A new nonmonotone Newton's modification for unconstrained Optimization

Aristotelis E. Kostopoulos[a] George S. Androulakis[b]
Theodoula N. Grapsa[a]
[a] Department of Mathematics, University of Patras, GR-265.04, Rio, Greece
[b] Department of Business Administration, University of Patras, GR-265.04, Rio, Greece

grapsa@math.upatras.gr

Technical Report 09-05

1 Introduction

In this paper we develop and analyze a variant of Newton's method for the unconstrained optimization problem for a twice continuously differentiable function $f(x)$. Mathematically, this problem is stated as

$$minf(x), \qquad for \quad x \in \mathcal{R}^n, \tag{1}$$

where $f : \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}$. Let $g(x) = \nabla f(x)$ and $H(x) = \nabla^2 f(x)$ be correspondingly the gradient and the Hessian matrix of the function f and let x^* be a minimizer of (1).

In a line search technique, given a current point x^k, the new approximation is determined by the relation

$$x^{k+1} = x^k + t_k d_k, \qquad k = 0, 1, 2, \ldots. \tag{2}$$

where t_k is the steplengh and d_k the current direction. To be ensured finding the minima points of (1), a descent direction must be used at each iteration. But, in some cases this affects to a slow converge. In order to avoid this situation some researchers have proposed a nonmonotone technique where the increment of the function value is allowed from time to time, [4, 13].

In unconstrained optimization it is important an algorithm to have both global convergence and a good convergence rate. Newton's method is one of the most effective methods for solving unconstrained optimization problems due to its attractive quadratic convergence, but it depends directly of the initial point which usually is difficult to be available. In such cases a slow convergence or a divergence may be happen.

Given a current point x^k, as it is well known, essentially in the original Newton's method a linearization of the form $g(x^k) + H(x^k)(x^k - x^{k+1})$ is used at each iteration (see for example [8]). Thus, the new iteration is given by

$$x^{k+1} = x^k - H(x^k)^{-1} g(x^k), \qquad k = 0, 1, 2, \ldots. \tag{3}$$

Recently, noting that the linearization form may probably plays an important role in the iterative process we proposed schemes, modifications of Newton's method (see [5, 6]) for solving nonlinear systems, by constructing proper linear approximations of function components.

Based on this idea, the proposed algorithm in this paper for the unconstrained optimization problem, generally fits in the concept of that given in [6] for nonlinear systems. In particular, a proper approximation $l(x^k)$ of gradient vector $g(x^k)$ is addressed in order to be defined a new direction of the form:

$$d_k = -H(x^k)^{-1} l(x^k), \qquad k = 0, 1, 2, \ldots. \tag{4}$$

This idea is enhanced of the usage of pivot points, [3, 1, 6]. These points are not randomly selected. They are extracted of each gradient component such that to vanish them and essentially replace the current point at each iteration. This selection is hopefully expected to contribute in the effectiveness of the algorithm.

Moreover, it is important that this modification does not destroy the theoretical properties of Newton's algorithm remaining its quadratic convergence and furthermore, it helps Newton's process to get over the difficulty of a slow convergence or divergence in remote points.

This paper is organized as follows. In Section 2, the extraction of pivot points is introduced in brief and continue producing the new AGP algorithm, describing it in detail and proving its quadratic convergence. In Section 3, the proposed approach is used to solve some unconstrained benchmark functions from the usual literature with efficient and promising results. Finally, Section 4 concludes the paper with an outlook on future work.

2 The AGP Proposed Algorithm

The key idea is to approximate the value of gradient components at each iteration by another value which carries more proper information. This approximation produce a new gradient and consequently a new direction.

Suppose that $x^k = (x_1^k, x_2^k, ..., x_n^k)$ is a current approximation of the optimal x^* in an iteration $k = 0, 1, \dots$ and $y^k = (x_1^k, x_2^k, ..., x_{n-1}^k)$.

As mentioned earlier, the pivot points are needed in the development of the new algorithm. Thus, in brief, their extraction and their important *quasi solution* property, to zero the corresponding gradient components are given (a detailed presentation of such points may be found in [1]).

2.1 The extraction of pivot points

A pivot point $x_{pivot}^{k,i}$, corresponding to the gradient component $g_i(x)$ in an iteration k, is a point which lie on its zero contour and simultaneously it is on a line passing through the current point x^k and parallel to the x_n-axis. Thus, they are of the form:

$$x_{pivot}^{k,i} = (x_1^k, x_2^k, ..., x_{n-1}^k, x_n^{k,i}) = (y^k; x_n^{k,i}), \qquad i = 1, 2, \dots \tag{5}$$

and it is obvious that its nth component $x_n^{k,i}$, may be extracted from the one dimensional equation:

$$g_i(x_1^k, x_2^k, ..., x_{n-1}^k, .) = g_i(x_{pivot}^{k,i}) = 0, \qquad i = 1, \dots, n \tag{6}$$

keeping the first $n-1$ components fix [1]. According to Implicit Function Theorem there are unique mappings φ_i such that

$$x_n = \varphi_i(y), \qquad g_i(y; \varphi_i(y)) = 0. \tag{7}$$

Thus, it is valid that $x_n^{k,i} = \varphi_i(y^k)$.

[1]Several methods can be used for solving the equation (6). In this paper, we use a bisection method based on the algebraic signs of the function values [11, 12]. Sequently, our proposed method, which is based in these points, is a sign-gradient component-based method.

2.2 The derivation of the new algorithm

Newton's method is a well-known method for solving the problem (1). In order to produce it, we obtain linear approximations to the given gradients g_i at the current point x^k using the Taylor series expansion of g_i, about this point, neglecting terms of second order and higher. So, we obtain

$$g_i(x) \approx g_i(x^k) + \sum_{j=1}^{n} \partial_j g_i(x^k)(x_j - x_j^k). \tag{8}$$

If we set the right-hand side of the above relation equal to zero, we have

$$g_i(x^k) + \sum_{j=1}^{n} \partial_j g_i(x^k)(x_j - x_j^k) = 0. \tag{9}$$

This linear system ends up at the general iterative form of Newton's method

$$x^{k+1} = x^k - H(x^k)^{-1} g(x^k), \tag{10}$$

with direction

$$p_k = x^{k+1} - x^k = -H(x^k)^{-1} g(x^k). \tag{11}$$

To produce our new iterative method a new direction is introduced. It is similar to Newton's one and results by replacing the value $g(x^k)$ of gradient at the current point in equation (18) by an appropriate approximation $l(x^k)$, at each iteration.

Targeting to this approximation we follow the next procedure. Utilizing Taylor's formula we expand every g_i, $i = 1, \ldots, n$ at the corresponding pivot points $x_{pivot}^{k,i} = (x_1^k, x_2^k, ..., x_{n-1}^k, x_n^{k,i}) = (y^k; x_n^{k,i}), i = 1, 2, \ldots$ about the current point x^k. Thus,

$$g_i(x_{pivot}^{k,i}) \approx g_i(x^k) + \sum_{j=1}^{n-1} \partial_j g_i(x^k)(y_j^k - x_j^k) + \partial_n g_i(x^k)(\varphi_i(y^k) - x_n^k). \tag{12}$$

Due to the definition of pivot points (5) and their property (6) to zero the corresponding gradient components, the above relation (12) becomes

$$0 \approx g_i(x^k) + \sum_{j=1}^{n-1} \partial_j g_i(x^k) 0 + \partial_n g_i(x^k)(\varphi_i(y^k) - x_n^k). \tag{13}$$

Hence,

$$g_i(x^k) \approx l_i(x^k) = \partial_n g_i(x^k)(x_n^k - \varphi_i(y^k)). \tag{14}$$

The relation (14) is the key point in which our new approach is based. It plays a notable, determinative role in the development of our new direction and influences on its characteristics such that a generally good behavior to be achieved by using the new method.

We continue by replacing $l_i(x^k)$, given in (14), in the relation (8) in order to transform Newton's method to the new one. Indeed, substituting the relation (14) in (8) we get the new linearization of gradient components, $h_i(x)$, from which the new method will be arisen. So,

$$\begin{aligned} g_i(x) \quad &\approx \quad g_i(x^k) + \sum_{j=1}^{n} \partial_j g_i(x^k)(x_j - x_j^k) \approx h_i(x) = l_i(x^k) + \sum_{j=1}^{n} \partial_j g_i(x^k)(x_j - x_j^k) = \\ &= \partial_n g_i(x^k)(x_n^k - \varphi_i(y^k)) + \sum_{j=1}^{n} \partial_j g_i(x^k)(x_j - x_j^k). \end{aligned} \tag{15}$$

If we set the right-hand side of the above relation equal to zero, we have

$$\partial_n g_i(x^k)(x_n^k - \varphi_i(y^k)) + \sum_{j=1}^{n} \partial_j g_i(x^k)(x_j - x_j^k) = 0. \tag{16}$$

Just as in Newton's method the above linear system ends up at the general iterative form of the new modified Newton's method, given by the form:

$$x^{k+1} = x^k - H(x^k)^{-1} l(x^k), \tag{17}$$

with direction

$$d_k = x^{k+1} - x^k = -H(x^k)^{-1} l(x^k). \tag{18}$$

taking $l(x^k)$ as the above defined approximation of $g(x^k)$:

$$l(x^k) \approx G(x^k) = \nabla f(x^k) = \left(\partial_n g_1(x^k)(x_n^k - \varphi_1(y^k)), \ldots, \partial_n g_i(x^k)(x_n^k - \varphi_n(y^k))\right). \tag{19}$$

This proposed method is named *AGP method (Approximated Gradient using Pivot points method).*

Of course, relative procedures for obtaining x^* can be constructed by replacing x_n with any one of the components $x_1, \ldots, x_{n-1}$, for example x_i, and taking $y = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

2.3 The convergence of the new method

The convergence and the order of convergence of the proposed method are proved in the following Convergence Theorem.

Theorem 2.1. *Suppose that the objective function $f = (f_1, \ldots, f_n) : D \subset R^n \to R$ is twice continuously differentiable in an open neighborhood $D^* \subset D$ of a point $x^* = (x_1^*, \ldots, x_n^*) \in \mathcal{D}^*$ for which $g(x^*) = 0$. Moreover, let $B_i, i = 1, 2, \ldots, n$, are those connected components of $g_i^{-1}(0)$, containing x^* on which $\partial_n g_i \neq 0$ for $i = 1, 2, \ldots$, respectively. Then the iterations x^k of the new method, given by Equation (17) will converge to x^* provided $H(x^*)$ is nonsingular and the initial guess x^0 is sufficiently close to x^*. Moreover, the order of convergence of the new method, will be two.*

Proof 2.1. *Let the matrix,*

$$V(x^k) = \begin{pmatrix} \partial_n g_1(x^k) & 0 & \cdots & 0 \\ 0 & \partial_n g_2(x^k) & \cdots & 0 \\ \vdots & \ddots & \cdots & 0 \\ 0 & 0 & \cdots & \partial_n g_n(x^k) \end{pmatrix} \tag{20}$$

¿From the assumptions of the theorem it is valid that the partial derivatives $\partial_n g_i(x^k) \neq 0$. Thus, the diagonal matrix $V(x^k)$ with diagonal elements $\partial_n g_i(x^k)$ has determinant non zero. So, it is invertible with

$$V(x^k)^{-1} = \begin{pmatrix} 1/\partial_n g_1(x^k) & 0 & \cdots & 0 \\ 0 & 1/\partial_n g_2(x^k) & \cdots & 0 \\ \vdots & \ddots & \cdots & 0 \\ 0 & 0 & \cdots & 1/\partial_n g_n(x^k) \end{pmatrix} \tag{21}$$

Now, we consider the mapping $w = (w_1, w_2, \ldots, w_n)^T : D \subset R^n \to R^n$, where

$$w_i(x) = x_n - \varphi_i(y) \tag{22}$$

and $\varphi_i(y)$ as they are given in Equation (7).

Utilizing the Implicit Function Theorem the elements $\partial_j w_i(x)$ of the matrix $w'(x)$ obviously are given by

$$\partial_j w_i(x) = \begin{cases} \partial_j g_i(x)/\partial_j g_i(x) & \texttt{if } \; j = 1, 2, \ldots, n-1 \\ 1 & \texttt{if } j = n \end{cases}$$

Hence, using the matrix $V(x)^{-1}$ given by 21 *we have:*

$$w'(x) = \begin{pmatrix} \frac{\partial_1 g_1(x)}{\partial_n g_1(x)} & \frac{\partial_2 g_1(x)}{\partial_n g_1(x)} & \cdots & \frac{\partial_{n-1} g_1(x)}{\partial_n g_1(x)} & 1 \\ \frac{\partial_1 g_2(x)}{\partial_n g_2(x)} & \frac{\partial_2 g_2(x)}{\partial_n g_2(x)} & \cdots & \frac{\partial_{n-1} g_2(x)}{\partial_n g_2(x)} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial_1 g_n(x)}{\partial_n g_n(x)} & \frac{\partial_2 g_n(x)}{\partial_n g_n(x)} & \cdots & \frac{\partial_{n-1} g_n(x)}{\partial_n g_n(x)} & 1 \end{pmatrix} = V(x)^{-1} H(x). \tag{23}$$

Furthermore, taking into account the definition of $l(x^k)$, $w(x^k)$ and the relation $w'(x^k) = V(x^k)^{-1} H(x^k)$ we result that

$$l(x^k) = V(x^k) w(x^k) \tag{24}$$

and the inverse of Hessian matrix, $H(x^k)^{-1}$, may be rewritten as:

$$H(x^k)^{-1} = w'(x^k)^{-1} V(x^k)^{-1} \tag{25}$$

Thus, using the relations (24) and (25), our iterative scheme given by (17) may be written in the form

$$x^{k+1} = x^k - w'(x^k)^{-1} V(x^k)^{-1} V(x^k) w(x^k), \tag{26}$$

and hence we take the equivalent form:

$$x^{k+1} = x^k - w'(x^k)^{-1} w(x^k). \tag{27}$$

Notice that, by the assumption of the theorem it is valid that $H(x^*)$ is nonsingular and by the form of matrix $V(x)$, which is given in (20), it is also valid that $detV(x^*)^{-1} \neq 0$. So, due to the relation (23) it is obvious that $detw'(x^*) \neq 0$ and thus the matrix $w'(x^*)$ is also nonsingular.

Consequently, by the Newton's convergence theorem (see [7, 9, 10]), applied on $w(x)$, for an initial guess x^0 sufficiently close to x^*, the iterations $x^k, k = 0, 1, 2, \ldots,$ of the new iterative scheme (27) (and hence the equivalent one given by (17)), converges to x^* and the order of convergence is two.

Notation 2.1. *One may notice that if x^* is an optimal point, for the gradient components it is valid that $g_i(x^*) = 0$. In addition, from the definition of the pivot points and relation (6) it is valid that $x_n^* = \varphi_i(y^*)$ and thus $x_n^* - \varphi_i(y^*) = 0$. Hence, by the definition of $w_i(x)$ it results that $w_i(x^*) = 0$ in any case and thus $w(x^*) = 0$.*

Notation 2.2. *Considering the form of $l(x)$ which is used in the iterative form of the proposed AGP method, given by (17), it is evident that the proposed modified Newton approach is directly independent of gradient values.*

Notation 2.3. *The new direction may also be posed as:*

$$d_k = -w'(x^k)^{-1} w(x^k) \tag{28}$$

which is hoping to enforce the standard Newton's method and to improve its behavior in difficult cases, such as in cases of faraway from the solution initial points.

A sketch of the proposed algorithm is given by Algorithm 2.4.

2.4 The AGP Algorithm

step 1:x^0 the initial point.
step 2:for $k = 0, 1, \ldots$ until converge
step 3:*The extraction of nth component of pivot points*
for $i = 1, \ldots, n$, solve the one-dimensional equations (6) holding the variables $y^k = (x_1^k, \ldots, x_{n-1}^k)$, fixed to extract the nth components $x_n^{k,i}, i = 1, \ldots, n$ of pivot points (5)
step 4: *The new direction*
Evaluate at each iteration the new direction according to the form $d_k = -H(x^k)^{-1} l(x^k)$, given by the relation (17). For this computation the below steps 4a and 4b are necessary.
step 4a: *The approximation of gradient components*
Approximate at each iteration the gradient complements at the current point x^k by using the relation (19).
step 4b: *The approximation of gradient components*
Evaluate the Hessian matrix at the current point x^k getting the

$H^k = [H^k_{ij}]$, where $H^k_{ij} = \partial_j \; g_i(x^k), \; \jmath = 1, \ldots, n$ for each $g_i, \quad i = 1, \ldots, n$.
step 5: *Calculation of new x*
Calculate $x^{k+1} = x^k + d_k = x^k - H(x^k)^{-1} l(x^k)$
step 6: go to step 2

3 The crucial points of the AGP approach

The AGP method may be presented in some different forms which however are equivalent. Studying these types separately some crucial results may be considered.

3.1 An optimization algorithm directly free of the objective function and of its gradient

Relation (17) due to the form of $l_i(x^K)$ as given by Relation (19) is obviously directly free of the values of the objective function and the values of gradient compontents.

3.2 An approximation of objective function via an approximation of its gradient

The AGP method is related to an approximation of the objective function, which their type there is no need to be known. The proposed approach is like working on this approximated objective function via an approximation of its gradient. In particular, the approximation $h(x)$ of the gradient $g(x)$ corresponds to an objective function. Evidently, this objective function consists an approximation of the original one, $f(x)$, and in our case there is no need be known. In other words, we work on an approximation of the objective function, without care about its form, working via an approximation of its gradient.

3.3 The AGP method and the distance between the current point and the corresponding zero contour

Taking the form of new proposed method as it is given in (27), which is equivalent to (17), we may firstly observe that the AGP algorithm is a variation of Newton's method. The main difference is the approximation of gradient $g(x)$ by $w(x)$ which uses information of pivot points $x^{k,i}_{pivot}$, and the Hessian matrix $H(x)$ is approximated by the new one, $w'(x)$. Thus, the problem of solving the system $g_i(x) = 0$ is transformed to an approximated one, given by $w_i(x) = 0$. Furthermore, at this point, it is important to notice that essentially the new posed gradient components w_i illustrate the distance between the current point x^k and the corresponding zero contour, in a selected direction. It is clear that for these new functions of gradient components, as n increases this distance tends to zero for all $i = 1, 2, \ldots, n$ and the optimal point is achieved. An exhaustive study for the selection of the best direction (component of pivot point) must be done.

3.4 The AGP as rotation of tangent planes to succeed better approximation of optimal point

Studying the AGP method through the new linear approximation $h(x)$ of $g(x)$, it is faced as a rotation of tangent planes such that their intersection point of approximations of gradient components, result to a better approximation of optimal point. Mainly,this process removes quickly from a bad current point or remains close to a good one. This point of view corresponds to an enlargement of basin of convergence area.

3.5 Increasing or decreasing the value of gradient components to remove from the current point or to move closer to it

Studying the AGP method through the form $x^{k+1} = x^k - H(x^k)^{-1}l(x^k)$, given by the equation (17), with $l(x^k)$ to be the approximation of $g(x^k)$, as it is given in (19) we result to another crucial point. This case is faced via a geometrical interpretation similar to Figure 1. in [6].

The value of $l_i(x^k)$ in comparison with $g_i(x^k)$ may be:

- increases, in cases the current point is a no good iteration, such that a removement to be done from this current point.
- decreases, in comparison to Newton's one, in cases the current point is a good one and there is no need to remove from the current point.

The movement far away from the current point corresponds to a quickly movement of a no good initial point and thus a contribution in the initialization dependence problem is done through this approach. This behavior is described as a "feeling" of the method to remove or to draw approach, contributing in the effectiveness of the algorithm and in its global convergence.

3.6 The hidden second order information in the form of AGP scheme

The most important information of the AGP method is the hidden second order information which is nested in its form, without anymore cost about it. This information is inherent in the form of the method and takes into account the morphology of gradient components in the neighborhood of the current point, without any second order information. This observation is very useful and must be under consideration in a future study.

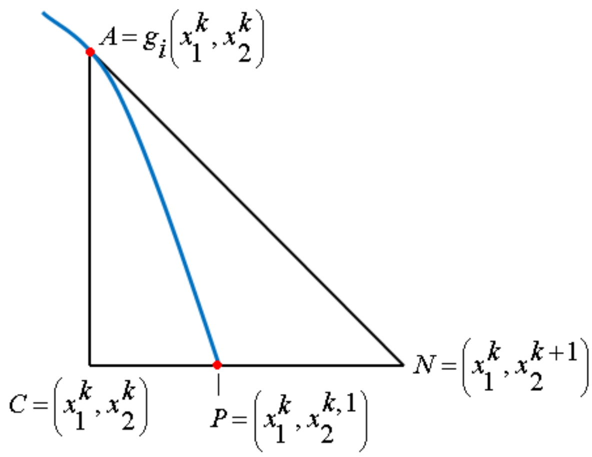

Figure 1: Zero contour close to the current point

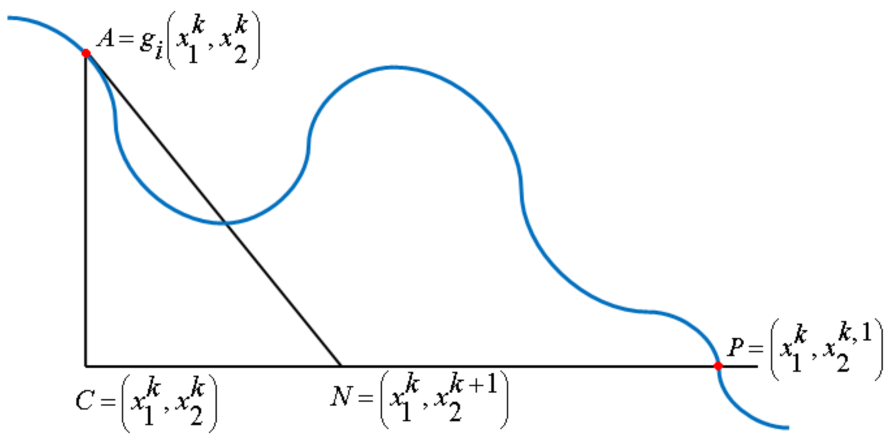

Figure 2: Zero contour far away from to the current point

In Figure 1, although the value of gradient component at the current point is large, its small distance from the zero contour shows that the curvature of the function is sharp and thus it may be close to the optimal point.

The value of the gradient component in combination with its distance of the zero contour show the information which in other cases is extracted only by using second order information to take into account the morphology of the function.

In Figure 2, the same big value of gradient component at this current point in combination to the remote pivot point show that a removement of this current point must be done. Moreover, this current point may be characterized as a 'bad' point for this component at this direction.

Newton's method, in both of the above cases, will give the same approximation point, because we have take to be the same value of the function and the same partial derivative. On the contrary, the AGP method in the first case will remain in the neighborhood of current point, while in the second case it will be moved far away. This is achieved automatically because it takes into account the curvature of the function, specifically the value of the function in combination to the distance between the current point and the zero contour in a direction.

3.7 How we can characterize an initial or current point as a "good" or as a "bad" one?

According to the above study we may, generally, characterize an initial point as a "good" or as a "bad" one, taking into account its distance of the zero contour of the corresponding gradient component, in a selected direction, and simultaneously, the value of partial derivative at this point. For convenience, one may see and the figure 1 in ([6]). In particular:

1. if $\left|l_i(x^k)\right| < \left|g_i(x^k)\right|$ then the point is a good one. It is valid because:

$$\left|l_i(x^k)\right| = \left|\partial_n g_i(x^k)\right| \left|x_n^k - \varphi_i(y^k)\right|, \qquad \left|g_i(x^k)\right| = \left|\partial_n g_i(x^k)\right| \left|x_n^k - x_n^{k+1}\right|. \quad (29)$$

   Obviously, if $\left|l_i(x^k)\right| < \left|g_i(x^k)\right|$ then it is evident that $\left|x_n^k - x_n^{k,i}\right| < \left|x_n^k - x_n^{k+1}\right|$. Thus, the $min(x_n^k, x_n^{k+1}) < x_n^{k,i} < min(x_n^k, x_n^{k+1})$. This means that the pivot point is closer to the current point than the point extracted by Newton's method. So, the current point may generally be considered as a 'good' one.

2. Similarly results are obtained if $\left|l_i(x^k)\right| > \left|g_i(x^k)\right|$. In fact, if $\left|l_i(x^k)\right| > \left|g_i(x^k)\right|$, due to the relation(29), it is evident that $\left|x_n^k - x_n^{k,i}\right| > \left|x_n^k - x_n^{k+1}\right|$. Thus, the corresponding pivot point is farther from the current point than the Newton's one. So, a removent far from it must be done and the current point may be characterized as a bad one.

   For a future work we may take into advance this notation, to estimate the behavior of the method in a point and to decide how to continue the progress.

   In Figure 3 we may see all the cases of value of $l_i(x^k)$ and $g_i(x^k)$ and moreover the possible pivot points P and the Newton's ones N

4 Numerical Results

The proposed AGP method, for unconstrained optimization has been applied to problems of various dimensions. The implementation has been done by using FORTRAN and tested on a PC IBM. The notation for the following tables is given below:

n	dimension,
x^0	starting point,
x^*	approximated local optimum,
IT	the total number of iterations required to obtain x^*,
D	indicates divergence or nonconvergence at the selected number of iterations.

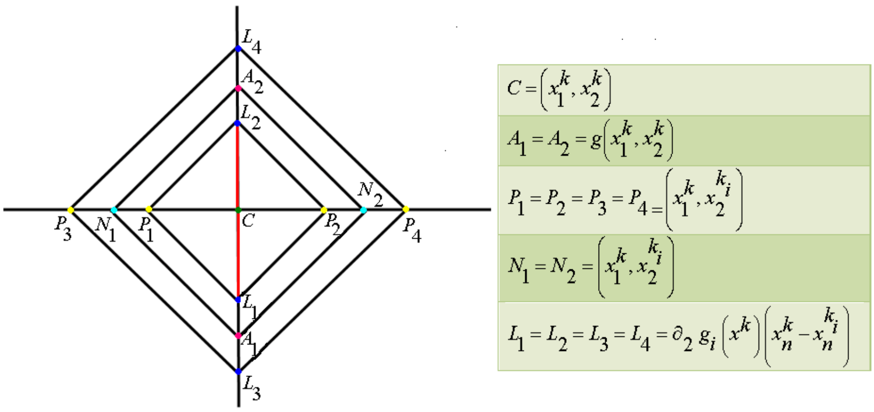

Figure 3: "Good" and "bad" cases.

The used convergence criteria are the following:

1. $\left\|\nabla g(x^{k+1})\right\|_1 \leq 10^{-10}$,
2. $\left\|x^{k+1} - x^k\right\|_1 \leq 10^{-10}$,
3. $MIT = 1000$, where MIT is the maximum number of iterations.

In this paper we give the results of two known problems: Fredenstein-Roth and almost Brown linear In Tables 1,2 are presented the results obtained by Armijo, BFGS Newton's method and the new one AGP.

Example 4.1. *(Freudenstein and Roth function). In this example the objective function f is given by :*

$$f(x) = \sum_{i=1}^{2} {f_i}^2(x), \tag{30}$$

where :

$$f_1(x) = -13 + x_1 + ((5 - x_2)\, x_2 - 2)\, x_2, \quad f_2(x) = -29 + x_1 + ((x_2 + 1)\, x_2 - 14)\, x_2, \tag{31}$$

with $f(x^*) = 0$ at $x^* = (5, 4)$ and $f(x^*) = 48.9842\ldots$ at $x^* = (11.41\ldots, -0.8968\ldots)$. As our starting values we utilized $x^0 = (0.5, -2)$ and obtained $x^* = (11.41\ldots, -0.8968\ldots)$ after IT $= 6$.

Example 4.2. *(Brown almost-linear function). In this case the objective function f is given by :*

$$f(x) = \sum_{i=1}^{n} {f_i}^2(x), \tag{32}$$

Table 1 Freudenstein and Roth function

	Armijo		FR		PR		BFGS		Newton		AGP		
x^0	IT	FE	IT	FE	IT	FE	IT	FE	IT	FE	IT	FE	ASG
$(0.5, -2)^\alpha$	1827	24155	18	356	8	187	7	138	7	42	6	24	120
$(0.5, 1000)$	1380	18770	D	D	D	D	D	D	32	192	12	48	240
$(10, 100)$	1845	24664	10	200	9	194	9	170	21	126	9	36	180
$(4.5, -8)$	D	D	D	D	15	75	14	84	13	78	5	20	100
$(10, -20)$	D	D	21	105	D	D	18	108	17	102	5	20	100
$(12, -24)$	38	190	D	D	D	D	18	108	18	108	5	20	100

Table 2. Brown almost-linear function, n=3

	Armijo		FR		PR		BFGS		Newton		AGP		
x^0	IT	FE	IT	FE	IT	FE	IT	FE	IT	FE	IT	FE	ASG
$(0.5, 0.5, 0.5)^\alpha$	177	1330	12	177	6	93	6	91	6	72	7	63	210
$(0, 0, 3)$	221	1612	27	389	9	143	9	131	1	12	1	9	10
$(.1, .1, -2)$	181	1349	52	742	9	149	9	145	9	108	3	27	90
$(0, 0, 17)$	13	117	22	198	34	306	D	D	1	12	1	9	30
$(-1, 0, -1)$	10	90	17	153	D	D	9	108	13	156	2	18	60
$(0.8, 0.7, -1.7)$	7	63	33	297	13	117	15	180	27	324	6	54	180

where :

$$f_i(x) = x_i + \sum_{j=1}^{n} x_j - (n+1), \quad 1 \le i < n, \quad f_n(x) = \left(\prod_{j=1}^{n} x_j \right) - 1, \tag{33}$$

with $f(x^*) = 0$ at $x^* = (\alpha, \dots, \alpha, \alpha^{1-n})$ where α satisfies the equation $n\alpha^n - (n+1)\alpha^{n-1} + 1 = 0$ and $f(x^*) = 1$ at $x^* = (0, \dots, 0, n+1)$.

This example has been tested for $n = 3$.

5 Conclusion and Further Research

In this paper a new Newton algorithm for the problem of unconstrained optimization was introduced. Its main goal is the improvement of some disadvantages of Newton's method, without loosing its quadratic convergence.

A new direction, based on a proper approximation of gradient, utilizing the pivot points instead of the current one, is determined. The key point for each gradient component approximation is its dependence from the distance between the current point and the pivot point in combination with the partial derivative at the current point.

The complication of current point in combination with its distance from the zero contour (pivot point) and its partial derivative plays a deterministic role such that the morphology of gradient component at this region to be taken into account, without any need of second order information. This constitutes the key of success in cases of a remote initial point.

The proposed method is presented in some alternative but equivalent forms where any one of them face the method from another point of view. From this perspective crucial points are arisen which although they are presented and explained. At least some of them have to be considered for a more deeper study in a future work.

Bibliography

[1] T.N. Grapsa, *Implementing the initialization-dependence and the singularity difficulties in newton's method*, Tech. rep., Division of Computational Mathematics and Informatics, Department of Mathematics, University of Patras, 2007.

[2] T.N. Grapsa, *A Quadratic Convergence Quasi-Newton Method for Unconstrained Optimization. In: Proceedings of Conference in Optimization (Optimization2007), Porto, Portugal, 2007.*

[3] T.N. Grapsa, M.N. Vrahatis, *A dimension-reducing method for unconstrained optimization, J. Comput. Appl. Math. 66, pp. 239-253, 1996*

[4] L. Grippo, F. Lampariello, S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal., 23, pp.707–716, 1986.

[5] T.N. Grapsa, E.N. Malihoutsaki, Newton's method without direct function evaluations, in: E. Lipitakis (ed.), Proceedings of 8th Hellenic European Research on Computer Mathematics & its Applications-Conference (HERCMA 2007), 2007.

[6] E.N. Malihoutsaki, I.A. Nikas, and T.N. Grapsa, *Improved newton's method without direct function evaluations*, J. Comput. Appl. Math. 227 (2009), pp. 206–212.

[7] J.M. Ortega, Numerical Analysis, Academic Press, New York, 1972.

[8] J.M. Ortega, W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.

[9] A. Ostrowski, Solution of Equations in Euclidean and Banach Spaces, 3rd ed., Academic Press, London, 1973.

[10] J.F. Traub, Iterative Methods for the Solution of Equations, Prentice-Hall. Inc., Englewood Cliffs, NJ, 1964.

[11] M.N. Vrahatis, CHABIS: A mathematical software package for locating and evaluating roots of systems of nonlinear equations, ACM Trans. Math. Software 14 (4) (1988) 330–336.

[12] M.N. Vrahatis, K.I. Iordanidis, A rapid generalized method of bisection for solving systems of non–linear equations, Numer. Math. 49 (2-3) (1986) 123–138.

[13] H. Zhang, W.W. Hager, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. Optim, 14(4), pp. 1043–1056, 2004.