Super linear Projected Structured Exact Penalty Secant Methods for Constrained Nonlinear Least Squares

Nezam Mahdavi-Amiri[*] and Narges Bidabadi

Faculty of Mathematical Sciences, Sharif University of Technology, Tehran, Iran
e-mails: nezamm@sina.sharif.edu; n_bidabadi@mehr.sharif.ir

Abstract

We present an exact penalty approach for solving constrained nonlinear least squares problems, using a new projected structured Hessian approximation scheme. We establish general conditions for the local two-step Q-superlinear convergence of our given algorithm. The approach is general enough to include the projected version of the structured PSB, DFP and BFGS formulas as special cases. The numerical results obtained by testing an implementation of our algorithm, as compared to existing competitive algorithms for nonlinear programs, confirm the efficiency and robustness of the proposed algorithm.

Keywords: Constrained nonlinear least squares; Exact penalty methods; Projected Hessian updating; Structured secant updates; Two-step superlinear convergence

1. Introduction

Consider the constrained nonlinear least squares (CNLLS) problem,

$$\begin{aligned} &\min \emptyset(x) = \tfrac{1}{2}F(x)^T F(x) \\ &\text{s.t. } c_i(x) = 0, i = 1, \dots, k \\ &c_i(x) \geq 0, i = k+1, \dots, k+m, \end{aligned} \tag{1}$$

where, $F(x) = [f_1(x), \dots, f_l(x)]^T, x \in R^n$, $f_\delta, 1 \leq \delta \leq l$, and $c_i, i = 1, \dots, k+m$, are functions from R^n to R, all assumed to be twice continuously differentiable. The gradient and Hessian of $\emptyset$ can be expressed as $\nabla\emptyset(x) = G(x)F(x)$, and $\nabla^2\emptyset(x) = G(x)G(x)^T + S(x)$, where, $G(x)$is the matrix whose columns are the gradients $\nabla f_\delta(x)$, and $S(x) = \sum_{\delta=1}^{l} f_\delta(x)\, \nabla^2 f_\delta(x)$.

An exact penalty function for nonlinear programming problems is defined to be

$$\psi(x,\mu) = \mu\phi(x) + \sum_{i=1}^{k}|c_i(x)| - \sum_{i=k+1}^{k+m}\min(0, c_i(x)), \tag{2}$$

where, $\mu > 0$ is a penalty parameter. If x^* is a stationary point of (1) and the gradients of the active constraints at x^* are linearly independent, then there exists a real number $\mu^* > 0$ such that x^* is also a stationary point of $\psi(x,\mu)$, for each $0 < \mu \leq \mu^*$. Mahdavi-Amiri and Bartels [10], based on Coleman and Conn's exact penalty approach for general nonlinear problems [3], proposed a special algorithm for minimization of y with a fixed value of

μ in solving the CNLLS problems. Let ε be a small positive number used to identify the near-active (ε- active) constraint set. The algorithm obtains search directions by using an ε-active set

[*] corresponding author

$AC(x,\varepsilon) = \{i : |c_i(x)| \le \varepsilon, 1 \le i \le k+m\}$,
and a corresponding merit function:

$$\psi_\varepsilon(x,\mu) = \mu\phi(x) + \sum_{i\in VE(x,\varepsilon)} sgn\left(c_i(x)\right)c_i(x) - \sum_{i\in VI(x,\varepsilon)} c_i(x) \quad (3)$$

where, $VE(x,\varepsilon) = \{i : |c_i(x)|j > \varepsilon, 1 \le i \le k\}$,
$VI(x,\varepsilon) = \{i : c_i(x) < -\varepsilon, k+1 \le i \le k+m\}$.
Step lengths are chosen by using ψ, which is ψ_ε with $\varepsilon = 0$. The optimality conditions are checked by using ψ, as well.

The gradient and Hessian of ψ_ε are:

$$\nabla\psi_\varepsilon(x,\mu) = \mu\nabla\square(x) + E(x)\sigma(x) - I(x)e$$
$$= \mu G(x)F(x) + E(x)\sigma(x) - I(x)e,$$

$$\nabla^2\psi_\varepsilon(x,\mu) =$$
$$\mu(G(x)G(x)^T + S(x) + \sum_{i\in VE(x,\varepsilon)} sgn\left(c_i(x)\right)\nabla^2 c_i(x) - \sum_{i\in VI(x,\varepsilon)} \nabla^2 c_i(x)$$

where,

$$e = [1 \dots 1 \dots]^T,$$
$$A(x) = [\dots \nabla c_i(x) \dots]_{i\epsilon AC(x.\varepsilon)},$$
$$E(x) = [\dots \nabla c_i(x) \dots]_{i\epsilon VE(x.\varepsilon)},$$
$$I(x) = [\dots \nabla c_i(x) \dots]_{i\epsilon VI(x.\varepsilon)},$$

and

$$\sigma(x) = [\dots sgnc_i(x) \dots]^T{}_{i\epsilon VE(x.\varepsilon)}.$$

It is well known that the necessary conditions for x^* to be an isolated local minimizer for ψ, under the assumptions made above on ϕ and the c_i, are that there exist multipliers, ${\lambda^*}_i$, for $i \epsilon AC(x^*,0)$, such that

$$\psi_0(x^*,\mu) = \sum_{i\epsilon AC(x^*,0)} {\lambda_i}^* \nabla c_i(x^*) = A(x^*)\lambda^*, \quad (4)$$
$$-1 \le {\lambda_i}^* \le 1, i\epsilon AC(x^*, 0) \cap \{1,\dots,k\},$$
$$0 \le {\lambda_i}^* \le 1, i\epsilon AC(x^*, 0) \cap \{k+1,\dots,k+m\}. \quad (5)$$

A point x for which only (4) above is satisfied is called a stationary point of ψ. A minimizer x^* must be a stationary point satisfying (5).

One major premise of the algorithm is that the multipliers are only worth estimating in the neighborhoods of stationary points. Nearness to a stationary point is governed by a stationary tolerance $\tau > 0$. The algorithm is considered to be in a local state, if norm of the projected or reduced gradient, i.e., $\|Z^T\nabla\psi_\varepsilon(x,\mu)\|$, is smaller than this tolerance, and it is in global state, otherwise.
Fundamental to the approach is the following quadratic problem:

$$min\ \nabla\psi_\varepsilon(x,\mu)^T h + \frac{1}{2}h^T H h$$
$$s.t. \nabla c_i(x)^T h = 0, i \in AC(x,\varepsilon), \quad (6)$$

where, $H = \nabla^2\psi_\varepsilon(x,\mu) - \sum_{i\epsilon AC(x,\varepsilon)} \lambda_i \nabla^2 c_i(x)$, the λ_i are the Lagrange multipliers assosiated with (6) in a local state (in the proximity of a stationary point) and the λ_i are taken to be zero when the algorithm is in its global state (far from stationary points).

Using the QR decomposition of $A(x) = Q[R^T\ 0]^T = [Y\ Z][R^T\ 0]^T$, where, Q is an orthonormal matrix and R is a square upper triangular matrix (being nonsingular when A(x) has full column rank), if we set $h = Zw$, for some $w \epsilon R^n$, then w is to be found by solving $Z^T HZw = -Z^T \nabla\psi_\varepsilon(x,\mu)$. Therefore, projected or reduced Hessians, the $Z^T HZ$, are needed for solving the quadratic problems (6).

Using the inherent structure of the Hessian matrix, we are to present a general approach for computing an effective projected structured secant approximation B_z to $Z^T HZ$, by providing a quasi-Newton approximation $A_z \approx Z^T\, S(x,\lambda)Z$, where,

$$S(x,\lambda) = \mu S(x) + \sum_{i\epsilon VE(x,\varepsilon)} sgn(c_i(x))\ \nabla^2 c_i(x) - \sum_{i\in VI(x,\varepsilon)} \nabla^2 c_i(x) - \sum_{i\in AC(x,\varepsilon)} \lambda_i \nabla^2 c_i(x)\ ,$$

with the setting $B_z = \mu Z^T G(x) G(x)^T Z + A_z$, and establish general conditions for an asymptotic two step superlinear convergence. Moreover, as special cases, we present general exact penalty methods using projected structured DFP, BFGS and PSB update schemes.

Consider the asymptotic case. Assume that the final active set has been identified, so that for all subsequent iterations k, with x^* designating the optimal point, we have $AC(x_k,\varepsilon) = AC(x^*,0)$, $VE(x_k,\varepsilon) = VE(x^*,0)$ and $VI(x_k,\varepsilon) = VI(x^*,0)$. Suppose that $A_{z,k} \approx Z^T{}_k\, S(x_k,\lambda_k)Z_k$, and we want to update $A_{z,k}$ to $A_{z,k+1}$ approximating $A_{z,k+1} \approx Z^T{}_{k+1}\, S(x_{k+1},\lambda_{k+1})Z_{k+1}$. Note that $A_{z,k}, A_{z,k+1} \epsilon\ R^{n-t,n-t}$, where t is the number of active constraints. We assume rank $(A(x_k)) = t$, for all k. Letting $s = \bar{Z}^T(\bar{x} - x)$ and $q = \bar{Y}^T(\bar{x} - x)$, we have $\bar{x} - x = \bar{Y}\,q + \bar{Z}s$, where, in order to simplify notation, the presence of a bar above a quantity indicates that it is taken at iteration $k+1$, and the absence of a bar indicates iteration k. If the constraints are linear, then $q = 0$, for $k > 0$. In the nonlinear case, asymptotically it is expected that $\bar{Y}\,q$ be negligible and thus we obtain [10]

$$\bar{Z}^T S(\bar{x},\bar{\lambda})\bar{Z}\, s \approx \bar{Z}^T [\mu(\bar{G} - G)\bar{F} + (\bar{E} - E)\bar{\sigma} - (\bar{I} - I)e + A\lambda] = y. \tag{7}$$

Thus, we use the quasi-Newton formula to update A_z to $\bar{A}_z$ according to the secant equation $\bar{A}_z s = y$, only if $\bar{Y}q$ has actually become negligible, that is, when

$$||q|| < \frac{\eta}{(k+1)^{1+v}} ||s||, \tag{8}$$

for $v = 0.01$, where, k is the iteration number and $||.||$ denotes the Euclidean norm as suggested by Nocedal and Overton in [11] for the general constrained nonlinear programs. Therefore, an approximation to $\bar{Z}^T \bar{H}\, \bar{Z}$ is $\bar{B}_z = \bar{A}_z + Q(\bar{x})$, where, $Q(\bar{x}) = \mu\bar{Z}^T G(\bar{x})G(\bar{x})^T\bar{Z}$.

We have recently established sufficient conditions for the two-step superlinearly convergence of the exact penalty algorithm for solving the CNLLS problems, when the projected structured secant update formulas are used for approximating the projected Hessian.

The remainder our work is organized as follows. In §2, we summarize an exact penalty algorithm for minimization of (2). In § 3, we briefly review structured secant update formulas. In § 4, we first give the sufficient conditions for a two-step superlinearly convergence of the algorithm and show that the projected structured BFGS update formula used in [2] satisfies these conditions. Then, we prove that these conditions also hold for our projected structured PSB and DFP update formulas, offered here, and therefore establish their asymptotic two-step superlinear convergence rate. Competitive numerical results are reported in § 5. The results are compared with the ones obtained by the f mincon function in MATLAB's toolbox and three algorithms in the KNITRO software package for solving general nonlinear programs. We conclude in § 6.

2. Exact Penalty Algorithm

A general exact penalty algorithm for minimizing (2) is given below; see [10].

Algorithm 1: An Exact Penalty Method.
Step 0: Give an initial point x, $\varepsilon > 0$ and $\tau > 0$.
Step 1: Determine $AC(x,\varepsilon)$, $VE(x,\varepsilon)$, $VI(x,\varepsilon)$, $A(x)$. Identify the matrix Z by computing the QR decomposition of $A(x)$ and set global = true, optimal = false.
Step 2: If $\|Z^T\nabla\psi_\varepsilon(x,\mu)\| > t$, then obtain the search direction d that is the solution to the quadratic problem (6), where, H approximates the global Hessian $\nabla^2\psi_\varepsilon(x,\mu)$, and go to Step 5.
Step 3: Determine the Lagrange multipliers $\lambda_i, i\epsilon\ AC(x,\varepsilon)$, as a solution of

$$min\|A(x)\lambda - \nabla\psi_\varepsilon(x,\mu)\|_2. \quad (9)$$

If conditions (5) are not satisfied, then choose an index r for which one of (5) is violated, and determine the search direction d that satisfies the system of equations $A(x)^T d = \sigma_r e_r$, where, e_r is the rth unit vector, and $\sigma_r = -sgn\lambda_r$, and go to Step 5.
Step 4: Set global = false. Determine the direction $d = h_A + v$, where, h_A is the solution to the quadratic problem (6), H approximates the local Hessian $\nabla^2\psi_\varepsilon(x,\mu) - \sum_{i\epsilon AC(x,\varepsilon)}\lambda_i\nabla^2 c_i(x)$, with the λ_i being the Lagrange multipliers assosiated with (6) as determined in Step 3. The vertical direction, v, is the solution to the system

$$A(x)^T v = -c_{AC(x,\varepsilon)}(x + h_A), \quad (10)$$

where, $c_{AC(x,\varepsilon)}$ is the vector of the constraint functions, ordered in accordance with the columns of A(x). Set the step length $\alpha = 1$ and go to Step 6.
Step 5: Determine step length a using a line search procedure on $\psi(x,\mu)$.
Step 6: Compute $\bar{x} = x + \alpha d$. If a sufficient decrease has been obtained, then set $x = \bar{x}$, else go to Step 8.
Step 7: If global = false, then check the optimality conditions for x. If x is optimal, then set optimal = true and stop, else go to Step 1.
Step 8: If global = true and $AC(x,\varepsilon) \neq AC(x;0)$, then reduce ε to change $AC(x,\varepsilon)$, else reduce τ so that $\|Z^T\nabla\psi_\varepsilon(x,\mu)\|$ becomes large tested against τ.
Step 9: If (global=true and $AC(x,\varepsilon) = AC(x,0)$) or ε is too small or τ is too small, then report failure and stop, else go to Step 1.

Remarks: *In Step 6, we assume there exists a line search strategy to determine the step length satisfying a sufficient decrease in y that is characterized by the line search assumption (see* [4], *P. 152 part (v)). In Step 7, the optimality conditions are checked as follows:*
If $\|Z^T \nabla\psi_0(x,\mu)\|$ *is small enough, determine the multipliers* λ_i, $i \in AC(x,0)$, *as the least squares solution of (9). If the conditions (5) are satisfied, then x is considered to satisfy the first order conditions, and thus x being a stationary point, the algorithm is stopped, as commonly practiced in optimization algorithms. Of course, second order conditions are needed to be checked to ascertain optimality of x.*

In the remainder of our work, we drop the index z from A_z and B_z, for simplicity.

3 Structured Secant Update Formulas

For the (unstructured) secant schemes, we have the secant equation $\bar{B}s = y$:Most interesting secant update formulas can be written in the form

$$\bar{B} = B + \Delta(s,y,B,v), \quad (11)$$

where,

$$\Delta(s,y,B,v) = \frac{(y-Bs)v^T + v(y-Bs)^T}{v^T s} - \frac{(y-Bs)^T s}{(v^T s)^2} vv^T, \quad (12)$$

for some choices of the vector v. The vector v will usually depend on s, y, or B, as is the case for the following well-known updates:

$$(BFGS): v = y + q\sqrt{\frac{s^T y}{s^T Bs}} Bs,$$
$$(PSB): v = s,$$
$$(DFP): v = y.$$

In what follows, we write $v(s,y,B)$ when the use of v alone may cause confusion. For our structured case here, we have the secant equation $\bar{A}s = y$. By setting $y^s = y + Q(\bar{x})s$ and $B^s = A + Q(\bar{x})$, we obtain the structured secant update of A as[5]

$$\bar{A} = A + \Delta\big(s,y,A,v(s,y^s,B^s)\big). \quad (13)$$

Therefore, for the structured BFGS, PSB and DFP update formulas, we set

$$(BFGS): v = y^s + q\sqrt{\frac{s^T y^s}{s^T B^s s}} B^s s,$$
$$(PSB): v = s,$$
$$(DFP): v = y^s. \quad (14)$$

Remark: Clearly, $\bar{A}$, *given by (13), satisfies the secant equation* $\bar{A}s = y$.

In [2], we used the projected structured BFGS formula and proved the local two-step superlinear convergence rate of the algorithm. Here, we give the results for a two-step superlinear convergence of the algorithm under a more general setting so that in addition to the BFGS scheme, the PSB and DFP schemes given here will also posses the superlinear local convergence property.

4 Local Convergence

We give a characterization of a local two-step superlinear convergence for the algorithm. To this end, we need the following usual assumptions.

Assumptions:
(A1) Problem (1) has a local solution x^*. Let $D_1 = \{x : \|x - x^*\| \le \varepsilon_1\}$.
(A2) $\{x_k\}$, generated by Algorithm 1 for minimizing $\psi(x, \mu)$, is so that $x_k \epsilon D_1$, for all k.
(A3) The function $\emptyset$ and c_i, $i = 1, \dots, k+m$, are twice continuously differentiable on the compact set D_1.
(A4) $\nabla^2 \emptyset(x)$ and $\nabla^2 c_i$(x) are locally Lipschitz continuous at x^*.
(A5) $Q(x)$ is locally Lipschitz continuous at x^*, that is, there exist a constant $L \ge 0$ such that $\|Q(x) - Q(x^*)\| \le L\|x - x^*\|$, for all $x \in D_1$.
(A6) The gradients of the active constraints at x_k, for all k, are linearly independent on D_1.
(A7) There exist positive constants b_1 and b_2 such that

$$b_1\|z\|^2 \le z^T B^* z \le b_2\|z\|^2 \quad \text{for all } z \in R^n, \tag{15}$$

where, B^* is the projected Hessian at x^*.

We will make use of the following two theorems from [8].

Theorem 1. [8] *Suppose that Algorithm 1 is applied with any update rule for approximating the projected Hessian matrix. Let τ_1 be a given constant. There exists $\varepsilon_2 > 0$ such that for any iteration k, if $\|B^{-1}{}_{k-1}\| \le \tau_1$, $\|B^{-1}{}_k\| \le \tau_1$ and $\|e_{k-1}\| \le \varepsilon_2$, then*

(i) $\|e_k\| \le C_1\|e_{k-1}\|$,
(ii) $\|Y^T{}_{k+1}(x_{k+1} - x_k)\| \le C_1(\|e_k\|^2 + \|e_{k-1}\|^2)$,
(iii) $\|e_{k+1}\| \le C_1\|e_{k-1}\|^2 + \|(B_k - B^*)Z^T{}_{k+1}e_k\|$,
(iv) $\|e_{k+1}\| \le C_1\|e_{k-1}\|^2 + \|(B_k - B^*)Z^T{}_{k+1}(x_{k+1} - x_k)\|$,

where, $e_k = x_k - x^*$, C_1 is a constant which depends on τ_1 and ε_2 but not on k (the assumption that $\|B^{-1}{}_k\| \le \tau_1$ is needed only for (iii)).

Theorem 2. [8] *There exists $\varepsilon_3 > 0$ such that if $\|e_k\| \le \varepsilon_3$ and $\|e_{k+1}\| \le \varepsilon_3$, then*

$\|y_k - A^*s_k\| \le C_2 max(\|e_k\|, \|e_{k+1}\|\|s_k\| + \|Z^*B^*Y^*q_k\|$,
where, $C_2 \ge 1$ is a constant independent of k, and if equation (8) holds, then

$$\|y_k - A^*s_k\| \le \gamma_k\|s_k\|, \tag{16}$$

where,

$$\sigma_k \le \gamma_k \le C_2(1+\eta)\sigma_k + \eta\frac{\|Z^*B^*Y^*\|}{(1+k)^{1+v}}, \tag{17}$$

and $\sigma_k = max(\|e_k\|, \|e_{k+1}\|)$.

Next, we give a two-step linear convergence result and specify conditions for a two-step superlinear convergence.

Theorem 3. *Suppose that Assumptions (A1)-(A7) hold. Let the sequence* $\{x_k\}$ *be generated by Algorithm 1 and let the* B_k *be obtained by any structured secant update rule for approximating the projected Hessian, that is,*

$$B_k = A_k + Q(x_k),$$

where, A_k *is the update formula for* A_{k-1} *obtained by using (13). Furthermore, suppose there exist positive constants* d_1 *and* d_2 *such that*

$$\|A_{k+1} - A^*\|_W \leq (1 + d_1\gamma_k)\|A_k - A^*\|_W + d_2\gamma_k, \quad (18)$$

for all $x_k, x_{k+1} \in \bar{D} = \{x : \|x - x^*\| \leq \varepsilon_4\}$, *where,* W *is a positive definite matrix and* $\|.\|_W = \|W^{\frac{-1}{2}}(.)W^{\frac{-1}{2}}\|_F$. *For any* $r \in (0,1)$, *there exist positive constants* ε *and* δ *such that if* $\|e_0\| \leq \varepsilon$ *and* $\|A_0 - A^*\|_W \leq \delta$, *then*

$$\|e_{k+1}\| \leq r\|e_{k-1}\|, k \geq 1, \quad (19)$$

that is, $\{x_k\}$ *converge to* x^*, *at least at a two-step linear rate.*
Moreover, if

$$lim_{k\to\infty} \frac{\|(A_k - A^*)s_k\|}{\|x_{k+1} - x_k\|} = 0, \quad (20)$$

then $x_k \to x^*$, *at a two-step superlinear rate.*

Corollary 4. *Suppose that Assumptions (A1)-(A7) hold. Let the sequence* $\{x_k\}$ *be generated by Algorithm 1 and let the* B_k *be obtained by any structured secant update rule for approximating the projected Hessian, that is,*

$$B_k = A_k + Q(x_k),$$

where, A_k *is the update formula for* A_{k-1} *obtained by using (13). Furthermore, suppose the deterioration inequality (18) holds. Then, we have* $\sum_{k\geq 0}\sigma_k < \infty$, *and*

$$\sum_{k\geq 0}\gamma_k < \infty \quad (21)$$

In §1, we pointed out that we use the structured quasi-Newton update formula for approximating the projected Hessian in Algorithm 1, if the inequality (8) holds (this is expected to happen when the algorithm is in its local phase with the iterate being close to a stationary point). Here, we give the superlinear convergence result for Algorithm 1. It can be proved that Algorithm 1 has a local two-step superlinear convergence rate, if the structured quasi-Newton update formula being used satisfies inequality (18) and conditions (20). It can also be shown that if the inequality (8) holds, then the structured BFGS, PSB and DFP update formulas satisfy inequality (18) and conditions (20). Hence, Algorithm 1 with these three structured update formulaes has a two-step local superlinear convergence rate.

Theorem 5. *Suppose that Assumptions (A1)-(A7) hold, let the sequence* $\{x_k\}$ *be generated by Algorithm 1 and* B_k *be obtained by*

$B_k = A_k + Q(x_k)$, *if inequality (8) holds,*
B_{k-1}, *otherwise,*

where, A_k *is the secant update formula of* A_{k-1} *obtained by using (13). Furthermore, suppose that the deterioration inequality (18) and conditions (20) hold. Then,* $\{x_k\}$ *converges to* x^* *with at least a twostep superlinear rate.*

In [2], we proved a local two-step superlinear convergence for the algorithm, in the special case of the projected Hessian being updated using a projected structured BFGS update formula. Here, we give the result that the structured BFGS update formula, as a special case, satisfies the general conditions of Theorem 5.

Lemma 6. *If* $\|e_k\| \leq \varepsilon_3$, $\|e_{k+1}\| \leq \varepsilon_3$, *the inequality (8) and Assumption (A5) hold, then there exists a positive constant* $\acute{L} > 0$ *such that*

$$\frac{\|{y^s}_k - B^* s_k\|}{\|s_k\|} \leq \acute{L}\gamma_k \tag{22}$$

Lemma 7. *Suppose that the inequality (8) and Assumptions (A5) and (A7) hold. Then, there exist positive constants* ε_4 *and* v *such that*

$${s^T}_k {y^s}_k \geq \varpi \|s_k\|^2 \text{ and } \frac{\|{y^s}_k\|\|s_k\|}{|{s^T}_k {y^s}_k|} \leq \frac{\|B^*\| + \acute{L}\gamma_k}{\varpi}$$

where, $x_k, x_{k+1} \in D_2 = \{x : \|x - x^*\| \leq \varepsilon_4\} \subseteq D_1$.

Theorem 8. *Suppose that the assumptions of Lemma 7 hold. Let* $\bar{B}$ *be the (unstructured) BFGS secant update,*

$$\bar{B} = B + \Delta\left(s, y, B, y + \sqrt{\frac{y^T s}{s^T B s}}\right).$$

Then, the bounded deterioration inequality

$$\|\bar{B} - B^*\|_{B^*} \leq \|B - B^*\|_{B^*} + \alpha_p \frac{\|y - B^* s\|}{\|s\|}$$

holds, where, $x, \bar{x} \in D_2$ *and* $\|.\|_M = \left\|M^{\frac{-1}{2}}(.)M^{\frac{-1}{2}}\right\|_F$,

for any positive definite matrix M.

Theorem 9. *Suppose that the inequality (8) and Assumptions (A5) and (A7) hold. Let* $\bar{B}$ *be the projected structured BFGS secant update, that is,* $\bar{B} = \bar{A} + Q(\bar{x})$,
where,

$$\bar{A} = A + \Delta\left(s, y, A, y^s + \sqrt{\frac{s^T y^s}{s^T B^s s}}\right).$$

Then, there exists a positive constant α_B *and a neighborhood* $D_2 \subseteq D_1$ *such that*

$$\|\bar{A} - A^*\|_{B^*} \leq \|A - A^*\|_{B^*} + \alpha_B \gamma,$$

for all $x, \bar{x} \in D_2$.

Theorem 10. If $K > 0$ exists so that for every iteration $k > K$, we have $\|q_k\| < \frac{\eta}{(k+1)^{1+v}}\|s_k\|$, and we update B_k using the projected structured BFGS secant formula for each $k > K$, that is, $B_k = A_k + Q(x_k)$, where,

$$A_k = A_{k-1} + \Delta\left(s_{k-1}, y_{k-1}, A_{k-1}, {y^s}_{k-1} + \sqrt{\frac{{s^T}_{k-1}{y^s}_{k-1}}{{s^T}_{k-1}{B^s}_{k-1}s_{k-1}}}{B^s}_{k-1}s_{k-1}\right).$$

then

$$lim_{k\to\infty}\frac{\|(A_k-A^*)s_k\|}{\|x_{k+1}-x_k\|}=0, \tag{23}$$

We have recently proved satisfaction of (20) in Theorem 3 for the other two update formulas, and hence established their two-step superlinear convergence. First, we established the key component for the superlinear convergence, that is, the bounded deterioration inequality (18) for the projected structured PSB and DFP secant approximations. To obtain the bounded deterioration inequality for the PSB and DFP secant updates, we made use of the ideas of Dennis and Walker [6] in a different context. They discussed the convergence analysis for least-change secant methods in solving nonlinear systems of equations,

$D(x)=0,\ D: R^n\to R^n,$

approximating a part of the derivative matrix $\acute{D}(x)$. For the specific case, they considered the (unconstrained) nonlinear least squares problem,

$$\min \emptyset(x)=\frac{1}{2}F(x)^T F(x)$$
$$s.t.\, x\in R^n$$

In this case, the system of equations to be solved is $D(x)=\Delta\emptyset(x)=G(x)F(x)=0$, and the associated derivative matrix is $\acute{D}(x)=\nabla^2\emptyset(x)=G(x)G(x)^T+S(x)$, where, $G(x)$ and $S(x)$ are as given before. They assumed that the $G(x)$ are available and approximated $S(x)$. Supposing $A\approx S(x)$, they approximated $S(\bar{x})$ by $\bar{A}$ as follows.

For any $s,y\in R^n$ with $s\neq 0$, $\mathcal{L}(y,s)=\{M: Ms=y\}$ and $\mathcal{A}=\{M: M=M^T\}$, they select $\bar{A}$ to uniquely solve

$\min\|\acute{A}-A\|_w$

$s.t.\acute{A}\in\mathcal{M}(\mathcal{A},\mathcal{L})$

where, $\mathcal{L}=\mathcal{L}(y,s)$, and $\mathcal{M}(\mathcal{A},\mathcal{L})$ is the set of elements of $\mathcal{A}$ for which the distance to $\mathcal{L}$ in the norm $\|.\|_w$ is minimal (this way of picking $\bar{A}$ is called the least-changes secant criterion). Let

$$\mathcal{N}=\mathcal{N}(s)=\{M: Ms=0\},$$

and suppose $\mathcal{P}$ be a parallel subspace to $\mathcal{A}$. We can write $\mathcal{A}=\{A_N+M: M\in\mathcal{P}\}$, where, $A_N\in\mathcal{P}^{\perp}$?. Dennis andWalker [6] proved that if we take $\mathcal{P}=\mathcal{A}=\{M: M=M^T\}$ and $W=I$, then the structured PSB update of A is obtained to be

$$\bar{A}=A+\frac{(y-As)s^T+s(y-As)^T}{s^Ts}-\frac{(y-As)^Ts}{s^Ts^2}s^Ts \tag{24}$$

and if we take $\mathcal{P}=\mathcal{A}=\{M: M=M^T\}$ and $W=B^*$, then the structured DFP update of A is given by

$$\bar{A}=A+\frac{(y-As)y^{s^T}+y^s(y-As)^T}{y^{s^T}s}-\frac{(y-As)^Ts}{(y^{s^T}s)^2}y^s y^{s^T} \tag{25}$$

where, $y^s=Q(\bar{x})s+y$ is a good approximation of B^*s, if s is small enough, in norm, and $\bar{x}$ is near x^*. Here, we use the least-change secant updates for approximating the projected structured Hessians in solving the constrained nonlinear least squares problem. We will make use of the following theorem from [6].

Theorem 11. [6] *Let there be given vectors* $s, y \in R^{n-t}$ *with* $s \neq 0$, *and an innerproduct norm* $\|.\|$. *Then*

$$\mathcal{M}\,(\mathcal{A},\mathcal{L}) = \left\{(I - P_{\mathcal{P}}P_{\mathcal{N}})^{-1}A_N + (I - P_{\mathcal{P}}P_{\mathcal{N}})^{-1}P_{\mathcal{P}}{P_{\mathcal{N}}}^{\perp}\left(\frac{ys^T}{s^Ts}\right) + M : M \in \mathcal{P} \cap \mathcal{N}\right\}.$$

Furthermore, if $G \in \mathcal{M}\,(\mathcal{A},\mathcal{L})$ *and* , $M \in R^{n-t,n-t}$, *then*

$$\|\bar{A} - M\| \leq \|P_{\mathcal{P}\cap\mathcal{N}}(A - M)\| + \|{P^{\perp}}_{\mathcal{P}\cap\mathcal{N}}(G - M)\|$$

where, P_S *is the orthogonal projection onto* S, *while* ${P^{\perp}}_S = I - P_S$.

Lemma 12. *If* $\|e_k\| \leq \varepsilon_3$, $\|e_{k+1}\| \leq \varepsilon_3$, *and the inequality (8) holds, with* $s_k \neq 0$, *then we have*

$$\|{P^{\perp}}_{\mathcal{P}\cap\mathcal{N}}(G - A^*)\| \leq 3\gamma_k, \tag{26}$$

for every $G \in \mathcal{M}\,(\mathcal{A},\mathcal{L})$.

Theorem 13. *Suppose that the inequality (8) and Assumptions (A5) and (A7) hold. Let* $\bar{B}$ *be the projected structured PSB secant update, that is,*

$$\bar{B} = \bar{A} + Q(\bar{x})$$

where, $\bar{A}$ *is given by (24). Then*

$$\|\bar{A} - A^*\|_F \leq \|A - A^*\|_F + 3\gamma, \tag{27}$$

for all $x, \bar{x} \in D_2$.

The next lemma, given as Lemma 4.1 in [6], is used for obtaining the bounded deterioration inequality for the structured DFP update formula.

Lemma 14. [6] *Let* W_* *be positive definite and symmetric, and let* κ *and* ϱ *be positive constants with* $\varrho \leq 1$. *Suppose that positive parameter* γ *and vectors* $s, v \in R^{n-t}$ *with* $s \neq 0$ *satisfy*

$$\|v - w_*s\|_2 \leq \kappa\gamma\|s\|_2,$$

$$\gamma \leq \frac{(1-\varrho)}{\kappa\|{w_*}^{-1}\|_2},\ \gamma \leq \frac{(1-\varrho)\varrho}{(1+\sqrt{\varrho})\kappa\|{w_*}^{-1}\|_2},$$

Then, W, *the BFGS update of* w_* *sending* s *to* v *(that is,* $Ws = v$*), is well defined by*

$$w_* = J_*{J_*}^T, w = \sqrt{\frac{v^Ts}{s^Tw_*s}}{J_*}^Ts, J = J_* + \frac{(v-J_*w)w^T}{w^Tw}, W = JJ^T,$$

and there exist positive constants β_1 *and* β_2, *independent of* γ, s *and* v, *such that*
$\|M\|_{w_*} \leq (1 + \beta_1\gamma)\|M\|_w$, $\|M\|_w \leq (1 + \beta_2\gamma)\|M\|_{w_*}$ *for every matrix* $M \in R^{n-t,n-t}$.

Now, using lemmas 14 and 6, we obtain the following lemma for use in our constrained projected structured Hessian.

Lemma 15. *If the inequality (8) holds, then there exist a neighborhood* $D_3 \subseteq D_2$ *of* x^* *and positive constants* β_1 *and* β_2 *such that*
$\|M\|_{B^*} \leq (1 + \beta_1\gamma)\|M\|_w$, $\|M\|_w \leq (1 + \beta_2\gamma)\|M\|_{B^*}$

for every matrix $M \in R^{n-t,n-t}$, where,W is the BFGS update of B^* sending s to y^s.

Theorem 16. *Suppose that the inequality (8) and Assumptions (A5) and (A7) hold. Let $\bar{B}$ be the projected structured DFP secant update, that is,*

$$\bar{B} = \bar{A} + Q(\bar{x})$$

where, $\bar{A}$ is given by (25). Then, there exist positive constants α_2 and α_3 such that

$$\|\bar{A} - A^*\|_{B^*} \leq (1 + \alpha_2\gamma)\|A - A^*\|_{B^*} + \alpha_3\gamma, \quad (28)$$

for all $x, \bar{x} \in D_3$.

Now, we point out the satisfaction of (20) in Theorem 3 for the two update formulas, and thus assure their two-step superlinear convergence.

Theorem 17. *If $K > 0$ exists so that for every iteration $k > K$, we have*

$$\|q_k\| < \frac{\eta}{(k+1)^{1+v}}\|s_k\|,$$

and we update the B_k using the structured PSB or DFP secant formulas for each $k > K$, that is,

$$B_k = A_k + Q(x_k),$$

where, A_k is the PSB or DFP update formula of A_{k-1} obtained by using (24) and (25), respectively. Then

$$\lim_{k\to\infty} \frac{\|(A_k - A^*)s_k\|}{\|x_{k+1} - x_k\|} = 0$$

Corollary 18. *The projected structured PSB and DFP update formulas satisfy the assumptions of Theorem 5 and thus Algorithm 1 with these updating schemes for approximating the projected structured Hessians possess at least a two-step superlinear convergence rate.*

5 Numerical Experiments

We coded our algorithm in MATLAB 7.6.0. In the global steps, a line search strategy is necessary.

In our implementation, for the line search strategy, we used the approach specially designed for nonlinear least squares by Mahdavi-Amiri and Ansari [9] and [7]. We put $\eta = 1$ in (8), as suggested in [11] and we set $\varepsilon = 10^{-2}$, $\tau = 10^{-1}$ and the initial matrix $A_{z,0}$, is set to be the identity matrix. For robustness, we followed the computational considerations provided by Mahdavi-Amiri and Bartels [10]. We test our algorithm on 7 randomly generated test problems using the test problem generation scheme given by Bartels and Mahdavi-Amiri [1]. The parameters of theses random problems are reported in Table 1. The column headed by "PN", "n", "p", "k", "m" and "v" give the problem number, the number of variables, the number of components in $F(x)$, the number of equality constraints, the number of inequality constraints and the number of active inequality constraints, respectively. The scalar "σ" dictates whether the Hessian of the Lagrangian of the generated problem is positive definite ($\sigma > 0$) or indefinite ($\sigma \leq 0$). All random numbers needed for the random problems were generated by the function

$rand$ in MATLAB. For the generated problem sets 1-3, 4-5 and 6-7, all quantities are exactly the same and only $\nabla^2 \in L(x^*, \lambda^*)$ differs in each set by having a different value of σ. In tables 2, 3 and 4, we report the results obtained by our program on these random problems. For comparison, the results obtained by $f\ mincon$ in MATLAB and the ones obtained by the three algorithms in KNITRO 6.0 (Interior-point/Direct, Interior-point/CG and Active set algorithms) on these random problems are separately reported in tables 5, 6, 7 and 8, respectively. In keeping the three algorithms of KNITRO in line with our computing features, we set the parameters 'GradObj' and 'GradConstr' to 'on', so that exact gradients are used, and the other parameters were set to the default parameter values (this way, the BFGS updating rule is used for Hessian approximations). In tables 2–8, the "SP" column shows the value to which all component of the starting point are set. The "OV" and "FV" columns give the optimal value and objective function value, respectively. The "NFE" column gives the number of times the objective function was computed. We computed the composite absolute relative feasibility error at the final point x as:

$$ref(x) = \frac{\sum_{i=1}^{k} |c_i(x)| - \sum_{j=k+1}^{k+m} \min(0, c_j(x))}{1 + \|C(x)\|_1}$$

where, $C(x)$ denotes the constraint vector evaluated at x. Rounded value of $rfe(x)$ is shown in the column headed by "RFE". According to the results in tables 2–8, we observe that our algorithm has shown to be substantially more efficient more often than f mincon and the three programs in KNITRO. The numerical results show that our algorithm terminated successfully at the specified local minima, for all the random test problems, while on a few test problems both KNITRO and $f\ mincon$ found other local minima having larger optimal values with excessive computing work. This may be due to the extra care the latter two programs take to ensure feasibility possibly at the expense of optimality. We also observe that KNITRO and $f\ mincon$ are more sensitive to the starting point and perform significantly worse than our algorithm as the dimension of the problem increase.

Table 1: The parameters of random problems

PN	**n**	**p**	**k**	**m**	**v**	**σ**
1	5	5	2	3	2	1
2	5	5	2	3	2	-1
3	5	5	2	3	2	-10
4	10	10	5	5	2	1
5	10	10	5	5	2	-1
6	20	20	8	12	2	1
7	20	20	8	12	2	-1

Table 2: Results on randomly generated test problems using the BFGS scheme

PN	SP	OV	FV	NFE	RFE
1	1	3.288834E+01	3.288834E+01	68	E-10
1	10	3.288834E+01	3.288834E+01	61	E-11
1	100	3.288834E+01	3.288834E+01	67	E-10
2	1	3.288834E+01	3.288834E+01	39	E-09
2	10	3.288834E+01	3.288834E+01	47	E-09
2	100	3.288834E+01	3.288834E+01	65	E-09
3	1	3.288834E+01	3.288834E+01	37	E-09
3	10	3.288834E+01	3.288834E+01	78	E-09
3	100	3.288834E+01	3.289961E+01	75	E-12
4	1	1.985654E+02	1.985577E+02	369	E-09
4	10	1.985654E+02	1.985577E+02	270	E-09
5	1	1.985654E+02	1.985654E+02	449	E-09
5	10	1.985654E+02	1.985654E+02	250	E-08
6	1	1.482277E+03	1.482278E+03	168	E-11
6	10	1.482277E+03	1.482277E+03	63	E-12
7	1	1.482277E+03	1.482283E+03	54	E-10
7	10	1.482277E+03	1.482277E+03	63	E-12

Table 3: Results on randomly generated test problems using the PSB scheme

PN	SP	FV	NFE	RFE
1	1	3.288834E+01	65	E-10
1	10	3.288834E+01	42	E-09
1	100	3.288834E+01	94	E-07
2	1	3.288834E+01	45	E-09
2	10	3.288834E+01	88	E-09
2	100	3.288834E+01	75	E-09
3	1	3.288834E+01	65	E-09
3	10	3.288834E+01	41	E-09
3	100	3.289961E+01	91	E-13
4	1	1.985654E+02	70	E-11
4	10	1.985654E+02	65	E-09
5	1	1.985654E+02	86	E-12
5	10	1.985652E+02	84	E-10
6	1	1.482277E+03	106	E-12
6	10	1.482277E+03	63	E-13
7	1	1.482283E+03	54	E-10
7	10	1.482277E+03	64	E-12

Table 4: Results on randomly generated test problems using the DFP scheme

PN	SP	FV	NFE	RFE
1	1	3.288834E+01	42	E-09
1	10	3.288834E+01	112	E-09
1	100	3.288834E+01	75	E-12
2	1	3.288834E+01	50	E-10
2	10	3.288834E+01	46	E-09
2	100	3.288834E+01	90	E-09
3	1	3.288834E+01	54	E-09
3	10	3.288834E+01	42	E-09
3	100	3.289961E+01	61	E-14
4	1	1.985654E+02	89	E-07
4	10	1.985654E+02	65	E-09
5	1	1.985654E+02	59	E-11
5	10	1.985655E+02	74	E-08
6	1	1.482281E+03	103	E-08
6	10	1.482277E+03	63	E-12
7	1	1.482283E+03	54	E-10
7	10	1.482277E+03	63	E-12

Table 5: Results on randomly generated test problems using the function f mincon in MATLAB

PN	SP	FV	NFE	RFE
1	1	8.340232E+01	505	E-03
1	10	3.288834E+01	283	E-07
1	100	3.288836E+01	362	E-08
2	1	3.293219E+01	502	E-05
2	10	3.288834E+01	402	E-07
2	100	3.288835E+01	389	E-12
3	1	3.289961E+01	349	E-07
3	10	3.289961E+01	392	E-10
3	100	1.252427E+02	505	E+01
4	1	1.985577E+02	369	E-09
4	10	1.985577E+02	270	E-09
5	1	1.985654E+02	449	E-09
5	10	1.985654E+02	250	E-08
6	1	2.805303E+10	680	E-13
6	10	2.805303E+10	369	E-13
7	1	2.805368E+10	424	E-12
7	10	2.805368E+10	479	E-14

Table 6: Results on randomly generated test problems obtained by three algorithms of KNITRO (Interiorpoint/ Direct)

PN	SP	FV	NFE	RFE
1	1	3.289E+01	35	E-07
1	10	3.299E+01	61	E-05
1	100	3.700E+01	62	E-07
2	1	3.289E+01	32	E-09
2	10	3.299E+01	45	E-07
2	100	3.565E+01	58	E-06
3	1	3.289E+01	31	E-07
3	10	3.298E+01	48	E-06
3	100	3.292E+01	49	E-05
4	1	1.986E+02	257	E-06
4	10	2.001E+02	50	E-05
5	1	1.987E+02	4092	E-06
5	10	2.023E+02	66	E-05
6	1	1.488E+03	201	E-06
6	10	1.490E+03	151	E-06
7	1	1.486E+03	279	E-06
7	10	1.485E+03	177	E-05

Table 7: Results on randomly generated test problems obtained by three algorithms of KNITRO (Interior-point/CG)

PN	SP	FV	NFE	RFE
1	1	3.289E+01	69	E-07
1	10	3.398E+01	53	E-06
1	100	4.534E+01	69	E-06
2	1	3.289E+01	65	E-07
2	10	3.427E+01	39	E-05
2	100	4.990E+01	67	E-06
3	1	3.290E+01	57	E-07
3	10	3.334E+01	84	E-07
3	100	4.583E+01	69	E-06
4	1	1.991E+02	63	E-06
4	10	2.110E+02	112	E-05
5	1	1.991E+02	63	E-06
5	10	2.140E+02	125	E-07
6	1	1.505E+03	1545	E-06
6	10	2.805E+10	80	E-10
7	1	1.526E+03	2388	E-10
7	10	2.805E+10	55	E-08

Table 8: Results on randomly generated test problems obtained by three algorithms of KNITRO (Active-Set)

PN	SP	FV	NFE	RFE
1	1	3.289E+01	191	E-11
1	10	3.321E+01	197	E-09
1	100	4.276E+01	91	E-09
2	1	3.289E+01	135	E-12
2	10	3.289E+01	15	E-04
2	100	4.292E+01	102	E-09
3	1	3.289E+01	98	E-09
3	10	3.291E+01	161	E-10
3	100	4.334E+01	89	E-09
4	1	1.987E+02	77	E-10
4	10	2.339E+02	455	E-10
5	1	1.987E+02	77	E-10
5	10	2.333E+02	490	E-10
6	1	1.511E+03	3934	E-12
6	10	2.805E+10	193	E-16
7	1	1.511E+03	3394	E-12
7	10	2.805E+10	205	E-13

6 Conclusions

We proposed a new general projected structured secant scheme for approximating the exact penalty projected structured Hessian matrix in an exact penalty method for solving constrained nonlinear least squares problems. We gave sufficient conditions for the local two-step superlinear convergence of the proposed algorithm and affirmed that the given special projected structured BFGS, PSB and DFP update formulas satisfy these conditions. Comparative numerical results showed the efficiency and robustness of the resulting algorithms.

Acknowledgements

The authors thank the Research Council of Sharif University of Technology for supporting this work.

References

[1] Bartels, R. H. and Mahdavi-Amiri, N., 1986. On generating test problems for nonlinear programming algorithms. *SIAM J. Sci. Stat. Comput.*, **7, 3** 769–798.

[2] Bidabadi, N. and Mahdavi-Amiri, N., 2011. A two-step superlinearly convergent projected structured bfgs method for constrained nonlinear least squares. Submitted.

[3] Coleman, T. F. and Conn, A. R., 1982a. Nonlinear programming via an exact penalty function: Asymptotic analysis. *Math. Program.*, **24** 123–136.

[4] Coleman, T. F. and Conn, A. R., 1982b. Nonlinear programming via an exact penalty function: Global analysis. *Math. Program.*, **24** 137–161.

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

[6] Dennis, J. E. and Walker, H. F., 1981. Convergence theorems for least-changes secant update methods. SIAM J. *Numer. Anal.*, **18, 6** 949–987.

[7] Mahdavi-Amiri, N. and Ansari, M. R., 2011a. A nonsmooth line search strategy and numerical results for a superlinearly convergent projected structured exact penalty method for constrained nonlinear least squares. Submitted.

[8] Mahdavi-Amiri, N. and Ansari, M. R., 2011b. Superlinearly convergent exact penalty projected structured Hessian updating schemes for constrained nonlinear least squares: asymptotic analysis. *Bulletin of the Iranian Mathematical Society*, Accepted.

[9] Mahdavi-Amiri, N. and Ansari, M. R., 2011c. Superlinearly convergent exact penalty projected structured Hessian updating schemes for constrained nonlinear least squares: global analysis. Submitted.

[10] Mahdavi-Amiri, N. and Bartels, R. H., 1989. Constrained nonlinear least squares: An exact penalty approach with projected structured quasi-Newton updates. *ACM Transactions on Mathematical Software*, **15, 3** 220– 242.

[11] Nocedal, J. and Overton, M. L., 1985. Projected Hessian updating algorithms for nonlinearly constrained optimization. *SIAM J. Numer. Anal.*, **22, 5** 821–850.