# **Partial pivoting:**

## **An empirical study into the effectiveness of the approach**

This study stems from the explanation of the method of partial-pivoting applied to gaussian elimination, as explained in *Linear Algebra - Gilbert Strang*. Certain queries arise during the course of the explanation which leads to this study. To prepare the context, a short description of partial-pivoting approach, as given in *Linear Algebra - Gilbert Strang*, will be presented. Based on that, some queries will be raised whose answers will require some empirical study. The study will finally show us the effectiveness of the approach of partial-pivoting.

## The context: partial-pivoting

This example is from *Linear Algebra - Gilbert Strang*.

Hence, the author makes the case that working with small pivots makes the solution tremendously unstable. To avoid this instability the approach of partial pivoting is used.

What is the approach of partial pivoting? At any point in elimination, if the pivot position entry is small compared to those below it, we perform a row exchange to ensure that the largest number is in the pivot position.

As an example -

Given this context certain questions arise which will be presented in the next section.

Queries

The problem with relatively (as compared to the other entries of the column) small pivot values, as presented above, is that they lead to instability in the solution. Instability in solutions denotes that tiny changes in the pivot values may lead to large changes in the solutions. Because these tiny changes may creep in unintentionally as rounding-off errors, the solutions are unreliable.

However, this above argument might hold for small values in any position. Then why this emphasis on the pivot position. The scheme of partial pivoting only shifts the small value from a pivot to a non-pivot position. Why does that cure the problem of instability in the solution? ***Is the solution more sensitive to small values in the pivot position than when they are in non-pivot position?***

## The methodology of study

To answer the above question, we can devise an emipirical form of investigation. We need to study the changes in the solution as a function of the relative values of the entry in a certain position. To achieve that we need to decide upon a way to quantify two parameters - ***changes in solution*** and ***relative value of the entry in a certain position***.

* **Changes in solution** - For solutions x1 and x2 (both are vectors), we define the change in solution by |x2 - x1|/|x1|.
* **Relative value of the entry in a certain position** - The relative value of an entry is its value relative to the other entries in the column. Thus we need a statistic to represent the overall entry sizes in a column. For our case we choose the minimum value of a column as the representative. Thus for an entry u in the column vector x, the relative value of u is u/min(x).

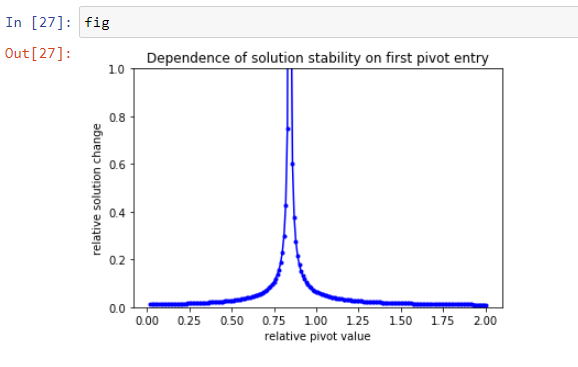
We now choose a random matrix A, column vector b of compatible sizes; find the solution x. We gradually change the relative value of the entry in position (0,0)in A and record the changes in solution x. A plot of this change vs relative value will give us an idea of the sensitivity of the solution to the relative size of pivot values at (0,0).

Initial study with a single matrix

The first study that we conduct with a single trial as per the above methodology is presented in the notebook **Partial\_Pivoting\_Project\Partial\_Pivoting.ipynb**. Please go through the notebook, at least through the section comments if not through the code, to get an idea of the way the study was conducted and the final conclusion we draw from it. It is necessary to understand all that follows from this point onward.

## Points to be addressed

The plot presented in **Partial\_Pivoting.ipynb** is being provided here just for quick reference -



Based on this plot below certain points arise which guide us toward further investigations.

* **Q1** - Does an excessive instability typically exist? Is a spike typically present in the plot? - only then there is a point in trying to avoid it via some strategy like partial pivoting.
* **Q2** - If answer to Q1 is true then - Where does the point of major instability typically lie? Is there a typical range of relative values where that spike typically falls? - then we might try to avoid that region.
* **Q3** - How do non-pivot points compare to pivot points in the light of Q1 and Q2? - that will explain whether or not we have to care equally about both pivot and non-pivot points. In fact that addresses the previously raised doubt - if a row exchange, leading to the small entry being placed in a non-pivot position, at all improves the stability of solution.

In order to find the “typical” nature of a n-sized matrix in terms of instability we will run 100 trials for each given size and position. For each of these 100 trails we will use some statistics to address each of Q1 and Q2.

## Measures of instability

**For Q1 (does an excessive instability typically exist?):**

Each of the 100 trials for a given size and position gives a sequence of relative solution change values, y\_data. We want to detect if a sharp spike is typical in the y\_data sequence across these 100 trials.

A **“sharp spike”** needs to have two properties -

* First, it should be thin otherwise if it is widely spread over the x-axis there is no point in trying to avoid it (and it is not a spike anymore).
* Second, the y\_data in the spiky region should be substantially high compared to the other portions. That is why we want to avoid the region.

We need to view the data from 100 trials in a way suited to identify the existence of these two above properties.

From this sequence of y\_data we can select the 90 and 100 percentile data-marks. Thus for 100 trials we have 100 pairs of these 90 and 100-percentile marks. We will have two box plots showing the spread of these 100 90-percentile and 100-percentile (or maximum instability value) marks respectively.

The 90-percentile and 100-percentile box-plots that we get - if they show sufficient vertical separation then that is an indicator of the above two properties being present. If there is a typical spiky nature, it ought to show up in the vertical separation of the box-plots for 90-percentile and 100-percentile marks.

If the above argument does not seem convincing consider how a spiky nature can be destroyed.

* First, the region of steep rise becomes fat - the spike gets blunt. That means more data is in this higher end. This goes to decrease the gap between the 90 and 100 percentile marks. This happens when the instability is widespread in the sense that for a wide range of relative entries the solution is very much instable. When that happens any attempt to avoid that region of relative entries is futile.
* Second, the spike might get shorter in height. That means that the higher end values are not so much greater than the rest of the points. This also goes to decrease the gap between the 90 and 100 percentile marks. This indicates that the instability, though peaking at someplaces, is not so much more than its typical value as to require any strategy to avoid it.

**For Q2 (is there a typical range of relative values where the spike lies?):**

For this across all the 100 trials we record the point displaying maximum instability. We again do a box-plot showing the spread of these max-instability points. The lower the spread of the box-plot the more typical the point of max-instability is for a given size and position. If it is not spread much, we can take steps to avoid that region of relative entry values.

## Comparing pivot and non-pivot points

The first measure of inequality, handled under the point “**For Q1 (does an excessive instability typically exist?)”**, can be compared for pivot and non-pivot points. That will help us understand how a row exchange, that moves a small relative entry from pivot to non-pivot position, affects the solution stability.

## Strategy to implement the above

The strategy to study for sizes from 2..10 and position (either 0,0 or 0,1) over a sequence of 100 trials has been implemented in **Partial\_Pivoting\_Project\ Partial\_Pivoting\_Aggregate.py**. There are certain basic elements in this implementation which are described below -

* **single\_matrix\_single\_trial** - This will take up a given position (either 0,0 or 1,0) and matrix size; change the relative value of that position within a LOW-HIGH range and return the sequences of relative entry values, relative solution changes, the point of maximum relative solution change, the 90 & 100-percentile positions respectively. This method will work as the atomic method for the entire solution. In fact it simulates 1 trial out of the 100 trials we plan to execute. This method is similar to what is present in **Partial\_Pivoting\_Project\Partial\_Pivoting.ipynb**.

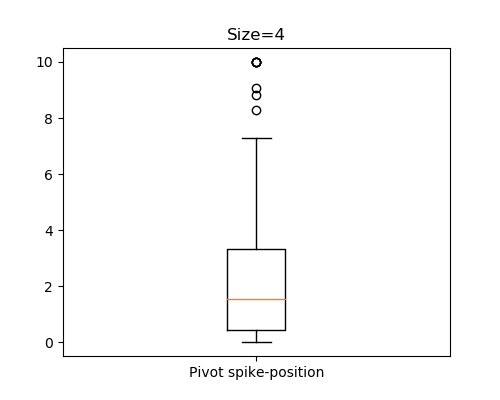
* **simulation\_table** - A table with a multi-indexed rows denoting the size and position (pivot and non-pivot) combination, while the multi-indexed column will contain the data -- point of maximum change, the 90 & 100-percentile relative solution change values -- for each of the hundred trails.
* **simulation\_data\_generation** - This method generates the simulation data **Partial\_Pivoting\_Project\SimulationData\ simulation\_table.csv.** For each size from 2..10 and for each position out of (0,0) and (1,0) it records the properties max\_spike\_point(relative entry value showing max solution instability), 90\_percentile and 100\_percentile relative solution changes.
* **plots** - There are two types of box-plots present here.
  + 90\_percentile/100\_percentile comparison for pivot and non-pivot position of each size, presented side by side
  + max\_spike\_point spread for pivot positions of each size

The plots are present in **Partial\_Pivoting\_Project\Plots\**. The 90\_percentile/100\_percentile comparisons are in the files named **SpikeExistence\_Sizek**. The max\_spike\_point spread is in the files named **SpikePosition\_Sizek**.

Conclusion

**Part I: Typical max-instability point**

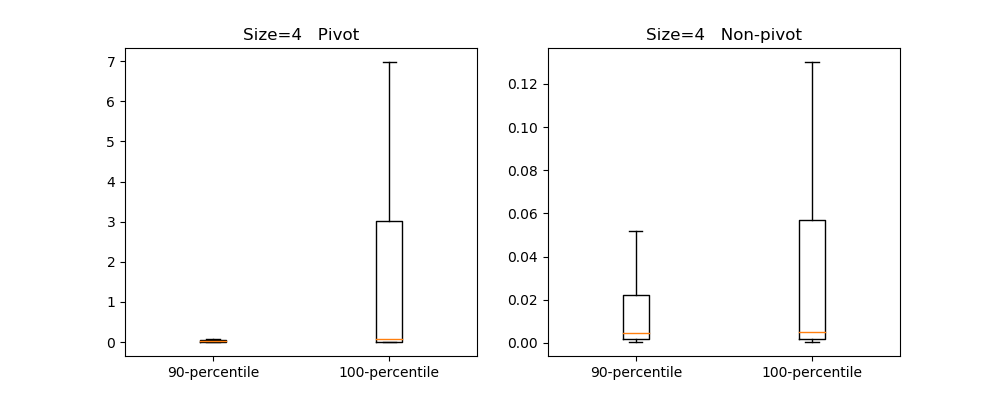
Below is a plot of the max-spike-position for size 4.



From this (and the plots for other sizes) it is evident that the max spike position typically lies at the lower end of relative entry values. **Hence, for pivot positions, small relative entry values lead to higher solution instability.**

**Part II: A small region of particularly high instability**

Below is a plot of the spike existence check for size 4.



From the plots for 90\_percentile/100\_percentile comparison it is evident that a spike typically exists for pivot positions. Also comparing the plots for pivot and non-pivot positions gives us a sense that the spike is much more higher for pivot positions than that of non-pivot positions (pay attention to the vertical scales of the two plots). **Hence smaller relative entries in pivot positions lead to more solution instability than at non-pivot positions.**

**These two observations establish the effectiveness of the partial pivoting strategy.**