Journal of Mathematical Extension
Vol. 13, No. 1, (2019), 103-116
ISSN: 1735-8299
URL: http://www.ijmex.com

Using a Modified Secant Equation for Unconstrained Optimization

R. Dehghani*
Yazd University

M. M. Hosseini
Yazd University

Abstract. We make some efficient modifications on the modified secant equation proposed by Zhang and Xu (2001). Then we introduce modified BFGS method using propose secant equation, and obtain some attractive results in theory and practice. We establish the global convergence property of the proposed method without convexity assumption on the objective function. Numerical results on some testing problems from CUTEr collection show the priority of the proposed method to some existing modified secant methods in practice.

AMS Subject Classification: 90C53; 49M37; 65K05
Keywords and Phrases: Unconstrained optimization, modified secant equation, BFGS method

1. Introduction

Consider the unconstrained nonlinear optimization problem

$$\min f(x), \quad x \in \mathbb{R}^n, \tag{1}$$

Received: February 2017; Accepted: August 2018
*Corresponding author

where f is twice continuously differentiable. Secant methods are popular iterative methods for solving (1), with the iterates being constructed as follows:

$$x_{k+1} = x_k + \alpha_k d_k,$$

where α_k is a step size and d_k is a searchdirection obtained by solving $B_k d_k = -g_k$, with $g_k = \bigtriangledown f(x_k)$ and B_k an approximation of the Hessian matrix of f at x_k satisfying the secant equation.

The standard secan equation can be established as follows (see [9]). We have

$$g_{k+1} - g_k = \int_0^1 \bigtriangledown^2 f(x_k + ts_k)dts_k, \tag{2}$$

where $s_k = x_{k+1} - x_k$. Since B_{k+1} is to approximate $G(x_{k+1}) = \nabla^2 f(x_{k+1})$, the secant equation is defined to be

$$B_{k+1}s_k = y_k, \tag{3}$$

where $y_k = g_{k+1} - g_k$. The relation (3) is sometimes called the standard secant equation. (see Dennise and Moré [8] for a comprehensive treatment of quasi-Newton methods particularly the secant methods).

A family of secant methods is Broyden family [2] in which the updates are defined by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \mu w_k . w_k^T,$$

$$w_k = (s_k^T B_k s_k)^{1/2}[\frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k}], \tag{4}$$

where μ is a scale parameter. The BFGS, DFP and SR1 updates are obtained by setting $\mu = 0$, $\mu = 1$ and $\mu = 1/(1 - s_k^T B_k s_k / s_k^T B_k)$, respectively.

Among the secant methods, the most efficient quasi-Newton method is the BFGS method which was proposed by Broyden [2], Fletcher [10], Goldfarb [12] and Shanno [19].

When f is convex, the global convergence of the BFGS method have been studied by some authors (see [4, 5, 15, 19, 21]). Dai [5] have constructed an example to show that the standard BFGS method may fail for non-convex functions with inexact line search. Mascarenhas [18] showed that the nonconvergence of the standard BFGS method even with exact line search. Li and Fukushima [16, 17] made a modification on the standard BFGS method and developed a modified BFGS method that is globally convergent without a convexity assumption on the objective function f.

The usual Secant equation employs only the gradients and the available function values are ignored. In order to get a higher order accuracy of approximating the Hessian matrix of the objective function, several researchers have modified the usual Secant equation (3) to make full use of both the gradient and function values (see [22]-[25]).
Zhang and Xu [25] using Taylor's Series modified (3) as follows:

$$B_{k+1}s_k = y_k^{\neq}, \tag{5}$$

where

$$y_k^{\neq} = y_k + \frac{\vartheta_k}{\|s_k\|^2}s_k, \qquad \vartheta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k. \tag{6}$$

Using modified secant equation (5), they proposed the following BFGS update formula:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k^{\neq} y_k^{\neq^T}}{s_k^T y_k^{\neq} k}. \tag{7}$$

They showed that the modified BFGS method is localy and superlinearly converge with the assumption f is uniformly convex function. However, if f is a general function may this method nonconvergence.
This motivated us to modification on modified secant equation (5).

Then, we make use of the new secant equation in a BFGS update formula. Under some proper assumptions, we prove the global convergence property for general functions.

The rest of our work is organized as follows: In Section 2, we introduce an alternative approximation of the secant equation. In Section 3, we investigate the convergence of our proposed method. In Section 4, we report the numerical results.

2. Modified Secant Equation

In this section, we first describe a modified secant equation in [25] that utilize both the available gradient and function values. Moreover, this method has a better theoretical feature than the usual secant equation and the secant equation introduced in [22].
Using the Taylor's series for $f(x)$, we have

$$f_k = f_{k+1} - g_{k+1}^T s_k + \frac{1}{2} s_k^T G_{k+1} s_k - \frac{1}{3!} s_k^T (T_{k+1} s_k) s_k + O(\| s_k \|^4), \quad (8)$$

and

$$s_k^T g_k = s_k^T g_{k+1} - s_k^T G_{k+1} s_k + \frac{1}{2} s_k^T (T_{k+1} s_k) s_k + O(\| s_k \|^4), \quad (9)$$

where

$$s_k^T (T_{k+1} s_k) s_k = \sum_{i,j,l=1}^{n} \frac{\partial^3 f(x_{k+1})}{\partial x^i \partial x^j \partial x^l} \partial s_k^i \partial s_k^j \partial_k s^l. \quad (10)$$

Cancelation of the terms which include the tensor yields

$$s_k^T G_{k+1} s_k \simeq (g_{k+1} - g_k)^T s_k + 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k. \quad (11)$$

Then using a new approximation B_{k+1}, we have

$$s_k^T B_{k+1} s_k = y_k^T s_k + \vartheta_k, \quad (12)$$

where

$$\vartheta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k. \quad (13)$$

This suggests the following new secant equation

$$B_{k+1} s_k = y_k^{\neq}, \quad (14)$$

where $y_k^{\neq} = y_k + \frac{\vartheta_k}{\|s_k\|^2} s_k$ and $\vartheta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k$.
One theoretical advantages of the new modified secant method can be seen from the following theorem [24].

Theorem 2.1. *Suppose that the function f is sufficiently smooth. If $\|s_k\|$ is small enough, then we have:*

$$s_k^T(G_{k+1}s_k - y_k^{\neq}) = O(\| s_k \|^4), \tag{15}$$

and

$$s_k^T(G_{k+1}s_k - y_k) = O(\| s_k \|^3). \tag{16}$$

Notice that if the objective function f is uniformly convex, then

$$\begin{aligned} y_k^{\neq T} s_k &= y_k^T s_k + 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k \\ &= 6(f_k - f_{k+1}) + 2(g_k + 2g_{k+1})^T s_k > 0. \end{aligned}$$

which guarantees the positive definite of the matrix B_k for uniformly convex function. However, if f is a general function, may happen $y_k^{\neq T} s_k \leqslant 0$. Hence the positive definiteness of the matrix B_k can not be guarantee for general function.

In addition, Theorem 2.1, demonstrate if $\|s_k\| > 1$, the standard secant equation is expected to be more accurate than the modified secant equation (14). In this case, the use of (14) does not seem to be suitable. To overcome these problems, we introduce some modification on (14) as follows:

$$B_{k+1}s_k = y_k^*, \tag{17}$$

with

$$y_k^* = \overline{y}_k + \gamma \|g_k\|^2 s_k + \max(-\frac{\overline{y}_k^T s_k}{\|s_k\|^2}, 0)s_k \tag{18}$$

where $\overline{y}_k = y_k + \rho_k \frac{\vartheta_k}{\|s_k\|^2} s_k$ and γ is a positive constant and

$$p_k = \begin{cases} e^{-||s_k||}, & \text{for} \quad ||s_k|| \leqslant 1, \\ 0, & \textit{otherwise}. \end{cases} \tag{19}$$

Clearly, if $\|s_k\| \longrightarrow 0$ then $\rho_k \longrightarrow 1$ and if $\|s_k\| > 1$ then $\rho_k = 0$.

It is easy to see

$$y_k^{*T} s_k \geqslant \rho_k \gamma \|g_k\|^2 \|s_k\|^2 > 0, \tag{20}$$

which guarantees that positive definite of the update matrix B_k, for the general function.

We can now give a new BFGS algorithm using our new secant relation for solving (1) as follows.

Algorithm 1: New modified BFGS method:

Step 1: given ε as a tolerance for convergence, $\sigma_1 \in (0,1)$, $\sigma_2 \in (\sigma_1, 1)$, a sarting point $x_0 \in \mathbb{R}^n$, and a positive definite matrix B_0. Set $k = 0$.

Step 2: If $\|g_k\| < \varepsilon$ then stop.

Step 3: Compute a search direction d_k : Solve $B_k d_k = -g_k$.

Step 4: Compute α_k by using the following Wolfe conditions:

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \sigma_1 \alpha_k g_k^T d_k, \tag{21}$$

and

$$g(x_k + \alpha_k d_k)^T d_k \geqslant \sigma_2 g_k^T d_k. \tag{22}$$

Step 5: Set $x_{k+1} = x_k + \alpha_k d_k$. Compute y_k^* by (18). Update B_{k+1} by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k^* {y_k^*}^T}{s_k^T y_k^*}. \tag{23}$$

Step 6: Set $k = k + 1$ and go to Step 2.

Next, we will investigate the global and superlinear convergence of the proposed Algorithm without convexity assumption on the objective function f.

3. Convergence analysis

Here, we establish the convergence of Algorithm 1. We need the following usual assumptions.

Assumption A. The level set $D = \{x \mid f(x) \leqslant f(x_0)\}$ is bounded.

Assumption B. The function f is continuously differentiable on D, and there is a constant $L \geqslant 0$ such that, for all $x, y \in D$, $\| g(x) - g(y) \| \leqslant L \| x - y \|$.

Clearly, these assumptions imply that there exists a constant $m > 0$ such that

$$\|g(x)\| \leqslant m, \qquad \forall x \in D. \tag{24}$$

From assumption A and the Wolfe conditions we deduce that $\{f(x_k)\}$ is a nonincreasing sequence, which ensures $\{x_k\} \subset D$ and the existence of x^* such that

$$\lim_{k\to\infty} f(x_k) = f(x^*). \tag{25}$$

In order to establish the global convergence of Algorithm 1, we present the following useful Lem- mas.

Lemma 3.1. *Suppose that Assumption B holds and y_k^* define by (14). Then, there exist*

$$\|y_k^*\| \leqslant M\|s_k\|, \tag{26}$$

where M is positive constant.

Proof. Considering Assumption B and definition of ϑ_k, we have

$$|\vartheta_k| \leqslant 3L\|s_k\| \qquad \text{(See the relations leading to (5.10) of [23])}. \tag{27}$$

Therefore, since $\rho_k \in [0,1]$ we can give

$$\|\overline{y}_k\| = \|y_k + \rho_k \frac{\vartheta_k}{\|s_k\|^2} s_k\| \leqslant \|y_k\| + \frac{|\vartheta_k|}{\|s_k\|} \leqslant L\|s_k\| + 3L\|s_k\| = 4L\|s_k\|. \tag{28}$$

Hence, from (24) and (28) we have

$$\|y_k^*\| \leqslant 2\|\overline{y}_k\| + \gamma\|g_k\|^2\|s_k\| \leqslant (2L + \gamma m^2)\|s_k\|. \tag{29}$$

This equation imply the (26) with $M = 2L + \gamma m^2$. $\square$

Lemma 3.2. *Let f satisfy assumptions A and B, and $\{x_k\}$ be generated by Algorithm 1 and there exist constants a_1 and a_2 such that*

$$\|B_k s_k\| \leqslant a_1\|s_k\|, \qquad s_k^T B_k s_k \geqslant a_2\|s_k\|^2, \tag{30}$$

for infinitely many k. Then, we have

$$\liminf_{k\to\infty} g(x_k) = 0. \tag{31}$$

Proof. Since $s_k = \alpha_k d_k$, it is clear that (30) holds true if s_k is replaced by d_k. From (30) and the relation $g_k = -B_k d_k$, we have

$$d_k^T B_k d_k \geqslant a_2 \|d_k\|^2, \qquad a_2\|d_k\| \leqslant \|g_k\| \leqslant a_1\|d_k\|. \tag{32}$$

Let Λ be the set of indices k for which (30) hold. By using (22) and Assumption B, we have

$$L\alpha_k \|d_k\|^2 \geqslant (g_{k+1} - g_k)^T d_k \geqslant -(1-\sigma_2) g_k^T d_k. \tag{33}$$

This implies that, for any $k \in \Lambda$,

$$\alpha_k \geqslant \frac{-(1-\sigma_2) g_k^T d_k}{L\|d_k\|^2} = \frac{(1-\sigma_2) d_k^T B_k d_k}{L\|d_k\|^2} \geqslant \frac{(1-\sigma_2) a_2}{L}. \tag{34}$$

Moreover, by (25), we obtain

$$\sum_{k=1}^{\infty} (f_k - f_{k+1}) = \lim_{N\to\infty} \sum_{k=1}^{N} (f_k - f_{k+1}) = \lim_{N\to\infty} f(x_1) - f(x_N)) = f(x_1) - f(x^*),$$

which yields

$$\sum_{k=1}^{\infty} (f_k - f_{k+1}) < \infty.$$

Using (21), we obtain

$$\sum_{k=1}^{\infty} \alpha_k g_k^T d_k < \infty,$$

which ensures

$$\lim_{k\to\infty} \alpha_k g_k^T d_k = 0.$$

This together with (34) lead to

$$\lim_{k\in\Lambda, k\to\infty} d_k^T B_k d_k = \lim_{k\in\Lambda, k\to\infty} -g_k^T d_k = 0.$$

which long with (32), yields (31). $\square$

The following Lemma is taken from [3], we represent it here but omit the proof.

Lemma 3.3. *(Theorem 2.1 of [3]) If there are positive constants M_1 and M_2 such that for all $k \geqslant 0$,*

$$\frac{\|y_k^*\|^2}{s_k^T y_k^*} \leqslant M_1 \quad \text{and} \quad \frac{s_k^T y_k^*}{\|s_k\|^2} \geqslant M_2, \tag{35}$$

then there exist constants a_1 and a_2 such that, for any positive integer t, (30) holds for at least $[\frac{t}{2}]$ values of $k \in \{1, 2, ..., t\}$.

Now, we prove the global convergence for Algorithm 1.

Theorem 3.4. *Let f satisfy the assumptions A and B, and $\{x_k\}$ be generated by Algorithm 1. Then (31) holds.*

Proof. Assume, to the contrary that the conclusion is not true. Then, there exists a positive constant δ such that, for all k,

$$\|g_k\| > \delta, \tag{36}$$

Hence, (20) imply that

$$y_{k^T}^* s_k \geqslant \delta^2 \|s_k\|^2, \tag{37}$$

Therefore, using Lemma 3.1 and (37), we obtain

$$\frac{\|y_k^*\|^2}{s_k^{*T} y_k^*} \leqslant M, \qquad \forall k \geqslant 0.$$

Hence, (35) holds for all k. Using Lemma 3.3 to the subsequence $\{B_k\}_{k \in K}$, clearly there exist constants $a_1 > 0$ and $a_2 > 0$ such that (30) holds for infinitely many k. Then Lemma 3.2 compltes the proof. $\square$

The above theorem shows a global convergence property of MBFGS method without convexity assumption on the objective function.

To establish the superlinear convergence of Algorithm 1, we need further the following assumption.

Assumption C. Suppose $x_k \to x^*$ at which $g(x^*) = 0$, $G(x)$ is positive definite, and $G(x)$ is Holder continuous at x^*, i.e., there exist constants $v \in (0, 1)$ and $L_2 > 0$ such that

$$\|G(x) - G(x^*)\| \leqslant L_2 \|x - x^*\|^v,$$

for all x in the neighborhood of x^*.

Similar to the proof of Theorem 3.8 in [17], it is not difficult to prove the superlinear convergence of Algorithm 1. Hence, we only present the theorem without giving the proof.

Theorem 3.5. *Suppose that assumptions A, B and C hold. Let the sequence $\{x_k\}$ be generated by the Algorithm 1. Then $\{x_k\}$ is superlinearly convergent.*

4. Numerical Results

We are to compare the performance of the following four methods on some unconstrained optimization problems:

BFGS: the usual BFGS method.
MBFGS: proposed method (Algorithm 1).

ZXMBFGS: the modified BFGS of Zhang and Xu using (5) [25].

BFGS$A_k(2)$: using the following Modified secant equation to update B_k which suggest by Wei et.al in [23]

$$B_{k+1} s_k = y_k + \frac{\theta_k}{\|s_k\|^2} s_k, \tag{38}$$

where $\theta_k = 2(f_k - f_{k+1}) + (g_k + g_{k+1})^T s_k$.

We have tested all the considered algorithms on 120 test problems from CUTEr library [14]. A summary of these problems are given in Table 1 of [7]. All codes were written in Matlab 2012 and run on PC with CPU Intel(R) Core(TM) i5-4200 3.6 GHz, 4 GB of RAM memory and Centos 6.2 server Linux operation system.

In all algorithms, the initial matrix is $B_0 = I$ and the steplengths α_k satisfyies the Wolfe conditions, with $\sigma_1 = 0.001$, and $\sigma_2 = 0.1$.

For all the test problems, the termination condition is $\|g_k\| \leqslant 10^{-5}$, or $\|f_{k+1} - f_k\| \leqslant 10^{-20} \max(1, f_k)$.

We use the profiles of Dolan and Moré (see [9]) to evaluate performance of these four algorithms with respect to number of iteration and the total number of function and gradient evaluations being equal to $N_f + nN_g$, where N_f and N_g denote the number of function and gradient evaluations respectively. Figs 1 and 2 demonstrate the results of the comparisons. From Figs. 1 and 2, it is easy to observe that the MBFGS method is the most efficient for solving these 120 test problems among the four methods. We see that MBFGS method solves about 75% and 74% of the test problems with the fewest number of iterations and function evaluations, respectively.

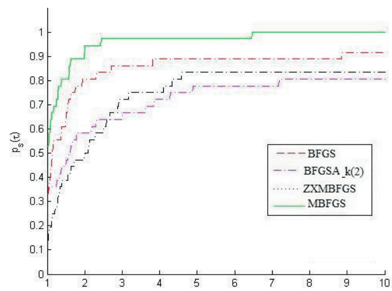

Figure 1. The Dolan-Moré performance profiles using number of function evoluations.

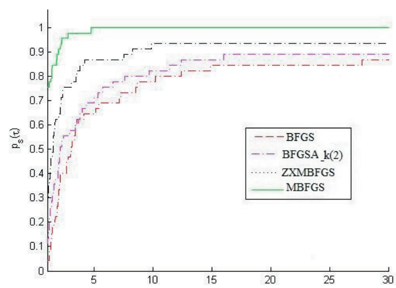

Figure 2. The Dolan-Moré performance profiles using number of iterations.

5. Conclusion

We introduced a modified BFGS (MBFGS) method using a new secant equation. An interesting feature of the proposed method is to take both the gradient and function values into account. Under suitable assumptions, we established the global convergence of the proposed method for the general functions. Numerical results on the collection of problems in CUTER show that the proposed method is efficient as compared to several proposed BFGS methods.

References

[1] C. G. Broyden, A class of methods for solving nonlinear simultaneous equations, *Math. Comp.,* 19 (1965), 577-593.

[2] C. G. Broyden, The convergence of a class of double rank minimization algorithms-Part 2: The new algorithm, *J. Inst. Math. Appl.,* 6 (1970), 222-231.

[3] R. Byrd and J. Nocedal, A tool for the analysis of quasi-Newton methods with application to unconstrained minimization, *SIAM J. Numerical Anal.,* 26 (1989), 727-739.

[4] R. Byrd, J. Nocedal, and Y. Yuan, Global convergence of a class of quasi-Newton methods on convex problems, *SIAM J. Numer. Analysis*, 24 (1987), 1171-1189.

[5] Y. Dai, Convergence properties of the BFGS algorithm, *SIAM J. Optim.,* 13 (2003), 693-701.

[6] R. Dehghani, R. Bidabadi, and M. M. Hosseini, A new modified BFGS method for unconstrained optimization problems, *Computational and Applied Mathematics*, 37 (2018), 5113-5125.

[7] J. E. Dennis and J. J. More, A characterization of superlinear convergence and its application to quasi-Newton methods, *Math. Comp.,* 28 (1974), 549-560.

[8] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

[9] E. Dolan and J. J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.,* 91 (2002), 201-213.

[10] R. Fletcher, A new approach to variable metric algorithms, *Comput,* 13 (1970), 317-322.

[11] J. A. Ford and A. F. Saadallah, On the construction of minimisation methods of quasi-Newton type, *J. Comput. Appl. Math.,* 20 (1987), 239-246.

[12] D. Goldfarb, A family of variable metric methods derived by variational means, *Math. Comp.,* 24 (1970), 23-26.

[13] N. I. M. Gould, D. Orban, and Ph. L. Toint, CUTEr, a constrained and unconstrained testing environment, revisited, ACM Trans. *Math. Softw.,* 29 (2003), 373-394.

[14] A. Griewank, The global convergence of partioned BFGS on problems with convex decompositons and Lipschitzian gradients, *Math. Program.,* 50 (1991), 141-175.

[15] D. H. Li and M. Fukushima, On the global convergence of BFGS method for nonconvex unconstrained optimization problems, *SIAM Journal on Optimization,* 11 (2001), 1054-1064.

[16] D. H. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, *J. Comput. Appl. Math.,* 129 (2001), 15-35.

[17] W. F. Mascarenhas, The BFGS method with exact line searches fails for non-convex objective functions, *Math. Program,* 99 (2004), 49-61.

[18] M. J. D. Powell, *Some Global Convergence Properties of a Variable Metric Algorithm for Minimization Without Exact Line Searches.* In: Cottle, R. W., Lemke, C. E. (eds.), Nonlinear Programming, SIAMAMS Proceedings, Philadelphia, (1976).

[19] D. F. Shanno, Conditioning of quasi-Newton methods for function minimization, *Math. Comp.,* 24 (1970), 647-656.

[20] Ph. L. Toint, Global convergence of the partioned BFGS algorithm for convex partially separable opertimization. *Math. Program,* 36 (1986), 290-306.

[21] Z. Wei, G. Li, and L. Qi, New quasi-Newton methods for unconstrained optimization problems, *Appl. Math. Comput.,* 175 (2006), 1156-1188.

[22] G. Yuan and Z. Wei, Convergence analysis of a modified BFGS method on convex minimizations, *Computational Optimization and Applications,* 47 (2010), 237-255.

[23] J. Z. Zhang, N. Y. Deng, and L. H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, *J. Optim. Theory Appl.,* 102 (1999), 147-167.

[24] J. Z. Zhang and C. X. Xu, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equation, *J. Comput. Appl. Math.,* 137 (2001), 269-278.

[25] D. T. Zhu, Nonmonotone backtracking inexact quasi-Newton algorithms for solving smooth nonlinear equations, *Appl. Math. Comput.,* 161(2005), 875-895.

Razieh Dehghani
Ph.D student of Mathematics
Department of Mathematics
Faculty of Mathematical Sciences
Yazd University
Yazd, Iran
E-mail:rdehghani@stu.yazd.ac.ir

Mohammad Mehdi Hosseini
Professor of Mathematics
Department of Mathematics
Faculty of Mathematical Sciences
Yazd University
Yazd, Iran
E-mail: hosse_ m@yazd.ac.ir.