DIMENSION REDUCING METHODS FOR SYSTEMS OF NONLINEAR EQUATIONS AND UNCONSTRAINED OPTIMIZATION: A REVIEW

T.N. Grapsa* **and M.N. Vrahatis****
Department of Mathematics, University of Patras, GR–26110 Patras, Greece
* e–mail: grapsa@math.upatras.gr
**e–mail: vrahatis@math.upatras.gr

Abstract. *The purpose of this report is to review a new class of methods we have proposed for solving systems of nonlinear equations and optimization problems, named Dimension Reducing Methods. These methods are based on reduction to simpler one-dimensional nonlinear equations. Although these methods use reduction to simpler one, they converge quadratically, and incorporate the advantages of nonlinear SOR and Newton's algorithms. Moreover, since they do not directly perform function evaluations, they can be applied to problems with imprecise function values.*

Keywords: Dimension–reducing method, systems of nonlinear equations, unconstrained optimization, reduction to one–dimensional equations, imprecise function values, quadratic convergence.

1 INTRODUCTION

Suppose that $F = (f_1, \ldots, f_n) : \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}^n$ is a continuously differentiable mapping on an open neighborhood $\mathcal{D}^* \subset \mathcal{D}$ of a solution $x^* \in \mathcal{D}$ of the system of nonlinear equations

$$F(x) = \Theta^n = (0, 0, \ldots, 0). \tag{1}$$

There is a class of methods for the numerical solution of the above system which arise from iterative procedures used for systems of linear equations[15, 17]. These methods use reduction to simpler one–dimensional nonlinear equations for the components $f_1, f_2, \ldots, f_n$ of F. The best–known method of this type is the *nonlinear successive overrelaxation (SOR)* method which solves the one–dimensional equation

$$f_i(x_1^{p+1}, \ldots, x_{i-1}^{p+1}, x_i, x_{i+1}^{p}, \ldots, x_n^{p}) = 0, \tag{2}$$

for x_i and then sets

$$x_i^{p+1} = x_i^p + \omega(x_i - x_i^p), \quad i = 1, \ldots, n, \quad p = 0, 1, \ldots \quad , \tag{3}$$

provided that $\omega \in (0, 1]$. The above process is independent of the value of ω and is called *SOR* process even though this nomenclature is sometimes reserved for the case $\omega > 1$. Now, a large variety of combined methods can be constructed depending on the secondary iteration and the number of steps required for solving (2). Thus, for example, one can obtain the *exact nonlinear SOR* or *m–step SOR–Newton process*[15, 17] and so on. If the Jacobian of F at the solution x^* of (1) is an $\mathcal{M}$–matrix[15] the iterates of the above processes will converge linearly to x^* provided that $\omega \in (0, 1]$[15]. Another well–known method is the Newton's method which, starting with an initial guess x^0 for the attainment of an approximation of the solution x^* of (1), is given by

$$x^{p+1} = x^p - F'(x^p)^{-1} F(x^p), \quad p = 0, 1, \ldots \tag{4}$$

If the Jacobian $F'(x^*)$ is nonsingular and $F'(x)$ is Lipschitz continuous then the iterates (4) converge quadratically to x^* provided the initial guess x^0 is sufficiently close to x^*. The quadratic convergence of Newton's method is attractive. However, the method depends on a good initial approximation[5] and requires, in general, $n^2 + n$ function evaluations per iteration, besides the solution of an $n \times n$ linear system. Moreover, the behavior of Newton's method is problematic when $F'(x^*)$ is singular, since, in that case, it does not converge quadratically and, in general, is not appropriate for approximations of x^* with a high accuracy. For this reason there are procedures[25, 26] which, under some assumptions (such as rank $F'(x^*) = n - 1$), can attain a highly accurate solution x^* by enlarging the system (1) to one which is at least $(2n + 1)$–dimensional[25, 26]. Also, Newton's method remains problematic when the

values of F can not be accurately achieved. Of course, this problem is common to all iterative procedures which directly depend on function evaluations. To overcome it, one may resort to generalized bisection methods[4, 11, 12, 20, 21, 22, 23] since they only make use of the algebraic sign of the function involved in the equations. However, these methods do not generally attain a quadratic convergence.
Recently, we have proposed methods[6, 7, 8, 9] for the numerical solution of a system of nonlinear algebraic and/or transcendental equations (1). These methods incorporate the advantages of Newton and nonlinear SOR algorithms[15]. Specifically, although these methods use reduction to simpler one–dimensional nonlinear equations, they converge quadratically.
At first, we have proposed the iterative procedure, DR-method[7], for the solution of a system of nonlinear algebraic and/or transcendental equations in $\mathcal{R}^n$, which generates a quadratically converging sequence of points in $\mathcal{R}^{n-1}$ which converges to the $n-1$ components of the solution while the remaining component of the solution is evaluated separately using the final approximations of the others. For this component an initial guess is not necessary and it is at the user's disposal to choose which will be the remaining component, according to the problem. Also this method does not directly need any function evaluation and it compares favorably with quadratically convergent methods. Moreover, it compares favorably with Newton's method when the Jacobian at the solution is singular (without making any enlargement of the system), or when it is difficult to evaluate the function values accurately.
Also, we have proposed a modification of the DR method which maintains the advantages of DR method. Specifically, we have proposed the method of Rotating Hyperplanes (RM)[8], which uses a "rotating" hyperplane in $\mathcal{R}^{n+1}$, whose rotation axis depends on the current approximation of $n-1$ components of the solution. This procedure has been applied on the traditional Newton's algorithm as well as on the DR method, whence two modified schemes have been obtained, the Modified Newton Method (MNM) and the Modified Dimension Reducing Method (MDR). We have proven that both of them hold the quadratic convergence.
The idea behind the other proposed method, the Perturbed Dimension Reducing Method (PDR)[9], is the reduction of the dimensionality of the system as well as the perturbation of the Jacobian of the reduced system, by using proper perturbation parameters $A'_j, j = 1, ..., n-1$. Also, PDR method converges quadratically, while a proper choice of the parameters A'_j accelerates the convergence even further.
In Optimization Problems there is a large variety of methods for unconstrained optimization of functions :

$$f : \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}, \tag{5}$$

which require precise function and gradient values. However, in many optimization problems of practical interest the values of the objective functions and the corresponding gradients are known only imprecisely. For example, when the function and gradient values depend on the results of numerical simulations, then it may be difficult or impossible to obtain very precise values. Or, in other cases, it may be necessary to integrate numerically a system of differential equations in order to obtain a function value, so the precision of the computed value is limited[13].
The basic idea of DR method can successfully be applied for the computation of an unconstrained optimum of functions (5), so we have presented a suitable iterative procedure named Dimension Reducing Optimization Method (DROPT)[10] to solve optimization problems. Of course this method is also based on the methods studied in[6, 7, 8, 9] and it incorporates the advantages of DR algorithms. This method does not directly need any gradient evaluation and it compares favorably with quadratically convergent optimization methods. Moreover, we have perturbed the matrix obtained by the DR procedure in order to transform it into a symmetric as well as into a diagonal one. Finally, we have presented the DROPT method utilizing finite difference approximations.
The convergence of all DR methods have been studied. The numerical results obtained by applying all the above DR methods for solving systems and optimization problems are very promising.

2 DIMENSION REDUCING METHODS FOR SOLVING SYSTEMS OF NONLINEAR EQUATIONS IN $\mathcal{R}^n$

2.1 The Dimension Reducing Method – DR-method

By applying the known Implicit Function Theorem[3, 15] we have derived a method for solving systems of nonlinear algebraic and/or transcendental equations in $\mathcal{R}^n$, named Dimension Reducing Method (DR

method)[7]. To do this, assume that $F = (f_1, \ldots, f_n)\colon \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}^n$ is twice–continuously differentiable on an open neighborhood $\mathcal{D}^* \subset \mathcal{D}$ of a solution $x^* = (x_1^*, \ldots, x_n^*) \in \mathcal{D}$ of the system of nonlinear equations. Our interest lies in obtaining an approximation of x^*. So, we consider the sets $\mathcal{B}_i,\ i = 1, \ldots, n$ to be those connected components of $f_i^{-1}(0)$ containing x^* on which $\partial_n f_i \neq 0$, for $i = 1, \ldots, n$ respectively. Next, applying the Implicit Function Theorem [15] for each one of the components $f_i, i = 1, \ldots, n$ of F we can find open neighborhoods $\mathcal{A}_1^* \subset \mathcal{R}^{n-1}$ and $\mathcal{A}_{2,i}^* \subset \mathcal{R}, i = 1, \ldots, n$ of the points $y^* = (x_1^*, \ldots, x_{n-1}^*)$ and x_n^* respectively, such that for any $y = (x_1, \ldots, x_{n-1}) \in \bar{\mathcal{A}}_1^*$ there exist unique mappings φ_i defined and continuous in $\mathcal{A}_1^*$ such that $x_n = \varphi_i(y) \in \bar{\mathcal{A}}_{2,i}^*,\quad i = 1, \ldots, n$, and $f_i\big(y; \varphi_i(y)\big) = 0,\quad i = 1, \ldots, n$. Moreover, the partial derivatives $\partial_j \varphi_i, j = 1, \ldots, n-1$ exist in $\mathcal{A}_1^*$ for each $\varphi_i, i = 1, \ldots, n$, they are continuous in $\bar{\mathcal{A}}_1^*$ and they are given by $\partial_j \varphi_i(y) = -\partial_j f_i\big(y; \varphi_i(y)\big) / \partial_n f_i\big(y; \varphi_i(y)\big),\quad i = 1, \ldots, n,\quad j = 1, \ldots, n-1$. We utilize Taylor's formula to expand the $\varphi_i(y), i = 1, \ldots, n$ about y^p. By straightforward calculations we can obtain the following iterative scheme for the computation of the $n-1$ components of x^*

$$y^{p+1} = y^p + A_p^{-1} V_p,\quad p = 0, 1, \ldots\ , \tag{6}$$

where

$$\begin{aligned} y^p &= [x_i^p],\ i = 1, \ldots, n-1\ , \\ A_p &= [a_{ij}] = [\partial_j f_i(y^p; x_n^{p,i}) / \partial_n f_i(y^p; x_n^{p,i}) \\ &\qquad - \partial_j f_n(y^p; x_n^{p,n}) / \partial_n f_n(y^p; x_n^{p,n})],\quad i, j = 1, \ldots, n-1\ , \\ V_p &= [v_i] = [x_n^{p,i} - x_n^{p,n}],\quad i = 1, \ldots, n-1\ . \end{aligned} \tag{7}$$

Finally, after a desirable number of iterations of the above scheme, say $p = m$, we can approximate the nth component of x^* by means of the following relationship

$$x_n^{m+1} = x_n^{m,n} - \sum_{j=1}^{n-1} (x_j^{m+1} - x_j^m) \partial_j f_n(y^m; x_n^{m,n}) / \partial_n f_n(y^m; x_n^{m,n})\ . \tag{8}$$

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 $y = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

Theorem 2.1 *Suppose that $F = (f_1, \ldots, f_n)\colon \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}^n$ is twice–continuously differentiable on an open neighborhood $\mathcal{D}^* \subset \mathcal{D}$ of a point $x^* = (x_1^*, \ldots, x_n^*) \in \mathcal{D}$ for which $F(x^*) = \Theta^n$. Let $\mathcal{B}_i, i = 1, \ldots, n$ be those connected components of $f_i^{-1}(0)$, containing x^* on which $\partial_n f_i \neq 0$ for $i = 1, \ldots, n$ respectively. Then the iterates of (6) and the relationship (8) will converge to x^* provided the matrix A_* which is obtained from the matrix A_p of (6) at x^* is nonsingular and also provided the initial guess $y^0 = (x_1^0, \ldots, x_{n-1}^0)$ is sufficiently close to $y^* = (x_1^*, \ldots, x_{n-1}^*)$. Moreover the iterates $y^p, p = 0, 1, \ldots$ of (6) have order of convergence two.*

Proof: See[7]

Notation 2.1 We would like to mention here that the above process does not require the expressions φ_i but only the values $x_n^{p,i}$ which are given by the solution of the one–dimensional equations $f_i(x_1^p, \ldots, x_{n-1}^p, \cdot) = 0$. So, by holding $y^p = (x_1^p, \ldots, x_{n-1}^p)$ fixed we can solve the equations

$$f_i(y^p; r_i^p) = 0,\quad i = 1, \ldots, n\ , \tag{9}$$

for r_i^p in the inteval $(\alpha, \alpha + \beta)$ with an accuracy δ. Of course, we can use any one of the well–known one–dimensional methods[15, 17, 19] to solve the above equations. Here we use the one–dimensional bisection (see[4, 18] for a discussion of its advantages), since frequently the steps β are long and also a few significant digits are required for the computations of the roots of the equations (9). Specifically here we use a simplified version of the bisection method, which can be found in[20, 21, 22, 23]. It requires only that the algebraic signs of the function values be correct, so it can be applied to problems with imprecise function values.

2.2 Solving Systems of Nonlinear Equations in $\mathcal{R}^n$ Using a Rotating Hyperplane in $\mathcal{R}^{n+1}$ The Method of Rotating Hyperplane – the RM-method

We have proposed in[8] a procedure which can accelerate the convergence of other algorithms used for the numerical solution of nonlinear systems. We have used a rotating hyperplane to bring the approximations of solution closer to the solution of the system and thus to achieve a more rapid convergence of the iterations.
We have introduced a rotating hyperplane, whose rotation axis depends on the current approximation of $(n-1)$ components of the solution. We have derived and determined this plane in[8] by:

$$x_{n+1} = \sum_{i=1}^{n-1} A_i^{'}(x_i^0 - x_i), \quad \text{for} \quad A_i^{'} = A_i/A_{n+1} \ , \quad i = 1, \ldots, n-1 \ , \tag{10}$$

with $n-1$ free parameters $A_i^{'} \ \ i = 1, \ldots, n-1$ and assuming that $A_{n+1} \neq 0$.
The above rotating hyperplane has been applied on the traditional Newton's algorithm as well as on the proposed method DR, whence two modified schemes have been obtained, MN method and MDR method, correspondingly.

The Modified Newton Method–MN method

To derive the Modified Newton method we have replaced in (4) the usual hyperplane $x_{n+1} = 0$, at every iteration, by the above hyperplane (10) to get:

$$F(x^p) + F^{'}(x^p)(x - x^p) = \begin{pmatrix} \sum_{i=1}^{n-1} A_i^{'}(x_i^p - x_i) \\ \vdots \\ \sum_{i=1}^{n-1} A_i^{'}(x_i^p - x_i) \end{pmatrix}$$

So, after some matrix manipulations we end up with the following modified Newton's scheme

$$x^{p+1} = x^p - G^{'}(x^p)^{-1} F(x^p), \quad p = 0, 1, \ldots \ . \tag{11}$$

where $G^{'}(x^p) = F^{'}(x^p) + \Xi$ and $\Xi = [\xi_{ij}]$ is the rank–1 $n \times n$ matrix with

$$\xi_{ij} = \begin{cases} A_j^{'} & if \quad j \neq n \\ 0 & if \quad j = n \end{cases}$$

Theorem 2.2 *Suppose that $F = (f_1, \ldots, f_n) \colon \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}^n$ is twice–continuously differentiable on an open neighborhood $\mathcal{D}^* \subset \mathcal{D}$ of a point $x^* = (x_1^*, \ldots, x_n^*) \in \mathcal{D}$ for which $F(x^*) = \Theta^n$ and $F^{'}(x^*)$ nonsingular. Let $\Xi = [\xi_{ij}]$ be the above rank–1 $n \times n$ matrix where the vector $A^{'} = [A_j^{'}], j = 1, \ldots, n$, $A_n^{'} = 0$ determine the parameters of the rotating hyperplane (10) such that the inner product $\langle x, A^{'} \rangle = 0 \ \forall \ x \in \mathcal{D}^*$ and that $\|F^{'}(x^*)^{-1}\| \ \|\Xi\| \ < 1$. Then the iterates $x^p, p = 0, 1, \ldots$ of (11) will converge to x^* provided the initial guess x^0 is sufficiently close to x^*. Moreover the order of convergence will be two.*

Proof: See[8]

The Modified Dimension Reducing Method–MDR method

We have also used the rotating hyperplane (10) to derive a modified scheme of the above mentioned DR method. This scheme is derived in such a way that it can incorporate the advantages of nonlinear SOR and Newton's method. It is important to note that, although we have used reduction to simpler one–dimensional nonlinear equations, we have produced a quadratically converging sequence of points in $\mathcal{R}^{n-1}$.
We have defined the mapping

$$G = (g_1, \ldots, g_n) \colon \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}^n, \quad \text{by} \quad g_i(x_1, \ldots, x_n) = f_i(x_1, \ldots, x_n) + \sum_{j=1}^{n-1} A_j^{'} x_j \ . \tag{12}$$

It is evident that the solutions of the equations $g_i(x_1^p, \ldots, x_{n-1}^p, \cdot) = 0$, for $i = 1, \ldots, n$ are identical with the corresponding solutions of $f_i(x_1^p, \ldots, x_{n-1}^p, \cdot) = 0$ in $\mathcal{D}^*$. Moreover, it is obvious that $g_i(x_1^*, \ldots, x_n^*) = f_i(x_1^*, \ldots, x_n^*) = 0$ for $i = 1, \ldots, n$. Working exactly as in[7], we have obtained the following iterative scheme for the computation of the $n-1$ components of x^*

$$y^{p+1} = y^p + U_p^{-1} V_p, \quad p = 0, 1, \ldots \quad , \tag{13}$$

where

$$\begin{aligned} y^p &= [x_i^p],\ i = 1, \ldots, n-1\ , \\ U_p &= [a_{ij}] = [\big(\partial_j f_i(y^p; x_n^{p,i}) + A_j^{'}\big)/\partial_n f_i(y^p; x_n^{p,i}) \\ &\qquad - \big(\partial_j f_n(y^p; x_n^{p,n}) + A_j^{'}\big)/\partial_n f_n(y^p; x_n^{p,n})], \quad i, j = 1, \ldots, n-1\ , \\ V_p &= [v_i] = [x_n^{p,i} - x_n^{p,n}], \quad i = 1, \ldots, n-1\ . \end{aligned} \tag{14}$$

Finally, after a desired number of iterations of the above scheme, say $p = m$, we can approximate the nth component of x^* using the following relationship

$$x_n^{m+1} = x_n^{m,n} - \sum_{j=1}^{n-1} \{(x_j^{m+1} - x_j^m)\big(\partial_j f_n(y^m; x_n^{m,n}) + A_j^{'}\big)/\partial_n f_n(y^m; x_n^{m,n}\} \ . \tag{15}$$

Theorem 2.3 *Suppose that $F = (f_1, \ldots, f_n)\colon \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}^n$ is twice–continuously differentiable on an open neighborhood $\mathcal{D}^* \subset \mathcal{D}$ of a point $x^* = (x_1^*, \ldots, x_n^*) \in \mathcal{D}$ for which $F(x^*) = \Theta^n$. Let $\mathcal{B}_i, i = 1, \ldots, n$ be those connected components of $g_i^{-1}(0)$, containing x^* on which $\partial_n g_i \neq 0$ for $i = 1, \ldots, n$ respectively where the function g_i are defined in (12). Then the iterates of (13) and the relationship (15) will converge to x^* provided the matrix U_* which is obtained from the matrix U_p of (14) at x^* is nonsingular and also provided the initial guess $y^0 = (x_1^0, \ldots, x_{n-1}^0)$ is sufficiently close to $y^* = (x_1^*, \ldots, x_{n-1}^*)$. Moreover the iterates $y^p, p = 0, 1, \ldots$ of (13) have order of convergence two.*

Proof: See[8]

2.3 The Perturbed Dimension Reducing Method – PDR-method

To accelerate further the convergence of DR method we have proposed in[9] a modified method, named Perturbation Dimension Reducing method (PDR method), in which we have perturbed the corresponding Jacobian matrix using proper perturbation parameters $A_j^{'}$, $j = 1, \ldots, n-1$. So we have proposed the following iterative scheme for the computation of the $n-1$ components of x^* :

$$y^{p+1} = y^p + U_p^{-1} V_p, \quad p = 0, 1, \ldots \tag{16}$$

where :

$$\begin{aligned} y^p &= [x_i^p],\ i = 1, \ldots, n-1\,, \\ U_p &= [a_{ij}] = [\partial_j f_i(y^p; x_n^{p,i})/\partial_n f_i(y^p; x_n^{p,i}) \\ &\qquad - \partial_j f_n(y^p; x_n^{p,n})/\partial_n f_n(y^p; x_n^{p,n}) + A_j^{'}], \quad i, j = 1, \ldots, n-1\,, \\ V_p &= [v_i] = [x_n^{p,i} - x_n^{p,n}], \quad i = 1, \ldots, n-1\,. \end{aligned} \tag{17}$$

Finally, after a desired number of iterations of the above scheme, say $p = m$, we can approximate the nth component of x^*, as in DR method, using Relation (8).

Remark 2.1 *The perturbation parameters $A_j^{'}$, $j = 1, \ldots, n-1$, can be estimated in each iteration from the equation : $\langle x^p, A^{'} \rangle = 0$, $p = 0, 1, \ldots$, by choosing $n-2$ arbitrary parameters and calculating at each iteration the $(n-1)$th parameter from this equation.*

Theorem 2.4 *Suppose that $F = (f_1, \ldots, f_n)\colon \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}^n$ is twice–continuously differentiable on an open neighborhood $\mathcal{D}^* \subset \mathcal{D}$ of a point $x^* = (x_1^*, \ldots, x_n^*) \in \mathcal{D}$ for which $F(x^*) = \Theta^n$. Let $\mathcal{B}_i, i = 1, \ldots, n$ be those connected components of $f_i^{-1}(0)$ containing x^* on which $\partial_n f_i \neq 0$ for $i = 1, \ldots, n$, respectively. Suppose further that the matrix A_*, which is obtained from the matrix A_p of (7) at $y^* = (x_1^*, \ldots, x_{n-1}^*)$, is nonsingular and that $\Xi = [\xi_{ij}] = A_j^{'}$ is the rank–1 $(n-1)\times(n-1)$ matrix where the vector $A^{'} = [A_j^{'}], j = 1, \ldots, n-1$, determines the perturbation parameters so that the inner product $\langle x, A^{'}\rangle = 0,\ \forall\ x \in \mathcal{D}^*$ and $\|A_*^{-1}\|\,\|\Xi\| < 1$. Then the iterates of (16) and Relation (8) will converge to x^*, provided the initial guess $y^0 = (x_1^0, \ldots, x_{n-1}^0)$ is sufficiently close to y^*. Moreover, the iterates $y^p, p = 0, 1, \ldots$, of (16) have order of convergence two.*

Proof: See[9]

3 THE DIMENSION REDUCING METHODS FOR UNCONSTRAINED OPTIMIZATION

3.1 The Dimension Reducing Optimization Method – DROPT-method

We have derived and applied in[10] an iterative procedure, for the computation of an unconstrained optimum of functions (1), named Dimension Reducing Optimization Method (DROPT). This method is also based on the methods studied in[6, 7, 8, 9] and it incorporates the advantages of DR algorithms. This method does not directly need any gradient evaluation and it compares favorably with quadratically convergent optimization methods.

Specifically, to obtain a sequence $\{x^p\}, p = 0, 1, \ldots$ of points in $\mathcal{R}^n$ which converges to a local optimum (critical) point $x^* = (x_1^*, \ldots, x_n^*) \in \mathcal{D}$ of the function (1), we have applied the Implicit Function Theorem[15] for each one of the components $g_i, i = 1, \ldots, n$, where $g = (g_1, g_2 \ldots, g_n)$ indicates the gradient of the objective function. So we have introduced[10] the following iterative scheme, for the computation of the $n-1$ components of x^* :

$$y^{p+1} = y^p + A_p^{-1} V_p, \quad p = 0, 1, \ldots \tag{18}$$

where :

$$\begin{aligned} y^p &= [x_i^p],\ i = 1, \ldots, n-1\,, \\ A_p &= [a_{ij}] = \left[\frac{\partial_j g_i(y^p; x_n^{p,i})}{\partial_n g_i(y^p; x_n^{p,i})} - \frac{\partial_j g_n(y^p; x_n^{p,n})}{\partial_n g_n(y^p; x_n^{p,n})}\right], \quad i, j = 1, \ldots, n-1\,, \\ V_p &= [v_i] = [x_n^{p,i} - x_n^{p,n}], \quad i = 1, \ldots, n-1\,, \end{aligned} \tag{19}$$

with $x_n^{p,i} = \varphi_i(y^p)$. After a desired number of iterations of (18), say $p = m$, the nth component of x^* is approximated by means of the following relation :

$$x_n^{m+1} = x_n^{m,n} - \sum_{j=1}^{n-1} \left\{ (x_j^{m+1} - x_j^m)\, \frac{\partial_j g_n(y^m; x_n^{m,n})}{\partial_n g_n(y^m; x_n^{m,n})} \right\}. \tag{20}$$

The proposed method is illustrated in the following algorithm in pseudo–code where x^0 is the starting point, $a = (a_1, a_2, \ldots, a_n)$, $b = (b_1, b_2, \ldots, b_n)$ indicate the endpoints in each coordinate direction which are used for the one–dimensional bisection method[10, 20, 21, 22, 23], δ the predetermined accuracy for applying this procedure, MIT the maximum number of iterations required and $\varepsilon_1, \varepsilon_2$ the predetermined desired accuracies.

Algorithm 1 : *Dimension–Reducing Optimization (DROPT).*

1. *Input* $\{x^0;\ a;\ b;\ \delta;\ MIT;\ \varepsilon_1;\ \varepsilon_2\}$.
2. Set $p = -1$.
3. If $p < MIT$ replace p by $p + 1$ and go to next step; otherwise, go to Step 14.
4. If $\|g(x^p)\| \leq \varepsilon_1$ go to Step 14.

5. Find a coordinate *int* such that the following relation holds :

$$\operatorname{sgn} g_i(x_1^p,\ldots,x_{int-1}^p,a_{int},x_{int+1}^p,\ldots,x_n^p)\cdot\operatorname{sgn} g_i(x_1^p,\ldots,x_{int-1}^p,b_{int},x_{int+1}^p,\ldots,x_n^p)=-1,$$

for all $i=1,2,\ldots,n$. If this is impossible, apply Armijo's method (see[1, 10, 24]) and go to Step 4.

6. Compute the approximate solutions r_i for all $i=1,2,\ldots,n$ of the equation

$$g_i(x_1^p,\ldots,x_{int-1}^p,r_i,x_{int+1}^p,\ldots,x_n^p)=0,$$

by applying the one–dimensional bisection method[10, 20, 21, 22, 23] in (a_{int},b_{int}) within accuracy δ. Set $x_{int}^{p,i}=r_i$.

7. Set $y^p=(x_1^p,\ldots,x_{int-1}^p,x_{int+1}^p,\ldots,x_n^p)$.
8. Set the elements of the matrix A_p of Relation (19) using x_{int} instead of x_n.
9. Set the elements of the vector V_p of Relation (19) using x_{int} instead of x_n.
10. Solve the $(n-1)\times(n-1)$ linear system $A_p s^p=-V_p$ for s^p.
11. Set $y^{p+1}=y^p+s^p$.
12. Compute x_{int} by virtue of Relation (20) and set $x^p=(y^p;x_{int})$.
13. If $\|s^p\|\le\varepsilon_2$ go to Step 14; otherwise return to Step 3.
14. *Output* $\{x^p\}$.

Theorem 3.1 *Suppose that the objective function $f:\mathcal{D}\subset\mathcal{R}^n\to\mathcal{R}$ is twice–continuously differentiable in an open neighborhood $\mathcal{D}^*\subset\mathcal{D}$ of a point $x^*=(x_1^*,\ldots,x_n^*)\in\mathcal{D}$ for which $g(x^*)=\Theta^n$. Let $\mathcal{B}_i, i=1,\ldots,n$ be those connected components of $g_i^{-1}(0)$, containing x^* on which $\partial_n g_i\neq 0$ for $i=1,\ldots,n$ respectively. Then the iterations of (18) and the relationship (20) will converge to x^* provided the Hessian of f at x^* is nonsingular and also provided the initial guess $y^0=(x_1^0,\ldots,x_{n-1}^0)$ is sufficiently close to $y^*=(x_1^*,\ldots,x_{n-1}^*)$. Moreover the iterations $y^p, p=0,1,\ldots$ of (18) have order of convergence two.*

Proof. See[10]

3.2 A Perturbed Dimension–Reducing Optimization Method – PDROPT-method

The above mentioned DROPT method computes any critical point x^* of f (minimum, maximum or saddle) and it minimizes it if the matrix A_p of (19) is symmetric and positive definite so that the critical point is a minimizer. But, in general, this matrix is not symmetric. A case where it is symmetric is given by the following Lemma :

Lemma 3.1 *If the elements H_{in}, $i=1,\ldots,n-1$ of the Hessian are equal to each other, then the matrix A_p defined by Relation (19) is symmetric.*

Proof. See [10]

In the sequel we have perturbed the matrix A_p using proper perturbation parameters Δ_{ij}, $i,j=1,\ldots,n-1$ in order to transform it to a symmetric one. To this end we have considered the mapping :

$$W=(w_1,\ldots,w_{n-1}):\ \bar{\mathcal{A}}_1^*\subset\mathcal{R}^{n-1}\to\mathcal{R}^{n-1},\quad\text{with}$$
$$w_i(y)=-\varphi_i(y)+\varphi_n(y)+\sum_{j=1}^{n-1}\Delta_{ij}x_j\,,\quad i=1,\ldots,n-1\,,\tag{21}$$

where the matrix $\Delta=\Delta_{ij}$ of the perturbation parameters is taken such that the inner products :

$$\langle x,\Delta_i\rangle=0,\quad\forall\ \ x\in\bar{\mathcal{A}}_1^*\ \ \text{and}\ \ \forall\, i=1,\ldots,n-1,\tag{22}$$

with $\Delta_i=(\Delta_{i1},\ldots,\Delta_{i,n-1})$.

By choosing $n(n-1)/2$ arbitrary parameters $\Delta_{ij}, \quad i > j$ and taking :

$$\Delta_{ij} = a_{ji} + \Delta_{ji} - a_{ij}, \quad i < j, \tag{23}$$

while the remaining parameters Δ_{ii} are computed so that $\langle x, \Delta_i \rangle = 0$, the matrix $A_p = [a_{ij}]$ is transformed into a symmetric one. Since $\Delta_{ij}, \ i > j$ are arbitrarily chosen, we are able to transform A_p to a diagonal one. Thus, by taking :

$$\Delta_{ij} = -a_{ij} = \left[-\frac{\partial_j g_i(y^p; x_n^{p,i})}{\partial_n g_i(y^p; x_n^{p,i})} + \frac{\partial_j g_n(y^p; x_n^{p,n})}{\partial_n g_n(y^p; x_n^{p,n})} \right], \quad i = 1, \ldots, n-1, \quad i > j, \tag{24}$$

from Relations (23) and (22) our method becomes :

$$y^{p+1} = y^p - D_p^{-1} M_p, \quad p = 0, 1, \ldots \ , \tag{25}$$

where D_p is a diagonal matrix with elements $d_{ii}, \quad i = 1, \ldots, n-1$, given by :

$$\begin{aligned} d_{ii} = a_{ii} + \Delta_{ii} = & \left[\frac{\partial_i g_i(y^p; x_n^{p,i})}{\partial_n g_i(y^p; x_n^{p,i})} - \frac{\partial_i g_n(y^p; x_n^{p,n})}{\partial_n g_n(y^p; x_n^{p,n})} \right] + \\ & + \frac{1}{x_i} \sum_{\substack{j=1 \\ j \neq i}}^{n-1} \left\{ x_j \left[\frac{\partial_j g_i(y^p; x_n^{p,i})}{\partial_n g_i(y^p; x_n^{p,i})} - \frac{\partial_j g_n(y^p; x_n^{p,n})}{\partial_n g_n(y^p; x_n^{p,n})} \right] \right\}. \end{aligned} \tag{26}$$

Finally, after a desired number of iterations of the above scheme, say $p = m$, the nth component of x^* is approximated using Relation (20).

Theorem 3.2 *Suppose that the objective function $f : \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}$ is twice–continuously differentiable in an open neighborhood $\mathcal{D}^* \subset \mathcal{D}$ of a point $x^* = (x_1^*, \ldots, x_n^*) \in \mathcal{D}$ for which $g(x^*) = \Theta^n$. Let $\mathcal{B}_i, i = 1, \ldots, n$ be those connected components of $w_i^{-1}(0)$, containing x^* on which $\partial_n w_i \neq 0$ for $i = 1, \ldots, n$ respectively where the functions w_i are defined by (21). Suppose further that*

$$\|A_*^{-1}\| \, \|\Delta\| \ < 1 \, , \tag{27}$$

where the matrix A_*, is obtained from the matrix A_p of (19) at x^* and Δ is the perturbation matrix. Then the iterations (25) and Relation (20) will converge to x^* provided the Hessian of f at x^* is nonsingular and the initial guess $y^0 = (x_1^0, \ldots, x_{n-1}^0)$ is sufficiently close to $y^* = (x_1^*, \ldots, x_{n-1}^*)$. Moreover the iterations $y^p, p = 0, 1, \ldots$ of (25) have order of convergence two.

Proof. See[10]

3.3 A Finite–Difference Derivative Dimension–Reducing Optimization Method – FDDROPT-method

Finally we have proposed the dimension–reducing method for unconstrained optimization (FDDROPT method) using finite difference gradients and Hessian.
For a detail development of FDDROPT method see[10].

Notation 3.1 For a geometric interpretation of the Dimension–Reducing methods and a corresponding illustration of the main differences between Newton's method and the Dimension–Reducing methods, we refer the interested reader to [6].

4 NUMERICAL RESULTS

The proposed DR, Modified DR, Modified Newton method and PDR methods for solving nonlinear systems and the proposed DROPT, PDROPT and FDDROPT methods for optimization problems have been applied to problems of various dimensions. Our experience is that the procedures behave predictably and reliably and the results are quite satisfactory. In this report we suggestively give the results of two examples, one for systems of nonlinear equations and one for optimization problems. For more examples see[6, 7, 8, 9, 10].

Firstly, in Table 1 we present the results obtained by Newton's method and the iterative procedure DR applied to the following system (the Jacobian of this system is singular at the root r)

$$\begin{aligned} f_1(x_1,x_2,x_3) &= x_1x_3 - x_3e^{x_1^2} + 10^{-4} = 0 \\ f_2(x_1,x_2,x_3) &= x_1(x_1^2+x_2^2) + x_2^2(x_3-x_2) = 0 \\ f_3(x_1,x_2,x_3) &= x_1^3 + x_3^3 = 0 \end{aligned} \tag{28}$$

Next, in Table 2 we present the results obtained by applying DROPT method and known optimization and rootfinding methods (Fletcher – Reeves (FR)[16], Polak – Ribiere (PR)[16] and Broyden – Fletcher – Goldfarb – Shanno (BFGS)[2], Brown's method, a Brent – Gay modification of Brown's method, specifically the modification of Brown's method suggested by Brent and followed by Gay (BBG), as well as with Brent and Choleski – Newton method (CN)[16]), for the following optimization problem:

Example Rosenbrock function[14]. We calculate the optimum of the following objective function f: $f(x) = \sum_{i=1}^{2} {f_i}^2(x)$, where $f_1(x) = 10(x_2 - x_1^2), \qquad f_2(x) = 1 - x_1$

Newton's method							*DR method*					
x_1^0	x_2^0	x_3^0	$\varepsilon = 10^{-7}$		$\varepsilon = 10^{-14}$		$\varepsilon = 10^{-7}$			$\varepsilon = 10^{-14}$		
			IT	*FE*	*IT*	*FE*	*IT*	*FE*	*AS*	*IT*	*FE*	*AS*
-2	-2	-2	34	408	35	420	3	27	90	4	36	120
-1	-1	-1	30	360	31	372	2	18	60	3	27	90
-1	1	1	42	504	43	516	7	63	210	8	72	240
-0.5	-0.5	-0.5	31	372	32	384	2	18	60	3	27	90
-0.5	-0.5	0.1	23	276	26	312	2	18	60	3	27	90
0.5	0.5	0.1	44	528	45	540	2	18	60	3	27	90
0.5	0.5	0.5	28	336	30	360	2	18	60	3	27	90
1	-2	1	39	468	40	480	3	27	90	4	36	120
1	-1	1	37	444	38	456	7	63	210	8	72	240
1	1	1	46	552	47	564	6	54	180	7	63	210
2	-2	2	41	492	42	504	6	54	180	7	63	210
2	2	2	47	564	48	576	2	18	60	3	27	90

Table 1: Comparison of DR method with Newton's method

	Armijo		FR		PR		BFGS		DROPT		
x^0	IT	FE	IT	FE	IT	FE	IT	FE	IT	FE	ASG
$(-1.2,1)^\alpha$	1881	21396	142	2545	19	364	22	343	1	4	20
$(-3,6)$	5960	74560	194	4462	23	455	28	436	4	16	80
$(-2,-2)$	1828	20852	29	480	15	290	20	305	4	16	80
$(3,3)$	5993	74364	130	2939	26	509	25	384	4	16	80
$(1,20)$	D	D	259	5732	32	689	32	689	1	4	20
$(10,10)$	18416	251611	310	7469	26	526	32	505	4	16	80
$(100,100)$	D	D	D	D	33	746	54	822	2	8	40
$(-2000,-2000)$	2542	35743	D	D	93	2466	173	2667	2	8	40
	Brown		BBG		Brent		CN		DROPT		
x^0	IT	FE	IT	FE	IT	FE	IT	FE	IT	FE	ASG
$(-1.2,1)^\alpha$	22	110	16	80	17	85	8	48	1	4	20
$(-7,1)$	D	D	31	155	29	145	6	36	1	4	20
$(-3,45)$	D	D	D	D	D	D	7	42	4	16	80
$(10,10)$	70	350	29	145	27	135	6	36	4	16	80
$(-100,1)$	D	D	D	D	46	230	6	36	1	4	20
$(100,-100)$	D	D	50	250	43	215	6	36	2	8	40
$(100,100)$	D	D	49	245	41	205	6	36	2	8	40
$(-2000,-2000)$	D	D	D	D	D	D	6	36	2	8	40

Table 2: Comparison of DROPT method with other known optimization and rootfinding methods

REFERENCES

[1] Armijo L.(1996), "Minimization of functions having Lipschitz continuous first partial derivatives", Pacific J. Math. Vol. 16, pp. 1–3.

[2] Dennis J.E., Schnabel Jr. and R.B.(1983), *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice–Hall, Inc., Englewood Cliffs, New Jersey.

[3] Dieudonné J.(1969), *Foundations of modern analysis*, Academic Press, New York.

[4] Eiger A., Sikorski K. and Stenger F.(1984), "A bisection method for systems of nonlinear equations", ACM Trans. Math. Software, Vol. 10, pp. 367–377.

[5] Gorn A.(1954), "Maximal convergence intervals and a Gibb's–type phenomenon for Newton's approximation procedure", Ann. Math., Vol 59, pp. 463–476.

[6] Grapsa T.N. and Vrahatis M.N.(1989), "The implicit function theorem for solving systems of nonlinear equations in $\mathcal{R}^2$", Inter. J. Computer Math., Vol. 28, pp. 171–181.

[7] Grapsa T.N. and Vrahatis M.N.(1990), "A dimension–reducing method for solving systems of nonlinear equations in $\mathcal{R}^n$", Inter. J. Computer Math., Vol. 32, pp. 205–216.

[8] Grapsa T.N., Vrahatis M.N. and Bountis T.C.(1990), "Solving systems of nonlinear equations in $\mathcal{R}^n$ using a rotating hyperplane in $\mathcal{R}^{n+1}$", Inter. J. Computer Math., Vol. 35, 133–151.

[9] Grapsa T.N. and Vrahatis M.N.(1995), "A new dimension–reducing method for solving systems of nonlinear equations", Inter. J. Computer Math. Vol. 55, pp. 235–244.

[10] Grapsa T.N. and Vrahatis M.N.(1996), "A dimension–reducing method for unconstrained optimization", Journal of Computational and Applied Mathematics, Vol. 66, pp. 239–253.

[11] Kearfott R.B.(1979), "An efficient degree–computation method for a generalized method of bisection", Numer. Math., Vol. 32, pp. 109–127.

[12] Kearfott R.B.(1987), "Some tests of generalized bisection", ACM Trans. Math. Software, Vol. 13, pp. 197–220.

[13] Kupferschmid M. and Ecker J.G.(1987), "A note on solution of nonlinear programming problems with imprecise function and gradient values", Math. Program. Study, Vol 31, pp. 129–138.

[14] Moré B.J., Garbow B.S. and Hillstrom K.E.(1981), "Testing unconstrained optimization", ACM Trans. Math. Software, Vol. 7, pp. 17–41.

[15] Ortega J.M. and Rheinbolt W.C.(1970), *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York.

[16] Press W.H., Teukolsky S.A., Vetterling W.T. and Flannery B.P.(1992), *Numerical Recipes, The Art of Scientific Computing, 2nd ed.*, Cambridge University Press, New York.

[17] Rheinboldt W.C.(1974), *Methods for solving systems of equations*, SIAM, Philadelphia.

[18] Sikorski K.(1982), "Bisection is optimal", Numer. Math., Vol. 40, pp. 111–117.

[19] Traub J.F.(1964), *Iterative methods for the solution of equations*, Prentice–Hall., Inc., Englewood Cliffs, NJ.

[20] Vrahatis M.N.(1988), "Solving systems of nonlinear equations using the nonzero value of the topological degree", ACM Trans. Math. Software, Vol. 14, pp. 312–329.

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

[22] Vrahatis M.N.(1989), "A short proof and a generalization of Miranda's existence theorem", Proc. Amer. Math. Soc., Vol. 107, pp. 701–703.

[23] Vrahatis M.N. and Iordanidis K.I.(1986), "A rapid generalized method of bisection for solving systems of non–linear equations", Numer. Math., Vol. 49, pp. 123–138.

[24] Vrahatis M.N., G.S. Androulakis and G.E. Manoussakis (1996), "A new unconstrained optimization method for imprecise function and gradient values", J. Math. Anal. Appl., Vol. 197, No 2, pp. 586-607

[25] Yamamoto N.(1984), "Regularization of solutions of nonlinear equations with singular Jacobian matrices", J. Infor. Processing, Vol. 7, pp. 16–21.

[26] Weber H. and Werner W.(1981), "On the accurate determination of nonisolated solutions of nonlinear equations", Computing, Vol. 26, pp. 315–326.

ΠΕΡΙΛΗΨΗ

Σκοπός της εργασίας αυτής είναι να κάνουμε μια ανασκόπιση σε μια νέα κλάση μεθόδων, που ονομάζονται Μέθοδοι Ελάττωσης Διάστασης, τις οποίες έχουμε προτείνει για συστήματα μη γραμμικών εξισώσεων και για προβλήματα βελτιστοποίησης. Οι μέθοδοι αυτές βασίζονται στην αναγωγή του αρχικού συστήματος σε απλούστερες μονοδιάστατες μη γραμμικές εξισώσεις. Αν και οι μέθοδοι αυτές χρησιμοποιούν ελάττωση διάστασης συγκλίνουν τετραγωνικά και ενσωματώνουν έτσι τα πλεονεκτήματα των μη γραμμικών SOR αλγορίθμων και της μεθόδου Newton. Επιπλέον, δεν απαιτούν απευθείας συναρτησιακούς υπολογισμούς, με αποτέλεσμα να μπορούν να εφαρμοστούν σε προβλήματα με μη ακριβείς συναρτησιακές τιμές.