Matrix Factorizations and Applications

We've taken the world apart but we have no idea what to do with the pieces.

LU factorization ($A = LU$ where A is an $n \times n$ *square matrix*)

Consider the following set of equations:

$$a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n = b_1 \ldots\ldots\ldots\ldots (i)$$
$$a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n = b_2 \ldots\ldots\ldots\ldots (ii)$$
$$\cdot \qquad \cdot = \cdot$$
$$\cdot \qquad \cdot = \cdot$$
$$\cdot \qquad \cdot = \cdot$$
$$a_{n1}x_1 + a_{n2}x_2 + \cdots a_{nn}x_n = b_n \ldots\ldots\ldots\ldots (\text{m}^{th} \text{ eq.})$$

With respect to the Gauss-elimination, the procedure to find the row-echelon form of the corresponding coefficient matrix, the following sequence of operations are performed $E_j : \ (E_j - m_{ji}E_i)$ which involves the calculation of the multipliers $m_{ji} := \frac{a_{ji}}{a_{ii}}$.

Further, the system of equations corresponding to the row-echelon form looks as follows:

$$a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + \cdots a_{1n}^{(1)}x_n = b_1^{(1)} \ldots\ldots\ldots\ldots (i)$$
$$0 + a_{22}^{(2)}x_2 + \cdots a_{2n}^{(2)}x_n = b_2^{(2)} \ldots\ldots\ldots\ldots (ii)$$
$$\cdot \qquad \cdot = \cdot$$
$$\cdot \qquad \cdot = \cdot$$
$$\cdot \qquad \cdot = \cdot$$
$$0 + 0 + 0 + \cdots a_{nn}^{(n)}x_n = b_n^{(n)} \ldots\ldots\ldots\ldots (\text{m}^{th} \text{ eq.})$$

Then the LU decomposition reads as follows: $A = LU = \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & 0 \\ m_{21} & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & 0 \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ m_{n1} & m_{n2} & \cdot & \cdot & m_{n(n-1)} & 1 \end{pmatrix} \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdot & \cdot & \cdot & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdot & \cdot & \cdot & a_{2n}^{(2)} \\ \cdot & 0 & \cdot & \cdot & & \cdot \\ \cdot & \cdot & 0 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & 0 & \cdot & a_{(n-1)n}^{(n-1} \\ 0 & \cdot & \cdot & \cdot & 0 & a_{nn}^{(n} \end{pmatrix}$.

Consequently,

$$A\mathbf{x} = \mathbf{b}$$
$$LU\mathbf{x} = \mathbf{b}$$
$$L\mathbf{y} = \mathbf{b} \qquad \text{where } \mathbf{y} = U\mathbf{x}$$

This entails that we can solve $L\mathbf{y} = \mathbf{b}$ first by *forward substitution* followed by solving for $\mathbf{x}$ in $U\mathbf{x} = \mathbf{y}$ by *backward substitution.*

Advantages of LU decomposition

Gauss elimination has a complexity of $O(n^3)$ while solving the system of equations by using the LU decomposition has a complexity of $O(n^2)$.

Reading Assignment: Complexity of Gauss-elimination and LU decomposition method:
https://235d9ee8-8e8c-4d7b-a842-264ad94cf102.filesusr.com/ugd/334434_5a3eab64a8b0442cabd729aa5defab45.pdf

Example 1: Solve the following system of linear equations by using LU factorization.

$$\begin{aligned} x_1 + x_2 + 0x_3 + 3x_4 &= 4 \\ 2x_1 + x_2 - x_3 + x_4 &= 1 \\ 3x_1 - x_2 - x_3 + 2x_4 &= -3 \\ -x_1 + 2x_2 + 3x_3 - x_4 &= 4 \end{aligned} \quad \text{i.e., } A\mathbf{x} = \mathbf{b}.$$

Soln: The following sequence of row operations reduces the above coefficient matrix A to the row-echelon form.

$$\begin{aligned} E_2 &: \ (E_2 - 2E_1) \\ E_3 &: \ (E_3 - 3E_1) \\ E_4 &: \ (E_4 - (-1)E_1) \\ E_3 &: \ (E_3 - 4E_2) \\ E_4 &: \ (E_4 - (-3)E_2) \end{aligned}$$

The row-echelon form of the coefficient matrix is given below.

$$\text{Row echelon form of } A = \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{pmatrix} = U$$

Now by inspection and following how the multipliers m_{ji}s constitute the L matrix, we have,

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{pmatrix}.$$

Now once the LU decomposition of the coefficient matrix A is accomplished, we can use this decomposition to solve any system of linear equations defined by the same coefficient matrix A (but different non-homogeneous "forcing" vector on the r.h.s.). This is where we leverage the most benefit from the LU factorization in terms of complexity.

Anyways, for the above system, we will first solve $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ -3 \\ 4 \end{pmatrix}$ by *forward substitution as follows:*

$$y_1 = 4, \;\; y_2 = 1 - 2y_1 = -7, \;\; y_3 = -3 - 3y_1 - 4y_2 = 13, \;\; y_4 = 4 + y_1 + 3y_2 = 8 - 21 = -13.$$

We can now solve the system $U\mathbf{x} = \mathbf{y}$, i.e. $\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 \\ -7 \\ 13 \\ -13 \end{pmatrix}$ by *backward substitution* and obtain the required solution:

$$x_4 = \frac{-13}{-13} = 1, \;\; x_3 = \frac{13 - 13 \times 1}{3} = 0, \;\; \text{etc...}$$

QR factorization

Coming up!

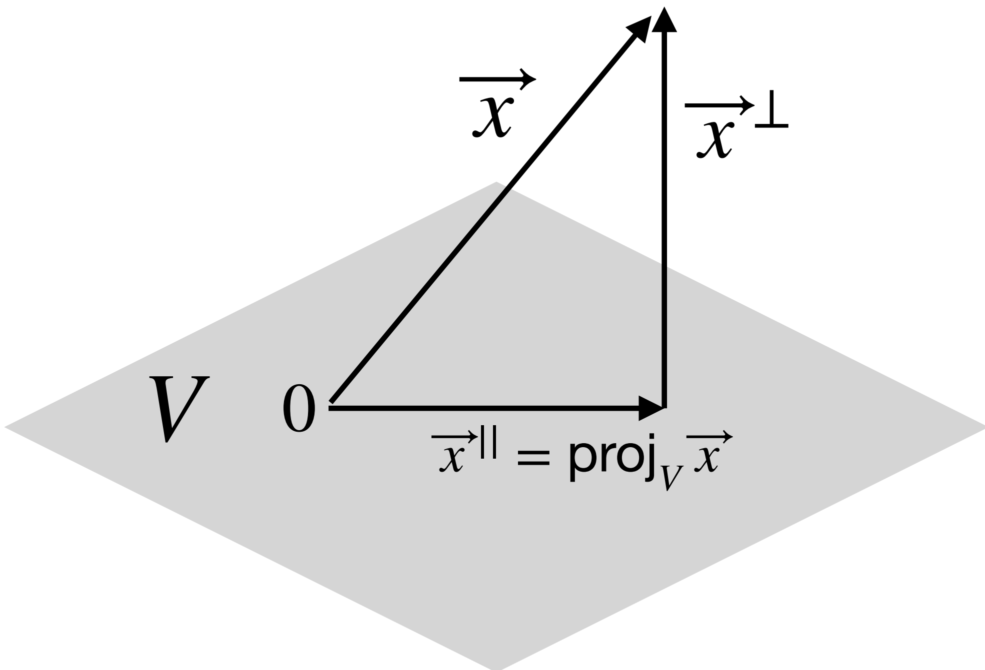
Orthogonal basis and Gram-Schmidt orthogonalization

Two vectors $\overrightarrow{u}_1$ and $\overrightarrow{u}_2$ are *orthogonal* if and only if $\langle \overrightarrow{u}_1, \overrightarrow{u}_2 \rangle = 0$.
The vectors $\overrightarrow{u}_1, \overrightarrow{u}_2, \cdots, \overrightarrow{u}_m \in \mathbb{R}^n$ are *orthonormal* if and only if $\langle \overrightarrow{u}_i, \overrightarrow{u}_j \rangle = \delta_{ij}$.

Example: The vectors $\overrightarrow{e}_1, \overrightarrow{e}_2, \ldots, \overrightarrow{e}_n \in \mathbb{R}^n$ are orthonormal.

Properties of orthonormal vectors:
1. Orthonormal vectors are (automatically) linearly independent.
2. Orthonormal vectors $\overrightarrow{u}_1, \overrightarrow{u}_2, \cdots, \overrightarrow{u}_n \in \mathbb{R}^n$ form a basis in $\mathbb{R}^n$.

The shaded area denoted by V in the figure below is an **infinite** plane through the origin.



Orthogonal projection and orthogonal complement:
Let $\vec{x} \in \mathbb{R}^n$ and a subspace V of $\mathbb{R}^n$. Then we can wrtie $\vec{x} = \vec{x}^{||} + \vec{x}^{\perp}$, where $\vec{x}^{||} \in V$ and $\vec{x}^{\perp} \in V^{\perp}$. The above representation is **unique**.

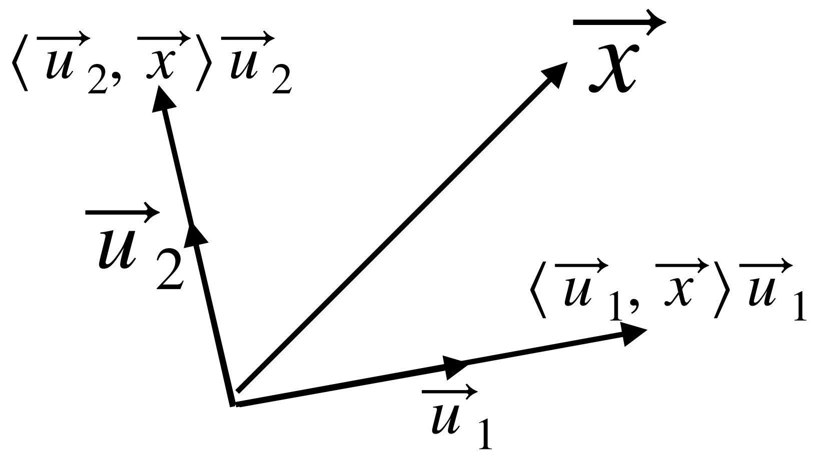
Here $V^{\perp} = \{\vec{x} \in \mathbb{R}^n : \langle \vec{v}, \vec{x} \rangle = 0, \ \ \forall \vec{v} \in V\}$. The transformation $T(\vec{x}) = \text{proj}_V \vec{x} = \vec{x}^{||}$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ is linear. $V^{\perp} = \text{Ker}(T)$.

How do we compute $\vec{x}^{||}$?
Consider an orthonormal basis of V: $\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_m \in V$ which is a subspace of $\mathbb{R}^n$. Then

$$\vec{x}^{||} = \langle \vec{u}_1, \vec{x} \rangle \vec{u}_1 + \cdots + \langle \vec{u}_m, \vec{x} \rangle \vec{u}_m; \quad \forall \vec{x} \in \mathbb{R}^n.$$

Consequently, consider an orthonormal basis of $\mathbb{R}^n$: $\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_n$. Then any $\vec{x} \in \mathbb{R}^n$,

$$\vec{x} = \langle \vec{u}_1, \vec{x} \rangle \vec{u}_1 + \cdots + \langle \vec{u}_n, \vec{x} \rangle \vec{u}_n.$$



Properties of orthogonal complement:

Consider a subspace $V \in \mathbb{R}^n$.

1. $V^\perp$ is a subspace of $\mathbb{R}^n$.
2. $V \cap V^\perp = \{\overrightarrow{0}\}$.
3. $\dim(V) + \dim(V^\perp) = n$.
4. $(V^\perp)^\perp = V$.

Example: Consider the subspace $V = Im(A)$ of $\mathbb{R}^4$, where $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$. Find $\overrightarrow{x}^{||}$ for $\overrightarrow{x} = \begin{pmatrix} 1 \\ 3 \\ 1 \\ 7 \end{pmatrix}$.

Solution: Recall that the column space of A is $Im(A)$. It can be easily checked that the column vectors of A are orthogonal by taking their scalar product. Thus we can construct an orthonormal basis of $Im(A)$. The basis vectors are: $\overrightarrow{u}_1 = \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{pmatrix}$ and $\overrightarrow{u}_2 = \begin{pmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{pmatrix}$.

Then $\overrightarrow{x}^{||} = \langle \overrightarrow{u}_1, \overrightarrow{x} \rangle \overrightarrow{u}_1 + \langle \overrightarrow{u}_2, \overrightarrow{x} \rangle \overrightarrow{u}_2 = 6\overrightarrow{u}_1 + 2\overrightarrow{u}_2 = \begin{pmatrix} 3 \\ 3 \\ 3 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 2 \\ 4 \end{pmatrix}$. In order to check that this answer is indeed correct, verify that $(\overrightarrow{x} - \overrightarrow{x}^{||}) \perp \overrightarrow{u}_1, \overrightarrow{u}_2$.

Why are orthonormal basis vectors useful?

1. We know that if we have some basis $\overrightarrow{v}_1, \overrightarrow{v}_2, \ldots, \overrightarrow{v}_n$ of an n-dimensional vector space W. Then any vector $\overrightarrow{x} \in W$ can be written as $\overrightarrow{x} = \alpha_1 \overrightarrow{v}_1 + \alpha_2 \overrightarrow{v}_2 + \cdots + \alpha_n \overrightarrow{v}_n$ (as a linear combination of the basis vectors) but there is no first-principles or convenient way of finding the unique coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ except by explicit guesswork calculations. Now instead if we have an orthonormal basis set $\overrightarrow{u}_1, \overrightarrow{u}_2, \ldots, \overrightarrow{u}_n$ then any vector can be written as a linear combination of this orthonormal basis set as follows:

   $\overrightarrow{x} = \beta_1 \overrightarrow{u}_1 + \beta_2 \overrightarrow{u}_2 + \cdots + \beta_n \overrightarrow{u}_n$ where the coefficients can now be uniquely determined as $\beta_i = \langle \overrightarrow{u}_i, \overrightarrow{x} \rangle, \quad \forall i = 1,2,...,n$

2. Orthogonality guarantees linear independence.

Why are orthogonal transformations useful?

1. Orthogonal transformations are metric preserving transformations, i.e. if $T : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal, then $||T(\overrightarrow{x})|| = ||\overrightarrow{x}||, \quad \forall \overrightarrow{x} \in \mathbb{R}^n$.[1]

2. Orthogonal transformations are angle preserving transformations for orthogonal vectors. If $\overrightarrow{u} \perp \overrightarrow{w}$, then $T(\overrightarrow{u}) \perp T(\overrightarrow{w})$.

[1] If $T(\overrightarrow{x}) = A\overrightarrow{x}$ is an orthogonal transformation, then we say that A is an orthogonal matrix.