Applied Mathematics and Computation 218 (2011) 413–418

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Scaled memoryless symmetric rank one method for large-scale optimization ☆

Wah June Leong [a,*,1], Malik Abu Hassan [b]

[a] Institute for Mathematical Research, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[b] Department of Mathematics, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ARTICLE INFO

Keywords:
Large-scale optimization
Memoryless quasi-Newton method
Symmetric rank one update
Optimal scaling

ABSTRACT

This paper concerns the memoryless quasi-Newton method, that is precisely the quasi-Newton method for which the approximation to the inverse of Hessian, at each step, is updated from the identity matrix. Hence its search direction can be computed without the storage of matrices. In this paper, a scaled memoryless symmetric rank one (SR1) method for solving large-scale unconstrained optimization problems is developed. The basic idea is to incorporate the SR1 update within the framework of the memoryless quasi-Newton method. However, it is well-known that the SR1 update may not preserve positive definiteness even when updated from a positive definite matrix. Therefore we propose the memoryless SR1 method, which is updated from a positive scaled of the identity, where the scaling factor is derived in such a way that positive definiteness of the updating matrices are preserved and at the same time improves the condition of the scaled memoryless SR1 update. Under very mild conditions it is shown that, for strictly convex objective functions, the method is globally convergent with a linear rate of convergence. Numerical results show that the optimally scaled memoryless SR1 method is very encouraging.

1. Introduction

Large-scale unconstrained optimization is concerned with the numerical solution of the following problem:

$$\min f(x); \quad x \in R^n, \tag{1}$$

where $f : R^n \to R$ is continuously differentiable function, and n, the dimension of the problem is assumed to be large. Usually, problem (1) is solved iteratively through a line search scheme:

$$x_{k+1} = x_k + \lambda_k d_k, \tag{2}$$

where d_k is the search direction and $\lambda_k > 0$ is the step length. The step length can be computed by an exact line search:

$$\lambda_k^* = \text{argmin}_{\lambda \in \Re}\{f(x_k + \lambda d_k)\}, \tag{3}$$

or by some line search conditions, such as Wolfe [16] conditions:

☆ This work was partially supported by the Malaysian MOHE-FRGS Grant No. 01-11-09-722FR.
* Corresponding author.
E-mail address: leong@math.upm.edu.my (W.J. Leong).
1 This work was done while the first author was visiting Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, PR China. During the visit, the author was supported by the CAS-TWAS Fellowship no. 3240157252.

0096-3003/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2011.05.080

$$f(x_k + \lambda_k d_k) \leqslant f(x_k) + \beta_1 \lambda_k g_k^T d_k, \tag{4}$$

$$g_{k+1}^T d_k \geqslant \beta_2 g_k^T d_k, \tag{5}$$

where $0 < \beta_1 < 1/2$, $\beta_1 < \beta_2 < 1$ and $g_k = \triangledown f(x_k)$ denotes the gradient vector of $f(x)$ at the current iteration point x_k.

In this paper, we are particularly interested in deploying methods for solving very large-scale cases, where the dimensions of the problems are up to 10^6. The need to solve these extremely large-scale optimization problems forces one to consider methods of $O(n)$ storage as the only methods of choice. This class of methods, includes those as the steepest descent method, conjugate gradient methods, limited memory quasi-Newton method and memoryless quasi-Newton method.

Memoryless quasi-Newton methods or one step quasi-Newton methods were first considered by Perry [13] and Shanno [14]. They are actually the quasi-Newton method for which at each iteration, a periodically restarted quasi-Newton correction is calculated from the initial approximation, commonly given by the identity matrix. Hence the memoryless quasi-Newton directions can be computed without the storage of matrices, namely $O(n^2)$ storages. Among the well-studied memoryless quasi-Newton methods is the memoryless BFGS method, which uses the BFGS update:

$$H_{k+1} = \left(I - \frac{y_k^T s_k}{s_k^T y_k}\right) H_k \left(I - \frac{y_k^T s_k}{s_k^T y_k}\right) + \frac{s_k s_k^T}{s_k^T y_k}, \tag{6}$$

where $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$. In fact, a result by Shanno [14] (see also [8]) shows that traditional CG methods such as the Fletcher–Reeves and Polak-Ribiére algorithm can be interpreted as a memoryless BFGS algorithm. This memoryless BFGS algorithm may then be scaled optimally by the scaling of Oren and Spedicato [12]. Besides the BFGS update, one can extend the idea of memoryless updating to SR1 update:

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{y_k^T (s_k - H_k y_k)} \tag{7}$$

and get the memoryless SR1 method. Minimization algorithms using SR1 update in both a line search and trust region context have been shown in computational experiments by Conn et al. [2] and Khalfan et al. [6] to be competitive with methods using the widely accepted BFGS update. Hence, it is reasonable to think that such encouraging results can be extended to the memoryless version of SR1 method as well. However, it is well-known that the SR1 update may not preserve positive definiteness even when updated from a positive definite matrix. Therefore, to overcome this drawback, we propose a scaled memoryless SR1 method, which uses a periodically restarted SR1 correction from a positive scaled identity matrix. The scaling factor is derived in such a way the positive definiteness of the updated SR1 matrix can be preserved naturally and the condition of the SR1 update is also improved.

This paper is organized as follows: in Section 2, we discuss the optimal scaling factor for the identity matrix. Section 3 gives the convergence result of the scaled memoryless SR1 method for a convex minimization. Finally we include some numerical experiments on a standard set of test problems in Section 4.

2. Scaling the identity matrix

Throughout this section, we will use the following notations:

$$\eta_k = y_k^T H_k y_k, \quad \upsilon_k = y_k^T s_k, \quad \text{and } \omega_k = s_k^T B_k s_k, \tag{8}$$

where B_k is the current Hessian approximation using the direct SR1 update:

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{s_k^T (y_k - B_k s_k)} \tag{9}$$

and $H_k = B_k^{-1}$. When we mention inverse SR1 update, we mean the updating formula (7), otherwise the direct SR1 update is given by (9). We assume that the curvature condition $\upsilon_k = y_k^T s_k > 0$ and B_k (or H_k) is positive definite.

Our primary motivation here is to find the best SR1 formula updated from current approximation, i.e. this update should satisfy the secant equation while preserving positive definiteness and as much information from the current update as possible. Because it is difficult to find the optimal scaling factor for SR1 update in l_2-norm condition number, one may consider obtaining it in some other measures. With this aim in mind, we first consider the following measure, which is suggested by Dennis and Wolkowicz [3]:

$$\sigma(A) = \frac{\xi_{\max}}{\det(A)^{1/n}}, \tag{10}$$

where A is an $n \times n$ positive definite matrix and $\xi_{\max}$ is the largest eigenvalue of A Here, the measure σ acts as a condition number in that it provides a deviation from a multiple identity as does the l_2-conditioned number. In fact, both Dennis and Wolkowicz [3] and Wolkowicz [17] had shown that any σ-optimal update will also be κ-optimal ($\kappa(A)$ denotes the l_2-condition number of A) and have a common spectral property. The κ-measure is used by Shanno and Phua [15] to derive the optimal scaling factor for the BFGS update.

We now give the following theorem which is due to Wolkowicz [17]: