On Lower Bounds for the Optimal Jacobian Accumulation Problem on linearized DAGs

Viktor Mosenkis
The Numerical Algorithms Group Ltd.
and
Uwe Naumann
LuFG Informatik 12: Software and Tools for Computational Engineering (STCE), RWTH Aachen, Germany

Motivation: Computing the Jacobian on a DAG

\[ F : \mathbb{R}^2 \rightarrow \mathbb{R} : \quad y = \exp(x_1 \ast x_2) + x_2 \]

Single assignment code

\[
\begin{align*}
  v_1 &= x_1 \ast x_2; \\
  v_2 &= \exp(v_1); \\
  y &= v_2 + x_2;
\end{align*}
\]

yields **linearized DAG (l-DAG)**

**Associativity of the chain rule**

\[
\left( \frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2} \right) = (c_1 c_3 c_5, c_4 + c_2 c_3 c_5), \quad 4m(ults)
\]

vs.

\[
\left( \frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2} \right) = (s c_1, (c_4 + s c_2)), \text{ where } s = c_5 c_3, \quad 3m
\]
Vertex Elimination\footnote{A. Griewank and S. Reese: \textit{On the Calculation of Jacobian Matrices by the Markowitz Rule}, AD1991.}
### Optimal Jacobian Accumulation

Computation of the Jacobian with minimal number of multiplications is intractable\(^2\)

### Complete Elimination Sequences

Vertex elimination sequence is complete if it transforms the I-DAG into a bipartite graph

### Optimal Vertex Elimination

Find a complete vertex elimination sequence with minimal costs \(VE(G)\)

---

Edge elimination

\[ G - (v_4, v_5)^+ \]

\[ G - (v_4, v_5)^- \]
### Complete Elimination Sequences

Edge elimination sequence is complete if it transforms the l-DAG into a bipartite graph.

### Optimal Edge Elimination

Find a complete edge elimination sequence with minimal costs $EE(G)$.
## Costs and Notation

### Notation

Consider l-DAG $G$

- **front elimination of** $(i, j) \in E(G)$: $G - (i, j)^+$
- **back elimination of** $(i, j) \in E(G)$: $G - (i, j)^-$
- **elimination of** $v \in V(G)$: $G - v$

### Costs

Consider l-DAG $G$

- **front elimination of** $(i, j) \in E(G)$: $\sigma_G((i, j)^+) = |S_G(j)|$
- **back elimination of** $(i, j) \in E(G)$: $\sigma_G((i, j)^-) = |P_G(i)|$
- **elimination of** $v \in V(G)$: $\sigma_G(v) = |P_G(v)| \cdot |S_G(v)|$
Obvious Lower Bounds

- $VE(G) \geq EE(G) \geq |Z(G)|$, where $Z(G)$ is set of all intermediate vertices
- $VE(G) \geq EE(G) \geq |E^*(G)| - |Z(G)|$, where $E^*(G)$ contains all edges of $G$, which do not connect inputs with outputs

- $|Z(G)| = 3$
- $|E^*(G)| - |Z(G)| = 11 - 3 = 8$
- $EE(G) = VE(G) = 22$

\[
\downarrow
\]
no reduction of the search space for the example
How can we get better lower bounds?
"Subgraph Property"

**Definition**
Consider DAGs $G$ and $G'$. We call $G' \subseteq G$ if $V(G') \subseteq V(G)$, $Z(G') \subseteq Z(G)$ and $E(G') \subseteq E(G)$.

**Lemma**
Consider DAGs $G$ and $G'$ with $G' \subseteq G$. Let $(i, j) \in E^*(G')$ and $v \in Z(G')$ then

\[
G' \setminus (i, j)^+ \subseteq G \setminus (i, j)^+ \quad \text{and} \quad G' \setminus v \subseteq G \setminus v
\]

**Lemma**
Consider DAGs $G$ and $G'$ with $G' \subseteq G$. Let $(i, j) \in E^*(G) \setminus E^*(G')$ and $v \in Z(G) \setminus Z(G')$ then

\[
G' \subseteq G \setminus (i, j)^+ \quad \text{and} \quad G' \subseteq G \setminus (i, j)^- \quad \text{and} \quad G' \subseteq G \setminus v
\]
Lower Bounds on “Spanning Trees”

Lemma

Let $T$ be a “spanning tree”\(^3\) of $G$, $v \in Z(T)$. Then

$$T - v \subseteq G - v.$$  

Corollary

Let $T$ be a “spanning tree” of $G$. Then

$$VE(T) \leq VE(G),$$

$$EE(T) \leq EE(G).$$

---

\(^3\) absorption-free subgraphs; V. Mosenkis and U. Naumann: *On Optimality Preserving Eliminations for the Minimum Edge Count and Optimal Jacobian Accumulation Problems in Linearized DAGs*. To appear in Optimization Methods and Software.
Markowitz-Degree

\[ VE(G) \geq EE(G) \geq \sum_{v \in Z(T)} |P_T(v)| \cdot |S_T(v)| \]

\( |Z(G)| = 3 \)
\( |E^*(G)| - |Z(G)| = 8 \)
Markowitz\((T_1) = 9\)
Markowitz\((T_2) = 11\)
Critical Edge Degree

**Definition**

Consider an l-DAG $G$. An edge $(i, j) \in E^*(G)$ is called $n$-critical for vertex (edge) elimination if

$$VE(G) - VE(G') \geq n \quad (EE(G) - EE(G') \geq n)$$

where $G' = G[E(G) \setminus \{(i, j)\}]$. 

Estimators for critical degree

On absorption-free l-DAGs we can estimate the critical degree as follows:

- $i \in Z(G)$ and $j \in Z(G')$
  - $VE(G) - VE(G') \geq |X(i)| \cdot |Y(j)| + \min\{|X(i)|, |Y(j)|\}$
  - $EE(G) - EE(G') \geq |X(i)| \cdot |Y(j)| + \min\{|X(i)|, |Y(j)|\}$
- $i \notin Z(G)$
  - $VE(G) - VE(G') \geq |Y(j)|$
  - $EE(G) - EE(G') \geq |Y(j)|$
- $j \notin Z(G)$
  - $VE(G) - VE(G') \geq |X(i)|$
  - $EE(G) - EE(G') \geq |X(i)|$

- $X(i)$ set of inputs that are connected to $i$
- $Y(j)$ set of outputs that are reachable from $j$
Critical Edge Degree

\[ VE(G) \geq |X(i)| \cdot |Y(j)| + \min\{|X(i)|, |Y(j)|\} + VE(T \setminus (i, j)) \]
\[ EE(G) \geq |X(i)| \cdot |Y(j)| + \min\{|X(i)|, |Y(j)|\} + EE(T \setminus (i, j)) \]

\[ C(T_1) = 20 \]
\[ C(T_2) = 15 + C(T_2 \setminus (4, 5)) \]
Example: Critical Edge Degree

\[ \begin{align*}
VE(G) & \geq |X(i)| \cdot |Y(j)| + \min\{|X(i)|, |Y(j)|\} + VE(T \setminus (i, j)) \\
EE(G) & \geq |X(i)| \cdot |Y(j)| + \min\{|X(i)|, |Y(j)|\} + EE(T \setminus (i, j))
\end{align*} \]

\[ T_1 \]
\[ T_2^1 := T_2 \setminus (4, 5) \]

\[ C(T_1) = 20 \]
\[ C(T_2) = 15 + 3 + C(T_2^1 \setminus (5, 6)) \]
Example: Critical Edge Degree

\[
VE(G) \geq |X(i)| \cdot |Y(j)| + \min\{|X(i)|, |Y(j)|\} + VE(T \setminus (i, j))
\]

\[
EE(G) \geq |X(i)| \cdot |Y(j)| + \min\{|X(i)|, |Y(j)|\} + EE(T \setminus (i, j))
\]

\[
C(T_1) = 20
C(T_2) = 15 + 3 + 2 = 20
Markowitz(T_2) = 11
|E^*(G)| - |Z(G)| = 8
VE(T_1) = EE(T_1) = 21
VE(T_2) = EE(T_2) = 21
\]
How good are the lower bounds?

Quality of the lower bounds depends on paths length in the l-DAG

▶ Good news
  ▶ we can solve absorption-free l-DAGs with maximal path length 3 in linear time

▶ Bad news
  ▶ the longer the paths in the l-DAG the worse are the lower bounds
### Results for solving $EE$ problem

<table>
<thead>
<tr>
<th>I-DAG</th>
<th>lower bounds</th>
<th>checked sequences</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>lion</td>
<td>simple</td>
<td>2594</td>
<td>0.02 sec</td>
</tr>
<tr>
<td>Markowitz</td>
<td></td>
<td>231</td>
<td>0.002 sec</td>
</tr>
<tr>
<td>crit. degree</td>
<td></td>
<td>57</td>
<td>0.002 sec</td>
</tr>
<tr>
<td>bat</td>
<td>simple</td>
<td>208701225</td>
<td>2538 sec</td>
</tr>
<tr>
<td>Markowitz</td>
<td></td>
<td>5367293</td>
<td>146 sec</td>
</tr>
<tr>
<td>crit. degree</td>
<td></td>
<td>2149352</td>
<td>68 sec</td>
</tr>
<tr>
<td>example</td>
<td>simple</td>
<td>116263115</td>
<td>1218 sec</td>
</tr>
<tr>
<td>Markowitz</td>
<td></td>
<td>1172296</td>
<td>31 sec</td>
</tr>
<tr>
<td>crit. degree</td>
<td></td>
<td>533</td>
<td>0.04 sec</td>
</tr>
</tbody>
</table>
Results for solving $EE$ problem

lion

bat
Optimality Preserving Eliminations (OPE)

**Definition (OPE)**

Consider an l-DAG $G$, $v \in V(G)$ and $(i, j) \in E^*(G)$. The elimination of $v$ ($(i, j)\{+,-\}$) is optimality preserving if

$$VE(G) = VE(G - v) + \sigma_G(v)$$

$$(EE(G) = EE(G - (i, j)\{+,-\}) + \sigma_G((i, j)\{+,-\}))$$

**Lemma (OPE of bridges on absorption-free l-DAGs)**

Consider an absorption-free l-DAG $G$, $a, b \in V(G)$, $S_G(a) = \{b\}$ and $P_G(a) = \{b\}$. If $|P_G(a)| \leq |S_G(b)|$, then elimination of $a$ preserves optimality. If $|P_G(a)| \geq |S_G(b)|$, then elimination of $b$ preserves optimality.
OPE of bridges: Example

\[ V_E(T_1) = 3 + V_E(T_1 - 5) \]
OPE of bridges: Example

\[ T_1^1 := T_1 - 5 \]

\[ V E(T_1) \geq 3 + 10 + V E(T_1 \setminus (4, 6)) \]
OPE of bridges: Example

\[ T_1^1 \setminus (4, 6) \]

\[ VE(T_1) \geq 3 + 10 + 4 + 4 = 21 \]
OPE of bridges and critical degree

... build a symbiosis. You can create a bridge by removing edges.
Absorption-free l-DAGs

Lemma

For every absorption-free l-DAG $G$ holds $VE(G) = EE(G)$
Outlook

▶ use local absorption-free l-DAG to determine lower bounds
▶ retract information from edges that were removed to build a "spanning tree"
▶ more OPE not only on absorption-free l-DAGs
▶ find optimal "spanning trees"
▶ strategy for choosing the edge for estimating the critical-degree
Thank you