The CPR Method and Beyond : Prologue

Shahadat Hossain[1], Trond Steihaug[2],

[1] University of Lethbridge, Department of Mathematics and Computer Science
Lethbridge, AB T1K 3M4, Canada
shahadat.hossain@iuleth.ca
[2] University of Bergen, Department of Informatics
Bergen, N-5020 Norway
Trond.Steihaug@ii.uib.no

Abstract. In 1974 A.R. Curtis, M.J.D. Powell, and J.K.Reid published a seminal paper on the estimation of Jacobian matrices which was later coined as the CPR method. Central to the CPR method is the effective utilization of a priori known sparsity information. It is only recently that the optimal CPR method in its general form is characterized and the theoretical underpinning for the optimality is shown. In this short note we provide an overview of the development of computational techniques and software tools for the estimation of Jacobian matrices.

Keywords. Structural Orthogonality, Optimal CPR, Sparse Jacobian Estimation Software.

1 An Overview of Sparse Jacobian Matrix Determination

The determination of the sparse Jacobian matrix with a priori known sparsity pattern of at least once continuously differentiable mapping $F : I\!R^n \to I\!R^m$ can be viewed as a computation of the form $AS \equiv B$ where A is an approximation of the Jacobian matrix $F'(x)$ but has the same sparsity pattern and $S \in I\!R^{n\times p}$ is a matrix of p directions. For example, using finite difference approximation we can write

$$\left.\frac{\partial F(x+ts)}{\partial t}\right|_{t=0} = F'(x)s \approx As = \frac{1}{\varepsilon}[F(x+\varepsilon s) - F(x)] \equiv b, \qquad (1)$$

where $s \in I\!R^n$ is a given direction. The forward mode of automatic differentiation allows us to compute the product $F'(x)s$ with one forward sweep at a cost equal to a small multiple of the cost of evaluating F at x. If the ρ_i rows of S defined by the ρ_i nonzero unknowns of $A(i,:)$ are linearly independent then $A(i,:)$ is uniquely determined from $B(i,:)$. This consistency condition is achieved, for example, by taking S to be the identity matrix ($p = n$), the Vandermonde matrix ($p = \rho$) [8], or the Pascal matrix ($p = \rho$) [5], ρ being the maximum number of nonzero entries in any row of A. The computational cost for recovering the nonzero unknowns from B in the Vandermonde

Dagstuhl Seminar Proceedings 09061
Combinatorial Scientific Computing
http://drops.dagstuhl.de/opus/volltexte/2009/2079

and Pascal directions involve solving m linear systems of size ρ. We have *direct determination* if the nonzero unknowns can be recovered from B without any extra arithmetic operations. The CPR [3] method yielding direct determination, exploits A's sparsity to define $S(:,k) = \sum_{j \in \mathcal{C}_k} e_j, k = 1, \ldots, p$ where $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_p\}$ is a partition of column indices such that for each pair of indices $j, l \in \mathcal{C}_k$, the product $a_{ij}a_{il}, i = 1, \ldots, m$ is identically 0. A greedy heuristic whereby the kth *structurally orthogonal* group $\mathcal{C}_k$ is formed by including in it the maximum number of indices such that the corresponding columns are mutually structurally orthogonal, is used to define the column partition. The next major contribution to the CPR method is attributed to Coleman and Moré [1] who analyze the partitioning problem as the coloring of $G(A)$ – the column intersection graph of A, and prove that finding minimum cardinality structurally orthogonal column partition is NP-hard; an efficient software tool, the DSM [2], implementing the graph coloring CPR was released. As observed by Stanley Eisenstat the chromatic number of graph $G(A)$, $\chi(G(A))$, does not necessarily represent the minimum number of extra function evaluations needed to determine A. A closer examination of the Eisenstat example suggest that "columns be structurally orthogonal to be determined in one matrix-vector product" is rather restrictive for effective utilization of available sparsity. Inspired by the Eisenstat example, Hossain and Steihaug [6] define orthogonal partitioning of column segments and settle the *optimal CPR* as the minimum cardinality orthogonal partitioning of the column segments.

Theorem 1 ([6]). *Matrix $A \in I\!R^{m \times n}$ is directly determined using p matrix-vector products $AS, S \in I\!R^{n \times p}$ if and only if the nonzero entries of A are partitioned into p groups of "isolated elements" [8]. The minimal number of matrix-vector products needed in any direct determination of A is $p = \chi(G_{\mathcal{I}}(A))$ where $G_{\mathcal{I}}(A)$ is the "element isolation graph [8]" or, the "column segments graph" corresponding to m block row partition.*

In [6] it is further shown that the minimum cardinality row partition Π such that $\chi(G_{\Pi}(A)) = \chi(G_{\mathcal{I}}(A))$ can always be found and in practical problems the number of row blocks is usually much less than m.

Table 1. Structurally Orthogonal Column Partition

Matrix	m	n	nnz	ρ	p (DSM) [2]	p [4]
af23560	23560	23560	484256	21	41	**32**
cage11	39082	39082	559722	31	**62**	81
cage12	130228	130228	2032536	33	**68**	96
e40r0100	9661	9661	306356	62	70	**66**
ihr34	14270	14270	307858	63	**63**	65
ihr71c	70304	70304	1528092	63	**63**	65
Total				273	367	405

Table 1 displays problem instances (m rows, n columns, nnz nonzero entries) and partitioning results (number of structurally orthogonal column groups) obtained by publicly available software tool DSM (a C++ implementation of the original F77 code) and the result of a greedy "partial distance 2 coloring" heuristic applied on bipartite graph representation of the test matrices as reported in [4]. We note that both the DSM and the bipartite model of [4] require $\Theta(nnz)$ storage – thus achieving an important design objective identified in [7] for solving large and sparse problems.

2 Concluding Remarks

We have provided a brief account of the significant milestones in the development of the CPR method. Mathematical derivatives constitute an essential algorithmic ingredient in many computational procedures for solving scientific and engineering problems. Computation of derivative information is therefore an important computational task for which efficient algorithms and software tools are needed especially with the ever increasing size of the scientific problems needed to be solved.

References

1. T. F. Coleman and J. J. Moré. Estimation of sparse Jacobian matrices and graph coloring problems. *SIAM J. Numer. Anal.*, 20(1):187–209, 1983.
2. T. F. Coleman, B. S. Garbow, and J. J. Moré. Software for estimating sparse Jacobian matrices. *ACM Trans. Math. Software*, 10(3):329–345, 1984.
3. A. R. Curtis, M. J. D. Powell, and J. K. Reid. On the estimation of sparse Jacobian matrices. *J. Inst. Math. Appl.*, 13:117–119, 1974.
4. A. H. Gebremedhin, F. Manne, and A. Pothen. What color is your Jacobian? Graph coloring for computing derivatives. *SIAM Rev.*, 47(4):629–705, 2005.
5. S. Hossain and T. Steihaug. Sparsity Issues in the Computation of Jacobian Matrices. In T. Mora, editor, *Proceedings of the International Symposium on Symbolic and Algebraic Computing (ISSAC)*, pages 123–130, New York, 2002. ACM.
6. S. Hossain and T. Steihaug. Optimal Direct Determination of Sparse Jacobian Matrices *(to appear in Optimization Methods and Software)*.
7. J. R. Gilbert, S. Reinhardt, and V. Shah. High-performance graph algorithms from parallel sparse matrices. In B. Kågström, E. Elmroth, J. Dongarra, and J. Wasniewski, editors, *PARA*, volume 4699 of *Lecture Notes in Computer Science*, pages 260–269. Springer, 2006.
8. G. N. Newsam and J. D. Ramsdell. Estimation of sparse Jacobian matrices. *SIAM J. Alg. Disc. Meth.*, 4(3):404–417, 1983.