Miroslav Tůma

A note on direct methods for approximations of sparse Hessian matrices

**Terms of use:**

# A NOTE ON DIRECT METHODS FOR APPROXIMATIONS
# OF SPARSE HESSIAN MATRICES

Miroslav Tůma

*Summary.* Necessity of computing large sparse Hessian matrices gave birth to many methods for their effective approximation by differences of gradients. We adopt the so-called direct methods for this problem that we faced when developing programs for nonlinear optimization. A new approach used in the frame of symmetric sequential coloring is described. Numerical results illustrate the differences between this method and the popular Powell-Toint method.

*Keywords:* Symmetric Graph Coloring, Hessian Matrix Estimation, Large Sparse Optimization, Numerical Differences.

*AMS Classification:* 65D15, 65F20, 65K05

## 1. INTRODUCTION

First we will review some of the known basic facts and recommend information sources on this topic. Many optimization methods require estimation of the Hessian matrices of the objective function $F: R^n \to R$. When $n$ becomes very large, it is clear that two difficulties can be encountered: there may be not enough storage locations, and the computation time required may become too long. Fortunately, however, the matrices of these very large problems have a very high proportion of zero elements. The ratio of non-zero elements to the total number of matrix elements is called the density; a matrix with a high proportion of zero entries is called sparse. Besides their low density, these matrices often have a fixed sparsity pattern and in this case we can drastically diminish the minimum number of gradient evaluations necessary to approximate the matrix elements.

When the Hessian matrix is dense or in the case when the Hessian matrix is sparse but its structure changes, we can obtain its approximation by differencing the gradient along the coordinates. With a fixed sparsity pattern we can form differences along certain special composite vectors. As a rule the number of these vectors for usual Hessian matrices of large problems is only a small fraction of the dimension of the problem so that we can spare much of the computation time when we are to approxi-

mate a Hessian matrix many times for some optimization problem using for instance a Newton-like method.

These composite vectors are chosen with the help of a tool of the discrete mathematics − the graph theory. Instead of the sparse matrix we will use its adjacency graph. If we denote its node set by $\{1, ..., n\}$, where $n$ is the order of the matrix, then an edge is between two nodes $i$ and $j$ iff an element of the matrix with these indices is structurally nonzero. The matrix symmetry guarantees that this definition is correct.

Methods used are usually divided into three classes: Direct Methods, Substitution Methods and the most general class of Elimination Methods (see [6], [7]). Direct methods need the greatest number of the composite directions, as well as the greatest number of gradient evaluations, but we obtain the matrix elements directly, without any additional effort. As we move towards the class of Elimination Methods, this effort increases, numerical stability possibly decreases, but the number of gradient evaluations needed decreases as well. The question of accumulation errors is well described in [7].

The most popular algorithm for the direct method is the one in [7]. Questions of the algorithms are studied in detail in [1], and our graph and matrix terminology will conform to that used there. Thapa's New Direct Method (see [8]) is one of the alternatives which may help reduce the drawbacks of the Powell-Toint algorithm.

We present here a method based on the symmetric sequential coloring (see [1]) that overcomes some of the drawbacks of the Powell-Toint algorithm. Its complexity is dominated by the complexity of the symmetric sequential coloring, that is $O(\sum_{i=1}^{n} r_i^3)$, where $r_i$ is the number of nonzero elements in the $i$-th row of the matrix. The Powell-Toint algorithm has the complexity dominated by approximately $O(\sum_{i=1}^{n} r_i^2)$ operations. As the density of the problems in practice tends to decrease as the size of the problem increases, to compare the results of the methods is the more important. For some problems it is even worth to try both these methods, as the Hessian matrix may have to be computed many times and thus the difference between their results is not insignificant any more. Our search was motivated by the desire to find a method advantageous not only for one type of problems but as generally as possible, and to use it in the program system for functional optimization that is being created. Although further tests will be done and the research will continue we can use the algorithm below as a possible alternative, the best we have tried yet.

## 2. MOTIVATION AND ALGORITHM

The system of programs for the functional optimization UFO (see [4]) will contain methods that require to approximate many time sparse Hessian of the same pattern. It is worthwhile to have available an effective way, and the possibility of making

one's option for the evaluation of the Hessian that is repeated so often. We examined the Powell-Toint method and found that it can be very good when used for various types of regular grids, but it can behave relatively badly in some cases. The generalization of a bad example from [7] is as follows:

Example 2.1. Graph $G = (V, E)$, $p \geq 3$, which is the adjacency graph of a matrix, as defined.

$$V = V_1 \cup V_2 \, ,$$
$$V_1 = \{1, \ldots, p\} \, ,$$
$$V_2 = \{p + 1, \ldots, 2p\} \, ,$$
$$E = E_1 \cup E_2 \, ,$$

$(V_1, E_1)$ is the complete graph on $p$ vertices,

$$E_2 = \bigcup_{i=1}^{p} \{i, p + i\} \, .$$

The Powell-Toint algorithm needs $p + 1$ colors, that is $p + 1$ function evaluations. Our algorithm will need $p$ colors for the graph. That is $p$ function evaluations.

Let $G = (V, E)$ denote the adjacency graph of some sparse matrix and $v \in V$. The symbol $\deg_2 v$ will denote the cardinality of the vertex set the members of which have the distance 2 from $v$. The ordering idea, the most promising for us, is the following. We want to order the verices for the symmetric sequential coloring so that the vertices with great $\deg_2 v$ might be processed by the sequential symmetric coloring algorithm as late as possible. On the other hand it is desirable to color the vertices with great degrees as early as possible. The third rule embodied is to process earlier the locally dense structures of the adjacency graph. We illustrate this point by the bipartite graph taken from [1].

Example 2.2. Let $G = (V, E)$, $V = \{u_1, u_2, v_1, \ldots, v_n\}$, $E = \{\{u_i, v_j\} \mid i \in \hat{2}, j \in \hat{n}\}$. We have $\deg_2 u_1 = \deg_2 u_2 = 2$; $\deg_2 v_i$ for $i \in \hat{n}$ is equal to $n$. If we color first one of the vertices $v_i$, $i \in \hat{n}$, using the symmetric sequential algorithm, we use only three colors and that is optimum.

We have found that the following compromise can be used with many sparse matrix structures. Our aim is to find the ordering $v_1, \ldots, v_n$ of the vertices of the graph $G = (V, E)$ corresponding to the adjacency graph of some symmetric sparse matrix.

Ordering 2.3.: Set $G_1 = G = (V, E)$, $k = 1$. Let all vertices be unmarked.

2) Find the vertex $v_k$ of $G_k$ with the maximum degree.

3) Order its unmarked neighbors using the smallest-last ordering (see $|5|$). Denote them $v_{k+1}, \ldots, v_{k+p}$.

4) Mark all unmarked neighbors of the set $v_k, \ldots, v_{k+p}$. Mark also the elements of this set.

5) $k = k + p + 1$, $G_k$ is the graph induces by the unmarked vertices of $G$.

6) If $G_k$ is not empty then go to 2).

7) Mark all the rest of the graph using the smallest-last ordering.

Using this ordering we found even small examples showing that it locally out-performs the others. The smallest of them is as follows.

Example 2.4. Let $G = (V, E)$, $V = \{1, \ldots, 12\}$, $E = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{4, 5\}, \{4, 6\}, \{11, 12\}, \{5, 6\}, \{1, 7\}, \{4, 8\}, \{7, 8\}, \{2, 9\}, \{5, 10\}, \{9, 10\}, \{3, 11\}, \{6, 12\}\}$.

Our approach needs three colors whereas the Powell-Toint algorithm needs five colors.

## 3. EXPERIMENTS WITH DIRECT METHODS

Experimental results in the cited papers are based on only a small part of the sparse patterns (see [2]), mostly on grids. Although the problem is important for many applications, there is still lack of experimental results for various classes of data sets. Moreover, the concomitant lack of analytic results on heuristics requires that any comparison of algorithms must be empirical. Our experiments are not exhaustive, either. We have concentrated on two main groups of problems. The first contains the problems on grids, similar to those used for the problems of heat transfer (see [3], [9]). The matrix structures were created by means of the Sparse Matrix Structure Generator, which facilitated the work and will be part of the program system UFO (see [4]). The results obtained appear in Table 1. For these grids our algorithm seems to be a bit less advantageous. The other group, on which we could see possible advantages of the symmetric sequential coloring with our ordering, consists of matrices whose elements were generated randomly; however, we did not use the uniformly random sparsity pattern but generated regions with the density increased to $50-70\%$, while the rest was random. The density of the whole matrix was kept at $0.15-1\%$. We assumed that for these problems our algorithm could be of advantage

Table 1. Grid Patterns

| $n$ | $d\%$ | $pt$ | $ss$ |
|------|-------|------|------|
| 180 | 1·2 | 6 | 6 |
| 220 | 1·2 | 8 | 8 |
| 330 | 1·7 | 8 | 8 |
| 505 | 1·8 | 9 | 9 |
| 520 | 1·8 | 9 | 10 |
| 660 | 1·2 | 7 | 7 |
| 755 | 1·7 | 9 | 9 |
| 856 | 1·1 | 6 | 6 |
| 869 | 1·2 | 6 | 6 |
| 940 | 0·9 | 5 | 5 |
| 960 | 1·2 | 8 | 8 |
| 1200 | 1·4 | 10 | 10 |

Notation.

*n*     number of graph nodes, that is, the order of the matrix;

$d\%$   matrix density;

*pt*    number of colors found by the Powell-Toint algorithm;

*ss*    number of colors obtained by our method.

Table 2. Randomly Generated Patterns (see the description in the text). The columns are denoted in the same way as in Table. 1.

| *n* | $d\%$ | *pt* | *ss* |
|---|---|---|---|
| 100 | 1 | 6 | 6 |
| 200 | 1 | 10 | 10 |
| 300 | 1 | 9 | 9 |
| 400 | 1 | 12 | 12 |
| 500 | 1 | 10 | 9 |
| 600 | 0·5 | 11 | 12 |
| 700 | 0·5 | 12 | 11 |
| 800 | 0·5 | 13 | 13 |
| 900 | 0·5 | 7 | 7 |
| 1000 | 0·25 | 6 | 5 |
| 1100 | 0·25 | 9 | 9 |
| 1200 | 0·15 | 10 | 9 |

We can see that the experiments are promising and that both the alternatives for the approximation of sparse Hessian could be useful when embedded in the program system UFO. Let us draw no exaggerated conclusion from the simple test. Nonetheless, it is seen that for less regular patterns we can offer other ways of solution to our problem, although the Powell-Toint algorithm has been found to be more stable.

*References*

[1] *T. F. Coleman, J. J. Moré:* Estimation of Sparse Hessian Matrices and Graph Coloring Problems, Math. Prog. 28 (1984), 243—270.

[2] *G. C. Everstine:* A Comparison of Three Resequencing Algorithms for the Reduction of Matrix Profile and Wavefront, International Journal on Numerical Methods in Engineering 14 (1979), 837—853.

[3] *A. George, J. W. H. Liu:* Computer Solution of Large Sparse Positive Definite Systems, Prentice-Hall, Inc. Englewood Cliffs. N. J. 07632, 1981.

[4] *P. Hanzálek, J. Hřebíček, J. Kučera:* A Conversational Program System for Mathematical Optimization, Computer Physics Communications 41 (1986), 403—408.

[5] *D. M. Matula, L. L. Beck:* Smallest-last Ordering and Clustering and Graph Coloring Algorithms, JACM 30 (1983), 417—427.

[6] *S. T. McCormick:* Optimal Approximation of Sparse Hessians and its Equivalence to a Graph Coloring Problem, Math. Prog. 26 (1983), 153—171.

[7] *M. J. D. Powell, Ph. L. Toint:* On the Estimation of Sparse Hessian Matrices, SIAM on Num. Anal. 16 (1979), 1060—1074.

[8] *M. N. Thapa:* Optimization of Unconstrained Functions with Sparse Hessian Matrices: Newton-type Methods, Math. Prog. 19 (1984), 156—186.

[9] *O. C. Zienkiewicz:* The Finite Element Method, McGraw Hill, London, 1977.

Souhrn

## POZNÁMKA K PŘÍMÝM METODÁM APROXIMACE ŘÍDKÝCH HESSOVÝCH MATIC

MIROSLAV TŮMA

V práci je uveden alternativní algoritmus pro efektivní aproximaci řídkých Hessových matic, který byl vyvinut a testován v SVT ČSAV v souvislosti s potřebou rychlého chodu optimalizačních programů pro rozsáhlé úlohy. Jsou zde uvedeny výsledky praktických experimentů a porovnání s algoritmem Powella a Tointa.


Резюме

## ЗАМЕЧАНИЕ ПО ПРЯМЫМ МЕТОДАМ АППРОКСИМАЦИИ РАЗРЕЖЕННЫХ МАТРИЦ ГЕССА

MIROSLAV TŮMA

В статье описан новый алгоритм для эффективной аппроксимации разреженных матриц Гесса, который был разработан и опробован в связи с необходимостью быстрой оптимизации больших систем. В конце статьи приведены результаты практических экспериментов и их сравнение с результатами алгоритма Повелла и Тоинта.

*Author's address:* Ing. *Miroslav Tůma*, SVT ČSAV, Pod vodárenskou věží 2, 182 07 Praha 8.