

A Nonmonotone ADMM-Based Diagonal Quasi-Newton Update with Application to the Compressive Sensing Problem

Zohre Aminifard^a and Saman Babaie-Kafaki^b

^a*Faculty of Mathematics, Statistics and Computer Science, Semnan University*

Semnan Province, 35131–19111 Semnan, Iran

^b*Faculty of Engineering, Free University of Bozen-Bolzano*

Piazza Università 5, 39100 Bolzano, Italy

E-mail: aminisor@semnan.ac.ir

E-mail(*corresp.*): saman.babaiekafaki@unibz.it

Received May 17, 2022; accepted October 3, 2023

Abstract. Considering a minimization problem according to the Byrd-Nocedal measure function together with the secant equation, a diagonal quasi-Newton updating formula is suggested. To find the optimal elements of the updating matrix, the well-known algorithm of the alternating direction method of multipliers (ADMM) is employed. Moreover, convergence analysis is conducted based on a modified nonmonotone Armijo line search incorporating the simulated annealing strategy. Lastly, performance of the method is numerically tested on a set of CUTER functions and on a smooth transcendental approximation of the compressive sensing problem. Across the computational spectrum, the given method turns out to be successful.

Keywords: large-scale optimization, quasi-Newton update, ADMM strategy, nonmonotone line search, simulated annealing, compressive sensing, smoothing technique.

AMS Subject Classification: 90C53; 90C59; 94A08.

1 Introduction

Quasi-Newton (QN) algorithms demonstrate potential competence for solving unconstrained optimization problems. So, enriching their efficiency has triggered the interest of optimization scholars. In particular, QN methods are

efficient because under standard circumstances, the algorithms generate descent search directions and have reasonable global as well as local superlinear convergence features [37].

In the QN methods, starting with some positive definite matrices, successive estimations of the (inverse) Hessian are updated to guarantee the secant (QN) equation [37]. To make the generated matrix approximations well-conditioned, the scaled QN updates have been developed based on the eigenvalue analyses [33,34]. Moreover, to adjust the methods for large-scale problems, memoryless QN techniques have been proposed with significantly reduced calculations [8].

Recently, the QN methods have been much heeded in practical applications such as image processing, time series prediction, neural networks training, document categorization, managing demands in the water distribution networks, machine learning, robotics, solving systems of nonlinear equations, curve fitting by B-splines, matrix approximation in Frobenius norm, computation of the matrix geometric mean and estimating unitary symmetric eigenvalues of the complex tensors; for more details see [11] and references therein. The methods have also been well-combined with classical optimization tools such as conjugate gradient methods [12] as well as metaheuristic algorithms [32].

Real-world applications of the QN algorithms motivated the researchers to ardently improve efficiency of the methods. Such attempts can be mainly categorized into modifying the secant equation to approximate the curvature more accurately [36] or to achieve convergence without convexity assumptions [30], improving the scaling scheme [10], and structuring the updating formulas to solve special problems such as nonlinear least squares models [1].

As an underpinning improvement devised to address big data models, researchers were eager to find diagonal matrices with positive diagonal elements to approximate the Hessian. Such efforts were principally begun by Zhu et al. [40] via considering various versions of the secant equation. Hassan et al. [27] developed another diagonal matrix based on the Barzilai-Borwein approach [14]. Leong et al. [29] advanced the issue by giving some effective diagonal QN matrices that preserve the positive definiteness. In another attempt, Andrei [6] proposed a diagonal QN update by minimizing the measure function of Byrd and Nocedal [17], based on forward and central finite differences [7]. A family of diagonal QN updates was suggested in [3] in accordance to the DFP (Davidon-Fletcher-Powell) and the BFGS (Broyden-Fletcher-Goldfarb-Shanno) updating formulas. A tridiagonal Hessian approximation has also been developed in [9].

Here, in an attempt to deliver progress in the diagonal Hessian approximations' scope, we mainly focus on the following issues in the given order:

1. Finding a diagonal QN updating formula is becoming pervasive, due to tying to the big data minimization problems. By the same token, Andrei [6] presented a minimization subproblem taking the weak secant condition into account to find elements of the Hessian matrix approximation. However, the weak secant equation may be far from the original secant equation. So, in a direct response to this issue, the least squares approach is employed to find an appropriate minimization model of the diagonal approximation of the Hessian matrix.

2. For solving the given subproblem and consequently, finding the optimal elements of the Hessian estimation effectively, an ADMM approach is applied.
3. To establish the global convergence of the proposed QN method without convexity assumption and what's more to upgrade the performance of the line search strategy, an improved version of the nonmonotone line search of [39] is developed. To do so, we incorporate it by a simulated annealing technique.
4. We evaluate performance of the proposed method on the CUTER test functions and the compressive sensing problem for which a smooth approximation model is rendered as well.

Mainstream theme of Section 2 is comprised of dealing with a diagonal QN updating formula by proposing an efficient minimization problem and solving it with the ADMM technique. Analyzing the global convergence according to a modification of the nonmonotone line search of [39] incorporated by the simulated annealing (SA) strategy is carried out in Section 3. Conducting practical tests is performed in Section 4.

2 A diagonal ADMM-based quasi-Newton update

QN iterations for finding the solution of the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{2.1}$$

are generally generated via

$$x_{k+1} = x_k + s_k, \quad k = 0, 1, \dots, \tag{2.2}$$

initiated by some $x_0 \in \mathbb{R}^n$, with $s_k = \alpha_k d_k$ where d_k is a descent direction and the step size $\alpha_k > 0$ is determined by a line search along d_k [35]. In QN methods, we have

$$d_0 = -g_0, \quad d_k = -B_k^{-1}g_k, \quad k = 1, 2, \dots, \tag{2.3}$$

where $g_k = \nabla f(x_k)$ and $B_k \in \mathbb{R}^{n \times n}$ is a positive definite approximation of $\nabla^2 f(x_k)$. Especially, in diagonal matrix estimation of the Hessian, it needs to find appropriate diagonal components $\mathcal{B}_{k_i}, i = 1, \dots, n$, to obtain B_k as follows: $B_k = \text{diag}(\mathcal{B}_{k_1}, \mathcal{B}_{k_2}, \dots, \mathcal{B}_{k_n})$.

Associated with the measurement of the well-conditioning of a positive definite matrix $A \in \mathbb{R}^{n \times n}$, Byrd and Nocedal [17] suggested the following function:

$$\psi(A) = \text{tr}(A) - \ln(\det(A)), \tag{2.4}$$

which has been essentially employed in the convergence analysis of the QN algorithms. Taking (2.4) into account, Andrei [6] dealt with the following minimization problem:

$$\min_{B_k \in \mathbb{D}} \psi(B_k), \quad \text{s.t.} \quad s_k^T B_k s_k = s_k^T y_k, \tag{2.5}$$

in which \mathbb{D} is the set of all diagonal $n \times n$ matrices. The constraint of (2.5) is called the weak secant equation [20], being a scalar (relaxed) version of the standard secant equation $B_k s_k = y_k$. The model (2.5) yields the following formula for the elements of B_k :

$$B_{k_i} = 1/(1 + \lambda s_{k_i}^2), \quad i = 1, \dots, n,$$

where λ is the Lagrange multiplier computed from an extended conjugacy condition [19] as follows:

$$\lambda = \begin{cases} r + \varepsilon, & \hat{\lambda} < r, \\ \hat{\lambda}, & \text{otherwise,} \end{cases}$$

in which $r = \max_{i=1, \dots, n, s_{k_i}^2 \neq 0} \{-1/s_{k_i}^2\}$, ε is a small positive constant and also, $\hat{\lambda} = (t s_k^T g_{k+1} - y_k^T g_{k+1}) / \sum_{i=1}^n y_{k_i} g_{k+1, i} s_{k_i}^2$, where $t > 0$ is the Dai-Liao parameter [19]. In another attempt, Babaie-Kafaki et al. [11] suggested the following penalized version of (2.5):

$$\min_{B_k \in \mathbb{D}} \psi(B_k) + \|B_k s_k - y_k\|^2, \tag{2.6}$$

where $\|\cdot\|$ stands for the ℓ_2 norm. By this model, doing some arithmetic we obtain

$$B_{k_i} = \frac{2s_{k_i} y_{k_i} - 1 + \sqrt{(2s_{k_i} y_{k_i} - 1)^2 + 8s_{k_i}^2}}{4s_{k_i}^2}, \quad i = 1, \dots, n,$$

which are positive and well-defined for $s_{k_i} \neq 0$.

As known, ADMM has been proven to be a powerful approach to address a wide range of practical fields, particularly machine learning and signal processing models [38]. Stimulated by merits of the ADMM approach, we intend to broaden the model (2.6) for finding an appropriate diagonal approximation for the Hessian of the cost function, in the sense of

$$\min_{B, C \in \mathbb{D}} \psi(B) + \|C s_k - y_k\|^2, \quad s.t. \ B = C,$$

for which the augmented Lagrangian function [35] can be rendered as

$$\mathcal{L}(B, C, \tau, \mu) = \psi(B) + \|C s_k - y_k\|^2 - \tau^T (B - C) + \frac{\mu}{2} \|B - C\|^2, \tag{2.7}$$

where $\tau \in \mathbb{R}^n$ is the vector of Lagrange multipliers (dual variables), B and C respectively denote the vectors which contain the diagonal elements of B and C , and $\mu_k > 0$ is a constant. The ADMM-based technique finds the solution of (2.7) by executing the next process in each iteration:

$$\tau_{k+1} = \tau_k - \mu_k (B_k - C_k), \quad B_{k+1} = \arg \min_{B \in \mathbb{D}} \mathcal{L}(B, C_k, \tau_{k+1}), \tag{2.8}$$

$$C_{k+1} = \arg \min_{C \in \mathbb{D}} \mathcal{L}(B_{k+1}, C, \tau_{k+1}), \quad \mu_{k+1} = \rho \mu_k, \tag{2.9}$$

where $\rho \geq 1$ is a constant.

To give some details for the solution process of (2.8) and (2.9), we have in view the following componentwise definition of (2.7):

$$\begin{aligned} \mathcal{L}(B, C, \tau, \mu) &= \sum_{i=1}^n \mathcal{B}_{k+1,i} - \ln \left(\prod_{i=1}^n \mathcal{B}_{k+1,i} \right) + \sum_{i=1}^n (\mathcal{C}_{k+1,i} s_{k,i} - y_{k,i})^2 \\ &\quad - \sum_{i=1}^n \tau_{k+1,i} (\mathcal{B}_{k+1,i} - \mathcal{C}_{k+1,i}) + \frac{\mu}{2} \sum_{i=1}^n (\mathcal{B}_{k+1,i} - \mathcal{C}_{k+1,i})^2. \end{aligned}$$

Therefore, answer of the mentioned problems can be obtained exactly by solving the equations $\frac{\partial \mathcal{L}}{\partial \mathcal{B}_i} = 0$ and $\frac{\partial \mathcal{L}}{\partial \mathcal{C}_i} = 0$, for $i = 1, \dots, n$, yielding

$$\begin{aligned} \mathcal{B}_{k+1,i} &= \frac{\mu_k \mathcal{C}_{k,i} + \tau_{k+1,i} - 1 + \sqrt{(\mu_k \mathcal{C}_{k,i} + \tau_{k+1,i} - 1)^2 + 4\mu_k}}{2\mu_k}, \tag{2.10} \\ \mathcal{C}_{k+1,i} &= (2s_{k,i} y_{k,i} + \mu_k \mathcal{B}_{k+1,i} - \tau_{k+1,i}) / (2s_{k,i}^2 + \mu_k). \end{aligned}$$

Note that for arbitrary constants a and $b \neq 0$, we have $a < \sqrt{a^2 + b^2}$. Thus, if we let $a = \mu_k \mathcal{C}_{k,i} + \tau_{k+1,i} - 1$ and $b = \sqrt{4\mu_k}$, then from (2.10) we have $\mathcal{B}_{k+1,i} > 0$ which ensures the matrix B_k is positive semidefinite and consequently, the search direction (2.3) is nonincreasing.

3 A nonmonotone line search technique with the simulated annealing strategy

Imposing monotone reduction on the successive function values in the classical iterative schemes for solving (2.1) may cause losing the efficiency when an iteration confines close to the bottom of a narrow curved valley of the cost function. Actually, a monotone scheme is forced to creep along the valley’s floor, making too short steps or even undesired zigzagging trajectories [39]. As a remedy, scholars put their best efforts into developing nonmonotone schemes which guarantee the global convergence. One of the pioneering nonmonotone schemes was proposed by Zhang and Hager [39], which specifies the smallest integer $h \geq 0$ fulfilling

$$f(x_k + \sigma^h a d_k) \leq C_k + \gamma (\sigma^h a) g_k^T d_k, \tag{3.1}$$

where $\gamma \in (0, 1)$ is the Armijo constant and

$$C_{k+1} = (\eta_k q_k C_k + f_{k+1}) / q_{k+1},$$

in which $f_{k+1} = f(x_{k+1})$, the parameter $\eta_k \in [\eta_{\min}, \eta_{\max}]$ with the constants $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$ handles the nonmonotonicity measure and

$$q_0 = 1, \quad q_{k+1} = \eta_k q_k + 1, \quad \forall k \geq 0. \tag{3.2}$$

In a backtracking scheme, the step size α_k is the largest member of $\{\sigma^j a\}_{j \geq 0}$ with $a > 0$ and $\sigma \in (0, 1)$.

As a well-known classic metaheuristic algorithm, SA is a probabilistic technique for approximating the global optimum of an optimization model [15],

being efficient practically and well-developed theoretically. It originates from the impact of effective control of the temperature in the annealing procedure of a substance. With a view of improving performance of the nonmonotone scheme, here we incorporate the line search strategy (3.1) with the SA algorithm. As a matter of fact, taking inspiration from SA, the line search accepts the iterate x_{k+1} with some probability specified by the temperature parameter even when it does not guarantee the condition (3.1). The probability of accepting the iterates is gradually decreased by decreasing the temperature during execution to enhance the exact exploration capabilities of the algorithm near the optimal solution. To provide a detailed description of the nonmonotone Armijo line search with the SA technique, the iterate $x_{k+1} = x_k + \sigma^h ad_k$ is accepted when σ^h fulfills (3.1) or $p_k \geq r_k$, in which

$$r_k \in (e^{-\vartheta}, e^{-\frac{1}{\vartheta}}), \tag{3.3}$$

for an integer $\vartheta \geq 1$, and $p_k = e^{-\Delta_k/T_k}$, where $\Delta_k = f(x_k + \sigma^h ad_k) - \mathfrak{F}_k - \gamma(\sigma^h a)g_k^T d_k$, and $T_k > 0$ is the temperature parameter. Hence, the incorporative line search can be presented as follows:

$$f(x_k + \sigma^h ad_k) \leq \mathfrak{F}_k + \gamma\sigma^h ag_k^T d_k - T_k \ln r_k, \tag{3.4}$$

in which $T_k = \theta^k T_0$ with $\theta \in (0, 1)$, and T_0 is the initial temperature, and also,

$$\mathfrak{F}_{k+1} = \frac{\eta_k q_k (\mathfrak{F}_k - T_k \ln r_k) + f_{k+1}}{q_{k+1}},$$

with $\mathfrak{F}_0 = f(x_0)$ and the parameter q_k is defined by (3.2). Now, we are in a position to spell out our algorithm.

Algorithm 1. The nonmonotone ADMM-based diagonal quasi-Newton algorithm (DQNADMM)

Step 0: Choose an initial point $x_0 \in \mathbb{R}^n$, the line search parameters $T_0 > 0$, $\vartheta > 0$, $a > 0$, $\gamma, \theta, \sigma \in (0, 1)$, and the tolerance $\epsilon > 0$. Set $d_0 = -g_0$ and $k = 0$.

Step 1: If $\|g_k\| < \epsilon$, **then stop**.

Step 2: Update the QN matrix B_k by (2.10).

Step 2: Compute the search direction d_k by (2.3).

Step 3: Compute the step length α_k satisfying (3.4).

Step 4: Set $x_{k+1} = x_k + \alpha_k d_k$.

Step 5: Set $k = k + 1$ and **goto** Step 1.

To establish global convergence of the iterative method (2.2) with the backtracking line search satisfying (3.4), hereafter, we suppose that the following assumption holds.

ASSUMPTION 1. (i) The cost function f is coercive. (ii) The gradient of f is Lipschitz continuous; that is, there exists a positive constant L such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

(iii) The search direction d_k satisfies the following conditions for some positive constants c_1 and c_2 :

$$d_k^T g_k \leq -c_1 \|g_k\|^2, \tag{3.5}$$

$$\|d_k\| \leq c_2 \|g_k\|. \tag{3.6}$$

Lemma 1. *The backtracking line search strategy based on (3.4) is well-defined.*

Proof. Firstly, we define $A_{k+1} : \mathbb{R}^+ \rightarrow \mathbb{R}$, given by

$$A_{k+1}(t) = (t(\mathfrak{F}_k - T_k \ln r_k) + f_{k+1}) / (t + 1),$$

which yields

$$A'_{k+1}(t) = (\mathfrak{F}_k - T_k \ln r_k - f_{k+1}) / (t + 1)^2.$$

Since, from (3.4) and (3.5), $f_{k+1} \leq \mathfrak{F}_k - T_k \ln r_k$, we get $A'_{k+1} \geq 0$ for all $t \neq -1$. Therefore, since $A_{k+1}(t)$ is nondecreasing, we have $f_{k+1} = A_{k+1}(0) \leq A_{k+1}(t)$, for all $t \geq 0$. Particularly, by taking $t = \eta_k q_k$ into account,

$$f_{k+1} = A_{k+1}(0) \leq A_{k+1}(\eta_k q_k) = \mathfrak{F}_{k+1}. \tag{3.7}$$

By contrary, assume that for all j , thanks to $T_k \ln r_k \leq 0$,

$$f(x_k + \sigma^j ad_k) > \mathfrak{F}_k + a\sigma^j \delta g_k^T d_k - T_k \ln r_k \geq f_k + a\sigma^j \delta g_k^T d_k.$$

Therefore,

$$\frac{f(x_k + \sigma^j ad_k) - f_k}{a\sigma^j} > \delta g_k^T d_k.$$

Hence, there exists $\varpi \in (0, 1)$ so that

$$g(x_k + \varpi \sigma^j ad_k)^T d_k > \delta g_k^T d_k.$$

Now, when k tends to infinity, $(1 - \delta)g_k^T d_k > 0$, which violates (3.5). In the following, we find the constant $\underline{\alpha} > 0$ such that $\alpha_k \geq \underline{\alpha}$, for all $k \geq 0$.

Exploiting the Cauchy-Schwarz inequality and the mean value theorem, we have

$$\begin{aligned} f(x_k + \alpha d_k) &= f_k + \alpha g_k^T d_k + \int_0^1 \alpha (g(x_k + t\alpha d_k) - g_k)^T d_k dt \leq f_k + \alpha g_k^T d_k \\ &+ \alpha \|d_k\| \int_0^1 \|g(x_k + t\alpha d_k) - g_k\| dt \leq \mathfrak{F}_k + \alpha g_k^T d_k + \frac{L}{2} \alpha^2 \|d_k\|^2. \end{aligned} \tag{3.8}$$

By setting $\alpha = \frac{\alpha_k}{\sigma}$ in (3.8) and applying the fact that

$$f\left(x_k + \frac{\alpha_k}{\sigma} d_k\right) > \mathfrak{F}_k + \frac{\alpha_k}{\sigma} \delta g_k^T d_k,$$

from (3.5) and (3.6) we get

$$(\delta - 1)g_k^T d_k < \frac{L}{2} \frac{\alpha_k}{\sigma} \|d_k\|^2 \leq \frac{L}{2} \frac{\alpha_k}{\sigma} \left(-\frac{c_2^2}{c_1} \right) g_k^T d_k.$$

Therefore,

$$\alpha_k > \frac{2\sigma c_1}{Lc_2^2} (1 - \delta) := \underline{\alpha},$$

which completes the proof. \square

In light of the convergence analysis conducted in [22, 39], now we should deal with the following theorem.

Theorem 1. *For the sequence $\{x_k\}_{k \geq 0}$ defined by (2.2) with the backtracking line search (3.4), $\{x_k\}_{k \geq 0} \subset \Omega_0 = \{x \in \mathbb{R}^n : f(x) \leq f(x_0) + \frac{T_0 \vartheta}{1 - \theta}\}$ and for every accumulation point \bar{x} of $\{x_k\}_{k \geq 0}$, we have $\nabla f(\bar{x}) = 0$.*

Proof. Considering Theorem 2 of [5], the continuity of f yields the closedness and consequently compactness of Ω_0 . From (3.4), we have

$$\begin{aligned} \mathfrak{F}_{k+1} &= \frac{\eta_k q_k}{q_{k+1}} (\mathfrak{F}_k - T_k \ln r_k) + \frac{1}{q_{k+1}} f_{k+1} \leq \frac{\eta_k q_k}{q_{k+1}} (\mathfrak{F}_k - T_k \ln r_k) \\ &+ \frac{1}{q_{k+1}} (\mathfrak{F}_k + \gamma \alpha_k g_k^T d_k - T_k \ln r_k) = \mathfrak{F}_k + \frac{\gamma \alpha_k}{q_{k+1}} g_k^T d_k - T_k \ln r_k. \end{aligned} \tag{3.9}$$

Now, since $\mathfrak{F}_0 = f(x_0)$, and due to (3.3), we get

$$f_{k+1} \leq \mathfrak{F}_k - T_k \ln r_k \leq \mathfrak{F}_0 - \sum_{i=0}^k T_i \ln r_i \leq f(x_0) + \sum_{i=0}^k T_0 \theta^i \vartheta \leq f(x_0) + \frac{\vartheta T_0}{1 - \theta},$$

which leads to $\{x_k\}_{k \geq 0} \subset \Omega_0$. Exploiting (3.7) together with continuity of f leads to $\mathfrak{F}_k \geq \beta$, for some constant β . Alternatively, from (3.9),

$$-\frac{\gamma \alpha_k}{q_{k+1}} g_k^T d_k \leq \mathfrak{F}_k - \mathfrak{F}_{k+1} - T_k \ln r_k. \tag{3.10}$$

Now, using Assumption 1(iii) and summing both sides of (3.10) from $k = 0$ to $K - 1$, we get

$$c_1 \gamma \sum_{k=0}^K \frac{\alpha_k}{q_{k+1}} \|g_k\|^2 \leq \mathfrak{F}_0 - \mathfrak{F}_K - \sum_{k=0}^K T_k \ln r_k \leq \mathfrak{F}_0 - \mathfrak{F}_K + \frac{\vartheta T_0}{1 - \theta}. \tag{3.11}$$

On the other hand, according to (3.4) and since $\eta_{\max} < 1$,

$$q_{k+1} = 1 + \sum_{j=0}^k \prod_{i=0}^j \eta_{k-i} \leq 1 + \sum_{j=0}^k \eta_{\max}^{j+1} \leq \sum_{j=0}^{\infty} \eta_{\max}^j = \frac{1}{1 - \eta_{\max}},$$

which coupled with (3.11),

$$c_1 \gamma (1 - \eta_{\max}) \sum_{k=0}^K \alpha_k \|g_k\|^2 \leq \mathfrak{F}_0 - \mathfrak{F}_K + \frac{\vartheta T_0}{1 - \theta}.$$

Now, when K tends to infinity, from (3.6),

$$\lim_{k \rightarrow \infty} \alpha_k \|g_k\|^2 = 0.$$

Besides, using Lemma 1, $\lim_{k \rightarrow \infty} \|g_k\| = 0$, and so, the proof is complete. \square

In account of supposition (iii) of Assumption 1, \mathcal{B}_{k_i} ($i = 1, 2, \dots, n$) given by (2.10) must be uniformly bounded. In order to ensure boundedness of \mathcal{B}_{k_i} , we can set $\mathcal{B}_{k_i} \leftarrow \min \{ \max \{ \mathcal{B}_{k_i}, \eta \}, 1/\eta \}$, for all $i = 1, 2, \dots, n$, where η is an enough small positive constant.

4 Numerical experiments

To uphold the theoretical results, here we compare performance of the DQ-NADMM algorithm with the two well-known QN methods, namely, the modified memoryless BFGS (MMLBFGS) method and the modified limited memory BFGS (MLMBFGS) method, both devised based on the modified secant condition proposed in [30]. It can be observed that MMLBFGS and MLMBFGS generate descent directions regardless of the line search. Furthermore, we consider the diagonal QN methods proposed in [6], [11] and [26] in our comparisons, here respectively named by DQNB1, DQNB2 and DQNMSG.

For DQNADMM, we set $\rho = 10$ and $\mu_0 = 1$ in (2.9). The scaling parameter of MSMLBFGS and MSLMBFGS was tuned by

$$v_k = \min \{ 10^{10}, \max \{ 10^{-10}, \|s_k\|/\|y_k\| \} \},$$

as suggested in [33]. Information of the hardware and software applied in the implementations have been provided in [2]. Besides, line search has been performed employing nonmonotone backtracking Armijo condition (3.4) using $a = 1$, $\gamma = 10^{-4}$, $\sigma = 0.85$, $T_0 = \|g_k\|$ and $\vartheta = 2$. The algorithms were ended up to an iterate satisfying $k > 10000$ or $\|g_k\| < 10^{-5}(1 + |f_k|)$. To compare efficiency of the algorithms, the CPU time (CPUT) and the total number of function and gradient evaluations (TNF), introduced in [25], have been assessed using the Dolan-Moré (DM) technique [21], following the notations of [2]. The test problems data including 147 functions of the CUTER library [23] is provided in Table 1.

The results of comparisons are illustrated by Figure 1 upon which DQ-NADMM decidedly outperforms the others with respect to the running time, while considering TNF into account, DQNADMM and DQNMSG are competitive and both of them are preferable to the other algorithms.

To assess the efficiency of the given nonmonotone line search technique, we investigated performance of DQNADMM together with the nonmonotone line search technique (3.4) and the nonmonotone approaches proposed in [39] and [24]. The corresponding methods are here respectively called DQNADMM-NM i , for $i = 1, 2, 3$. The results have been shown in Figure 2. According to Figure 2(a), DQNADMM-NM1 is preferable to the others with respect to TNF. Nevertheless, from Figure 2(b), it can be concluded that performing extra

Table 1. Test problems data

Function	n	Function	n	Function	n
AKIVA	2	DIXMAANH	3000	HUMPS	2
ALLINIT	4	DIXMAANK	3000	JENSMP	2
ALLINITU	4	DIXMAANL	3000	KOWOSB	4
ARGLINA	200	DIXMAANM	15	LOGHAIRY	2
ARGLINB	200	DIXMAANN	15	MANCINO	100
BARD	3	DIXON3DQ	10000	MARATOSB	2
BDEXP	5000	DJTTL	2	MEXHAT	2
BDQRTIC	5000	DJTTL	2	MOREBV	5000
BEALE	2	DMN15102	66	MSQRTALS	1024
BENNETT5	3	DMN15103	99	MSQRTBLS	1024
BIGGS3	6	DMN37142	66	NONCVXU2	5000
BIGGS5	6	DMN37143	99	NONDIA	5000
BIGGS6	6	DQRTIC	5000	NONDQUAR	5000
BIGGSB1	5000	DRCAV1LQ	4489	OSBORNEB	11
BOX	10000	DRCAV2LQ	4489	OSCIPTH	10
BOX3	3	DRCAV3LQ	4489	PALMER1D	7
BQP1VAR	1	ECKERLE4	3	PALMER2C	8
BQPGABIM	50	EG2	1000	PALMER3C	8
BQPGASIM	50	EIGENALS	2550	PALMER4C	8
BQPGAUSS	2003	EIGENCLS	2652	PALMER6C	8
BRATU1D	5003	ENGVAL1	5000	PALMER7C	8
BRKMCC	2	ENGVAL2	3	PALMER8C	8
BROWNAL	200	ERRINROS	50	PARKCH	15
BROWNBS	2	EXPFIT	2	PENALTY1	1000
BROWNDEN	4	EXTROSNB	1000	PENALTY2	200
BROYDN7D	5000	FLETCBV2	5000	PENALTY3	200
CHAINWOO	4000	FLETCBV3	5000	POWER	10000
CHENHARK	5000	FLETCHBV	5000	QUARTC	5000
CHNROSNB	50	FMINSRF2	5625	S308	2
CHWIRUT2	3	FMINSURF	5625	SINEVAL	2
CLIFF	2	GAUSS1LS	8	SINQUAD	5000
CLPLATEB	5041	GAUSS3LS	8	SISSER	2
CRAGGLVY	5000	GENHUMPS	5000	SNAIL	2
CUBE	2	GENROSE	500	SPARSINE	5000
CURLY10	10000	GROWTHLS	3	SPARSQR	10000
CURLY20	10000	GULF	3	TESTQUAD	5000
CURLY30	10000	HAIRY	2	TOINTGOR	50
DECONVU	63	HATFLDD	3	TOINTGSS	5000
DENSCHNA	2	HATFLDE	3	TOINTPSP	50
DENSCHNB	2	HATFLDFL	3	TOINTQOR	50
DENSCHNC	2	HEART6LS	6	TQUARTIC	5000
DENSCHND	3	HEART8LS	8	TRIDIA	5000
DENSCHNE	3	HELIX	3	VARDIM	200
DENSCHNF	2	HILBERTA	2	VAREIGVL	50
DIXMAANA	3000	HILBERTB	10	VIBRBEAM	8
DIXMAANB	3000	HIMMELBB	2	WATSON	12
DIXMAANC	3000	HIMMELBF	4	WOODS	4000
DIXMAAND	3000	HIMMELBG	2	YFITU	3
DIXMAANG	3000	HIMMELBH	2	ZANGWIL2	2

computations to evaluate simulated annealing condition (to be used in (3.4)) is time-consuming. As a result, DQNADMM-NM1 is slightly competitive with the other algorithms in the perspective of CPU.

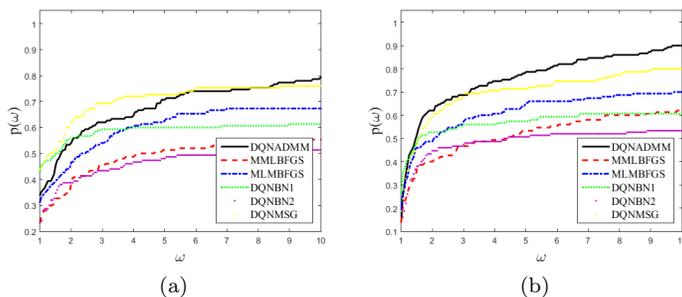


Figure 1. DM performance profile outputs for TNF (a) and CPUT (b) of DQNADMM, MMLBFGS, MLMBFGS, DQNB1, DQNB2 and DQNMSG.

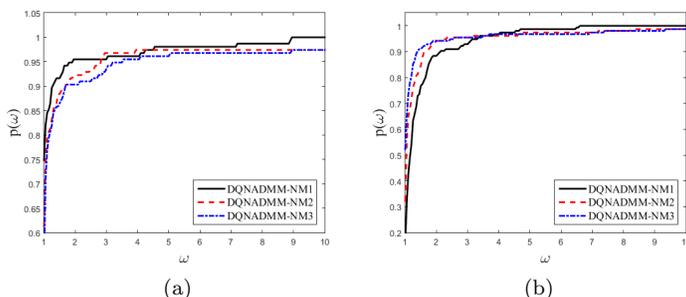


Figure 2. DM performance profile outputs for TNF (a) and CPUT (b) of DQNADMM-NM1, DQNADMM-NM2 and DQNADMM-NM3.

We also render an application of the presented method for the compressive sensing (CS) problem as a practical issue. The CS problem is a rapidly growing field that has picked up considerable attention in a broad-ranging of scientific areas. CS addresses a framework for simultaneous sensing and compression of finite-dimensional vectors. To be more specific, CS exploits the sparse signal by solving an underdetermined linear system in the following unified formulation:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|\psi x - y\|_2^2 + \varsigma \|x\|_1, \tag{4.1}$$

with the sampling matrix $\psi \in \mathbb{R}^{m \times n}$ ($m \ll n$), and the measurement vector $y \in \mathbb{R}^m$ [16], where $\|\cdot\|_1$ represents the ℓ_1 norm and also, the regularization parameter $\varsigma > 0$ keeps the balance between the sparsity and the measurements coupled with the additive white Gaussian noise. Smoothing strategies, as a remedy to combat nonsmoothness of the ℓ_1 regularizer, are one of the methods for solving problem (4.1) [28, 31].

Smooth approximations for optimization problems have been scrutinized for decades including complementarity problems, variational inequalities, second-order cone complementarity problems, semidefinite programming, semi-infinite programming, optimal control, eigenvalue optimization, penalty methods and

mathematical programs with equilibrium constraints; see [18] and references therein, for instance. The smoothing techniques are advantageous thanks to the rich theory and powerful methods provided for continuously differentiable cost functions, and the guarantee for finding a local minimizer or stationary point.

Recently, Bagul [13] proposed the following smooth transcendental approximation of $|x|$:

$$\varphi(x) = x \tanh(x/\nu), \quad (4.2)$$

where the small constant $\nu > 0$ is called smoothing parameter. As established in Theorem 1 of [13], $\|x| - \varphi(x)| < \nu$. Hence, the CS cost function can be approximated by

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|\psi x - y\|_2^2 + \varsigma \sum_{i=1}^n \varphi(x_i),$$

which is (smooth and) solvable via the given algorithm.

We generated the random test functions by choosing the signal dimension equal to $n = 2^{13}$ and the sampling matrix ψ as the Hadamard matrix, utilizing the approach of [4]. The initial point has been set as $0 \in \mathbb{R}^n$. We also set $\varsigma = \max(0.001 \|\psi^T y\|_\infty, 2^{-8})$ in (4.1) and $\nu = 0.001$ in (4.2). To make a meaningful comparison on the performance of the methods, we scrutinized the outputs with respect to relative error (RelErr) [4]. Results are shown by Figure 3. In the figure, the initial and the noisy signals have been pictured by subfigures (a) and (b), respectively. Also, reconstructed signals (marked by the red circles) obtained by DQNADMM, MMLBFGS, MLMBFGS, DQNB1, DQNB2 and DQNMSG over the initial signals (marked by the blue endpoints) have been pictured by subfigures (c)–(h), respectively. The figure reveals that DQNADMM, DQNB2 and MLMBFGS work better than the other techniques.

5 Conclusions

To take advantage of the significant aspects of the diagonal quasi-Newton updates for handling the big data optimization models, inspired by [6], we have proposed a minimization problem founded upon the popular Byrd-Nocedal measure function as well as the secant equation. The proposed minimization problem has been solved utilizing the alternating direction method of multipliers strategy. Besides, we have addressed convergence of the algorithm using the classic nonmonotone line search of [39] attached by the simulated annealing algorithm. To evaluate the effect of our theoretical arguments, we performed some computational tests using a class of problems of the CUTer library. Results showed that the given method is computationally promising.

A special numerical experiment has been also performed on the well-known compressive sensing problem for which a smooth transcendental function has been employed to approximate the ℓ_1 regularizer term. The outputs showed that our algorithm is capable of delivering progress in sparse signal recovery.

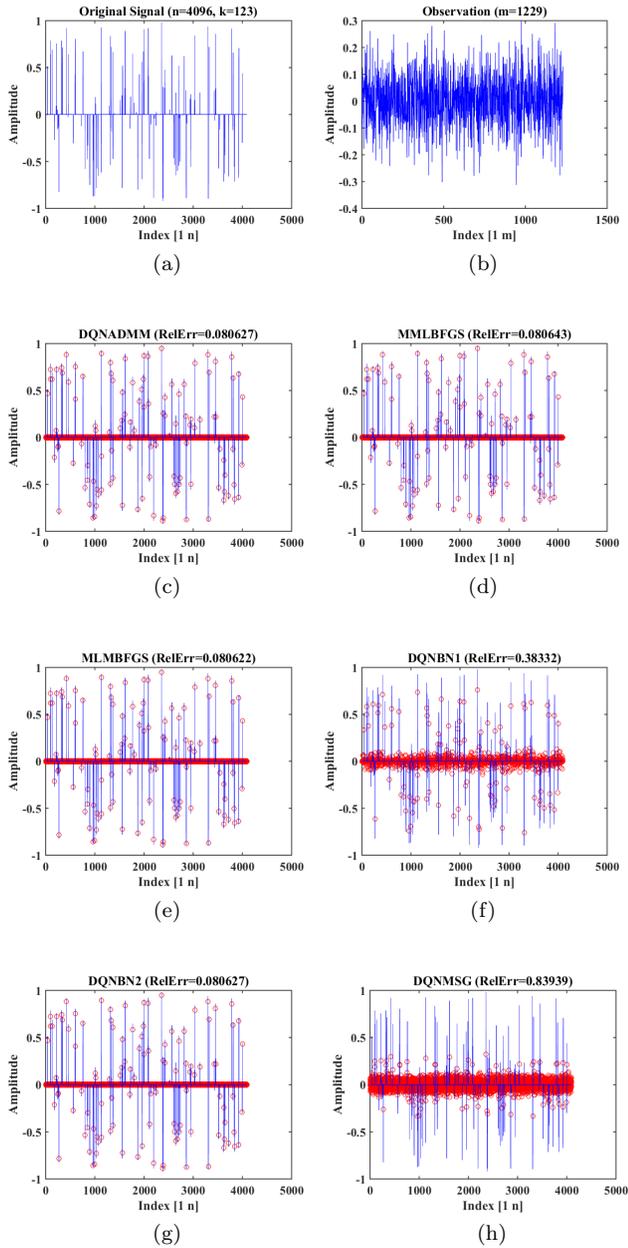


Figure 3. CS visual output data.

Acknowledgements

This research was in part supported by the grant no. 4005578 from Iran National Science Foundation (INSF), and in part by the Research Councils of

Semnan University and Free University of Bozen-Bolzano. The authors thank the anonymous reviewers for their valuable comments and suggestions that helped to improve the quality of this work.

References

- [1] K. Amini and A. Ghorbani Rizi. A new structured quasi-Newton algorithm using partial information on Hessian. *J. Comput. Appl. Math.*, **234**(3):805–811, 2010. <https://doi.org/10.1016/j.cam.2010.01.044>.
- [2] Z. Aminifard and S. Babaie-Kafaki. A modified descent Polak–Ribière–Polyak conjugate gradient method with global convergence property for nonconvex functions. *Calcolo*, **56**(2):16, 2019. <https://doi.org/10.1007/s10092-019-0312-9>.
- [3] Z. Aminifard and S. Babaie-Kafaki. Diagonally scaled memoryless quasi-Newton methods with application to compressed sensing. *J. Ind. Manag. Optim.*, **19**(1):437–455, 2023. <https://doi.org/10.3934/jimo.2021191>.
- [4] Z. Aminifard, A. Hosseini and S. Babaie-Kafaki. Modified conjugate gradient method for solving sparse recovery problem with nonconvex penalty. *Signal Process.*, **193**:108424, 2022. <https://doi.org/10.1016/j.sigpro.2021.108424>.
- [5] N. Andrei. Convex functions. *Adv. Model. Optim.*, **9**(2):257–267, 2007.
- [6] N. Andrei. A diagonal quasi-Newton updating method based on minimizing the measure function of Byrd and Nocedal for unconstrained optimization. *Optimization*, **67**(9):1553–1568, 2018. <https://doi.org/10.1080/02331934.2018.1482298>.
- [7] N. Andrei. Diagonal approximation of the Hessian by finite differences for unconstrained optimization. *J. Optim. Theory Appl.*, **185**(3):859–879, 2020. <https://doi.org/10.1007/s10957-020-01676-z>.
- [8] S. Babaie-Kafaki. On optimality of the parameters of self-scaling memoryless quasi-Newton updating formulae. *J. Optim. Theory Appl.*, **167**(1):91–101, 2015. <https://doi.org/10.1007/s10957-015-0724-x>.
- [9] S. Babaie-Kafaki. A monotone preconditioned gradient method based on a banded tridiagonal inverse Hessian approximation. *Univ. Politehnica Bucharest Sci. Bull. Ser. A: Appl. Math. Phys.*, **80**(1):55–62, 2018.
- [10] S. Babaie-Kafaki and Z. Aminifard. Two-parameter scaled memoryless BFGS methods with a nonmonotone choice for the initial step length. *Numer. Algor.*, **82**(4):1345–1357, 2019. <https://doi.org/10.1007/s11075-019-00658-1>.
- [11] S. Babaie-Kafaki, Z. Aminifard and S. Ghafoori. Nonmonotone diagonally scaled limited-memory BFGS methods with application to compressive sensing based on a penalty model. *Appl. Numer. Math.*, **181**:618–629, 2022. <https://doi.org/10.1016/j.apnum.2022.07.008>.
- [12] S. Babaie-Kafaki and R. Ghanbari. A linear hybridization of the Hestenes–Stiefel method and the memoryless BFGS technique. *Mediterr. J. Math.*, **15**:86, 2018. <https://doi.org/10.1007/s00009-018-1132-x>.
- [13] Y.J. Bagul. A smooth transcendental approximation to $|x|$. *Inter. J. Math. Sci. & Engg. Appls.*, **11**(II):213–217, 2017.
- [14] J. Barzilai and J.M. Borwein. Two-point stepsize gradient methods. *IMA J. Numer. Anal.*, **8**(1):141–148, 1988. <https://doi.org/10.1093/imanum/8.1.141>.
- [15] D. Bertsimas and J.N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, Massachusetts, 1997.

- [16] A.M. Bruckstein, D.L. Donoho and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, **51**(1):34–81, 2009. <https://doi.org/10.1137/060657704>.
- [17] R.H. Byrd and J. Nocedal. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM J. Numer. Anal.*, **26**(3):727–739, 1989. <https://doi.org/10.1137/0726042>.
- [18] X. Chen. Smoothing methods for nonsmooth, nonconvex minimization. *Math. Programming*, **134**:71–99, 2012. <https://doi.org/10.1007/s10107-012-0569-0>.
- [19] Y.H. Dai and L.Z. Liao. New conjugacy conditions and related nonlinear conjugate gradient methods. *Appl. Math. Optim.*, **43**(1):87–101, 2001. <https://doi.org/10.1007/s002450010019>.
- [20] J.E. Dennis, Jr. and H. Wolkowicz. Sizing and least-change secant methods. *SIAM J. Numer. Anal.*, **30**(5):1291–1314, 1993. <https://doi.org/10.1137/0730067>.
- [21] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Math. Programming*, **91**(2, Ser. A):201–213, 2002. <https://doi.org/10.1007/s101070100263>.
- [22] W.L. Dong, X. Li and Z. Peng. A simulated annealing-based Barzilai-Borwein gradient method for unconstrained optimization problems. *Asia-Pac. J. Oper. Res.*, **36**(4):1950017, 2019. <https://doi.org/10.1142/S0217595919500179>.
- [23] N.I.M. Gould, D. Orban and Ph.L. Toint. CUTER: a constrained and unconstrained testing environment, revisited. *ACM Trans. Math. Software*, **29**(4):373–394, 2003. <https://doi.org/10.1145/962437.962439>.
- [24] L. Grippo, F. Lampariello and S. Lucidi. A nonmonotone line search technique for Newton’s method. *SIAM J. Numer. Anal.*, **23**(4):707–716, 1986. <https://doi.org/10.1137/0723046>.
- [25] W.W. Hager and H. Zhang. Algorithm 851: CG-Descent, a conjugate gradient method with guaranteed descent. *ACM Trans. Math. Software*, **32**(1):113–137, 2006. <https://doi.org/10.1145/1132973.1132979>.
- [26] L. Han, G. Yu and L. Guan. Multivariate spectral gradient method for unconstrained optimization. *Appl. Math. Comput.*, **201**(1-2):621–630, 2008. <https://doi.org/10.1016/j.amc.2007.12.054>.
- [27] M.A. Hassan, W.J. Leong and M. Farid. A new gradient method via quasi-Cauchy relation which guarantees descent. *J. Comput. Appl. Math.*, **230**(1):300–305, 2009. <https://doi.org/10.1016/j.cam.2008.11.013>.
- [28] P.J. Huber. Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Stat.*, **1**(5):799–821, 1973. <https://doi.org/10.1214/aos/1176342503>.
- [29] W.J. Leong, M. Farid and M.A. Hassan. Scaling on diagonal quasi-Newton update for large-scale unconstrained optimization. *B. Malays. Math. Sci. So.*, **35**(2):247–256, 2012.
- [30] D.H. Li and M. Fukushima. On the global convergence of the BFGS method for nonconvex unconstrained optimization problems. *SIAM J. Optim.*, **11**(4):1054–1064, 2001. <https://doi.org/10.1137/S1052623499354242>.
- [31] Y. Nesterov. Smooth minimization of nonsmooth functions. *Math. Programming*, **103**(1):127–152, 2005. <https://doi.org/10.1007/s10107-004-0552-5>.

- [32] A.M. Nezhad, R.A. Shandiz and A.E. Jahromi. A particle swarm-BFGS algorithm for nonlinear programming problems. *Comput. Oper. Res.*, **40**(4):963–972, 2013. <https://doi.org/10.1016/j.cor.2012.11.008>.
- [33] S.S. Oren and D.G. Luenberger. Self-scaling variable metric (SSVM) algorithms. I. Criteria and sufficient conditions for scaling a class of algorithms. *Management Sci.*, **20**(5):845–862, 1973. <https://doi.org/10.1287/mnsc.20.5.845>.
- [34] S.S. Oren and E. Spedicato. Optimal conditioning of self-scaling variable metric algorithms. *Math. Programming*, **10**(1):70–90, 1976. <https://doi.org/10.1007/BF01580654>.
- [35] W. Sun and Y.X. Yuan. *Optimization Theory and Methods: Nonlinear Programming*. Springer, New York, 2006.
- [36] Z. Wei, G. Li and L. Qi. New quasi-Newton methods for unconstrained optimization problems. *Appl. Math. Comput.*, **175**(2):1156–1188, 2006. <https://doi.org/10.1016/j.amc.2005.08.027>.
- [37] C. Xu and J.Z. Zhang. A survey of quasi-Newton equations and quasi-Newton methods for optimization. *Ann. Oper. Res.*, **103**(1–4):213–234, 2001.
- [38] Y. Xu, M. Liu, Q. Lin and T. Yang. ADMM without a fixed penalty parameter: faster convergence with new adaptive penalization. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [39] H. Zhang and W.W. Hager. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM J. Optim.*, **14**(4):1043–1056, 2004. <https://doi.org/10.1137/S1052623403428208>.
- [40] M. Zhu, J.L. Nazareth and H. Wolkowicz. The quasi-Cauchy relation and diagonal updating. *SIAM J. Optim.*, **9**(4):1192–1204, 1999. <https://doi.org/10.1137/S1052623498331793>.