LLNL IM Release number: LLNL-JRNL-745068

A STRUCTURED QUASI-NEWTON ALGORITHM FOR OPTIMIZING WITH INCOMPLETE HESSIAN INFORMATION

COSMIN G. PETRA*, NAIYUAN CHIANG†, AND MIHAI ANITESCU‡

Abstract. We present a structured quasi-Newton algorithm for unconstrained optimization problems that have unavailable second-order derivatives or Hessian terms. We provide a formal derivation of the well-known BFGS secant update formula that approximates only the missing Hessian terms, and we propose a line-search quasi-Newton algorithm based on a modification of Wolfe conditions that converges to first-order optimality conditions. We also analyze the local convergence properties of the structured BFGS algorithm and show that it achieves superlinear convergence under the standard assumptions used by quasi-Newton methods using secant updates. We provide a thorough study of the practical performance of the algorithm on the CUTEr suite of test problems and show that our structured BFGS-based quasi-Newton algorithm outperforms the unstructured counterpart(s).

Key words. structured quasi-Newton, secant approximation, BFGS, incomplete Hessian

AMS subject classifications. 90C53, 90C30, 90C06

1. Introduction and motivation.

We propose an algorithm for solving unconstrained optimization problems of the form

$$\min_{x\in\mathbb{R}^n} f(x), \quad f(x) := k(x) + u(x) \tag{1}$$

for which the Hessian of $k(x)$, $K(x) := \nabla^2 k(x)$, is available but the Hessian of $u(x)$, $U(x) := \nabla^2 u(x)$ is not available or is expensive to evaluate. Such a situation occurs, for example, when f comprises sums of functions that are either algebraically described (and derivatives can be evaluated) or simulated (and second-order derivatives are costly to compute), as is the case of adjoint-based computations. A notable example is security-constrained operations in power grids, when the security criteria are expressed by constraints on dynamic contingencies [18]. When these transient behavior requirements are expressed by penalty or Lagrangian approaches, the optimization problem has a logical two-stage structure [1]. Evaluation of the contribution of the second-stage, transient components involves obtaining the sensitivity information of a differential algebraic equation, for which the forward simulation alone is much costlier than the evaluation of the terms depending only on the first-stage variables. Another example is nonlinear least-squares problems, for which the Hessian is the sum of a term containing only first-order Jacobian information, which is available, and a term containing Hessian information, which is costly to compute and evaluate. On the other hand, we work under the assumption that the gradient $\nabla f(x)$ is available.

The goal of this work is to develop algorithms of a quasi-Newton flavor that are capable of combining the existing Hessian information and secant updates for the unavailable part of the Hessian. The motivation behind this work is that one can reasonably expect that such algorithms will perform better than general quasi-Newton counterparts that do not consider the structure in the Hessian.

*Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550. E-mail: petra1@llnl.gov. Corresponding author.

†United Technologies Research Center, 411 Silver Lane, East Hartford, CT 06118. E-mail:chiangn@utrc.utc.com

‡Mathematics and Computer Science Division, Argonne National Laboratory, 9700 Cass Avenue, Lemont, IL 60439. E-mail: anitescu@mcs.anl.gov

Our approach is in the spirit of quasi-Newton secant methods equipped with a line-search mechanism. Such iterative numerical procedures produce a sequence of iterates $\{x_k\}_{k\geq 0}$ of the form $x_{k+1} = x_k + \alpha_k p_k$, where the quasi-Newton search direction p_k is given by $p_k = B_k^{-1}\nabla f(x_k)$, with B_k being an approximation of the Hessian $\nabla^2 f(x_k)$. The step length α_k is found by an appropriate search along the direction p_k that ensures the convergence to a stationary point of the gradient, for example, a line search based on Wolfe conditions [22].

The secant Hessian approximation B_k is updated during the line-search algorithm accordingly to a closed-form formula

$$B_{k+1} = \mathbb{B}(x_k, s_k, y_k, B_k),$$

where s_k and y_k are the changes in the variables and gradient, respectively, and are given by

$$s_k = x_{k+1} - x_k \text{ and } y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

The formula $\mathbb{B}$ ensures that B_{k+1} is symmetric and satisfies the secant equation

$$B_{k+1}s_k = y_k. \tag{2}$$

In order to derive the update formula $\mathbb{B}$ and to uniquely determine B_{k+1}, additional conditions need to be posed on B_{k+1} [19]. These conditions take the form of the so-called proximity criterion, which requires that B_{k+1} or its inverse, B_{k+1}^{-1}, is the closest approximation in some norm to B^k or B_k^{-1}, respectively. Weighted Frobenius matrix norms have been used to define proximity [19]; they are a convenient choice since they allow an analytical expression for $\mathbb{B}$. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) formulas are obtained by imposing the proximity criterion on the inverse. The BFGS formula for the inverse $H_{k+1} = B_{k+1}^{-1}$ is obtained as the solution to the following variational characterization problem,

$$\begin{aligned} H_{k+1}^{BFGS} = \operatorname{argmin}_H \quad & \|H - H_k\|_W \\ \text{s.t.} \quad & H_k y_k = s_k, \\ & H = H^T, \end{aligned} \tag{3}$$

as the analytical expression (see [14, Corollary 2.3])

$$H_{k+1}^{BFGS} = (I - \gamma_k y_k s_k^T)H_k(I - \gamma_k s_k y_k^T) + \gamma_k y_k y_k^T. \tag{4}$$

Here the weight matrix W can be chosen as any symmetric positive definite matrix satisfying the secant equation $Wy_k = s_k$. The weighted Frobenius norm $\|A\|_W$ is the Frobenius norm of $W^{1/2}AW^{1/2}$. We observe that the choice of W does not influence the inverse formula (4). Also, we use the notation $\gamma_k = 1/(s_k^T y_k)$. The BFGS update for the Hessian can be obtained by using the Sherman–Morrison–Woodbury formula, and it is given by

$$B_{k+1}^{BFGS} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T y_k}. \tag{5}$$

When the proximity criterion, together with the secant equation and symmetry conditions, is enforced for the matrix B_k, one can similarly obtain the Davidon–

Fletcher–Powell (DFP) secant update as the solution to

$$
\begin{aligned}
(6) \qquad B_{k+1}^{DFP} = \text{argmin}_B \quad & \|B - B_k\|_W \\
\text{s.t.} \quad & B_k s_k = y_k, \\
& B = B^T,
\end{aligned}
$$

in the form of the analytical expression $B_{k+1}^{DFP} = (I - \gamma_k y_k s_k^T) B_k (I - \gamma_k s_k y_k^T) + \gamma_k y_k y_k^T$. The weight matrix W is chosen to satisfy $W s_k = y_k$ and to be positive definite.

In this paper we first derive in Section 2 *structured* counterparts to the BFGS formulas (4) and (5). The formal derivation of the structured formulas uses the apparatus of Güler et al. [14] to derive two structured BFGS formulas coming from two different perspectives, least-square and trace-determinant variational characterizations, on BFGS updates.

We then investigate in Section 3 line-search globalization strategies. We find that for one of the structured BFGS formulas, one can always find a subset of the Wolfe points that ensures the positive definiteness of the structured update; furthermore, such Wolfe points can be identified at low computational cost. For the other structured BFGS update, we suggest using a standard inertia perturbation (by a multiple of identity) of the Hessian approximation to obtain global convergence. We show that the perturbation parameter can be computed in a computationally efficient manner a priori, and thus no extra factorization of the linear systems involving the quasi-Newton approximation is needed. As a result, both approaches are globally convergent to a first-order stationary point, as we discuss in Section 3.2.

Structured BFGS updates have been investigated before, for example, for nonlinear least-squares problems [8, 6, 16, 2] and for Lagrangian functions in sequential quadratic programming (SQP) methods for constrained optimization [20, 15]. To the best of our knowledge, however, no attempts have been made before to provide a globalization strategy in conjunction with this class of BFGS updates. Furthermore, we show that our formal derivation of structured secant formulas (see discussions of the least-square and trace-determinant variational characterizations for equations (9) and (10)) is a rigorous tool that requires no intuition or any other type of empirical input in finding structured secant formula updates that are optimal with respect to various variational characterizations.

In Section 3.3 we show the local superlinear convergence of the two structured BFGS formulas. We work under the standard assumptions used in analyzing the local convergence rates of certain quasi-Newton secant methods; for example, see the review paper of Dennis and Moré [7]. Our analysis is also standard in the sense that the superlinear convergence is proved by using a characterization (e.g., bounded deterioration property) of superlinear convergence due to Dennis and Moré[7, 3] and it employs intermediary results and techniques from [3, 13, 6].

We also provide a thorough performance evaluation of various computational strategies with the structured BFGS formulas proposed in this paper. This includes a comparison with the unstructured BFGS formula using a similar line-search algorithm, which reveals that the structured quasi-Newton algorithms outperform the unstructured counterparts and confirms that our structured formulas are capable of using the existing, incomplete Hessian information to improve numerical performance.

2. Derivation of structured quasi-Newton secant updates. In this section we derive "structured" BFGS formulas for structured Hessians $\nabla^2 f(x) = K(x) + U(x)$ of our problem of interest (1). We drop the k indexing and use the subscript $+$ to denote the quantities updated at each iteration (indexed by $k+1$ in the preceding section).

Our goal is to derive structured formulas that use the exact Hessian information $K(x_+)$ and only approximate the missing curvature, $U(x_+)s$, along the direction s. We are specifically looking for *structured* secant update formulas in the form

$$B_+ = K(x_+) + A_+, \tag{7}$$

where A_+ approximates the unknown Hessian $U(x_+)$ in the spirit of the secant equation (2). That is, we require that

$$A_+ s = \bar{y} := \nabla u(x_+) - \nabla u(x). \tag{8}$$

We remark that one also can use an "unstructured" right-hand side in the secant equation (2), that is, require that A_+ satisfies $(K(x_+)+A_+)s = y := \nabla f(x_+) - \nabla f(x)$. Doing so would imply that $A_+ s = \nabla u(x_+) - \nabla u(x) + [\nabla k(x_+) - \nabla k(x) - K(s_+)s]$, which forces A_+ to inadvertently incorporate the curvature variation $\nabla k(x_+) - \nabla k(x) - K(s_+)s$ corresponding to the term $k(\cdot)$ for which the curvature is available. Other structure-exploiting choices for $\bar{y}$ also are possible, for example, for nonlinear least-squares problems [8, 6, 16] and the Lagrangian function in SQP methods for constrained optimization [20, 15].

Similarly to the preceding section, we derive the analytical formulas for A_+ by imposing a proximity criterion together with the symmetry condition and the secant equation. We observe that, in contrast to the classical quasi-Newton update, the proximity condition (3) can be formulated in two ways, depending on where $K(x)$ is evaluated. That is, it can be posed as either

$$A_+^M = \operatorname{argmin}_X \|(X + K(x_+))^{-1} - (A + K(x))^{-1}\|_W \quad \text{or} \tag{9}$$

$$A_+^P = \operatorname{argmin}_X \|(X + K(x_+))^{-1} - (A + K(x_+))^{-1}\|_W. \tag{10}$$

We also denote $B_+^M = K(x_+) + A_+^M$ and $B_+^P = K(x_+) + A_+^P$ the structured updates of the form (7) corresponding to the two variational characterizations cited. We observe that the latter variational characterization seems more reasonable from a least-squares perspective since it uses updated Hessian information $K(x_+)$ and, therefore, A_+^M is not required to approximate the change in the known part of the Hessian.

On the other hand, the trace-determinant variational characterization [4, 10, 14] indicates that solving (9) and (10) is in fact analogous to enforcing the the eigenvalues of $(A_+^M + K(x_+))^{-1}(A^M + K(x))$ and $(A_+^P + K(x_+))^{-1}(A^P + K(x_+))$ to be as close to 1 as possible. In this respect, the update based on (9) would be more suitable since B_+^M is more likely to inherit the spectral properties (e.g., positive definiteness) of $B^M = A^M + K(x)$; in contrast, B_+^P has spectral properties similar to $A^P + K(x_+) = B^P + K(x_+) - K(x)$, which is a matrix that is not necessarily positive definite when $K(x_+) - K(x)$ is large.

Based on the discussion of the preceding two paragraphs, we are uncertain which of the variational characterizations (9) and (10) has superior properties. For this reason we consider both updates in this paper.

Before deriving the structured BFGS formulas, we observe that $\bar{y} = A_+ s$ can be equivalently written as $\bar{y} = (H_+^{-1} - K(x_+))s$, where H_+ denotes the inverse of B_+^M or B_+^P. This leads to the "structured" inverse secant condition

$$H_+(\bar{y}_k + K(x_+)s) = s.$$

Therefore (9) and (10) with the symmetry condition and the inverse secant equation become

$$
\begin{aligned}
H_+ = \operatorname{argmin}_Y \quad & \|Y - H\|_W \\
\text{s.t.} \quad & Y(\bar{y} + K(x_+)s) = s, \\
& Y = Y^T,
\end{aligned} \tag{11}
$$

where H is either $(B^M)^{-1}$ or $(B^P)^{-1}$ (corresponding to (9) or (10), respectively). The solution to this variational characterization is (see [14, Corrolary 2.5])

$$
H_+ = (I - \bar{\gamma}s(\bar{y} + K(x_+)s)^T)H(I - \bar{\gamma}(\bar{y} + K(x_+)s)s^T) + \bar{\gamma}ss^T, \tag{12}
$$

where we define

$$
\bar{\gamma} = \left[(\bar{y} + K(x_+)s)^T s\right]^{-1}. \tag{13}
$$

The BFGS update for the Hessian matrix can be obtained again by using the Sherman–Morison–Woodbury formula, namely,

$$
H_+^{-1} = H^{-1} - \frac{H^{-1}ss^TH^{-1}}{s^TH^{-1}s} + \bar{\gamma}(\bar{y} + K(x_+)s)(\bar{y} + K(x_+)s)^T.
$$

This allows us to write the BFGS update corresponding to (9) as

$$
\begin{aligned}
A_+^M = & A^M + K(x) - K(x_+) \\
& - \frac{(A^M + K(x))ss^T(A^M + K(x))}{s^T(A^M + K(x))s} + \bar{\gamma}(\bar{y} + K(x_+)s)(\bar{y} + K(x_+)s)^T
\end{aligned}
$$

and the BFGS update corresponding to (10) as

$$
A_+^P = A^P - \frac{(A^P + K(x_+))ss^T(A^P + K(x_+))}{s^T(A^P + K(x_+))s} + \bar{\gamma}(\bar{y} + K(x_+)s)(\bar{y} + K(x_+)s)^T.
$$

To simplify notation, we introduce

$$
\mathbb{B}(s, y, M) = -\frac{Mss^TM}{s^TMs} + \frac{yy^T}{y^Ty}.
$$

We can then express A_+^M and A_+^P as

$$
A_+^M = A^M + K(x) - K(x_+) + \mathbb{B}(s, \bar{y} + K(x_+)s, A^M + K(x)), \tag{14}
$$

$$
A_+^P = A^P + \mathbb{B}(s, \bar{y} + K(x_+)s, A^P + K(x_+)). \tag{15}
$$

The corresponding structured BFGS updates for the Hessian matrix become

$$
B_+^M = A_+^M + K(x_+) = B^M + \mathbb{B}(s, \bar{y} + K(x_+)s, B^M), \tag{16}
$$

$$
\begin{aligned}
B_+^P = & A_+^P + K(x_+) = B^P + K(x_+) - K(x) + \\
& + \mathbb{B}(s, \bar{y} + K(x_+)s, B^P + K(x_+) - K(x)).
\end{aligned} \tag{17}
$$

Note on structured DFP update formulas. One can use this formal approach to derive a DFP structured update. The structured counterpart of the DFP variational form (6) would be

$$\begin{aligned}\min_{Y} \quad & \|Y-A\|_W \\ \text{s.t.} \quad & Ys = \bar{y}^k, \\ & Y = Y^T,\end{aligned}$$

which has the solution

$$A_+^{DFP} = (I - \bar{\gamma}'\bar{y}s_k^T)A^{DFP}(I - \bar{\gamma}' s\bar{y}^T) + \bar{\gamma}'\bar{y}\bar{y}^T,$$

where $\bar{\gamma}'$ is the scalar $(s^T\bar{y})^{-1}$. We observe that the corresponding structured DFP update $B_+^{DFP} = A_+^{DFP} + K(x_+)$ does not explicitly incorporate the approximation B^{DFP}. Therefore, it is difficult to investigate the properties of the Hessian approximation (e.g., positive definiteness, bounded deterioration) that are needed in our line-search algorithmic approach. Trust-region methods can be, for example, an alternative avenue for structured DFP updates; we defer such investigation to future work.

3. A line-search quasi-Newton algorithm for the structured updates.

In this section we propose and analyze a line-search algorithm that employ the structured BFGS updates B^M and B^P introduced in the preceding section. We first investigate the hereditary positive definiteness of the structured updates and, motivated by this investigation, propose a line-search algorithm derived from the Wolfe conditions enhanced by an additional condition to ensure positive definiteness of the update away from a neighborhood of the solution. Furthermore, we show that the algorithm converges to stationary points, and we prove that the algorithm is locally superlinear convergent.

3.1. Hereditary positive definiteness.

The positive definiteness of the BFGS update is needed to ensure that the quasi-Newton direction is a descent direction [19]. The BFGS *unstructured* updates remain positive definite (p.d.) as long as a set of conditions, namely, the Wolfe conditions, are satisfied by the line-search [19]. As we show in this section, this is not necessarily the case for our *structured* BFGS updates B^M and B^P. For the B^M update we propose a line-search mechanism that ensures the positive definiteness of the update throughout the algorithm. For the B^P update such modification is not possible, and we will ensure positive definiteness by using the traditional approach of inertia/eigenvalue regularization.

3.1.1. Structured B^M update.

Provided B^M is p.d., a sufficient condition for B_+^M to be p.d. is $\bar{\gamma} > 0$. To see this, let $u \neq 0$, and use (12) to write

$$\begin{aligned}u^T B_+^M u = u^T\left(H(x_+) + A_+^M\right)^{-1} u = \left[u - \bar{\gamma}(s^Tu)(\bar{y} + K(x_+)s)\right]^T \cdot \\ \cdot (B^M)^{-1}\left[u - \bar{\gamma}_k(s^Tu)(\bar{y}_k + K(x_+)s)\right] + \bar{\gamma}(s^Tu)^2 \geq 0. \end{aligned} \tag{18}$$

Equality holds if and only if $u - \bar{\gamma}(s^Tu)(\bar{y} + K(x_+)s) = 0$ and $s^Tu = 0$, which implies that $u = 0$. Therefore the inequality is strict, showing that B_+^M is positive definite.

In general $\bar{\gamma} = 1/\left[(\bar{y} + K(x_+)s)^Ts\right]$ is not necessarily positive. To see this, we use the mean value theorem for the function $t \mapsto s^T\nabla u(x + t(x_+ - x))$ on $[0, 1]$ to write

$s^T\bar{y} = s^T(\nabla u(x_+) - \nabla u(x)) = s^T\nabla^2 u(\bar{x})(x_+ - x) = s^T\nabla^2 u(\bar{x})s$, where $\bar{x} \in [x, x_+]$; this implies that

$$\bar{\gamma} = 1/\left[s^T(U(\bar{x}) + K(x_+)s\right], \tag{19}$$

which can be negative since $U(\bar{x})+K(x_+)$ is not necessarily positive definite. However, if there exist $x_+ = x + \alpha x$ for which $s^T\nabla^2 f(x_+)s^T = s^T(U(x_+) + K(x_+))s > 0$, the positiveness of $\bar{\gamma}$ can be ensured provided $\bar{x}$ is close to x_+. Below we show that $s^T\nabla^2 f(x_+)s^T > 0$ holds for a subset of x_+ satisfying the (weak) Wolfe conditions, and we enforce the latter condition in the line-search algorithm.

We introduce a simplifying notation and denote

$$x_+ = x_+(\alpha) = x + \alpha p, \quad \phi(\alpha) = f(x + \alpha p).$$

In other words, ϕ is f on the search direction p. Observe that

$$\phi'(\alpha) = \nabla f(x + \alpha p)^T p, \tag{20}$$

$$\phi''(\alpha) = (p)^T\nabla^2 f(x + \alpha p)p. \tag{21}$$

The following lemma shows that, for the purpose of leading to a $\bar{\gamma} > 0$, a subset of points satisfying the Wolfe conditions [19] also satisfy the inequality $\phi''(\alpha) > 0$.

Lemma 1. *If p is a descent direction, namely, $\phi'(0) < 0$, and if $\phi(\alpha)$ is bounded from below on $[0, \infty)$ then there exists an interval of points α^* such that $\phi''(\alpha^*) > 0$ and $x + \alpha^* p$ satisfies the Wolfe conditions:*

$$\phi(\alpha^*) \le \phi(0) + c_1\alpha^*\phi'(0), \tag{22}$$

$$\phi'(\alpha^*) \ge c_2\phi'(0), \tag{23}$$

where $0 < c_1 < c_2 < 1$.

Proof. First observe that there is an $\alpha_1 > 0$ for which $\phi'(\alpha_1) = c_1\phi'(0)$. To see this, consider $l_1(\alpha) = \phi(0) + c_1\alpha\phi'(0)$. Since $\phi'(0) < 0$, then $l_1(\alpha)$ is unbounded; however, $l_1(0) = \phi(0)$, and $l_1'(0) > \phi'(0)$ indicates that that there is at least one positive α where l_1 and ϕ intersect. We denote the smallest such value by α_0. Therefore we have $\phi(\alpha_0) = l_1(\alpha_0)$, or $\phi(\alpha_0) - \phi(0) = c_1\alpha_0\phi'(0)$. By the mean value theorem we have that there exists α_1 such $\phi(\alpha_0) - \phi(0) = \alpha_0\phi(\alpha_1)$. The last two equalities show that $\phi'(\alpha_1) = c_1\phi'(0)$.

Since $\phi'(0) < c_2\phi'(0) < c_1\phi'(0) = \phi'(\alpha_1)$, one can similarly show that there exists $\alpha_2 \in (0, \alpha_1)$ such that $\phi'(\alpha_2) = c_2\phi'(0)$. Let α_2 denote the largest such value.

The mean value theorem applied to $\phi(\alpha)$ on $[\alpha_2, \alpha_1]$ indicates that there exists $\alpha^* \in [\alpha_1, \alpha_2]$ such that $(\alpha_1 - \alpha_2)\phi''(\alpha^*) = \phi'(\alpha_1) - \phi'(\alpha_2)$, which implies $\phi''(\alpha^*) > 0$.

It remains to prove that α^* satisfies Wolfe conditions (22) and (23). The sufficient decrease condition (22) is satisfied in fact by any $\alpha \in (0, \alpha_0)$ ($l_1(\alpha)$ dominates $\phi(\alpha)$ on $(0, \alpha_0)$, as pointed out above) and hence by α^* also. To show that the curvature condition (23) is satisfied, we note that $\phi'(\alpha^*) \le \phi'(\alpha_2) = c_2\phi'(0)$ cannot hold since it would imply that $\phi'(\cdot)$ takes the value $c_2\phi'(0)$ on (α^*, α_1), contradicting the maximality α_2 in this respect. Therefore $\phi'(\alpha^*) > c_2\phi'(0)$ which proves (23). ∎

According to Lemma 2, the Hessian is p.d. (and thus $\bar{\gamma} > 0$) along the direction p at some step length α^* if $\bar{x}$ from (19) is close to $x_+ = x + \alpha^* p$. This suggests a modification of the secant equation (11), namely, that it should be applied at $\bar{x} = x + \bar{\alpha}p$ and x_+ instead of at x and x_+. This modification ensures that the BFGS update remains p.d. provided $\bar{\alpha}$ is close to α^*. This will be shown by Theorem 2. Before that step, we introduce the following notation:

$$\bar{y}(\alpha_1, \alpha_2) = \nabla u(x + \alpha_1 p) - \nabla u(x + \alpha_2 p), \tag{24}$$

$$s(\alpha_1, \alpha_2) = (x + \alpha_1 p) - (x + \alpha_2 p) = (\alpha_1 - \alpha_2)p, \tag{25}$$

$$\bar{\gamma}(\alpha_1, \alpha_2) = \left[(\bar{y}(\alpha_1, \alpha_2) + K(x + \alpha_1 p)s(\alpha_1, \alpha_2))^T s(\alpha_1, \alpha_2)\right]^{-1}. \tag{26}$$

Here α_1 and α_2 are positive numbers. Observe that $\bar{y}$, s, and $\bar{\gamma}$ correspond to (24), (25), and (26) for $\alpha_1 = \alpha^*$ and $\alpha_2 = 0$.

THEOREM 2. *Let α^* denote one of the step lengths given by Lemma 1 and $x_+ = x + \alpha^* p$. Then there exists α_* close to α^* such that the modified structured BFGS update*

$$\begin{aligned} A^M &= A^M - \frac{\left[A^M + K(x_+)\right] s(\alpha^*, \alpha_*) s(\alpha^*, \alpha_*)^T \left[A^M + K(x_+)\right]}{s(\alpha^*, \alpha_*)^T \left[A^M + K(x_+)\right] s(\alpha^*, \alpha_*)} \\ &\quad + \bar{\gamma}(\alpha^*, \alpha_*) \left[\bar{y}(\alpha^*, \alpha_*) + K(x_+) s(\alpha^*, \alpha_*)\right] \left[\bar{y}(\alpha^*, \alpha_*) + K(x_+) s(\alpha^*, \alpha_*)\right]^T, \\ B_+^M &= A_+^M + K(x_+) = B^M + \mathbb{B}\left(s(\alpha^*, \alpha_*), \bar{y}(\alpha^*, \alpha_*) + K(x_+) s(\alpha^*, \alpha_*), B^M\right) \end{aligned}$$

is positive definite.

Proof. Since $\phi''(\alpha^*) > 0$ by Lemma 1, one can easily see that for any $\hat{\alpha}$

$$s(\alpha^*, \hat{\alpha})^T \left[U(x_+) + K(x_+)\right] s(\alpha^*, \alpha_2) > 0. \tag{27}$$

On the other hand, one can write

$$\bar{\gamma}(\alpha^*, \alpha_*) = 1/\left[s(\alpha^*, \alpha_*)^T \left[U(x + \hat{\alpha} p) + K(x + \alpha^* p)\right] s(\alpha^*, \alpha_*)\right], \tag{28}$$

with $\hat{\alpha} \in [\alpha_*, \alpha^*]$ given by the mean value theorem, similar to obtaining (19).

Then (27), (28), and the continuity of $U(\cdot)$ imply that $\bar{\gamma}_k(\alpha^*, \alpha_*) > 0$ (and thus B_+^M is p.d.) provided that α^* (and hence $\hat{\alpha}$) is close to α^*. ☐

The Wolfe point satisfying $\phi''(\alpha^*) > 0$ (required by Lemma 1) can be found by modifying a line-search procedure such as Moré-Thuente [17]. The test itself, that is, checking $\phi''(\alpha^*) > 0$, or equivalently $\bar{\gamma}(\alpha^*, 0) > 0$, has low computational cost and does not require function or gradient evaluations. In our computational experiments, the first Wolfe point encountered generally satisfies this condition, with the exception of a few test problems that require one or two additional searches for α_*.

3.1.2. Structured B_+^P update. For the structured B_+^P, the situation is different from that for the B^M because

$$\begin{aligned} u^T B_+^P u = u^T \left(H(x_+) + A_+^P\right)^{-1} u &= \left[u - \bar{\gamma}(s^T u)(\bar{y} + K(x_+)s)\right]^T \cdot \\ &\cdot (B^P - K(x) + K(x_+))^{-1} \left[u - \bar{\gamma}_k(s^T u)(\bar{y}_k + K(x_+)s)\right] + \bar{\gamma}(s^T u)^2 \end{aligned} \tag{29}$$

can be negative since $B^P - K(x) + K(x_+)$ is not necessarily p.d. even if B^P is p.d.

To ensure the positive definiteness of the structured B_+^P update, we use a regularization approach that consists of adding a positive multiple of the identity, namely,

$$B_+^P = A_+^P + K(x_+) + \sigma I, \tag{30}$$

where $\sigma > 0$ is chosen such that B_+^P is positive definite.

In the context of our structured quasi-Newton line-search, one can easily compute the diagonal perturbation σ that will maintain positive definiteness. One needs only to choose σ such that $(\bar{y} + K(x_+)s)^T s + \sigma s^T s \geq \varepsilon > 0$, where $\varepsilon > 0$ is a safeguard to avoid division by small numbers; in particular, such a σ would be $\sigma = (\varepsilon + (\bar{y} + K(x_+)s)^T s)/\|s\|^2$. With this computationally cheap choice $\bar{\gamma} = 1/\left[(\bar{y} + K(x_+)s)^T s + \sigma s^T s\right]$ is positive, and hence the update B_+^P is p.d.

3.2. Global convergence. The satisfaction of the Wolfe conditions (22)–(23) at each iteration implies that the sequence of iterates has a limit point x_* that satisfies $\|\nabla f(x_*)\| = 0$ [7, Theorem 6.3] (also see [22, 23].

3.3. Local superlinear convergence. Of great interest in computational practice is also how fast the algorithm converges. In this section we prove that the two structured BFGS updates we derived in Section 2 together with the Wolfe-based line-search mechanism of Section 3 have *local* convergence properties specific to quasi-Newton method with secant updates, for example, q-superlinear convergence. Specifically, we show in this section that the iterates produced by our algorithm satisfy

$$\lim_{k\to\infty} \frac{x_{k+1} - x_*}{x_k - x_*} = 0.$$

We work under the following "standard" assumptions that have been used in many publications dedicated to the analysis of local convergence rates of certain quasi-Newton secant methods; for example, see the review paper [7].

Assumption 3. *There exists a local minimizer x_* to the minimization problem (1).*

Assumption 4. *The objective f is of class C^2, and the Hessians $\nabla^2 f$ and $\nabla^2 k$ are locally Lipschitz continuous at x_*; that is, there exist $L \geq 0$ and $L_\kappa \geq 0$ and a neighborhood $\mathcal{N}_1(x_*)$ of x_* such that for any $x \in \mathcal{N}_1(x_*)$*

$$\|\nabla^2 f(x) - \nabla^2 f(x_*)\| \leq L\|x - x^*\|, \tag{31}$$

$$\|\nabla^2 k(x) - \nabla^2 k(x_*)\| \leq L_\kappa\|x - x^*\|. \tag{32}$$

Under this assumption, u is also C^2, and its Hessian is locally Lipschitz. We denote the Lipschitz constant by L_u. From the triangle inequality it follows that $L_u \leq L + L_\kappa$.

Assumption 5. *The Hessian is continuously bounded from below and above in a neighborhood $\mathcal{N}_2(x_*)$ of x_*; that is, there exist $m > 0$ and $M > 0$ such that*

$$m\|x\|^2 \leq x^T \nabla^2 f(x_*) x \leq M\|x\|^2, \forall x \in \mathcal{N}_2(x^*). \tag{33}$$

Our analysis is also standard in the sense that the superlinear convergence is proved by using a characterization of superlinear convergence due to Dennis and Moré [7], namely, Theorem 6 below. In fact, the remainder of this section essentially shows that our structured BFGS updates satisfy the hypotheses of the Dennis–Moré characterization of superlinear convergence.

Theorem 6. *[7, Theorem 3.1] Let B_k be a sequence of nonsingular matrices, and consider updates of the form $x_{k+1} = x_k - B_k^{-1}\nabla f(x_k)$ such that x_k remains in a neighborhood of x_*, $x_k \neq 0$ for all $k \geq 0$, and converges to x_*. Then the sequence $\{x_k\}$ converges superlinearly to x_* if and only if*

$$\lim_{k\to\infty} \frac{\|(B_k - B_*)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0. \tag{34}$$

Here $B_* = \nabla^2 f(x^*)$.

For our line search approach, we note that the Wolfe conditions (22)–(23) hold with with unit steps ($\alpha = 1$) in the limit, according to a well-known result [7, Theorem 6.4]. Moreover, from Assumptions 4 and 5 we have that $\phi''(t) > 0, \forall t \in [0, 1]$, where ϕ is the function f along the search direction as described in Lemma 1. Therefore the possible change in the points used in the secant update as described in and after Theorem 2 will not be necessary. Since the (modified) Moré–Thuente [17] line search used in this work always starts with unit step, it follows that in a neighborhood of x^* our algorithm uses updates of the form $x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f(x_k)$ and the secant update is applied at x_k and x_{k+1}. Therefore, condition (34) is equivalent to

$$\lim_{k \to \infty} \frac{\|(B_k - B_*) s_k\|}{\|s_k\|} = 0, \tag{35}$$

a result that we prove in the remainder of this section. We remark that the remaining conditions required by Theorem 6 are satisfied. In particular, our updates remain positive definite, and the convergence was proved in Section 3.2.

Key to showing (35) is proving a "bounded deterioration" property for B_+^M and B_+^P, namely, Theorem 11. This property implies q-linear convergence, which in turn is needed to show (35) [3].

In what follows we denote $\sigma(u, v) = \max\{\|u - x_*\|, \|v - x_*\|\}$ and $B_* = \nabla^2 f(x_*)$.

LEMMA 7. *Let* $x, x_+ \in \mathcal{N}_3(x_*)$. *Under Assumptions* (3)–(5)*, the following inequalities hold:*

(i) $\|(\bar{y} + K(x_+) s) - B_* s\| \leq (L + 2L_k) \sigma(x, x_+) \|s\|$;
(ii) $\|\bar{y} + K(x_+) s\| \leq (M + (L + 2L_k) \sigma(x, x_+)) \|s\|$;
iii) $s^T (\bar{y} + K(x_+) s) \leq (M + (L + 2L_k) \sigma(x, x_+)) \|s\|^2$;
(iv) $s^T (\bar{y} + K(x_+) s) \geq [m - (L + 2L_k) \sigma(x, x_+)] \|s\|^2$.

Proof. (i) We first write

$$\begin{aligned} \bar{y} + K(x_+) s - B_* s &= \nabla u(x_+) - \nabla u(x) + K(x_+) s - K(x_*) s - U(x_*) s \\ &= \int_0^1 [U((1-t)x + t x_+) s] dt - U(x_*) s \; + \; [K(x_+) - K(x_*)] s. \end{aligned}$$

Then, standard norm inequalities and Assumption 5 imply that

$$\begin{aligned} \|(\bar{y} + K(x_+) s) - B_* s\| &\leq \int_0^1 [L_u (1-t) \|x - x_*\| + L_u t \|x_+ - x_*\|] \, \|s\| dt + L_\kappa \|x_+ - x_*\| \|s\| \\ &\leq \int_0^1 L_u \sigma(x, x_+) \|s\| dt + L_\kappa \|x_+ - x_*\| \|s\| \\ &\leq L_u \sigma(x, x_+) \|s\| + L_\kappa \sigma(x, x_+) \|s\| \\ &\leq (L + 2L_k) \sigma(x, x_+) \|s\|. \end{aligned}$$

∎

(ii) The inequality follows from the triangle inequality, (i), and Assumption 5, namely, $\|\bar{y} + K(x_+) s\| \leq \|\bar{y} + K(x_+) s - B_* s\| + \|B_* s\| \leq (M + (L + 2L_k) \sigma(x, x_+)) \|s\|$.

(iii) The inequality follows from the Cauchy-Schwartz inequality and (ii).

(iv) Similar to the proof of (i) we write

$$
\begin{aligned}
s^T\left[U(x_*)s-\bar{y}\right] &= s^T\left[U(x_*)s-\int_0^1 U((1-t)x+tx_+)sdt\right] \\
&= s^T\int_0^1\left[U(x_*)-U((1-t)x+tx_+)\right]sdt \\
&\le \|s\|\int_0^1\left[(1-t)L_u\|x_*-x\|+tL_u\|x_*-x_+\|\right]\|s\|dt \\
&= L_u\sigma(x,x_+)\|s\|^2 \le (L+L_\kappa)\sigma(x,x_+)\|s\|^2.
\end{aligned}
$$

This allows us to use Assumptions (4) and (5) to write

$$
\begin{aligned}
m\|s\|^2 &\le s^T B_* s \\
&= s^T(\bar{y}+K(x_+))s + s^T\left[U(x_*)s-\bar{y}\right] + s^T(K(x_*)-K(x_+))s \\
&\le s^T(\bar{y}+K(x_+))s + (L+L_\kappa)\sigma(x,x_+)\|s\|^2 + L_k\|x_*-x_+\|\|s\|^2 \\
&\le s^T(\bar{y}+K(x_+))s + (L+2L_\kappa)\sigma(x,x_+)\|s\|^2.
\end{aligned}
$$

This completes the proof of (iv). □

We introduce the "theoretical" update $B' = B + \mathbb{B}(s, B_*s, B) = B - \frac{Bss^TB}{s^TBs} + \frac{B_*ss^TB_*}{s^TB_*s}$, which uses the curvature at the optimal solution along the current update direction, namely, B_*s, instead of the structured curvature $\bar{y}+K(x_+)s$. This theoretical update plays an important role in establishing the bounding inequalities needed to prove the bounded deterioration of B_+^M and B_+^P. The theoretical update of B^M is denoted by ${B^M}'$.

To simplyf the proofs, we also transform the B^P update to a slightly different form and notation. In particular, we observe from (17) that the B_+^P formula can be written as $B_+^P = \bar{B}^P + \mathbb{B}(s, \bar{y}+K(x_+)s, \bar{B}^P)$, where $\bar{B}^P = B^P + K(x_+) - K(x)$. One can easily see that a recurrence similar to (17) holds:

$$
\bar{B}_+^P = B_+^P + K(x_{++}) - K(x_+) = \bar{B}^P + \mathbb{B}(s, \bar{y}+K(x_+)s, \bar{B}^P) + [K(x_{++})-K(x_+)]. \tag{36}
$$

This new notation simplifies the use of the theoretical update in the proofs, as will become apparent later in this section. We denote the theoretical update of $\bar{B}^P$ by ${\bar{B}^P}'$.

LEMMA 8. *([6, Lemma 2.1.]) Consider a symmetric matrix M and vectors u, z such that $u^Tu = 1$ and $u^TMu = (u^Tz)$. If we define $M_+ = M + uu^T - zz^T$, then*

$$
\|M_+ - I\|_F^2 = \|M-I\|_F^2 - \left[(1-z^Tz)^2 + 2\left(z^TMz - (z^Tz)^2\right)\right]. \tag{37}
$$

Moreover, if M is also positive definite, if $u = v/\|v\|$, and if $z = Mv/\sqrt{v^TMv}$ for some $v \neq 0$, then

$$
z^TMz \ge (z^Tz)^2, \tag{38}
$$

and

$$
\|M_+ - I\|_F \le \|M-I\|_F. \tag{39}
$$

The following results are a direct consequence of Lemma 8 and reveal interesting properties of B'. Here we denote $\|M\|_* \triangleq \|M\|_{B_*} = \|B_*^{-0.5}MB_*^{-0.5}\|_F$.

Lemma 9. *Let B denote B^M or $\bar{B}^P$, and let B' denote $B^{M'}$ or $\bar{B}^{P'}$, respectively, and consider $z = \frac{B_*^{-0.5}Bs}{\sqrt{s^TBs}}$. Then one can write*

(i) $\|B' - B_*\|_*^2 = \|B - B_*\|_*^2 - \left[(1 - z^Tz)^2 + 2\left(z^T B_*^{-0.5} B B_*^{-0.5} z - (z^Tz)^2\right)\right]$,

(ii) $z^T B_*^{-0.5} B B_*^{-0.5} z \geq (z^Tz)^2$, *and*

(iii) $\|B' - B_*\|_* \leq \|B - B_*\|_*$.

Proof. We first observe that

$$\|B' - B_*\|_*^2 = \left\| B_*^{-0.5} B B_*^{-0.5} - I + \frac{B_*^{0.5} s s^T B_*^{0.5}}{s^T B_* s} - \frac{B_*^{-0.5} B s s^T B B_*^{-0.5}}{s^T B s} \right\|_F^2. \tag{40}$$

The inequality (i) follows from (37) by taking $M = B_*^{-0.5} B B_*^{-0.5}$ and $u = B_*^{0.5} s / \sqrt{s^T B_* s}$ in Lemma 8 and by observing that the lemma's condition $u^T M z = u^T z$ is fulfilled.

Similarly, (ii) follows from (38) of Lemma 8 with $v = B_*^{0.5} s \neq 0$ since one can easily verify that the conditions $u = v/\|v\|$ and $z = Mv/\sqrt{v^T M v}$ are satisfied.

Furthermore, since $\|B - B_*\|_* = \|B_*^{-0.5} B B_*^{-0.5} - I\|_F$, (iii) is just (39). ∎

We now bound the variation in our updates from the theoretical updates.

Lemma 10. *If the Assumptions 3, 4, and 5 hold, then*

(i) there exists a constant $C_1^M > 0$ such that $\|B_+^M - B^{M'}\|_* \leq C_1^M \sigma(x, x_+)$ for any x, x_+ in a nontrivial neighborhood of x_* included in $\mathcal{N}_3(x_*)$;

(ii) there exists a constant $\bar{C}_1^P > 0$ such that $\|B_+^P - \bar{B}^{P'}\|_* \leq C_1^P \sigma(x, x_+)$ for any x, x_+ in a nontrivial neighborhood of x_* included in $\mathcal{N}_3(x_*)$.

As a consequence of (ii),

(iii) $\|\bar{B}_+^P - \bar{B}^{P'}\|_* \leq \bar{C}_1^P \sigma(x, x_+) + L_\kappa \sigma(x_+, x_{++})$ for any x, x_+, and x_{++} in a nontrivial neighborhood of x_* included in $\mathcal{N}_3(x_*)$

Proof. (i) We first observe that $B_+^M - B^{M'}$ is a rank 2 update, namely,

$$B_+^M - B^{M'} = \frac{(\bar{y} + K(x_+)s)(\bar{y} + K(x_+)s)^T}{(\bar{y} + K(x_+)s)^T s} - \frac{B_* s s^T B_*}{s^T B_* s}.$$

Then we perform the following manipulation:

$$\begin{aligned}
B_+^M - B^{M'} &= \frac{(\bar{y} + K(x_+)s)(\bar{y} + K(x_+)s - B_*s)^T}{(\bar{y} + K(x_+)s)^T s} + \frac{(\bar{y} + K(x_+)s - B_*s)s^T B_*}{s^T B_* s} \\
&\quad + \frac{(\bar{y} + K(x_+)s)s^T B_*}{(\bar{y} + K(x_+)s)^T s} - \frac{(\bar{y} + K(x_+)s)s^T B_*}{s^T B_* s} \\
&= \frac{(\bar{y} + K(x_+)s)(\bar{y} + K(x_+)s - B_*s)^T}{(\bar{y} + K(x_+)s)^T s} + \frac{(\bar{y} + K(x_+)s - B_*s)s^T B_*}{s^T B_* s} \\
&\quad + (\bar{y} + K(x_+)s)s^T B_* \frac{\left(B_*s - (\bar{y} + K(x_+)s)\right)^T s}{s^T B_* s \cdot (\bar{y} + K(x_+)s)^T s}.
\end{aligned}$$

Using the fact that $\|uv^T\|_F = \|u\|\|v\|$ and the standard norm inequalities, we write

$$\begin{aligned}
\|B_+^M - B^{M'}\|_F &\leq \frac{\|\bar{y} + K(x_+)s\|\|\bar{y} + K(x_+)s - B_*s\|}{\|(\bar{y} + K(x_+)s)^T s\|} + \frac{\|(\bar{y} + K(x_+)s - B_*s)\|\|B_*s\|}{s^T B_* s} \\
&\quad + \|\bar{y} + K(x_+)s\|\|B_*s\| \frac{\|\bar{y} + K(x_+)s - B_*s\|\|s\|}{s^T B_* s \cdot (\bar{y} + K(x_+)s)^T s}.
\end{aligned} \tag{41}$$

Furthermore, we use Assumption 5 and inequalities of Lemma 7 to obtain

$$
\begin{aligned}
\|B_+^M - B^{M'}\|_F \le & \frac{[M + (L+2L_k)\sigma(x,x_+)]}{[m-(L+2L_k)\sigma(x,x_+)]}(L+2L_k)\sigma(x,x_+) \\
& + \frac{M}{m}(L+2L_k)\sigma(x,x_+) \\
& + \frac{[M+(L+2L_k)\sigma(x,x_+)]\,M}{m\,[m-(L+2L_k).\sigma(x,x_+)]}(L+2L_k)\sigma(x,x_+)
\end{aligned}
$$

In this inequality we assume that $\mathcal{N}_3(x_*)$ was potentially shrinked to $\mathcal{N}_4(x_*)$ to ensure $m-(L+2L_k)\sigma(x,x_+) = \frac{1}{2}m > 0$. More specifically, $\mathcal{N}_4(x_*)$ should contain a ball of radius $\epsilon_4 = \frac{1}{2}\frac{m}{L+2L_k}$ around x_* since this implies that $m-(L+2L_k)\sigma(x,x_+) \ge m-(L+2L_k)\epsilon_4 = \frac{1}{2}m > 0$. We obtain

$$
\|B_+^M - B^{M'}\|_F \le \tilde{C}_1\,\sigma(x,x_+), \tag{42}
$$

$$
\text{where } \tilde{C}_1 = \left[\frac{M+\frac{1}{2}m}{\frac{1}{2}m} + \frac{M}{m} + \frac{(M+\frac{1}{2}m)M}{\frac{1}{2}m^2}\right](L+2L_k).
$$

Since $\|B_+^M - B^{M'}\|_* \le \|B_*^{-0.5}\|^2\|B_+^M - B^{M'}\|_F \le \frac{1}{m}\|B_+^M - B^{M'}\|_F$, we conclude that the lemma holds with $C_1^M = \tilde{C}_1/m$.

(ii) Once we observe that $B_+^P - \bar{B}^{P'}$ is also a rank two update,

$$
B_+^P - \bar{B}^{P'} = \frac{(\bar{y}+K(x_+)s)(\bar{y}+K(x_+)s)^T}{(\bar{y}+K(x_+)s)^T s} - \frac{B_* s s^T B_*}{s^T B_* s},
$$

we can see that the proof is identical to (i).

(iii) By definition, $\bar{B}_+^P = B_+^P + K(x_{++}) - K(x_+)$ (see (36)). This allows to write, based on Assumption 4 and the triangle inequality, that

$$
\|\bar{B}_+^P - \bar{B}^{P'}\|_* \le \|B_+^P - \bar{B}^{P'}\|_* + L_\kappa\|x_{++} - x_+\|,
$$

which, based on (ii), proves (iii). Here, as before, we use the fact that $\|x_{++} - x_+\| \le \sigma(x_+, x_{++})$. ∎

Theorem 11. *Under Assumptions 3, 4, and 5, there exist constants $C_1^M > 0$ and $C_2^P > 0$ independent of x and x_+ such that*

$$
\|B_+^M - B_*\|_* \le \|B^M - B_*\|_* + C_1^M\sigma(x,x_+) \quad \text{and} \tag{43}
$$

$$
\|B_+^P - B_*\|_* \le \|B^P - B_*\|_* + C_2^P\sigma(x,x_+) \tag{44}
$$

for any $x, x_+ \in \mathcal{N}_3(x_*)$.

Proof. To obtain (43), we use the triangle inequality together with (i) of Lemma 10 and (iii) of Lemma 9:

$$
\|B_+^M - B_*\|_* \le \|B_+^M - B^{M'}\|_* + \|B^{M'} - B_*\|_* \le C_1^M\sigma(x,x_+) + \|B^M - B_*\|_*.
$$

To prove (44), we follow an argument similar to the one above:

$$
\|B_+^P - B_*\|_* \le \|B_+^P - \bar{B}^{P'}\|_* + \|\bar{B}^{P'} - B_*\|_* \le C_1^P\sigma(x,x_+) + \|\bar{B}^P - B_*\|_*.
$$

The last term can be bounded based on the definition of $\bar{B}^P$ and using the triangle inequality and Assumption 4:

$$\|\bar{B}^P - B_*\|_* = \|B^P + K(x_+) - K(x) - B_*\|_* \le \|B^P - B_*\|_* + L_\kappa \sigma(x, x_+). \qquad \square$$

Therefore (44) follows with $C_2^P = C_1^P + L_\kappa$.

The bounded deterioration property proved above implies q-linear convergence, namely, Theorem 12 below. This is only an intermediary step, which is needed to prove the asymptotic convergence (35) of the search direction by using the structured updates B_+ and B_+^P to the Newton direction in Theorem 14.

Theorem 12 is known to hold (for example, see [3, Theorem 3.2]) for any BFGS update sequence that satisfies a bounded deterioration property, such as the one we proved in Theorem 11. Hence, we state the result without a proof. We still work under Assumptions 3–5.

THEOREM 12. *[3, Theorem 3.2] Consider updates $x_{k+1} = x_k - B_k^{-1}\nabla f(x_k)$, where B_k are symmetric and satisfy a bounded deterioration property of (43) or (44). Then for each $r \in (0,1)$, there are positive constants $\epsilon(r)$ and $\delta(r)$ such that for $\|x_0 - x_*\| \le \epsilon(r)$ and $\|B_0 - B_*\|_* \le \delta(r)$, the sequence $\{x_k\}$ converges to x_*. Furthermore,*

$$\|x_{k+1} - x_*\| \le r\|x_k - x_*\| \tag{45}$$

for each $k \ge 0$. In addition, the sequences $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are uniformly bounded.

To prove the asymptotic convergence (35), we need the following result that is similar to the bounded deterioration property but involves the theoretical updates. We drop the "+" subscripts and use "k" indexes for the remainder of this section. We remark that $\sigma(x_k, x_{k+1}) = \max\{\|x_k - x_*\|, \|x_{k+1} - x_*\|\} = \|x_k - x_*\|$ and $\|x_k - x_*\| \le \epsilon(r) r^k$, for all $k \ge 0$ when one works under the conditions of Theorem 12.

LEMMA 13. *For a given $r \in (0,1)$ there exist constants $C_2^M(r) > 0$ and $\bar{C}_2^P(r) > 0$ independent of the sequence $\{x_k\}$ and k such that*

$$\|B_{k+1}^M - B_*\|_*^2 \le \|B_k^{M'} - B_*\|_*^2 + C_2^M r^k \quad \text{and} \tag{46}$$

$$\|\bar{B}_{k+1}^P - B_*\|_*^2 \le \|\bar{B}_k^{P'} - B_*\|_*^2 + \bar{C}_2^P r^k. \tag{47}$$

Proof. Based on the triangle inequality and (i) of Lemma 10, we first write

$$\begin{aligned} \|B_{k+1}^M - B_*\|_* &\le \|B_{k+1}^M - B_k^{M'}\| + \|B_k^{M'} - B_*\|_* \\ &\le C_1^M \sigma(x_k, x_{k+1}) + \|B_k^{M'} - B_*\|_* \le C_1^M \epsilon(r) r^k + \|B_k^{M'} - B_*\|_*. \end{aligned} \tag{48}$$

On the one hand, based on (iii) of Lemma (9), this inequality implies that

$$\|B_{k+1}^M - B_*\|_* \le C_1^M \epsilon(r) r^k + \|B_k^M - B_*\|_*,$$

which, in turn, inductively implies that

$$\|B_k^M - B_*\|_* \le \|B_0^M - B_*\|_* + C_1^M \epsilon(r)(r^{k-1} + \ldots + 1) \le \delta(r) + C_1^M \epsilon(r) \frac{1 - r^k}{1 - r}. \square \tag{49}$$

On the other hand, squaring (48) implies

$$\|B_{k+1}^M - B_*\|_*^2 \le \|B_k^{M'} - B_*\|_*^2 + 2C_1^M \epsilon(r) r^k \|B_k^{M'} - B_*\|_* + (C_1^M)^2 \epsilon^2(r) r^{2k},$$

which, by (iii) of Lemma (9) and inequality (49), implies that

$$\|B_{k+1}^M - B_*\|_*^2 \leq \|B_k^{M'} - B_*\|_*^2 + r^k \left[2C_1^M \epsilon(r) \left(\delta(r) + C_1^M \epsilon(r) \frac{1-r^k}{1-r} \right) + (C_1^M)^2 \epsilon^2(r) r^k \right].$$

We remark that the quantity inside the brackets in the second term in this inequality is bounded from above by

$$C_2^M = 2C_1^M \epsilon(r) \left(\delta(r) + C_1^M \epsilon(r) \frac{1}{1-r} \right) + (C_1^M)^2 \epsilon^2(r),$$

which completes the proof of (47).

The proof of (47) is almost identical. The slight difference is the inequality (48), which translates into

$$\|\bar{B}_{k+1}^P - B_*\|_* \leq (\bar{C}_1^P + L_\kappa r)\epsilon(r) r^k + \|\bar{B}_k^{P'} - B_*\|_* \leq (\bar{C}_1^P + L_\kappa)\epsilon(r) r^k + \|\bar{B}_k^{P'} - B_*\|_*$$

since (iii) of Lemma 10 needs to be used. The proof then is identical, and (47) holds with

$$\bar{C}_2^M = 2(\bar{C}_1^P + L_\kappa)\epsilon(r) \left(\delta(r) + (\bar{C}_1^P + L_\kappa)\epsilon(r) \frac{1}{1-r} \right) + (\bar{C}_1^P + L_\kappa)^2 \epsilon^2(r).$$

We are now in position to prove the final result.

THEOREM 14. *Under Assumptions 3–5 the update sequences* $\{B_k^M\}$ *and* $\{B_k^P\}$ *satisfy the asymptotic convergence limit* (35).

Proof. The proof has two parts.

(i) We first prove that the limit (35) holds for $\{B_k^M\}$ and $\{\bar{B}_k^P\}$;

(ii) We then use the convergence of $\{\bar{B}_k^P\}$ and a continuity argument to show that $\{B_k^P\}$ has a similar convergence behavior.

Part (1) of the proof relies on Lemma 13 and, for this reason, is identical for both $\{B_k^M\}$ and $\{\bar{B}_k^P\}$ sequences. B_k denotes either B_k^M or $\bar{B}_k^P$.

First let us fix $r \in (0,1)$. Lemma (13) implies

$$\|B_k - B_*\|_*^2 - \|B_k' - B_*\|_*^2 \leq \|B_k - B_*\|_*^2 - \|B_{k+1} - B_*\|_*^2 + C_2 r^k,$$

where C_2 is one of the constants provided by Lemma (13). We sum this over k and, based on Theorem 12, we obtain

$$\sum_{k=0}^{\infty} \left[\|B_k - B_*\|_*^2 - \|B_k' - B_*\|_*^2 \right] \leq \|B_0 - B_*\|_*^2 + C_2 \sum_k^{\infty} r^k < \infty.$$

Therefore $\lim_{k\to\infty} \left[\|B_k - B_*\|_*^2 - \|B_k' - B_*\|_*^2 \right] = 0$. Then Lemma 9 implies that $z_k^T z_k$ and $z_k^T B_*^{-0.5} B_k B_*^{-0.5} z_k$ converge to 1. These are exactly

$$\gamma_k := \frac{s_k^T B_k B_*^{-1} B_k s_k}{s_k^T B_k s_k} \to 1 \text{ and } \theta_k := \frac{s_k^T B_k B_*^{-1} B_k B_*^{-1} B_k s_k}{s_k^T B_k s_k} \to 1.$$

Furthermore, we observe that

$$\frac{s_k^T (B_k - B_*) B_*^{-1} B_k B_*^{-1} (B_k - B_*) s_k}{s_k^T B_k s_k} = \gamma_k - 2\theta_k + 1 \to 0 \text{ as } k \to 0.$$

However, this expression is exactly

$$\frac{\|B_k^{0.5} B_*^{-1}(B_k - B_*)s_k\|^2}{\|B_k^{0.5} s_k\|^2} \to 0,$$

which, by the boundedness of $\{B_k\}$ (Theorem 12) and of B_* (Assumption 5), implies the asymptotic convergence limit (35):

$$\frac{\|(B_k - B_*)s_k\|}{\|s_k\|} \to 0.$$

This completes (i).

For the proof of part (ii), we observe that B_k^P is by definition a $K(x_{k+1}) - K(x_k)$ perturbation of $\bar{B}_k^P$. Then Theorem 12 and the continuity of K given by Assumption 4 imply that $\{B_k^P\}$ also satisfies the limit above. ∎

3.4. Hereditary (strong) positive definiteness. We proved that the inverses of the updates B and B^P are bounded. This proof implies that the updates are positive definite and their eigenvalues are bounded away from zero. The consequence is that asymptotically regularizations and extra factorization are not needed for the updates, a desirable property since it reduces the computational cost.

3.5. Discussion for quadratic functions. We investigate the behavior of our line-search quasi-Newton method for strongly convex quadratic functions since this may provide insight into the behavior for the general nonquadratic case. For such functions the known and unknown Hessian parts are constant and denoted by K and U, and $Q := \nabla^2 f(x) = K + U$ is positive definite. We immediately observe that our structured update formulas (16) and (17) are identical and reduce to the unstructured BFGS formula

$$B_+ = \mathbb{B}(s, y, B), \tag{50}$$

where y is the variation in the gradient for the (unstructured) function and defined in (2). Nevertheless, our line-search algorithm starts with a different initial BFGS approximation, namely, with $K + A_0$, where A_0 is the initial approximation for U. Consequently, our algorithm will produce different iterations from those of an unstructured counterpart. The number of iterations of our algorithm is given by the following result.

PROPOSITION 15. *The quasi-Newton structured BFGS algorithm with an exact line search and A_0 as the initial approximation will generate a sequence of iterates identical to the conjugate gradient method (CG) preconditioned with $K + A_0$.*

Proof. Formulas like (50) produce a certain conjugacy property among the quasi-Newton iterates for strongly convex functions when *exact* line searches are used. This translates into iterations identical with the preconditioned CG preconditioned with B_0; see [19, Theorem 6.4] and the discussion afterwards. Our proposition follows from the fact that $B_0 = K + A_0$ in the case of our structured algorithm. ∎

Let us consider the situation when the unknown Hessian U is of rank r. We take $A_0 = 0$. The preconditioned CG iteration count will be at most the number of distinct eigenvalues of the preconditioned matrix $(K + A_0)^{-1}(K + U) = I + K^{-1}U$, which is bounded from above by $r + 1$. This points toward a dramatically improved performance of the structured quasi-Newton method over the unstructured counterpart (which would "blindly" start at a multiple of identity and would take n iterations [19, Theorem 6.4]).

One can argue that the unstructured BFGS algorithm can be started with $B_0 = K$ and thus would reproduce the behavior of the structured algorithm. While this is a valid argument for the quadratic case, it does not generalizes to nonquadratic nonlinear functions. In this case $B_0 = K(x_0)$ may not be positive definite and, hence, not suitable as a general initial choice for the initial approximation B_0.

4. Numerical Experiments.

In this section we investigate the performance (number of iterations and function evaluations) of our structured updates and perform comparisons with Newton and unstructured BFGS methods. We also study the behavior of our updates on nonlinear problems that have low-rank missing Hessians $\nabla^2 u(x)$, along the lines of the discussion following Proposition 15.

4.1. Test problems.

We use the CUTEr test set [12] specified in AMPL [11]. Of 73 unconstrained problems with nonlinear objectives in CUTEr, we consider 61 problems that have less than $5,000$ variables, a threshold dictated by the limitations our MATLAB implementation based on dense linear algebra. Of these 61 small- to medium-scale problems we have selected the problems on which IPOPT [21] requires fewer than 200 iterations in order to obtain an optimal solution. The selection yielded a total of 55 test problems. To mimic the unavailability of Hessian information, we modified the test problems to include a term $u(x)$ in (1) along the lines of the following two experiments.

Experiment 1: Given the CUTEr objective $f(x)$ and a *ratio* parameter $p \in [0,1]$, we consider the problem of minimizing $\bar{f}(x) = p * k(x) + (1-p) * u(x)$, where $k(x) = u(x) = f(x)$. Our structured BFGS algorithms use only $\nabla^2 k(x)$, the unstructured BFGS use no Hessian, and the Newton's method uses full Hessian $\nabla^2 \bar{f}(x) (= \nabla^2 f(x))$.

Experiment 2: Given the CUTEr objective $f(x)$, we set $u(x) = \sum_{i \in \mathcal{L}} \log[(x_i - x_i^*)^2 + 0.25]$, where x^* is the minimizer of $f(x)$ and $\mathcal{L}$ is an index set specifying the variables included in $u(x)$. In this experiment we minimize $\bar{f}(x) = k(x) + u(x)$, where $k(x) = f(x)$. The term $u(x)$ is a smooth nonlinear nonconvex function having the same minimizer x^* as $f(x)$ (when $\mathcal{L} = \{1, \ldots, n\}$). We observe that x^* is a local minimizer of $\bar{f}(x)$ for any subset $\mathcal{L}$ of $\{1, \ldots, n\}$. In this experiment, $u(x)$ is taken to be *full-space*, that is, $\mathcal{L} = \{1, \ldots, n\}$, as well as *subspace*, that is, of $\mathcal{L} \subset \{1, \ldots, n\}$ and of low cardinality $|\mathcal{L}| = O(1)$. The subspace approach is used to study the performance of our structured BFGS formulas on problems for which the unknown Hessian is of low rank and to investigate whether the theoretical result given by Proposition 15 for quadratic objectives is observed empirically for nonquadratic nonlinear objective functions.

When the term $u(x)$ is defined on a subspace, we compute subspace variants of the structured updates (14) and (15) of the form

(51)

$$A_{k+1}^M = A_k^M + K_k^L - K_{k+1}^L - \frac{\left[A_k^M + K_k^L\right](s_k^L)(s_k^L)^T\left[A_k^M + K_k^L\right]}{(s_k^L)^T\left[A_k^M + K_k^L\right](s_k^L)} + \gamma_k^L (y_k^L)(y_k^L)^T,$$

(52)

$$B_{k+1}^M = K_{k+1} + P A_{k+1}^M P^T,$$

and, respectively,

$$A_{k+1}^P = A_k^P - \frac{\left[A_k^P + K_{k+1}^L\right](s_k^L)(s_k^L)^T\left[A_k^P + K_{k+1}^L\right]}{(s_k^L)^T\left[A_k^P + K_{k+1}^L\right](s_k^L)} + \gamma_k^L (y_k^L)(y_k^L)^T, \tag{53}$$

$$B_{k+1}^P = K_{k+1} + PA_{k+1}^P P^T., \tag{54}$$

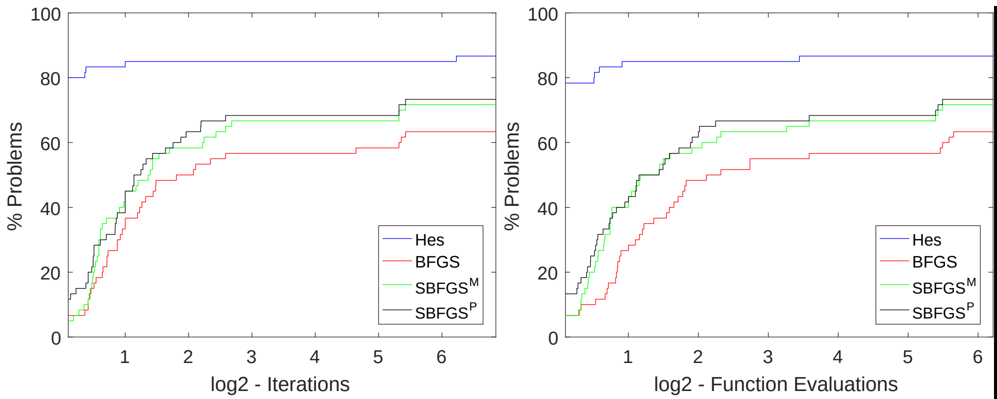
where $K_k^L = PK(x^k)P^T$, $s_k^L = Ps_k$, $y_k^L = \bar{y}_k + K_{k+1}^L s_k^L$, and $\gamma_k^L = ((y_k^L)^T(s_k^L))^{-1}$. Also, $P : \mathbb{R}^n \to \mathbb{R}^{|\mathcal{L}|}$ denotes the projection (matrix) from the full space to the subspace considered by $u(x)$.

Table 1 lists a summary of the algorithms and experimental setups used in investigating and comparing the performance of the structured BFGS algorithms of this paper.

Table 1: Summary of the six different methods. As discussed in the text, all the methods use the same line-search mechanism; they differ only in the type of Hessian approximation used.

Acronym	**Description**
Hes	Newton method with exact Hessian. Regularization is required to have positive definiteness.
BFGS	Unstructured BFGS. Positive definiteness is naturally guaranteed.
SBFGS_M	*Structured* BFGS with *full-space* update B^M from (18). Positive definiteness is naturally guaranteed.
SubBFGS_M	*Structured* BFGS with *subspace* update B^M from (52). Regularization is required to have positive definiteness. (Experiment 2 only)
SBFGS_P	*Structured* BFGS with *full-space* update B^P from (29). Regularization is required to have positive definiteness.
SubBFGS_M	*Structured* BFGS with *subspace* update B^P from (54). Regularization is required to have positive definiteness. (Experiment 2 only)

4.2. Implementation details. The numerical experiments were performed on a desktop with a quad-core 2.66 GHz Intel CPU Q8400 and with 3.8 GB memory and using MATLAB R2013a as an environment for implementation. We implemented Newton's method, (unstructured) BFGS method [19], and structured BFGS methods; all the algorithms use the line-search algorithm proposed by Moré and Thuente [17], which finds a point that satisfies the strong Wolfe conditions by using zoom-in procedure and both quadratic and cubic interpolation. For the Moré-Thuente line search we initially set the upper bound on the step to $\alpha_{max} \leftarrow 2$. The line search can terminate at the upper bound α_{max}, and the strong Wolfe conditions do not hold. This situation can occur when the objective function is not bounded from below along the search direction and also, in practice, when the search direction is "short" (for example, caused by distorted quasi-Newton approximations). In these situations, we increase $\alpha_{max} \leftarrow 100$ and restart the line search to rule out a short search direction.



(a) number of iterations (b) number of function evaluations

Fig. 1: Experiment 1 for ratio $p = 0.25$.

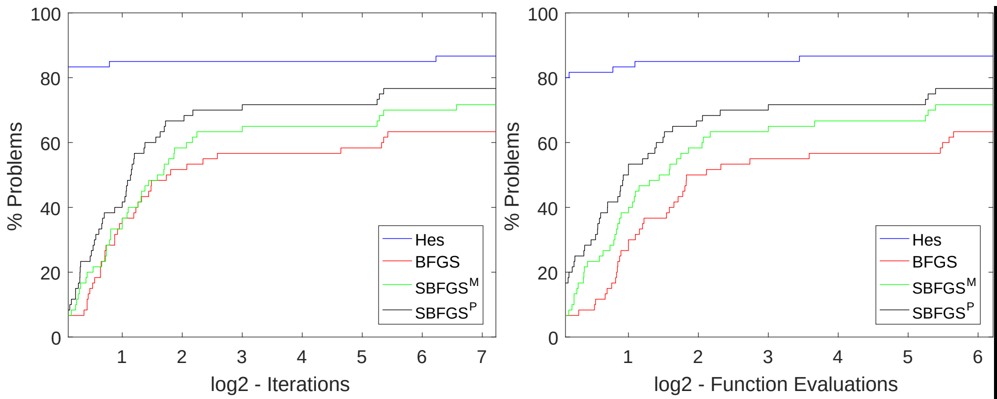
In a couple of instances the second line search fails, in which case we conclude that the function is likely unbounded and stop the algorithm.

In order to guarantee that the search direction is of descent, the line-search procedure requires that the Hessian (for Newton's method) or Hessian secant approximation (for BFGS-based algorithms) is p.d. For our (full-space) structured BFGS update B^M_+, the Moré-Thuente line-search was modified to find a Wolfe point satisfying the conditions of Theorem 2; for the (full-space) structured BFGS update B^P_+, a diagonal perturbation of the update is used, as discussed in Section 3.1.2. For Newton's method, whenever the Hessian is not p.d., we use an inertia-correction mechanism present in IPOPT [21], in which we repeatedly factorize and increase the diagonal perturbation until the (perturbed) Hessian becomes p.d. We remark that the unstructured BFGS algorithm does not need inertia regularization since the unstructured update is p.d. at Wolfe points. Clearly, the subspace structured BFGS updates (52) and (54) can be rank deficient, and thus we use inertia regularizations to guarantee positive definiteness of the updates. When the search direction is not a descent direction, which can occur for all algorithms because of numerical round-off error, we terminate the algorithm with an error message.

The subspace structured updates A^M_{k+1} and A^P_{k+1} given by (51) and (53) are set to zero whenever our implementation detects that there is no change in the gradient of the unknown part and in the variables in the subspace, namely, $s^L_k = 0$ and $y^L_k = 0$ numerically. This occurs in Experiment 2, when $u(x)$ is defined on a subspace and it is a manifestation of the fact that missing curvature $\nabla^2 u(x)$ is zero.

4.3. Discussion of results. We show comparisons of the six algorithms in Figures 1, 2, and 3. We employ the performance profiling technique of Dolan and Moré [9] in these figures; the x-axis shows the logarithm of the performance *metric*, namely, the number of iterations ((a) figures) or the number of function evaluations ((b) figures); the y-axis represents the percentage of problems solved with a given value/number of the performance metric on the x-axis.

Figures 1 and 2 show the performance profile for Experiment 1 for $p = 0.25$ and $p = 0.75$, respectively. Unsurprisingly, *Hes* (Newton's method) dominates all other



(a) number of iterations (b) number of function evaluations

Fig. 2: Experiment 1 for ratio $p = 0.75$.

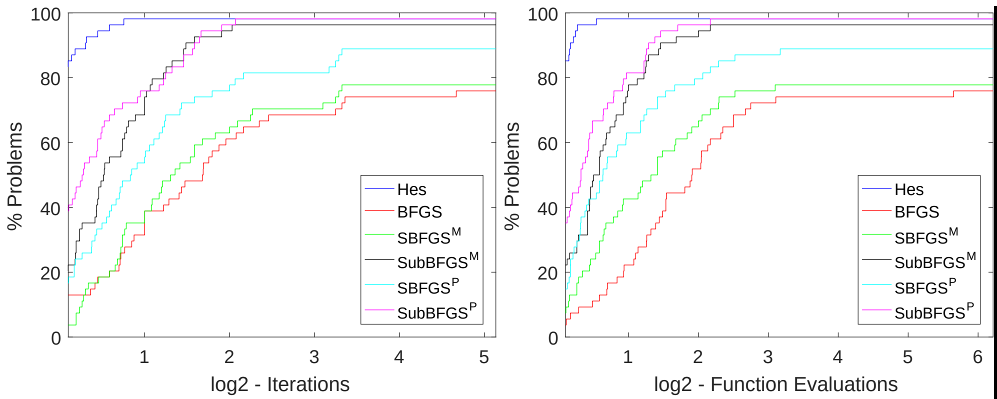
methods and can solve more than 80% of the given problems within 200 iterations. However, the structured BFGS methods outperform the unstructured BFGS method in both metrics (the number of iterations and the number of function evaluations). The performance gap between the structured and unstructured BFGS methods seems to be larger for $p = 0.75$ than for $p = 0.25$, possibly as the result of "more" Hessian information used by the structured updates. Also, we remark that the B^P update slightly outperforms B^M.

However, we point out that for some classes of problems, for example, with a large number of variables, the cost of factorizations needed by the inertia regularization technique necessary for B^P can be higher than the difference in the number of iterations/function evaluations over B^M, and thus B^M can outperform B^P in terms of execution time.

For Experiment 2 we consider subspaces of two (randomly chosen) variables for $u(x)$. Then we compare the six methods on the problems that can be solved by Hes or $BFGS$, for a total of 54 test problems. Figure 3 shows the Dolan and Moré performance profile for this comparison. We observe that the subspace update strategies perform better than the full space counterparts, as we anticipated. However, the subspace structured updates B^M and B^P perform about the same. We also observe that, as in Experiment 1, the four structured BFGS approaches perform better than the nonstructured BFGS algorithm.

Detailed results for each CUTER test problem and for each algorithm are displayed in Appendix A in Table 2 for Experiment 1 and in Table 3 for Experiment 2. In these tables, column $Stat$ shows the convergence status: 1 indicates success, while 2 denotes that the maximum number of (200) iterations has been reached; $Stat = 0$ denotes the failure of convergence because of the two conditions discussed in the preceding section: unboundedness of the objective function along the search direction or the occurrence of an ascent search direction. Columns $\#Iter$ and $\#Func$ show the numbers of iteration and function evaluation, respectively.

5. Conclusion and further developments. We derived structured secant formulas that incorporate available second-order derivative information. We also ana-



(a) number of iterations (b) number of function evaluations

Fig. 3: Experiment 2 with full-space and subspace structured updates B_+^M and B_+^P, as well as unstructured BFGS and Newton's method

lyzed the convergence (global and local) properties of a line-search algorithm equipped with the structured formulas. In particular we have shown convergence to stationary points and superlinear local convergence properties. Our numerical experiments were aimed at investigating the benefits of incorporating Hessian terms in the secant update formula and, indeed, show that the performance (number of iterations and number of function evaluations) of the unstructured BFGS can be improved; furthermore, the improvement can be considerable when the Hessian terms that are not known or available have low rank.

In this work we have used a straightforward rank two update recursive computation for the Hessian approximation matrix that results in *dense* linear system; obviously, this approach has severe computational limitations for large-scale problems. Limited-memory methods equipped with compact (e.g., low-rank) representations for the dense matrix in the idea of [5] is a natural direction of further development. A plausible alternative direction to investigate is the use of iterative methods for the solution of the quasi-Newton linear system since, as we discussed in Section 3.5, the known part of the Hessian is a natural preconditioner for the quasi-Newton linear system. We mention that further development is needed to extend or adapt such computational techniques to *constrained* problems.

Acknowledgments. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This material was also based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR), under Contract DE-AC02-06CH11347. C. P. also acknowledges partial support from the LDRD Program of Lawrence Livermore National Laboratory under the projects 16-ERD-025 and 17-SI-005. M. A. acknowledges partial NSF funding through awards FP061151-01-PR and CNS-1545046.

Appendix A. Detailed listing of the performance of our methods for CUTEst problems.

Table 2: CUTEr results of the four methods. Compute $u(x)$ by Experiment 1 with p=0.5.

	Hes			$BFGS$			$SBFGS^M$			$SBFGS^P$		
Prob	Stat	It	Fun	Stat	It	Fun	Stat	It	Fun	Stat	It	Fun
arglina	1	1	1	1	2	3	1	1	1	1	1	1
arglinb	0	1	72	0	1	71	1	2	3	1	2	3
arglinc	1	2	2	0	1	69	1	2	3	1	2	3
bard	1	7	7	1	19	24	1	19	19	1	21	26
bdqrtic	1	10	10	0	123	317	0	37	102	0	15	58
beale	1	9	10	1	14	18	1	12	12	1	12	12
biggs6	1	38	51	1	72	81	1	34	35	1	60	62
box3	1	8	9	1	20	21	1	21	21	1	19	19
brownal	1	7	7	1	9	15	1	9	13	1	10	13
brownbs	1	8	8	1	11	28	1	11	27	1	13	16
brownden	1	8	8	0	20	123	0	17	43	1	13	13
chainwoo	1	90	145	2	200	834	2	200	238	2	200	269
chnrosnb	1	52	92	1	173	327	1	171	187	1	76	104
cube	1	28	36	1	27	44	1	40	50	1	34	40
deconvu	1	44	50	1	103	157	1	50	65	1	41	46
denschnb	1	6	12	1	8	10	1	7	8	1	7	8
denschnc	1	11	12	1	16	22	1	18	21	1	12	12
denschnd	1	35	52	1	80	116	1	85	89	1	88	90
denschne	1	10	11	1	24	39	1	22	33	1	22	33
denschnf	1	6	6	1	11	26	1	8	8	1	7	7
eigenals	1	23	27	1	97	135	1	108	111	1	104	107
eigenbls	1	119	155	2	200	331	2	200	213	1	170	172
engval2	1	16	20	1	31	46	1	33	38	1	29	34
errinros	1	29	40	2	200	409	2	200	207	1	80	100
expfit	1	6	9	1	11	18	1	12	17	1	12	17
extrosnb	1	0	0	1	0	0	1	0	0	1	0	0
genhumps	1	47	93	1	77	99	1	101	145	1	90	156
genrose	2	200	302	2	200	464	2	200	292	2	200	302
growthls	1	75	98	1	1	9	0	120	196	1	99	107
gulf	1	21	27	1	28	39	1	30	33	1	30	36
hatfldd	1	20	21	1	27	34	1	30	33	1	29	31
hatflde	1	19	21	1	27	37	1	28	31	1	26	30
heart6ls	2	200	252	2	200	298	2	200	252	2	200	209
heart8ls	1	74	99	2	200	279	2	200	233	2	200	202
helix	1	10	12	1	28	35	1	24	27	1	23	26
himmelbf	1	10	10	0	30	87	0	31	85	1	30	38
humps	2	200	506	1	86	213	1	193	466	2	200	486
kowosb	1	8	18	1	28	32	1	27	29	1	27	27
mancino	1	5	5	1	125	251	1	33	37	1	13	17
meyer3	2	200	267	2	200	298	2	200	252	2	200	208
msqrtals	1	33	57	2	200	499	2	200	205	2	200	203
msqrtbls	1	28	39	2	200	509	2	200	205	2	200	203
nonmsqrt	1	137	189	2	200	260	0	148	191	0	141	191
osbornea	0	0	0	0	0	0	0	0	0	0	0	0
osborneb	1	19	22	1	53	72	1	50	62	1	63	71
palmer5c	1	1	1	1	6	12	1	13	13	1	13	13
palmer6c	1	1	1	1	43	48	1	43	44	1	43	44
palmer7c	1	1	1	1	40	44	1	39	40	1	39	40
palmer8c	1	1	1	1	41	45	1	40	41	1	40	41
penalty1	1	40	42	2	200	266	1	71	78	1	54	55
penalty2	1	19	20	0	0	0	0	60	115	0	29	75
pfit1ls	2	200	268	2	200	267	2	200	257	1	135	145
pfit2ls	1	119	160	2	200	275	1	72	94	1	43	51
pfit3ls	2	200	271	0	0	0	2	200	251	2	200	206
pfit4ls	2	200	278	0	0	0	2	200	254	2	200	212
rosenbr	1	22	28	1	36	52	1	27	33	1	27	30
s308	1	9	10	1	14	18	1	14	14	1	11	11
sensors	1	30	54	2	200	636	1	43	53	1	32	38
sineval	1	44	61	1	69	100	1	55	70	1	53	61
watson	1	12	12	1	61	80	1	56	56	1	53	53
yfitu	1	38	48	1	63	86	1	50	63	1	46	50

Table 3: CUTEr results of the six methods for Experiment 2.

	Hes			$BFGS$			$SBFGS^M$			$SubBFGS^M$			$SBFGS^P$			$SubBFGS^P$		
Prob	Stat	#Iter	#Fun	Stat	#Iter	#Fun	Stat	#Iter	#Fun	Stat	#Iter	#Fun	Stat	#Iter	#Fun	Stat	#Iter	#Fun
arglina	1	5	5	1	5	7	1	6	6	1	5	5	1	6	6	1	5	5
arglinb	1	3	3	1	2	3	1	8	8	0	17	26	1	8	8	0	17	26
arglinc	1	4	4	1	4	6	1	9	9	1	15	18	1	9	9	1	15	18
bard	1	7	7	1	14	19	1	11	11	1	14	14	1	7	7	1	8	8
bdqrtic	1	10	10	0	111	308	0	47	101	1	10	10	1	13	13	1	10	10
beale	1	7	8	1	9	13	1	8	11	1	8	11	1	7	10	1	7	10
biggs6	1	17	20	1	45	58	1	39	46	1	20	30	1	40	45	1	18	27
box3	1	5	5	1	9	11	1	8	8	1	6	6	1	9	9	1	7	7
brownal	1	15	16	1	15	21	1	11	11	1	15	15	1	11	12	1	15	15
brownbs	1	8	8	1	16	33	1	17	32	1	17	32	1	12	13	1	12	13
brownden	1	8	8	0	22	101	1	24	30	1	14	14	1	9	9	1	8	8
chainwoo	1	87	136	2	200	841	2	200	223	1	70	115	1	141	186	1	84	128
chnrosnb	1	54	94	1	171	321	2	200	207	1	53	95	1	57	78	1	56	102
cube	1	21	25	1	17	27	1	28	34	1	28	34	1	23	31	1	23	31
deconvu	1	16	16	1	88	138	1	137	137	1	22	22	1	144	144	1	25	25
denschnb	1	7	9	1	7	9	1	8	8	1	8	8	1	7	9	1	7	9
denschnc	1	10	10	1	17	24	1	17	19	1	17	19	1	10	10	1	10	10
denschnd	1	32	45	1	78	118	1	74	81	1	46	75	1	52	56	1	46	55
denschne	1	9	10	1	25	41	1	18	28	1	13	16	1	16	24	1	11	14
denschnf	1	6	6	1	12	27	1	9	9	1	9	9	1	6	6	1	6	6
eigenals	1	23	27	1	97	135	1	109	112	1	24	28	1	103	104	1	24	28
eigenbls	1	60	76	2	200	329	2	200	203	1	56	68	2	200	200	1	57	71
engval2	1	14	18	1	28	43	1	28	31	1	20	24	1	28	32	1	27	31
errinros	1	22	36	2	200	414	2	200	211	1	25	38	1	25	34	1	26	39
expfit	1	6	9	1	11	19	1	10	14	1	10	14	1	10	14	1	10	14
extrosnb	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
growthls	1	54	71	1	32	136	1	79	90	2	200	309	1	69	82	1	100	123
gulf	1	11	26	1	18	42	1	18	28	1	19	26	1	18	32	1	15	26
hatfldd	1	9	10	1	21	29	1	19	22	1	18	27	1	12	15	1	17	26
hatflde	1	12	18	1	16	22	1	20	22	1	20	24	1	19	20	1	27	34
heart8ls	1	61	108	2	200	282	2	200	232	1	113	174	1	127	157	1	74	135
helix	1	10	12	1	27	35	1	29	30	1	28	33	1	27	28	1	25	28
himmelbf	0	9	49	1	29	45	1	32	38	1	12	17	0	28	87	1	9	12
kowosb	1	5	6	1	16	21	1	15	16	1	5	8	1	15	19	1	5	8
mancino	1	5	5	1	127	251	1	16	16	1	7	7	1	5	5	1	5	5
msqrtals	1	28	40	2	200	503	2	200	237	1	30	39	2	200	206	1	31	41
msqrtbls	1	26	38	2	200	512	2	200	234	1	27	38	2	200	206	1	28	40
nonmsqrt	1	140	194	2	200	260	0	120	149	1	129	184	0	158	182	1	127	175
osborneb	1	17	22	1	52	70	1	64	68	1	15	19	1	52	55	1	16	19
palmer5c	1	3	3	1	11	17	1	8	8	1	7	7	1	8	8	1	7	7
palmer6c	1	4	8	1	41	51	1	40	46	1	11	19	1	40	46	1	11	19
palmer7c	1	4	10	1	38	51	1	38	45	1	11	19	1	38	45	1	11	19
palmer8c	1	4	9	1	40	51	1	39	44	1	12	22	1	39	44	1	12	22
penalty1	1	40	42	2	200	301	0	94	162	1	54	56	1	45	49	1	40	42
penalty2	1	19	20	0	0	0	0	59	114	1	38	39	1	29	30	1	19	20
pfit2ls	1	25	38	1	97	143	1	43	55	1	102	122	1	47	59	1	105	124
pfit3ls	1	42	61	1	188	268	1	53	68	1	100	121	1	97	119	1	133	168
pfit4ls	1	60	87	0	0	0	1	74	93	1	122	154	1	88	115	1	177	208
rosenbr	1	17	20	1	28	43	1	20	24	1	20	24	1	24	27	1	24	27
s308	1	8	8	1	12	20	1	11	12	1	11	12	1	8	8	1	8	8
sensors	1	30	56	2	200	685	0	78	118	1	34	61	0	60	131	1	34	61
sineval	1	16	26	1	54	73	1	41	49	1	40	62	1	38	51	1	38	51
watson	1	11	11	1	56	74	1	53	54	1	23	24	1	46	46	1	14	14
yfitu	1	37	47	1	60	84	1	45	52	1	50	71	1	48	53	1	37	49

REFERENCES

[1] Shrirang Abhyankar, Vishwas Rao, and Mihai Anitescu, *Dynamic security constrained optimal power flow using finite difference sensitivities*, in PES General Meeting— Conference & Exposition, 2014 IEEE, IEEE, 2014, pp. 1–5.

[2] Keyvan Amini and Ashraf Ghorbani Rizi, *A new structured quasi-Newton algorithm using partial information on Hessian*, Journal of Computational and Applied Mathematics, 234 (2010), pp. 805–811.

[3] C. G. Broyden, J. E. Dennis Jr, and J. J. Moré, *On the local and superlinear convergence of quasi-newton methods*, IMA Journal of Applied Mathematics, 12 (1973), pp. 223–245.

[4] Richard H. Byrd and Jorge Nocedal, *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM Journal on Numerical Analysis, 26 (1989), pp. 727–739.

[5] Richard H. Byrd, Jorge Nocedal, and Robert B. Schnabel, *Representations of quasi-newton matrices and their use in limited memory methods*, Mathematical Programming, 63 (1994), pp. 129–156.

[6] J. E. Dennis Jr., H. J. Martinez, and R. A. Tapia, *Convergence theory for the structured BFGS secant method with an application to nonlinear least squares*, Journal of Optimization Theory and Applications, 61 (1989), pp. 161–178.

[7] J. E. Dennis Jr. and J. Moré, *A characterization of superlinear convergence and its application to quasi-Newton methods*, Mathematics of Computation, 28 (1974), pp. 549–560.

[8] J. E. Dennis Jr. and H. F. Walker, *Convergence theorems for least-change secant update methods*, SIAM Journal on Numerical Analysis, 18 (1981), pp. 949–987.

[9] Elizabeth D. Dolan and Jorge J. Moré, *Benchmarking optimization software with performance profiles*, Mathematical Programming, 91 (2002), pp. 201–213.

[10] R. Fletcher, *An optimal positive definite update for sparse Hessian matrices*, SIAM Journal on Optimization, 5 (1995), pp. 192–218.

[11] R. Fourer, D. M. Gay, and B. Kernighan, *Algorithms and model formulations in mathematical programming*, Springer-Verlag New York, Inc., New York, NY, USA, 1989, ch. AMPL: A Mathematical Programming Language, pp. 150–151.

[12] N. I. M. Gould, D. Orban, and P. L. Toint, *CUTEr and SifDec: A constrained and unconstrained testing environment, revisited*, ACM Trans. Math. Softw., 29 (2003), pp. 373–394.

[13] A. Griewank and Ph. L. Toint, *Local convergence analysis for partitioned quasi-Newton updates*, Numerische Mathematik, 39 (1982), pp. 429–448.

[14] O. Güler, F. Gurtuna, and O. Shevchenko, *Duality in quasi-Newton methods and new variational characterizations of the DFP and BFGS updates*, Optimization Methods Software, 24 (2009), pp. 45–62.

[15] J. Huschens, *Exploiting additional structure in equality constrained optimization by structured SQP secant algorithms*, Journal of Optimization Theory and Applications, 77 (1993), pp. 343–358.

[16] L. Lukšan and E. Spedicato, *Variable metric methods for unconstrainted optimization and nonlinear least squares*, J. Comput. Appl. Math., 124 (2000), pp. 61–95.

[17] Jorge J. Moré and David J. Thuente, *Line search algorithms with guaranteed sufficient decrease*, ACM Trans. Math. Softw., 20 (1994), pp. 286–307.

[18] Tony B Nguyen and MA Pai, *Dynamic security-constrained rescheduling of power systems using trajectory sensitivities*, IEEE Transactions on Power Systems, 18 (2003), pp. 848–854.

[19] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 2nd ed., 2006.

[20] Richard Tapia, *On secant updates for use in general constrained optimization*, Mathematics of Computation, 51 (1988), pp. 181–202.

[21] A. Wächter and L. T. Biegler, *On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming*, Mathematical Programming, 106 (2006), pp. 25–57.

[22] Philip Wolfe, *Convergence conditions for ascent methods*, SIAM Review, 11 (1969), pp. 226–235.

[23] ———, *Convergence conditions for ascent methods, II: Some corrections*, SIAM Review, 13 (1971), pp. 185–188.