# Reducing the number of AD passes for computing a partitioned sparse Jacobian matrix

**Shahadat Hossain**
Department of Mathematics and Computer Science
University of Lethbridge
**Trond Steihaug**
Department of Informatics
University of Bergen

AD 2000 The 3rd International Conference/Workshop on Automatic Differentiation:
From Simulation to Optimization
June 19-23, 2000
Nice, France





## Outline

- The Jacobian Computation Problem
- Direct Methods and Partitioning
- Can we improve when the CPR is optimal?
  Curtis, Powell, and Reid 74, Coleman and Moré 83
- Substitution Ordering
- Computational Results
- Concluding Remarks





## The Jacobian Computation Problem

Given a nonlinear function $F : R^n \to R^m$
The Jacobian matrix $F'(x)$ can be approximated by

- Finite-Difference
- Automatic Differentiation (AD)

**Efficient Computation**:

- Utilize sparsity structure of $F'$





## Assumptions:

- The Jacobian matrix is sparse
- The sparsity pattern of the Jacobian matrix is known a priori or can be determined and is independent of the actual values of $x$
- If we need one or more components of $F$ at $x$ we have to compute the whole vector $F(x)$
  - Evaluation $F(x)$ costs as much as computing $F_i(x)$ separately (common sub-expressions are evaluated only once)
  - $F$ is a computer subroutine that returns the vector $F(x)$

**In the following** $A = F'(x)$

## Partitioning and Direct Methods

**The Problem**:

Given the sparsity structure of a matrix $A \in R^{m \times n}$ obtain $X \in R^{n \times p}$ such that nonzeros of $A$ can be recovered "easily" with $p$ as small as possible.

Solve for nonzero $a_{ij}$ in

$$X^T A^T = B^T.$$

where the columns of $B$ are matrix-vector products of $A$ computed from one AD pass or a difference in function evaluations $B \leftarrow AX$

- The maximum number of nonzeros ($\rho$) in any row of $A$ is a lower bound on the number of AD passes or function evaluations ($p$) required





## Partitioning and Direct Methods

A group of columns are said to be *structurally orthogonal* if they do not contain nonzeros in the same row position.

*Let the columns be partitioned into structurally orthogonal groups* $C_k, \ k = 1, 2, \ldots, p$

**Direct Computation**:

Define

$$x_{jk} = 1 \text{ if } j \in C_k \text{ and } x_{jk} = 0 \text{ otherwise}$$

$$\boxed{X^T A^T = B^T \text{ defines } m \text{ diagonal linear systems of order } p}$$

**Question:** Can we improve when the partitioning is minimal?





## Substitution Ordering: Example

**Example**:

$$A = \begin{bmatrix} a_{11} & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ 0 & a_{32} & a_{33} \end{bmatrix}, \text{ and choose } X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Obtain $B = AX$ using forward mode of AD.

**Determine row** $2$ **of** $A$:

$$\begin{bmatrix} a_{21} & a_{22} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} b_{21} & b_{22} \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} b_{21} \\ b_{22} \end{bmatrix}$$





## Substitution Ordering: Example (2)

- The column intersection graph $G(A)$ is a complete graph (3 forward passes are needed)
- Each row contains 2 nonzeros
- Only two forward passes are needed using substitution ordering
- Solve for $A$ row by row

## Substitution Ordering

- $A \in R^{m\times n}$ contains at most $\rho$ nonzeros in a row.
- Columns are partitioned into $p \geq \rho + 1$ groups $C_i,\ i = 1, 2, \ldots, p$
- Define a new grouping

$$\mathcal{C}_i = \{C_i, C_{i+1}\}, \quad i = 1, 2, \ldots, p-1. \tag{1}$$

Notation: $\{C_i, \ldots, C_k\} = C_i \cup \cdots \cup C_k$.

**Claim: (1) defines a substitution ordering of the unknowns**





## Substitution Ordering

Assume $p = \rho + 1$, and let
$\alpha_1, \alpha_2, \ldots, \alpha_p$ row $i$ of $A$, and $\beta_1, \beta_2, \ldots, \beta_{p-1}$ row $i$ of $B$.
Then

$$\begin{bmatrix} 1 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \\ \alpha_k \\ \alpha_{k+1} \\ \vdots \\ \alpha_{p-1} \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \\ \beta_k \\ \vdots \\ \beta_{p-2} \\ \beta_{p-1} \end{bmatrix} \tag{2}$$





## Substitution Ordering

Let $\alpha_k = 0$ for some $k \in \{1, 2, \ldots, p\}$ (every row has at least one zero since $p > \rho$). Then $(1 < k < p)$

$$\alpha_{k-1} + \alpha_k = \beta_{k-1}$$

$$\alpha_k + \alpha_{k+1} = \beta_k,$$

$\Longrightarrow$ $\alpha_{k-1}$ is determined directly and $\alpha_{k-2} \ldots \alpha_1$ are determined by substitution and

$\Longrightarrow$ $\alpha_{k+1}$ is determined directly and thus $\alpha_{k+2} \ldots \alpha_p$ are determined by substitution

**Lemma 1** *If $p > \rho$ then $\{C_1, C_2\}, \{C_2, C_3\}, \cdots, \{C_{p-1}, C_p\}$ defines a substitution ordering of the nonzero elements of $A$.*





## Substitution Ordering: When $p \gg \rho$

- Any arbitrary combination of $\rho + 1$ column groups saves one AD pass
- If $p = k(\rho + 1)$ we save $k$ passes
- If $p$ *modulo* $(\rho + 1) \neq 0$, the total AD passes is at most $p - \lfloor \frac{p}{(\rho+1)} \rfloor$

**Example**:

$$\begin{bmatrix} \times & \times & & & & \\ & & \times & \times & & \\ & & & & \times & \times \\ \times & & \times & & & \\ \times & & & \times & & \\ & \times & \times & & & \\ & \times & & \times & & \\ \times & & & & \times & \\ \times & & & & & \times \\ & \times & & & \times & \\ & \times & & & & \times \\ & & \times & & \times & \\ & & \times & & & \times \\ & & & \times & \times & \\ & & & \times & & \times \end{bmatrix},$$

New grouping:
$\{C_1, C_2\},\ \{C_2, C_3\}$ and $\{C_4, C_5\},\ \{C_5, C_6\}$
where $C_j$ contains column
$j$ of $A$ for $j = 1, 2, \ldots, 6$
saves 2 forward passes

## Exploiting Row (Column) Sparsity

- Including more than two structurally orthogonal column groups in each of the new column groups will reduce the number of new groups
- The sparsity structure of the compressed Jacobian rows determine whether or not a substitution ordering is obtained

**Example**: Let $p = 8$, $\rho \leq 6$, and define

$$\mathcal{C}_i = \{C_i, C_{i+1}, C_{i+2}\} \quad i = 1, 2, \ldots, p-2.$$

Assume $\alpha_3 = \alpha_4 = 0$. Then

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix}$$





## Exploiting Row (Column) Sparsity

**Example**:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix}$$

**Observation**:

If at least two consecutive elements are known to be zero, then the system would split and the unknowns could be recovered by substitution

**Result**:

**Lemma 2** *If $p > \rho$ and each row of the compressed matrix has at least $d$ consecutive zero entries, $d \leq p - \rho$, then the $p - d$ groups $\{C_1, \ldots, C_{d+1}\}$, $\{C_2, \ldots, C_{d+2}\}, \cdots, \{C_{p-d}, \ldots C_p\}$ defines a substitution ordering of the nonzero elements.*





## Computational Results

| Name | Direct(DSM) | | | | | | Substitution | |
|---|---|---|---|---|---|---|---|---|
| | $A$ | | | $A^T$ | | | $A$ | $A^T$ |
| | $\rho$ | mngp | $p$ | $\rho$ | mngp | $p$ | $d$ | $d$ |
| abb313 | 6 | 10 | 10* | 26 | 26 | 26 | 1 | 0 |
| ash219 | 2 | 3 | 4* | 9 | 9 | 9 | 1 | 0 |
| ash331 | 2 | 6 | 6* | 12 | 12 | 12 | 2 | 0 |
| ash608 | 2 | 5 | 6* | 12 | 12 | 12 | 2 | 0 |
| ash958 | 2 | 6 | 6* | 13 | 13 | 13 | 2 | 0 |
| will199 | 6 | 7 | 7* | 9 | 9 | 9 | 1 | 0 |
| str | 34 | 34 | 34 | 34 | 34 | 35* | 0 | 1 |
| str200 | 34 | 34 | 34 | 26 | 26 | 31* | 0 | 2 |
| str400 | 33 | 33 | 33 | 34 | 36 | 36* | 0 | 1 |
| str600 | 33 | 33 | 33 | 34 | 36 | 36* | 0 | 1 |
| fs541-1 | 11 | 11 | 13* | 541 | 541 | 541 | 1 | 0 |
| fs541-2 | 11 | 11 | 13* | 541 | 541 | 541 | 1 | 0 |
| lundA | 12 | 12 | 12 | 12 | 12 | 13* | 0 | 1 |
| lundB | 12 | 12 | 12 | 12 | 12 | 13* | 0 | 1 |
| Total | 200 | 217 | 223 | 1315 | 1319 | 1327 | 11 | 7 |

Table 1: Results for SMTAPE collection

- $d$: Number of consecutive zeros in the compressed Jacobian matrix obtained from the corresponding consistent partition





## Concluding Remarks

- The proposed substitution techniques can be extended to the complete direct cover method (Hossain and Steihaug [HS98]) or the bipartition (Coleman and Verma [CV98])
- The technique of Newsam and Ramsdell (Newsam and Ramsdell [NR83], Geitner, Utke, and Griewank [GUG96]) requires only $\rho$ AD passes ; the proposed substitution is close to this minimum and avoids the possible ill-conditioning
- Research Question: Will a reordering of the groups or an interchange of columns (of the Jacobian matrix) between groups increase the number of consecutive zeros in the compressed matrix?