PanguLU Users' Guide v5.0.0

July 31, 2025

PanguLU Developer Team

Contents

1 Introduction

This chapter includes an introduction to PanguLU and its code structure.

1.1 Introduction to PanguLU

PanguLU [1] is an open source software package for solving a linear system $Ax = b$ on heterogeneous distributed platforms. The library is written in pure C, and exploits parallelism from MPI, OpenMP and CUDA. The sparse LU factorisation algorithm used in PanguLU splits the sparse matrix into multiple equally-sized sparse matrix blocks and computes them by using sparse basic linear algebra subprograms (sparse BLAS). PanguLU also uses a synchronisation-free communication strategy to reduce the overall latency overhead, and a variety of block-wise sparse kernels have been adaptively called to improve efficiency on CPUs and GPUs. Currently, PanguLU supports data types of float, double, and complex floating points.

Pangu [2] is a primordial being and creation figure in Chinese mythology who separated heaven and earth. This is very like that an LU factorisation method decomposes a matrix A into two triangular matrices L and U. This is why we named our work PanguLU and designed the logo shown in the cover page of this guide.

Our team at the Super Scientific Software Laboratory (SSSLab, https://www.ssslab.cn/), China University of Petroleum-Beijing, is constantly optimising and updating PanguLU.

1.2 Code Structure

Field	Description
`PanguLU/README.md`	Instructions on installation
`PanguLU/src`	C and CUDA source code, to be compiled into libpangulu.a and libpangulu.so
`PanguLU/examples`	Example code
`PanguLU/include`	Contains headers archive pangulu.h
`PanguLU/lib`	Contains library archive libpangulu.a and libpangulu.so
`PanguLU/hunyuan_omp`	Hunyuan - parallel graph partitioning and fill-reducing matrix ordering
`PanguLU/Makefile`	Top-level Makefile that does installation and testing
`PanguLU/make.inc`	Compiler, compiler flags included in all Makefiles

2 Installation

This chapter shows you how to install and configure PanguLU.

2.1 Compilers

Before you install the PanguLU software package by using the ‘make’ command, you need to have the following compilers installed on your system:

- **MPICC** (Tested version: MPICH-4.3.0)
  Download link: https://www.mpich.org/
- **NVCC** (Tested version: CUDA-12.9)
  Download link: https://developer.nvidia.com/cuda-downloads

2.2 Paths to Dependent Packages

Before compiling, you need to specify the library and header file paths of OpenBLAS in the make.inc file and example/Makefile file.

You can refer to the following content for configuration:

- **Sample make.inc file**

```
COMPILE_LEVEL = -O3
CUDA_COMPILE_LEVEL =

#0201000,GPU_CUDA
CUDA_INC = -I/path/to/cuda/include
NVCC = nvcc $(CUDA_COMPILE_LEVEL) $(COMPILE_LEVEL)
NVCCFLAGS = $(PANGULU_FLAGS) -w -Xptxas -dlcm=cg -gencode=arch=compute_61,
code=sm_61 -gencode=arch=compute_61,code=compute_61 $(CUDA_INC)

#general
CC = gcc $(COMPILE_LEVEL)
MPICC = mpicc $(COMPILE_LEVEL)
OPENBLAS_INC = -I/path/to/openblas/include
CFLAGS = $(OPENBLAS_INC) $(CUDA_INC) -fopenmp -lpthread -lm
MPICCFLAGS = $(CFLAGS)
HUNYUAN_INC = -I../hunyuan_omp/include
#METIS_INC = -I/path/to/metis/include
PANGULU_FLAGS = -DPANGULU_LOG_INFO -DCALCULATE_TYPE_R64 -DGPU_OPEN
#-DMETIS #-DPANGULU_MC64 #-DPANGULU_PERF
```

- **Sample example/Makefile file**

```
LINK_HUNYUAN = ../hunyuan_omp/lib/libmynd.so
#LINK_METIS = /path/to/metis/lib/libmetis.so
LINK_OPENBLAS = /path/to/openblas/lib/libopenblas.so
LINK_CUDA = -L/path/to/cuda/lib64 -lcudart -lstdc++ -lcusolver -lcublas
LINK_PANGULU = ../lib/libpangulu.a

all: pangulu_example.elf

pangulu_example.elf:example.c
        mpicc -O3 $< -DCALCULATE_TYPE_R64 -I../include $(LINK_PANGULU)
    $(LINK_CUDA) $(LINK_HUNYUAN) $(LINK_METIS) $(LINK_OPENBLAS)
    -fopenmp -lpthread -lm -o $@

clean:
        rm -f *.elf
```

2.3 Compile Options

PanguLU supports matrix calculations using real and complex types, and can be executed on CPU or GPU.

The following are the specific configuration steps:

- **Configuration options for real and complex types and precisions**
  To configure real and complex types, as well as single and double precisions, update the corresponding compilation flags by adding one of the following options to both the PANGULU_FLAGS variable in the make.inc file and the compilation command in example/Makefile:

	Real	Complex
Single Precision	-DCALCULATE_TYPE_R32	-DCALCULATE_TYPE_CR32
Double Precision	-DCALCULATE_TYPE_R64	-DCALCULATE_TYPE_CR64

- **Enable hyperthreading function**
  If your machine has enabled the hyperthreading feature, adding the following option to PANGULU_FLAGS in the make.inc file may improve performance:

```
-DHT_IS_OPEN
```

Note: If the option is enabled, in the numeric factorisation phase, only even numbered cores are bound when the thread is bound to the CPU. If user can disable the hyperthreading, please try to disable it and not use this option.

- **Use GPU for calculation**
  To use GPU, please modify PANGULU_FLAGS in the make.inc file and add the following option:

```
-DGPU_OPEN
```

  Note: GPU is enabled by default. If you want to compile and run PanguLU on machines without CUDA library, please remove "GPU_CUDA" from build_list.csv and comment "LINK_CUDA" from examples/Makefile.
- **Call METIS [3] or Hunyuan to reorder the matrix**
  If you want to use METIS, modify PANGULU_FLAGS in the make.inc file and add the following option; otherwise, Hunyuan will be used by default for matrix reordering.

```
-DMETIS
```

- **Call MC64 [4] to reorder the matrix**
  If you want to use MC64 to improve numerical stability, modify PANGULU_FLAGS in the make.inc file and add the following option:

```
-DPANGULU_MC64
```

  Note: Our implementation of MC64 currently does not support the calculation of complex numbers.
- **Output information selection**
  If you want to output execution information to stdout, modify PANGULU_FLAGS in the make.inc file and add the following options:

```
-DPANGULU_LOG_INFO
```

  If you just want to output warning and error messages, modify PANGULU_FLAGS in the make.inc file and add the following option:

```
-DPANGULU_LOG_WARNING
```

  If you only want to output error messages that cause PanguLU to fail to run, modify PANGULU_FLAGS in the make.inc file and add the following option:

```
-DPANGULU_LOG_ERROR
```

- **Output additional performance information**
  If you want to output additional performance information, such as kernel time per gpu and gflops of numeric factorisation, you can use the following option:

  ```
  -DPANGULU_PERF
  ```

  Note: This option will slow down the speed of numeric factorisation.

After you have completed the desired configuration in the make.inc file, you can use the ‘make’ command to automatically compile PanguLU. If you need to change the configuration options later, use the ‘make update’ command to recompile after the change.

3 Interface

This chapter describes the main function interfaces of the PanguLU library, covering the steps from matrix initialisation to the solution process. The parameters and return values of each function are described.

3.1 pangulu_init()

- **Description**
  pangulu_init() is used to initialise a handle of PanguLU, and to distribute the matrix to all MPI ranks. Please note that if you need to factorise more than one different matrices, call pangulu_finalize() after completing the solution of one matrix, and then use pangulu_init() to to initialise the next matrix.
- **Function**

```
void pangulu_init(
  sparse_index_t pangulu_n,
  sparse_pointer_t pangulu_nnz,
  sparse_pointer_t *csc_colptr,
  sparse_index_t *csc_rowidx,
  sparse_value_t *csc_value,
  pangulu_init_options *init_options,
  void **pangulu_handle);
```

- **Parameters**
  pangulu_n: Specifies the number of rows in the CSC matrix.
  pangulu_nnz: Specifies the total number of non-zero elements in the CSC matrix.
  csc_colptr: Points to an array that stores pointers to columns of the CSC matrix.
  csc_rowidx: Points to an array that stores indices to rows of the CSC matrix.
  csc_value: Points to an array that stores the values of non-zero elements of the CSC matrix.

init_options: Pointer to a pangulu_init_options structure containing initialisation parameters for the solver. Detailed information refer to following initialisation parameters.
pangulu_handle: On return, contains a handle pointer to the library's internal state.

- **Initialisation parameters**
  nthread: Number of CPU threads to use for OpenMP parallel regions during preprocessing and numeric factorisation.
  nb: Block size for partitioning the matrix.
  gpu_kernel_warp_per_block: Number of warps per CUDA block to use in GPU kernel computations. If set to 0, a default value (typically 4) is used.
  gpu_data_move_warp_per_block: Number of warps per CUDA block to use during GPU data movement operations. Defaults to 4 if set to 0.
  hunyuan_nthread: Number of CPU threads used specifically for Hunyuan reordering. If set to 0, defaults to 4. This parameter will be ignored if METIS is used.
  sizeof_value: Specifies the size (in bytes) of the value type used in the matrix (e.g., 4 for `float`, 8 for `double`, 8/16 for complex types).
  is_complex_matrix: A flag indicating whether the input matrix is complex-valued (nonzero: complex; zero: real).
  mpi_recv_buffer_level: Scaling factor for estimating the number of buffer slots used in MPI communication. Higher values allocate more memory for receiving data blocks. Effective only when MPI is enabled with more than one process.

3.2 pangulu_gstrf()

- **Description**
  pangulu_gstrf() performs distribute sparse LU factorisation. Note that you should call pangulu_init() before calling pangulu_gstrf() to create a handle of PanguLU.
- **Function**

```
void pangulu_gstrf(
  pangulu_gstrf_options *gstrf_options,
  void **pangulu_handle);
```

- **Parameters**
  gstrf_options: Pointer to pangulu_gstrf_options structure.
  pangulu_handle: Pointer to the library internal status handle returned from initialisation.

3.3 pangulu_gstrs()

- **Description**
  pangulu_gstrs() solves linear equation with factorised L and U, and right-hand side vector b. Note that you should call pangulu_gstrf() before calling pangulu_gstrs() to ensure that L and U are available.
- **Function**

```
void pangulu_gstrs(
  sparse_value_t *rhs,
  pangulu_gstrs_options *gstrs_options,
  void** pangulu_handle);
```

- **Parameters**
  rhs: Pointer to the right-hand side vector.
  gstrs_options: Pointer to the pangulu_gstrs_options structure.
  pangulu_handle: Pointer to the library internal state handle returned from initialisation.

3.4 pangulu_gssv()

- **Description**
  pangulu_gssv() solves the linear equation with A and right-hand size b. This function is equivalent to calling pangulu_gstrs() after pangulu_gstrf().
- **Function**

```
void pangulu_gssv(
  sparse_value_t *rhs,
  pangulu_gstrf_options *gstrf_options,
  pangulu_gstrs_options *gstrs_options,
  void **pangulu_handle);
```

- **Parameters**
  rhs: Pointer to the right-hand side vector.
  gstrf_options: Pointer to a pangulu_gstrf_options structure.
  gstrs_options: Pointer to a pangulu_gstrs_options structure.
  pangulu_handle: Pointer to the library internal status handle returned from initialisation.

3.5 pangulu_finalize()

- **Description**
  pangulu_finalize() is used to destroy a handle of PanguLU.
- **Function**

```
void pangulu_finalize(
  void **pangulu_handle);
```

- **Parameters**
  pangulu_handle: Pointer to the library internal state handle returned on initialisation.

4 Execution of an Example

4.1 example.c

The following is the content of example.c in the examples folder calling the PanguLU interfaces:

```
typedef unsigned long long int sparse_pointer_t;
#define MPI_SPARSE_POINTER_T MPI_UNSIGNED_LONG_LONG
#define FMT_SPARSE_POINTER_T "%llu"

typedef unsigned int sparse_index_t;
#define MPI_SPARSE_INDEX_T MPI_UNSIGNED
#define FMT_SPARSE_INDEX_T "%u"

#if defined(CALCULATE_TYPE_R64)
typedef double sparse_value_t;
#elif defined(CALCULATE_TYPE_R32)
typedef float sparse_value_t;
#elif defined(CALCULATE_TYPE_CR64)
typedef double _Complex sparse_value_t;
typedef double sparse_value_real_t;
#define COMPLEX_MTX
#elif defined(CALCULATE_TYPE_CR32)
typedef float _Complex sparse_value_t;
typedef float sparse_value_real_t;
#define COMPLEX_MTX
#else
typedef double sparse_value_t;
#error [example.c Compile Error] Unknown value type. Set -DCALCULATE_TYPE_CR64
or -DCALCULATE_TYPE_R64 or -DCALCULATE_TYPE_CR32 or -DCALCULATE_TYPE_R32 in
compile command line.
#endif
```

```
#include "../include/pangulu.h"
#include <sys/resource.h>
#include <getopt.h>
#include <stdio.h>
#include <mpi.h>
#include <math.h>
#include "mmio_highlevel.h"

#ifdef COMPLEX_MTX
sparse_value_real_t complex_fabs(sparse_value_t x)
{
    return sqrt(__real__(x) * __real__(x) + __imag__(x) * __imag__(x));
}

sparse_value_t complex_sqrt(sparse_value_t x)
{
    sparse_value_t y;
    __real__(y) = sqrt(complex_fabs(x) + __real__(x)) / sqrt(2);
    __imag__(y) = (sqrt(complex_fabs(x) - __real__(x)) / sqrt(2)) * (__imag__
    (x) > 0 ? 1 : __imag__(x) == 0 ? 0 : -1);
    return y;
}
#endif

void read_command_params(int argc, char **argv, char *mtx_name, char *rhs_
name, int *nb)
{
    int c;
    extern char *optarg;
    while ((c = getopt(argc, argv, "nb:f:r:")) != EOF)
    {
        switch (c)
        {
        case 'b':
            *nb = atoi(optarg);
            continue;
        case 'f':
            strcpy(mtx_name, optarg);
            continue;
```

```
        case 'r':
            strcpy(rhs_name, optarg);
            continue;
        }
    }
    if ((nb) == 0)
    {
        printf("Error : nb is 0\n");
        exit(1);
    }
}

int main(int ARGC, char **ARGV)
{
    // Step 1: Create varibles, initialise MPI environment.
    int provided = 0;
    int rank = 0, size = 0;
    int nb = 0;
    MPI_Init_thread(&ARGC, &ARGV, MPI_THREAD_MULTIPLE, &provided);
    MPI_Comm_rank(MPI_COMM_WORLD, &rank);
    MPI_Comm_size(MPI_COMM_WORLD, &size);
    sparse_index_t m = 0, n = 0, is_sym = 0;
    sparse_pointer_t nnz;
    sparse_pointer_t *colptr = NULL;
    sparse_index_t *rowidx = NULL;
    sparse_value_t *value = NULL;
    sparse_value_t *sol = NULL;
    sparse_value_t *rhs = NULL;

    // Step 2: Read matrix and rhs vectors.
    if (rank == 0)
    {
        char mtx_name[200] = {'\0'};
        char rhs_name[200] = {'\0'};
        read_command_params(ARGC, ARGV, mtx_name, rhs_name, &nb);

        switch (mtx_name[strlen(mtx_name) - 1])
        {
        case 'x':
```

```
            // mtx read (csc)
            printf("Reading mtx matrix %s\n", mtx_name);
            mmio_info(&m, &n, &nnz, &is_sym, mtx_name);
            colptr = (sparse_pointer_t *)malloc(sizeof(sparse_pointer_t) * (n
            + 1));
            rowidx = (sparse_index_t *)malloc(sizeof(sparse_index_t) * nnz);
            value = (sparse_value_t *)malloc(sizeof(sparse_value_t) * nnz);
            mmio_data_csc(colptr, rowidx, value, mtx_name);
            printf("Read mtx done.\n");
            break;
        case 'd':

            // lid read
            printf("Reading lid matrix %s\n", mtx_name);
            FILE *lid_file = fopen(mtx_name, "r");
            fread(&m, sizeof(sparse_index_t), 1, lid_file);
            fread(&n, sizeof(sparse_index_t), 1, lid_file);
            fread(&nnz, sizeof(sparse_pointer_t), 1, lid_file);
            sparse_pointer_t *rowptr = (sparse_pointer_t *)malloc(sizeof(
            sparse_pointer_t) * (n + 1));
            sparse_index_t *colidx = (sparse_
            index_t *)malloc(sizeof(sparse_index_t) * nnz);
            sparse_value_t *value_csr = (sparse
            _value_t *)malloc(sizeof(sparse_value_t) * nnz);
            fread(rowptr, sizeof(sparse_pointer_t), n + 1, lid_file);
            fread(colidx, sizeof(sparse_index_t), nnz, lid_file);
            fread(value_csr, sizeof(sparse_value_t), nnz, lid_file);
            fclose(lid_file);

            colptr = (sparse_pointer_t *)malloc(sizeof(sparse_pointer_t) * (n
            + 1));
            rowidx = (sparse_index_t *)malloc(sizeof(sparse_index_t) * nnz);
            value = (sparse_value_t *)malloc(sizeof(sparse_value_t) * nnz);
            memset(colptr, 0, sizeof(sparse_pointer_t) * (n + 1));
            sparse_pointer_t *trans_aid = (sparse_pointer_t *)malloc(sizeof(
            sparse_pointer_t) * n);
            memset(trans_aid, 0, sizeof(sparse_pointer_t) * n);
            for (sparse_index_t row = 0; row < n; row++)
            {
```

```
            for (sparse_pointer_t idx = rowptr[row]; idx < rowptr[row +
            1]; idx++)
            {
                sparse_index_t col = colidx[idx];
                colptr[col + 1]++;
            }
        }
        for (sparse_index_t row = 0; row < n; row++)
        {
            colptr[row + 1] += colptr[row];
        }
        memcpy(trans_aid, colptr, sizeof(sparse_pointer_t) * n);
        for (sparse_index_t row = 0; row < n; row++)
        {
            for (sparse_pointer_t idx = rowptr[row]; idx < rowptr[row +
            1]; idx++)
            {
                sparse_index_t col = colidx[idx];
                rowidx[trans_aid[col]] = row;
                value[trans_aid[col]] = value_csr[idx];
                trans_aid[col]++;
            }
        }
        free(rowptr);
        free(colidx);
        free(value_csr);
        free(trans_aid);
        printf("Read lid done.\n");

        break;
    }

    sol = (sparse_value_t *)malloc(sizeof(sparse_value_t) * n);
    rhs = (sparse_value_t *)malloc(sizeof(sparse_value_t) * n);
    memset(rhs, 0, sizeof(sparse_value_t) * n);
    if (rhs_name[0])
    {
        FILE *rhs_file = fopen(rhs_name, "r");
        if (!rhs_file)
        {
```

```
        fprintf(stderr, "Error: Failed to open rhs file '%s'\n",
                rhs_name);
        exit(1);
    }

    char line[512];
    int found_data = 0;
    while (fgets(line, sizeof(line), rhs_file))
    {
        if (line[0] != '%')
        {
            found_data = 1;
            break;
        }
    }
    if (!found_data)
    {
        fprintf(stderr, "Error: File '%s' contains only comments or
        is empty\n", rhs_name);
        fclose(rhs_file);
        exit(1);
    }
    int rhs_len;
    if (sscanf(line, "%d", &rhs_len) != 1)
    {
        fprintf(stderr, "Error: Failed to read vector dimension from
        '%s'\n", rhs_name);
        fclose(rhs_file);
        exit(1);
    }
    if (rhs_len != n)
    {
        fprintf(stderr, "Error: Vector dimension mismatch - expected
        %d, got %d\n", n, rhs_len);
        fclose(rhs_file);
        exit(1);
    }
    for (int i = 0; i < n; i++)
    {
        int read_success = 0;
```

```
#ifdef COMPLEX_MTX
                double real_part, imag_part;
                if (fscanf(rhs_file, "%le %le", &real_part, &imag_part) == 2)
                {
                    __real__(rhs[i]) = real_part;
                    __imag__(rhs[i]) = imag_part;
                    read_success = 1;
                }
#else
                if (fscanf(rhs_file, "%lf", &rhs[i]) == 1)
                {
                    read_success = 1;
                }
#endif
                if (!read_success)
                {
                    fprintf(stderr, "Error: Failed to read vector element %d
                    from '%s'\n", i, rhs_name);
                    fclose(rhs_file);
                    exit(1);
                }

                sol[i] = rhs[i];
            }

            fclose(rhs_file);
            printf("Successfully read rhs from '%s'\n", rhs_name);
        }
        else
        {
            if (!colptr || !value)
            {
                fprintf(stderr, "Error: Invalid matrix data for rhs generation
                \n");
                exit(1);
            }

            for(int i=0; i < n; i++){
                rhs[i] = 0;
            }
```

```
            for (int i = 0; i < n; i++)
            {
                for (sparse_pointer_t j = colptr[i]; j < colptr[i + 1]; j++)
                {
                    rhs[rowidx[j]] += value[j];
                }
            }
            for(int i=0; i < n; i++){
                sol[i] = rhs[i];
            }
            printf("Successfully generated rhs from matrix\n");
        }
        if((m != n) || m == 0){
            printf("Matrix A is %d * %d. Exit.\n", m, n);
            exit(1);
        }
    }
    MPI_Bcast(&n, 1, MPI_SPARSE_INDEX_T, 0, MPI_COMM_WORLD);
    MPI_Bcast(&nb, 1, MPI_INT, 0, MPI_COMM_WORLD);
    MPI_Bcast(&nnz, 1, MPI_LONG_LONG_INT, 0, MPI_COMM_WORLD);
    MPI_Barrier(MPI_COMM_WORLD);

    // Step 3: Initialise PanguLU solver.
    pangulu_init_options init_options;
    init_options.nb = nb;
    init_options.gpu_kernel_warp_per_block = 4;
    init_options.gpu_data_move_warp_per_block = 4;
    init_options.nthread = 1;
    init_options.hunyuan_nthread = 4;
    init_options.sizeof_value = sizeof(sparse_value_t);
    #ifdef COMPLEX_MTX
    init_options.is_complex_matrix = 1;
    #else
    init_options.is_complex_matrix = 0;
    #endif
    init_options.mpi_recv_buffer_level = 0.5;
    void *pangulu_handle;
    pangulu_init(n, nnz, colptr, rowidx, value, &init_options, &pangulu_handle
    );
```

```
// Step 4: Execute LU factorisation.
pangulu_gstrf_options gstrf_options;
pangulu_gstrf(&gstrf_options, &pangulu_handle);

// Step 5: Execute triangle solve using factorise results.
pangulu_gstrs_options gstrs_options;
pangulu_gstrs(sol, &gstrs_options, &pangulu_handle);
MPI_Barrier(MPI_COMM_WORLD);

// Step 6: Check the answer.
sparse_value_t *rhs_computed;
if (rank == 0)
{
    // Step 6.1: Calculate rhs_computed = A * x.
    rhs_computed = (sparse_value_t *)malloc(sizeof(sparse_value_t) * n);
    memset(rhs_computed, 0, sizeof(sparse_value_t) * n);
    for (int i = 0; i < n; i++)
    {
        for (sparse_pointer_t j = colptr[i]; j < colptr[i + 1]; j++)
        {
            rhs_computed[rowidx[j]] += value[j] * sol[i];
        }
    }

    // Step 6.2: Calculate residual residual = rhs_comuted - rhs.
    sparse_value_t *residual = rhs_computed;
    for (int i = 0; i < n; i++)
    {
        residual[i] = rhs_computed[i] - rhs[i];
    }
    sparse_value_t sum, c;
    // Step 6.3: Calculte norm2 of residual.
    sum = 0.0;
    c = 0.0;
    for (int i = 0; i < n; i++)
    {
        sparse_value_t num = residual[i] * residual[i];
        sparse_value_t z = num - c;
        sparse_value_t t = sum + z;
        c = (t - sum) - z;
```

```
            sum = t;
        }
#ifdef COMPLEX_MTX
        sparse_value_real_t residual_norm2 = complex_fabs(complex_sqrt(sum));
#else
        sparse_value_t residual_norm2 = sqrt(sum);
#endif
        // Step 6.4: Calculte norm2 of original rhs.
        sum = 0.0;
        c = 0.0;
        for (int i = 0; i < n; i++)
        {
            sparse_value_t num = rhs[i] * rhs[i];
            sparse_value_t z = num - c;
            sparse_value_t t = sum + z;
            c = (t - sum) - z;
            sum = t;
        }
#ifdef COMPLEX_MTX
        sparse_value_real_t rhs_norm2 = complex_fabs(complex_sqrt(sum));
#else
        sparse_value_t rhs_norm2 = sqrt(sum);
#endif
        // Step 6.5: Calculate relative residual.
        double relative_residual = residual_norm2 / rhs_norm2;
        printf("|| Ax - b || / || b || = %le\n", relative_residual);
    }
    // Step 7: Clean and finalize.
    pangulu_finalize(&pangulu_handle);
    if (rank == 0)
    {
        free(colptr);
        free(rowidx);
        free(value);
        free(sol);
        free(rhs);
        free(rhs_computed);
    }
    MPI_Finalize();
}
```

4.2 Execution

After installing PanguLU, the example.c file is automatically compiled as part of the library. Subsequently, the following instructions can be referenced for execution:

```
mpirun -np process_count ./pangulu_example.elf -nb block_size -f path_to_mtx
-r path_to_rhs
```

process_count: MPI process number to launch PanguLU;
block_size: Rank of each non-zero block;
path_to_mtx : The matrix path in mtx format;
path_to_rhs : The path of the right-hand side vector. (optional)
You can also use the run.sh command:

```
bash ./run.sh path_to_mtx block_size process_count
```

In the following example, four MPI processes are used, the matrix block size is 10, and the matrix file is Trefethen_20b.mtx with the id 2203 in the SuiteSparse Matrix Collection [5]:

```
mpirun -np 4 ./pangulu_example.elf -nb 10 -f Trefethen_20b.mtx
```

or use the run.sh in this way:

```
./run.sh Trefethen_20b.mtx 10 4
```

4.3 Output

If the execution above runs successfully, you will get an output similar to the following:

```
Reading mtx matrix Trefethen_20b.mtx
Read mtx done.
Successfully generated rhs from matrix
[PanguLU Info]
[PanguLU Info] --- PanguLU Configuration & Matrix Info ---
[PanguLU Info]       Matrix Order:             19
[PanguLU Info]       #NNZ:                     147
[PanguLU Info]       Matrix Block Order (nb):  10
[PanguLU Info]       MPI Processes Count:      4
[PanguLU Info]       MPI Recv Buffer Level:    0.50
[PanguLU Info]       Hunyuan Index Type:       i32
[PanguLU Info]       Hunyuan Thread Count:     4
[PanguLU Info]       GPU:                      Enabled
[PanguLU Info]       GPU Kernel Warp/Block:    4
[PanguLU Info]       GPU Data Move Warp/Block: 4
[PanguLU Info] ----------------------------------------
[PanguLU Info]
[PanguLU Info] Reordering time: 0.001703 s
[PanguLU Info] Symbolic nonzero count: 243
[PanguLU Info] Symbolic factorisation time: 0.000015 s
[PanguLU Info] Preprocessing time: 0.094246 s
[PanguLU Info] Numeric factorisation time: 0.225403 s
[PanguLU Info] Host memory usage: 1514 MiB (bytes = 1587941376)
[PanguLU Info] GPU memory usage: 1228 MiB (bytes = 1288175616)
[PanguLU Info] Solving time (SpTRSV): 0.000037 s
|| Ax - b || / || b || = 1.572899e-16
```

5 History Versions

Version 5.0.0 (Jul. 31, 2025)

- Added a task aggregator to increase numeric factorisation performance.
- Optimised performance of preprocessing phase.
- Added the Hunyuan reordering algorithm on CPU as the default reordering algorithm.
- Optimised GPU memory layout to reduce GPU memory usage.

Version 4.2.0 (Dec. 13, 2024)

- Updated preprocessing phase to distributed data structure.

Version 4.1.0 (Sep. 01, 2024)

- Optimised memory usage of numeric factorisation and solving.
- Added parallel building support.

Version 4.0.0 (Jul. 24, 2024)

- Optimised user interfaces of solver routines.
- Optimised performance of numeric factorisation phase on CPU platforms.
- Added support on complex matrix solving.
- Optimised pre-processing performance.

Version 3.5.0 (Aug. 06, 2023)

- Updated the pre-processing phase with OpenMP.
- Updated the compilation method, compilling libpangulu.so and libpangulu.a at the same time.
- Updated timing for the reordering phase, the symbolic factorisation phase, and the pre-processing phase.
- Computed GFLOPS for the numeric factorisation phase.

Version 3.0.0 (Apr. 02, 2023)

- Used an adaptive method for selecting sparse BLAS in the numeric factorisation phase.

- Added the reordering phase.
- Added the symbolic factorisation phase.
- Added the MC64 algorithm in the reordering phase.
- Added an interface for 64-bit METIS package in the reordering phase.

Version 2.0.0 (Jul. 22, 2022)

- Used a synchronisation-free scheduling strategy in the numeric factorisation phase.
- Updated the MPI communication method in the numeric factorisation phase.
- Added single precision in the numeric factorisation phase.

Version 1.0.0 (Oct. 19, 2021)

- Used a rule-based 2D LU factorisation scheduling strategy.
- Used sparse BLAS for floating point calculations on GPUs.
- Added the pre-processing phase.
- Added the numeric factorisation phase.
- Added the triangular solve phase.

6 License

PanguLU uses the GNU Affero General Public License 3.0, details of which can be found at the following link:
https://github.com/SuperScientificSoftwareLaboratory/PanguLU?tab=AGPL-3.0-1-ov-file

7 Contributors

The project is leaded by Prof. Weifeng Liu. Active contributors are Yida Li, Siwei Zhang, Wenxuan Li, Helin Cheng, Yiduo Niu and Yang Du.

If you have any questions on PanguLU, please contact the major developer Yida Li by emailing yida.li@student.cup.edu.cn.

Before the year 2025, the contributors include Xu Fu, Bingbin Zhang, Tengcheng Wang, Wenhao Li, Yuechen Lu, Enxin Yi, Jianqi Zhao, Xiaohan Geng, Fangying Li, Jingwen Zhang, Zhou Jin, Chisen Wang and Zhengye Ye.

8 Citing PanguLU

```
@inproceedings{10.1145/3581784.3607050,
author = {Fu, Xu and Zhang, Bingbin and Wang, Tengcheng and Li,
Wenhao and Lu, Yuechen and Yi, Enxin and Zhao, Jianqi and Geng, Xiaohan
and Li, Fangying and Zhang, Jingwen and Jin, Zhou and Liu, Weifeng},
title = {PanguLU: A Scalable Regular Two-Dimensional Block-Cyclic
Sparse Direct Solver on Distributed Heterogeneous Systems},
year = {2023},
isbn = {9798400701092},
publisher = {Association for Computing Machinery},
address = {New York, NY, USA},
url = {https://doi.org/10.1145/3581784.3607050},
doi = {10.1145/3581784.3607050},
booktitle = {Proceedings of the International Conference for High
Performance Computing, Networking, Storage and Analysis},
articleno = {51},
numpages = {14},
keywords = {sparse LU, regular 2D block, distributed heterogeneous systems},
location = {Denver, CO, USA},
series = {SC '23}
}
```

Bibliography

[1] Xu Fu, Bingbin Zhang, Tengcheng Wang, Wenhao Li, Yuechen Lu, Enxin Yi, Jianqi Zhao, Xiaohan Geng, Fangying Li, Jingwen Zhang, Zhou Jin, and Weifeng Liu: “PanguLU: A Scalable Regular Two-Dimensional Block-Cyclic Sparse Direct Solver on Distributed Heterogeneous Systems”, Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC ’23 (2023) doi:10.1145/3581784.3607050

[2] Wikipedia: *Pangu*, https://en.wikipedia.org/wiki/Pangu, 2007

[3] Karypis, George and Kumar, Vipin: “A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs”, SIAM Journal on Scientific Computing **20**, 359–392 (1998) doi:10.1137/S1064827595287997

[4] Iain S. Duff and Jacko Koster: “On Algorithms For Permuting Large Entries to the Diagonal of a Sparse Matrix”, SIAM Journal on Matrix Analysis and Applications **22**, 973–996 (2000) doi:10.1137/S0895479899358443

[5] Timothy A. Davis and Yifan Hu: “The University of Florida sparse matrix collection”, ACM Transactions on Mathematical Software **38**, 1–25 (2011) doi:10.1145/2049662.2049663