The purpose of this discussion will be to better establish the ideas from Lecture #13 *Sparse matrices and optimized parallel implementations*, as related to working with sparse matrices and optimizing sparse matrix-vector product.

File *matrix.output* contains a sparse matrix in ASCII AIJ format. The first line gives 3 numbers

```
number_of_columns  number_of_rows number_of_nonzeroes
```

followed by the matrix. We will discuss

1. Writing routines (and a driver to use them; in C/C++ or Fortran) that
   1. Read the matrix and store it in CRS or CCS format.
   2. Perform matrix-vector product where the matrix is in CRS/CCS format.
   3. Test that the code is correct
   4. Report Mflop/s rate using PAPI

2. Optimizing the routines from part 1. Two optimization strategies can be suggested for this particular case. It is known (e.g. information about the matrix structure) that certain reordering may be beneficial for the performance, and in particular, the following index reordering produces a reordered matrix that is composed of 3x3 dense blocks. Namely, old index $i_{old}$ becomes

   $i_{new} = (i_{old} - 1) / 8660 + 3 * (i_{old} - 1) \% 8660 + 1$.

Here '/' is integer division and '%' is the modulo operation. Note that reordering here means that if the old matrix (before reorder) had a nonzero element

```
2 8661 4.5
```

the new matrix will have

```
4 2 4.5
```

1. Will a matrix-vector product with the reordered matrix be faster than the mat-vec product with the original matrix and why?
2. Will saving the reordered matrix in BCRS format with blocks of size 3x3 improve on performance and why?

**Can we think of other ways to optimize the performance?**

3. How would we implement a sparse matrix-vector product for GPU (NVIDIA GPU with CUDA).