ALN – Sistemas Lineales dispersos

In. Co.
Facultad de Ingeniería
Universidad de la República

Factorización de matrices dispersas

- La clave al trabajar con matrices dispersas es el ahorro en:
  - Espacio de memoria
  - Cantidad de cálculos

- Si factorizamos una matriz dispersa A en LU, nada nos garantiza que las matrices L y U sean dispersas !!!!

Factorización de matrices dispersas

- Si A es dispersa y las matrices L y U no, estaríamos cometiendo varios errores:
  - Utilizando estructuras dispersas para matrices completas.
  - Engañándonos con respecto al número de operaciones realizadas.
- Nos encontramos frente al problema de “llenado” (*fill-in* en la literatura anglosajona)

Factorización de matrices dispersas

- La resolución de sistemas completos consta de 2 etapas:
  - Factorización (escalerización).
  - Sustituciones.

- En el caso de las matrices dispersas cuales son las etapas ??!

Factorización de matrices dispersas

- Además de las tareas de la resolución de sistemas completos, necesitamos:
  - Minimizar, o por lo menos controlar, el fill-in.
  - Prever la estructura de las matrices L y U.
- La mayoría de los trabajos comenzaron resolviendo matrices simétricas y definidas positivas.
- Factorización de Cholesky: $A = LL^T$
  - L es una matriz triangular inferior

Motivación

- Si tenemos la siguiente matriz:

$$\begin{bmatrix} x & 0 & 0 & x \\ 0 & x & 0 & x \\ 0 & 0 & x & x \\ x & x & x & x \end{bmatrix}$$

¿cuanto cuesta factorizarla?

- Y la siguiente?!

$$\begin{bmatrix} x & x & x & x \\ x & x & 0 & 0 \\ x & 0 & x & 0 \\ x & 0 & 0 & x \end{bmatrix}$$

Conocimientos Previos

- Antes de comprender las distintas etapas de la factorización sobre matrices dispersas es necesario introducir algunas herramientas.
- Para estudiar las estrategias de ordenamiento y factorización simbólica se utilizan grafos.

Definiciones

La envolvente de una matriz A se define como:

$$f_i = \min\{j : a_{ij} \neq 0\}$$

$$l_i = \max\{j : a_{ij} \neq 0\}$$

$$Env(A) = \{(i,j) : f_i \leq j \leq l_i, 1 \leq i \leq n\}$$

$$\boldsymbol{A} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \left[\begin{array}{ccccccc} \times & \times & \times & \times & & & \\ \times & \times & 0 & \times & & & \\ & & \times & 0 & \times & \times & \\ & \times & 0 & \times & & & \\ & & & \times & \times & \times & \times \\ & \times & \times & 0 & 0 & \times & \times \\ & & & & & & \times \end{array}\right] \end{array}$$

Sistemas dispersos y grafos

Definiciones básicas de grafos:

- Un grafo $G(V,E)$ es un conjunto de vértices (o nodos) $V = v_1, v_2, ..., v_n$ y un conjunto de aristas $E = (v_i, v_j)$.
- Si las aristas del grafo son pares ordenados el grafo de define como dirigido, en caso contrario se denomina grafo no dirigido.
- El grafo es "completo" si el conjunto de aristas cumple que existe una arista (u, v) perteneciente a E para todo u, v pertenecientes a V.

Sistemas dispersos y grafos

- Un camino en un grafo del nodo v_i al nodo v_k se define como una secuencia de aristas $(v_i, v_j), (v_j,v_l),\ldots, (v_p, v_k)$.
  - Si $v_i = v_k$ el camino se denomina ciclo.
  - Se dice que un grafo es conexo si existe por lo menos un camino entre todo par de nodos.
  - Una cuerda en un camino es una arista que une dos vértices no consecutivos.

- Un grafo **Cordal** es aquel que no posee ciclos sin cuerdas. Un ciclo sin cuerdas es un ciclo de largo mayor o igual a 4 sin cuerdas.

- Un **árbol** es un grafo acíclico.

Sistemas dispersos y grafos

- $H(U, F)$ es un subgrafo de $G(V,E)$ si H es un grafo y el conjunto de vértices U está incluido en V y F está incluido E.

  Clique es un sub-grafo completo.

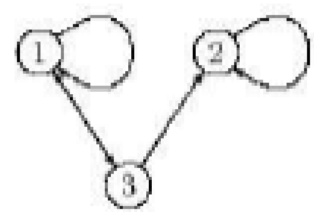
- Se define el conjunto de vértices adyacentes al vértice v, ($AdjG(v)$) a aquellos vértices que poseen aristas incidentes desde el vértice v.

- El grado de un vértice v del grafo G se define como la cardinalidad del conjunto de vértices adyacentes: $gradoG(v) = |AdjG(v)|$.

Sistemas dispersos y grafos

- Una matriz dispersa A cuadrada de dimensión n se puede representar mediante un grafo etiquetado dirigido $G = (V,E)$ donde el conjunto de vértices es $V = \{1..n\}$ y el conjunto de aristas se compone de (i, j) si $a_{ij} \neq 0$.

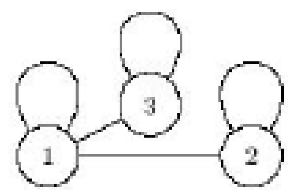
$$\begin{pmatrix} x & & x \\ & x & \\ x & x & \end{pmatrix}$$



Sistemas dispersos y grafos

- Para representar una matriz con estructura simétrica se puede utilizar un grafo etiquetado no dirigido $G = (V,E)$ donde $V = \{1..n\}$ y (i, j) pertenece a E si $a_{ij} \neq 0$ $(a_{ji} \neq 0)$.

$$\begin{pmatrix} x & x & x \\ x & x & \\ x & & x \end{pmatrix}$$



- Los vértices adyacentes a cierto vértice x son los índices de columna/fila de los elementos no nulos en la fila/columna x

Sistemas dispersos y grafos

Árbol de eliminación

- Permite describir las dependencias entre las filas y columnas en la factorización de Cholesky. Los nodos del árbol de eliminación son enteros entre 1 y n, que representan las columnas (filas) de la matriz. El padre de la fila i es el menor j tal que $j > i$ y que el elemento $L(i, j)$ sea distinto de cero.

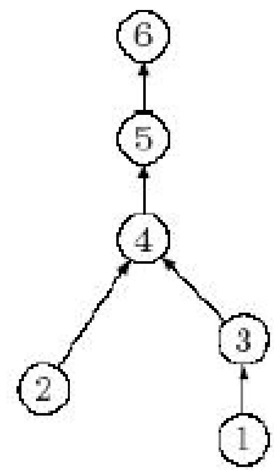
- El patrón de L a partir de A:
  - Los elementos no 0 de A serán elementos no 0 de L.
  - $i<j<k$, $L_{ji} \neq 0$ y $L_{ki} \neq 0$ entonces $L_{kj} \neq 0$

Sistemas dispersos y grafos

Árbol de eliminación

- Los primeros autores en utilizarlo fueron Schreiber, Duff y Reid
- Permite identificar dónde aparecerán aristas de fill-in
  - Si (i,j) es una arista en el grafo final y i < j, entonces j es ancestro de i en el árbol de eliminación.
- Son útiles en la factorización simbólica
  - Por ejemplo, como reflejan las dependencias entre filas/columnas, son útiles para sincronizar el trabajo en factorizaciones paralelas.

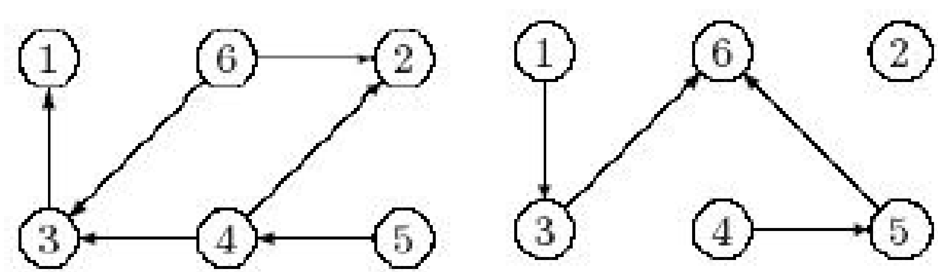
$$\begin{pmatrix} 1 & & & & & \\ & 2 & & & & \\ x & & 3 & & & \\ & x & x & 4 & & \\ & x & & x & 5 & \\ x & x & & & & 6 \end{pmatrix}$$



$$A = \begin{pmatrix} 1 & & x & & & \\ & 2 & & & & \\ x & & 3 & & & x \\ & x & x & 4 & x & \\ & x & & x & 5 & \\ x & x & & & & 6 \end{pmatrix} = \begin{pmatrix} 1 & & & & & \\ & 2 & & & & \\ x & & 3 & & & \\ & x & x & 4 & & \\ & x & & x & 5 & \\ x & x & + & & & 6 \end{pmatrix} \begin{pmatrix} 1 & & x & & & \\ & 2 & & & & \\ & & 3 & & & x \\ & & & 4 & x & + \\ & & & & 5 & + \\ & & & & & 6 \end{pmatrix}$$

DAG

- Gilbert y Liu extendieron el concepto de árbol de eliminación para matrices no simétricas. La propuesta se basa en la utilización de dos grafos dirigidos acíclicos (directed acyclic graph - DAG) que llaman DAGs de eliminación. Los DAGs se obtienen mediante reducciones transitivas de los grafos (dirigidos) asociados a las matrices L y U.



Etapas de la resolución de matrices dispersas

- Ordenamiento
- Factorización simbólica
- Factorización numérica
- Sustitución

Ordenamientos

- Lo vamos a ver en forma separada en la próxima clase.
- En forma resumida, reordenan las filas/columnas de la matriz para facilitar la tarea de factorización
- El número de posibles de ordenaciones es n!

Factorización simbólica

Motivación

- Predecir la estructura de las matrices resultado.
- Esta etapa permite prever y reservar en forma anticipada la memoria de las distintas estructuras de datos a utilizar.
- Acceder en forma “sencilla” a las distintas estructuras.
- No es necesario realizar operaciones en punto flotante!!!!

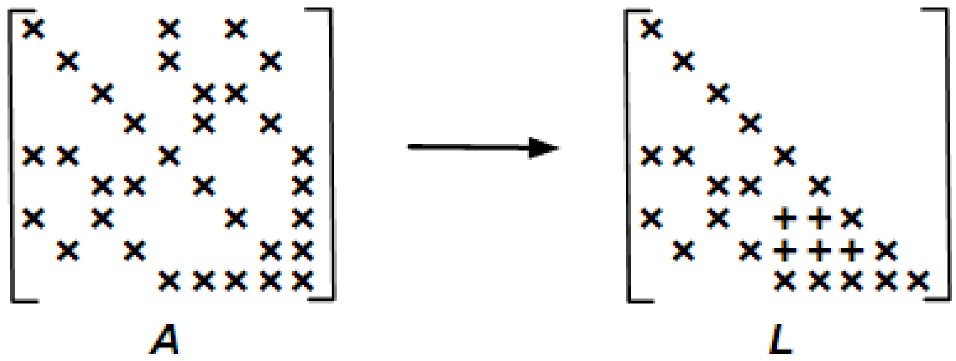
Factorización simbólica

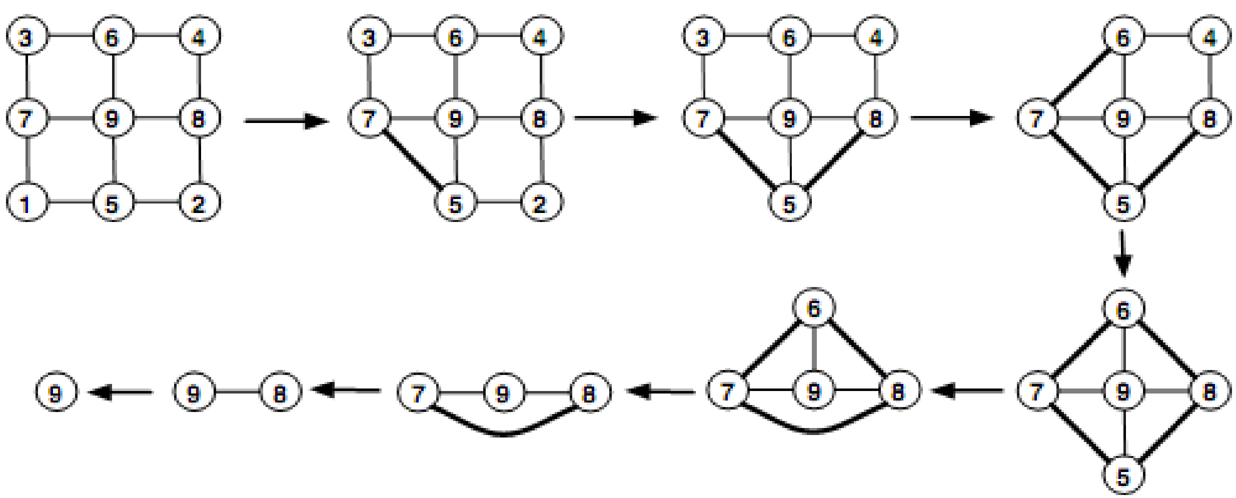
Como se hace:

- En forma simbólica, trabajar con la estructura de la matriz (las posiciones no nulas) y no con los coeficientes. Se asume que toda operación sobre elementos no nulos resultara no nula (se obtiene una estimación de la estructura, cota superior).
- Schreiber fijo la base de muchos de los trabajos sobre factorización simbólica utilizando el árbol de eliminación.

Factorización simbólica

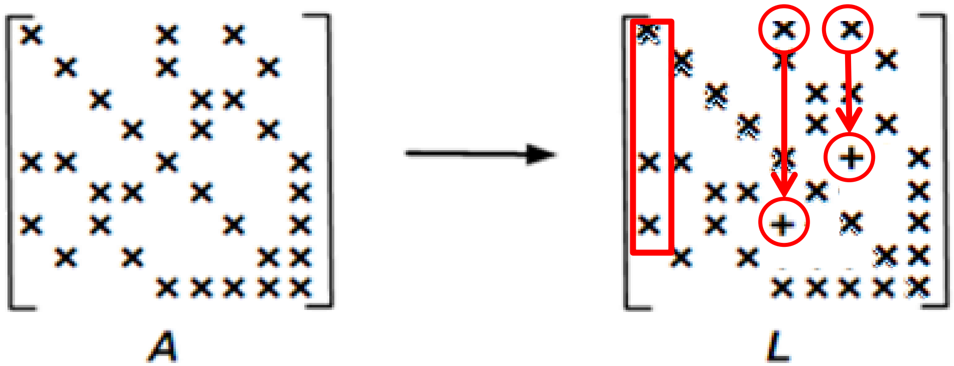
Como se hace (grafos):





Factorización simbólica

Como se hace (grafos):



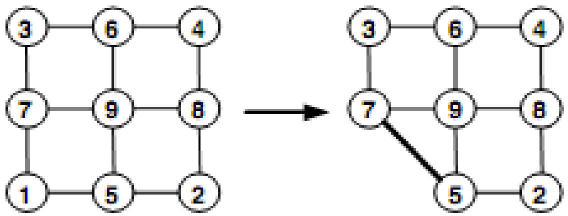
Durante el paso k de la eliminación, se resta la fila k multiplicada por un coeficiente a cada una de las filas que tienen elementos no nulos en la columna k (debajo del pivote).

Los elementos no nulos de la fila k “bajan” a estas filas y pasan a ser elementos no nulos de las mismas.

En el grafo de adyacencia, esto equivale a que los vértices adyacentes a k pasan a ser adyacentes a los vértices correspondientes a las filas mencionadas.

Factorización simbólica

Como se hace (grafos):



Por esta razón, al eliminar un vértice se forma un clique entre sus adyacentes

Factorización simbólica

Como se hace (con conjuntos):

Sea $S_i = \{x_1, \ldots, x_i\}$

$Alc(x_i, S_{i-1})$ es el conjunto de vértices alcanzables desde x_i a través de vértices con numeración menor (que serán eliminados antes que i)

Si x_j pertenece a Alc(x_i , S_{i-1}) hay un camino desde x_i a x_j a través S_{i-1}, por lo que cuando se eliminen todos los vértices de S_{i-1} se generará una arista de fill-in entre x_i y x_j.

Entonces, el grafo de adyacencia que incluye las aristas de fill-in queda caracterizado por:

$$E^F = \{(x_i, x_j) | x_j \in Alc(x_i, S_{i-1})\}$$

Factorización simbólica

- Usando conjuntos alcanzables expresamos el fill-in solamente en función de las aristas del grafo original
- En cada paso i de la eliminación gaussiana, los nodos adjacentes a cualquier nodo *y* en el grafo de eliminación $G^i(X_i, E_i) = Alc(y, S_i)$

Factorización numérica

- Existen varias sub-técnicas dependiendo del acceso de memoria:
  - Super-nodo
  - Frontal
  - Multifrontal
  - Otras

Más info…

- Alan George, Joseph Liu, and Esmond Ng. "Computer Solution of Sparse Linear Systems." *Academic, Orlando* (1994). Disponible en: http://web.engr.illinois.edu/~heath/courses/cs598mh/george_liu.pdf
- A. Pothen, S. Toledo, Elimination structures in scientific computing (2001) https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.490.5817&rep=rep1&type=pdf