Sparsity in Derivative Free Optimization

Lennart Frimannslund* Trond Steihaug†

Abstract

We consider the unconstrained optimization problem. If the objective function is two times continuously differentiable, it is well known that a sparse Hessian matrix implies that the function is separable. The property of the function being partially separable may be regarded as a structure of the function. In derivative free optimization (DFO) we do not assume that the function is smooth, but we still have structure. We will consider a large class of DFO methods using independent search directions where sparsity appears in a matrix used to derive the search directions. In the smooth case this matrix will be a similarity transformed sparse Hessian matrix. We will discuss the use of techniques from automatic differentiation to compute the structure of the matrix. We present a theorem regarding the average curvature properties of partially separable functions which need not be differentiable. This has implications for derivative-free optimization methods which make use of average curvature information to select the set of search directions.

1 Background

A Hessian matrix contains local curvature information. One can informally define an average curvature information matrix as a matrix which contains average, rather than local information about curvature. This matrix will then be defined even if the function is not two times continuously differentiable. Just as Hessian information can speed up derivative-based methods, average curvature information can speed up derivative-free methods [2].

Generating Set Search (GSS) methods [5] are a class of methods which only make use of function values. In a previously published paper [2] the authors presented a GSS method which gathers average curvature information through calculations of the form

$$(C_Q)_{ij} = \frac{f(x + hq_i + kq_j) - f(x + hq_i) - f(x + kq_j) + f(x)}{hk}, \tag{1}$$

where the q-vectors are orthogonal, but not necessarily the coordinate directions, and h and k are relatively large. $(C_Q)_{ij}$ denotes the ij-th entry of the $n \times n$ matrix C_Q. The information contained in this matrix is converted to the standard coordinate system by the formula

$$C = QC_QQ^T, \tag{2}$$

where Q is the matrix with q_i as its ith column, and C contains average curvature information. If the function is quadratic then C is the exact Hessian. Using the eigenvectors of C as search directions the method is often able to make rapid progress towards the optimal solution compared with a standard GSS method called compass search, of which it is an extension.

It is possible to enhance the method further by introducing the notion of partial separability for nondifferentiable functions.

2 Average curvature of nondifferentiable partially separable functions

It is well known [1, 4] that if a function has a sparse Hessian it is partially separable, and vice versa. We now give a similar result for nondifferentiable functions.

Theorem 1 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be deterministic and defined for all $x \in \mathbb{R}^n$, but not necessarily differentiable or continuous. If f can be written as a sum of k element functions f_k, where each f_k depends only on the indices of x in the set $\mathcal{I}_k$, $|\mathcal{I}_k| < n,\ k = 1, \ldots, r$, that is:*

$$f = \sum_{k=1}^{r} f_k,\ f_k : \mathbb{R}^{|\mathcal{I}_k|} \mapsto \mathbb{R},$$

---

[*] Department of Informatics, University of Bergen, Box 7803, N-5020 Bergen, Norway. E-mail: lennart.frimannslund@ii.uib.no

[†] Department of Informatics, University of Bergen.Box 7803, N-5020 Bergen, Norway. E-mail: trond.steihaug@ii.uib.no

then for all i, j *such that* $\{i, j\} \notin \mathcal{I}_k,\ k = 1, \dots, r$, *we have*

$$f(x + \delta_1 e_i + \delta_2 e_j) - f(x + \delta_1 e_i) - f(x + \delta_2 e_j) + f(x) = 0, \tag{3}$$

for all values of x, δ_1 *and* δ_2, *and* e_i *being the vector with a 1 in position* i *and zero everywhere else and similarly for* e_j. *The converse is also true, if (3) holds for some* (i, j) *and all values of* x, δ_1 *and* δ_2, *then the function is partially separable.*

Theorem 1 has implications for computing the matrix of average curvature used by the method of [2].

The theorem shows that if the curvature information matrix is sparse, then the function has a structure. However, in the case of smooth problems this structure can also be recovered using nonlinear interaction domains [6]. In the case of non differentiable functions nonlinear interaction domains must be reformulated, however the algorithm II in [6] may still be valid.

3 Application to the optimization method

If we define a *covariation graph* G of f to be a graph of n nodes with an edge form node i to j if and only if (3) holds, then the adjacency matrix $\mathcal{A}_G$ of this graph will in general be a sparse matrix, and we impose this structure onto C, the matrix of average curvature. If we rewrite (2) as

$$(Q^T \otimes Q^T)\mathbf{vec}(C) = \mathbf{vec}(C_Q), \tag{4}$$

where $\otimes$ is the Kronecker product and $\mathbf{vec}(\cdot)$ stacks the columns of a matrix in a vector, then we can easily enforce that some entries of the vector $\mathbf{vec}(C)$ are to be zero. Doing so results in a more effective optimization method on many functions, since forcing elements of $\mathbf{vec}(C)$ to be zero means that the matrix in (4) has more rows than columns and has full rank, and therefore the whole right-hand side of (4) needs not be computed. A consequence of this is that the optimization method is able to update its search directions more frequently than it would have otherwise, and often converges more rapidly than when separability is not taken into account. See [3] for details.

References

[1] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. An introduction to the structure of large scale nonlinear optimization problems and the LANCELOT project. In R. Glowinski and A. Lichnewsky, editors, *Computing Methods in Applied Sciences and Engineering*, pages 42–54. SIAM, Philadelphia, USA, 1990.

[2] L. Frimannslund and T. Steihaug. A generating set search method using curvature information. *Computational Optimization and Applications*, 38(1):105–121, 2007.

[3] L. Frimannslund and T. Steihaug. A new generating set search algorithm for separable functions. To be submitted, 2007.

[4] A. Griewank and Ph. L. Toint. On the unconstrained optimization of partially separable functions. In M. Powell, editor, *Nonlinear Optimization 1981*, pages 301–312. 1982.

[5] T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003.

[6] A.Walther Computing sparse Hessianswith automatic differentiation. *TOMS, 34(1):1–15, 2008.*