NA 565 - Fall 2023

Linear Algebra Review

August 30, 2023

UNIVERSITY OF MICHIGAN

Motivation

- Linear algebra is the workhorse of modern science and engineering.
- Digital systems and modern computations use algebraic or time-difference (discrete) equations for implementation.
- Most successful algorithms in practice boils down to doing linear algebra (machine learning, optimization, control, estimation, computer vision, etc.).
- A secret to share: If you know least squares and can solve problems, you could claim to be an engineer (unless you're asked to show proof!).

Identity Matrix

- The identity matrix is a square matrix denoted I that has ones down the main diagonal and zeroes elsewhere.

- Here are some examples of 1×1, 2×2, 3×3, and 4×4 identity matrices.

$$\begin{bmatrix} 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

- The notation I_n means an $n \times n$ identity matrix.

We identify $\mathbb{R}^n$ with the set of all n-column vectors with real entries

$$\{(x_1, x_2, \ldots, x_n) \mid x_i \in \mathbb{R}, 1 \leq i \leq n\}$$

$$\iff \mathbb{R}^n \iff \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \;\middle|\; x_i \in \mathbb{R}, 1 \leq i \leq n \right\}$$

For all real numbers α and β, and all vectors x and y in $\mathbb{R}^n$ we have scalar multiplication and vector addition:

$$\begin{aligned}\alpha x + \beta y &= \alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \beta \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix} + \begin{bmatrix} \beta y_1 \\ \beta y_2 \\ \vdots \\ \beta y_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \\ \vdots \\ \alpha x_n + \beta y_n \end{bmatrix} \in \mathbb{R}^n.\end{aligned}$$

Euclidean Norm or "Length" of a Vector

Definition

Let $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be a vector in $\mathbb{R}^n$. The Euclidean norm of v, denoted $\|v\|$, is defined as

$$\|v\| := \sqrt{(v_1)^2 + (v_2)^2 + \cdots + (v_n)^2} = \sqrt{\sum_{i=1}^{n} (v_i)^2} = \sqrt{v^\mathsf{T} v}$$

Properties of the Norm of a vector

All norms satisfy the following properties

1. For all vectors $v \in \mathbb{R}^n$, $\|v\| \geq 0$ and moreover, $\|v\| = 0 \iff v = 0$.

2. For any real number $\alpha \in \mathbb{R}$ and vector $v \in \mathbb{R}^n$,
$$\|\alpha v\| = |\alpha| \cdot \|v\|.$$

3. For any pair of vectors v and w in $\mathbb{R}^n$,
$$\|v + w\| \leq \|v\| + \|w\|.$$

A vector $v \in \mathbb{R}^n$ is a *linear combination of* $\{u_1, u_2, \ldots, u_m\} \subset \mathbb{R}^n$ if there exist real numbers $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that

$$v = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_m u_m.$$

Remark

$A \subset B$ means A is a subset of B. This means that B includes or contains A. For example, $\{1,2\} \subset \{1,2,3\}$.

- The vectors $\{v_1, v_2, ..., v_m\}$ are *linearly independent* if the *only* real numbers $\alpha_1, \alpha_2, \ldots, \alpha_m$ yielding a linear combination of vectors that adds up to the zero vector,

  $$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_m v_m = 0_{n\times 1},$$

  are $\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_m = 0$.

Suppose that $V \subset \mathbb{R}^n$ is a nonempty subset of $\mathbb{R}^n$.

Definition

V is a subspace of $\mathbb{R}^n$ if any linear combination constructed from elements of V and scalars in $\mathbb{R}$ is once again an element of V. One says that V is *closed under linear combinations.*

In symbols, $V \subset \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ if for all real numbers α and β, and all vectors v_1 and v_2 in V

$$\boxed{\alpha v_1 + \beta v_2 \in V.}$$

Span of a Set of Vectors

Definition

Suppose that $S \subset \mathbb{R}^n$, then S is a set of vectors. The set of all possible linear combinations of elements of S is called the span of S,

$\mathrm{span}\{S\} :=$ {all possible linear combinations of elements of S}.

Basis Vectors and Dimension

Suppose that V is a subspace of $\mathbb{R}^n$. Then $\{v_1, v_2, \ldots, v_k\}$ is a *basis for* V if

- the set $\{v_1, v_2, \ldots, v_k\}$ is linearly independent, and
- $\mathrm{span}\{v_1, v_2, \ldots, v_k\} = V$.
- The maximum number of vectors in any linearly independent set contained in V is the *dimension* of V (here k).

Definition

Let $n \geq 1$ and, as before, define $e_i :=$ i-th column of the $n \times n$ identity matrix, I_n. Then

$$\{e_1, e_2, \ldots, e_n\}$$

is a basis for the vector space $\mathbb{R}^n$.
Its elements e_i are called both natural (standard) basis vectors and canonical basis vectors.

We write a general system of linear equations as

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\vdots \qquad &= \vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{aligned}$$

We can write this system as:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

where:

$$A := \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, x := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, b := \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

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

- All terms above the diagonal of the matrix A are zero.
- More precisely, the condition is $a_{ij} = 0$ for all $j > i$.
- Such matrices are called *lower-triangular*.

Lower Triangular Systems and Forward Substitution

We will solve this example using a method called *forward substitution*.

$$\begin{aligned} 3x_1 &= 6 \\ 2x_1 - x_2 &= -2 \\ x_1 - 2x_2 + 3x_3 &= 2 \end{aligned} \iff \underbrace{\begin{bmatrix} 3 & 0 & 0 \\ 2 & -1 & 0 \\ 1 & -2 & 3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 6 \\ -2 \\ 2 \end{bmatrix}}_{b}.$$

Upper Triangular Matrices

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

- All terms below the diagonal of the matrix A are zero.
- More precisely, the condition is $a_{ij} = 0$ for $i > j$.
- Such matrices are called *upper-triangular*.

We solve the upper triangular systems using a method called *back substitution*.

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &= 6 \\ 2x_2 + x_3 &= -2 \\ 3x_3 &= 4, \end{aligned} \quad \Longleftrightarrow \quad \underbrace{\begin{bmatrix} 1 & 3 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 6 \\ -2 \\ 4 \end{bmatrix}}_{b}.$$

LU Factorization for Solving Linear Equations

- We wish to solve the system of linear equations $Ax = b$.
- If we can factor $A = L \cdot U$, where U is upper triangular and L is lower triangular. Then

$$L \cdot Ux = b.$$

- Define $U \cdot x =: y$, then

$$\begin{aligned} Ly &= b \\ Ux &= y. \end{aligned}$$

- We first solve for y via forward substitution. Given y, we solve for x via back substitution.

Least Squares Solutions to $A_{n\times m} \cdot x_{m\times 1} = b_{n\times 1}$

- Assume $A^\mathsf{T}A$ is invertible, i.e., the columns of A are linearly independent.

- Then there is a unique vector $x^* \in \mathbb{R}^m$ achieving $\min_{x\in\mathbb{R}^m} \|Ax-b\|^2$ and it satisfies the equation (called *the normal equations*)

$$\left(A^\mathsf{T}A\right) x^* = A^\mathsf{T}b.$$

-

$$x^* = (A^\mathsf{T}A)^{-1}A^\mathsf{T}b \iff x^* = \underset{x\in\mathbb{R}^m}{\arg\min} \|Ax-b\|^2 \iff \left(A^\mathsf{T}A\right) x^* = A^\mathsf{T}b.$$

Suppose that A is an $n \times m$ matrix with linearly independent columns.

Fact

Then there exists an $n \times m$ *matrix* Q *with orthonormal columns,* $Q^{\mathsf{T}} Q = I$*, and an upper triangular,* $m \times m$*, invertible matrix* R *such that* $A = Q \cdot R$.

Least Squares via the QR Factorization

Whenever the columns of A are linearly independent, a least squared error solution to $Ax = b$ is computed as

- factor $A =: QR$,
- compute $\bar{b} := Q^{\mathsf{T}} b$, and then
- solve $Rx = \bar{b}$ via back substitution.

- A function (or a map) view of a matrix defines two subspaces:
  1. its *null space* and
  2. its *range*.

- Let A be an $n \times m$ matrix.

- We can then define a function $f : \mathbb{R}^m \to \mathbb{R}^n$ by, for each $x \in \mathbb{R}^m$

$$f(x) := Ax \in \mathbb{R}^n.$$

The following subsets are naturally motivated by the function view of a matrix.

Definition

1. $\mathrm{null}(A) := \{x \in \mathbb{R}^m \mid Ax = 0_{n\times 1}\}$ is the *null space* of A.
2. $\mathrm{range}(A) := \{y \in \mathbb{R}^n \mid y = Ax \text{ for some } x \in \mathbb{R}^m\}$ is the *range* of A.

Range of A Equals Column Span of A

Let A be an $n \times m$ matrix, its columns are vectors in $\mathbb{R}^n$,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} =: \begin{bmatrix} a_1^{\rm col} & a_2^{\rm col} & \ldots & a_m^{\rm col} \end{bmatrix}$$

Then

$$\boxed{\mathrm{range}(A) := \{Ax \mid x \in \mathbb{R}^m\} = \mathrm{span}\{a_1^{\rm col}, a_2^{\rm col}, \ldots, a_m^{\rm col}\} =: \mathrm{col\ span}\{A\}.}$$

Definition

For an $n \times m$ matrix A,

1. $\operatorname{rank}(A) := \dim \operatorname{range}(A)$.

2. $\operatorname{nullity}(A) := \dim \operatorname{null}(A)$.

Because $\operatorname{range}(A) \subset \mathbb{R}^n$, we see that $\operatorname{rank}(A) \leq n$.

Theorem

For an $n \times m$ matrix A, we have the property

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = m \quad \textit{number of columns of } A.$$

- *Since $\mathrm{rank}(A)$ is equal to the number of linearly independent columns of A, it follows that $\mathrm{nullity}(A)$ is counting the number of linearly dependent columns of A.*

- *If all of the columns of A are linearly independent, then none are dependent, and hence $\mathrm{null}(A) = \{0_{m\times 1}\}$.*

Multi-objective least squares

- goal: choose n-vector x so that k norm squared objectives

$$J_1 = \|A_1 x - b_1\|^2, \ \ldots, \ J_k = \|A_k x - b_k\|^2$$

  are all small

- A_i is an $m_i \times n$ matrix, b_i is an m_i-vector, $i = 1, \ldots, k$
- J_i are the objectives in a *multi-objective optimization problem* (also called a *multi-criterion problem*)
- could choose x to minimize any one J_i, but we want *one* x that makes them all small

Weighted sum objective

- choose positive *weights* $\lambda_1, \ldots, \lambda_k$ and form *weighted sum objective*

$$J = \lambda_1 J_1 + \cdots + \lambda_k J_k = \lambda_1 \|A_1 x - b_1\|^2 + \cdots + \lambda_k \|A_k x - b_k\|^2$$

- we'll choose x to minimize J
- we can take $\lambda_1 = 1$, and call J_1 the *primary objective*
- interpretation of λ_i: how much we care about J_i being small, relative to primary objective
- for a bi-criterion problem, we will minimize

$$J_1 + \lambda J_2 = \|A_1 x - b_1\|^2 + \lambda \|A_2 x - b_2\|^2$$

Weighted sum minimization via stacking

- write weighted-sum objective as

$$J = \left\| \begin{bmatrix} \sqrt{\lambda_1}(A_1 x - b_1) \\ \vdots \\ \sqrt{\lambda_k}(A_k x - b_k) \end{bmatrix} \right\|^2$$

- so we have $J = \|\tilde{A}x - \tilde{b}\|^2$, with

$$\tilde{A} = \begin{bmatrix} \sqrt{\lambda_1} A_1 \\ \vdots \\ \sqrt{\lambda_k} A_k \end{bmatrix}, \qquad \tilde{b} = \begin{bmatrix} \sqrt{\lambda_1} b_1 \\ \vdots \\ \sqrt{\lambda_k} b_k \end{bmatrix}$$

- so we can minimize J using basic ('single-criterion') least squares

Weighted sum solution

- assuming columns of $\tilde{A}$ are independent,

$$\begin{aligned}\hat{x} &= (\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T\tilde{b} \\ &= (\lambda_1 A_1^T A_1 + \cdots + \lambda_k A_k^T A_k)^{-1}(\lambda_1 A_1^T b_1 + \cdots + \lambda_k A_k^T b_k)\end{aligned}$$

- can compute $\hat{x}$ via QR factorization of $\tilde{A}$
- A_i can be wide, or have dependent columns

Optimal trade-off curve

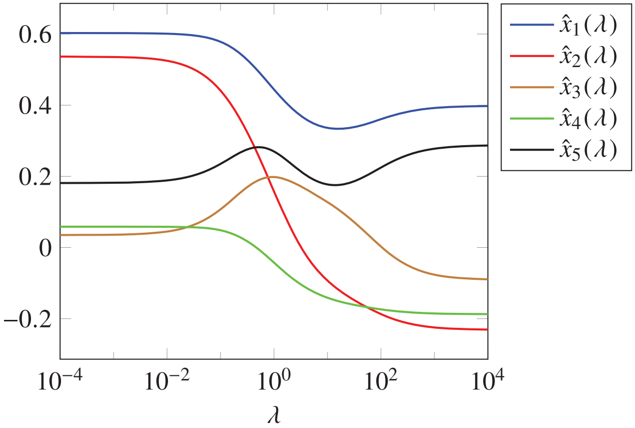
- bi-criterion problem with objectives J_1, J_2
- let $\hat{x}(\lambda)$ be minimizer of $J_1 + \lambda J_2$
- called *Pareto optimal*: there is no point z that satisfies

$$J_1(z) < J_1(\hat{x}(\lambda)), \quad J_2(z) < J_2(\hat{x}(\lambda))$$

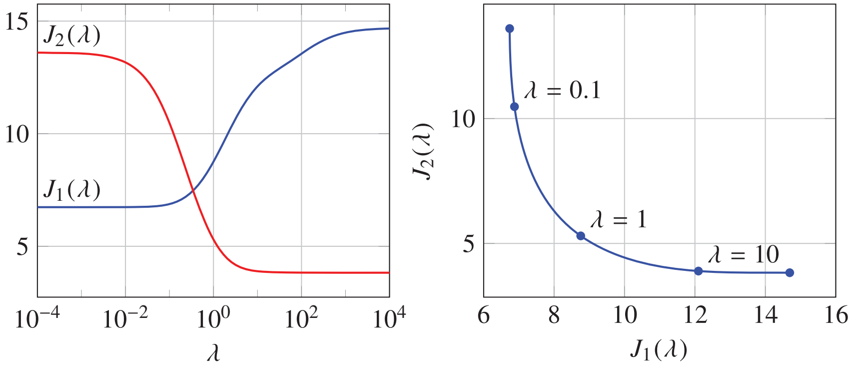
  i.e., no other point x beats $\hat{x}$ on both objectives
- *optimal trade-off curve*: $(J_1(\hat{x}(\lambda)), J_2(\hat{x}(\lambda)))$ for $\lambda > 0$

Example

A_1 and A_2 both 10×5



Objectives versus λ and optimal trade-off curve



Using multi-objective least squares

- identify the primary objective
  - the basic quantity we want to minimize
- choose one or more secondary objectives
  - quantities we'd also like to be small, if possible
  - *e.g.*, size of x, roughness of x, distance from some given point
- tweak/tune the weights until we like (or can tolerate) $\hat{x}(\lambda)$
- for bi-criterion problem with $J = J_1 + \lambda J_2$:
  - if J_2 is too big, increase λ
  - if J_1 is too big, decrease λ

Image de-blurring

- x is an image
- A is a blurring operator
- $y = Ax + v$ is a blurred, noisy image
- least squares de-blurring: choose x to minimize

$$\|Ax - y\|^2 + \lambda(\|D_\mathrm{v}x\|^2 + \|D_\mathrm{h}x\|^2)$$

  D_v, D_h are vertical and horizontal differencing operations
- λ controls smoothing of de-blurred image

Example

blurred, noisy image

regularized inversion with $\lambda = 0.007$

Image credit: NASA

Regularization path

$\lambda = 10^{-6}$

$\lambda = 10^{-4}$

Regularization path

$\lambda = 10^{-2}$

$\lambda = 1$