\|x\|^2 \leq -\frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)}v, \quad \forall\ t \geq 0$$

Applying the fact that for some constant $\mu \in \mathbb{R}$,

$$\dot{v} \leq \mu v(t),\ \forall t \geq t_0 \quad \implies \quad v(t) \leq e^{\mu(t-t_0)}v(t_0),\ \forall t \geq t_0$$

we get that

$$v(t) \leq e^{-\lambda(t-t_0)}v(t_0),\ \forall t \geq 0, \quad \lambda = -\frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)}$$

which shows that $v(t)$ converges to zero exponentially fast and so does $\|x(t)\|$. $\square$

There are discrete time versions of the above results. Indeed, consider

$$x_{k+1} = Ax_k$$

then we have a similar theorem.

Theorem 6. The following four conditions are equivalent:

1. The DT LTI system is asymptotically (exponentially) stable.
2. All the eigenvalues of A have magnitude strictly smaller than 1.
3. For every symmetric positive definite Q, there exists a unique solution P to the following Stein equation (aka the discrete-time Lyapunov equation):

$$A^T PA - P = -Q$$

   Moreover, P is symmetric and positive-definite.

4. There exists a symmetric positive-definite matrix P for which the following Lyapunov matrix inequality holds:

$$A^T PA - P < 0$$