User's Guide for ParU, an unsymmetric multifrontal multithreaded sparse LU factorization package

Mohsen Aznaveh[*], Timothy A. Davis

VERSION 0.0.0, May 20, 2022

Abstract

ParU is an implementation of the multifrontal sparse LU factorization method. Parallelism is exploited both in the BLAS and across different frontal matrices using OpenMP tasking, a shared-memory programming model for modern multicore architectures. The package is written in C++ and real sparse matrices are supported.

1 Introduction

The algorithms used in ParU will be discussed in a companion paper, ?. This document gives detailed information on the installation and use of ParU. ParU is a parallel sparse direct solver. This package uses OpenMP tasking for parallelism. ParU calls UMFPACK for the symbolic analysis phase, after that some symbolic analysis is done by ParU itself and then the numeric phase starts. The numeric computation is a task parallel phase using OpenMP and each task calls parallel BLAS; i.e. nested parallelism. The performance of BLAS has a heavy impact on the performance of ParU. However, depending on the input problem performance of parallelism in BLAS sometimes does not have effects in ParU.

1.0.1 Instructions on using METIS

SuiteSparse is now on METIS 5.1.0, which is distributed along with SuiteSparse itself. Its use is optional, however. ParU is using METIS as the default ordering. METIS tends to give orderings that are good for the parallelism. You can compile and run your code without using METIS; We recommend using METIS along with ParU.

Note that METIS is not bug-free; it can occasionally cause segmentation faults, particularly if used when finding basic solutions to underdetermined systems with many more columns than rows. ParU does not solve such systems anyway but you might see some problems with other SuiteSparse packages.

[*]email: aznaveh@tamu.edu. http://www.suitesparse.com.

2 Using ParU in C and C++

ParU relies on CHOLMOD for its basic sparse matrix data structure, a compressed sparse column format. CHOLMOD provides interfaces to the AMD, COLAMD, and METIS ordering methods, writing a matrix to a file, and many other functions. ParU also relies on UMFPACK Version 6.0 or higher for symbolic analysis.

2.1 Installing the C/C++ library on Linux/Unix

Before you compile the ParU library and demo programs, you may wish to edit the `SuiteSparse/SuiteSparse_config/SuiteSparse_config.mk` configuration file. The defaults should be fine on most Linux/Unix systems and on the Mac. It automatically detects what system you have and sets compile parameters accordingly.

The configuration file defines where the LAPACK and BLAS libraries are to be found. Selecting the right BLAS is critical. There is no standard naming scheme for the name and location of these libraries. The defaults in the `SuiteSparse_config.mk` file use `-llapack` and `-lblas`; the latter may link against the standard Fortran reference BLAS, which will not provide optimal performance. For best results, you should use the OpenBLAS at openblas.net (based on the Goto BLAS) [10], or high-performance vendor-supplied BLAS such as the Intel MKL, AMD ACML, or the Sun Performance Library. Selection of LAPACK and the BLAS is done with the `LAPACK=` and `BLAS=` lines in the `SuiteSparse_config.mk` file.

There are two parts that are important in chosing the compiler and `BLAS` library.

`'AUTOCC`?= yes' This line let `SuiteSparse_config` choose the compiler automatically. If there is an Intel compiler available it will be chosen. If you change `yes` to `no` then GCC will be used for the compilation.

`'BLAS`?= -lopenblas' This line let `SuiteSparse_config` choose the `BLAS` library. By default ParU uses `openBLAS`. If you comment out this line ParU will look for the Intel Math Kernel Library.

After you decide about the compiler and `BLAS` library, type `make` at the Linux/Unix command line, in either the `SuiteSparse` directory (which compiles all of SuiteSparse) or in the `SuiteSparse/ParU` directory (which just compiles ParU and the libraries it requires)???. ParU will be compiled, and a set of simple demos will be run (including the one in the next section).

To test the lines of ParU, go to the `Tcov` directory and type `make`. To fully test 100% of the lines of ParU you should define `PARU_ALLOC_TESTING` and `PARU_COVERAGE` in `ParU\Source\paru_internal.hpp`. This will work for Linux only.

To install the shared library into /usr/local/lib and /usr/local/include, do `make install`. To uninstall, do `make uninstall`. For more options, see the `SuiteSparse/README.txt` file.

2.2 C/C++ Example

The C++ interface is written using only real matrices. The simplest function computes the MATLAB equivalent of `x=A\b` and is almost as simple: Below is a simple C++ program that illustrates the use of ParU. The program reads in a problem from `stdin` in MatrixMarket format [3], solves it, and prints the norm of `A` and the residual. Some error testing code is omited to only show how the program works. The full program can be found in `Paru/Demo/paru_demo.cpp`

```cpp
#include "ParU.hpp"
int main(int argc, char **argv)
{
    cholmod_common Common, *cc;
    cholmod_sparse *A;
    ParU_Symbolic *Sym = NULL;

    //~~~~~~~~~Reading the input matrix and test if the format is OK~~~~~~~~~~~~
    // start CHOLMOD
    cc = &Common;
    int mtype;
    cholmod_l_start(cc);

    // A = mread (stdin) ; read in the sparse matrix A
    A = (cholmod_sparse *)cholmod_l_read_matrix(stdin, 1, &mtype, cc);
    //~~~~~~~~~~~~~~~~~~~Starting computation~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
    ParU_Control Control;
    ParU_Ret info;
    info = ParU_Analyze(A, &Sym, &Control);
    ParU_Numeric *Num;
    info = ParU_Factorize(A, Sym, &Num, &Control);
    double my_time = omp_get_wtime() - my_start_time;
    //~~~~~~~~~~~~~~~~~~~Test the results ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
    Int m = Sym->m;
    if (info == PARU_SUCCESS)
    {
        double *b = (double *)malloc(m * sizeof(double));
        double *xx = (double *)malloc(m * sizeof(double));
        for (Int i = 0; i < m; ++i) b[i] = i + 1;
        info = ParU_Solve(Sym, Num, b, xx, &Control);
        printf("Solve time is %lf seconds.\n", my_solve_time);
        double resid, anorm;
        info = ParU_Residual(A, xx, b, m, resid, anorm, &Control);
        printf("Residual is |%.2lf| and anorm is %.2e and rcond is %.2e.\n",
                resid == 0 ? 0 : log10(resid), anorm, Num->rcond);
        free(b);
        free(xx);
    }
    //~~~~~~~~~~~~~~~~~~~End computation~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
    Int max_threads = omp_get_max_threads();
    BLAS_set_num_threads(max_threads);
```

```
    //~~~~~~~~~~~~~~~~~~~Free Everything~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
    ParU_Freenum(&Num, &Control);
    ParU_Freesym(&Sym, &Control);

    cholmod_l_free_sparse(&A, cc);
    cholmod_l_finish(cc);
}
```

2.3 C/C++ Syntax

`ParU_Ret` is the output structure of all ParU routines. The user must check the output before continuing and using the result of prior routine. The following is a list of user-callable C++ functions and what they can do:

1. `ParU_Version`: return the version of the ParU package you are using.

2. `ParU_Analyze`: Symbolic analysis is done in this routine. UMFPACK is called here and after that, some more specialized symbolic computation is done for ParU. `ParU_Analyze` called once and used for different `ParU_Factorize` calls.

3. `ParU_Factorize`: Numeric factorization is done in this routine. Scaling and making Sx matrix, computing factors and permutations is here. `ParU_Symbolic` structure is computed `ParU_Analyze` and is an input in this routine.

4. `ParU_Solve`: Using symbolic analysis and factorization phase output to solve $Ax = b$. In all the solve routines Num structure must come with the same Sym struct that comes from `ParU_Factorize`. This routine is overloaded and can solve differently. It has versions that keep a copy of x or overwrite it. Also, it can solve multiple right-hand side problems.

5. `ParU_Freenum`: frees the numerical part of factorization.

6. `ParU_Freesym`: frees the symbolic part of factorization.

2.4 Details of the C/C++ Syntax

For further details on how to use the C/C++ syntax, please refer to the definitions and descriptions in the following files:

1. `SuiteSparse/ParU/Include/ParU.hpp` describes each C++ function. Only `double` and square matrices are supported.

2. `SuiteSparse/ParU/Include/ParU.h` describes the C-callable functions.

There are C/C++ options to control ParU which is an input argument to several routines. When you make `ParU_Control` object it is initialized with default values. The user can change the values. Here is the list of control options:

Other parameters, such as `opts.ordering` and `opts.tol`, are input parameters to the various C/C++ functions. Others such as `opts.solution='min2norm'` are separate functions in the C/C++ interface. Refer to the files listed above for details. Output statistics include:

`ParU_Control`	default value	explanation
`mem_chunk`	$1024 * 1024$	chunk size for memset and memcpy
`umfpack_ordering`	`UMFPACK_ORDERING_METIS`	default UMFPACK ordering
`umfpack_strategy`	`UMFPACK_STRATEGY_AUTO`	default UMFPACK strategy
`relaxed_amalgamation_threshold`	32	threshold for relaxed amalgamation
`scale`	1	if 1 matrix will be scaled using `max_row`
`panel_width`	32	width of panel for dense factorizaiton
`paru_strategy`	`PARU_STRATEGY_AUTO`	default strategy for ParU
`piv_toler`	0.1	tolerance for accepting sparse pivots
`diag_toler`	0.001	tolerance for accepting symmetric pivots
`trivial`	4	Do not call BLAS for smaller dgemms
`worthwhile_dgemm`	512	dgemms bigger than worthwhile are tasked
`worthwhile_trsm`	4096	trsm bigger than worthwhile are tasked
`paru_max_threads`	0	initialized with `omp_max_threads`

The first row of the options is used in the symbolic analysis. In the symbolic analysis phase, only the pattern of the matrix and numerical values are not probed. The second row has an impact on numerical analysis.

`paru_max_threads` is initalized by `omp_max_threads` if the user do not provide a smaller number. If `paru_strategy` is set to `PARU_STRATEGY_AUTO` ParU uses the same strategy as UMFPACK, however the user can ask UMFPACK for a unsymmetric strategy but use a symmetric strategy for ParU.

3 Requirements and Availability

ParU requires several Collected Algorithms of the ACM: CHOLMOD [4, 7] (version 1.7 or later), AMD [1, 2], COLAMD [5, 6] and UMFPACK [8] for its ordering/analysis phase and for its basic sparse matrix data structure, and the BLAS [9] for dense matrix computations on its frontal matrices. An efficient implementation of the BLAS is strongly recommended, either vendor-provided (such as the Intel MKL, the AMD ACML, or the Sun Performance Library) or other high-performance BLAS such as those of [10]. Note that while ParU uses nested parallelism heavily the right options for BLAS library must be chosen.

The use of OpenMP tasking is optional, but without it, only parallelism within the BLAS can be exploited (if available). See ParU/Doc/LICENSE for the license. Alternative licenses are also available; contact the author for details.

References

[1] P. R. Amestoy, T. A. Davis, and I. S. Duff. An approximate minimum degree ordering algorithm. *SIAM J. Matrix Anal. Appl.*, 17(4):886–905, 1996.

[2] P. R. Amestoy, T. A. Davis, and I. S. Duff. Algorithm 837: AMD, an approximate minimum degree ordering algorithm. *ACM Trans. Math. Software*, 30(3):381–388, 2004.

[3] R. F. Boisvert, R. Pozo, K. Remington, R. Barrett, and J. J. Dongarra. The Matrix Market: A web resource for test matrix collections. In R. F. Boisvert, editor, *Quality of Numerical Software, Assessment and Enhancement*, pages 125–137. Chapman & Hall, London, 1997. (`http://math.nist.gov/MatrixMarket`).

[4] Y. Chen, T. A. Davis, W. W. Hager, and S. Rajamanickam. Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate. *ACM Trans. Math. Software*, 35(3), 2009.

[5] T. A. Davis, J. R. Gilbert, S. I. Larimore, and E. G. Ng. Algorithm 836: COLAMD, a column approximate minimum degree ordering algorithm. *ACM Trans. Math. Software*, 30(3):377–380, 2004.

[6] T. A. Davis, J. R. Gilbert, S. I. Larimore, and E. G. Ng. A column approximate minimum degree ordering algorithm. *ACM Trans. Math. Software*, 30(3):353–376, 2004.

[7] T. A. Davis and W. W. Hager. Dynamic supernodes in sparse Cholesky update/downdate and triangular solves. *ACM Trans. Math. Software*, 35(4), 2009.

[8] Timothy A. Davis. Algorithm 832: Umfpack v4.3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Softw.*, 30(2):196–199, jun 2004.

[9] J. J. Dongarra, J. J. Du Croz, I. S. Duff, and S. Hammarling. A set of Level 3 Basic Linear Algebra Subprograms. *ACM Trans. Math. Software*, 16:1–17, 1990.

[10] K. Goto and R. van de Geijn. High performance implementation of the level-3 BLAS. *ACM Trans. Math. Software*, 35(1):4, July 2008. Article 4, 14 pages.