Journal of Computational and Applied Mathematics 273 (2014) 25–28

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Letter to the editor

On the similarities between the quasi-Newton least squares method and GMRes

R. Haelterman [a,*], B. Lauwens [a], F. Van Utterbeeck [a], H. Bruyninckx [a], J. Vierendeels [b]

[a] Royal Military Academy, Renaissancelaan 30, B-1000 Brussels, Belgium

[b] Ghent University, Department of Flow, Heat and Combustion Mechanics, St.-Pietersnieuwstraat 41, B-9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 23 January 2014

MSC:
65F10
90C53
90C56

Keywords:
Quasi-Newton method
Least squares
GMRes

ABSTRACT

We show how the quasi-Newton least squares method (QN-LS) relates to Krylov subspace methods in general and to GMRes in particular.

1. Introduction

After having showed that the quasi-Newton inverse least squares method (QN-ILS) shows similarities with Krylov methods [1], we complete our study with a similar result for the quasi-Newton least squares method (QN-LS). Logically, the approach is analogous to that in [1], but the details differ. Surprisingly, we find that the Krylov space of the iterates is the same for QN-LS, QN-ILS and GMRes, but that for the residuals differs.

2. The quasi-Newton least squares method

The quasi-Newton least squares (QN-LS) method [2,3] is an iterative method that has been developed to solve a non-linear problem of the form $K(p) = 0$, where $K : \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$. In this study we will limit ourselves to the case where K is affine, i.e. $K(p) = A_K p - b$, with $A_K \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^{n\times 1}$. We assume A_K is non-singular.

We only give the most important characteristics of the method here; more details can be found in [3]. The iterations in the QN-LS method start from p_o and generate the sequence $p_{s+1} = p_s - (\hat{K}'_s)^{-1} K(p_s)$ $(s = 1, 2, \ldots)$, where

$$\hat{K}'_o = -I \quad \text{and} \quad \hat{K}'_s = W_s (V_s^T V_s)^{-1} (V_s)^T - I \quad \text{for } s > 0, \tag{1}$$

and

- $W_s = [\delta H_o \; \delta H_1 \ldots \delta H_{s-1}]$;
- $V_s = [\delta p_o \; \delta p_1 \ldots \delta p_{s-1}]$;

* Corresponding author. Tel.: +32 2 742 6527; fax: +32 2 742 6512.
E-mail address: robby.haelterman@rma.ac.be (R. Haelterman).

http://dx.doi.org/10.1016/j.cam.2014.05.020
0377-0427/

- $\delta H_k = H(p_{k+1}) - H(p_k)$ $(k = 0, 1, \ldots, s-1)$;
- $\delta p_k = p_{k+1} - p_k$ $(k = 0, 1, \ldots, s-1)$;
- $H : \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ is defined such that $\forall x \in \mathbb{R}^{n\times 1} : H(x) = K(x) + x$.

In the affine case (1) is equivalent to $\hat{K}'_s = A_H V_s(V_s^T V_s)^{-1}(V_s)^T - I$, with $A_H = A_K + I$. We note that $V_s(V_s^T V_s)^{-1}(V_s)^T$ is an orthogonal projection matrix on the range of V_s and that it has already been proven that $\hat{K}'_s$ cannot become singular before the solution is reached [4].

If we write p^* for the solution of $K(p) = 0$, $e_s = p_s - p^*$ and $r_s = K(p_s)$, then we obtain

$$e_{s+1} = e_s - (\hat{K}'_s)^{-1}A_K e_s \quad \text{and} \quad r_{s+1} = r_s - A_K(\hat{K}'_s)^{-1} r_s. \tag{2}$$

3. Comparing QN-LS with Krylov subspace methods

Definition 3.1. A Krylov (subspace) method to solve the linear system $A_K p = b$ is a method where for the sth iterate p_s we have $p_s \in p_o + \mathcal{K}_s\{A_K; r_o\}$ and $r_s \perp \mathcal{Z}$, where $\mathcal{Z} \vec{\subset} \mathbb{R}^{n\times 1}$ has dimension s. The choice of $\mathcal{Z}$ defines the particular Krylov subspace method; e.g. for GMRes this is $A_K \mathcal{K}_s\{A_K; r_o\}$ [5].

We will now show that the QN-LS method shows similarities with Krylov subspace methods when applied to the affine problem. More specifically $p_s \in p_o + \mathcal{K}_s\{A_K; r_o\}$, but $r_s \perp (A_H^T)^{-1}\mathcal{K}_{s-1}\{A_K; r_o\}$. As r_s is orthogonal only to an $s-1$-dimensional subspace, it is not a Krylov method in the strict sense.

Theorem 3.1. *For QN-LS applied to the affine problem we have that* $\forall j \in \{0, 1, \ldots, s\} : (I - (\hat{K}'_s)^{-1}A_K)(e_s - e_j) = 0$.

For the proof of this theorem we refer to [3].

Corollary 3.1. *For QN-LS applied to the affine problem we have that* $e_1 = A_H e_o$, $r_1 = A_H r_o$ *and for* $s \geq 1$ $\exists\{\gamma_{i,s+1}\}_{i=1}^{i=s}$, *such that*

$$e_{s+1} = A_H e_o + A_H \sum_{i=1}^{s} \gamma_{i,s+1}(e_i - e_o) \tag{3}$$

$$r_{s+1} = A_H r_o + A_H \sum_{i=1}^{s} \gamma_{i,s+1}(r_i - r_o). \tag{4}$$

Proof. From (2) and Theorem 3.1 it follows that

$$\begin{aligned} e_{s+1} &= e_s - (\hat{K}'_s)^{-1}A_K e_s = (I - (\hat{K}'_s)^{-1}A_K)e_s \\ &= (I - (\hat{K}'_s)^{-1}A_K)e_j \quad (j = 0, 1, \ldots, s) \\ &= e_o - (\hat{K}'_s)^{-1}(A_H - I)e_o \end{aligned}$$

$$(A_H V_s(V_s^T V_s)^{-1}(V_s)^T - I)(e_{s+1} - e_o) = -(A_H - I)e_o$$

$$e_{s+1} = A_H V_s(V_s^T V_s)^{-1}(V_s)^T(e_{s+1} - e_o) + A_H e_o.$$

As $V_s(V_s^T V_s)^{-1}(V_s)^T$ is a projection operator on $\text{span}\{p_s - p_{s-1}, p_{s-1} - p_{s-2}, \ldots, p_1 - p_o\} = \text{span}\{p_1 - p_o, p_2 - p_o, \ldots, p_s - p_o\} = \text{span}\{e_1 - e_o, e_2 - e_o, \ldots, e_s - e_o\}$, the latter expression can be written as $e_{s+1} = A_H e_o + A_H \sum_{i=1}^{s} \gamma_{i,s+1}(e_i - e_o)$, which proves (3). To prove (4) we only need to multiply (3) by A_K and note that A_K and A_H commute.

Theorem 3.2. *For QN-LS applied to the affine problem we have that*

$$e_s \in e_o + \mathcal{K}_s\{A_K; r_o\} \tag{5a}$$

$$p_s \in p_o + \mathcal{K}_s\{A_K; r_o\} \tag{5b}$$

$$r_s \in r_o + A_K\mathcal{K}_s\{A_K; r_o\}. \tag{5c}$$

Proof. Let $\mathbb{R}_o^k[x] = \{q(x) \in \mathbb{R}[x] : q(x) = \sum_{i=1}^{k} \kappa_i x^i\}$, i.e. the space of real polynomials of degree k, or lower, with zero constant term. We first note that $\mathbb{R}_o^k[x]$ over $\mathbb{R}$ is a vector-space of dimension k and that $\forall l \leq k = \mathbb{R}_o^l[x] \vec{\subset} \mathbb{R}_o^k[x]$.

- We will first prove, by induction, that $e_s = e_o + q_s(A_K)e_o$ $(s = 1, 2, \ldots)$, where $q_s(x) \in \mathbb{R}_o^k[x]$.

  We know that $e_1 = A_H e_o = e_o + (A_H - I)e_o = e_o + q_1(A_K)e_o$, where $q_1 \in \mathbb{R}_o^1[x]$. We also know (from Corollary 3.1) that $\exists \gamma_{1,2} \in \mathbb{R}$ such that

$$e_2 = A_H \gamma_{1,2}(e_1 - e_o) + A_H e_o \tag{6}$$

$$= \gamma_{1,2}(A_H - I)^2 e_o + (1 + \gamma_{1,2})(A_H - I)e_o + e_o \tag{7}$$

$$= e_o + q_2(A_K)e_o, \tag{8}$$

  where $q_2 \in \mathbb{R}_o^2[x]$.

We now prove that, if we have $e_k = e_o + q_k(A_K)e_o$ for $k = 1, 2, \ldots s-1$, where $q_k \in \mathbb{R}_o^k[x]$, it follows that $e_s = e_o + q_s(A_K)e_o$,where $q_s \in \mathbb{R}_o^s[x]$.

We have (from Corollary 3.1) that $\exists \gamma_{k,s} \in \mathbb{R}$, $(k = 1, 2, \ldots, s-1)$; such that

$$e_s = A_H e_o + A_H \sum_{k=1}^{s-1} \gamma_{k,s}(e_k - e_o) = A_H \sum_{k=1}^{s-1} \gamma_{k,s}(q_k(A_K)e_o) + A_H e_o. \tag{9}$$

Knowing that $\forall k \le s-1 : q_k \in \mathbb{R}_o^k[x] \overset{\rightarrow}{\subset} \mathbb{R}_o^{s-1}[x]$, and posing $\tilde{q}_{s-1} = \sum_{k=1}^{s-1} \gamma_{k,s}(q_k(x)) \in \mathbb{R}_o^{s-1}[x]$, we can write

$$e_s = A_H \tilde{q}_{s-1}(A_K)e_o + A_H e_o = \underbrace{(A_H - I)}_{=A_K} \tilde{q}_{s-1}(A_K)e_o + \tilde{q}_{s-1}(A_K)e_o + (A_K)e_o + e_o. \tag{10}$$

As $\forall q(x) \in \mathbb{R}_o^{s-1}[x] : xq(x) \in \mathbb{R}_o^s[x]$ and as $x \in \mathbb{R}_o^s[x]$ we can finally write

$$e_s = q_s(A_K)e_o + e_o, \tag{11}$$

where $q_s(x) = x\tilde{q}_{s-1}(x) + \tilde{q}_{s-1}(x) + x \in \mathbb{R}_o^s[x]$.

- From (11) we see that $e_{s+1} \in e_o + \mathrm{span}\{A_K e_o, A_K^2 e_o, A_K^3 e_o, \ldots, A_K^s e_o\}$.

  Noting that $\mathrm{span}\{A_K e_o, A_K^2 e_o, A_K^3 e_o, \ldots, A_K^s e_o\} = \mathrm{span}\{r_o, A_K r_o, A_K^2 r_o, \ldots, A_K^{s-1} r_o\}$ we have proven (5a)–(5c) follow immediately.

Lemma 3.1. *For QN-LS applied to the affine problem we have that* $r_{s+1} = A_H \bar{L}_{s+1} \bar{L}_{s+1}^T \delta p_s$ $(s = 0, 1, 2, \ldots)$, *where* $\mathcal{L}_{s+1} = [\bar{L}_1 | \bar{L}_2 | \ldots | \bar{L}_{s+1}]$ *is an orthonormal matrix with the same range as* V_{s+1}, *constructed such that* $\mathcal{L}_{s+1} = [\mathcal{L}_s | \bar{L}_{s+1}]$.

For a proof of this corollary we refer to [3].

Theorem 3.3. *For QN-LS applied to the affine problem we have that* $r_s \perp (A_H^T)^{-1} \mathcal{K}_{s-1}\{A_K; r_o\}$.

Proof. From Lemma 3.1 we know that $r_s = A_H \bar{L}_s \bar{L}_s^T \delta p_{s-1}$.

Setting $\bar{L}_s^T \delta p_{s-1} = \kappa \in \mathbb{R}$, we have $r_s = \kappa A_H \bar{L}_s$ and as $\forall y \in \mathcal{R}(V_{s-1}) : \langle \bar{L}_s, y \rangle = 0$, it follows that $\forall y \in \mathcal{R}(V_{s-1}) : \langle r_s, (A_H^T)^{-1} y \rangle = 0$.

From the definition of V_s and Eq. (5b), we see that $\mathcal{R}(V_{s-1}) = \mathcal{K}_{s-1}\{A_K; r_o\}$. r_s is thus orthogonal to $(A_H^T)^{-1} \mathcal{K}_{s-1}\{A_K; r_o\}$.

We note that

- for QN-LS p_s lies in the same Krylov subspace $\mathcal{K}_s\{A_K; r_o\}$ as for GMRes and QN-ILS. r_s also lies in the same subspace for the QN-LS, QN-ILS and GMRes methods [1,5].
- for QN-LS r_s is orthogonal to $(A_H^T)^{-1} \mathcal{K}_{s-1}\{A_K; r_o\}$, whereas for QN-ILS this is $(A_H^T)^{-1} A_K \mathcal{K}_{s-1}\{A_K; r_o\}$ [1]. Both are only subspaces of dimension $s-1$; hence these methods do not comply with the definition of a Krylov method. For GMRes r_s is orthogonal to $A_K \mathcal{K}_s\{A_K; r_o\}$ [5].

As the residual and the iterate already share the subspaces of their equivalents in the GMRes method it is fairly easy to adapt the QN-LS method in order to make it algebraically identical to GMRes, based on the ideas in [1]. For this we adapt the iterations as follows: $p_{s+1} = p_s - \sum_{i=0}^s \theta_{s,i} \Delta_i$, and thus $r_{s+1} = r_s - \sum_{i=0}^s \theta_{s,i} q_i$, where $\Delta_i = (\hat{K}_i')^{-1} r_i$ and $q_i = A_K \Delta_i$ $(i = 0, 1, \ldots, s)$.

To find the optimal parameters $\theta_{s,i}$ $(i = 0, 2, \ldots s)$ we define $\Theta_s = [\theta_{s,o} \ \theta_{s,1} \ \ldots \ \theta_{s,s}]^T$ and impose $r_{s+1} \in Q^\perp$ where $Q = [q_0 | q_1 | \ldots | q_s]$. This leads to $\Theta_s = (Q^T Q)^{-1} Q^T r_s$.

As $\hat{K}_i$ $(i = 0, 1, \ldots, s)$ is non-singular [4] we have that $\{\Delta_i\}_{i=0}^s$ forms a basis for the Krylov subspace $\mathcal{K}_s\{A_K; r_o\}$. It follows that $\{q_i\}_{i=0}^s$ forms a basis for the Krylov subspace $A_K \mathcal{K}_s\{A_K; r_o\}$ to which r_{s+1} is now orthogonal. Hence this modification makes the QN-LS method algebraically identical to GMRes.

4. Conclusions

We have shown that for an affine problem the iterates for QN-LS share the same Krylov search subspace as those of GMRes. It is also shown that adding suitable step-length parameters can make QN-LS equivalent to GMRes.

References

[1] R. Haelterman, J. Degroote, D. Van Heule, J. Vierendeels, On the similarities between the quasi-Newton inverse least squares method and GMRes, SIAM J. Numer. Anal. 47 (6) (2010) 4660–4679.

[2] J. Degroote, R. Haelterman, S. Annerel, A. Swillens, P. Segers, J. Vierendeels, An interface quasi-Newton algorithm for partitioned simulation of fluid–structure interaction, in: Proc. of the International Workshop on Fluid-Structure Interaction: Theory, Numerics and Applications, Kassel University Press, ISBN: 978-3-89958-666-4, 2009, pp. 55–64. Herrsching am Ammersee, Germany. September 29-October 1, 2008.

[3] R. Haelterman, J. Degroote, D. Van Heule, J. Vierendeels, The Quasi-Newton Least Squares method: a new and fast secant method analyzed for linear systems, SIAM J. Numer. Anal. 47 (3) (2009) 2347–2368.
[4] R. Haelterman, J. Petit, B. Lauwens, H. Bruyninckx, J. Vierendeels, On the non-singularity of the QN-LS method, J. Comput. Appl. Math. 257 (2014) 129–131. http://dx.doi.org/10.1016/j.cam.2013.08.020.
[5] Y. Saad, Iterative Methods for Sparse Linear Systems, second ed., Society for Industrial and Applied Mathematics (SIAM), Philadelphia, ISBN: 0-89871-534-2, 2003.