

Un Método Secante Estructurado con Estrategia Globalizadora no Monótona para Resolver el Problema de Cuadrados Mínimos

Graciela Croceri, Gonzalo Pizarro y Graciela Sottosanto

Departamento de Matemática, Universidad Nacional del Comahue, Santa Fe 1400, Neuquén, Argentina

Resumen. En este trabajo se propone resolver el problema de cuadrados mínimos, mediante la aplicación de un algoritmo que combina un método secante estructurado con una estrategia globalizadora no monótona, de región de confianza. La matriz Hessiana para conformar el modelo cuadrático, se elige usando un método secante que aprovecha la estructura del problema, y el radio de la región de confianza se actualiza siguiendo una técnica adaptativa.

La experimentación numérica preliminar, en la que se comparan diferentes formas de elegir la matriz Hessiana, pone de manifiesto la eficiencia del método.

1 Introducción

En este trabajo proponemos un algoritmo para resolver el problema de cuadrados mínimos no lineales sin restricciones. El problema de cuadrados mínimos aparece frecuentemente en diversos campos de la ciencia, como la ingeniería, física, etc., en aplicaciones prácticas como estimación de parámetros, ajuste de datos, aproximación de funciones y análisis estadísticos.

El problema a considerar es el siguiente:

$$\min f(x) = \frac{1}{2} \|F(x)\|_2^2 = \frac{1}{2} \sum_{i=1}^m F_i(x)^2, \quad (1)$$

donde la función residuo es $F(x) = (F_1(x), \dots, F_m(x))^T$, las componentes $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$ para $i = 1, \dots, m$, son funciones dos veces continuamente diferenciables y, en general, no lineales y $\|\cdot\|$ denota la norma Euclídeana.

El gradiente de la función f es:

$$\nabla f(x) = J(x)^T F(x), \quad (2)$$

donde $J(x)$ es la matriz Jacobiana de F evaluada en x y la matriz Hessiana de f está dada por

$$\nabla^2 f(x) = J(x)^T J(x) + \sum_{i=1}^m F_i(x) \nabla^2 F_i(x). \quad (3)$$

En lo sucesivo vamos a denotar $\nabla f_k = \nabla f(x_k)$, $J_k = J(x_k)$, $F_k = F(x_k)$, $s_k = x_{k+1} - x_k$.

Si bien para resolver el problema (1) se pueden utilizar los métodos clásicos de los problemas de minimización sin restricciones, se han desarrollado diversos algoritmos específicos. Los más eficientes hacen uso de su estructura especial y utilizan algoritmos cuasi-Newton estructurados. En particular, vamos a presentar un algoritmo de región de confianza no monótono adaptativo utilizando una aproximación secante estructurada de una porción de la matriz Hessiana.

Este trabajo está organizado como sigue, en la sección siguiente presentamos la estrategia de región de confianza no monótona adaptativa, en la sección 3 se definen diferentes aproximaciones secantes estructuradas, el algoritmo propuesto se presenta en la sección 4, en la sección 5 describimos las experiencias numéricas y finalmente, en la sección 6, se plantean el futuro trabajo y conclusiones.

2 Región de Confianza no Monótona Adaptativa

La estrategia clásica de región de confianza es un proceso de globalización iterativo [1] muy popular que requiere, en cada iteración, obtener la solución del siguiente subproblema

$$\begin{aligned} m_k(x+s) &= f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s \\ \text{s.a.} \quad & \|s\|_2 \leq \delta_k \end{aligned} \quad (4)$$

donde $B_k \in \mathbb{R}^{n \times n}$ es una matriz simétrica.

Como se observa, m_k es un modelo cuadrático de f alrededor de x_k y δ_k es el radio de la región de confianza. Si se consigue suficiente reducción de la función en el minimizador del modelo cuadrático, se acepta el paso de prueba y se obtiene el nuevo iterado $x_{k+1} = x_k + s_k$, de lo contrario el radio de la región se reduce y se actualiza siguiendo un esquema estandar [2,3].

Debido a la robustez y a las importantes propiedades de convergencia global, el método de región de confianza ha sido ampliamente estudiado y aplicado.

Como esta estrategia construye una sucesión de iterados $\{x_k\}$ tales que $\{f(x_k)\}$ es monótonamente decreciente, algunas investigaciones, indican que la tasa de convergencia puede ser muy lenta ya que se puede generar pérdida de eficiencia cuando la función objetivo desciende en valles muy pronunciados, pues pueden realizarse pasos muy cortos y trayectorias zigzagueantes.

En los últimos años, se han propuesto variantes del método de modo que al relajar el requerimiento de monotonicidad se obtiene un algoritmo más eficiente. La idea de construir métodos no monótonos se remonta a 1986 donde Grippo et al. [4] presentan una técnica de búsqueda lineal no monótona combinada con un método de Newton para optimización sin restricciones. En 1993 Deng et al. [5] extienden las técnicas no monótonas de búsqueda lineal a algoritmos de región de confianza. Posteriormente, numerosos autores han propuesto variantes del método de región de confianza no monótono [6,7,8,9] y han podido mostrar que posee propiedades de convergencia similares al algoritmo de región

de confianza usual. El objetivo básico de estas modificaciones es disminuir el número de iteraciones ineficientes, reduciendo la cantidad de subproblemas que se resuelven.

Una vez que resolvemos el subproblema (4) con una estrategia *double-dogleg* y encontramos un punto x_{k+1} , necesitamos decidir si el punto hallado es o no aceptado. Relajando la condición clásica de aceptación del paso, vamos a comparar la reducción real de la función objetivo

$$A_{redk} = f_{k+1} - f_{l_k} \quad (5)$$

con la reducción predicha

$$P_{predk} = m_{k+1} - f_{l_k}, \quad (6)$$

donde

$$f_{l_k} = f(x_{l(k)}) = \max_{0 \leq j \leq p(k)} f(x_{k-j}), \quad (7)$$

$p(0) = 0$ y $0 \leq p(k) \leq \min\{p(k-1)+1, N\}$, $k \geq 1$ y N es un entero no negativo.

De modo que si

$$r_k = \frac{A_{redk}}{P_{predk}} \geq \eta \quad (8)$$

aceptamos el paso, de otra manera lo rechazamos. Observemos que este criterio no genera una sucesión $\{f(x_k)\}$ monótonamente decreciente. Sin embargo, cuando $N \geq 1$, el criterio de aceptación del paso debería garantizar cierto decrecimiento de $f(x_k + s_k)$ en comparación con $f(x_{l(k)})$. Si $N = 0$ se reduce a la región de confianza tradicional.

2.1 Determinación del Radio de la Región de Confianza

Discutimos a continuación una versión no monótona del método de región de confianza, en la cual, el radio de la región se ajusta de manera adaptativa. En los últimos años, varios autores, entre ellos [10,11,12,13], han propuesto estrategias en las cuales, el ajuste del radio en cada iteración interna, se realiza usando algún procedimiento que involucra valores de la función, del gradiente o de la matriz Hessiana en el punto actual.

En este caso proponemos un método que lo ajusta automáticamente de acuerdo a la información que se tiene del punto actual, y el cálculo de una fórmula simple que involucra valores de la función objetivo. En nuestro caso, para δ_k proponemos

$$\delta_k = \begin{cases} e^{q_k} R_k & \text{si } k = 0 \\ e^{q_k} \max\{R_k, \delta_{k-1}\} & \text{si } k \geq 1 \end{cases} \quad (9)$$

con $R_k = \theta_k f_{l_k} + (1 - \theta_k) \|F_k\|$, para $0 \leq \theta_k \leq 1$, q_k es un entero positivo, que se incrementa en cada iteración interna cuando la relación entre la reducción alcanzada y la predicha no es la esperada, y $0 \leq \theta_k \leq 1$.

Se construye de esta manera un método adaptativo no monótono de región de confianza.

3 Aproximación Secante Estructurada

Como ya mencionamos, se han desarrollado muchos algoritmos específicos para resolver el problema de mínimos cuadrados que explotan su particular estructura. Los métodos Gauss Newton y de Levenberg Marquadt [1,14] se basan en la observación que la parte de segundo orden de la matriz Hessiana tiende a cero en el óptimo cuando los problemas son de residuo nulo. En ese caso tienen una tasa de convergencia q-cuadrática, pero su performance es pobre cuando el residuo en la solución es no nulo. Para mejorar esta situación fueron propuestos los métodos cuasi-Newton estructurados. A partir de la actualización estructurada BFGS que fue propuesta por Al Baali et al. [15], en 1989, Dennis et al. [16] derivaron un principio que tiene en cuenta la estructura con el que consiguieron convergencia local superlineal. Mas recientemente Huschens [17] propuso un nuevo método cuasi Newton estructurado que converge cuadráticamente para problemas de residuo cero y superlinealmente para los de residuo no nulo. Otros trabajos han sido publicados posteriormente tales como [18,19].

Consideramos la matriz Hessiana compuesta por un término que contiene información de primer orden, que es la parte disponible, y otro que contiene la información de segundo orden, es decir

$$\nabla^2 f(x) = J(x)^T J(x) + S(x) = C(x) + S(x), \quad (10)$$

donde $C : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ y $S(x) = \sum_{i=1}^m F_i(x) \nabla^2 F_i(x)$.

El objetivo es construir una aproximación secante de la porción que contiene la información de segundo orden.

Si B_{k+1} es una aproximación de $\nabla^2 f$ y A_{k+1} es una aproximación de S entonces

$$B_{k+1} = C_{k+1} + A_{k+1}, \quad (11)$$

con $A_{k+1} = A_k + \Delta(s_k, y_k^\sharp, A_k, v_k)$.

Si utilizamos una actualización secante del tipo BFGS

$$\Delta(s, y^\sharp, A, v) = \frac{(y^\sharp - As)v^T + v(y^\sharp - As)^T}{v^T s} - \frac{(y^\sharp - As)^T s}{(v^T s)^2} v v^T, \quad (12)$$

donde

$$v(s, y, B) = y + \left(\frac{y^T s}{s^T B s} \right)^{1/2} B s, \quad (13)$$

$$y_k = y_k^\sharp + J_{k+1}^T J_{k+1} s_k, \quad (14)$$

e y_k^\sharp es una aproximación de $S_{k+1} s_k$. De modo que

$$B_{k+1} = C_{k+1} + A_k + \Delta(s_k, y_k^\sharp, A_k, v(s_k, y_k, B_k)), \quad (15)$$

y entonces es razonable definir

$$A_{k+1} = A_k + \Delta(s_k, y_k^\sharp, A_k, v(s_k, y_k, B_k)) \quad (16)$$

como la actualización secante estructurada de A .

3.1 Elecciones de y^\sharp

En este trabajo se realizaron evaluaciones numéricas para tres elecciones diferentes de y^\sharp .

Dennis [20] y Bartholomew-Biggs [21] introdujeron independientemente la siguiente aproximación

$$y_k^\sharp = (J_{k+1} - J_k)^T F_k. \quad (17)$$

Ellos derivaron esta expresión considerando la ecuación secante de A_{k+1} y observando que

$$\nabla^2 F_i(x_{k+1})s \approx \nabla F_i(x_{k+1}) - \nabla F_i(x_k), \quad (18)$$

para $i = 1, \dots, m$. Entonces

$$\left[\sum_{i=1}^m F_i(x_{k+1}) \nabla^2 F_i(x_{k+1}) \right] s = \sum_{i=1}^m F_i(x_{k+1}) [\nabla^2 F_i(x_{k+1})s] \quad (19)$$

$$\approx \sum_{i=1}^m F_i(x_{k+1}) [\nabla F_i(x_{k+1}) - \nabla F_i(x_k)] \quad (20)$$

$$= (J_{k+1} - J_k) F_k. \quad (21)$$

Por otro lado, en su trabajo Huschens [17], utilizando un escalamiento sobre la matriz Hessiana del problema

$$\nabla^2 f(x) = C(x) + \|F(x)\| \sum_{i=1}^m \frac{F_i(x)}{\|F(x)\|} \nabla^2 F_i(x) \quad (22)$$

y

$$B(x) = C(x) + \|F(x)\|A, \quad (23)$$

mostró cómo la técnica del producto estructurado puede ser explotado para obtener valores de y^\sharp que permiten conseguir mejores resultados de convergencia. A partir de esta consideración, en los trabajos de Zhang et al. [18] y el de Zhou y Chen [19], se obtuvieron las siguientes aproximaciones:

$$y_k^\sharp = \frac{(J_{k+1} - J_k)^T F_k}{\|F_k\|} + \frac{3F_k}{\|s_k\|^2} [(J_{k+1} - J_k)s_k - 2(F_{k+1} - F_k)], \quad (24)$$

$$y_k^\sharp = (J_{k+1} - J_k)^T F_{k+1} \frac{\|F_{k+1}\|}{\|F_k\|}. \quad (25)$$

4 Descripción del Algoritmo Propuesto

Dados $x_k \in \mathbb{R}^n$, las constantes positivas δ_0, δ_{max} , $0 \leq \eta_1 \leq \eta_2 < 1$, y una matriz simétrica $B_k \in \mathbb{R}^{n \times n}$:

Paso I. Si x_k es un punto estacionario del problema el algoritmo termina, de otro modo, ir al Paso II.

Paso II. (Cálculo del paso de prueba)
Resolver aproximadamente el problema (4) y obtener s_k .

Paso III. (Evaluación del paso de prueba)
Calcular A_{redk} , P_{predk} y r_k de (5) y (6).

Paso IV. (Actualización de la región de confianza)

- Si $r_k < \eta_1$ entonces $x_{k+1} = x_k$ y actualizar δ_k de acuerdo a (9).
- Si $r_k > \eta_2$ entonces $x_{k+1} = x_k + s_k$ $\delta_k = \min(2\delta_k, \delta_{max})$.
- De lo contrario $x_{k+1} = x_k + s_k$ y $\delta_{k+1} = \delta_k$.

Paso V. (Actualización de la información)
Calcular $B_{k+1} = J_{k+1}^T J_{k+1} + \alpha A_{k+1}$ de acuerdo a (16) e ir al Paso I.

Nota 1. Observemos que el valor de α es 1 si utilizamos (17) y (25) y $\|F_{k+1}\|$ si utilizamos (24).

5 Resultados Numéricos

A fin de exhibir el comportamiento del algoritmo consideramos un conjunto de 26 problemas test tomados de la literatura [22,18]. Cada problema fue considerado con los puntos iniciales propuestos en [18]. Entre los problemas resueltos, 16 de ellos pertenecen al grupo de problemas de residuo no nulo y en la Tabla 1 están señalados con (*).

El algoritmo se codificó en Scilab con sistema operativo Linux. El código se ha ejecutado en una computadora personal con procesador INTEL Core i3 con 3.8 Gb de memoria RAM usando Ubuntu 14.04.

El subproblema de región de confianza se resolvió usando un método cuasi-Newton con las diferentes aproximaciones de la matriz Hessiana descritas en la sección 3. El objetivo de la experimentación numérica es mostrar el comportamiento del algoritmo con las tres formas de estructurar la matriz.

Los parámetros del algoritmo fueron elegidos de la siguiente manera: $c = 0.5$, $\eta_1 = 0.01$ y $\eta_2 = 0.75$.

En la tabla 1, el número de variables del problema se indica con n , y m es el número de funciones componentes del residuo. Se reportan el número de

iteraciones para cada una de las actualizaciones secantes estructuradas elegidas. Con ASDMT nos referimos al algoritmo propuesto con la actualización secante de Dennis, Martínez y Tapia [16] que corresponde a la elección de $y^\#$ (17), ASZXZ corresponde a la propuesta de Zhang, Xue y Zhang [18], con $y^\#$ de acuerdo a (24), y, finalmente, ASZCH indica que la actualización secante estructurada utiliza la expresión (25), propuesta por Zhou y Chen [19]. Hemos señalado con (–) cuando el problema no pudo ser resuelto por el algoritmo usando una actualización determinada para el número máximo de iteraciones establecido.

Para visualizar el comportamiento de las distintas versiones del algoritmo hemos usado, además, gráficos de perfiles de rendimiento [23], donde la cantidad de iteraciones externas ha sido el parámetro de contraste.

Claramente, sobre los problemas de residuo no nulo, el algoritmo ASDMT es el más eficiente, mientras que sobre los de residuo nulo resulta que ASZCH es el más eficaz: resuelve con menor o igual número de iteraciones en el 90% de los casos.

Si el usuario desconoce *a priori* si el problema es de residuo nulo o no nulo, la fig. 3 muestra que ASDMT es el algoritmo más confiable para resolverlo.

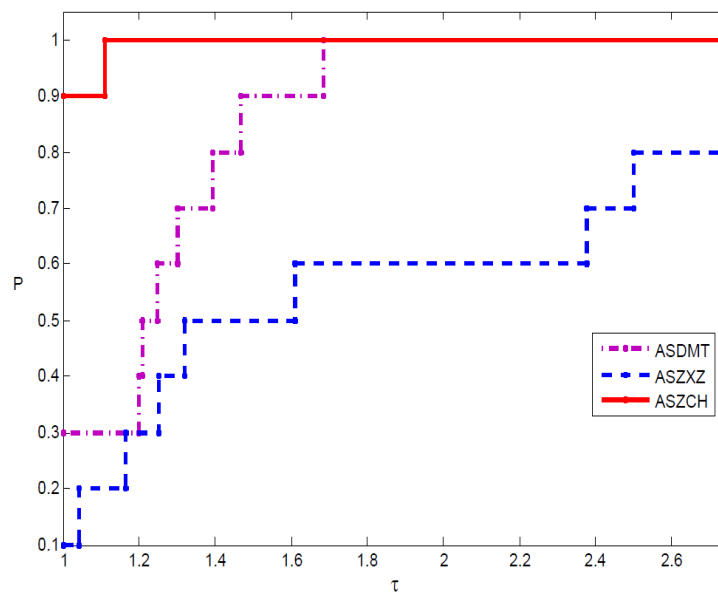


Fig. 1. Perfil de rendimiento para los problemas de residuo nulo

Tabla 1. Problemas test considerados

Problema	n	m	ASDMT	ASZXZ	ASZCH
Chained Rosembrock	40	78	113	183	77
Chained Wood	100	186	165	158	116
Chained Powell singular	40	76	13	25	10
Chained Cragg and Levy (*)	100	245	34	95	129
Generalized Broyden triangular	60	60	5	11	4
Generalized Broyden banded	60	60	6	5	5
Chained Freudenstein and Roth (*)	140	248	17	–	22
Wright and Holt zero residual	60	300	6	7	6
Toint quadratic merging (*)	60	174	21	33	21
Chained exponential (*)	3	5	24	–	196
Chained serpentine	80	158	196	203	162
Chained and modified (HS47) (*)	11	18	91	–	347
Chained and modified (HS47) (*)	83	162	223	–	337
Chained and modified (HS48) (*)	29	63	53	–	52
Chained and modified (HS48) (*)	32	70	56	–	43
Chained and modified (HS48) (*)	38	84	62	–	62
Sparse signomial (*)	4	4	11	14	12
Sparse exponential (*)	16	28	5	–	5
Sparse trigonometric (*)	20	36	10	128	28
Countercurrent reactors (*)	40	40	355	–	343
Countercurrent reactors (*)	80	80	121	311	193
Tridiagonal system	140	140	39	37	28
Structured Jacobian problem	60	60	19	–	19
Shifted discrete boundary value	180	180	18	29	20
Modified discrete boundary value (*)	41	41	22	–	–
Attracting Repelling (*)	20	38	148	182	–

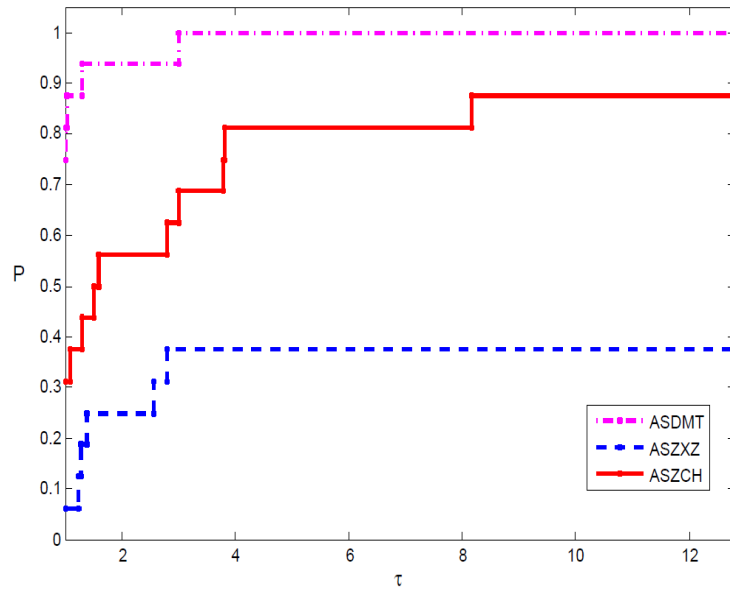


Fig. 2. Perfil de rendimiento para los problemas de residuo no nulo

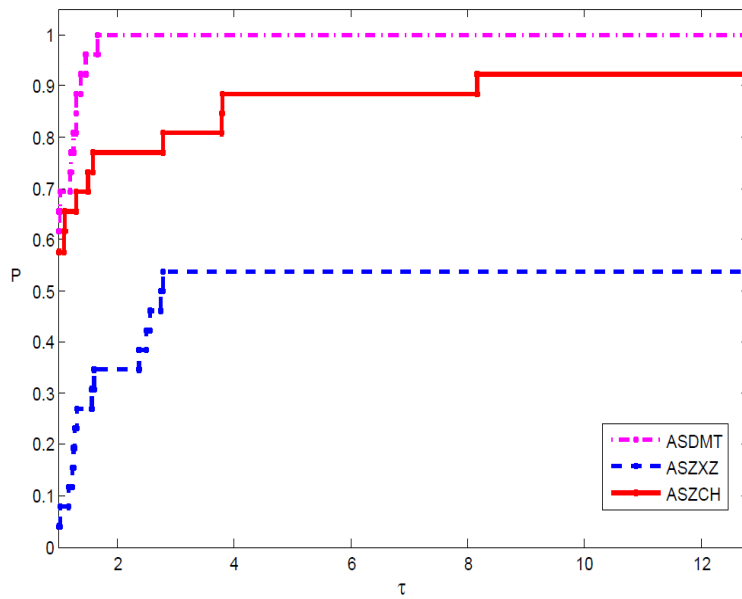


Fig. 3. Perfil de rendimiento para todos los problemas analizados

6 Conclusiones y Futuro Trabajo

En este trabajo hemos presentado un algoritmo para resolver el problema de cuadrados mínimos no lineales. Se utiliza una estrategia de región de confianza no monótona y un método cuasi Newton con diferentes elecciones de la estructura de la matriz Hessiana. Los resultados numéricos iniciales muestran que la técnica es eficiente para resolver tanto problemas de residuo nulo como no nulos.

El algoritmo se usó para resolver un conjunto de problemas test, que si bien puede considerarse pequeño, permite conjeturar que el método propuesto resulta eficiente.

Continuando con esta investigación, pretendemos realizar una discusión de los resultados teóricos y propiedades de convergencia del algoritmo y extender el método, de modo tal de poder abordar problemas con restricciones.

Bibliografía

1. Dennis, J., Schnabel, R.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. SIAM, Philadelphia (1996)
2. Conn, A., Gould, N., Toint, P.: Trust Region Methods. SIAM, Philadelphia (2000)
3. Moré, J.J.: Recent developments in algorithms and software for trust region methods. In A. Bachem, M. Grottschel and B. Korte, eds., *Mathematical Programming: The state of the art*, Springer-Verlag, New York, 258–287 (1984)
4. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.* 23 (4), 707–716 (1986)
5. Deng, N., Xiao, Y, Zhou, F.: Nonmonotonic trust region algorithm. *Journal of Optimization Theory and Applications* 76, 259–285 (1993)
6. Sun, W.: Nonmonotone trust region method for solving optimization problems. *Applied Mathematics and Computation* 156, 159–74 (2004)
7. Mo, J., Zhang, K., Wei, Z.: A nonmonotone trust region method for unconstrained optimization. *Applied Mathematics and Computation* 171, 371–384 (2005)
8. Ahookhosh, M., Amini, K.: An efficient nonmonotone trust-region method for unconstrained optimization. *Numerical Algorithms* 59 (4), 523–540 (2012)
9. Chen, J., Sun, W., Yang, Z.: A non-monotone retrospective trust-region method for unconstrained optimization. *Journal of Industrial and Management Optimization* 9 (4), 919–944 (2013)
10. Zhang, J., Zhang, X.: A nonmonotone adaptive trust region method and its convergence. *Computers and Mathematics with Applications* 45, 1469–1477 (2003)
11. Fu, J., Sun, W.: Nonmonotone adaptive trust-region method for unconstrained optimization problems. *Appl. Math. Comput.* 163, 489–504 (2005)
12. Esmaeili, H., Kimiaei, M.: A new adaptive trust-region method for system of nonlinear equations. *Appl. Math. Modelling* 38, 3003–3015 (2014)
13. Kamandi, A., Amini, K.: An improved adaptive trust-region algorithm. *Optim. Lett.* DOI 10.1007/s11590-016-1018-4 (2016)
14. Fletcher, R.: *Practical Methods of Optimization*. John Wiley, Chichester and New York, (1987)
15. Al-Baali, M., Fletcher, R.: Variational methods for non-linear least squares. *J. Oper. Res. Soc.* 36, 405–421 (1985)

16. Dennis, J., Martínez, H.J. and Tapia, R.: Convergence theory for the structured BFGS secant methods with an application to nonlinear least squares. *Journal of Optimization Theory and Applications* 61 (2), 161–178 (1989)
17. Huschens, J.: On the use of product structure in secant methods for nonlinear least square problems. *SIAM J. Optim.* 4, 108–129 (1994)
18. Zhang, J., Xue, Y., Zhang, K.: A structured secant methods based on a new quasi-Newton equation for nonlinear least squares problems. *BIT Numerical Mathematics* 43, 217–229 (2003)
19. Zhou, W., Chen, X.: Global Convergence of a New Hybrid Gauss-Newton Structured BFGS Method for Nonlinear Least Squares Problems. *SIAM J. Optim.* 20 (5), 2422–2441 (2010)
20. Dennis, J.: A brief survey of convergence results for quasi Newton methods. *Non-linear Programming SIAM-AMS Proceedings* 9, 185–199 (1976)
21. Bartholomew–Biggs, M.: The estimation of the Hessian matrix in nonlinear least square problem with non zero residuals. *Mathematical Programming* 12, 67–80 (1977)
22. Luksan, L., Vlcek, J.: Sparse and partially separable test problems for unconstrained and equality constrained optimization. Technical report No. 767, Institute of Computer Science, Academy of Sciences of the Czech Republic (1999)
23. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. *Math. Program.*, 91, 201–213 (2002)