21 Sparse Cholesky

Post-lecture version. There are some changes to the verion given-out in class, particularly in Sections 21.4 and 21.5.

Today we want to focus on how the Cholesky Factorisation Algorithm works in the case where the matrix A is "sparse". There are many possible definitions for what it means for a matrix to the sparse, but there is one that most people agree with: *a matrix is sparse if it there is a benefit to exploiting the fact that it has many nonzero entries* (Wilkinson). If a matrix is not sparse we say it is *full.* Typical examples of sparse matrices include

- Diagonal matrices, such as the identity matrix.
- Tridiagonal matrices, such as the one-dimensional finite difference matrix, T_N.
- The two-dimensional finite difference matrix, T_{NN}.
- The matrix representing, say, a social network.

21.1 Cholesky of a tridiagonal matrix

Recall that the Cholesky factorisation of

$$T_N = \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 \end{pmatrix}.$$

is given by the formula

$$L_{i,i} = \sqrt{\frac{i+1}{i}}, \quad \text{and} \quad L_{i,i-1} = -\sqrt{\frac{i-1}{i}}.$$

So we can see that Cholesky factor is bidiagonal.

In fact, it is always the case the Cholesky factor of an s.p.d. tridiagonal matrix is bidiagonal. There are many ways of showing this, but we'll use the idea the led to Alg. 20.2.

21.2 Banded matrices

A matrix is *banded* if the only non-zero entries are found within a certain distance of the main diagonal. More precisely, the matrix $m \times n$ matrix A is *banded* with *band-width* b if

$$a_{ij} = 0 \quad \text{whenever} \quad |i-j| > b.$$

Usually we care about the case where B is much less than $\min(m, n)$.

The most important example we have met so far is the two-dimensional Laplacian:

$$T_{N,N} = I \otimes T_N + T_N \otimes I.$$

This is an $m \times m$ banded matrix, with $m = N^2$ and a band-width of N. It has only (roughly) $4m$ non-zero entries, so it is quite sparse, considering that a full $m \times m$ matrix would have m^2 non-zeros.

If $A = LL^T$, would we expect that L is similarly sparse? There are three possible scenarios

Most optimistic: L will have non-zeros only where A has non-zeros. (This is not the case, unfortunately).

Least optimistic: L will be full. (This is not the case, fortunately).

Realistic: L will have many more non-zeros than A: roughly $m^{3/2}$.

We'll explain in class why this is likely to be. A precise justification is left as an exercise.

21.3 General concepts

Not every banded matrix has the simple structure of T_{NN}. The more general case is (see [Dem97, §2.7.3] for a detailed discussion):

Definition 21.1 (Banded)**.** A is a banded matrix with lower band-width b_L and upper band-width b_U if

$$a_{ij} = 0 \quad \text{whenever} \quad i > j + b_L \text{ or } i < j - b_U.$$

But we'll just focus on the simple case $b_L = b_U = b$.

Definition 21.2 (Fill-in)**.** If $A = LL^T$. Entries that appear in $L + L^T$ that correspond to zeros in A are called *fill-in.*

21.4 The arrow matrices

Two famous example are used to highlight the hazards associated with fill-in during factorisation:A

$$A_1 = \begin{pmatrix} 5 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 5 \end{pmatrix}$$

If $A_1 = L_1 L_1^T$, and $A_2 = L_2 L_2^T$, then all entries of L_1 on and below the main diagonal will be non-zero, where as L_2 will only have non-zeros in the main diagonal and in the last row. So it is pretty sparse. However, A_1 and A_2 are permutation equivalent.

21.5 Exercises

Exercise 21.1. In class, we outlined a proof that if A is tridiagonal and $A = LL^T$, then L is bidiagonal, using Alg. 20.2. Deduce your own proof.

Exercise 21.2. Show that if $A = T_{N,N}$, and $A = LL^T$, then L has the same band-width as A.