Sparse Matrices in Numerical Mathematics

Miroslav Tůma

Faculty of Mathematics and Physics

Charles University

mirektuma@karlin.mff.cuni.cz

Praha, September 18, 2022

Outline

Basic Terminology

- Interest in solving linear systems of equations

$$Ax = b, \tag{1}$$

- $A \in \mathbb{R}^{n\times n}, 1 \leq i \leq n,$ is nonsingular
- A is sparse
- $b \in \mathbb{R}^n$ (sparse or dense), $x \in \mathbb{R}^n$ is the solutions
- Used throughout:

$$A = (a_{ij}), \;\; 1 \leq i, j \leq n.$$

- Matlab-like notation: nonzero (set a priori), $A_{:,j}$, $A_{i,:}$, $A_{i:j,k:l}$, A_{*j}, A_{i*}.

Basic Terminology

- $$\binom{X}{2} = \{Y \subseteq X \mid |Y| = 2\}$$
- Vectors denoted by small letters as v, u, x, matrices by capital letters as $A, B, \ldots$
- $A \in \mathbb{R}^{n \times n}, 1 \leq i \leq n$, is nonsingular, the right-hand side vector $b \in \mathbb{R}^n$ is given and $x \in \mathbb{R}^n$ is the required solution vector. n is the order (or dimension) of A.

Basic Terminology

- A is diagonal if for all $i \neq j$, $a_{ij} = 0$;
- A is lower triangular if for all $i < j$, $a_{ij} = 0$
- A is upper triangular if for all $i > j$, $a_{ij} = 0$.
- A is unit triangular if it is triangular and all the entries on the diagonal are equal to one.
- A is structurally symmetric if for all i and j for which a_{ij} is nonzero the entry a_{ji} is also nonzero.
- A is symmetric if

  $$a_{ij} = a_{ji}, \text{ for all } i, j.$$

  Otherwise, A is nonsymmetric.
- The symmetry index $s(A)$ of A: the number of nonzeros a_{ij}, $i \neq j$, for which a_{ji} is also nonzero divided by the total number of off-diagonal nonzeros. Small values of $s(A)$: A is far from symmetric.

Basic Terminology: special matrix classes

- A is symmetric positive definite (SPD) if it is symmetric and satisfies
$$v^T A v > 0 \text{ for all nonzero } v \in \mathbb{R}^n.$$
- Otherwise, A is symmetric indefinite.
- Symmetric and (typically) indefinite saddle point matrices have the form
$$A = \begin{pmatrix} G & R^T \\ R & B \end{pmatrix},$$
where $G \in \mathbb{R}^{n_1 \times n_1}$, $B \in \mathbb{R}^{n_2 \times n_2}$, $R \in \mathbb{R}^{n_2 \times n_1}$ with $n_1 + n_2 = n$, G is a SPD matrix and B is a symmetric positive semidefinite matrix (that is $v^T B v \geq 0$ for all nonzero $v \in \mathbb{R}^{n_2}$). In some applications, $B = 0$.

Basic Terminology: blocks

- Symmetric block structure of A:
- $$A = (A_{ib,\,jb}), \quad A_{ib,\,jb} \in \mathbb{R}^{n_i \times n_j}, \quad 1 \le ib, jb \le nb, \tag{2}$$

  that is,

  $$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,nb} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,nb} \\ \vdots & \vdots & \ddots & \vdots \\ A_{nb,1} & A_{nb,2} & \cdots & A_{nb,nb} \end{pmatrix}.$$
- Assuming the square blocks $A_{jb,\,jb}$ on the diagonal are nonsingular.
- Special cases: A is block diagonal if $A_{ib,\,jb} = 0$ for all $ib \neq jb$, A is block lower triangular if $A_{1:jb-1,\,jb} = 0$, $2 \le jb \le nb$, block upper triangular if $A_{jb+1:nb,\,jb} = 0$, $1 \le jb \le nb - 1$.

Basic Terminology: blocks and reducibility

Definition

Matrix $A \in \mathbf{R}^{n\times n}$ is **reducible**, if there is a permutation matrix P such that

$$P^T A P = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \tag{3}$$

where A_{11} and A_{22} are square nontrivial matrices (of dimension at least 1). If A is not reducible, it is called **irreducible**. Matrices of dimension 1 are always considered to be irreducible.

Remark

Symmetric reducible matrix is block diagonal.

Basic Terminology: sparsity

- A is a sparse matrix if many of its entries are zero.
- Attempts to formalize matrix sparsity more precisely: matrix of order n may be said to be sparse if it has $O(n)$ nonzeros.
- Our choice: A is sparse if it is advantageous to exploit its zero entries. Otherwise, A is dense.
- The sparsity pattern $\mathcal{S}\{A\}$ of A is the set of nonzeros, that is,

$$\mathcal{S}\{A\} = \{(i,j) \,|\, a_{ij} \neq 0,\ 1 \leq i,j \leq n\}.$$

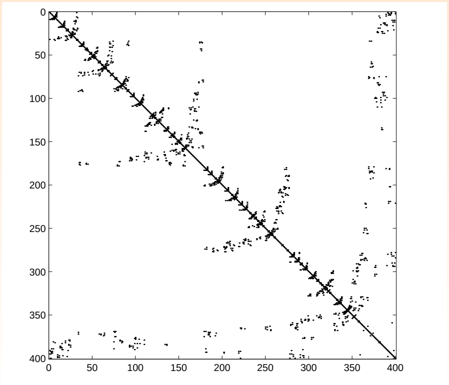
- $\mathcal{S}\{A\}$ is symmetric if for all i and j, $a_{ij} \neq 0$ if and only if $a_{ji} \neq 0$ (the values of the two entries need not be the same). If $\mathcal{S}\{A\}$ is symmetric then A is structurally symmetric.
- The number of nonzeros in A: denoted by $nz(A)$ (or $|\mathcal{S}\{A\}|$). A is structurally (or symbolically) singular if there are no values of the $nz(A)$ entries of A whose row and column indices belong to $\mathcal{S}\{A\}$ for which A is nonsingular.

Basic Terminology: sparsity

Sparsity: taking into account the structure of matrix nonzeros

Definition

Matrix $A \in \mathbb{R}^{m \times n}$ is said to be sparse if it has $O(\min\{m, n\})$ entries.



Basic Terminology: sparsity

Definition

Matrix $A \in \mathbb{R}^{m \times n}$ is said to be sparse if it has row counts bounded by $r_{max} << n$ or column counts bounded by $c_{max} << m$.

Definition

Matrix $A \in \mathbb{R}^{m \times n}$ is said to be sparse if its number of nonzero entries is $O(n^{1+\gamma})$ for some $\gamma < 1$.

Definition

(pragmatic, application-based definition: J.H. Wilkinson) Matrix $A \in \mathbb{R}^{m \times n}$ is said to be sparse if we can exploit the fact that a part of its entries is equal to zero.

Basic Terminology: sparsity

An example showing importance of small exponent γ for $n = 10^4$

γ	$n^{1+\gamma}$
0.1	25119
0.2	63096
0.3	158489
0.4	398107
0.5	1000000

Basic Terminology: sparsity

Rough comparison of dense and sparse (dimension, storage, time for decomposition)

Dense matrix

dim	space	dec time (s)
3000	4.5M	5.72
4000	8M	14.1
5000	12.5M	27.5
6000	18M	47.8

Sparse matrix

dim	space	dec time (s)
10000	40k	0.02
90000	0.36M	0.5
1M	4M	16.6
2M	8M	49.8

Basic Terminology: sparsity

- Sparse vectors The sparsity pattern of $v \in \mathbb{R}^n$ is given by

$$\mathcal{S}\{v\} = \{i \,|\, v_i \neq 0\},$$

  and $|\mathcal{S}\{v\}|$ is the length of v.
- Matrix A is factorizable (or strongly regular) if its principal leading minors (the determinants of its principal leading submatrices) are nonzero, that is, if its LU factorization without row/column interchanges does not break down.
- SPD matrices are factorizable.
- For more general A, in exact arithmetic the following standard result holds.

Theorem

If A is nonsingular then the rows of A can be permuted so that the permuted matrix is factorizable. The row permutations do not need to be known in advance. They can be constructed on-the-fly as the factorization proceeds.

Basic Terminology: factorizations

- For symmetric positive definite A, the Cholesky factorization $A = LL^T$, where L is a lower triangular matrix with positive diagonal entries.
  - Rewritten as $A = \widehat{L}D\widehat{L}^T$, where $\widehat{L}$ is a unit lower triangular matrix and D is a diagonal matrix with positive diagonal entries: square root-free Cholesky factorization.
- For nonsymmetric A, the LU factorization $A = LU$, where L is a unit lower triangular matrix and U is an upper triangular matrix. Gaussian elimination is one process to put a matrix into LU form.
  - Rewritten as $A = LD\widehat{U}$, where $\widehat{U}$ is a unit upper triangular matrix and D is a diagonal matrix. This is called the LDU factorization.

Basic Terminology: direct solver phases

- First look: The matrix A is factorized and then, given the right-hand side b, the factors used to compute the solution x.
- Second look:
- Most approaches further split the factorization into a symbolic phase (also called the analyse phase) and a numerical factorization phase that computes the factors.
- The symbolic phase: typically uses only $\mathcal{S}\{A\}$ to compute the nonzero structure of the factors of A without computing the numerical values of the nonzeros.
- The solve phase uses the factors to solve for a single b or for a block of multiple right-hand sides or for a sequence of right-hand sides one-by-one.
- Historically, the symbolic phase was much faster than the factorization phase. But parallelising the factorization $\rightarrow$ timings are much more closer.
- Series of problems in which the numerical values of the entries of A change but $\mathcal{S}\{A\}$ does not: symbolic phase just once.

Basic Terminology: computational environment

- Basic sequential model: the von Neumann architecture:union of a central processing unit (CPU) and the memory, interconnected via input/output (I/O) mechanisms.

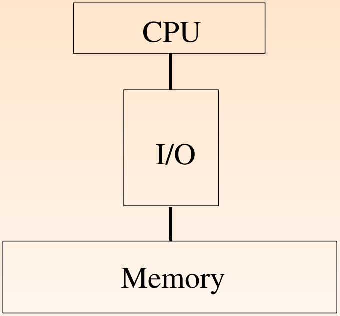


Figure: *A simple uniprocessor von Neumann computer model.*

Basic Terminology: computational environment

- Nowadays: CPU $\rightarrow$ a mixture of powerful processors, co processors, cores, GPUs, and so on.
- Furthermore, performing arithmetic operations on the processing units is much faster than communication-based operations.
- Moreover, improvements in the speed of the processing units outpaces those in the memory-based hardware. Moore's Law is an example of an experimentally derived observation of this kind.

Basic Terminology: computational environment

- Important milestones in processor development have been multiple functional units that compute identical numerical operations in parallel and data pipelining (also called vectorization) that enables the efficient processing of vectors and matrices.
- Vectorization often supported by additional tools like instruction pipelining, registers and by memory architectures with multiple layers, including small but fast memories called caches.
- Superscalar processors that enable the overlapping of identical (or different) arithmetic operations during run-time have been a standard component of computers since the 1990s.
- The ever-increasing heterogeneity of processing units and their hardware environment inside computers: expressing the code via units of scheduling and execution called threads.

Basic Terminology: computational environment

- Computer-based limitations:
  - Compute throughput, that is, the number of arithmetic operations that can be performed per cycle.
  - Memory throughput, that is, the number of operands than can be fetched from memory/cache and/or registers each cycle
  - Latency, which is the time from initiating a compute instruction or memory request before it is completed and the result available for use in the next computation.
- Distinguishing: algorithms compute-bound, memory-bound or latency-bound.
- More ways to hide latency (blocks, prefetch, threads)
- Measuring computational intensity: the ratio of the number of operations to the number of operands read from memory.
- Most chips designed such that dense matrix-matrix multiply, which typically performs n^3 operations on n^2 data (with ratio k for a blocked algorithm with block size k), can run at full compute throughput, whilst matrix-vector multiply performs n^2 operations on n^2 data (ratio 1) and is limited by the memory throughput.

Basic Terminology: computational environment

- The development of basic linear algebra subroutines (BLAS) for performing common linear algebra operations on dense matrices partially motivated by obtaining a high ratio. efficiently.
- Other important motivations behind using the BLAS (standardization, portability).
- Machine-specific optimized BLAS libraries available for a wide variety of computer architectures.

procedure	*comm*	*ops*	ratio
BLAS 1: AXPY: $y = y + \alpha x$	$3n + 1$	$2n$	$2/3$
BLAS 2: GEMV: $y = Ax$	$n^2 + 2n$	$n(2n - 1)$	2
BLAS 3: GEMM: $C = AB$	$3n^2$	$n^2(2n - 1)$	$n/2$

- Consequently, exploiting Level 3 BLAS when designing and implementing matrix algorithms (for both sparse and dense matrices) can improve performance compared to using Level 1 and Level 2 BLAS.

Basic Terminology: finite precision arithmetic

- The IEEE standard (1985) expresses real numbers as $a = \pm d_1 . d_2 \ldots d_t \times 2^k$, where k is an integer and $d_i \in \{0, 1\}, 1 \le i \le t$, with $d_1 = 1$ unless $d_2 = d_3 = \ldots = d_t = 0$.
- $t = 24$ (single precision), $t = 53$ (double precision), exponent k satisfies $-126 \le k \le 127$ (single precision) and $-1022 \le k \le 1023$ (double precision).
- Floating-point (FP) operations:

$$fl(a\ op\ b) = (a\ op\ b)(1 + \delta), \qquad |\delta| \le \epsilon,$$

  (op is a mathematical operation (such as $=, +, -, \times, /, \sqrt{}$) and $(a\ op\ b)$ is the exact result), ϵ is the machine epsilon.
- $2 \times \epsilon$ is the smallest FP number which when added to the FP number 1.0 gives a result different from 1.0.
- ϵ is $2^{-24} \approx 10^{-7}$ (single precision), $\epsilon = 2^{-53} \approx 10^{-16}$ (double precision).
- rounding errors, truncation errors.
- catastrophic errors $\rightarrow$ numerical instability

Basic Terminology: bit compatibility

- Bit compatibility is essential for some users because of regulatory requirements (such as within the nuclear or financial industries) or to build trust in their software from non technical users.
- The critical issue is the way in which N numbers (or, more generally, matrices) are assembled:

  $$sum = \sum_{j=1}^{N} S_j,$$

  where the S_j are computed using one or more processors. The assembly is commutative but, because of the potential rounding of the intermediate results, is not associative so that the result sum depends on the order in which the S_j are assembled.
- A straightforward approach to achieving bit compatibility is to enforce a defined order in such operations.
- This may adversely limit the scope for parallelism.

Basic Terminology: complexity

- The computational complexity of a numerical algorithm is typically based on estimating asymptotically the number of integer or floating-point operations or the memory usage.

Definition

A real function $f(k)$ of a nonnegative real k satisfies $f = O(g)$ if there exist positive constants c_u and k_0 such that

$$f(k) \leq c_u g(k) \text{ for all } k \geq k_0. \tag{4}$$

We say that $f = \Theta(g)$ if, additionally, there exists a positive constant c_l such that

$$0 \leq c_l \, g(k) \leq f(k) \leq c_u \, g(k) \text{ for all } k \geq k_0.$$

Basic Terminology: complexity

- While $O(g)$ bounds f asymptotically from above, $\Theta(g)$ represents an asymptotically tight bound.
- As a simple illustration, consider the quadratic function

  $$f(k) = \alpha * k^2 + \beta * k - \gamma.$$

  Provided $\alpha \neq 0$, $f(k) = \Theta(k^2)$ and the coefficient of the highest asymptotic term is α. Computational complexity can estimate quantities related to the worst-case behaviour of an algorithm (worst-case complexity), or it can express average behaviour (average-case complexity).
- Unit costs, Sparse matrix algorithms that are $\Theta(n^3)$ are considered to be computationally expensive.

Basic Terminology: complexity

Complexity here and in CS

- Because of the development in computations, MFLOPs may be misleading
- Still terminology $O(.)$ (bounding from above) or $\Theta(.)$ (bounding from both sides) sometimes relevant - consists in replacing the bound (bounds) by $constant \times simpler\ function$ (etalon).
- Simpler functions are, e.g., $n^2, n^3, \log n, \ldots$
- Distinguish **worst case** and **average case** analysis
- Inverse Ackermann function will be introduced in exercises
- In CS: polynomial complexity versus superpolynomial complexity. Our case: even n^3 may be too much.
- Decision problems, polynomial reduction, class $\mathcal{NP}$, etc.

Basic Terminology: complexity and sparsity

- Absolutely crucial for direct methods: complexity for generally dense matrices, sequential case: $O(n^3)$ factorization, $O(n^2)$ substitutions
- Useful for iterative methods as well: repeated multiplications and solve steps. But, expecting rather sparse matrices and (typically) dense vectors.
- Complexity in the sparse case depends on the decomposition model and computer architecture (implementation, completeness/incompleteness)