Huda Nassar hnassar@purdue.edu
Notes on Matrix Factorizations
April 14, 2016

Most of these notes are based on my personal notes from CS 515 from Fall 2014 and personal experience with matrices. There might be a few missing details since this document was prepared to be used in a Matrix Factorizations Workshop that I offered on Monday April 14 2016 (the few missing details were discussed on the board with the attendees). I will keep adding more information and extra details to it whenever possible.

Outline

- A Brief Background
- Preliminaries
- Singular Value Decomposition
- LU factorization
- QR factorization
- Some applications

A Brief Background

I start this document by quoting this introduction from the chapter 39 from "50 Mathematical Ideas You Really Need to Know"

> This is the story of 'extraordinary algebra' - a revolution in mathematics which took place in the middle of the 19th century. Mathematicians had played with blocks of numbers for centuries, but the idea of treating blocks as a single number took off 150 years ago with a small group of mathematicians who recognized its potential.

The problem $\boldsymbol{A}\mathbf{x} = \mathbf{b}$ shows up in many research disciplines and has been studied extensively over the past ¿ 100 years. The point I intend to make in this section is that there is no right answer when it comes to matrix algebra. One question I learned to ask is, can I look at the matrix? I try to answer questions relating to its structure, sparsity and known previous knowledge about it. Figure 1 shows the flowchart that Matlab uses to solve a square linear system for example. Below is a brief historical timeline of SVD,LU, and QR factorizations.

- 1873 Eugenio Beltrami (SVD)
- 1974 Camille Jordan (SVD)
- 1889 James Joseph Sylvester (SVD)
- 1948 Alan Turing (LU)
- 1959 John Francis (QR)
- 1963 Carl Eckart and Gale Young (SVD)

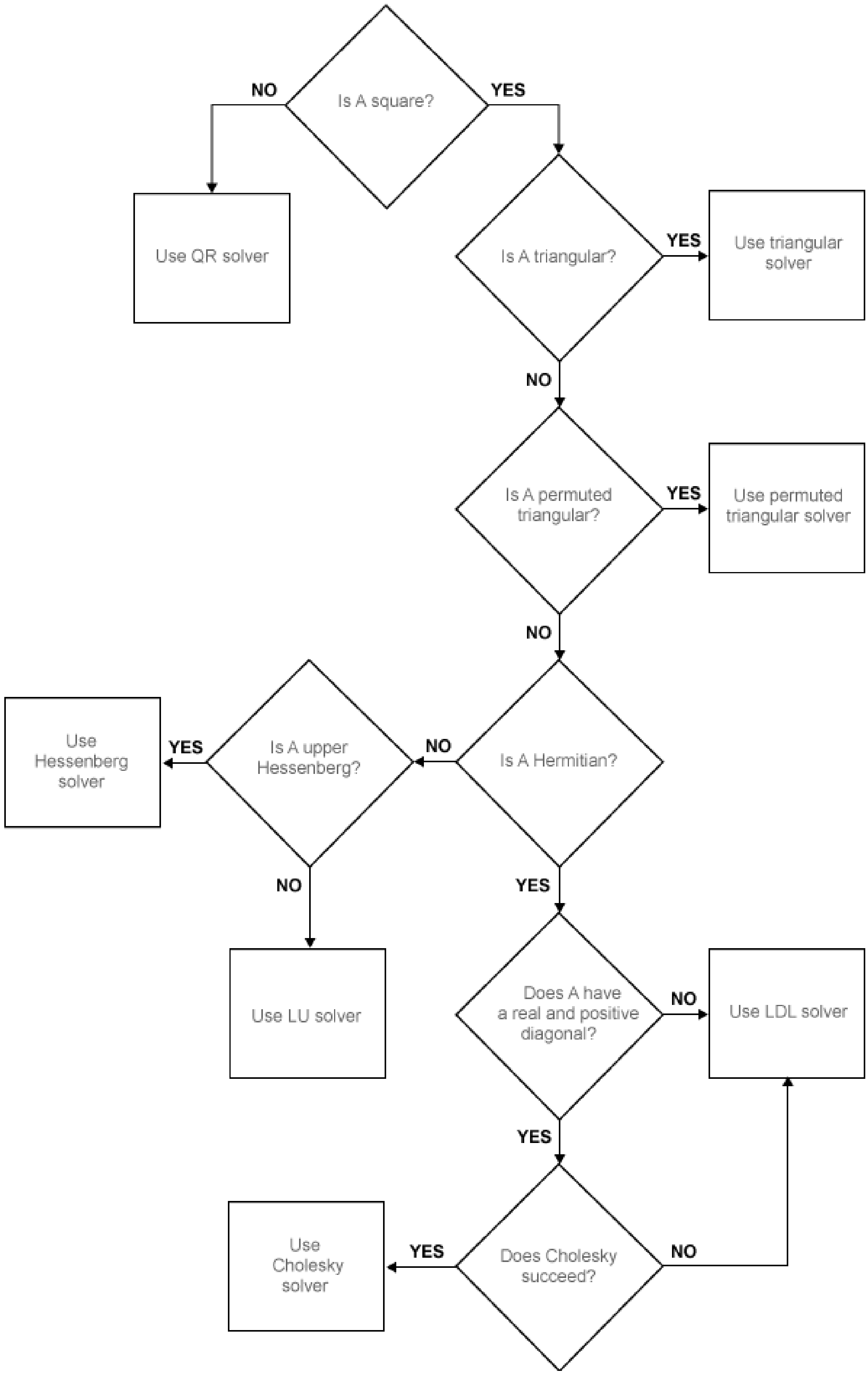

Figure 1: Flow chart of Matlab's backslash

Preliminaries

Matrix Inverses

We denote the inverse of $\boldsymbol{A}$ by $\boldsymbol{A}^{-1}$ where,

1. $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$ and $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix of all ones on the diagonal. This implies uniqueness of the matrix inverse.
2. $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

Properties of an invertible matrix

Let $\boldsymbol{A}$ be an $n - by - n$ matrix. The following are equivalent:

1. $\boldsymbol{A}$ is invertible
2. $\boldsymbol{A}$ is full rank
3. Range($\boldsymbol{A}$) $= \mathbb{R}^n$
4. Null space($\boldsymbol{A}$) $= \{0\}$
5. 0 is not an eigenvalue of $\boldsymbol{A}$
6. 0 is not a singular value of $\boldsymbol{A}$
7. $\det(A) \neq 0$

Vector Norms

Insight

- $\|\mathbf{x}\|$ is a vector norm to measure the "distance" from $\mathbf{x}$ to $\mathbf{0}$.
- $\|\mathbf{x} - \mathbf{y}\|$ is a vector norm to measure the "distance" from $\mathbf{x}$ to $\mathbf{y}$.

Properties

$\mathbb{R}^n \to \mathbb{R}$

1. $f(\mathbf{x}) \geq 0$
2. $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$
3. $f(\alpha\mathbf{x}) = |\alpha| f(\mathbf{x})$

Matrix Norms

Insight

- How "big" is your matrix
- vec($\boldsymbol{A}$): $\mathbb{R}^{m\times n} \to \mathbb{R}^{mn\times 1}$ represent the Frobenius norms.
- Induced p-norms look at a quantity that maximizes the matrix

Frobenius Norms

1. $\|\boldsymbol{A}\| \geq 0$
2. $\|\boldsymbol{A}+\boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$
3. $\|\alpha\boldsymbol{A}\| = |\alpha|\|\boldsymbol{A}\|$

Example: $\|\boldsymbol{A}\|_F = \|\mathrm{vec}(\boldsymbol{A})\|_2 = \sqrt{\Sigma \boldsymbol{A}_{ij}^2}$

Induced p-norms

Let $\|\mathbf{x}\|$ be a vector norm.
$\|\boldsymbol{A}\| = \max_{\mathbf{x} \neq = 0} \frac{\|\boldsymbol{A}\mathbf{x}\|}{\|\mathbf{x}\|}$ Famous induced p-norms:

1. $\|\boldsymbol{A}\|_1$ = maximum absolute column sum
2. $\|\boldsymbol{A}\|_\infty$ = maximum absolute row sum

Singular Value Decompostion

SVD statement

Let $\boldsymbol{A}$ be an $m - by - n$ matrix, then there exists orthogonal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, a diagonal matrix Σ such that $\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$, where $\boldsymbol{U}$ is $m - by - m$ and $\boldsymbol{V}$ is $n - by - n$.

Let $\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \sigma_n \end{bmatrix}$, then σ_1 is the largest singular value and σ_n is the smallest singular value.

SVD related statements

2-norm of a matrix

Recall that $\|\boldsymbol{A}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|\boldsymbol{A}\mathbf{x}\|_2$. Observe the following:

$$\begin{aligned}
\|\boldsymbol{A}\|_2^2 &= \max_{\|\mathbf{x}\|_2=1} \|\boldsymbol{A}\mathbf{x}\|_2^2 \\
&= \max_{\|\mathbf{x}\|_2=1} (\boldsymbol{A}\mathbf{x})^T(\boldsymbol{A}\mathbf{x}) \\
&= \max_{\mathbf{x}^T\mathbf{x}=1} \mathbf{x}^T\boldsymbol{A}^T\boldsymbol{A}\mathbf{x} \\
&= \max_{\mathbf{x}^T\mathbf{x}=1} \mathbf{x}^T(\boldsymbol{U}\Sigma\boldsymbol{V}^T)^T(\boldsymbol{U}\Sigma\boldsymbol{V}^T)\mathbf{x} \\
&= \max_{\mathbf{x}^T\mathbf{x}=1} \mathbf{x}^T\boldsymbol{V}\Sigma\boldsymbol{U}^T\boldsymbol{U}\Sigma\boldsymbol{V}^T\mathbf{x} \\
&= \max_{\mathbf{x}^T\mathbf{x}=1} \mathbf{x}^T\boldsymbol{V}\Sigma^2\boldsymbol{V}^T\mathbf{x}
\end{aligned}$$

Let $\mathbf{y} = \boldsymbol{V}^T\mathbf{x}$, and observe that $\mathbf{y}^T\mathbf{y} = 1$

$$\|\boldsymbol{A}\|_2^2 = \max_{\mathbf{y}^T\mathbf{y}=1} \mathbf{y}^T\Sigma^2\mathbf{y}$$

Frobenius norm of a matrix

Since $\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$, then $\Sigma = \boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}$, and thus:

$$\|\boldsymbol{A}\|_2 = \|\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}\|_2 = \|\Sigma\|_2 = \sqrt{\sigma_1^2 + \ldots + \sigma_k^2}$$

where k is the rank of $\boldsymbol{A}$.

Determinant

$$\begin{aligned} \det(A) &= \det(\boldsymbol{U}\Sigma\boldsymbol{V}^T) \\ &= \det(\boldsymbol{U})\det(\Sigma)\det(\boldsymbol{V}) \\ &= (\pm 1)\det(\Sigma)(\pm 1) \end{aligned}$$

$\Rightarrow |\det(A)| = |\det(\Sigma)|$

Eckart-Young Theorem

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, without loss of generality, assume $m \geq n$. Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ be the SVD decomposition of $\boldsymbol{A}$. Let $\boldsymbol{U} = \begin{bmatrix}\mathbf{u}_1, \ldots, \mathbf{u}_m\end{bmatrix}$, $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_n)$, and $\boldsymbol{V} = \begin{bmatrix}\mathbf{v}_1, \ldots, \mathbf{v}_n\end{bmatrix}$. Define

$$\boldsymbol{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Then,

$$\|\boldsymbol{A} - \boldsymbol{A}_k\| \leq \|\boldsymbol{A} - \boldsymbol{B}\|$$

for any matrix $\boldsymbol{B}$ of rank k.

Singular Value Decomposition - Special Cases

Singular Value Decomposition of a Column Vector

Here, we derive the general form of the SVD of a row vector or a column vector.

Let $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_2 \end{bmatrix}$ **be a column vector, below is the derivation of its SVD decomposition:**

We want to find $\boldsymbol{U}$, $\boldsymbol{V}$ and Σ such that $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} u_{1,1} & \ldots & u_{1,n} \\ \vdots & \ddots & \vdots \\ u_{n,1} & \ldots & u_{n,n} \end{bmatrix}_{n\times n} \begin{bmatrix} N \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1} \begin{bmatrix} l \end{bmatrix}_{1\times 1}$$

We know that $\begin{bmatrix} l \end{bmatrix}_{1\times 1}$ has to be orthonormal so the only possible values of l will be ± 1. Choose the value to be $+1$ (since the sign will be reflected somewhere else in the multiplication of the matrices).
This means that:
$a_1 = u_{1,1} \times N \Rightarrow a_1^2 = u_{1,1}^2 \times N^2$

$\vdots$

$a_n = u_{n,1} \times N \Rightarrow a_n^2 = u_{n,1}^2 \times N^2$, but

$u_{1,1}^2 + \ldots + u_{n,1}^2 = 1$ (Definition of orthogonal matrix). Multiply this equation by N^2
$N^2 \times (u_{1,1}^2 + \ldots + u_{n,1}^2) = N^2$
$\Rightarrow a_1^2 + \ldots + a_n^2 = N^2 \Rightarrow N = \pm\|\mathbf{a}\|$, let $N = \|\mathbf{a}\|$
This means that $u_{1,1} = \frac{a_1}{\|\mathbf{a}\|} \ldots u_{n,1} = \frac{a_n}{\|\mathbf{a}\|}$. We still need to form the rest of the matrix $\boldsymbol{U}$ in a way that all its columns and rows are orthogonal to each other.
We know that we can form a matrix $\boldsymbol{U}'$ where $\boldsymbol{U} = \begin{bmatrix}\mathbf{u}_1 & \boldsymbol{U}'\end{bmatrix}$ is an orthogonal.
Let v be a row vector, to find its SVD decomposition, we can find the SVD decomposition of $\mathbf{v}^T$ and then transpose the result.

Singular Value Decomposition Demo

Demo time!

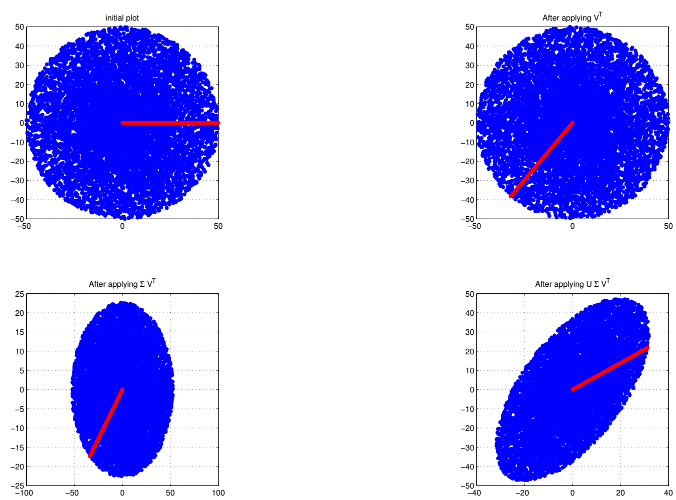

Figure 2: The stages of SVD on a random plot

LU Factorization

To solve a system $\boldsymbol{A}\mathbf{x} = \mathbf{b}$, Gaussian Elimination can be used. The idea is to convert a system $\boldsymbol{A}\mathbf{x} = \mathbf{b}$ into a triangular system. Backward and forward solving are then used to solve such systems. Before introducing the LU factorization, let's take a look at a basic example. Assume we want to solve the following system:

$$x - y = 30$$
$$x + y = 20$$

Geometrically, this problem is illustrated in figure 3. Algebraically, if we subtract equation 2 from equation 1, we get

$$0x + 2y = -10 \Rightarrow y = -5$$
$$\therefore x = 25$$

If we encode the above system in a matrix format, we get,

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 30 \\ 20 \end{bmatrix} \Rightarrow$$
$$\begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 30 \\ -10 \end{bmatrix}$$

LU Factorization Statement

Let $\boldsymbol{A}$ be an $m-by-n$ matrix, then we can write the following equality $\boldsymbol{PA} = \boldsymbol{LU}$, where $\boldsymbol{L}$ is a lower triangular matrix, $\boldsymbol{U}$ is an upper triangular matrix and $\boldsymbol{P}$ is a permutation matrix. To realize why a permutation matrix is important, think of a matrix where the $a_1 1$ entry is zeros.
Row permutations is sufficient for the LU factorization and is called LU factorization with partial pivoting. LU factorization with full pivoting involves permuting the rows and columns of the matrix. It can be stated as follows, $\boldsymbol{PAQ} = \boldsymbol{LU}$

LU Factorization Example - The 2-by-2 case

Let us take the matrix from the previous example and try to find its LU factorization.

$$\begin{aligned} \boldsymbol{A} &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix} \end{aligned}$$

Then,

$$\begin{aligned} 1 &= l_{11}u_{11} \\ -1 &= l_{11}u_{12} \\ 1 &= l_{21}u_{11} \\ 1 &= l_{21}u_{12} + l_{22}u_{22} \end{aligned}$$

Set $l_{11} = 1$ and $l_{22} = 1$. This leads to:

$$\begin{aligned} u_{11} &= 1 \\ u_{12} &= -1 \\ l_{21} &= 1 \\ u_{22} &= 2 \end{aligned}$$

Substituting these values in the original equation leads to:

$$\begin{aligned} \boldsymbol{A} &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} \end{aligned}$$

LU Factorization For Solving Linear Systems

Here is an illustration that forward and backward solving is nothing but a few lines of code.

```
function x = backsolve_huda(A,b)
% input: A is an n by n square full rank upper triangular matrix
%        b is an n by 1 column vector
% output: x such that Ax = b

% obtain the number of rows/cols in A
n = size(A,1);
x=zeros(n,1);

% obtain the nth entry in x which is expressed by the formula below
x(n)=b(n)/A(n,n);

% start going upwards in the matrix
% in general, x(i) = b(i) - A(i,i+1:n)*x(i+1:n))/A(i,i)
% and this is why we start with the loop from n-1 and then go backwards
for i=n-1:-1:1
   x(i)=(b(i)-A(i,i+1:n)*x(i+1:n))/A(i,i);
end


end
```

```
function v = forwardsolve_huda(L,y)
% input: A is an n by n square full rank lower triangular matrix
%        b is an n by 1 column vector
% output: x such that Ax = b


n = size(L,1);
% In solving a lower triangular matrix, we start by filling up the first
% element in x and then go all the way to n
v=zeros(n,1);
v(1)=y(1)/L(1,1);
for i=2:n
   v(i)=(y(i)-L(i,1:i-1)*v(1:i-1))/L(i,i);
end



end
```

LU Rank - 1 update

LU Decomposition

Since the three systems in each Simplex iteration involve the same matrix $\boldsymbol{B}$, then computing the $\boldsymbol{LU}$ decomposition of $\boldsymbol{B}$ once at each iteration appears a natural option. Solving the three systems could then be written using the following commands:

```
x_B = U\(L\(P*b));
l_vec = P'*(L'\(U'\c_B));
d = U\(L\(P*a_q));
```

Compellingly, MATLAB's backslash operator is so powerful that storing the matrix `U\(L\(P)` is not a good idea. This goes back to the way `'\'` is implemented, where its performance on sparse triangular matrices is significantly faster than a matrix-vector product.

Updating The LU Factorization

Let $\boldsymbol{B}$ be a matrix involved in one of the iterative methods where only one column changes at each iteration (an example of such methods is the Simplex Method). One column leaves the basis and another one enters. This could be expressed as a rank-1 update of the $\boldsymbol{LU}$ factors of $\boldsymbol{B}$. In this section, we describe the simple rank-1 update, move on to the rank-k update, and then analyze the rank-k update in the context of the simplex algorithm.

General LU rank-1 update

The problem we address here is the following: Given the $\boldsymbol{LU}$ decomposition of a matrix $\boldsymbol{A}$, find some sort of factorization of the rank-1 update of $\boldsymbol{A}$. In the following, w.l.o.g. denote $\boxed{1}$ as a rank 1 matrix. Note that if $\boxed{1}$ stays from one step to another, this doesn't necessarily mean that it is still the same matrix, yet it still qualifies as a rank 1 matrix.

$$\begin{gathered}
\text{Let } \boldsymbol{PA} = \boldsymbol{LU} \\
\text{Let } \boldsymbol{A}' = \boldsymbol{A} + \boxed{1}, \text{ then} \\
\boldsymbol{PA}' = \boldsymbol{PA} + \boxed{1}, \\
\Rightarrow \boldsymbol{PA}' = \boldsymbol{LU} + \boxed{1} \\
\Rightarrow \boldsymbol{L}^{-1}\boldsymbol{PA}' = \boldsymbol{U} + \boxed{1}
\end{gathered}$$

Interestingly, the righthand side has a special format. Let $\boldsymbol{U}^+ = \boldsymbol{U} + \boxed{1}$. It is an upper triangular matrix with a spike. Applying simple permutations on $\boldsymbol{U}^+$ leads to a matix that is upper triangular except for the last row. Call this matrix $\boldsymbol{G} = \Pi^T \boldsymbol{U}^+ \Pi$. Decomposing $\boldsymbol{G}$ into its $\boldsymbol{LU}$ factors is obvious and we provide an illustration below.

$$\text{Let } \boldsymbol{U}^+ = \begin{bmatrix} * & * & * & * & * \\ & * & * & * & * \\ & * & * & * & * \\ & * & & * & * \\ & * & & & * \end{bmatrix} \Rightarrow \text{for a certain } \Pi,\ \boldsymbol{G} = \begin{bmatrix} * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & * \\ & * & * & * & * \end{bmatrix}.$$

Then by permuting the second column and placing it at the location of the fifth column and then permuting the rows to preserve an upper triangular form, we reach to a matrix $\boldsymbol{G}$ that has the form presented above.
To illustrate the computation of the $\boldsymbol{LU}$ factors of $\boldsymbol{G}$,

$$\text{Let } \boldsymbol{G} = \begin{bmatrix} g_{1,1} & g_{1,2} & g_{1,3} & g_{1,4} & g_{1,5} \\ & g_{2,2} & g_{2,3} & g_{2,4} & g_{2,5} \\ & & g_{3,3} & g_{3,4} & g_{3,5} \\ & & & g_{4,4} & g_{4,5} \\ & g_{5,2} & g_{5,3} & g_{5,4} & g_{5,5} \end{bmatrix}, \text{ Then the } \boldsymbol{LU} \text{ factors of } \boldsymbol{G} \text{ are:}$$

$$\boldsymbol{L} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & l_{5,2} & l_{5,3} & l_{5,4} & 1 \end{bmatrix}, \text{ and } \boldsymbol{U} = \begin{bmatrix} g_{1,1} & g_{1,2} & g_{1,3} & g_{1,4} & g_{1,5} \\ & g_{2,2} & g_{2,3} & g_{2,4} & g_{2,5} \\ & & g_{3,3} & g_{3,4} & g_{3,5} \\ & & & g_{4,4} & g_{4,5} \\ & & & & u_{5,5} \end{bmatrix}.$$

Equating the last row of the system $\boldsymbol{G} = \boldsymbol{LU}$, yields to solutions for the last row of $\boldsymbol{L}$ and the entry $u_{5,5}$ in $\boldsymbol{U}$.

General LU rank-k update

The general $\boldsymbol{LU}$ rank-k update follows directly from the rank-1 update. We dervie the rank-1, and rank-2 updates of a general matrix $\boldsymbol{A}$, and then present the the general case:

Let $\boldsymbol{P}_1\boldsymbol{A}_1 = \boldsymbol{L}_1\boldsymbol{U}_1$. Then, if $\boldsymbol{A}_2 = \boldsymbol{A}_1 + \boxed{1}$ Then, $\boldsymbol{P}_1\boldsymbol{A}_2 = \boldsymbol{P}_1\boldsymbol{A}_1 + \boxed{1}$
Thus, $\boldsymbol{P}_1\boldsymbol{A}_2 = \boldsymbol{L}_1\boldsymbol{U}_1 + \boxed{1}$ and performing the steps discussed in (General LU rank-1 update), we reach:

$$\boldsymbol{P}_1\boldsymbol{A}_2 = \boldsymbol{L}_1\Pi_1\boldsymbol{L}_2\boldsymbol{U}_2\Pi_1^T$$

If $\boldsymbol{A}_3 = \boldsymbol{A}_2 + \boxed{1}$ Then, $\boldsymbol{P}_1\boldsymbol{A}_3 = \boldsymbol{P}_1\boldsymbol{A}_2 + \boxed{1} = \boldsymbol{L}_1\Pi_1\boldsymbol{L}_2\boldsymbol{U}_2\Pi_1^T + \boxed{1}$
Thus, $\boldsymbol{L}_2^{-1}\Pi_1\boldsymbol{L}_1^{-1}\boldsymbol{P}_1\boldsymbol{A}_3\Pi_1 = \boldsymbol{U}_2 + \boxed{1}$ and performing the steps discussed in (General LU rank-1 update), we reach:

$$\boldsymbol{L}_2^{-1}\Pi_1\boldsymbol{L}_1^{-1}\boldsymbol{P}_1\boldsymbol{A}_3\Pi_1 = \Pi_2\boldsymbol{L}_3\boldsymbol{U}_3\Pi_2^T$$

This leaves us with the following general form:

$$\boldsymbol{P}_1\boldsymbol{A}_i\Pi_1\ldots\Pi_{i-1} = \boldsymbol{L}_1\Pi_1\boldsymbol{L}_2\Pi_2\ldots\boldsymbol{L}_{i-1}\Pi_{i-1}\boldsymbol{L}_i\boldsymbol{U}_i$$

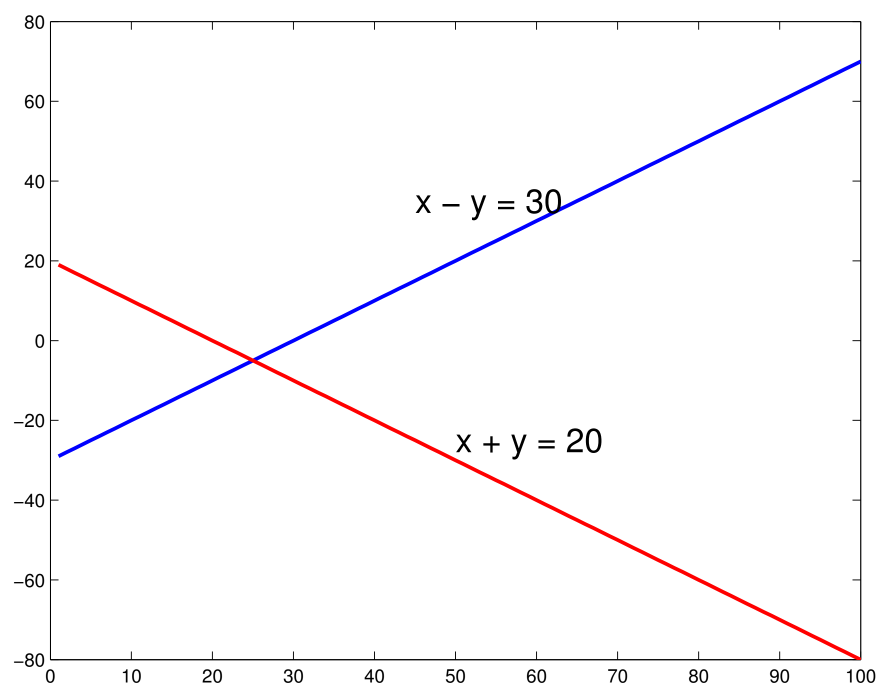

Figure 3: A linear system of two equations illustrated geometrically

QR Decomposition

Let $\boldsymbol{A}$ be an $m-by-n$ matrix, then the QR decomposition of A is, $\boldsymbol{A} = \boldsymbol{QR}$ where $\boldsymbol{Q}$ is an $m-by-m$ orthogonal matrix and $\boldsymbol{R}$ is an $m-by-n$ upper triangular matrix. Note that if we are trying to solve a linear system $\boldsymbol{A}\mathbf{x} = \mathbf{b}$, the system can be written as follows:

$$\begin{aligned} \boldsymbol{A}\mathbf{x} &= \mathbf{b} \\ \boldsymbol{Q}\boldsymbol{R}\mathbf{x} &= \mathbf{b} \\ \boldsymbol{R}\mathbf{x} &= \boldsymbol{Q}^T\mathbf{b} \end{aligned}$$

QR deomposition of a column vector

Let $\mathbf{a}$ be a vector $\in \mathbb{R}^n$.

$\mathbf{a} = \boldsymbol{Q}\boldsymbol{R} = \boldsymbol{Q}\begin{bmatrix} \gamma \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \boldsymbol{Q}\gamma\mathbf{e}_1$

$\gamma = \pm\|\mathbf{a}\|$

$\mathbf{a} = \boldsymbol{Q}\gamma\mathbf{e}_1 = \mathbf{q}_1\|a\|\mathbf{e}_1$

$\mathbf{q}_1 = \frac{\mathbf{a}}{\|\mathbf{a}\|}$

Now, the question becomes, can we find an orthogonal matrix $\boldsymbol{Q}$ such that

$$\boldsymbol{Q}\mathbf{a} = \|\mathbf{a}\|\mathbf{e}_1$$

And the answer is yes - use householder reflector.

Householder Reflections

Let $\mathbf{v} \in \mathbb{R}^m$ be nonzero. An $m - by - m$ matrix $\boldsymbol{P}$ of the form

$$\boldsymbol{P} = \boldsymbol{I} - \beta\mathbf{v}\mathbf{v}^T, \beta = \frac{2}{\mathbf{v}^T\mathbf{v}}$$

is the Householder refelection.

QR Related Statements

Let $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ be a QR factorization of a full column rank $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and

$$\begin{aligned} \boldsymbol{A} &= \begin{bmatrix}\mathbf{a}_1 | \dots | \mathbf{a}_n\end{bmatrix} \\ \boldsymbol{Q} &= \begin{bmatrix}\mathbf{q}_1 | \dots | \mathbf{q}_m\end{bmatrix} \end{aligned}$$

then, $\text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_k\} = \text{span}\{vq_1, \dots, \mathbf{q}_k\}$,
where k is any scalar such that $1 \le k \le \min(m, n)$

On the uniqueness of the QR decomposition

Let $\boldsymbol{A}$ be a full rank $m - by - n$ matrix with $m > n$. $\text{Rank}(\boldsymbol{A}) = n$.
Suppose:

$$\boldsymbol{A} = \boldsymbol{Q}_1\hat{\boldsymbol{R}}_1 = \boldsymbol{Q}_2\hat{\boldsymbol{R}}_2$$

Without loss of generality, let us assume that $\hat{\boldsymbol{R}}_1$ and $\hat{\boldsymbol{R}}_2$ are unipotent.

$$
\begin{aligned}
\boldsymbol{Q}_1\boldsymbol{D}_1\boldsymbol{R}_1 &= \boldsymbol{Q}_2\boldsymbol{D}_2\boldsymbol{R}_2 \\
\boldsymbol{Q}_2^T\boldsymbol{Q}_1\boldsymbol{D}_1\boldsymbol{R}_1 &= \boldsymbol{D}_2\boldsymbol{R}_2 \\
\boldsymbol{Q}_2^T\boldsymbol{Q}_1\boldsymbol{D}_1\boldsymbol{R}_1\boldsymbol{R}_1^{-1}\boldsymbol{D}_1^{-1} &= mD_2\boldsymbol{R}_2 mR_1^{-1}\boldsymbol{D}_1^{-1} \\
(\boldsymbol{Q}_2^T\boldsymbol{Q}_1)^T(\boldsymbol{Q}_2^T\boldsymbol{Q}_1) &= (\boldsymbol{D}_2\boldsymbol{R}_2\boldsymbol{R}_1^{-1}\boldsymbol{D}_1^{-1})^T(\boldsymbol{D}_2\boldsymbol{R}_2\boldsymbol{R}_1^{-1}\boldsymbol{D}_1^{-1})
\end{aligned}
$$

Thus, the RHS should be the identity matrix, more specifically each of $(\boldsymbol{D}_2\boldsymbol{R}_2\boldsymbol{R}_1^{-1}\boldsymbol{D}_1^{-1})^T$ and $(\boldsymbol{D}_2\boldsymbol{R}_2\boldsymbol{R}_1^{-1}\boldsymbol{D}_1^{-1})$ must be a diagonal matrix. Then, $\boldsymbol{D}_2\boldsymbol{R}_2\boldsymbol{R}_1^{-1}\boldsymbol{D}_1^{-1} = \boldsymbol{D}_3$, thus $\boldsymbol{R}_2\boldsymbol{R}_1^{-1} = \boldsymbol{D}_2^{-1}\boldsymbol{D}_3\boldsymbol{D}_1$.

Recall that $\hat{\boldsymbol{R}}_1 = \boldsymbol{D}_1\boldsymbol{R}_1$ and $\hat{\boldsymbol{R}}_2 = \boldsymbol{D}_2\boldsymbol{R}_2$ where $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ have 1's on the diagonal.

$$
\begin{aligned}
\boldsymbol{R}_2\boldsymbol{R}_1^{-1} &\text{have 1's on the diagonal} \\
\boldsymbol{D}_2^{-1}\boldsymbol{D}_3\boldsymbol{D}_1 &\text{have 1's on the diagonal}
\end{aligned}
$$

Then,

$$
\begin{aligned}
\boldsymbol{D}_2^{-1}\boldsymbol{D}_3\boldsymbol{D}_1 &= \boldsymbol{I} \\
\boldsymbol{R}_2\boldsymbol{R}_1^{-1} &= I
\end{aligned}
$$

This leads to, $\boldsymbol{Q}_1\boldsymbol{D}_1 = \boldsymbol{Q}_2\boldsymbol{D}_2 \Rightarrow \boldsymbol{Q}_2^T\boldsymbol{Q}_1^T = \boldsymbol{D}_2\boldsymbol{D}_1^{-1}$. Multiply this equation by its inverse:

$$
\begin{aligned}
(\boldsymbol{Q}_2^T\boldsymbol{Q}_1^T)^T(\boldsymbol{Q}_2^T\boldsymbol{Q}_1^T) &= (\boldsymbol{D}_2\boldsymbol{D}_1^{-1})^T(\boldsymbol{D}_2\boldsymbol{D}_1^{-1}) \\
\boldsymbol{I} &= \boldsymbol{D}_1^{-1}\boldsymbol{D}_2\boldsymbol{D}_2\boldsymbol{D}_1^{-1}
\end{aligned}
$$

Conclusion: QR factorization is essentially unique (up to $\pm$)

Some Applications

Demo time.