Prof. Dr. Stefan Funken
Prof. Dr. Karsten Urban
Lewin Ernst

High Performance Computing II
Institute of Numerical Mathematics
Sheet 1, 34 Points

Exercises in High Performance Computing II *

Introduction

1. We start leisurely by considering the following tasks.

(a) Don't use C-code for the answer.

```
// 1. Get the value of a and b as well as of *c, *d and **e.
// 2. What is the value of c, d and e?
int a = 2, b = 5, *c = &a, *d = &b;
int* e[1] = {&a};
 a  =  *c * *d;
 d  =   c;
*d *=  *c * *d;
 a *= **e = 10;

```

(b) Have a look at the material and make yourself familiar with the `structs`, `functions`, `data types` and `utilities`.

(c) Read this brief introduction to `makefiles`.

(d) Run the `makefile` and explain the output.

(4+Karma-Point+Karma-Point+2 = 6 Punkte + 2 Karma-Points)

First impressions of storage formats, algorithms

2. In this exercise we want to get a first impression why storage formats are crucial. We concern ourselves with the storage formats from the lecture. As an example, we introduce a first simple sparse matrix storage format, namely the `COO` (coordinate) format, which is also known as the *triplet format*. If a matrix M is given in the `COO` format, then

- $M \to m$; denotes the number of rows.
- $M \to n$; denotes the number of columns.
- $M \to nz$; is the number of entries.
- $M \to nzmax$; is the number of numerical values which is greater or equal to nz. The reason is that we allocate a puffer to store at least nz values.
- $Mx = M \to x$; is an array, which contains the numerical values of M and has the length $nzmax$.
- $M \to p$; is an array, which contains the column indices for Mx and has the length $nzmax$.
- $M \to ind$; is an array, which contains the row indices for Mx and has the length $nzmax$.

The `COO` format is not ordered and permits duplicate entries. If the matrix gets larger during runtime, the new entries are appended to the data array. For instance, the data

$$\begin{aligned} m = 3, \quad n = 3, \quad nzmax = 8, \\ Mx = [1.0\ \ 9.0\ \ 2.0\ \ 5.4\ \ 5.5\ \ *\ \ *\ \ *], \\ p = [0\ 0\ 1\ 2\ 0\ \ *\ \ *\ \ *], \quad ind = [0\ 1\ 2\ 1\ 2\ \ *\ \ *\ \ *], \quad nz = 5. \end{aligned}$$

* Generally: Results are to be explained and code is to be commented.

gives the matrix M

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 9 & 0 & 5.4 \\ 5.5 & 2 & 0 \end{pmatrix}. \tag{1}$$

In C, we combined the matrix formats `COO`, `CRS` and `CCS` in one type `cs`.

Explain the following sparse matrix storage formats:

(a) `CRS` format (*Compressed Row Storage*), `CCS` format (*Compressed Column Storage*),

(b) `SED` format (*Sparse Extracted Diagonal*), also called Modified Column Storage,

(c) `CDS` format (*Compressed Diagonal Storage*),

(d) `JDS` format (*Jagged Diagonal Storage*),

(e) `SKY` format (*Skyline Storage*).

What are the advantages of each format? (2+4+2+2+2 = 12 Punkte)

Parallel matrix vector product

3. In this exercise we will implement the parallel matrix vector product and in the course of this we revise some mpi functions, e.g. `MPI_Bcast` or `MPI_Scatterv`. In order to do so, let $A \in \mathbb{R}^{m \times n}$ be a dense (row major) matrix, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ vectors. We want to compute

$$y \leftarrow \beta y + \alpha A x, \quad \alpha, \beta \in \mathbb{R}.$$

To develop our parallel algorithm, we have to think about data decomposition. There are a variety of ways to partion the matrix and vector elements. Each data decomposition results in a different parallel algorithm. We will use a straightforward way: the rowwise block striping. With this strategy, each of the p processes is responsible for a contiguous group of either $\lfloor m/p \rfloor$ or $\lceil m/p \rceil$ rows of the matrix A. The vectors y and x are replicated, meaning all the vector elements are copied on all of the nodes. Each process then calculates inner products of their respective rows of A and x, multiplies the results with α and add it to the scaled entries of y. After the inner product computation each process has some elements of the result y. However, the vectors are supposed to be replicated so that each process has to communicate its results to all other processes (Hint: there is a mpi function for this task: openmpi doc).

Parallel matrix vector multiplication:

(a) Why is it acceptable to store the vectors x and y in their entirety on each node, but not the matrix A?

(b) Implement the parallel matrix vector multiplication explained above. Assume that m is dividable by p, so that `MPI_Scatter` can be used.

(c) Extend your code such that m do not need to be dividable by p. Therefore `MPI_Scatter` can not be used anymore and a function `slices` is necessary, which calculates for the input `problem_size`, `number_of_processes` and `process_rank` the `offset` and `size` of the slice (for process `process_rank`).

(d) Test your functions.

(1+5+6+4 = 16 Punkte)