Optimal Bi-directional Determination of Sparse Jacobian Matrices

Mini Goyal and Shahadat Hossain
Department of Mathematics and Computer Science
University of Lethbridge, Canada

Second International Workshop on Combinatorial Scientific Computing
(CSC05), Toulouse, France
Outline

• The Problem
• Bi-directional Determination
• Optimal Bi-directional Determination
• Integer Linear Programming Model
• Complexity
• Preliminary Experimental Results
Introduction

Let

$$F = \left( \begin{array}{cccc} f_1 & f_2 & \ldots & f_m \end{array} \right)^T$$

be a mapping $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Assume that $F$ is continuously differentiable in the domain of interest and let $F'(x)$ denote the Jacobian matrix of $F$ at $x$.

Given vectors $s \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$, we can compute

$$b = F'(x)s$$

via one forward pass of automatic differentiation (AD).

$$c^T = w^T F'(x)$$

via one reverse pass of AD.
Assumptions

- Jacobian matrix is sparse
- The sparsity pattern of the Jacobian matrix is known a priori and independent of the actual values of $x$. 
Let $F'(x) \equiv A$,

$$A = \begin{array}{cccc}
  i_1 & a_{i_1 j} & a_{i_1 l} & 0 \\
  k_1 & 0 & 0 & a_{k_1 j} \\
  i_2 & 0 & 0 & a_{k_1 l} \\
  k_2 & a_{i_2 j} & a_{i_2 l} & 0 \\
  i_3 & 0 & 0 & a_{k_2 j} \\
  k_3 & a_{i_3 j} & a_{i_3 l} & 0 \\
\end{array}$$

Columns $j$ and $l$ are **structurally orthogonal** i.e. there does not exist a row index $i$ for which both $a_{ij} \neq 0$ and $a_{il} \neq 0$. Determine the unknowns in columns $j$ and $l$ of matrix $A$ from the product $As = b$ (obtained via one forward pass).
Examples

Partition the columns of $A$ into structurally orthogonal groups of columns,

\[
A = \begin{bmatrix}
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
\end{bmatrix},
S = \begin{bmatrix}
  1 & 0 \\
  0 & 1 \\
  0 & 1 \\
  0 & 1 \\
\end{bmatrix}
\]

Partition the columns of $A^T$ into structurally orthogonal groups of columns,

\[
A^T = \begin{bmatrix}
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times & \times \\
\end{bmatrix},
W^T = \begin{bmatrix}
  1 & 0 & 0 & 0 & 0 \\
  0 & 1 & 1 & 1 & 1 \\
\end{bmatrix}
\]
The Arrowhead Example

\[ A = \begin{bmatrix}
  \times & \times & \times & \times & \times \\
  \times & \times    \\
  \times   \\
  \times & \times \\
  \times & \times 
\end{bmatrix} \]

We need 5 matrix-vector multiplications either by forward or by reverse AD.
Bi-directional Determination of Sparse Jacobian Matrices

Obtain vectors \( s_1, s_2, \ldots, s_{p_c} \) and \( w_1, w_2, \ldots, w_{p_r} \) such that matrix-vector products

\[
b_i = A s_i, \quad i = 1, 2, \ldots, p_c \quad \text{or} \quad B = A S
\]

and the vector-matrix product

\[
c_j^T = w_j^T A, \quad j = 1, 2, \ldots, p_r \quad \text{or} \quad C^T = W^T A
\]

determine the \( m \times n \) matrix \( A \) uniquely.
Computing the Arrowhead Matrix

\[ A = \begin{bmatrix}
  \times & \times & \times & \times & \times \\
  \times & \times \\
  \times & \times \\
  \times & \times \\
  \times & \times \\
\end{bmatrix} \]

\[ S = \begin{bmatrix}
  1 & 0 \\
  0 & 1 \\
  0 & 1 \\
  0 & 1 \\
  0 & 1 \\
\end{bmatrix} \]

\[ W^T = \begin{bmatrix}
  1 & 0 & 0 & 0 & 0 \\
\end{bmatrix} \]

Two forward passes and one reverse pass are sufficient to determine \( A \). If the seed matrices \( S \) and \( W \) are such that the nonzero entries of \( A \) can be read-off from the products \( AS = B \) and \( W^T A = C^T \) than we have **direct determination**.
Efficient Bi-directional Determination of Sparse Jacobian Matrices

Obtain vectors $s_1, s_2, \ldots, s_{p_c}$ and $w_1, w_2, \ldots, w_{p_r}$ such that matrix-vector products $B = AS$ and the vector-matrix product $C^T = W^T A$ determine the $m \times n$ matrix $A$ uniquely and $p_r + p_c$ is minimized.

A $p$-coloring of graph $G = (V, E)$ is a function $\phi : V \rightarrow \{1, \ldots, p\}$ such that $\phi(v_i) \neq \phi(v_j)$ if $\{v_i, v_j\} \in E$.

Let $A \in \mathbb{R}^{m \times n}$. Define $G_b(A) = (U \cup V, E)$ where $U$ corresponds the set of column vertices and $V$ corresponds the set of row vertices and for $u_j \in U$ and $v_i \in V$, $\{v_i, u_j\} \in E$ if $a_{ij} \neq 0$. 
A Graph Coloring Formulation

**Bi-directional p-coloring:** A mapping \( \phi : \{U \cup V\} \to \{1, 2, \ldots, p\} \) is called a *bi-directional p-coloring* of bipartite graph \( G_b = (U \cup V, E) \) if the following conditions apply:

1. \( \phi \) is \( p \)-coloring.

2. The set of colors used on vertices in \( U \) and \( V \) are disjoint, i.e. for \( u_j \in U \) and \( v_i \in V \):
\[
\phi(u_j) \neq \phi(v_i).
\]

3. Every path of length 3 in \( G_b(A) \) uses at least 3 different colors.

The *bi-chromatic number*, \( \chi_b \), of \( G_b(A) \) is the smallest \( p \) for which \( G_b(A) \) has a bi-directional \( p \)-coloring.
Example

Given a sparse matrix $A$, obtain a bi-directional $p$-coloring of $G_b(A)$ such that $p = p_r + p_c$ is minimized.

Figure 1: Optimal bi-directional $p$-coloring of the arrowhead example
Bi-directional Determination of Sparse Jacobian Matrices

- Bi-coloring is NP-hard.
- Heuristic methods
- Exact methods
  - Let
    \[ \rho_{\text{max}} : \text{maximum number of nonzeros in any row}, \]
    \[ \kappa_{\text{max}} : \text{maximum number of nonzeros in any column}. \]
    A lower bound on the number of matrix-vector (or vector-matrix) products in one dimensional determination of \( A \) is \( \min(\kappa_{\text{max}}, \rho_{\text{max}}) \).
  - Find a good lower bound on the number of matrix-vector (vector-matrix) products in bi-directional determination.
Optimal Bidirectional Determination - An Integer Linear Programming Formulation (ILP)

Variables used in the ILP formulation of bi-directional $p$-coloring follows.

- 0-1 variable $w_j$ denotes whether ($w_j = 1$) or not ($w_j = 0$) color $j$, $1 \leq j \leq p_U$ has been assigned to some vertex $u \in U$.

- 0-1 variable $w_j$ denotes whether ($w_j = 1$) or not ($w_j = 0$) color $j$, $p_U + 1 \leq j \leq p_U + p_V$ has been assigned to some vertex $v \in V$.

- 0-1 variable $x_{i,j}$ denotes whether ($x_{i,j} = 1$) or not ($x_{i,j} = 0$) vertex $i$, $1 \leq i \leq n$ has been assigned color $j$, $1 \leq j \leq p_U$.

- 0-1 variable $x_{i,j}$ denotes whether ($x_{i,j} = 1$) or not ($x_{i,j} = 0$) vertex $i$, $n + 1 \leq i \leq m + n$ has been assigned color $j$, $p_U + 1 \leq j \leq p_U + p_V$. 
An ILP Model for Optimal Bi-directional Determination

\[
\text{minimize} \quad \sum_{j=1}^{p_U+p_V} w_j \tag{1}
\]

\[
\sum_{j=1}^{p_U} x_{i,j} = 1, \text{ for } i \in U \tag{2}
\]

\[
\sum_{j=p_U+1}^{p_U+p_V} x_{i,j} = 1, \text{ for } i \in V \tag{3}
\]

\[
x_{i,j} + x_{l,j'} + x_{l',j} + x_{l',j'} \leq (w_j + w_{j'} + 1) \tag{4}
\]

(for every path \(v_i - u_l - v_{i'} - u_{l'}\) of length 3)
\[ w_j \leq \sum_{i \in U} x_{i,j} \quad \text{for} \quad j = 1, \ldots, p_U \quad (5) \]

\[ w_j \leq \sum_{i \in V} x_{i,j} \quad \text{for} \quad j = p_U + 1, \ldots, p_U + p_V \quad (6) \]

\[ \sum_{i \in U} x_{i,j} \leq n w_j \quad \text{for} \quad j = 1, \ldots, p_U \quad (7) \]

\[ \sum_{i \in V} x_{i,j} \leq m w_j \quad \text{for} \quad j = p_U + 1, \ldots, p_U + p_V \quad (8) \]

\[ w_{j+1} \leq w_j \quad \text{for} \quad j = 1, \ldots, p_U - 1 \quad (9) \]

\[ w_{j+1} \leq w_j \quad \text{for} \quad j = p_U + 1, \ldots, p_U + p_V - 1 \quad (10) \]

\[ w_j \in \{0, 1\}, \quad \text{for} \quad 1 \leq j \leq p_U + p_V \quad (11) \]

\[ x_{i,j} \in \{0, 1\}, \quad \text{for} \quad i \in U \cup V, 1 \leq j \leq p_U + p_V \quad (12) \]
Null Color Elimination

**Null Color:** Consider a $p$-coloring problem with colors 1...$p$ for a graph $G(V,E)$. Assuming that $G$ can be optimally colored with $p - 1$ colors, consider a solution where color $i$ is not used: $(n_1, n_2, ..., n_{i-1}, n_i(=0), n_{i+1}, ..., n_p)$, where $n_i$ denotes the number of vertices colored with color $i$. The color $i$ for which $n_i = 0$ is known as the *null color*.

Example, the assignment (1,0,2,3) is equivalent to (1,3,2,0), (0,1,2,3), (1,2,0,3).

The constraints (9) and (10) ensures that in a feasible solution, the null colors will not be present.
Complexity

Number of variables:

\[(n + 1)p_U + (m + 1)p_V\]

Number of 3-paths:

\[\text{num3paths} = \sum_{i=1}^{m} (\rho_i - 1) \left[ \sum_{j: a_{ij} \neq 0} (\kappa_j - 1) \right]\]

Number of constraints:

\[(\text{num3paths} \times p_U \times p_V) + (m + n) + 2(p_U + p_V) + (p_U + p_V - 2)\]
### Experimental Results

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Statistics</th>
<th>One-directional*</th>
<th>Bi-directional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho_{max}$</td>
<td>$\kappa_{max}$</td>
<td>$DSM$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RG</td>
</tr>
<tr>
<td>ibm32</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>ash219</td>
<td>2</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>ash331</td>
<td>2</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>ash608</td>
<td>2</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>impcol-a</td>
<td>8</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>impcol-c</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

* column partitioning
RG : total number of row groups
CG : total number of column groups
TG : RG + CG
$\rho_{max}$ : maximum number of nonzeros in any row
$\kappa_{max}$ : maximum number of nonzeros in any column
Conclusion

- Formulation of optimal bi-directional determination.
- Large problems are difficult solve:
  - Memory constraints
  - Symmetry
- More elaborate numerical tests are needed.