Development a new conjugate gradient algorithm for improvement feedforward artificial neural networks for predicting the numbers of the Iraqi population

Neaam H. alfahady¹, Khalil K. Abbo², Edrees M. Nori Mahmood³

¹University of Mosul, Department of Operations Research and Intelligent Techniques, Mosul, Iraq
²University of Telafer, Department of Mathematics, Telafer, Iraq,
³University of Mosul, Department of Operations Research and Intelligent Techniques, Mosul, Iraq
Email: Neam.alfahady@uomosul.edu.iq,
    dr.khalil-khuder@uotelafer.edu.iq,
    edreesnori@uomosul.edu.iq

Abstract. A conjugate gradient method has been proposed for use in the training of Artificial Neural Networks with the feedforward technique. The proposed method demonstrated that the proposed CG method is accurate in predicting the numbers of the Iraqi population, the Our numerical results where very close to the exact solution of Tomas Malthose Model.

1. Introduction
As the human population is not a static factor, population growth is one of the key concerns of the world today. Instead, it is rising at an unprecedented pace. Despite the growing population of the planet, the resources of the earth remain constant. The need to maintain sustainable development is, thus becoming a big challenge for humanity today. All societies still need an accurate understanding of the future size of multiple entities for their future lives, such as population, wealth, demands, and consumption. Mathematical Modeling is a broad multidisciplinary science that uses mathematical and computational techniques to model and explains life science phenomena. A population model is a type of mathematical model that is applied to the study of population dynamics.
In this study, we have modeled the population growth in Iraq using the differential equation according to Thomas's theory, as well as the use of artificial neural networks, and then compared the two methods.

2. Thomas's theory [3][4]
Thomas theory is based on the assumption that when population growth is not controlled, it increases according to a geometric progression. Where, during short periods of time, the number of births and deaths is proportional to the size of the population and the length of time. While the food increases according to the sequence of my account.

3. Modeling the numbers of the population

Single population models using Thomas Malthus' theory is a matter of building a mathematical model of the population of a particular country or region by using differential equations to study the movement (dynamics) of living organisms.

3.1 Analysis of Population Growth in Mathematical Model of Thomas
According to Thomas Malthus' hypothesis that during time periods the numbers of births and deaths are proportional to both the size of society and the length of time model of population growth. It is based on the assumption that the population grows at a rate proportional to the size of the population. This is a reasonable assumption for a group of organisms under ideal conditions such as an indeterminate environment, adequate nutrition, and so on.

3.2 Mathematical formula
Assuming that:
- $N(t)$ Population variable with respect time $t$
- $\delta t$ Variable for a small-time specified.
- $A$ denotes the constant growth rate per capita of the population

$$\frac{\delta N(t)}{\delta t} = A N(t)$$

3.3 Mathematical solution[1][2]
The Mathematical model for the numbers of the population is a first-order differential equation.
\[ N(t) = Ce^{At} \]  \hspace{1cm} (2)

Here \( C \) is a constant. By solving this equation, we obtain the basic equation of Thomas's unconstrained growth model. It is stated that the size of the population would grow exponentially under these conditions by Malthus[2]

\[ N(t) = e^{At} \cdot N(0) \]  \hspace{1cm} (3)

Equation (3) is the population exponential of the fundamental equation of unfettered growth models Thomas. The population model is a deterministic model, A Thomas model (3) depends on a known constant quantity \( N(0) \) and an unknown constant “\( A \)”.

\[ A = \ln \frac{N(1)}{N(0)} \]  \hspace{1cm} (4)

4 Artificial Neural Networks:

Artificial neural networks (ANNs) are among the most powerful artificial intelligence techniques, as they are simulations of the biological neural network in the human brain. ANNs have many applications in different fields. One of the most amazing facts about ANNs is that they can compute any function at all.

In this section, we model the population growth of Iraq using ANNs (Figure 2. for the structure of the ANN used in this paper). A new gradient-based conjugate CG method is proposed in order to train the proposed ANNs. The simulation outcomes show that the suggested techniques are very effective in predicting the population growth of Iraq.
Fig. (1) a multi-layered network

4.1 Developed new CG algorithm [7][8][9]

In this section, we will show some of the conjugate gradient methods and then suggest a new algorithm for the conjugate gradient algorithm

\[ w_{k+1} = w_k + \alpha_k d_k \quad k \geq 1 \quad (5) \]

Where \( \alpha_k \) step-size that satisfy the standard Wolfe conditions

\[ E(w_k + \alpha_k d_k) \leq E(w_k) + \delta \alpha_k g_k^T d_k \quad (6) \]
\[ d_k^T g(w_k + \alpha_k d_k) \geq \sigma d_k^T g_k \quad (7) \]

or strong Wolfe conditions

\[ E(w + \alpha_k d_k) \leq E(w) + \delta \alpha_k g_k^T d_k \quad (8) \]
\[ |d_k^T g(w_k + \alpha_k d_k)| \leq -\sigma d_k^T g_k \quad (9) \]
\[ d_{k+1} = \begin{cases} -g_k, & k = 1 \\ -g_{k+1} + \beta_k d_k, & k \geq 1 \end{cases} \quad (10) \]

The Fletcher and Reeves (FR), Fletcher (CD), Polak and Ribiere (PRP) and \( \beta \) is scalar.

\[ \beta_{FR} = \frac{||g_{k+1}||^2}{||g_k||^2}, \text{ see [7]} \]
\[ \beta_{CD} = \frac{||g_{k+1}||^2}{g_k^T d_k}, \text{ see [8]} \]
\[ \beta_{PRP} = \frac{g_{k+1}^T g_k}{g_k^T g_k}, \text{ see [9]} \]

Where \( g_k = \nabla E(w_k) \), and let \( y_k = g_{k+1} - g_k \).
Now we suggest a new conjugate gradient algorithm the search direction can be viewed as a four-term extension of the FR method. When
\[ \beta^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \text{ see [7]} \]
and the Wolfe condition
\[
\begin{align*}
E(w_k + \alpha_k d_k) &\leq E(w_k) + \delta \alpha_k g_k^T d_k & (11) \\
d_k^T g(w_k + \alpha_k d_k) &\geq \sigma d_k^T g_k & (12)
\end{align*}
\]
where \(0 < \delta < \alpha\).

Then the search direction
\[
d_{k+1} = -g_{k+1} + \beta^{FR} d_k - t_k g_{k+1}^{T} d_{k+1} d_k - g_{k+1}^{T} \frac{d_{k+1} d_k}{d_k^T y_k} y_k
\]
(13)

Then from pure conjugacy condition
\[d_{k+1}^T y_k = 0\]

\[
d_{k+1} = -g_{k+1} + g_{k+1}^{T} g_k d_k + g_{k+1} g_k d_k y_k - t_k g_{k+1} d_k d_k^T y_k y_k - d_k^T y_k y_k = 0
\]
(14)

4.2 Algorithm

Step1. Given \(w \in R^n, \varepsilon > 0, d_1 = -g_1, K = 1\)

Step2. If \(\|g_{K+1}\| \leq \varepsilon\), stop, else go to Step 3

Step3. Find \(\alpha_k\) satisfying Wolfe condition (8) and (9).

Step4. Compute new iterative \(w_{k+1}\) by \(w_{k+1} = w_k + \alpha_k d_k\)

Step5. Compute \(\beta^{FR}\) and \(d_{k+1}\) from (13) and \(t_k\) from (14) and set
\[K = K + 1\]

Go to step (2).
4.3 The Descent Property of a CG New Method

The descent property for our proposed new conjugate gradient scheme is shown below, denoted by $\beta_k^{NEW}$.

4.3.1 Theorem (1)

The search direction $d_{k+1}$ and $\beta_k^{FR}$ with $t_k$ given as follows

$$d_{k+1} = -g_{k+1} + \beta_k^{FR} d_k - t_k \frac{g_{k+1}^T d_k}{d_k^T y_k} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k$$

for all $k \geq 1$

Proof

The proof is by using induction mathematical

1- If $k = 1$ then $g_1^T d_1 < 0$, $d_1 = -g_1 \rightarrow 0$.

2- Let the relation $g_k^T d_k < 0$ for all $k$.

3- We’re going to prove that the relation is true when $k = k + 1$ multiplying the equation in $g_{k+1}$ we obtain

$$d_{k+1} = -g_{k+1} + \beta_k^{FR} d_k - t_k \frac{g_{k+1}^T d_k}{d_k^T y_k} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k$$

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$$

$$t_k = \frac{1}{g_{k+1}^T d_k} \left[ g_{k+1}^T g_{k+1} - g_k^T g_k \right] - \frac{1}{d_k^T y_k} y_k^T y_k$$

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + B_{FR} d_k g_{k+1}^T - t_k \frac{g_{k+1}^T d_k}{d_k^T y_k} g_{k+1}^T d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} g_{k+1}^T y_k$$

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + B_{FR} d_k g_{k+1}^T - \frac{1}{g_{k+1}^T d_k} \left[ g_{k+1}^T g_{k+1} d_k^T y_k - g_{k+1}^T y_k \right]$$

$$- \frac{1}{d_k^T y_k} \left( g_{k+1}^T d_k \frac{g_{k+1}^T d_k}{d_k^T y_k} g_{k+1}^T d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} g_{k+1}^T y_k \right)$$

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + B_{FR} d_k g_{k+1}^T - \frac{1}{g_{k+1}^T d_k} \left[ g_{k+1}^T g_{k+1} d_k^T y_k - g_{k+1}^T y_k \right]$$

$$- \frac{1}{d_k^T y_k} \left( g_{k+1}^T d_k \frac{g_{k+1}^T d_k}{d_k^T y_k} g_{k+1}^T d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} g_{k+1}^T y_k \right) - \frac{g_{k+1}^T d_k}{d_k^T y_k} g_{k+1}^T y_k$$
4.3.2 Global convergence study

We will display that CG method with $\beta_k^{FR}$ and $t_k$ convergences globally. We need some assumptions for the convergence of the proposed new algorithm.

4.3.3 Assumption [10] [11] [12]

1- Assume $f$ is bound below in the level set $S = \{ w \in \mathbb{R}^n : \varphi(w) \leq \varphi(w_0) \}$; in some Initial point.
2- $\varphi$ is continuously differentiable and its gradient is Lipchitz continuous, there exist $L > 0$ such that:

$$\| g(w) − g(y) \| \leq L \| w − y \| \quad \forall w, y \in S .$$

On the other hand, under Assumption (1), it is clear that there exist positive constants $B$ such

$$\| w \| \leq B, \forall w \in S$$

$$\| \nabla E(w) \| \leq \bar{\gamma}, \forall w \in S$$

$$g_{k+1}^{\tau}d_{k+1} = -g_{k+1}^{\tau}g_{k+1} + B^{FR}d_kg_{k+1}^{\tau} - \left( \frac{1}{g_{k+1}^{\tau}d_k} g_{k+1}^{\tau}d_k g_{k+1}^{\tau}d_k y_k - g_{k+1}^{\tau}y_k \right)$$

$$g_{k+1}^{\tau}d_{k+1} = -g_{k+1}^{\tau}g_{k+1} + B^{FR}d_kg_{k+1}^{\tau} + \left( \frac{g_{k+1}^{\tau}d_k}{d_k y_k} - g_{k+1}^{\tau} \right)$$

$$g_{k+1}^{\tau}d_{k+1} = -g_{k+1}^{\tau}g_{k+1} + B^{FR}d_kg_{k+1}^{\tau} - \left( \frac{g_{k+1}^{\tau}d_k}{d_k y_k} - g_{k+1}^{\tau} \right)$$

$$g_{k+1}^{\tau}d_{k+1} = -g_{k+1}^{\tau}g_{k+1} - \frac{g_{k+1}^{\tau}d_k}{d_k y_k} y_k < 0$$

$$g_{k+1}^{\tau}d_{k+1} < 0$$
4.3.4 Lemma
Assume that Assumption (1) and equation (15) hold, take into consideration any conjugate gradient method in from (13) and (14), where \( d_k \) is a decent direction and \( \alpha_k \) is obtained by the S.W.L.S. If
\[
\sum_{k \geq 1} \frac{1}{\| d_{k+1} \|^2} = \infty
\]
then we have \( \lim_{k \to \infty} \| g_k \| = 0 \)
More details can be found in [13][14][15][16].

4.3.5 Theorem (2)
Assume that Assumption (1) and equation (15) and the descent condition hold. Consider a conjugate gradient scheme in the form
\[
d_{k+1} = -g_{k+1} + \beta_k^{NEW} d_k
\]
where \( \alpha_k \) is computed from strong Wolfe line search condition for more details see[17][18], If the objective function is uniformly on set S, then
\[
\lim_{n \to \infty} (in E \| g_k \|) = 0.
\]

Proof
\[
d_{k+1} = -g_{k+1} + \beta_k^{FR} d_k - t_k \frac{g_{k+1}^T d_k}{d_k^T y_k} d_k - \frac{g_{k+1}^T d_k}{y_k}
\]
\[
\beta_k^{FR} = \frac{g_{k+1}^T d_k}{g_k d_k}
\]
\[
t_k = \frac{1}{g_{k+1}^T d_k} \left[ g_{k+1}^T g_{k+1} g_k d_k y_k - g_{k+1}^T y_k \right] - \frac{1}{d_k^T y_k} y_k
\]
\[
\| d_{k+1} \| \leq \| g_{k+1} \| + \| B^{FR} d_k \| + t_k \frac{g_{k+1}^T d_k}{d_k^T y_k} d_k + \| \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k \| + \| \frac{g_{k+1}^T d_k}{d_k^T} \| + \| \frac{g_{k+1}^T d_k}{y_k} \|
\]
\[
\| d_{k+1} \| \leq \| g_{k+1} \| + \frac{\| g_{k+1} \|^2}{\| g_k \|} \| d_k \| + \| g_{k+1}^T d_k \| + \| \frac{g_{k+1}^T d_k}{d_k^T} \| + \| \frac{g_{k+1}^T d_k}{y_k} \| + \| \frac{g_{k+1}^T d_k}{y_k} \| + \| \frac{g_{k+1}^T d_k}{y_k} \|
\]
\[
\| d_{k+1} \| \leq \| g_{k+1} \| + \frac{\| g_{k+1} \|^2}{\| g_k \|} \| d_k \| + \| g_{k+1}^T d_k \| + \| \frac{g_{k+1}^T d_k}{d_k^T} \| + \| \frac{g_{k+1}^T d_k}{y_k} \| + \| \frac{g_{k+1}^T d_k}{y_k} \| + \| \frac{g_{k+1}^T d_k}{y_k} \|
\]
\[
\| d_{k+1} \| \leq \| g_{k+1} \| + \frac{\| g_{k+1} \|^2}{\| g_k \|} \| d_k \| + \| g_{k+1}^T d_k \| + \| \frac{g_{k+1}^T d_k}{d_k^T} \| + \| \frac{g_{k+1}^T d_k}{y_k} \| + \| \frac{g_{k+1}^T d_k}{y_k} \| + \| \frac{g_{k+1}^T d_k}{y_k} \|
\]
\[
\| d_{k+1} \| \leq 2 \| g_{k+1} \| + \frac{\| g_{k+1} \|^2}{\| g_k \|^2} \| d_k \| + t_k \frac{\| g_{k+1} \| \| d_k \|}{y_k} \\
\| d_{k+1} \| \leq (2 + \frac{\| g_{k+1} \|}{\| g_k \|^2} \| d_k \| + t_k \frac{\| d_k \|}{y_k}) \| g_{k+1} \| \\
\]

Let \( \alpha = (2 + \frac{\| g_{k+1} \|}{\| g_k \|^2} \| d_k \| + t_k \frac{\| d_k \|}{y_k}) \)

\[
\| d_{k+1} \| \leq \alpha \| g_{k+1} \| \\
\sum_{k \geq 1} \frac{1}{d_{k+1}} \geq \frac{1}{\alpha y} \sum_{k \geq 1} 1 = \infty
\]

Then

\[
\lim_{k \to \infty} g_k = 0
\]

5- Numerical results of the developed ANNs algorithm
Results were shown by using simulation by PC-Laptop Dell win10 compatible with MATLAB (2018) using the core i7 processor with RAM 8GB and 7.89HDD. The results of the simulation artificial neural network were obtained by using the developed new CG algorithm training that illustrates the temporal behavior of population growth. It turns out that the network is effective when used to forecast in the short and long term because it is close to the real values of the population census for subsequent years table 3.

| Year | numbers of the population P(t) | Tomas model N(t) | ANN |
|------|-------------------------------|------------------|-----|
| 1947 | 4.5600                        | 4.5600           | 4.5591 |
| 1957 | 6.3403                        | 6.3403           | 6.3414 |
| 1967 | 8.8156                        | 8.8156           | 8.8155 |
| 1977 | 12.2573                       | 12.2573          | 12.2577 |
| 1987 | 17.0427                       | 17.0427          | 17.0427 |
| 1997 | 23.6964                       | 23.6964          | 23.6963 |
| 2007 | 32.9477                       | 32.9477          | 32.9475 |
| 2017 | 45.8109                       | 45.8109          | 45.8109 |
| 2027 | -                             | 63.6961          | 63.6948 |
| 2037 | -                             | 88.5637          | 88.5649 |
| 2047 | -                             | 123.1401         | 123.1379 |
6- Conclusions
1. The main results of this study are summarized in the following.
2. A mathematical model for population growth is constructed based on Thomas’s theory.
3. ANNs have been employed to model and simulate population growth. Where a new gradient-based
   conjugate CG method is proposed in order to train the ANNs.
4. Simulation results showed that the ANNs are effective and accurate in predicting the population growth
   of Iraq.
5. The results of simulation of ANN prediction are better than Thomas' prediction model

References
[1] R. S. Robeva et al., “An Invitation to Biomathematics,” PhD Propos., 2015.
[2] Al-Khayat Basil Younis Thanoon, Marcov modeling with practical applications. Mosul: Ibn Al-
   Atheer House for Printing and Publishing, university of Mosul, 2011.
[3] S. M. Henson, F. Brauer, and C. Castillo-Chavez, “Mathematical Models in Population Biology and
   Epidemiology,” Am. Math. Mon., 2003, doi: 10.2307/3647954.
[4] D. K. Chaturvedi, Modeling and simulation of systems using MATLAB® and simulink®. 2017.
[5] A. Shrestha and A. Mahmood, “Optimizing deep neural network architecture with enhanced genetic
   algorithm,” 2019, doi: 10.1109/ICMLA.2019.00222.
[6] J. Bongaarts, “Human population growth and the demographic transition,” Philos. Trans. R. Soc. B
   Biol. Sci., 2009, doi: 10.1098/rstb.2009.0137.
[7] R. Fletcher and C. M. Reeves, “Function minimization by conjugate gradients,” Comput. J., vol. 7,
   no. 2, pp. 149–154, 1964, doi: 10.1093/comjnl/7.2.149.
[8] L. C. W. Dixon, “Conjugate gradient algorithms: quadratic termination without linear searches,” IMA J. Appl. Math., vol. 15, no. 1, pp. 9–18, 1975.

[9] E. Polak and G. Ribiere, “Note sur la convergence de méthodes de directions conjuguées,” ESAIM Math. Model. Numer. Anal. Mathématique Anal. Numérique, vol. 3, no. R1, pp. 35–43, 1969.

[10] K. K. Abbo and H. M. Khudhur, “New A hybrid Hestenes-Stiefel and Dai-Yuan conjugate gradient algorithms for unconstrained optimization,” Tikrit J. Pure Sci., vol. 21, no. 1, pp. 118–123, 2015.

[11] Y. A. Laylani, K. K. Abbo, and H. M. Khudhur, “Training feed forward neural network with modified Fletcher-Reeves method,” J. Multidiscip. Model. Optim., vol. 1, no. 1, pp. 14–22, 2018, [Online]. Available: http://dergipark.gov.tr/jmmo/issue/38716/392124#article_cite.

[12] Z. M. Abdullah, M. Hameed, M. K. Hisham, and M. A. Khaleel, “Modified new conjugate gradient method for Unconstrained Optimization,” Tikrit J. Pure Sci., vol. 24, no. 5, pp. 86–90, 2019.

[13] H. Tomizuka and H. Yabe, “A Hybrid Conjugate Gradient method for unconstrained Optimization,” 2004.

[14] Y. Dai and L. Liao, “New conjugacy conditions and related non-linear CG methods’ Appl,” Math optim.,(43). Spring-verlag, New York, 2001.

[15] H. N. Jabbar, K. K. Abbo, and H. M. Khudhur, “Four--Term Conjugate Gradient (CG) Method Based on Pure Conjugacy Condition for Unconstrained Optimization,” kirkuk Univ. J. Sci. Stud., vol. 13, no. 2, pp. 101–113, 2018.

[16] K. K. Abbo, Y. A. Laylani, and H. M. Khudhur, “Proposed new Scaled conjugate gradient algorithm for Unconstrained Optimization,” vol. 5, no. 7, 2016.

[17] M. Al-Baali, “Descent property and global convergence of the Fletcher—Reeves method with inexact line search,” IMA J. Numer. Anal., vol. 5, no. 1, pp. 121–124, 1985.

[18] L. Zhang and W. Zhou, “Two descent hybrid conjugate gradient methods for optimization,” J. Comput. Appl. Math., vol. 216, no. 1, pp. 251–264, 2008.

[19] Dean C. Karnopp, "SYSTEM DYNAMICS Modeling, Simulation, and Control of Mechatronic stems", Fifth Edition, WILEY, 2012,Canada.

[20] Chai T. and R. R. Draxler R. R., "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature" GE=eoscientific model Development, 2014.

[21] Yang I. sheng, "A Loss-Function for Causal Machine-Learning"2020

[22] Ali Abed , S., & A. Salim, M. (2019). THE FIRST OBSERVATION OF THE BANK MYNA ACRIDOTHERES GINGINIANUS (LATHAM, 1790) (FAMILY: STURNIDAE, CLASS: AVES) IN THE WILD IN IRAQ. Al-Qadisiyah Journal Of Pure Science, 24(1), 45 - 48.