Linear Algebra: Introduction / Review / Crash Course

We start off by a quick review(?) of basic linear algebra concepts and algorithms.

Depending on your background this will either be a review of things you know, possibly in a new notation / framework, or a crash course.

... · · · · ·

An \( n \)-dimensional vector \( \mathbf{x} \) is an \( n \)-tuple of either real \( \mathbf{x} \in \mathbb{R}^n \) or complex \( \mathbf{x} \in \mathbb{C}^n \) numbers, in this class all vectors are column vectors, i.e.

\[
\mathbf{x} \in \mathbb{R}^n \Rightarrow \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \text{ where } x_i \in \mathbb{R}, \ i = 1, 2, \ldots, n.
\]

We express a row vector using the transpose, i.e.

\[
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^T.
\]

Vector addition and subtraction

\[
\mathbf{x}, \mathbf{y} \in \mathbb{R}^n \Rightarrow \mathbf{x} \pm \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \pm \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 \pm y_1 \\ x_2 \pm y_2 \\ \vdots \\ x_n \pm y_n \end{bmatrix},
\]

or \( \mathbf{z} = \mathbf{x} + \mathbf{y} \) where

\[
z_i = x_i + y_i, \ i = 1, 2, \ldots, n.
\]
Matrices

Matrix-Vector Product

An $m \times n$ matrix ($m$ rows, $n$ columns) $A$ with real or complex entries is represented

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{bmatrix}, \quad \{ a_{ij} \in \mathbb{R}, \text{ or } a_{ij} \in \mathbb{C} \}$$

Sometimes we write $A \in \mathbb{R}^{m \times n}$ (or $A \in \mathbb{C}^{m \times n}$).

If $A \in \mathbb{R}^{m \times n}$ and $\bar{x} \in \mathbb{R}^n$, then the **matrix-vector product**, $\bar{b} = A\bar{x}$, is well defined, and $\bar{b} \in \mathbb{R}^m$, where

$$b_i = \sum_{j=1}^{n} a_{ij}x_j, \quad i = 1, 2, \ldots, m.$$ 

**Note:** We need $n$ multiplications and $(n-1)$ additions to compute each entry in $\bar{b}$. In total we need $m \cdot n$ multiplications and $m \cdot (n-1)$ additions. We say that the matrix-vector product requires $\mathcal{O}(m \cdot n)$ operations.

Matrix-Vector Product... 

...Functional Definition

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1 a_{11} + x_2 a_{12} + x_3 a_{13} + \cdots + x_n a_{1n}$$

$$= x_1 a_{21} + x_2 a_{22} + x_3 a_{23} + \cdots + x_n a_{2n}$$

$$= \cdots$$

$$= x_1 a_{m1} + x_2 a_{m2} + x_3 a_{m3} + \cdots + x_n a_{mn}$$

$$= A_{11} x_1 + A_{12} x_2 + A_{13} x_3 + \cdots + A_{1n} x_n$$

$$= A_{21} x_1 + A_{22} x_2 + A_{23} x_3 + \cdots + A_{2n} x_n$$

$$= \cdots$$

$$= A_{mn} x_1 + A_{m2} x_2 + A_{m3} x_3 + \cdots + A_{mn} x_n$$

$$= \sum_{i=1}^{m} A_{i1} x_1 + \sum_{i=1}^{m} A_{i2} x_2 + \cdots + \sum_{i=1}^{m} A_{in} x_n$$

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

**Note:** Every linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be expressed as multiplication by an $m \times n$-matrix.

Matrix-Vector Product: Linearity

The map $\bar{x} \rightarrow A\bar{x}$ (from $\mathbb{R}^n$ to $\mathbb{R}^m$, or from $\mathbb{C}^n$ to $\mathbb{C}^m$) is **linear**, i.e. $\forall \bar{x}, \bar{y} \in \mathbb{R}^n$ ($\mathbb{C}^n$), and $\alpha, \beta \in \mathbb{R}$ ($\mathbb{C}$)

$$A(\bar{x} + \bar{y}) = A\bar{x} + A\bar{y}$$

$$A(\alpha \bar{x}) = \alpha A\bar{x}$$

$$A(\alpha \bar{x} + \beta \bar{y}) = \alpha A\bar{x} + \beta A\bar{y}$$

**Note:** Every linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be expressed as multiplication by an $m \times n$-matrix.
Example: The Vandermonde Matrix

Given a set of points \( \{x_1, x_2, \ldots, x_m\} \), we can express the evaluation of the polynomial

\[
p(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1} x^{n-1}
\]

at those points using the \( m \times n \) Vandermonde matrix \( A \), and the vector \( \bar{c} \), containing the polynomial coefficients

\[
A = \begin{bmatrix}
1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\
1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_m & x_m^2 & \cdots & x_m^{n-1}
\end{bmatrix}, \quad \bar{c} = \begin{bmatrix}
c_0 \\
c_1 \\
c_2 \\
\vdots \\
c_{n-1}
\end{bmatrix}
\]

Forming \( \bar{p} = A\bar{c} \) gives us an \( m \)-vector containing the values of \( p(x_i) \), \( i = 1, 2, \ldots, m \).

The Vandermonde Matrix...

Evaluating polynomials using matrix notation may seem cute and useless?!

But, wait a minute — this notation looks vaguely familiar from the discussion of linear least squares (LLSQ) problems from Math 541.

In case you forgot (or never studied) LLSQ: The goal is to find the best model in a class (i.e. low-dimensional polynomials) to measured data (observations \( y_i \), made at the points \( x_i \)).

The discrepancy (error) between the model and the observations is measured in the sum-of-squares norm.

Linear Least Squares: Explicit Example

Find the best straight line \( p(x) = c_0 + c_1 x \) fitting the observations \( (x, y) \in \{(0, 1), (1, 2), (2, 2.5), (3, 4), (4, 7)\} \).

We have the \( 5 \times 2 \) Vandermonde matrix \( A = [\bar{1} \bar{x}] \), the 2-vector \( \bar{c} \) (of polynomial coefficients) and the 5-vector \( \bar{y} \) (of measurements):

\[
A = \begin{bmatrix}
1 & 0 \\
1 & 1 \\
1 & 2 \\
1 & 3 \\
1 & 4
\end{bmatrix}, \quad \bar{c} = \begin{bmatrix}
c_0 \\
c_1
\end{bmatrix}, \quad \bar{y} = \begin{bmatrix}
1 \\
2 \\
2.5 \\
4 \\
7
\end{bmatrix}
\]

The Linear Least Squares Problem: Find the \( \bar{c} \) which minimizes the least squares error \( \| A\bar{c} - \bar{y} \|_2^2 \).

Figure: The data points \((x_i, y_i)\) and the straight line corresponding to the best fit (in the least-squares-sense), i.e. \( p^*(x) = c_0^* + c_1^* x \).
Given a model $\bar{c}$, we can evaluate to corresponding linear polynomial $p(x) = c_0 + c_1 x$ at the points $x_i$: $\bar{p} = A\bar{c}$. The pointwise error in the model is $\bar{e} = \bar{p} - \bar{y}$:

$$
\bar{p} = \begin{bmatrix}
c_0 + 0c_1 \\
c_0 + 1c_1 \\
c_0 + 2c_1 \\
c_0 + 3c_1 \\
c_0 + 4c_1
\end{bmatrix}, \quad \bar{e} = \begin{bmatrix}
c_0 + 0c_1 - 1 \\
c_0 + 1c_1 - 2 \\
c_0 + 2c_1 - 2.5 \\
c_0 + 3c_1 - 4 \\
c_0 + 4c_1 - 7.5
\end{bmatrix}
$$

The least squares error is given by

$$r_{\text{LSQ}} = \|\bar{e}\|_2^2 = \sum_{i=1}^{5} e_i^2 = \|A\bar{c} - \bar{y}\|_2^2$$

In order to identify the optimal choice of $\bar{c}$, we compute the partial derivatives with respect to the model parameters, and set those expressions to be zero (in order to identify the optimum)

$$\frac{\partial r_{\text{LSQ}}}{\partial c_0} = \frac{\partial r_{\text{LSQ}}}{\partial c_1} = 0$$

After some work (which is not central to this discussion), we get the Normal Equations

$$A^T\bar{c} = A^T\bar{y} \iff A^T(A\bar{c} - \bar{y}) = 0$$

Even though the matrix $A$ is (usually) tall and skinny (here $5 \times 2$), the matrix $A^T A$ is square (here $2 \times 2$). The (formal) solution $\bar{c} = [A^T A]^{-1} A^T \bar{y}$, to this linear system gives us the coefficients for the optimal polynomial (the red line on slide 12).

The previous LLSQ example raises more questions than it answers, the most important one "Would anyone in his/her right mind form the matrix $A^T A$, then invert it $[A^T A]^{-1}$, then multiply the vector $A^T \bar{y}$ by the inverse?"

The answer is "No!" ...which raises even more questions!

This class is all about how to solve linear systems... taking issues like speed, accuracy, and stability into consideration.

We will revisit the questions raised by the example in more detail later... However, we will use the example to introduce some further linear algebra functionality and terminology...

The matrix-matrix product $B = AC$ is well defined if the matrix $C$ has as many rows as the matrix $A$ has columns

$$B_{k \times n} = A_{k \times m} \cdot C_{m \times n}$$

The elements of $B$ are defined by

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

Sometimes it is useful to think of the columns of $B$, $\bar{b}_j$ as linear combinations of the columns of $A$:

$$\bar{b}_j = A\bar{c}_j = \sum_{k=1}^{m} c_{kj} \bar{a}_k$$
The Transpose of a Matrix ($A^T$)

The transpose of a matrix $A = \{a_{ij}\}$ is the matrix $A^T = \{a_{ji}\}$, e.g.:

$$
A = \begin{bmatrix}
    a_{11} & a_{12} & a_{13} & a_{14} \\
    a_{21} & a_{22} & a_{23} & a_{24} \\
    a_{31} & a_{32} & a_{33} & a_{34} \\
    a_{41} & a_{42} & a_{43} & a_{44}
\end{bmatrix}, \quad A^T = \begin{bmatrix}
    a_{11} & a_{21} & a_{31} & a_{41} \\
    a_{12} & a_{22} & a_{32} & a_{42} \\
    a_{13} & a_{23} & a_{33} & a_{43} \\
    a_{14} & a_{24} & a_{34} & a_{44}
\end{bmatrix}
$$

The operation is intellectually simple — just mirror across the diagonal — but can be quite (memory-access) expensive, especially for large matrices.

For complex matrices $C = \{c_{ij}\}$, the complex (Hermitian) transpose is given by $C^H = \{c_{ji}^*\}$, where $c^*$ is the complex conjugate of $c$:

$$
c = a + bi, \quad c^* = a - bi.
$$

The Range and Nullspace of a Matrix $A$

The range of a matrix, written $\text{range}(A)$, is the set of vectors that can be expressed as a linear combination of the columns of $A_{m \times n}$, i.e.

$$
\text{range}(A) = \{ \bar{y} \in \mathbb{R}^m : \bar{y} = A\bar{x}, \text{ for some } \bar{x} \in \mathbb{R}^n \}
$$

we say “range$(A)$ is the space spanned by the columns of $A$.”

The nullspace of a matrix $A$, written $\text{null}(A)$, is the set of vectors that satisfy $A\bar{x} = 0$, i.e.

$$
\text{null}(A) = \{ \bar{x} \in \mathbb{R}^n : A\bar{x} = 0 \}
$$

The Rank of a Matrix $A_{m \times n}$

The column rank of a matrix is the dimension of $\text{range}(A)$, its “column space.” The row rank of a matrix is the dimension of its “row space,” or $\text{range}(A^T)$.

The column rank is always equal to the row rank (we will see the proof of this in a few lectures), hence we only refer to the Rank of a matrix

$$
\text{rank}(A)
$$

An $m \times n$ matrix is of full rank if it has the maximal possible rank $\min(m, n)$.

An $m \times n$ $(m \geq n)$ matrix $A$ with full rank must have $n$ linearly independent columns.

Recall: The Normal Equations

$$
A^T A \bar{c} = A^T \bar{y} \iff A^T (A\bar{c} - \bar{y}) = 0
$$

Due to the “tall-and-skinniness” of $A$, the equation $A\bar{c} - \bar{y} = 0$ does not necessarily have a solution.

Given a vector $\bar{c}$ we can define the residual, $\bar{r}(\bar{c}) = A\bar{c} - \bar{y}$, which measures how far from solving the system we are.

We notice that the solution to the normal equations requires that the residual is in the nullspace of $A^T$.

The solution is in $\text{range}(A)$ such that the residual is orthogonal (perpendicular) to $\text{range}(A^T)$. 

Peter Blomgren, ⟨blomgren.peter@gmail.com⟩

Linear Algebra Introduction / Review — (17/25)

— (18/25)

— (19/25)

— (20/25)
An invertible or nonsingular matrix $A$ is a square matrix of full rank.

The $m$ columns of an invertible matrix form a basis for the whole space $\mathbb{R}^m$ (or $\mathbb{C}^m$) — any vector $\vec{x} \in \mathbb{R}^m$ can be expressed as a unique linear combination of the columns of $A$.

In particular we can express the unit vector $\vec{e}_j$ (which has a 1 in position $j$ and zeros in all other positions):

$$\vec{e}_j = \sum_{i=1}^{m} z_{ij} \vec{a}_i, \iff \vec{e}_j = A\vec{z}_j$$

If we play this game for $j = 1 \ldots m$, we get

$$\begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \ldots & \vec{e}_m \end{bmatrix} = A \begin{bmatrix} \vec{z}_1 & \vec{z}_2 & \ldots & \vec{z}_m \end{bmatrix}$$

We have

$$I_{m \times m} = A \cdot Z$$

The $m \times m$ matrix $I_{m \times m}$ which has ones on the diagonal and zeros everywhere else is the identity matrix.

The matrix $Z$ is the inverse of $A$.

Any square nonsingular matrix $A$ has a unique inverse, written $A^{-1}$, which satisfies

$$A \cdot A^{-1} = A^{-1} \cdot A = I$$

For a matrix $A \in \mathbb{C}^{m \times m}$ the following are equivalent

- $A$ has an inverse $A^{-1}$
- $\text{rank}(A) = m$
- $\text{range}(A) = \mathbb{C}^m$
- $\text{null}(A) = \{ \vec{0} \}$
- 0 is not an eigenvalue of $A$
- 0 is not a singular value of $A$
- $\det(A) \neq 0$

Note: The determinant is rarely useful in numerical algorithms — it is usually too expensive to compute.
Figure: Note that the illustration is not necessarily complete. In particular, recall Newton’s 3rd Law of Motion: *When one body exerts a force on a second body, the second body simultaneously exerts a force equal in magnitude and opposite in direction to that of the first body.*