Peak Memory Consumption Minimization Problem

Leonardo Lima

Lisbon School of Economics and Management

July 21, 2019

Outline

Formulations

- (**Matrix** bandwidth minimization problem) Let $A = \{a_{ij}\}$ be a sparse symmetric matrix, the bandwidth of A is defined as

$$B(A) = \max\{|i - j| : a_{ij} \neq 0\}$$

  Thus, we want to minimize the bandwidth $B(A)$

- (**Graph** bandwidth minimization problem) Let $G = (V, E)$ be a finite undirected graph and an injective function $f : V \to \{1, 2, \ldots, n\}$ labeling its vertices, where V is the set of vertices and E is the set of edges. The bandwidth of a vertex v and of a graph G are respectively defined as

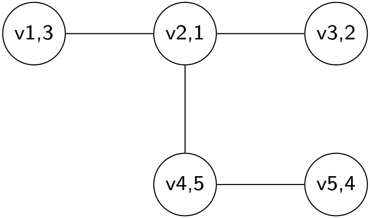
$$B_f(v) = \max_{i:(i,j)\in E}\{|f(i) - f(j)|\}$$

$$B_f(G) = \max_{v\in V} B_f(v)$$

  Thus, we want to find a labeling f such that the graph bandwidth $B_f(G)$ is minimum

Example

Let $G = (V, E)$ with $|V| = 5$, labeling f such that $f(v_1) = 3$, $f(v_2) = 1$, $f(v_4) = 5$, $f(v_5) = 4$, represented by the following graph and adjacency matrix:



$$A(f) = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

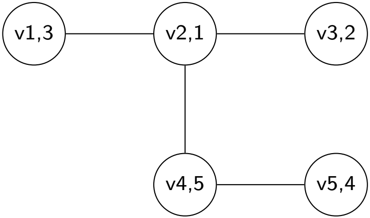
We can compute the bandwidth of each vertex and the bandwidth of the graph G under f:

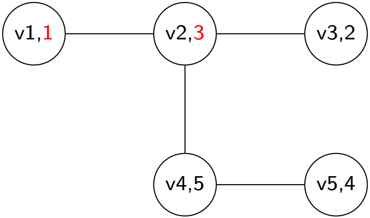
$$B_f(v_1) = \max\{|1-3|\} = 2 \qquad B_f(v_2) = \max\{|3-1|, |2-1|, |5-1|\} = 4$$
$$B_f(v_3) = \max\{|1-2|\} = 1 \qquad B_f(v_4) = \max\{|1-5|, |4-5|\} = 4$$
$$B_f(v_5) = \max\{|5-4|\} = 1$$

$$B_f(G) = \max_{v \in V} B_f(v) = \max\{2, 4, 1, 4, 1\} = 4$$

Example

One can easily notice that we can exchange the label of vertex v_1 with the label of vertex v_2, resulting in the graph and adjacency matrix on the right side:





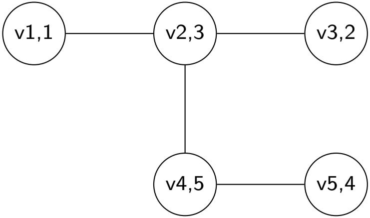
$$A(f) = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Before

$$A(f') = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

After

Example



$$A(f') = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Computing again the bandwidth of each vertex and the bandwidth of the graph G under f':

$$B_{f'}(v_1) = \max\{|3-1|\} = 2 \qquad B_{f'}(v_2) = \max\{|1-3|, |2-3|, |5-3|\} = 2$$
$$B_{f'}(v_3) = \max\{|3-2|\} = 1 \qquad B_{f'}(v_4) = \max\{|3-5|, |4-5|\} = 2$$
$$B_{f'}(v_5) = \max\{|5-4|\} = 1$$

$$B_{f'}(G) = \max_{v \in V} B_{f'}(v) = \max\{2, 2, 1, 2, 1\} = 2$$

Hence, the bandwidth of the graph has been reduced

Complexity

- This is a long-established combinatorial optimization problem
- Unfortunately, it is NP-complete, which means that it cannot be solved in polynomial time in any known way
- For a graph with n vertices, the number of possible labeling is $n!$
- The brute-force method has a running time complexity of $O(n!)$
- Hence, this approach is impractical even for small graphs with only 10 vertices

Relevance

- The main application of the bandwidth minimization problem is to solve large linear systems, because Gaussian elimination can be performed in $O(nb^2)$ time on matrices of dimension n and bandwidth b, which is a big win over the regular $O(n^3)$ if b is smaller than n
- Arranging a set of n circuit components in a line on a circuit board in such a way to minimize the length of the longest wire, which directly impacts time delay
- Finite element methods for approximating solutions of partial differential equations
- Hypertext layout
- Chemical kinetics
- Numerical geophysics

Formulation

(Peak Memory Consumption Minimization Problem) Let $G = (V, A)$ be a finite **directed** graph and an injective function $f : V \to \{1, 2, \dots, n\}$ labeling its vertices, where V is the set of vertices and A a set of **ordered pairs of vertices** representing **arcs**. The bandwidth of a vertex v and of a graph G are respectively defined as

$$B_f(v) = \max_{i:(i,j) \in A} \{|f(i) - f(j)|\}$$

$$B_f(G) = \max_{v \in V} B_f(v)$$

Thus, we want to find a labeling f such that the graph bandwidth $B_f(G)$ is minimum, with the constraint that **if there exists an arc from v to v', then $f(v) < f(v')$**

Example

(**Case 1**) All variables are deallocated at the end:

Variable	**Expression**	**Mem. Incr.**	**Mem. Total**	**Variables**
A	initA()	2	2	1
B	f(A)	7	9	2
C	f(A)	5	14	3
D	f(B)	3	17	4
E	f(C)	2	19	5
F	f(D, E)	1	20	6
-A, -B, -C, -D, -E	free A, B, C, D, E	-19	1	1
-F	return F	-1	0	0
		Average	**13.5**	**3.5**
		Peak	**20**	**6**

Example

(**Case 2**) Variables are deallocated when they are no longer necessary:

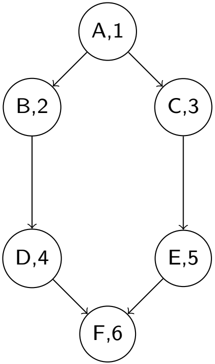
Variable	Expression	Mem. Incr.	Mem. Total	Variables
A	initA()	2	2	1
B	f(A)	7	9	2
C	f(A)	5	14	3
D	f(B)	3	17	4
-B	free B	-7	10	3
E	f(C)	2	12	4
-C	free C	-5	7	3
F	f(D, E)	1	8	4
-A	free A	-2	6	3
-D	free D	-3	3	2
-E	free E	-2	1	1
-F	return F	-1	0	0
		Average	**8.09**	**2.86**
		Peak	**17**	**4**

Example

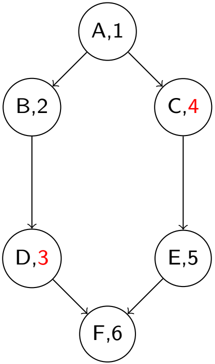
(**Case 3**) Variables are deallocated when they are no longer necessary and instructions are reordered:

Variable	Expression	Mem. Incr.	Mem. Total	Variables
A	initA()	2	2	1
B	f(A)	7	9	2
D	f(B)	3	12	3
-B	free B	-7	5	2
C	f(A)	5	10	3
-A	free A	-2	8	2
E	f(C)	2	10	3
-C	free C	-5	5	2
F	f(D, E)	1	6	3
-D	free D	-3	3	2
-E	free E	-2	1	1
-F	return F	-1	0	0
		Average	**6.82**	**2.18**
		Peak	**12**	**4**

Example



Before reordering (Cases 1 and 2)



After reordering (Case 3)

The problem

- Given a set of instructions in a particular order, the problem consists of minimizing peak memory consumption
- The algorithm to perform this task can be derived from the bandwidth minimization problem
- However, it is necessary to consider that if instruction B depends on the output of instruction A, then vertex A must have lower labeling than vertex B

Bibliography

- Miklós Bóna. *A walk through combinatorics: an introduction to enumeration and graph theory*. World Scientific, 2002.
- Elizabeth Cuthill and James McKee. "Reducing the bandwidth of sparse symmetric matrices". In: *Proceedings of the 1969 24th national conference*. ACM. 1969, pp. 157–172.
- Ch H Papadimitriou. "The NP-completeness of the bandwidth minimization problem". In: *Computing* 16.3 (1976), pp. 263–270.
- Chen Wang, Chuan Xu, and Abdel Lisser. "Bandwidth minimization problem". In: 2014.

Finally

Questions?