UNIVERSITY OF MINNESOTA TWIN CITIES

C S C I 8314 Spring 2023

SPARSE MATRIX COMPUTATIONS

Class time	:	*MW 9:45 – 11:00am*
Room	:	*Ackerman Hall 211*
Instructor	:	*Yousef Saad*

January 17, 2023

About this class: Objectives

Set 1 An introduction to sparse matrices and sparse matrix computations.

- Sparse matrices;
- Sparse matrix direct methods ;
- Graph theory viewpoint; graph theory methods;

Set 2 Iterative methods and eigenvalue problems

- Iterative methods for linear systems
- Algorithms for sparse eigenvalue problems and the SVD
- Possibly: nonlinear equations

Set 3 Applications of sparse matrix techniques

- Applications of graphs; Graph Laplaceans; Networks ...;
- Standard Applications (PDEs, ..)
- Applications in machine learning
- Data-related applications
- Other instances of sparse matrix techniques

➤ Please fill out (now if you can)

This survey

Short link url:

https://forms.gle/i5MCBg3X289JMAHd8

Who is in this class today?

➤ Out of 20 [as of Tuesday] - registered

- 5 in Computer Science
- 5 in Aerospace Engineering
- 2 Electrical Engineering
- 2 Civil Engineering
- 2 Chemical Engineering/ Materials Science
- 2 Mathematics
- 1 Statistics
- 1 Industrial & Systems Eng.

Logistics:

➤ Lecture notes and minimal information will be located here:
8314 at cselabs class web-sites

URL:

https://www-users.cselabs.umn.edu/classes/Spring-2023/csci8314

[also follow: 'teaching' at www.cs.umn.edu: /saad]

➤ There you will find :

- Lecture notes, Schedule of assignments/ tests, class info

➤ Canvas will contain the rest of the information: assignments, grades, etc.

About lecture notes:

➤ Lecture notes (like this first set) will be posted on the class web-site – usually before the lecture.

➤ Review them to get some understanding if possible before class.

➤ Read the relevant section (s) in the texts or provided references

➤ Lecture note sets are grouped by topics rather than by lecture.

➤ In the notes the symbol ✍1 indicates suggested easy exercises or questions – often [not always] done in class.

➤ Also: occasional practice exercises posted

Matlab, Python-Numpy, etc..

➤ Important to use either Matlab (mostly) or Python to quickly illustrate and test algorithms.

➤ Scripts in either matlab or python will be posted in the 'matlab' section of the class web-site.

➤ Also: matlab or python demos seen in class will be posted

Roadmap – [subject to itinerary change!]

Part 1

1. Sparse matrices;
2. Graph representations;
3. Sparse direct methods for linear systems;

Part 2

4. Iterative methods for linear systems ;
5. Projection methods and Krylov subspace methods;
6. Eigenvalue problems;

Part 3

7. Back to Graphs; Paths in graphs; Markov Chains;
8. Graph centrality;
9. Graph Laplaceans and applications; Clustering;
10. Graph embeddings.

CSCI 8314: SPARSE MATRIX COMPUTATIONS

GENERAL INTRODUCTION

- *General introduction - a little history*
- *Motivation*
- *Resources*
- *What will this course cover*
- *Examples of problems leading to sparse matrix computations*

Historical Perspective: Focus of numerical linear algebra

➤ Linear algebra took many direction changes in the past

1940s–1950s: Major issue: flutter problem in aerospace engineering → eigenvalue problem [cf. Olga Taussky Todd] → LR, QR, .. → 'EISPACK'

1960s: Problems related to the power grid promoted what we would call today general sparse matrix techniques

1970s– Automotive, Aerospace, ..: Computational Fluid Dynamics (CFD)

Late 1980s: Thrust on parallel matrix computations.

Late 1990s: Spur of interest in "financial computing"

Current: Machine Learning

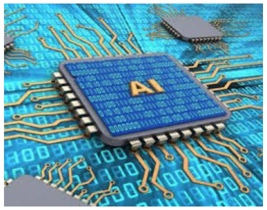
Solution of PDEs (e.g., Fluid Dynamics) and problems in mechanical eng. (e.g. structures) major force behind numerical linear algebra algorithms in the past few decades.

➤ Strong new forces are now reshaping the field today: Applications related to the use of "data"

➤ Machine learning is appearing in unexpected places:

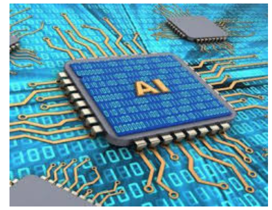
- design of materials
- machine learning in geophysics
- self-driving cars, ..
- ....

Big impact on the economy



- New economy driven by Google, Facebook, Netflix, Amazon, Twitter, Ali-Baba, Tencent, ..., and even the big department stores (Walmart, ...)
- Huge impact on **Jobs**

Big impact on the economy



- New economy driven by Google, Facebook, Netflix, Amazon, Twitter, Ali-Baba, Tencent, ..., and even the big department stores (Walmart, ...)
- Huge impact on **Jobs**

- Old leaders - e.g., Mining; Car companies; Aerospace; Manufacturing; offer little growth – Some instances of renewal driven by new technologies [e.g. Tesla]

- Look at what you are doing under new lenses: DATA

Sparse matrices: a brief history

Sparse matrices have been identified as important early on – origins of terminology is quite old. Gauss defined the first method for such systems in 1823. Varga used explicitly the term 'sparse' in his 1962 book on iterative methods.

https://www-users.cs.umn.edu/~saad/PDF/icerm2018.pdf

- Special techniques used for sparse problems coming from Partial Differential Equations
- One has to wait until to the 1960s to see the birth of the general methodology available today
- Graphs introduced as tools for sparse Gaussian elimination in 1961 [Seymour Parter]

- Early work on reordering for banded systems, envelope methods
- Various reordering techniques for general sparse matrices introduced.
- Minimal degree ordering [Markowitz - 1957] ...
- ... later used in Harwell MA28 code [Duff] - released in 1977.
- Tinney-Walker Minimal degree ordering for power systems [1967]
- Nested Dissection [A. George, 1973]
- SPARSPAK [commercial code, Univ. Waterloo]
- Elimination trees, symbolic factorization, ...

History: development of iterative methods

➤ 1950s up to 1970s : focus on "relaxation" methods

➤ Development of 'modern' iterative methods took off in the mid-70s. but...

➤ ... The main ingredients were in place earlier [late 40s, early 50s: Lanczos; Arnoldi ; Hestenes (a local!) and Stiefel; ....]

➤ The next big advance was the push of 'preconditioning': in effect a way of combining iterative and (approximate) direct methods – [Meijerink and Van der Vorst, 1977]

https://www-users.cs.umn.edu/~saad/PDF/NIST75th.pdf

History: eigenvalue problems

➤ Another parallel branch was followed in sparse techniques for large eigenvalue problems.

➤ A big problem in 1950s and 1960s : flutter of airplane wings.. This leads to a large (sparse) eigenvalue problem

➤ Overlap between methods for linear systems and eigenvalue problems [Lanczos, Arnoldi]

Resources

➤ SuiteSparse site (Formerly : Florida collection)

https://sparse.tamu.edu/

➤ SPARSKIT, etc. [SPARSKIT = old written in Fortran. + more recent 'solvers']

http://www.cs.umn.edu/~saad/software

Resources – continued

Books: on sparse direct methods.

➤ Book by Tim Davis [SIAM, 2006] see syllabus for info

➤ An old reference [Still a great book]: Alan George and Joseph W-H Liu, Computer Solution of Large Sparse Positive Definite Systems, Prentice-Hall, 1981.

➤ Of interest mostly for references:

• I. S. Duff and A. M. Erisman and J. K. Reid, Direct Methods for Sparse Matrices, Clarendon press, Oxford, 1986.

• Some coverage in Golub and van Loan [John Hopinks, 4th Ed., Chap. 10 to end]

BACKGROUND: PROBLEMS LEADING TO SPARSE MATRICES

Background: Examples leading to sparse matrices

➤ The classical: CFD, electrical networks,

➤ ... and the modern:

- Graph algorithms and tools (Sparse graphs, graph coarsening, graphs and sparse methods). ..
- Dimension reduction methods; Graph embeddings;
- Specific machine learning algorithms; unsupervised/ supervised learning;
- Deep learning;
- Network analysisl
- ...

Example: Fluid flow

Physical Model

↓

Nonlinear PDEs

↓

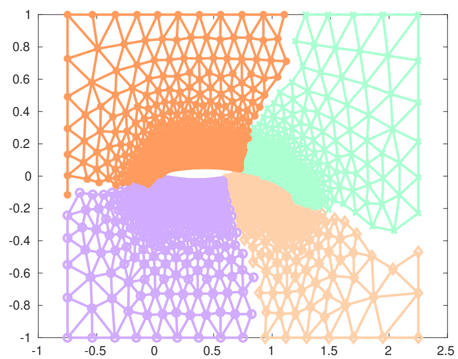
Discretization

↓

Linearization (Newton)

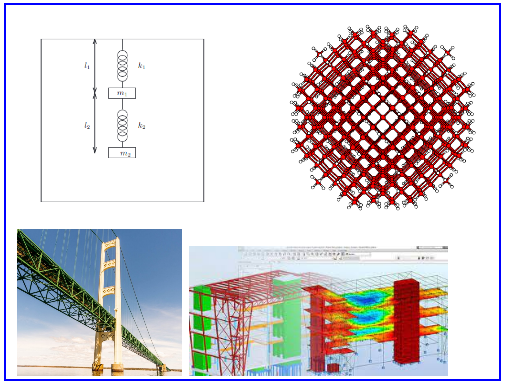
↓

Sparse Linear Systems $Ax = b$



Example: Eigenvalue Problems

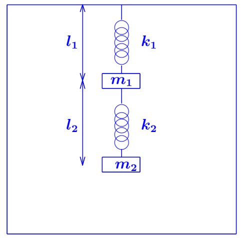
➤ Many applications require the computation of a few eigenvalues + associated eigenvectors of a matrix $\boldsymbol{A}$



- Structural Engineering – (Goal: frequency response)
- Electronic structure calculations [Schrödinger equation..] – Quantum chemistry
- Stability analysis [e.g., electrical networks, mechanical system,..]
- ...

Example: Vibrations

- Vibrations in mechanical systems. See: www.cs.umn.edu/~saad/eig_book_2ndEd.pdf

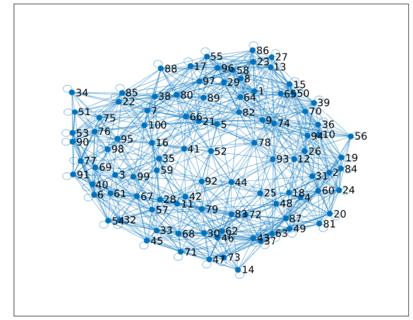


Problem: Determine the vibration modes of the mechanical system [to avoid resonance]. See details in Chapter 10 (sec. 10.2) of above reference.

- Problem type: Eigenvalue Problem

Example: Google Rank (pagerank)

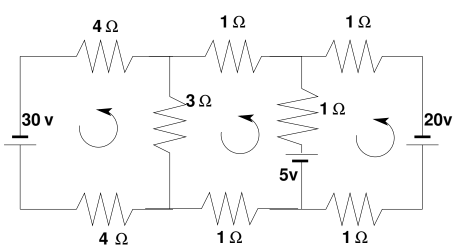
If one were to do a random walk from web page to web page, following each link on a given web page at random with equal likelihood, which are the pages to be encountered this way most often?



- Problem type: (homogeneous) Linear system. Eigenvector problem.

Example: Power networks

- Electrical circuits .. [Kirchhiff's voltage Law]

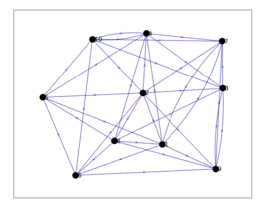


Problem: Determine the loop currents in a an electrical circuit - using Kirchhoff's Law ($V = RI$)

- Problem: Sparse Linear Systems [at the origin Sparse Direct Methods]

Example: Economics/ Marketing/ Social Networks

- Given: an influence graph G: g_{ij} = strength of influence of j over i
- Goal: charge member i price p_i in order to maximize profit
- Utility for member i: [x_i = consumption of i]



$$u_i = ax_i - bx_i^2 + \sum_{j \neq i} g_{ij}x_j - p_i x_i$$

- 1: 'Monopolist' fixes prices; 2: agent i fixes consumption x_i

Result: Optimal pricing proportional to Bonacich centrality: $(I - \alpha G)^{-1} \mathbb{1}$ where $\alpha = \frac{1}{2b}$ [*Candogan et al., 2012* + many refs.]

➤ 'centrality' defines a measure of importance of a node (or an edge) in a graph

➤ Many other ideas of centrality in graphs [degree centrality, betweenness centrality, closeness centrality, ...]

➤ Important application: Social Network Analysis

Example: Method of least-squares

➤ First use of least squares by Gauss, in early 1800's:

A planet follows an elliptical orbit according to $ay^2 + bxy + cx + dy + e = x^2$ in cartesian coordinates. Given a set of noisy observations of (x, y) positions, compute $a, b, c, d, e,$ and use to predict future positions of the planet. This least squares problem is nearly rank-deficient and hence very sensitive to perturbations in the observations.

➤ Problem type: Least-Squares system

Read Wikipedia's article on planet ceres:

http://en.wikipedia.org/wiki/Ceres_(dwarf_planet)

Example: Dynamical systems and epidemiology

A set of variables that fill a vector y are governed by the equation

$$\frac{dy}{dt} = Ay$$

Determine $y(t)$ for $t > 0$, given $y(0)$ [called 'orbit' of y]

➤ Problem type: (Linear) system of ordinary differential equations.

Solution: $$y(t) = e^{tA}y(0)$$

➤ Involves exponential of A [think Taylor series], i.e., a matrix function

➤ This is the simplest form of dynamical systems (linear).

➤ Consider the slightly more complex system:

$$\frac{dy}{dt} = A(y)y$$

➤ Nonlinear. Requires 'integration scheme'.

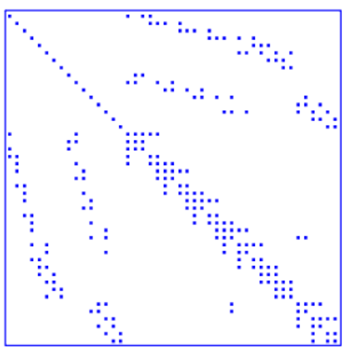
General Problems in Numerical Linear Algebra (dense & sparse)

- Linear systems: $Ax = b$. Often: A is large and sparse
- Least-squares problems $\min \|b - Ax\|_2$
- Eigenvalue problem $Ax = \lambda x$. Several variations -
- SVD .. and
- ... Low-rank approximation
- Tensors and low-rank tensor approximation
- Matrix equations: Sylvester, Lyapunov, Riccati, ..
- Nonlinear equations – acceleration methods
- Matrix functions and applications
- Many many more ...

SPARSE MATRICES

- *See the "links" page on the class web-site*
- *See also the various sparse matrix sites.*
- *Introduction to sparse matrices*
- *Sparse matrices in matlab –*
- *See Chap. 3 of text*

What are sparse matrices?



Pattern of a small sparse matrix

➤ Vague definition: matrix with few nonzero entries

➤ For all practical purposes: an $m \times n$ matrix is sparse if it has $O(\min(m, n))$ nonzero entries.

➤ This means roughly a constant number of nonzero entries per row and column -

➤ This definition excludes a large class of matrices that have $O(\log(n))$ nonzero entries per row.

➤ Other definitions use a slow growth of nonzero entries with respect to n or m.

"..matrices that allow special techniques to take advantage of the large number of zero elements." (J. Wilkinson)

A few applications which lead to sparse matrices:

Structural Engineering, Computational Fluid Dynamics, Reservoir simulation, Electrical Networks, optimization, Google Page rank, information retrieval (LSI), circuit similation, device simulation, ...

Goal of Sparse Matrix Techniques

➤ To perform standard matrix computations economically i.e., without storing the zeros of the matrix.

Example: To add two square dense matrices of size n requires $O(n^2)$ operations. To add two sparse matrices A and B requires $O(nnz(A) + nnz(B))$ where $nnz(X)$ = number of nonzero elements of a matrix X.

➤ For typical Finite Element /Finite difference matrices, number of nonzero elements is $O(n)$.

Remark: A^{-1} is usually dense, but L and U in the LU factorization may be reasonably sparse (if a good technique is used)

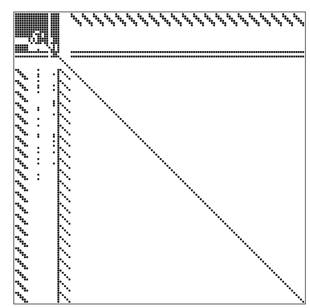
✍2 Look up Cayley-Hamilton's theorem if you do not know about it.

✍3 Show that the inverse of a matrix (when it exists) can be expressed as a polynomial of A, where the polynomial is of degree $\leq n-1$.

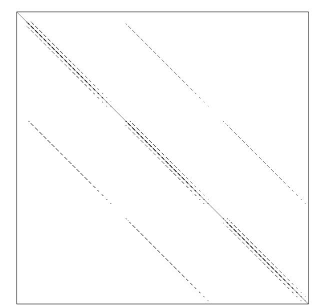
✍4 When is the degree $< n-1$? [Hint: look-up minimal polynomial of a matrix]

✍5 What is the pattern of the inverse of a tridiagonal matrix? a bidiagonal matrix?

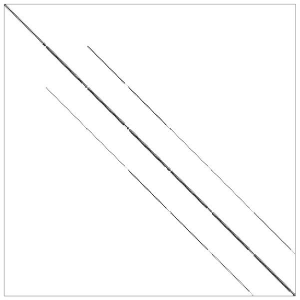
Nonzero patterns of a few sparse matrices



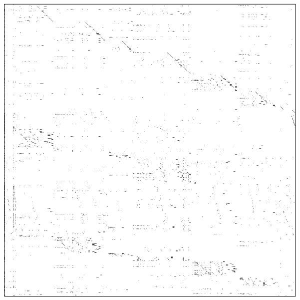
ARC130: Unsymmetric matrix from laser problem. a.r.curtis, oct 1974



SHERMAN5: fully implicit black oil simulator 16 by 23 by 3 grid, 3 unk



PORES3: Unsymmetric MATRIX FROM PORES



BP_1000: UNSYMMETRIC BASIS FROM LP PROBLEM BP

Types of sparse matrices

➤ Two types of matrices: **structured** (e.g. Sherman5) and **unstructured** (e.g. BP_1000)

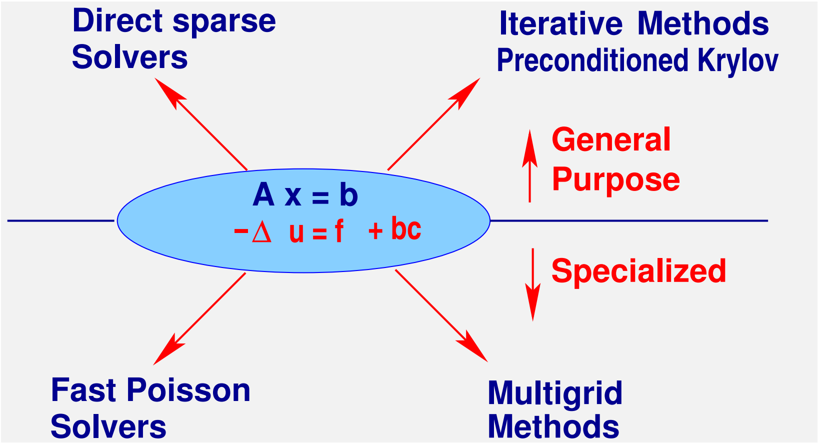
➤ The matrices PORES3 and SHERMAN5 are from Oil Reservoir Simulation. Often: 3 unknowns per mesh point (Oil , Water saturations, Pressure). Structured matrices.

➤ 40 years ago reservoir simulators used rectangular grids.

➤ Modern simulators: Finer, more complex physics ➤ harder and larger systems. Also: unstructured matrices

➤ A naive but representative challenge problem: $100 \times 100 \times 100$ grid + about 10 unknowns per grid point ➤ $N \approx 10^7$, and $nnz \approx 7 \times 10^8$.

Solving sparse linear systems: existing methods



Two types of methods for general systems:

➤ Direct methods : based on sparse Gaussian eimination, sparse Cholesky,..

➤ Iterative methods: compute a sequence of iterates which converge to the solution - preconditioned Krylov methods..

Remark: These two classes of methods have always been in competition.

➤ 40 years ago solving a system with $n = 10,000$ was a challenge

➤ Now you can solve this in a fraction of a second on a laptop.

➤ Sparse direct methods made huge gains in efficiency. As a result they are very competitive for 2-D problems.

➤ 3-D problems lead to more challenging systems [inherent to the underlying graph]

Difficulty:

- No robust 'black-box' iterative solvers.
- At issue: Robustness in conflict with efficiency.

➤ Iterative methods are starting to use some of the tools of direct solvers to gain 'robustness'

Consensus:

1. Direct solvers are often preferred for two-dimensional problems (robust and not too expensive).

2. Direct methods loose ground to iterative techniques for three-dimensional problems, and problems with a large degree of freedom per grid point,

Sparse matrices in matlab

➤ Matlab supports sparse matrices to some extent.

➤ Can define sparse objects by conversion

```
A = sparse(X) ; X = full(A)
```

➤ Can show pattern

```
spy(X)
```

➤ Define the analogues of `ones, eye`:

```
speye(n,m),   spones(pattern)
```


➤ A few reorderings functions provided.. [will be studied in detail later]

```
symrcm, symamd, colamd, colperm
```

➤ Random sparse matrix generator:

`sprand(S)` or `sprand(m,n, density)`

(also textttsprandn(...))

➤ Diagonal extractor-generator utility:

```
spdiags(A) ,spdiags(B,d,m,n)
```

➤ Other important functions:

```
spalloc(..)  ,find(..)
```

Graph Representations of Sparse Matrices

➤ Graph theory is a fundamental tool in sparse matrix techniques.

DEFINITION. A graph G is defined as a pair of sets $G = (V, E)$ with $E \subset V \times V$. So G represents a binary relation. The graph is undirected if the binary relation is reflexive. It is directed otherwise. V is the vertex set and E is the edge set.

Example: Given the numbers 5, 3, 9, 15, 16, show the two graphs representing the relations

R1: Either $x < y$ or y divides x.

R2: x and y are congruent modulo 3. [mod(x,3) = mod(y,3)]

➤ Adjacency Graph $G = (V, E)$ of an $n \times n$ matrix A :

- Vertices $V = \{1, 2, ..., n\}$.
- Edges $E = \{(i, j) | a_{ij} \neq 0\}$.

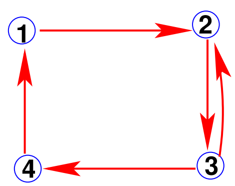
➤ Often self-loops (i, i) are not represented [because they are always there]

➤ Graph is undirected if the matrix has a symmetric structure:

$$a_{ij} \neq 0 \quad \text{iff} \quad a_{ji} \neq 0.$$

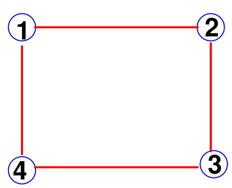
Example: (directed graph)

	⋆		
		⋆	
	⋆		⋆
⋆			



Example: (undirected graph)

	⋆		⋆
⋆		⋆	
	⋆		⋆
⋆		⋆	



✍6 Graph of a tridiagonal matrix? Of a dense matrix?

✍7 Adjacency graph of:

$$A = \begin{bmatrix} \star & \star & & & \star & \\ \star & \star & \star & & & \star \\ & \star & \star & & & \\ & & & \star & \star & \\ \star & & & \star & \star & \star \\ & \star & & & \star & \star \end{bmatrix} ?$$

✍8 Recall what a star graph is. Show a matrix whose graph is a star graph. Consider two situations: Case when center node is labeled first and case when it is labeled last.

- Note: Matlab now has a `graph` function.
- `G = graph(A)` creates adjacency graph from A
- G is a matlab class/
- `G.Nodes` will show the vertices of G
- `G.Edges` will show its edges.
- `plot(G)` will show a representation of the graph

✍9 Do the following:

- Load the matrix 'Bmat.mat' located in the class web-site (see 'matlab' folder)
- Visualize pattern (`spy(B)`) + find: Number of nonzero elements, size, ...
- Generate graph - without self-edges:

```
G = graph(B,'OmitSelfLoops'
```

- Plot the graph –
- $1M question: Any idea on how this plot is generated?