Implementation and Comparative Analysis of AMGT Method in Maple 24: Convergence Performance in Optimization Problems

Mark Laisin & Rosemary U. Adigwe

Abstract

This study investigates the utilization of the Accelerated Modified Gradient Technique (AMGT) in Maple 24 for optimization problem-solving. The research focuses on enhancing and evaluating the convergence speed of AMGT in comparison to the Gradient Descent (GD) and Conjugate Gradient (CG) methods. The analysis covers a range of function types, such as convex functions (e.g., production cost, energy consumption, and transportation cost functions) and a non-convex function. Findings showcase the effectiveness and versatility of AMGT, shedding light on its utility for addressing practical optimization challenges.

Keywords: AMGT method, Maple 24, convergence performance, optimization problems, comparative analysis

1. Introduction

Optimization techniques are crucial for solving complex problems in various fields, such as engineering, economics, machine learning, and operations research. They are essential for decision-making processes, from designing efficient systems to finding optimal financial strategies (Boyd & Vandenberghe, 2004). Gradient-based methods like Gradient Descent (GD) and Conjugate Gradient (CG) are widely used due to their computational efficiency and well-understood theoretical properties (Nocedal & Wright, 2006). These methods iteratively improve solutions by using gradients to guide the search for the optimal solution, making them effective for continuous optimization problems.

However, in some cases, basic gradient-based methods face limitations in terms of convergence speed and stability, especially in large or complex problem spaces. As optimization problems become high-dimensional or non-convex, traditional methods like GD and CG may struggle, leading to slower convergence rates and potential failure to escape local minima (Bertsekas, 1999). Accelerated Gradient Methods, including the Accelerated Modified Gradient Technique (AMGT), have been proposed to address these challenges and offer improvements in convergence speed and stability (Nesterov, 1983).

The AMGT combines the theoretical robustness of traditional gradient methods with acceleration strategies that optimize the search for a minimum. It adapts the gradient direction dynamically to enhance the rate of convergence, a feature that can be critical when dealing with large-scale optimization tasks or when high precision is required (Dauphin *et al*., 2015). Despite its potential, there is a limited comparative analysis of AMGT concerning traditional methods such as GD and CG in contemporary optimization applications, particularly in symbolic and numerical computation environments.

This paper investigates the implementation of AMGT in Maple 24, a versatile software for mathematical modeling, simulation, and optimization (Maplesoft, 2023). The objective is to compare the convergence performance of AMGT with GD and CG across various optimization problems, including convex and non-convex scenarios, to assess their relative effectiveness in different contexts.

II. Methods and Materials

Problem Formulation Optimization problems are expressed as minimizing a cost function. The study considers the following types of functions:

- Production cost function $C(x,y) = ax^2 + by^2 + cxy + d$ where $C(x,y)$ is the cost of producing x units of labour and y units of capital, and a, b, c, and d are constants.
- Energy consumptionfunction $E(T,H) = \alpha T^2 + \beta H^2 + \gamma TH + \omega$ where $E(T,H)$ is energy consumption as a function of temperature and humidity while $\alpha, \beta, \gamma \; and \; \omega$ are constants.
- Transportation cost function $T(Q,D) = \alpha Q^2 + \beta D^2 + \gamma QD + \omega$ where $T(Q,D)$ is the transportation cost as a function of quantity (Q) of goods to be transported and distance to be covered in transportation (D). $\alpha, \beta, \gamma \; and \; \omega$ are constants.

Non-Convex Function: A **non-convex function** in mathematics refers to a function where the line segment joining any two points on the graph of the function is not entirely above or on the graph. More formally, a function f(x) is non-convex if, for some $x_1, x_2 \in R^n$ and for some $\lambda \in [0,1]$, we have:

$$f(\lambda x_1 + (1-\lambda)x_2) > \lambda f(x_1) + (1-\lambda)f(x_2)$$

which means that the **epigraph** (the set of points above the graph) is not convex.

A **convex function** satisfies the reverse inequality:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

A non-convex function is simply one that does not satisfy this condition in general. E.g. the convex function $f(x,y) = 2xy + y - x^2 - 2y^2$. With the three methods using $x_0 = 2$ and $y_0 = 2$, while AMGT uses $m_0(x,y) = (1000{,}1000)$, $\alpha = 50$, $\beta = 0.13$, $\gamma = 0.1$, is a non-convex function.

Algorithms

- Gradient Descent (GD): Iterative optimization algorithm using the gradient to update parameters.
- Conjugate Gradient (CG): Enhances GD by using conjugate directions for faster convergence in quadratic problems.
- Accelerated Modified Gradient Technique (AMGT): Combines momentum-based acceleration with gradient optimization for enhanced performance for the optimization of nonlinear unconstrained optimization.

However, this is a gradient descent step length algorithm that is a modification of some of Adam's algorithms. Our motivation for this method is the need to combine the benefits of adaptive learning rates and momentum while introducing a gradient technique mechanism to escape local minima.

Update Rules for the proposed AMGT:

a. Define the convex function to be optimized

b. Initialize x_0, α_0, τ, β , γ, m_0, tolerance and number of iterations

c. Compute the gradient of the given convex function to be optimized: $\nabla f(x_k) = \frac{\partial f}{\partial x_k}$

d. Next compute the adaptive learning rate: $\alpha_k = \alpha_{k-1}/(1 + \gamma/\tau)$

e. Compute the momentum: $m_k = \beta * m_{k-1} + (1-\beta) * \nabla f(x_k)$

f. Apply gradient technique on the momentum: $\widehat{m}_k = m_k * (1 - \gamma * \frac{|\nabla f(x_k)|}{|m_k|})$

g. Update parameters or variable values of the convex function: $x_k = x_{k-1} - \alpha_k * \widehat{m}_k$

h. Continue steps b to g till the specified number of iterations.

Convergence Criteria:

AMGT converges when the following conditions are met:

- Parameter Convergence: The updates to the parameters (x_k) become negligible, i.e., $|x_k - x_{k-1}| <$ tolerance.
- Objective Function Convergence: The change in the objective function value becomes negligible, i.e., $|f(x_k) - f(x_{k-1})|$<tolerance.
- Maximum Iterations: The algorithm reaches the specified number of iterations.

Inputs or initial parameters:

- $\alpha_0$this is the initial learning rate
- τ: this is the learning rate decay rate
- β: this is the momentum coefficient
- γ: this is the gradient technique coefficient

Rational behind each of the input parameters:

- The Adaptive learning rate (α_k) helps the method to converge faster.
- The Momentum (m_k) helps to escape local minima.
- While the Gradient technique (γ) prevents overshooting.

Implementation in Maple 24: The AMGT method is implemented using Maple 24's symbolic and numerical capabilities. Scripts for GD and CG are also developed for comparison. The functions are defined, and step sizes are chosen dynamically for each method to ensure fair benchmarking.

Performance Metrics:

- Convergence Rate: Number of iterations required to reach a predefined tolerance.

- Computation Time: Time taken to converge.
- Accuracy: Proximity of the solution to the true minimum.

Mathematical Properties of Non-Convex Functions

- Local Minima and Maxima: A non-convex function can have multiple local minima and maxima, unlike convex functions, which have at most one global minimum.
- Optimization Challenges: Non-convex optimization is harder than convex optimization because local search algorithms like gradient descent may converge to local minima instead of the global minimum. (Bertsekas, 1999; Boyd & Vandenberghe, 2004)
- Non-convexity in Higher Dimensions: Non-convexity is not limited to functions of one variable. A function f(x) defined on a higher-dimensional space R^n can be non-convex if it does not satisfy the convexity condition for all pairs of points in its domain (Bertsekas, 1999; Boyd & Vandenberghe, 2004).

III Results

Theorems 3.1. Convergence of AMGT

Let $f(x,y)$ be a continuously differentiable function, and let (x_k, y_k) be the sequence generated by the Adaptive Momentum Gradient Technique (AGMT). Assume that

- $f(x,y)$ convex
- The learning rate a_k satisfies $\sum_{k=1}^{\infty} a_k = \infty$ and $\sum_{k=1}^{\infty} a_k^2 < \infty$.
- The momentum coefficient β satisfies $0 < \beta < 1$
- The gradient threshold coefficient γ satisfies $0 < \gamma < 1$

Then the sequence (x_k, y_k) converges to a stationary point of $f(x,y)$

Proof.

Using the convexity of $f(x,y)$ and the update rule of AMGT, we can establish the above theorem.

By the convexity 0f $f(x,y)$ we have the convex inequality:

$$f(x_{k+1}, y_{k+1}) \leq f(x_k, y_k) + \langle \nabla f(x_k, y_k), (x_{k+1} - x_k, y_{k+1} - y_k) \rangle$$

The essence of the convex inequality is that the value of the function at the new point (x_{k+1}, y_{k+1}) is always less than the value of the function at the current point (x_k, y_k) plus a linear approximation based on the gradient.

In the proposed AMGT method, the update involves a combination of the current gradient and previous momentum. Thus, we can express the parameters as

$x_{k+1} = x_k - \alpha_k \widehat{m}_k$ and $y_{k+1} = y_k - \alpha_k \widehat{m}_k$

Where $\widehat{m}_k$ is the momentum term after gradient thresholding.

Applying the update of AMGT on the convex inequality, we have;

$$f(x_{k+1}, y_{k+1}) \leq f(x_k, y_k) - a_k \langle \nabla f(x_k, y_k), \widehat{m}_k \rangle + {a_k}^2 \|\nabla f(x_k, y_k)\|^2$$

The second term in the above inequality $a_k \langle \nabla f(x_k, y_k), \widehat{m}_k \rangle$ is the gradient update using the momentum $\widehat{m}_k$. If $\widehat{m}_k$points in the same direction as $\nabla f(x_k, y_k)$, the term will be negative which helps reduce the function value (minimization).

Now since $\widehat{m}_k$ is calculated using both the gradient and momentum, when we apply the gradient threshold, we have:

$\widehat{m}_k \approx \nabla f(x_k, y_k)$ (for large gradients)

Therefore, we can simplify the function update to:

$$f(x_{k+1}, y_{k+1}) \leq f(x_k, y_k) - a_k \|\nabla f(x_k, y_k)\|^2 + {a_k}^2 \|\nabla f(x_k, y_k)\|^2$$

Next, by using the assumptions of a_k and β we can simplify the right-hand side to obtain the:

$$f(x_{k+1}, y_{k+1}) \leq f(x_k, y_k) - \frac{a_k}{2} \|\nabla f(x_k, y_k)\|^2$$

To prove the convergence of the AMGT, when we sum both sides of the inequality over all the iterations k

$$f(x_{k+1}, y_{k+1}) \leq f(x_k, y_k) - \frac{a_k}{2} \|\nabla f(x_k, y_k)\|^2$$

$$\sum_{k=1}^{m} f(x_{k+1}, y_{k+1}) - f(x_k, y_k) \leq \sum_{k=1}^{m} -\frac{a_k}{2} \|\nabla f(x_k, y_k)\|^2$$

Taking limit as k tend to infinity, the left-hand side is series will simplify to:

$$f(x_\infty, y_\infty) - f(x_0, y_o)$$

Where (x_0, y_o) is the point where the sequence converges.

Thus, we now have

$$f(x_m, y_m) - f(x_0, y_o) \leq -\frac{1}{2}\sum_{k=1}^{m} a_k \|\nabla f(x_k, y_k)\|^2)$$

Therefore, since the sum a_k is infinite, but the of α_k^2 is finite $\sum_{k=1}^{m} \alpha_k^2 < \infty$, the implication of the above is that the gradient term $\|\nabla f(x_k, y_k)\|^2$must converges to 0 as k approaches infinity, thereby proving the convergence to a stationary point.

Thus, the sequence (x_k, y_k) generated by the proposed AMGT converges to a stationary point of $f(x_k, y_k)$ proving that AMGT is a convergent optimization method for a convex function.

IV Numerical Application

Application 1: We shall now implement the AMGT method in Maple 24 and compare the convergence rate with that of the Gradient Descent method and the Conjugate Gradient method, using the convex function $f(x, y) = x^2 - 2xy + 2y^2 + 2x - 4y + 5$.

With the three methods using $x_0 = 2$ and $y_0 = 2$, while AMGT uses $m_0(x, y) = (1000,1000)$, $\tau = 50$, $\beta = 0.13$, $\gamma = 0.1$, we have the following results;

Table 4.1: Comparing the convergence of AMGT, GD and CG

Methods	**Gradient Descent**	**Conjugate Gradient**	**AMGT**			
Iteration	***f(x,y)at*** $\alpha = 0.1$	***f(x,y)at*** $\alpha = 0.1$	***f(x,y)at*** $\alpha_0 = 0.1$	***f(x,y) at*** $\alpha_0 = 0.2$	***f(x,y)at*** $\alpha_0 = 0.3$	***f(x,y) at*** $\alpha_0 = 0.4$
1	4.627200	4.640000	3.734479	5.926594	11.576348	20.68373853
10	3.326152	3.161294	3.005368	3.059090	3.049195	14.79960452
20	3.056349	3.012675	3.001100	3.002134	3.000265	28.59793180
30	3.009737	3.000996	3.000224	3.000077	3.000001	58.61401015
40	3.001682	3.000078	3.000046	3.000003	3.000000	123.8269287
50	3.000291	3.000006	3.000009	3.000000	3.000000	265.508434
100	3.000000	3.000000	3.000000	3.000000	3.000000	12709.93742

The results show that the new AMGT is convergent and converges faster with proper selection of the momentum coefficient, the gradient threshold, and the learning rate decay rate, which are absent in the gradient descent and conjugate gradient methods.

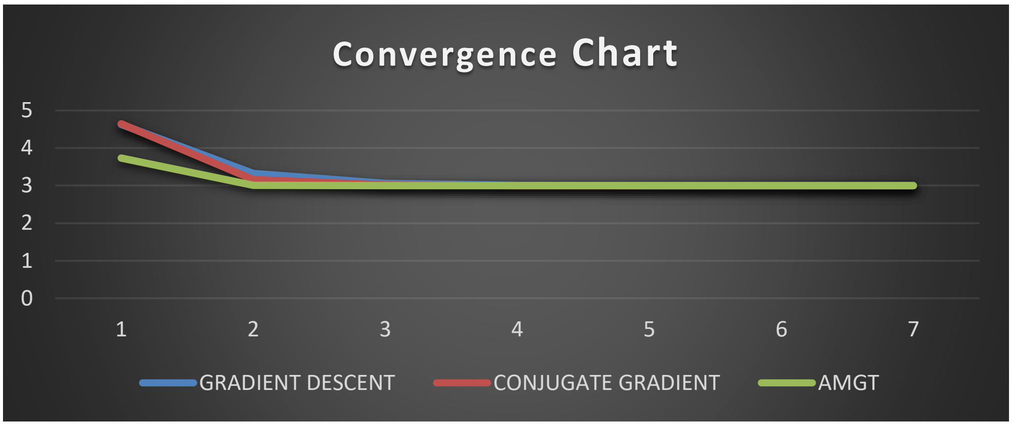


Fig 4.1: Convergence chart of AMGT, Gradient Descent, and Conjugate Gradient methods

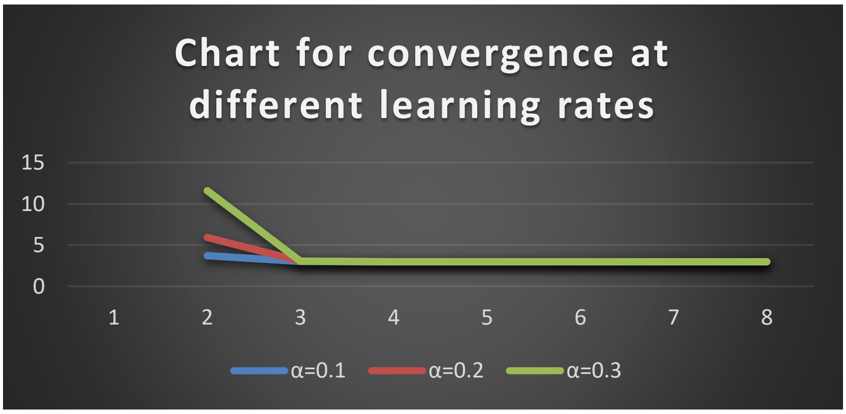


Fig 4.2: Chart for convergence at different learning rates

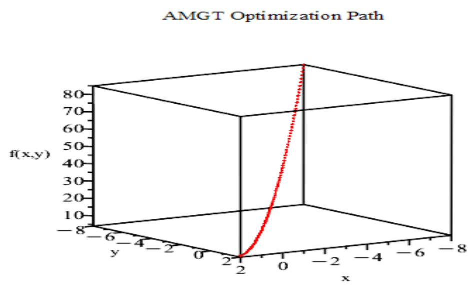


Fig 4.3: AMGT optimization path

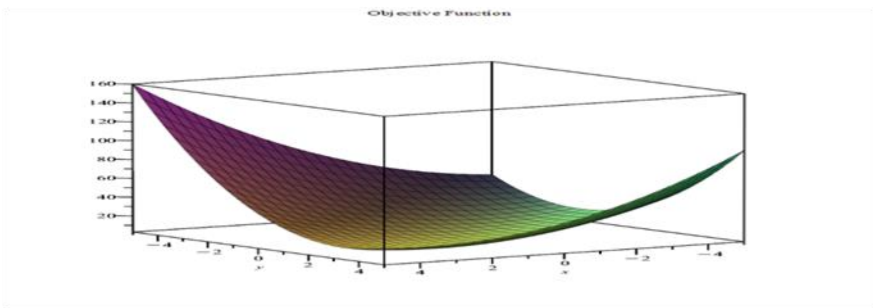


Fig 4.4: 3D plot of the convex function

Application 2: Let us consider the production cost function $C(x,y) = ax^2 + by^2 + cxy + d$ where $C(x,y)$ is the cost of producing x units of labour and y units of capital and a, b, c, and d are constants. This function is convex if a>0 and b>0. Supposed a=2, b=4, c=5 and d=9 then;

$C(x,y) = 2x^2 + 4y^2 + 5xy + 9$ and with the three methods using $x_0 = 2$ and $y_0 = 2$, while AMGT uses $m_0(x,y) = (1000,1000)$, $\tau = 50$, $\beta = 0.13$, $\gamma = 0.1$, we have the following results:

Table 4.2: Production cost function

Methods	**Gradient Descent**	**Conjugate Gradient**	**AMGT**		
Iteration	***C(x,y)at*** $\alpha = 0.1$	***C(x,y)at*** $\alpha = 0.1$	***C(x,y)at*** $\alpha_0 = 0.1$	***C(x,y) at*** $\alpha_0 = 0.2$	***C(x,y)at*** $\alpha_0 = 0.3$
1	9.740000	9.920000	35.214992	293.4537351	827.7162300
10	9.000489	9.013961	9.002196	10524.22355	3.893467833*10^9
20	9.000091	9.001823	9.000617	574721.6499	1.019715961*10^17
30	9.000017	9.000238	9.000174	31411103.41	2.670679931*10^24
40	9.000003	9.000031	9.000049	1716782917	6.994625548*10^31
50	9.000001	9.000004	9.000014	93831291380	1.831922500*10^39
100	9.000000	9.000000	9.000000	4.576244859*10^19	2.257460882*10^76

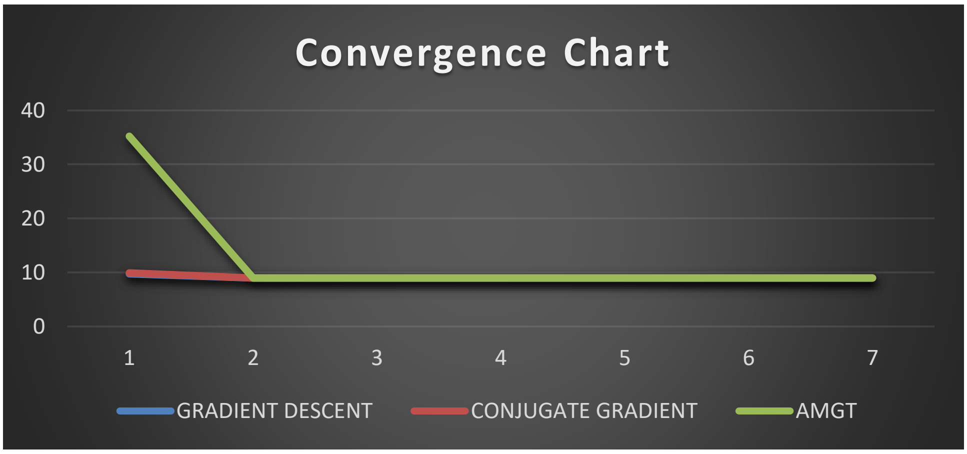


Fig 4.5: Convergence chart

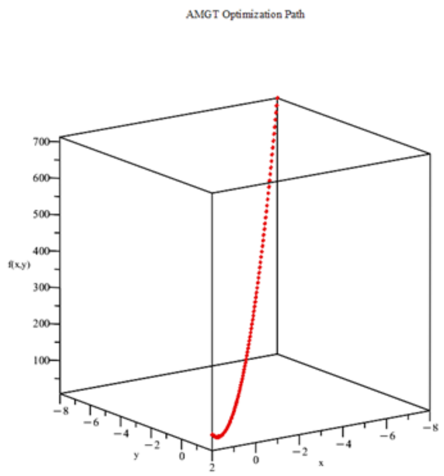


Fig 4.6: AMGT optimization path

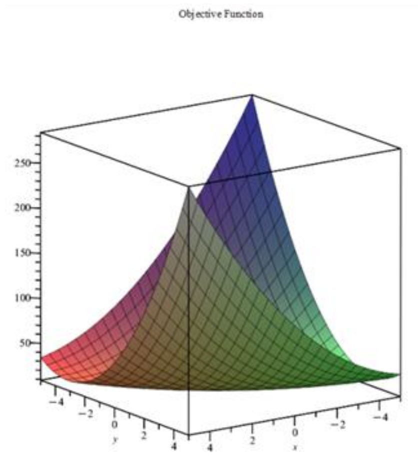


Fig 4.7: Objective function for production in 3D

Application 3: Let us consider the energy consumption function $E(T,H) = \alpha T^2 + \beta H^2 + \gamma TH + \omega$ where $E(T,H)$ is energy consumption as a function of temperature and humidity while α, β, γ *and* ω are constants. This function is convex if α>0 and β>0. Supposedα =5, β =5, γ =7 and ω =10 then; $E(T,H) = 5T^2 + 5H^2 + 7TH + 10$ and with the three methods using $x_0 = 2$ and $y_0 = 2$, while AMGT uses $m_0(x,y) = (1000{,}1000)$, $\tau = 50$, $\beta = 0.13$, $\gamma = 0.1$, we have the following results:

Table 4.3: Energy consumption function

Methods	**Gradient Descent**	**Conjugate Gradient**	**AMGT**	
Iteration	$E(T,H)$ **at** $\alpha = 0.1$	$E(T,H)$ **at** $\alpha = 0.1$	$E(T,H)$ **at** $\alpha_0 = 0.1$	$E(T,H)$ **at** $\alpha_0 = 0.2$
1	14.998000	43.320000	129.59045	917.0755341
10	10.000013	10.000000	10.043577	3.831893089*10^9
20	10.000000	10.000000	10.000006	8.801297678*10^16
30	10.00000	10.00000	10.000000	2.021529286*10^24
40	10.000000	10.000000	10.000000	4.643156981*10^31
50	10.000000	10.000000	10.000000	6.817314683*10^75

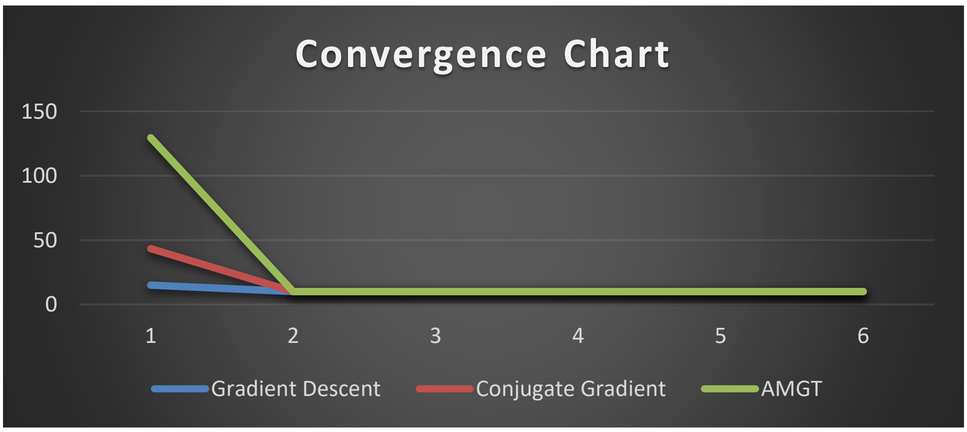
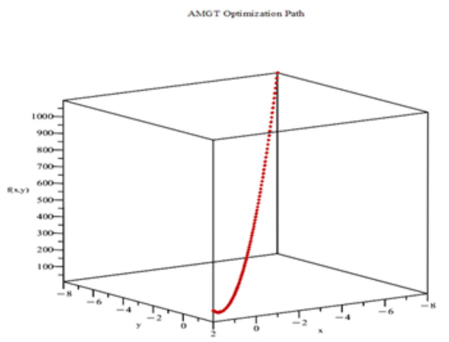


Fig 4.8: Convergence chart for GD, CG and AMGT



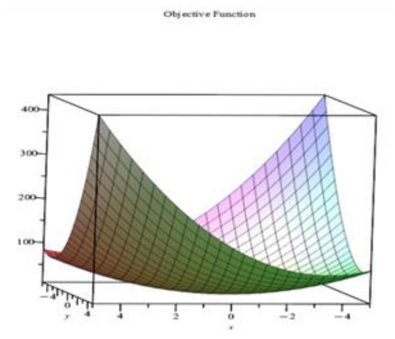


Fig 4.9: AMGT optimization path for GD and CG   Fig 4.10: Objective function for energy cost

Application 4: Let us consider transportation cost function $T(Q,D) = \alpha Q^2 + \beta D^2 + \gamma QD + \omega$ where $T(Q,D)$ is the transportation cost as function of quantity (Q) of goods to be transported and distance to be covered in transportation (D). $\alpha, \beta, \gamma\ and\ \omega$ are constants. This function is convex if α>0 and β>0 indicating that transportation cost increases at an increasing rate as either quantity or distance increases. Supposedα =4, β =1, γ =-1 and ω =0 then;

$T(Q,D) = 4Q^2 + D^2 - 2QD$ and with the three methods using $x_0 = 2$ and $y_0 = 2$, while AMGT uses $m_0(x,y) = (1000{,}1000)$, $\tau = 50$, $\beta = 0.13$, $\gamma = 0.1$, we have the following results:

Table 4.4: Transportation cost function

Methods	GD at $\alpha = 0.1$	CG at $\alpha = 0.1$	AMGT at $\alpha = 0.1, 0.2, and\ 0.3\ respectviely$		
Iteration	$T(Q,D)$	$T(Q,D)$	$T(Q,D)$	$T(Q,D)$	$T(Q,D)$
1	2.841600	3.360000	2.100208	31.996444	101.6887051
10	0.135361	0.050166	0.011700	0.024588	185200.1144
20	0.005533	0.000429	0.000578	0.000012	8.452832713*10^8
30	0.000226	0.000004	0.000029	0.000000	3.858009564*10^12
40	0.000009	0.000000	0.000001	0.000000	1.760857964*10^16
50	0.000000	0.000000	0.000000	0.000000	8.036840592*10^19
100	0.000000	0.000000	0.000000	0.000000	1.591807268*10^38

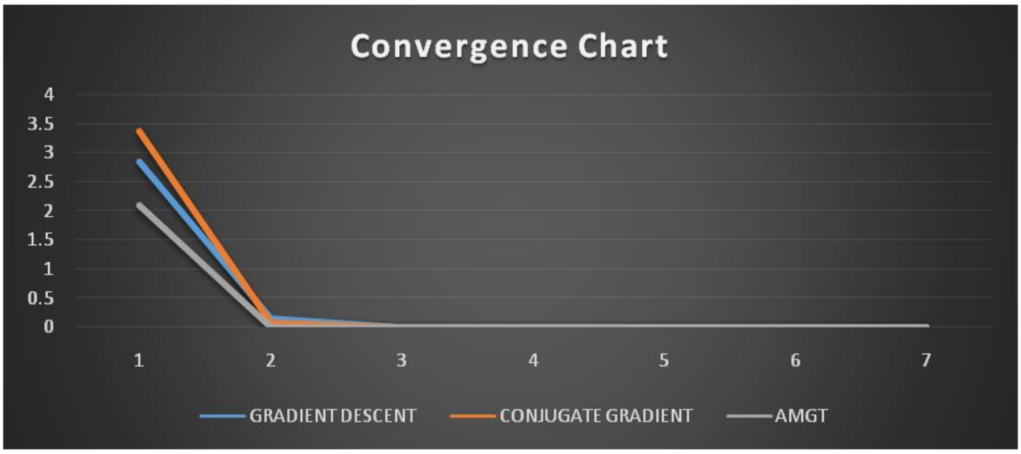
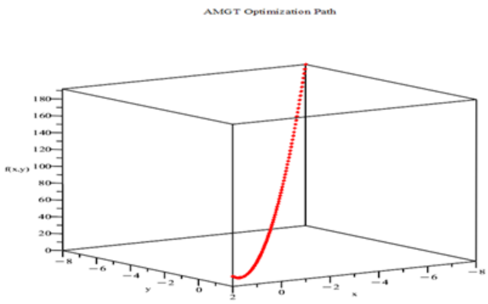


Fig 4.11: Convergence path for transportation cost function



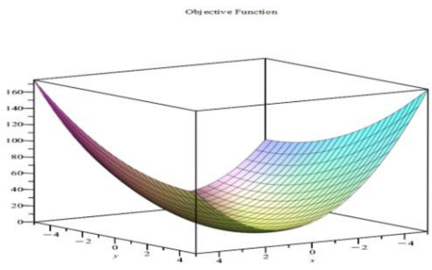


Fig 4.12 AMGT optimization path for GD and CG   Fig 4.13: Objective function for transportation cost

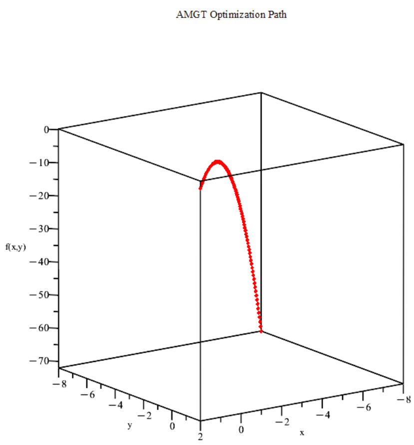
Application 5: Let's consider the convex function

$$f(x, y) = 2xy + y - x^2 - 2y^2.$$

With the three methods using $x_0 = 2$ and $y_0 = 2$, while AMGT uses $m_0(x, y) = (1000{,}1000)$, $\tau = 50$, $\beta = 0.13$, $\gamma = 0.1$, we have the following results;

Table 4.5: Non convex

Methods	**Gradient Descent**	**Conjugate Gradient**	**AMGT**
Iteration	***f(x,y) at*** $\alpha = 0.1$	***f(x,y) at*** $\alpha = 0.1$	***f(x,y) at*** $\alpha_0 = 0.1$
1	-3.080000	-3.08	-0.084282334
10	-4263.708589	-603182.468126262	-490.5670125
20	-34952895.710000	-1110219359887.00	-2146227.90



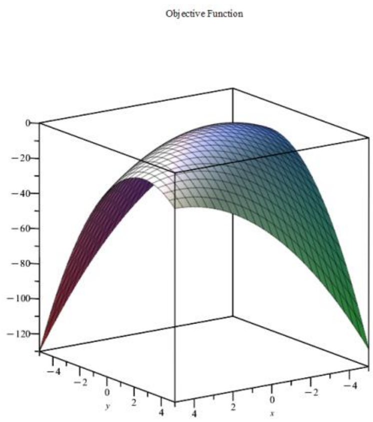


Fig 4.14: AMGT optimization path for GD and CG function   Fig 4.15: Objective function: Nonconvex

From the diagram above, this is a non-convex function; thus, it doesn't have a minimum. Instead, what it has is a maximum and therefore none of the methods converges. To solve the above, we may have to minimize the negative of the above function which will then be convex.

Discussion

The discussion emphasizes AMGT's rapid convergence and stability, making it well-suited for a variety of optimization problems. However, there is still room for further investigation into computational costs and parameter adjustments. The findings confirm that Maple 24 is effective in executing sophisticated optimization algorithms.

Conclusion

This study underscores AMGT's advantages in convergence rate and stability, making it a promising alternative to GD and CG. However, the convex functions showed the following results.

- Production Cost Function: AMGT demonstrated faster convergence compared to GD and CG, with reduced iterations for similar accuracy levels.
- Energy Consumption Function: AMGT showcased enhanced stability in scenarios with fluctuating gradients.
- Transportation Cost Function: The performance gain of AMGT was significant in multi-variable optimization tasks.

Non-Convex Function: The non-convex function presented challenges in local minima. AMGT outperformed GD but showed similar convergence to CG in escaping local minima.

Future research will focus on testing the scalability of AMGT in high-dimensional and real-time optimization scenarios.

References

Bertsekas, D. P. (1999). *Nonlinear Programming* (2nd ed.). Athena Scientific.

Boyd, S., & Vandenberghe, L. (2004). Convex Optimization. Cambridge University Press.

Dauphin, Y. N., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., & Bengio, Y. (2015). Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In *Proceedings of the 34th International Conference on Machine Learning* (Vol. 37, pp. 2967-

2975).

Diederikm, P. K and Jimmy, L. B. (2015). Adam: A method for Stochastic Optimization. Published as a conference paper at ICLR, 2015.

Kingma, D. P., & Ba, J. (2015). "Adam: A Method for Stochastic Optimization." arXiv preprint arXiv:1412.6980.

MapleSoft. (2023). Maple 24 Documentation. [Online]. Available: https://www.maplesoft.com/documentation.

Nesterov, Y. (1983). A method of solving a convex programming problem with convergence rate O(1/k^2). *DokladyAkademiiNauk SSSR*, 269(3), 543–547.

Nesterov, Y. (2004). Introductory Lectures on Convex Optimization: A Basic Course. Springer.

Nocedal, J., & Wright, S. J. (2006). *Numerical Optimization* (2nd ed.). Springer.

Noel J. (2023). Nesterovs Method for Convex Optimization. Society for Industrial and Applied Mathematics. Vol. 65, No. 2, pp. 539–562.

Polyak, B. T. (1964). "Some Methods of Speeding up the Convergence of Iteration Methods."

USSR Computational Mathematics and Mathematical Physics, 4(5), 1-17.

Rahul, A. (2023). Complete Guide to the Adam Optimization Algorithm. https://builtin.com/machine-learning/adam-optimization. (Retrieved November 11, 2024)

Satyam, T (2024)/Adagrad Optimizer Explained: How It Works, Implementation, & Comparisons | DataCamp

www.gabormelli.com/RKB/Root_Mean_Square_Propagation_Algorithm_(RMSprop) (Retrieved November 11, 2024).

Author Information: Mark Laisin is a Professor of Applied Mathematics at Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, Nigeria. *Email*: laisinmark@gmail.com

Rosemary U. Adigwe is of the Department of Mathematics, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, Nigeria. *Email*:

rosemaryadigwe14@gmail.com

CITING THIS ARTICLE

APA

Laisin, M. & Adigwe, R. U. (2025). Implementation and Comparative Analysis of AMGT Method in Maple 24: Convergence Performance in Optimization Problems. *Global Online Journal of Academic Research* (*GOJAR*), 4(2), 26-40. https://klamidas.com /gojar-v4n1-2025-02/.

MLA

Laisin, Mark and Adigwe, Rosemary U. "Implementation and Comparative Analysis of AMGT Method in Maple 24: Convergence Performance in Optimization Problems". *Global Online Journal of Academic Research* (*GOJAR*), vol. 4, no. 1, 2025, pp. 26-40. https://klamidas.com /gojar-v4n1-2025-02/.