18-819F: Introduction to Quantum Computing
47-779/47-785: Quantum Integer Programming & Quantum Machine Learning

Review of Linear Algebra I

Lecture 01

2022.08.31.

Agenda

- Review of linear algebra
  - Ordinary vector spaces
  - Inner product
  - Hilbert space
- Matrices
  - Decomposition of matrices
  - Quadratic form
  - Singular value decomposition
  - Gram-Schmidt Orthonormalization
  - Solution of linear systems of equations

Vectors in Euclidean Space

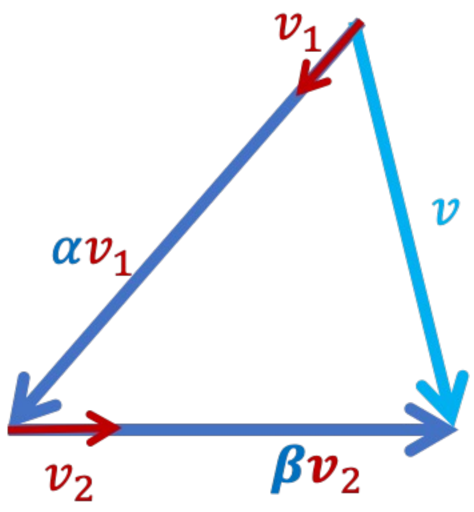
- In ordinary Euclidean space, vectors are mathematical objects that have magnitude and direction; they can be combined by **addition** and scaled by **multiplying** by a number (scalar) in $\mathbb{R}$;

- We can add and scale vectors as illustrated in the graphic on the right;

- The sum of the two vectors in 2D space is given as

$$\boldsymbol{v} = \alpha\boldsymbol{v_1} + \beta\boldsymbol{v_2} \text{ Eqn. (1.1);}$$

- The vector $\boldsymbol{v}$ is a **linear combination** (superposition) of the two vectors $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$, which are considered unit measures of distance along the appropriate directions; α and β are scaling factors (scalars);
- Clearly any arbitrary vector v in the 2D plane can be written as a linear combination (superposition) of $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ for any arbitrary scalars α and β, which are just numbers in $\mathbb{R}^2$;
- A collection of ***all*** vectors formed for arbitrary α and β constitute a set of vectors in a real ***vector space in 2D*** spanned by the ***basis*** vectors $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$.



Inner Product

- The inner (or scalar) product of two vectors $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ is defined as $\boldsymbol{v_1}.\boldsymbol{v_2} = \|v_1\|\|v_2\| \cos\vartheta$, where $|\boldsymbol{v_1}|$ and $|\boldsymbol{v_1}|$ are the lengths of the vectors $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$, and ϑ is the angle between them;

- If v_1 and v_2 are unit vectors, then $v_1 = v_2 = 1$ and if the angle ϑ between them is 0, then the inner product $\boldsymbol{v_1}.\boldsymbol{v_2} = \|v_1\|\|v_2\| \cos 0^o = 1 = \boldsymbol{v_1}.\boldsymbol{v_1} = \boldsymbol{v_2}.\boldsymbol{v_2}$. This is a condition for parallel vectors;

- Now suppose the angle ϑ between the unit vectors is 90^o; then $\boldsymbol{v_1}.\boldsymbol{v_2} = \|v_1\|\|v_2\| \cos 90^o = 0$. This is a condition for ***orthogonality*** of the unit vectors;

- Orthogonal unit vectors, as defined above, can be the ***basis vectors*** that span a 2D vector space. Any vector in this 2D space can be written, as before, as $v = \alpha v_1 + \beta v_2$ but this time v_1 and v_2 are orthogonal unit (***orthonormal***) vectors, and the scalars α and β provide an answer to the question "how much" of each unit vector is required in this linear combination.

Projection and inner product

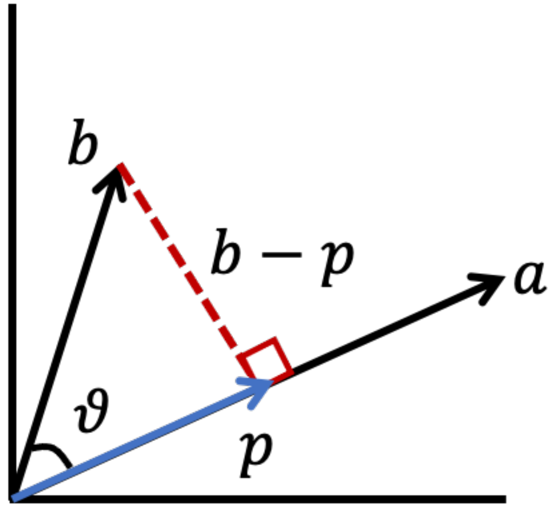
- The projection of vector $\boldsymbol{b}$ onto vector $\boldsymbol{a}$ is the vector

$\boldsymbol{p} = \gamma \boldsymbol{a}$;

- Vector $\boldsymbol{b} - \boldsymbol{p}$ is perpendicular to vector $\boldsymbol{a}$; this means $\boldsymbol{a}.(\boldsymbol{b} - \boldsymbol{\gamma a}) = 0$ or $\boldsymbol{a}.\boldsymbol{b} - \gamma \boldsymbol{a}.\boldsymbol{a} = 0$;
- We deduce that $\gamma = \frac{a.b}{a.a} = \frac{a^T.b}{a^T.a}$; note a^T is row vector;
- Projection of $\boldsymbol{b}$ onto $\boldsymbol{a}$ is therefore $p = \frac{a^T b}{a^T a} a$;
- From $p = \boldsymbol{a}\gamma = \boldsymbol{a}\frac{a^T b}{a^T a} = \frac{a a^T}{a^T a}\boldsymbol{b} = P\boldsymbol{b}$

$$\Longrightarrow \text{Projection matrix } P = \frac{a a^T}{a^T a}\ ;$$

- NB: aa^T is the outer product of vector a, which is a matrix (more on this later).



Higher Dimensions

- In ordinary 3D Euclidian space, one typically denotes orthonormal unit vectors as $\boldsymbol{i}, \boldsymbol{j},$ and $\boldsymbol{k}$; we can generalize these to 4, 5, and higher dimensional spaces and write them as $u_1, u_2, u_3 \ldots . u_n$, where the parallelism and orthogonality conditions become

$$\text{Eqn. (1.2)} \begin{cases} \boldsymbol{u_n}.\boldsymbol{u_m} = 1, n = m \\ \boldsymbol{u_n}.\boldsymbol{u_m} = 0, n \neq m \end{cases};$$

- In generalized n-dimensional "space" any vector $\boldsymbol{u}$ can then be written with c_n a scalar, and u_n an orthonormal vector, as

$$\boldsymbol{u} = c_1 u_1 + c_2 u_2 + \cdots + c_n u_n = \sum_n c_n u_n \quad \text{Eqn. (1.2);}$$

- By orthonormality we can determine "how much" of each basis vector contributes to $\boldsymbol{u}$, by the operation

$$c_n = \boldsymbol{u_n}.\boldsymbol{u} \quad \text{Eqn. (1.3);}$$

- From here on, the notion of a "vector" and a "vector space" as geometric constructions lose their meaning as it becomes impossible to visualize the space;

- All the usual mathematical operations on vectors, however, still have meaning on what is now an ***abstract*** vector space.

Defining a Vector Space

- From what we now know, it is apparent that a vector space, V, is a non-empty set with vector elements u, v, for which the operations of addition and multiplication are defined in the following way:

- (1) $w = u + v$, with $w \in V$;
- (2) For a scalar (number) $\alpha \in \mathbb{R}, \alpha u \in V$;

<u>A vector space also has the following properties</u>:

- For vectors u, v, addition is commutative: $u + v = v + u$;
- For vectors u, v, w, addition is associative: $(u + v) + w = u + (v + w)$;
- A vector space V has a zero vector, 0 such that $u + 0 = 0 + u$;
- For v in vector space V, there is an additive inverse $-v$ such that $v + (-v) = 0$.

Complex Vectors

- If we decide to allow the multipliers α and β in 2D and c_n in nD space to be complex numbers in $\mathbf{C}^n$, the concept of linear combination (superposition) and *all* the other properties of vectors still hold true in this new abstract vector space;

- Given a complex vector in 2D space: $\boldsymbol{v} = (2 + j)\boldsymbol{u_1} + (4 + 2j)\boldsymbol{u_2}$ we can use the concept of ***inner product*** to define the length of the vector $\boldsymbol{v}$, which must be a real number given by

$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}^*.\boldsymbol{v}} = \sqrt{29} \quad \text{Eqn. (1.4);}$$

- Note that in Eqn. (1.4), we have used the fact that $u_1^*.u_2 = u_2^*.u_1 = 0$ because of the orthonormality of the basis vectors, and we have further used the fact that $u_1^*.u_1 = u_2^*.u_2 = 1$;
- The result above is a consequence of Eqn. (1.4) which makes evident the validity of ***complex basis*** vectors.

- The concept of vectors and vector spaces can be generalized to complex vectors in abstract vector spaces of arbitrary dimensions;

- If ***objects*** can be called vectors and form a vector space, then the rules for ordinary vectors carry over.

Inner Product of complex vectors

- If we write the inner product of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ as $(\boldsymbol{u},\boldsymbol{v})$, then for complex numbers α and β the following is true

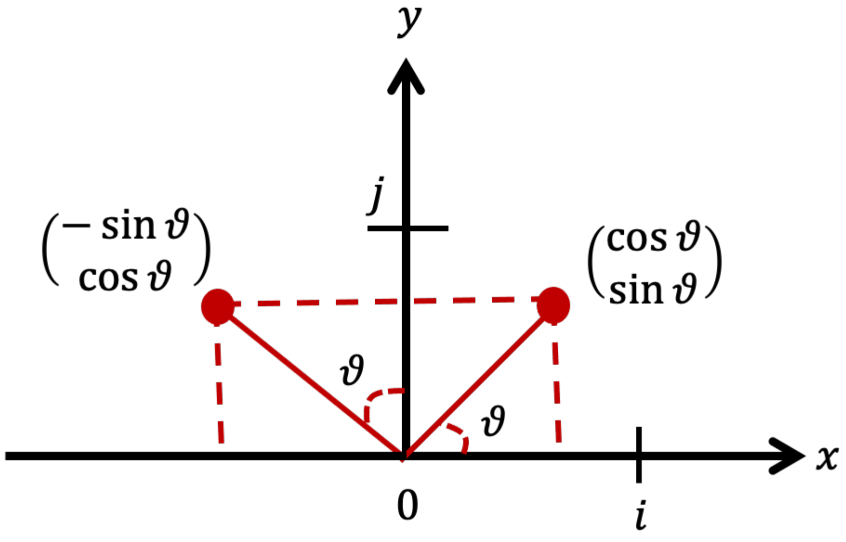
$$\text{Eqn. (1.5)}\begin{cases} (\boldsymbol{u},\boldsymbol{v}) = \text{real number} \\ (u,u) \geq 0 \\ (u,v)^* = (v,u) \\ (\boldsymbol{w}, \alpha\boldsymbol{u} + \beta v) = \alpha(w,u) + \beta(w,v) \end{cases};$$

- The definition of the inner product with a bracket was what motivated the Dirac <bra| and |ket> notation for vectors. What goes inside $<|$ is the complex transpose of the vector, and what goes inside $|\;>$ is the column vector;
- Given the ket $|v> = \begin{pmatrix} j \\ 2 \end{pmatrix}$, the bra is given by $< v| = (-j \quad 2)$; The inner product or the bra-ket is then given by: $< v|v> = (-j \quad 2)\begin{pmatrix} j \\ 2 \end{pmatrix} = (-j)(j) + (2)(2) = 5$;
- We will extensively discuss kets and bras later.

Hilbert Space

- Complex vectors, like the ordinary vectors we have discussed, can also form a set, which is a complex vector space in which addition and multiplication are defined;
- Unlike ordinary Euclidean space, a complex vector space is allowed to be infinite dimensional;
- In addition to the foregoing conditions, when there is a properly defined inner product for the *infinite dimensional*, complex vector space, we call it *Hilbert Space*;
- Hilbert space is convenient for discussing and describing the properties of quantum particles (and hence quantum information science, including computing and communication);

Operations (transformations) on vectors



- We can transform one vector to another by performing an operation or a transformation on it;
- In two-dimensional (x, y) space, the operator is a 2×2 matrix;
- An anti-clockwise rotation of the (i, j) orthonormal coordinate point, for example, by an angle of ϑ is indicated on the graphic on the right;
- After rotation, new coordinates for the unit vectors become $(\cos\vartheta, \sin\vartheta)$ for i and $(-\sin\vartheta, \cos\vartheta)$ for j;
- The rotation operator in matrix form is:

$$R(\vartheta) = \begin{bmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{bmatrix} \quad \text{Eqn. (2.5)}$$

Generalized vectors

- A general vector is an ordered array of n numbers in one dimension that may represent properties of a system; the list of numbers can be real in $\mathbb{R}^n$ or complex in C^n;

- These numbers could represent, for examples, the velocity, momentum, and kinetic energy of a tennis ball hit by Roger Federer at the 2021 US Tennis Open;

- A general abstract vector can be written as

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix};$$

- The transpose of the vector v is written as $v^T = [v_1, v_2, \dots, v_n]$ for real components;
- All the rules for geometric vectors apply equally well to this abstract kind of vector.

Vectors as Linear Combination of Orthonormal Basis Vectors

- In a conventional 2D vector spaces, when we write a vector such as $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, an implicit assumption is that this vector is written in the standard *orthogonal unit vector coordinate system*, which technically is called the *standard basis;*
- In 2D vector space, the standard basis is: $\hat{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\hat{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, where

$$(\hat{u}_1, \hat{u}_2) = \hat{u}_1^T \hat{u}_2 = < \hat{u}_1 | \hat{u}_2 > = 0;$$

- Any vector $v = \begin{bmatrix} v_1 \\ v_1 \end{bmatrix}$ can be written as linear combination of the standard basis, thus

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = v_1 \hat{u}_1 + v_2 \hat{u}_2;$$

- The concept of linear combination can be extended to $3D$…and nD vector spaces, thus

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + v_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Matrices

- In 2D space, we define a matrix as an ordered 2×2 array of numbers that transform abstract vectors to other vectors; we can extend this to a higher $n \times n$ dimension to define an $n \times n$ array;

- A matrix A can therefore be written as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ Eqn. (1.6);}$$

- If the matrix elements A_{mn} are real, then $A \in \mathbb{R}^{m \times n}$;

- We define the transpose of real matrix A as the mirror image along the main diagonal;

- If $A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \implies A^T = \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}$ Eqn. (1.7);

Complex Matrices

- If the matrix M has complex matrix elements, for example,

$$M = \begin{bmatrix} 2 & j \\ 0 & 0 \\ e^{j(\pi/4)} & 3j \end{bmatrix} \text{ then the transpose is } M^T = \begin{bmatrix} 2 & 0 & e^{j(\pi/4)} \\ j & 0 & 3j \end{bmatrix} \text{ Eqn. (1.8);}$$

- A useful version of the matrix of M is the ***conjugate transpose*** written as

$$M^\dagger = \begin{bmatrix} 2 & 0 & e^{-j(\pi/4)} \\ -j & 0 & -3j \end{bmatrix} \text{ Eqn. (1.9);}$$

- Complex matrices can transform vectors in the same way that real matrices do.

Operations with Matrices

- The matrix A linearly transforms the vector x to a new vector b; this is written as:

$$Ax = b$$

- Or alternatively as $\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ Eqn. (1.10);

- This is one of the most important operations in machine learning and quantum mechanics;
- For known A and b, one can solve for the unknown x in the following manner:

$$A^{-1}Ax = A^{-1}b \quad \text{Eqn. (1.11).}$$

Inverse and Identity Matrices

- An identity matrix does not change the value of a vector; thus, if I is an identity matrix, where $I \in \mathbb{R}^{n\times n}$ then for all vectors $x \in \mathbb{R}^n$ $Ix = x$;
- An $n\times n$ identity matrix $I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$ Eqn. (1.12)
- An ***inverse of a square matrix*** is then defined through $\boldsymbol{A}^{-1}\boldsymbol{A} = \mathbf{I}$;
- Solving for an unknown vector x in a linear system of equations, $Ax = b$, means
$$A^{-1}Ax = A^{-1}b \Rightarrow Ix = A^{-1}b \Rightarrow x = A^{-1}b \quad \text{Eqn. (1.13);}$$
- Our task then becomes that of finding the matrix A^{-1}, if it exists.

Decomposition of Matrices

- One can decompose a matrix into factors that expose more universal properties about it;
- One common decomposition of a matrix A is to find the eigenvectors and eigenvalues;
- An eigenvector v of a matrix A is a vector that, when multiplied by A, only scales the value of v: thus $Av = \lambda v$, where the scalar λ is the eigenvalue;
- For $Av = \lambda v \implies (A - \lambda I)v = 0$; a solution for v exists if $|A - I\lambda| = 0$;
- One can factor the polynomial that results from above to get

$$|A - \lambda I| = (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n);$$

- The roots of the polynomial $\lambda = \lambda_1 \dots \lambda = \lambda_n$ are the eigenvalues of A.

Eigen-decomposition

- A matrix A can be written as: $A = V\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \vdots & 0\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}V^{-1}$ Eqn. (1.14)

where V is a matrix composed of all the independent eigenvectors of A and $\lambda_1 \dots \lambda_2$ are the eigenvalues of A;

- The process of assembling the eigenvector and eigenvalues of A as shown above is known as the eigen-decomposition of A.

- If A is symmetric ($A = A^T$), then $A = V \wedge V^T$ where $V = \begin{bmatrix}\vdots & \vdots & \vdots & \\ v_1^{(1)} & v_2^{(2)} & v_3^{(3)} & \dots\\ \vdots & \vdots & \vdots & \end{bmatrix}$ is an orthogonal matrix comprised of all the eigenvectors of A and $\wedge$ is a diagonal matrix comprised of the eigenvalues of A; note that orthogonality of V means $\left(v_2^{(1)}\right)^T v_2^{(2)} = 0$ for each pair of columns of V.

Spectral Decomposition

- As we stated, the task of eigen-decomposition is to compute the eigenvalues and eigenvectors of A;

- If A is square ($n \times n$) and symmetric ($A = A^T$), then one can write

$$AV = V\Lambda,$$

where V is matrix comprised of columns of the eigenvectors of A, and Λ is a diagonal matrix comprised of the eigenvalues of A;

- The set of eigenvalues is called the spectrum of A, hence the spectral decomposition.

Spectral Decomposition

- One can reformulate A as eigenvalue-eigenvector pairs to arrive at

$$A = \sum_{i=1}^{k} \lambda_i e_i e_i^T = B\Lambda B^T \quad \text{Eqn. (1.15)}$$

Where λ_i are the eigenvalues and e_i the normalized eigenvectors; the matrix B is constructed from the columns of the normalized eigenvectors; furthermore, $B^T B = BB^T = I$, and Λ is a diagonal matrix of the eigenvalues;

- The inverse of A can be computed as

$$A^{-1} = B\Lambda^{-1}B^T = \sum_{i=1}^{k} \frac{1}{\lambda_i} e_i e_i^T \quad \text{Eqn. (1.16).}$$

Quadratic Form

- Quadratic form is a generalization of the inner product of a vector with itself;

- Simply defined, a quadratic form Q on $\mathbb{R}^n$ is a function $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ of the form

$$Q(x) = x^T A x \text{ where } A \in \mathbb{R}^{n\times n} \text{ is a symmetric matrix of the quadratic form;}$$

- Consider $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$: $\mathrm{Q(x)} = [x_1\ x_2]\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2 = x^T I x = x^T x = x.x;$
- Now consider $A = \begin{bmatrix} 5 & -3 \\ -3 & 4 \end{bmatrix}$: $Q(x) = [x_1\ x_2]\begin{bmatrix} 5 & -3 \\ -3 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 5x_1^2 + 4x_2^2 - 6x_1x_2;$
- Notice that the diagonal elements of matrix A are the coefficients of the quadratic terms of the polynomial and the off-diagonal elements are one-half of the sum of the cross-terms of the polynomial (when A is symmetric as required in our definition above).

Writing a polynomial in quadratic form

- Write the polynomial $Q(x) = -3x_1^2 + 4x_2^2 - 5x_3^2 - 6x_1x_2 + 4x_2x_3$ in quadratic form;

- From the previous slide, we know that the diagonal elements of the matrix A must be the coefficients of the quadratic terms of the polynomial. The polynomial has 3 variables for the vector x, implying a 3×3 symmetric matrix; off-diagonal terms are symmetric and are one-half of the cross-terms; thus

- $Q(x) = x^T A x = [x_1\ x_2\ x_3]\begin{bmatrix} -3 & -3 & 0 \\ -3 & 4 & 2 \\ 0 & 2 & -5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = [x_1\ x_2\ x_3]\begin{bmatrix} -3x_1 - 3x_2 \\ -3x_1 + 4x_2 + 2x_3 \\ 2x_2 - 5x_3 \end{bmatrix}$ Eqn. (1.17);

- We simplify this to:

- $$\begin{aligned} Q(x) &= x_1(-3x_1 - 3x_2) + x_2(-3x_1 + 4x_2 + 2x_3) + x_3(2x_2 - 5x_3) \\ &= -3x_1^2 - 3x_1x_2 - 3x_1x_2 + 4x_2^2 + 2x_2x_3 + 2x_2x_3 - 5x_3^2 \qquad \text{Eqn. (1.18).} \\ &= -3x_1^2 + 4x_2^2 - 5x_3^2 - 6x_1x_2 + 4x_2x_3 \quad \text{QED} \end{aligned}$$

Quadratic Optimization

- Suppose we are given the variables x_1, x_2, x_3, which model a binary yes/no (or 1/0) decision. Now suppose a combination of these variables in a non-trivial decision process is written as the function

$$Q(x) = -2x_1 + x_2 + 4x_3 + 2x_1x_2 + 4x_1x_3 + 6x_2x_3 \text{ Eqn. (1.19);}$$

- The objective is to minimize the function $Q(x)$, i.e., find the yes/no decision values for each x_i; there might or might not be constraints imposed on some of the x_i;

- Since the variables are binary, it is true that $x_i = x_i^2$, which means the terms linear in x_i in Eqn. (1.19) can be replaced by quadratic versions so that (1.19) becomes

$$Q(x) = -2x_1^2 + x_2^2 + 4x_3^2 + 2x_1x_2 + 4x_1x_3 + 6x_2x_3 \text{ Eqn. (1.20);}$$

$$\Longrightarrow Q(x) = x^T Ax = [x_1\ x_2\ x_3] \begin{bmatrix} -2 & 1 & 2 \\ 1 & 1 & 3 \\ 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \text{ Eqn. 1.21);}$$

Singular Value Decomposition (SVD)

- SVD can be viewed as (i) a method for transforming correlated variables into uncorrelation ones to better expose relationships among the original dataset; (ii) a method for identifying dimensions along which data varies the most; or (iii) a method to finding an approximation to an original dataset with fewer dimensions;

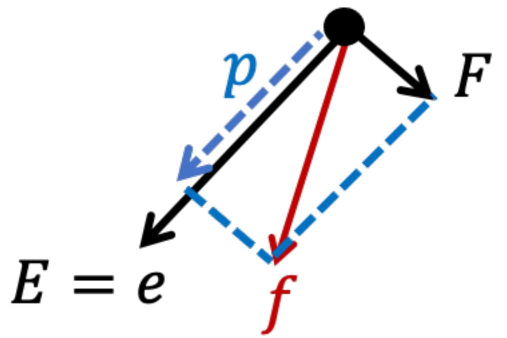
- Formally, SVD is a method for decomposing a matrix A_{mn} into a product of 3 simpler, but special matrices, U_{mm}, D_{mn}, V_{nn}^T, (note: D_{mn} is a diagonal matrix of singular values):

$$A_{mn} = U_{mm} D_{mn} V_{nn}^T \quad \text{Eqn. (1.22);}$$

- The matrix U is comprised of the orthonormal eigenvectors of AA^T and V is comprised of the orthonormal eigenvectors of $A^T A$; D is a diagonal matrix comprised of the square roots of the eigenvalues of U or V in descending order; furthermore, $U^T U = I, V^T V = I$;

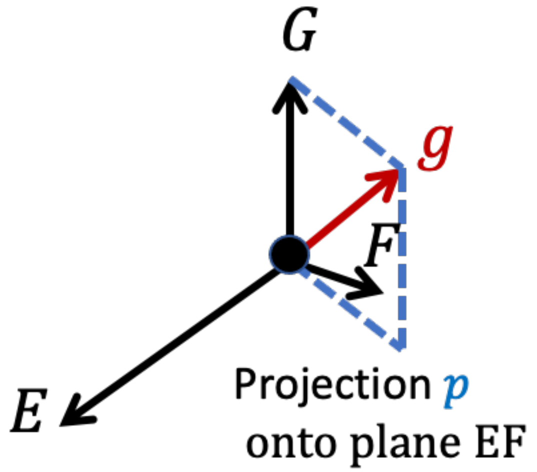
Gram-Schmidt Orthogonal Vector Construction

- Given three independent vectors ***e, f, g,*** we can construct three orthogonal vectors ***E, F, G;*** the process is illustrated in the graphic on the right (where we only show ***e*** *and* ***f***);



- Choose $\boldsymbol{E} = \boldsymbol{e}$; since ***F*** must be perpendicular to ***E,*** we take the projection of $\boldsymbol{f}$ on $\boldsymbol{E}$; this is $\boldsymbol{p} = \frac{\boldsymbol{E}^T\boldsymbol{f}}{\boldsymbol{E}^T\boldsymbol{E}}\boldsymbol{E}$; to get F, we must subtract the projection $\boldsymbol{p}$ from $\boldsymbol{f}$;
- Thus $\boldsymbol{F} = \boldsymbol{f} - \frac{\boldsymbol{E}^T\boldsymbol{f}}{\boldsymbol{E}^T\boldsymbol{E}}\boldsymbol{E}$ Eqn. (1.23);
- In the same manner, we get the 3rd perpendicular vector

$$G = g - \frac{E^T g}{E^T E}E - \frac{F^T g}{F^T F}F \quad \text{Eqn. (1.24).}$$



Gram-Schmidt Orthonormalization

- Once the orthogonal vectors E, F, G have been created, they are normalized in the following manner:

$$\hat{e} = \frac{E}{||E||}, \hat{f} = \frac{F}{||F||}, \text{ and } \hat{g} = \frac{G}{||G||};$$

- Given the vectors $a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $b = \begin{bmatrix} 4 \\ 0 \\ -4 \end{bmatrix}$, and $c = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$, construct orthogonal vectors from them;

- Accept $A = a$, then $A^T A = 2$, thus $B = b - \frac{A^T b}{A^T A} A = b - \frac{4}{2} A = \begin{bmatrix} 2 \\ 2 \\ -4 \end{bmatrix}$;

- Finally, $C = c - \frac{A^T c}{A^T A} A - \frac{B^T c}{B^T B} B = c - \frac{2}{2} A + \frac{1}{6} B = \frac{1}{6} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Gram-Schmidt Orthonormalization

- The orthogonal vectors A, B, C must now be normalized by dividing by their lengths $||A|| = \sqrt{2}$, $||B|| = 2\sqrt{3}$, $||C|| = 1/2\sqrt{3}$;

- Thus $\hat{a} = \frac{A}{||A||} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $\hat{b} = \frac{B}{||B||} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ and $\hat{c} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$;

- Determination of orthogonal vectors from linearly independent vectors is important for SVD;

- Normalized orthogonal vectors are the columns of the matrices that form the U and V matrices in SVD.

Numerical Example of SVD

- Perform the SVD of $A = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 2 & 1 \end{bmatrix}$;
- $AA^T = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix} \Longrightarrow$ eigenvalues of AA^T are $\lambda_1 = 5, \lambda_2 = 7$, and eigenvectors are $w_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $w_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$;
- By the Gram-Schmidt method, we calculate the normal vectors $u_1 = w_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and

$$u_2 = w_2 - \frac{u_1^T w_2}{u_1^T u_1} u_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{[1\ -1]\begin{bmatrix} 1 \\ 1 \end{bmatrix}}{[1\ -1]\begin{bmatrix} 1 \\ -1 \end{bmatrix}} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$; the normalized vectors are

$$\hat{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \text{ and } \hat{u}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \Longrightarrow \text{Matrix } U = [\hat{u}_1 \quad \hat{u}_2] = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix};$$

Numerical Example of SVD

- We now calculate $A^TA = \begin{bmatrix} 2 & -1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ -1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 1 \\ 0 & 5 & 3 \\ 1 & 3 & 2 \end{bmatrix}$

$\Longrightarrow$ Eigenvalues of A^TA are $\lambda_1 = 0, \lambda_2 = 5, \lambda_3 = 7$ and eigenvector for $\lambda_3 = 7$ is $z_3 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ and for $\lambda_2 = 5$, it is $z_2 = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}$; and for $\lambda_1 = 0$, it is $z_1 = \begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix}$;

Numerical Example of SVD

- By the Gram-Schmidt process, we calculate the orthogonal vector $v_1 = z_1 = \begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix}$, then

$v_2 = z_2 - \frac{v_1^T z_2}{v_1^T v_1} v_1 = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} - \frac{0}{35} \cdot \begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}$, and

$$v_3 = z_3 - \frac{v_1^T z_3}{v_1^T v_1} v_1 - \frac{v_2^T z_3}{v_2^T v_2} v_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} - 0 - 0 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix};$$

Orthonormalized vectors are therefore $\hat{v}_1 = \frac{1}{\sqrt{35}} \begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix}, \hat{v}_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}, \hat{v}_3 = \frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$

Numerical Example of SVD

- The matrix V of the SVD is therefore

$$V = [\hat{v}_1 \quad \hat{v}_2 \quad \hat{v}_3] = \begin{bmatrix} 1/\sqrt{35} & 3/\sqrt{10} & 1/\sqrt{14} \\ 3/\sqrt{35} & -1/\sqrt{10} & 3/\sqrt{14} \\ -5/\sqrt{35} & 0/\sqrt{10} & 2/\sqrt{14} \end{bmatrix};$$

- One can now take the transpose of $V \longrightarrow V^T$;

- All the matrices for decomposition of A have now been determined; observe that

$$A_{23} = V_{22} D_{23} V_{33}^T$$

where $D_{23} = \begin{bmatrix} \sqrt{7} & 0 & 0 \\ 0 & \sqrt{5} & 0 \end{bmatrix}$.

SVD and Linear System of Equations

- We already stated that matrices are important for solving a system of linear equations of the form

$$Ax = b \quad \text{Eqn. (1.25)}$$

- In most cases, A is an $m \times n$ matrix and b is an $m \times 1$ vector (in the general non-geometric sense); if $m = n$, there is a good chance we can find the unknown vector x, in which case

$$A^{-1}Ax = A^{-1}b \Longrightarrow x = A^{-1}b \quad \text{Eqn. (1.26);}$$

- Another way of solving for x is to use the transpose of A, thus

$$A^T Ax = A^T b \Longrightarrow x = (A^T A)^{-1} A^T b \quad \text{Eqn. (1.27)}$$

- This approach is the least-squares sense of the solution with

$$A^{\dagger} = (A^T A)^{-1} A^T \quad \text{Eqn. (1.28),}$$

where $A^{\dagger}$ is defined as the Moore-Penrose pseudo-inverse of A.

Moore-Penrose Pseudoinverse

- Together with SVD, the Moore-Penrose pseudoinverse permits solution of linear system of equations where the matrix maybe (i) square but singular or degenerate, (ii) underdetermined (fewer equations than unknowns), or (iii) overdetermined (more equations than unknowns);

- Consider the matrix A below which is for an overdetermined system

$$A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \\ 0 & 0 \end{bmatrix} \Rightarrow AA^T = \begin{bmatrix} 8 & 8 & 0 \\ 8 & 8 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ with eigenvalues of } \lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 16;$$

$$A^T A = \begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix} \text{ with eigenvalues of } \lambda_1 = 0, \lambda_2 = 16;$$

- Eigenvectors of AA^T are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, and for $A^T A$ they are $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

SVD and Moore-Penrose Pseudoinverse

- SVD of the previous matrix A using the calculated eigenvalues and eigenvectors is

$$A = U\Lambda V^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0; & 0 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^T = \begin{bmatrix} 2 & 2 \\ 2 & 2 \\ 0 & 0 \end{bmatrix};$$

- It is now relatively easy to calculate the pseudoinverse of A so the linear system $Ax = b$ can be solved as indicated below

$$x = (A^T A)^{-1} A^T b = U\Lambda^{-1} V^T b \quad \text{Eqn. (1.29);}$$

- The U and V^T can be directly read off from above and the reciprocal of Λ^{-1} is trivial to obtain since Λ is available from above.

SVD Partitioning of Information stored in a Matrix

- Matrix M stores viewership information of Netflix movies by a group of friends who have either seen (1) or not seen (0) a movie;
- Information contained in M can be analyzed by Netflix to create a recommender algorithm using SVD;
- Many or large singular values in D, which are the square roots of eigenvalues of U or V, tell us about the strength of interactions between what is stored in U and V.

Netflix Movies Watched

	M_1	M_2	M_3	M_4
Alice	1	0	0	1
Bob	1	1	1	1
Chuck	1	0	1	1
Dave	1	1	0	0

Matrix M

Rank of M is k

D is a diagonal matrix of square roots of eigenvalues of U or V^T

$$\underset{m\times n}{M} \qquad \underset{m\times k}{U} \qquad \underset{\mathrm{k}\times\mathrm{k}}{D} \qquad \underset{\mathrm{k}\times m}{V^T}$$

$$\begin{bmatrix} M & M & M & M \\ M & M & M & M \\ M & M & M & M \end{bmatrix} = \underbrace{\begin{bmatrix} U & U & U \\ U & U & U \\ U & U & U \end{bmatrix}} \overbrace{\begin{bmatrix} D & & \\ & D & \\ & & D \end{bmatrix}} \underbrace{\begin{bmatrix} V^* & V^* & V^* & V^* \\ V^* & V^* & V^* & V^* \\ V^* & V^* & V^* & V^* \end{bmatrix}}$$

Columns of U are orthonormal

Rows of V^T are orthonormal

Data Partitioning and Compression by SVD

- In the previous Netflix example of the data stored in the matrix M and subsequent analysis by SVD: $M = UDV^T$, useful information about the relationship of viewership to movies is contained in matrix U, which tells us about the viewers and matrix V which contains information about the movies;

- The diagonal matrix D, comprised of the singular values, indicates the level of importance of the interaction between viewers and movies; many singular values mean there is lots of interaction between viewers and movies;

- Since D is typically of low-dimensional space (telling us the rank of matrix M), it provides a way to compress the data contained in M.

- The graphic illustration provides additional information about the meaning of SVD;

Role of SVD in Linear Transformations

- The role of an $m \times n$ matrix M is to perform linear transformations on points by taking them from $\mathbb{R}^n$ to $\mathbb{R}^m$; the matrix $M = U\Lambda V^T$ encodes rotations and rescaling;

- SVD essentially factors out the transformations into V^T which is a rotation, Λ a rescaling along the principal axes, and U which is another rotation;

- SVD can be used to compress data in the matrix M; for example, when M is comprised of pixel data of an image; without SVD, the matrix M requires mn storage values; by decomposing M into U_{mr}, Λ_r, and V_{rn}, the values required become $mr + r + nr$; since usually $r < m, n$, it is evident that the new storage requirement $mr + r + nr < mn$.

Summary

- Reviewed linear algebra
  - Ordinary vector spaces
  - Inner product
  - Hilbert space
- Reviewed matrices and introductions some applications of decomposed matrices
  - Decomposition of matrices
  - Quadratic form
  - Singular value decomposition
  - Gram-Schmidt Orthonormalization
  - Solution of linear systems of equations