

## USPARS direct linear solver

D. Bykov, M. Cherepanov, V. Kostin, A. Semenov, S. Solovyev

### Unipro

6 октября 2022 г.

## Оглавление

Введение

Архитектура

PyUspars

Тестирование

Результаты

Текущий статус работ

Вопросы

## USPARS - прямой решатель разреженных СЛАУ

Поддерживаемые архитектуры и реализации BLAS:

- ▶ **x86** - **mkl**, FLAME, OpenBlas
- ▶ arm - Apple Accelerate, OpenBlas
- ▶ e2k - EML (Elbrus Math Library)

## Известные прямые решатели

Прямые решатели разреженных матриц (\* - с открытым исходным кодом):

- ▶ \*TAUCS <https://www.tau.ac.il/~stoledo/taucs/> [v 2.2, 2003].
- ▶ \*UMFPACK <https://people.engr.tamu.edu/davis/suitesparse.html> [v. 5.10.1, 2021];
- ▶ PARDISO <https://www.pardiso-project.org/>
- ▶ MKL PARDISO <https://software.intel.com/content/www/us/en/develop/tools/oneapi/components/onemkl.html#gs.613p2m> [v. 2021.3, 2021].
- ▶ \*MUMPS <http://mumps.enseeiht.fr/> [v. 5.5.0, 2022].
- ▶ \*SuperLU <https://portal.nersc.gov/project/sparse/superlu/>

## Предпосылки к разработке собственного прямого решателя

- ✓ Компетенции в разработке ПО.
- ✓ Наличие заказчиков.
- ✓ Оперативная и надежная поддержка для российских потребителей
- ✓ Использование в собственных продуктах.

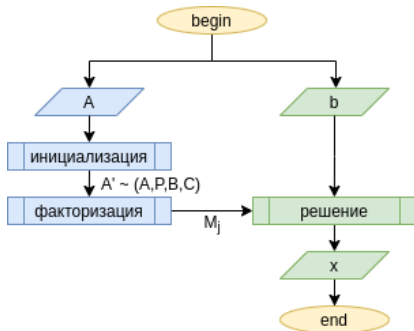
## Схема вычислений

$$Ax = b$$

$$A' = BPAP^T C$$

$P$  – матрица перестановок,  $B, C$  – невырожденные

$$A' = \prod_{j \in J} M_j, M_j \text{ – треугольные}$$



## Инициализация

$A' = BPA P^T C$ , для положительно определенных  $B = C = E$

- ▶  $P$  перестановка, минимизирующая заполняемость ненулевыми элементами треугольных факторов (fill-in)
  - Metis (<http://glaros.dtc.umn.edu/gkhome/metis/metis/download>).
  - Atlant (собственная разработка).
- ▶ В случае неопределенных матриц  $B = D_l * P_1$ ,  $C = D_r$ , где:
  - $P_1$ : перемещение больших элементов  $A$  на главную диагональ (matching);
  - $D_l, D_r$  - масштабирование элементов матрицы  $A$  (scaling)

## Минимизация fill-in эффекта

Пример:

$$A = \begin{pmatrix} 9 & 1.5 & 6 & 0.75 & 3 \\ 1.5 & 0.5 & & & \\ 6 & & 12 & & \\ 0.75 & & & 0.625 & \\ 3 & & & & 16 \end{pmatrix}.$$

$A = MM^T$ , где:

$$M = \begin{pmatrix} 3 & & & & \\ 0.5 & 0.5 & & & \\ 2 & -2 & 2 & & \\ 0.25 & -0.25 & -0.5 & 0.5 & \\ 1 & -1 & -2 & -3 & 1 \end{pmatrix}.$$



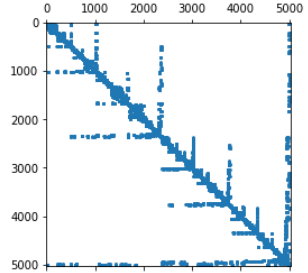
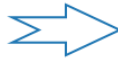
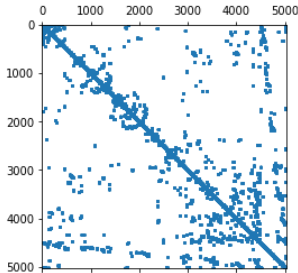
$$A' = PAP^T = M' M'^T = \begin{pmatrix} 16 & & & & 3 \\ & 0.5 & & & 0.75 \\ & & 12 & & 6 \\ & & & 0.625 & 1.5 \\ 3 & 1.5 & 6 & 0.75 & 9 \end{pmatrix}$$

где:

$$M' = \begin{pmatrix} 4 & & & & \\ & \frac{1}{\sqrt{2}} & & & \\ & & 2\sqrt{3} & & \\ & & & \frac{\sqrt{10}}{4} & \\ \frac{3}{4} & \frac{3}{\sqrt{2}} & \sqrt{3} & \frac{3}{\sqrt{10}} & \sqrt{\frac{3}{80}} \end{pmatrix}.$$

# Atlant

- ▶ Основан на методе вложенных сечений<sup>1</sup>
- ▶ Помимо вектора перестановок возвращает дерево сепараторов



<sup>1</sup>Gilbert, J. R., & Tarjan, R. E. (1986). The analysis of a nested dissection algorithm. *Numerische mathematik*, 50(4), 377-404.

## Типы факторизаций

- ▶ Матрицы общего вида:

$$A = LU$$

- ▶ Положительно определенные матрицы:

$$A = \begin{cases} LL^T & \text{симметричная действительная} \\ LL^H & \text{комплексная Эрмитова} \end{cases}$$

- ▶ Неопределенные матрицы:

$$A = \begin{cases} LDL^T & \text{симметричная} \\ LDL^H & \text{комплексная Эрмитова} \end{cases}$$

$L$ ,  $U$  - ниже- и верхнетреугольная матрицы.

$D$  - диагональная.



## Python-пакет. Преимущества

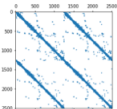
- ▶ Удобный пользовательский интерфейс
  - Простой (близкий к естественному) синтаксис задания запросов на решение
  - Автоматический контроль за выделением, распределением и освобождением динамической памяти
- ▶ Расширенный функционал, направленный на верификацию полученного решения
  - Расчет невязки, погрешности
  - Визуализации исходной матрицы, решения, перестановок
  - Нахождение числа обусловленности матрицы
- ▶ Используется в качестве инструмента для написания и тестирования нового функционала

Пример решения системы с визуализацией матрицы:

Считывание матрицы, задание вектора решения и правой части

```

[3]: A = mread("1.mat").tocsr()
     x_exact = [1]*A.shape[0]
     b = 4*x_exact
     all spy(A, markersize=1)
[4]: outplotlib.lines.Line2D at 0x7f79467ca340
    
```



Решение

```

[4]: usp = PyUspars()
     usp.mgl=100
     usp.init(A, ftype='i')
     usp.fact()
     x = usp.solve(b)
[5]: print("Error: ", usp.calc_norm_error(x_exact))
     print("Residual: ", usp.calc_norm_residual())
     Error = 1.4950543368204521e-12
     Residual = 7.109688709564373e-14
    
```



## Стек технологий

c++, python, gtest, git, Teamcity

## Тестовые данные

- Sparse Suite Matrix Collection (<https://sparse.tamu.edu>) - порядка 900 матриц.
- Матрицы заказчика
- Специальные параметризованные тесты (уравнения Лапласа, Гельмгольца, Сильвестра)

## Методика

- Ночное тестирование (Teamcity)
- Сравнение с референсным солвером (MKL Pardiso, MUMPS).
- Web-сервис генерации отчетов сравнения производительности
- Встроенные инструменты профилирования
- Динамическое конфигурирование тестовой сьюты

## Характеристики солверов

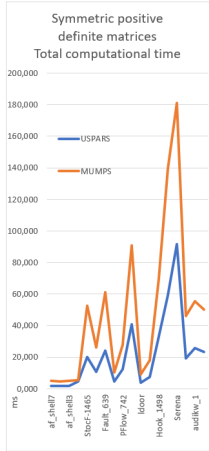
	USPARS	MUMPS
HW	Server 2 × 12 cores Intel <sup>®</sup> Xeon <sup>®</sup> CPU E5-2680 v3 @2.50 GHz	
Version	1.0.2	5.5.0
Math libs	MKL 2021.4	
Compiler	Intel <sup>®</sup> C Compiler version 2021.5.0	
Parallel	24 OpenMP threads	2 MPI processes with 12 OpenMP threads on each
Reorder	Atlant	METIS, SCOTCH (auto)



# Симметричные положительно определенные матрицы

matrix	nnz	nnz(L)		ratio1	relative error		ratio2
		MUMPS	USPARS		MUMPS	USPARS	
af_shell7	9046865	102946987	99031758	0,962	3,14E-13	1,61E-12	5,127
af_shell4	9046865	102946987	99018958	0,962	3,52E-13	2,22E-12	6,304
af_shell3	9046865	102946987	99012958	0,962	3,54E-13	2,22E-12	6,258
bundle_adj	10360701	16507619	11258613	0,682	8,32E-06	6,08E-06	0,731
StocF-1465	11235263	1132720311	1072093660	0,946	2,36E-10	4,03E-10	1,705
nd24k	14393817	368963954	323383089	0,876	1,04E-11	9,58E-11	9,180
Fault_639	14626683	1166161107	1096124979	0,940	9,55E-13	1,4E-12	1,469
inline_1	18660027	194845761	177012139	0,908	9,38E-11	3,96E-11	0,422
PFlow_742	18940627	524283329	528559035	1,008	1,1E-09	8,43E-09	7,663
Emilia_923	20964171	1775495567	1670123520	0,941	4,24E-12	1,78E-11	4,208
ldoor	23737339	160068701	153212616	0,957	1,03E-13	9,37E-14	0,908
boneS10	28191660	308344731	287554709	0,933	1,01E-11	5,73E-11	5,656
Hook_1498	31207734	1584091368	1528293986	0,965	1,56E-13	1,92E-13	1,234
Geo_1438	32297325	2700735099	2484253965	0,920	1,97E-14	8,17E-14	4,154
Serena	32961525	2959377990	2721768419	0,920	3,78E-15	8,37E-15	2,212
bone010	36326514	1150148226	1087379536	0,945	1,06E-11	4,44E-10	41,764
audikw_1	39297771	1322970878	1244750851	0,941	8,06E-13	8,43E-13	1,045
Flan_1565	59485419	1593631358	1510009271	0,948	1,76E-12	2,67E-12	1,512

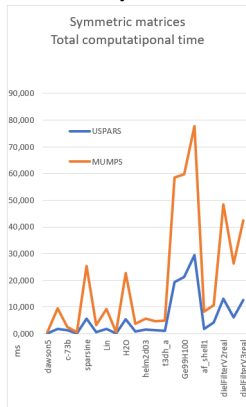
Время (мс).  
Меньше - лучше



## Симметричные матрицы

matrix	nnz	nnz(L)		ratio1	relative error		
		MUMPS	USPARS		MUMPS	USPARS	ratio2
dawson5	531157	4721299	4300301	0,91	3,00E-11	2,31E-13	7,72E-03
SiO	675528	100363897	85543671	0,85	1,08E-11	5,64E-15	5,22E-04
c-73b	724348	4346469	1698020	0,39	4,75E-12	3,25E-12	6,84E-01
nemeth24	758028	1299752	2069033	1,59	3,09E-13	2,37E-15	7,68E-03
sparsine	799494	237817606	197894507	0,83	4,41E-07	6,10E-13	1,38E-06
gas_sensor	885141	25763054	23104318	0,90	3,00E-11	4,61E-14	1,54E-03
Lin	1011200	117075659	96477401	0,82	1,29E-11	8,71E-14	6,76E-03
bcsstk39	1068033	8727759	8332118	0,95	2,25E-11	2,12E-16	9,40E-06
H2O	1141880	268234557	230757614	0,86	4,08E-14	6,01E-15	1,48E-01
filter3D	1406808	21603250	19718284	0,91	2,25E-13	5,92E-15	2,63E-02
helm2d03	1567096	21157163	20051775	0,95	3,55E-11	1,44E-15	4,05E-05
t3dh	2215638	52667533	47520559	0,90	5,09E-05	8,90E-14	1,75E-09
t3dh_a	2215638	52667533	47520559	0,90	5,09E-05	9,16E-14	1,80E-09
Ga10As10H30	3114357	704999994	607125460	0,86	1,25E-10	4,14E-13	3,32E-03
Ge99H100	4282190	685164834	618819749	0,90	9,81E-11	1,79E-12	1,82E-02
Ga19As19H42	4508981	872647798	792103664	0,91	7,01E-11	1,16E-13	1,65E-03
af_shell1	9046865	102946987	99012958	0,96	9,28E-09	9,32E-14	1,00E-05
F1	13590452	195634237	177356513	0,91	3,60E-13	7,16E-16	1,99E-03
dielFilterV2real	24848204	588430085	560750420	0,95	6,77E-11	3,89E-16	5,75E-06
af_shell10	27090195	379263437	366868082	0,97	3,11E-09	5,04E-15	1,62E-06
dielFilterV3real	45204422	608652052	564191186	0,93	7,02E-11	1,40E-15	1,99E-05

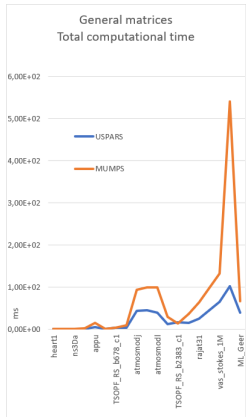
Время (мс).  
Меньше - лучше



## Матрицы общего вида

matrix	nnz	Nnz in factors		ratio1	relative error		ratio2
		MUMPS (LU)	USPARS (L)		MUMPS	USPARS	
heart1	1387773	3627829	2429037	1,34	2,94E-13	6,14E-11	2,09E+02
TSOPF_RS_b300_c1	1474325	13023742	5553581	0,86	1,39E-13	2,81E-13	2,03E+00
ns3Da	1679599	16847462	9586970	1,14	2,26E-15	1,09E-11	4,81E+03
bbmat	1771722	40233520	18567567	0,92	6,31E-07	1,06E-01	1,68E+05
appu	1853104	183983754	85009023	0,92	1,81E-14	1,70E-14	9,37E-01
sme3Db	2081063	15310395	7527021	0,98	8,46E-12	3,87E-12	4,57E-01
TSOPF_RS_b678_c1	4396289	39553712	15952209	0,80	5,94E-11	6,33E-06	1,06E+05
torso3	4429042	245961594	117351614	0,96	3,04E-16	3,85E-16	1,27E+00
atmosmodj	8814880	2012471658	908934059	0,90	3,4E-13	4,02E-14	1,18E-01
atmosmodd	8814880	2012471658	908934059	0,90	2,3E-14	2,31E-14	1,00E+00
atmosmodl	10319760	2025401512	926688713	0,92	3,95E-15	5,52E-14	1,40E+01
memchip	14810202	68222032	36728178	1,08	3,86E-15	3,15E-15	8,17E-01
TSOPF_RS_b2383_c1	16171169	146435070	57582501	0,78	7E-06	2,45E-03	3,51E+02
Freescal1	18920347	53386061	30957569	1,16	4,18E-11	3,77E-11	9,03E-01
rajat31	20316253	220198562	115880629	1,06	3,32E-13	1,71E-13	5,16E-01
Transport	23500731	2621344403	1301199003	1,00	6,51E-13	3,23E-12	4,96E+00
vas_stokes_1M	34767207	3099869506	1505182461	0,98	8,78E-14	2,32E-13	2,64E+00
circuit5M	59524291	102781446	36268611	0,70	2,04E-10	6,04E-10	2,95E+00
ML_Geer	1,11E+08	2053642052	1051106528	1,02	1,99E-14	4,86E-14	2,44E+00

Время (мс).  
Меньше - лучше



▶ 2022

- **USPARS 1.0** (выпуск первой релизной версии)
- Итерационное уточнение
- Написание собственного реордеринга
- Параллелизация стадии инициализации
- Параллелизация стадии решения с одной и с несколькими правыми частями
- Блочные варианты параллельной факторизации

▶ 2023-...

- Портирование и оптимизация работы USPARS на **других платформах** (arm, e2k)
- Переход к **кластерному MPI** солверу и реордерингу
- Методы для решения систем с **прямоугольными** матрицами
- Расширение функционала и дополнительные оптимизации

Спасибо за внимание