PROCEEDING

The 5th IMT - GT International Conference on Mathematics, Statistics and Their Applications

ICMSA 2009

Editors :
I Made Arnawa, Muhafzan, Maiyastri, Susila Bahri

Organized by :

June 9-11, 2009
The Hills Hotel
Bukittinggi, Indonesia

Proceedings of

The 5th IMT-GT

International Conference on Mathematics, Statistics, and their Applications 2009

(ICMSA 2009)

"Mathematics and Statistics for Industry and Community Development"

June 09 – 11, 2009

The Hills Hotel, Bukittinggi - Indonesia

Organized by :
Department of Mathematics, Andalas University,
Padang – Indonesia

Preface

First of all, I would like to say welcome to Bukittinggi, Indonesia to all of you. It is an honour for us to host this conference. We are very happy and proud because the participants of this conference come from many countries; we have participants from Libya, Japan, Qatar, India, Malaysia, Singapore, Thailand, Iran, and many more.

Ladies and gentlemen, according to constructivism theory, mathematics comes out as a result of social construction; that's why, the outcomes of our researches in mathematics, like theorem or formula of mathematics, should be communicated in a scientific forum such as seminar or conference. Through this kind of seminar or conference, we could refine the existing theorems or we could get new ideas to produce a new one. Seminar or conference which is held annually enables us to continually develop the science of mathematics until the end of the time.

That's way, in this two-day conference, we are going to discuss around 250 papers coming from diverse aspects of mathematics ranging from analysis, applied mathematics, statistics, algebra, Computational Mathematics, mathematics education, and other related topics.

For all of us here, I would like to convey my endless appreciation and gratitude for your participation in this conference.

Thank you very much

Dr. I Made Arnawa
Chairman of the Conference

Message from Rector Andalas University

It gives me great pleasure to extend my sincere and warm welcome to the participants of the 5th International Conference on Mathematics Statistics and Application (The IMT GT's 5th ICMSA 2009) - A Joint Scientific Program organized by Universities over Indonesia, Malaysia and Thailand Growth Triangle Region. On behalf of Andalas University, let me welcome all of you to the conference in Bukittinggi, West Sumatra Province, the land of Minang kabau.

We believe that from this scientific meeting, all of participants will have time to discuss and exchange ideas, findings, creating new networking as well as strengthen the existing collaboration in the respective fields of expertise. In the century in which the information is spreading in a tremendous speed and globalization is a trend, Andalas University must prepare for the tough competition that lay a head. One way to succeed is by initiating and developing collaborative work with many partners from all over the world. Through the collaboration in this conference we can improve the quality of our researches as well as teaching and learning process in mathematics and to achieve standards and requirements applied in many developed countries. I strongly believe that this conference is and extraordinary testimony to our capacity building at international, regional and local level.

I would like to express my deep gratitude to International Scientific Committee of who has honored the Mathematics Department, Faculty of Mathematics and Natural Sciences, Andalas University to host this prestigious conference. This is a very special opportunity for us to be involved in a regional community of knowledgeable scientist in the field of mathematics, statistics and their applications. I would also like to extend my gratitude to keynote speakers, participants, and organizer of this conference for their contribution to this event. My special thank is also rendered to the local government of West Sumatra for various supports and facilities.

Finally I wish all participants a fruitful deliberation at the conference. I also wish all participants and accompanying spouses a pleasant and enjoyable stay in Bukittinggi City, West Sumatra.

Prof. Dr. Ir. Musliar Kasim, MS
Rector

Organizing Committee

ORGANIZING COMMITTEE

Patrons : Rector Andalas University
Vice Rector of Academic Affair Andalas University

Steering : Dean Faculty of Sciences Andalas university
Vice Dean of Academic Affairs Andalas University

International Scientific Committee

Prof. Dr. Novesar Jamarun (Andalas University – Indonesia)
Prof. Dr. Herman Mawengkang (University of Sumatera Utara – Indonesia)
Prof. Dr. Dato Rosihan Ali (University Sains Malaysia)
Assoc. Prof. Dr. Ahmad Izani Md. Ismail (University Sains Malaysia)
Assoc. Prof. Dr. Putipong Bookkamana (Chiang Mai university – Thailand)
Assoc. Prof. Dr. Anton Abdulbasah Kamil (University Sains Malaysia)
Assoc. Prof. Dr. Pachitjanut Siripanitch (School of Applied Statistics – Thailand)
Dr. Saib Suwilo (University of Sumatera Utara – Indonesia)
Dr. Hizir Sofyan (Syiah Kuala University – Indonesia)
Dr. Tarmizi Usman (Syiah Kuala University – Indonesia)

Local Scientific Committee

Dr. Maiyastri (Andalas University – Indonesia)
Dr. Susila Bahri (Andalas University – Indonesia)
Dr. Muhafzan (Andalas University – Indonesia)
Adek Tasri, Ph.D (Andalas University – Indonesia)

Committee

Chairman	:	Dr. I Made Arnawa.
Secretary	:	Ir. Yudiantri Asdi, M.Sc.
	:	Izzati Rahmi, H. G., M.Si.
Transportation/Accomodation	:	Drs. Syafruddin, Zulakmal, M.Si.
Treasurer	:	Ir. Werman Kasoep, M.Kom., Jenizon, Msi, Ir. Hazmira Yozza, M.Si., Ir. Insanul Kamil, MS., Drs. Zulkarnain Khaidir, MS., Drs. Mulyadi, MS.
Technical Programme	:	Dr. Muhafzan, Dr. Maiyastri, M.Si.
Secretariat / Documentation	:	Budi Rudianto, M.Si., Narwen,M.Si., Arrival Rince Putri, M.T., Budi Rahmadya, S.Kom, Mohammad Hafiz Hersyah, S.Kom., Rahmi Eka Putri, S.Kom., Ratna Aisuwarya, S.Kom, Fajar Wisga, Berkah Fajar.

Published by :

Department of Mathematics, Andalas University
Kampus UNAND Limau Manis
Padang 25163, Sumatera Barat,
INDONESIA.

ISBN 978-602-95343-0-6

June 2009

Contents

CONTENTS

Computational Mathematics 85 - 180

Statistić 181 - 390

Computational Mathematics

Nonlinear Equations

Taufiq Iskandar[1], Marlan[2], Sanggam P. Gultom[3],
Herman Mawengkang[4],

[1]Matematika MIPA Unsyiah, [2]Universitas Simalungun Pematangsiantar
[3]STMIK Sisingamangaraja XII,
[4]Mathematics Graduate program, the University of Sumatera Utara

Abstract

In this paper we develop secant methods for nonlinear system using a population of previous iterates. Contrarily to classic secant methods, where exact interpolation is used, we prefer a least square approach to calibrate the linear model. We address an explicit control of the numerical stability of the method. We show that our approach can lead to an update formula. In that case, we prove that the local convergence of the corresponding undamped quasi-Newton method.

1. Introduction

We consider the standard problem of identifying the solution of a system of nonlinear equations

$$F(x) = 0 \tag{1}$$

where $F : R^n \rightarrow R^n$ is a diferentiable function. Since Newton, this problem has received a tremendous amount of attention. Newton's method and its many variations are still intensively analyzed and used in practice. The idea of Newton-like methods is to replace the nonlinear function F by a linear model, which approximates F in the neighborhood of the current iterate. The original Newton method invokes Taylor's theorem and uses the gradient matrix (the transpose of which is called the Jacobian) to construt the linear model. based on the secant equation. Because secant methods exhibit a q-superlinear rate of convergene, they have been intensively analyzed in the literature.

The secant equation imposes that the linear model exactly matches the nonlinear function F at two successive iterates. If the number of unknowns η is strictly greater than 1, an infinite number of linear models verify the secant equation. Therefore, each secant method derives a specific update formula which arbitrarily picks one linear model among them. The most common strategies are called "least-hange updates" and select the linear model whih minimizes the difference between two successive models.

In this paper, we provide a class of algorithms generalizing these ideas. Instead of using only two successive iterates to determine this linear model, we maintain a "population" of previous iterates. This approach allows all the available information colleted through the iterations to be explicitly used for calibrating the model.

An important feature of our method is that we do not impose an exact math between the model and the function. Instead, we use a least squares approach to request that the model fits the function "as well as possible". In this paper, we present the class of algorithms based on our method (Section 2.2) and prove that they are locally convergent (Section 3). This class of algorithms exhibits a faster convergene and a greater robustness than quasi-Newton methods for most numerial tests that we have performed (Setion 4) at a cost of substantial linear algebra computation. Therefore it is valuable when the cost of evaluating F is high in comparison with the numerical algebra overhead.

2. Quasi-Newton methods

Quasi-Newton methods consider at each iteration the linear model

$$L_k(x, B_k) = F(x_k) + B_k (x - x_k) \tag{2}$$

whih approximates F(x) in the neighborhood of x_k and computes x_{k+1} as a solution of the linear system $L_k(x; = B_k)$ 0. Consistently with most of the publications on this topic, quasi-Newton methods can be summarized as methods based on the following iterations:

$$x_{k+1} = x_k - B^{-1}_k F(x_k), \tag{3}$$

followed by the computation of B_{k+1}. The pure Newton method is obtained with $B_k = J(x_k) = \nabla F(x_k)^T$, the Jacobian of F evaluated at x_k, that is a $n \times n$ matrix such that entry (i, j) is $\partial F_i / \partial x_j$. We refer the reader to Dennis and Schnabel (1996) for an extensive analysis of Newton and quasi-Newton methods.

2.1. Secant methods

Broyden (1965) proposes a quasi-Newton method based on the secant equations, imposing the linear model L_{k+1} to exactly match the nonlinear function at iterates x_k and x_{k+1}, that is

$$\begin{aligned} L_{k+1}(x_k, B_{k+1}) &= F(x_k), \\ L_{k+1}(x_{k+1}, B_{k+1}) &= F(x_{k+1}) \end{aligned} \tag{4}$$

Subtracting these two equations and defining $y_k = F(x_{k+1}) - F(x_k)$ and $s_k = x_{k+1} - x_k$ we obtain the classical secant equation:

$$B_{k+1}s_k = y_k. \tag{5}$$

Clearly, if the dimension η is stritly greater than 1, there is an infinite number of matrices B_{k+1} satisfying (5). An arbitrary decision must consequently be made. The "least-change secant update" strategy, Clearly, if the dimension η is stritly greater than 1, there is an infinite number of matrices B_{k+1} satisfying (5). An arbitrary decision must consequently be made. The "least-change secant update" strategy, proposed by Broyden (1965), consists in seleting among the matrices verifying (5) the one minimizing variations (in Frobenius norm) between two successive matries B_k and B_{k+1}. It leads to the following update formula

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)s_k^T}{s_k^T\, s_k} \tag{6}$$

This method has been very sucessful, and has been widely adopted in the field. However, we believe that the idea of interpolating the linear model at only two iterates and ignoring previous iterates could be too restritive. Therefore, we propose to use more than two iterates to build the linear model.

This idea has already been considered. Dennis and Schnabel (1996) say that "Perhaps the most obvious strategy is to require the model to interpolate F(x) at other past points... One problem is that the directions tend to be linearly dependent or lose to it, making the computation of (the approximation matrix) a poorly posed numerical problem". Later, they write "In fact, multivariable generalizations of the secant method have been proposed ... but none of them seem robust enough for general use."

There are few attempts to generalize this approach in the literature. A first generalization of the secant method is the *sequential secant method* proposed by Wolfe (1959) and disussed by Ortega and Rheinboldt (1970). The idea is to impose exact interpolation of the linear model on $n + 1$ iterates instead of 2:

$$L_{k+1}(x_{k+1-j}, B_{k+1}) = F(x_{k+1-j}),\ \ j = 0, 1, \ldots, n. \tag{7}$$

or, equivalently,

$$B_{k+1}\, s_{k-j} = y_{k-j},\ \ j = 0, 1, \ldots, n-1, \tag{8}$$

where $s_i\ x_{k+1} = x_i$, and $y = F(x_{k+1})\ F(x_i)$, for all i. If the vectors $s_k, s_{k-1}, \ldots, s_{k-n+1}$ are linearly independent, exists exactly one matrix B_{k+1} satisfying (8), which is

$$B_{k+1} = Y_{k+1}S^{-1}{}_{k+1} \tag{9}$$

Where $Y_{k+1} = (y_k, y_{k-1}, \ldots, y_{k-n+1})$ and $S_{k+1} = (s_k, s_{k-1}, \ldots, s_{k-n+1})$.

Quoting Ortega and Rheinboldt (1970) "...(sequantial methods) are prone to unstable behavior and ... no satisfatory convergene results can be given". Nevertheless Gragg and Stewart (1976) propose a method which avoids instabilities by working with orthogonal factorizations of the involved matries. Martinez (1979) gives three implementations of the idea proposed by Gragg and Stewart (1976) and some numerical experiments.

Multi-step quasi-Newton methods have been proposed by Moghrabi (1993), Ford and Moghrabi (1997) and Ford (1999) in the context of nonlinear programming. An interpolating path is built based on previous iterates, and used to produce an alternative secant equation. Interestingly, the best numerical results were obtained with no more than two steps.

We believe that the comments about the poor numerical stability of those methods found in major reference texts such as Dennis and Schnabel (1996) and Ortega and Rheinboldt (1970) have not enouraged researchers to pursue these investigatations. We provide here a successful multi-iterates appoach with robust convergene properties and exhibiting an exellent behavior on numerical examples. The idea of using a least squares approach is similar to an idea proposed in the physis litterature by Vanderbilt and Louie (1984), whih has inspired other authors in the same field (Johnson, 1988, Eyert, 1996). Bierlaire and Crittin (forthoming) have used a similar approach for solving noisy large scale transportation problems.

2.2. Population-based approach

We propose a class of methods calibrating a linear model based on several previous iterates. The difference with existing approaches is that we do not impose the linear model to interpolate the function. Instead, we prefer to identify the linear model whih is as close as possible to the nonlinear function, in the least squares sense.

At each iteration, we maintain a finite population of previous iterates. Without loss of generality, we present the method assuming that all previous iterates $x_0, \ldots, x_{k+1}$ are considered. Our method belongs also to the quasi-Newton framework defined by (3), where B_{k+1} is computed as follows.

$$B_{k+1} = \arg\min_J \left(\sum_{i=0}^{k} \left\| \omega_{k+1}^i F(x_i) - \omega_{k+1}^i L_{k+1}(x_i; j) \right\|_2^2 + \left\| J\Gamma - B_{k+1}^0 + \tau \right\|_F^2 \right) \tag{10}$$

Where L_{k+1} defined by (2) and $B_{k+1}^0 \in \mathrm{R}^{n \times n}$ is an a priori approximation of B_{k+1}. The role of the second term is to overcome the under-determination of the least squares problem based on the first term and also control the numerical stability of the method. The matrix Γ contains weights assoiated with the arbitrary term B_{k+1}^0, and the weights $\omega_{k+1}^i \in \mathrm{R}^+$ are associated with the previous iterates. Equation (10) can be written in matrix form as follows: $B_{k+1} =$

$$\arg\min_J \left\| J\left(S_{k+1}\ I_{n\times n}\right) \begin{pmatrix} \Omega & o_{k\times n} \\ o_{n\times k} & \Gamma \end{pmatrix} - \left(Y_{k+1}\ B_{k+1}^0\right) \begin{pmatrix} \Omega & o \\ o & \Gamma \end{pmatrix} \right\|_F^2$$

where $\Omega \in \square^{k+1}$ is a diagonal matrix with weights ω_{k+1}^i on the diagonal for $i = 0, \cdots, k$. The normal equations of this least squares problem lead to the following formula:

The role of the a priori matrix B^0_{k+1} to overcome the possible underdetermination of problem (10). For example, choosing $B^0_{k+1} = B_k$ (similarly to classial Broyden-like methods) exhibits good properties. In that case, (11) becomes an update formula, and local convergene can be proved (see Section 3).

The weights ω^i_{k+1} capture the relative importance of each iterate in the population. Roughly speaking, they should be designed in the lines of the assumptions of Taylor's theorem, that is assigning more weight to points close to x_{k+1}, and less weight to points which are far away. The matrix Γ captures the importane of the arbitrary terms defined by B^0_{k+1} for the identification of the linear model. The weights have to be finite, and Γ must be such that

$$\Gamma\Gamma^T + S_{k-1}\Omega^2 S^T_{k+1} \tag{12}$$

is safely positive definite. To ensure this property we describe below three possible approaches for choosing $\Gamma\Gamma^T$: the *geometrical approach*, based on specific geometri properties of the population, the subspace decomposition approach, decomposing R^n into the subspae spanned by the columns of S_{k+1} and its orthogonal complement, and the numerical approach, designed to guarantee a numerically safe positive definiteness of (12).

The *geometrical approach* assumes that $n + 1$ members of the population form a simplex, so that the columns of S_{k+1} span R^n, and (12) is positive definite with $\Gamma\Gamma^T = 0$. In that case, (11) becomes

$$B_{k+1} = Y_{k+1}\Omega^2 S^T_{K=1}(S_{k+1}\Omega^2 S^T_{K=1})^{-1} \tag{13}$$

If there are exactly $n + 1$ iterates forming a simplex, the geometrical approach is equivalent to the interpolation method proposed by Wolfe (1959), and (13) is exactly (9), as S_{k+1} is square and non singular in that ase. This approach have not shown good numerical behavior in practice as mentioned in Section 2. Also, it requires at least $n + 1$ iterates, and may not be appropriate for large-scale problems.

The *subspace deomposition* approach is based on the QR decomposition of S_{k+1}. We denote by r the rank of S_{k+1}, with $r \leq$ n, and we have $S_{k+1} = QR$, where

$$Q = (Q1\ \ Q2) \tag{14}$$

With Q_1 is $(n \times r)$, Q_2 is $(n \times n - r)$, and R is $(n \times k + 1)$. The columns of Q_1 form an orthogonal basis of the range of S_{k+1}. We define now Γ such that

$$\Gamma = (\,0_{n\times r} Q_2\,) \tag{15}$$

that is Q where Q_1 has been replaced by a null matrix. With this construction $\Gamma\Gamma^T + S_{k+1}\Omega^2 S^T_{k+1}$ is invertible and $S_{k+1}\,\Gamma\Gamma^T = 0$. In the case where S_{k+1} spans the entire space then $r = n$, Γ is a null matrix and (11) is equivalent to (13).

With the subspace decomposition approach, the changes of F predited by B_{k+1} in a direction orthogonal to the range of S_{k+1} is the same as the one predited by the arbitrary matrix B_{k+1}.This idea is exactly the same as the one used by Broyden (1965) to construct his so called *Broyden's Good method.*

Numerical problems may happen when the columns of S_{k+1} are close to linear dependence. These are the problems already mentioned in the introduction, and reported namely by Ortega and Rheinboldt (1970) and Dennis and Schnabel (1996). Clearly, such problems do not occur when S_{k+1} has exactly one column, which leads to the classical Broyden method.

The numerical approach is designed to address both the problem of overcoming the under-determination, and of guaranteeing numerical stability. It is directly inspired by the modified Cholesky factorization proposed by Schnabel and Eskow (1991). The modified Cholesky fatorization of a square matrix A creates a matrix E such that A + E is safely positive definite, while computing its Cholesky factorization. It may namely happen that A has full rank, but with smallest eigenvalue very small with regard to machine preision. In that case, E is non zero despite the fact that A is non singular. We apply this technique with $A = S_{k+1}\,\Omega^2 S^T_{k+1}$ and $E = \Gamma\Gamma^T$ .So, if the matrix $S_{k+1}\,\Omega^2\,S^T_{k+1}$ is safely positive definite, $\Gamma\Gamma^T = 0$ and (11) reduces to (13). If not, the modified Cholesky factorization guarantees that the role of the arbitrary term Γ is minimal.

We now emphasize important advantages of our generalization combined with the *numerical approach*. Firstly, contrarily to interpolation methods, our least squares model allows to use more than p points to identify a model in a subspace of dimension p (where p $\leq$ n). This is very important when the objective function is expensive to evaluate. Indeed, we make an efficient use of all the available information about the function to calibrate the secant model. It is namely advantageous compared to Broyden's method, where only two iterates are explicitly used to build the model, while previous iterates only play an implicit role due to the "least-change" principle. Secondly, the numerical approach proposed above controls the numerical stability of the model constrution process, when a sequene of iterates may be linearly dependent. Finally, the fact that existing methods are special cases of our approach allows to generalize the theoretical and pratical properties already published in the literature, and simplifies their extension to our context. We apply this principle to the local convergene analysis in section 3. The main drawbak is the increase in numerical linear algebra as the least squares problem (10) must be solved at each iteration. Therefore, it is particularly appropriate for problems where F is very expensive to compute.

numerical linear algebra as the least squares problem (10) must be solved at each iteration. Therefore, it is particularly appropriate for problems where F is very expensive to compute.

We conclude this section by showing that our population-based update formula is a generalization of Broyden update. Actually, the classical Broyden update (6) is a special case of our update formula (11), if $B^0_{k+1} = B_k$, the population contains just two iterates x_k and x_{k+1}, and the *subspace decomposition* approach is used. The secant equation (5) completely defines the linear model in the one-dimensional subspace spanned by $s_k = x_{k+1} - x_k$, while an arbitrary decision is made for the rest of the model. If we define $\omega^k_{k+1} = 1$ and Γ is given by (15) with r = 1, we can write (11) as

$$B_{k+1} = B_k + (Y_k - B_k s_k)\, s_k^T \left(\Gamma\Gamma^T + s_k s_k^T\right)^{-1} \tag{16}$$

The equivalence with (6) is due to the following equality

$$s_k^T \left(\Gamma\Gamma^T + s_k s_k^T\right)^{-1} = s_k^T \frac{1}{s_k^T s_k} \tag{17}$$

obtained from the fat that $s_k^T\ \Gamma\Gamma^T = 0$, by (15)

3. Local convergence analysis

We show that if $\Gamma\Gamma^T$ is determined by the numerical approach described in Section 2.2, then the undamped algorithm described in Section 3.1, where B_{k+1} is defined by (11) in its update form (i.e. $B^0_{k+1} = B_k$), locally converges to a solution of (1) if the following assumptions are verified. Note that the assumptions made on the problem are similar to those given by Broyden (1965).

Assumptions on the problem:

(P_1) $F: R^n \to R^n$ is continuosly differentiable in an open convex set D

(P_2) The System of equations has a solution, that is $\exists\, x^* \in D$ such that $F(x^*) = 0$

(P_3) $J(x)$ is Lipschitz continuosly at x^* with constant K_{lip}, that is

$$\| J(X) - J(x^*) \| \le K_{lip} \| x - x^* \|\ \forall x \in D \tag{18}$$

In the neihgboorhood D

(P_4) $J(x^*)$ is non singular and three exist $\gamma > 0$ such that $\| J(x^*)^{-1} \| < \gamma$

Assumptions on the algorithm :

(A1) The algorithm is based on the iteration (3) with x_0 and B_0 as initial guess

(A2) B_k is generated by (11) with $B^0_{k+1} = B_k$.

(A3) $\Gamma\Gamma^T$ is computed using the numerical approach.

(A4) $\forall i \le k$, we have $\omega^i_{k+1} \le M_\omega$ for all k and some constant $M_\omega > 0$.

(A5) The size of the population ρ is bounded above by M_ρ where $M_\rho > 0$ is a constant

The notation $\|\cdot\|$ is used for the l_2 vector norm $\|x\| = (x^T x)^{\frac{1}{2}}$ as well as for the Frobenius matrix norm $\|A\|$. The notation $\|\cdot\|_2$ is used for the l_2 matrix norm $\|A\|_2$. For the sake of simplification, we denote $\omega^i_{k+1} = \omega_i$, $S = S_{k+1}$, $Y = Y_{k+1}$, and $I_p = \{0, 1, \ldots, P\}$. The proof uses some lemmas. Lemma 1 and 2 are classical results from the literature. Lemmas 3-5 are tehnical lemmas related to our method. Their proofs are provided in the appendix.

Lemma 1 *Let* $F: R^n \to R^n$ *be continuously diferentiable in the open convex* $D \subset R^n$, $x \in D$, *and let j be Lipshitz continuous at x in the neighborhood D with constant* K_{lip}. *Then for any* $u, v \in D$,

$$\|F(v) - F(u) - J(x)(v-u)\| \le K_{lip} \frac{\|v - x\| + \|u - x\|}{2} \|v - u\|. \tag{19}$$

Proof. See, for example, Dennis and Schnabel, 1996.

Lemma 2 *Let* A, C $\in R^{n\times n}$ *and assume that A is invertible, with* $\|A^{-1}\| \le \mu$. *If* $\|A - c\| \le \beta$ *and* $\beta\ \mu < 1$, *then C is also invertible and*

$$\|C^{-1}\| \le \frac{\mu}{1 - \beta\mu} \tag{20}$$

Proof. This lemma is known as the Banach Perturbation Lemma. (See, for example, Ortega and Rheinboldt, 1970).

Lemma 3 *If assumptions* (A_4) - (A_5) *are verified, then*

$$\|S\ \Omega^2\ S^T\| \le 2 M_\rho\ M_\omega^2 \max_{i \in I_{k+1}} \|x_i - x^*\|^2, \tag{21}$$

Lemma 4 *If assumptions* (P_1), (P_2) *and* (P_3) *are verified then:*

$$\left\|\left(Y-J(x^*)S\right)\right\| \le \sqrt{2\,\mathrm{M}_\rho}\; K_{lip} \max_{i\in I_{k+1}}\left(\left\|x_i - x^*\right\|^2\right) \tag{23}$$

Where x* is solution of (1).

Lemma 5 *If assumption* (A_3) *is verified, then*

$$\left\|\left(\Gamma\Gamma^T + S\,\Omega^2\,S^T\right)^{-1}\right\|_2 \le \frac{1}{\tau} \tag{24}$$

where $\tau > 0$

The parameter τ in Lemma 5 controls the way we perturb $S\Omega^2S^T$. It guarantees that the smallest eigenvalue of $(\Gamma\Gamma^T + S\Omega^2S^T)$ is strictly greater than τ and, therefore, safely positive in a finite arithmatic context if τ is properly chosen. Schnabel and Eskow (1991) suggest to choose τ $(\text{macheps})^{1/3}$ where macheps is the machine epsilon.

Theorem 6 *Let assumptions* (P_1) *to* (P_4) *hold for the problem and assumptions* (A_1) *to* (A_5) *hold for the algorithm. Then there exists two non-negative constants* α_1 *and* α_2 *such that for each* x_k *and* B_k*:*

$$\left\|B_{k+1} - J(x^8)\right\| \le \left(1 + \alpha_1 \max\nolimits_{i\in Ik+1}\left\|x_i - x^*\right\|^2\right)\left\|B_k - J(x^*)\right\|$$
$$+ \alpha_2 \max\nolimits_{i\in I_{k+1}}\left\|x_i - x^*\right\|^3.$$

Proof. *From the update formula (11), and defining*

$$T_1 = I - S\Omega^2 S^T(\Gamma\Gamma^T + S\Omega^2 S^T)^{-1}$$
$$T_2 = (Y - J(x^*)S)\,\Omega^2 S^T\,(\Gamma\Gamma + S\Omega^2 S^T)^{-1},$$

we obtain

$$\begin{aligned}\left\|B_{k+1} - J(x^*)\right\| &= \left\|B_k - J(x^*) + \left[(J(x^*)S - J(x^*)S) + (Y - B_k S)\right]\Omega^2 S^T(\Gamma\Gamma^T + S\Omega^2 S^T)^{-1}\right\| \\ &\le \|T_1\|\left\|B_k - J(x^*)\right\| + \|T_2\|.\end{aligned} \tag{25}$$

from Lemmas 3 and we obtain

$$\|T_1\| \le \|I\| + \left\|S\Omega^2 S^T\right\|\left\|\left(\Gamma\Gamma^T + S\Omega^2 S^T\right)^{-1}\right\| \tag{26}$$

$$\le 1 + \alpha_1 \max_{i\in I_{k+1}}\|x_i - x^*\|^2 \tag{27}$$

with

$$\alpha_1 = \frac{2\sqrt{n}}{\tau}\mathrm{M}_\rho \mathrm{M}_\omega^2 > 0$$

we conclude the proof using Lemmas 3, 4 and 5 to show that :

$$\|T_2\| \le \left\|\left(Y - J(x^*)S\right)\right\|\left\|\Omega^2\;S^T\right\|\left\|\left(\Gamma\Gamma^T + S\Omega^2 S^T\right)^{-1}\right\| \tag{28}$$

$$\le \alpha_2 \max_{i\in I_{k+1}}\left\|x_i - x^*\right\|^3 \tag{29}$$

with

$$a_2 = \frac{2\sqrt{n}}{\tau}\,K_{lip}\,\mathrm{M}\rho\mathrm{M}_\omega^2 > 0$$

Theorem 7 *Let assumptions* (P_1) *to* (P_3) *hold for the problem and assumptions* (A_1) *to* (A_5) *hold for the algorithm. Then for each* $r \in]0,\ 1[$, *there exists* $\varepsilon(r)$ *and* $\delta(r)$ *such that for*

$$\left\|x_0 - x^*\right\| \le \varepsilon(r) \tag{30}$$

and

$$\left\|B_0 - J(x^*)\right\| \le \delta(r) \tag{31}$$

the sequence $x_{k+1} = x_k - B^{-1}_k\,F(x_k)$ *is well defined and converges q-linearly to x* with q-factor at most r. Furthermore, the sequences* $\left\{\|B_k\|\right\}_k$ *and* $\left\{\|B^{-1}{}_k\|\right\}_k$ *are uniformly bounded.*

Proof. The structure of the demonstration is similar to the proof of Theorem 3.2 in Broyden et al. (1973). We have purposedly skipped some identical tehnical details.

First choose $\varepsilon(r) = \varepsilon$ and $\delta(r) = \delta$ such that

$$\gamma(1 + r)(K_{lip}\;\varepsilon + 2\delta) \le r \tag{32}$$

and

and

$$\left(2a_1 + a_2 \frac{\varepsilon}{1-r}\right)\frac{\varepsilon^2}{1-r^2} \leq \delta. \tag{33}$$

We invoke Lemma 2 with $\mu = \gamma$ and $\beta = 2\delta$ to prove that B_0 is non-singular and

$$\left\|B_0^{-1}\right\| < \gamma\,(1+r). \tag{34}$$

Note that assumption $2\,\delta\,\gamma < 1$ for Lemma 2 is directly deduced from (32).

The improvement after the first iteration, that is

$$\left\|x_1 - x^*\right\| \leq r\left\|x_0 - x^*\right\| \tag{35}$$

is independent of the specific update formula and, therefore, is proven in Broyden et al. (1973).

The result for iteration k is proven with an induction argument based on the following recurrence assumptions:

$$\left\|B_m - J^*\right\| \leq 2\delta \tag{36}$$

$$\left\|x_{m+1} - x^*\right\| \leq r\left\|x_m - x^*\right\| \tag{37}$$

for all $m = 1, \ldots., k-1$

we first prove that $\left\|B_k - J^*\right\| \leq 2\delta$ using Theorem 6. From (25) we deduce

$$\begin{aligned}&\left\|B_{m+1} - J(x^*)\right\| - \left\|B_m - J(x^*)\right\| \\ &\leq \alpha_1 \max_{i\in I_{m+1}}\left\|x_i - x^*\right\|^2 \left\|B_m - J(x^*)\right\| + \alpha_2 \max_{i\in I_{m+1}}\left\|x_i - x^*\right\|^3 \\ &\leq a_1\, r^{2(m+1)}\,\varepsilon^2\, 2\delta + a_2\, r^{3(m+1)}\,\varepsilon^3.\end{aligned} \tag{38}$$

Summing both sides of (38) for m ranging from 0 to k - 1, we deduce that

$$\left\|B_k - J(x^*)\right\| \leq \left\|B_0 - J(x^*)\right\| + \left(2\alpha_1\,\delta + \alpha_2 \frac{\varepsilon}{1-r}\right)\frac{\varepsilon^2}{1-r^2} \tag{39}$$

where (40) derives from (31) and (33).

The fact that B_k is invertible and $\left\|B_k^{-1}\right\| \leq \gamma(1+r)$ is again a direct application of the Banach Perturbation Lemma 2. Following again Broyden et al. (1973), we can now obtain (37) for m = k, conluding the induction proof.

3.1. Undamped and damped quasi-Newton methods

All the algorithms presented in Section 2.1 and 2.2 are based on the following structure.

- Given $F : R^n \to R^n$, x_0 □R^n and B_0 □$R^{n\times n}$
- While stopping criteria is not verified:
  – Find s solving $B_{ks} =$ $-F(x_k)$,
  - Evaluate $F(x_{k+1})$ where $x_{k+1} = x_k + s$,
  - Compute B_{k+1}.

This general algorithm is often called undamped quasi-Newton method, i. e. without any step control or globalization methods. It allows to compare different type of algorithms, in term of number of funtion evaluations, and their robustness without introducing a bias due to the step control or the globalization method. Consequently, the algorithms differ only by the method used to compute B_{k+1}.

The main drawback of undamped methods is that we cannot ensure convergene from remote starting points. Moreover, Newton-like methods without any control on the step lengths may encounter several other soures of failure. For instance, the components of the unknown vector (x) or the funtion vector (F) or the Jacobian approximate (B_k) may become arbitrarily large.

Globalization strategies can be grouped into two disticnt frameworks: linesearch and trust-region. Linesearch approaches are applied to a merit function based on F, used to measure progress toward a solution of F(x) = 0 (see for instance Noedal and Wright, 1999). Trust-region methods and filter-trust-region methods (see Gould et al., 2005) can be used to solve the associated nonlinear least squares problem:

$$\min_{x\in R^n} \frac{1}{2}\left\|F(x)\right\|_2^2 \tag{41}$$

The main disadvantage of the second type of globalization is that the iterates can be stucked in a local minimum of (41), which is not a solution of F(x) = 0. As we want to keep solving the original problem F(x) = 0, we adopt in this paper the linesearch approach.

When integrating a linesearch strategy to the previous undamped quasi- Newton framework, we obtain the following structure.

- Given $F : R^n \to R^n$, $x_0 \in R^n$ and $B_0 \in R^{n\times n}$
- While stopping criteria is not verified :

- Compute B_{k+1}.

This general method is called *damped* quasi-Newton method. In the following, we describe how we determine the step α_k at each iteration of the algorithm using the classical sum-of-squares merit function

$$m(x_k) = \frac{1}{2} \| F(x_k) \|_2^2 = \frac{1}{2} \sum_{i=1}^{n} F_i^2 (x_k)$$

to measure progress toward a solution of the system F. We choose a step α_k satisfying the following Armijo-type condition with $\beta \in (0,1)$:

$$m(x_k + \alpha_k s) \le m(x_k) + \alpha_k \beta \nabla m(x_k)^T s .$$

Note that β is a parameter which defines the quality of the decrease we want to obtain. Condition (42) is valid only if the quasi-Newton direction s is a descent diretion for min x_k, that is:

$$\nabla m(x_k)^T s < 0 . \tag{43}$$

If condition (43) holds, we find a step α_k satisfying (42) using a backtraking strategy. Unfortunately, we do not have the guarantee that our quasi-Newton direction $s = -B^{-1}_k F(x_k)$ is a descent direction for m, unless B_k is close enough to the real Jacobian at x_k, $J(x_k) = \nabla F(x_k)^T$, and $\nabla m(x_k)^T s$ is bounded below. Consequently, we use the following sequential procedure to find a desent direction for the merit function in the current iterate x_k:

- Check whether the quasi-Newton direction $s = -B^{-1}_k F(x_k)$ is a descent direction for min x_k;
- If not, compute using the modified Cholesky factorization (see Schnabel and Eskow, 1999) can auxiliary direction š
  where $\tau > 0$ and I is the identity matrix in dimension n. Aording to Nocedal and Wright (1999), we an always choose τ to ensure that $m(\nabla xk)^T s$ is bounded below.
- Check whether the quasi-Newton direction s is a descent direction for m in x_k;
- If not, do the following:
  - Update the current approximation of the Jacobian B_k with a new point close to x_k to get B_k. More precisely, we take a step of length 1e -4 in the direction s. The goal is to try to get a good local approximation of J (x_k);
  - Compute the direction $s^+ = -(B_k)^{-1} F(x_k)$;

  and restart the process with s^+.

Note that we compute the directional derivative of the merit function m in a direction s, $\nabla m(x)^T s$, using a finite differences procedure.

4. Numerical Results

4.1. General behavior

We present here an analysis of the performance of our method, in comparison to classical algorithms. All algorithms and test functions have been implemented with the package Octave (Eaton, 1997) and computations have been done on a desktop equipped with 3GHz CPU in double precision. The machine epsilon is about 2.2e-16.

The numerical experiments were carried out on a set of 43 test functions. For 37 of them, we consider five instances of dimension n = 6, 10, 20, 50, 100. We obtain a total of 191 problems. This set is composed of the four standard nonlinear systems of equations proposed by Dennis and Schnabel (1996) (that is, Extended Rosenbrok Function, Extended Powell Singular Function, Trigonometric Function, Helial Valley Function), three functions from Broyden (1965), five functions proposed by Kelley (2003) in his book on Newton's method (that is, Arc tangent Function, a Simple Two-dimensional Function, Chandrasekhar H-equation, Ornstein -Zernike Equcations, Right Preconditioned Convetion-Diffusion Equcation), three linear systems of equcations (see Appendix), the test functions given by Spediato and Huang (1997) and some test functions of the colletion proposed by More et al. (1981). For each problem, we have used the starting point proposed in the original paper. Note that the results include all these problems.

The algorithms are based on both the damped and undamped quasi-Newton framewok given in Section 3.1 with the following characteristics: the initial Jacobian approximation B_o is the same for all algorithms and equal to the identity matrix. The stopping criterion is a composition of three conditions: small residual, that is $\|F(x_k)\| / \|F(x_0)\| \le 10_e$-6, maximum number of iterations ($k \ge 200$ for problems of size $n \le 20$ and $k \ge 500$ for problems of size $n > 20$), and divergene, diagnosed if $\|F(x_k)\| \ge 10_e10$or if a descent direction has not been found after several updates of the approximate Jacobian in the linesearch procedure (meaning that we have not been able to find a sufficently good approximation of the Jacobian).

We consider four quasi-Newton methods:

1. Broyden's Good Method (BGM), using the update (6).
2. Broyden's Bad Method (BBM), also proposed by Broyden (1965). It is based on the following secant equation:

$$s_k = B_{k+1}^{-1} y_k \tag{44}$$

and dircetly computes the inverse of B_k:

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1}\ y_k)\,y_k^T}{y_k^T\ y_k} \tag{45}$$

Broyden (1965) describes this method as "bad", that is numerially unstable. However, we have decided to include it in our tests for the sake of completeness. Moreover, as discussed below, it does not always deserve its name.

3. The Hybrid Method (HMM) proposed by Martinez (1982). At each iteration, the algorithm decides to apply either BGM or BBM. Martinez (2000) observes a systematic improvement of the Hybrid approach with respect to each individual approach. As discussed below, we reach similar conclusions.
4. Our population-based approach, called Generalized Secant Method (GSMz defined by (11) in its update form with B_{k+1}^0 = B_k using the numerical approach described in Section 2.2, with τ = (macheps)$^{1/3}$ and a maximum of p = max(n, 10) previous iterates in the population. Indeed, including all previous iterates, as proposed in the theoretical analysis, may generate memory management problems, and anyway does not significantly affect the behavior of the algorithm. The weights are defined as

$$\omega_{k+1}^i = \frac{1}{\left\| X_{k+1} + X_i \right\|^2} \quad \forall i \in I_p$$

The measure of performance is the number of function evaluations to reach convergence. Indeed we are interested in applying the method on computation-nally expensive systems, where the running time is dominated by the function evaluations. We are presenting the results following the *performance profiles* analysis method proposed by Dolan and More (2002).

If $f_{p,a}$ is the performance index (the number of function evaluations in our case) of algorithm a on problem p, then the performance ratio is defined by

$$r_{p,a} = \frac{f_{p,a}}{\min_a \{f_{p,a}\}}, \tag{47}$$

if algorithm a has converged for problem p, and $r_{p,a} = r_{fail}$ otherwise, where r_{fail} must be strictly larger than any performance ratio (47). For any given threshold π, the overall performance of algorithm a is given by

$$\rho_a(\pi) = \frac{1}{n_p}\,\Phi_a(\pi)$$

where n_p the number of problems considered, and $\Phi_a(\pi)$ is the number of problems for wchih $r_{p,a} \le \pi$.

In particular, the value $\rho_a(1)$ gives the probability that algorithm a wins over all other algorithms. The value $\lim \pi \rightarrow r_{fail}\ \rho_a(\pi)$ gives the probability that algorithm a solves a problem and, consequently, provides a measure of the robustness of eah method.

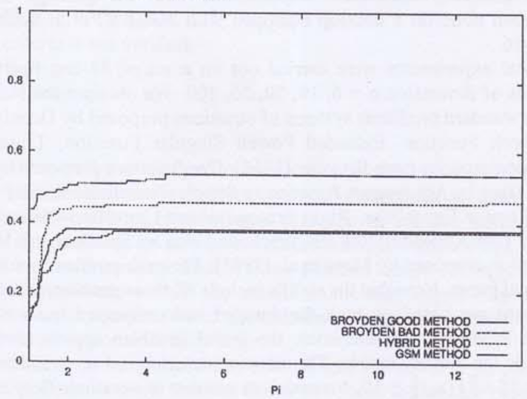


Figure1: Perfomance Profile

We first analyze the performance profile of all algorithms desribed above without globalization strategy on all problems. The performance profile is reported on Figure 1. A zoom on πbetween 1 and 5 is provided in Figure 2.

The results are very satisfactory for our method. Indeed, we observe that GSM is the most efficent and the most robust algorithm among the challenged quasi-Newton methods.

We also confirm results by Martinez (2000) showing that the Hybrid method is more reliable than BGM and BBM. Indeed, it converges on almost 50% of the problems, while each Broyden method converges only on less than 40% of the cases. Moreover, HMM wins more often than BGM and BBM does, and is also more robust, as its performance profile grows faster than the profile for BGM and BBM. The relative robustness of BGM and BBM is comparable.

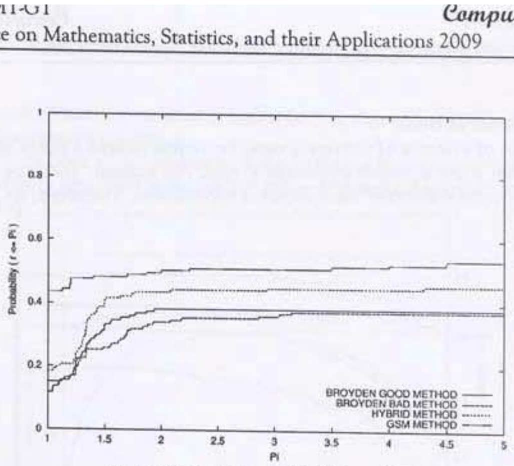


Figure 2: Perfomance Profileon (1,5)

Even if GSM is the most reliable algorithm, note that it only converges on 55% of the 191 runs. We now present the performance profile for all algorithms in their damped version, that is making use of the linesearch strategy presented in Section 3.1, on Figure 3. A zoom for π between 1 and 3 is provided in Figure 4. Firstly we observe that the globalization technique significantly improves the robustness of all four presented algorithms as expected. Secondly and most importantly, GSM remains the best algorithm in terms of efficieny and robustness. More precisely, GSM is the best algorithm on more than 60% of the problems and is able to solve more than 80% of the 191 considered problems. From Figure 4, we note also that when GSM is not the best method, it converges within a factor of 2 of the best algorithm for most problems.

The performance profile analysis depends on the number of methods that are being compared. Therefore, we like to present a comparison between BGM and GSM only, as BGM is probably the most widely used method. The significant improvement provided by our method over Broyden's method is illustrated by Figure 5 considering the undamped version of both algorithms. Figure 6 shows the superiority of GSM as well, when both algorithms are globalized using the linesearch strategy.

In this paper, in the context of solving systems of nonlinear equations, we foused on quasi-Newton methods whcih do not use information about the derivative of the system to be solved. We have already shown that GSM is a very competitive derivative-free algorithm. To conclude our numerical experiments, we like to compare our method with an algorithm using derivative information.

We consider a method belonging to the family of inexact Newton methods

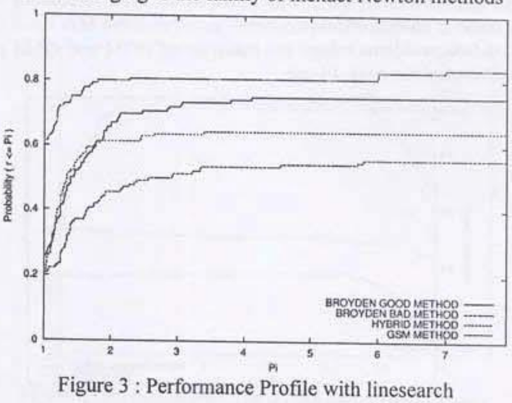


Figure 3 : Performance Profile with linesearch

which identify a direction d_k satisfying the inexact Newton condition:

$$\| F(x_k) + J(x_k)d_k \| \leq \eta_k \| F(x_k) \| \tag{49}$$

for some $\eta_k \in [0,1)$. The most conventional inexact Newton method uses iterative techniques to compute the Newton step d_k using (49) as a stopping criterion. Among these iteratives techniques, Krylov-based linear solvers are generally chosen. Newton-Krylov methods need to estimate Jacobian-vector products using finite differenes approximations in the appropriate Krylov subspace.

We now challenge GSM against the Newton-Krylov method presented by Kelley (2003). The considered version of this method uses the iterative linear GMRES (proposed by Saad and Schultz, 1986) and a parabolic linesearch via three interpolation points. Similarly to the Newton-Krylov algorithm, we allow GSM to use a finite differences approximation of the initial Jacobian. From Figure 7, we observe that GSM is competitive with Newton-Krylov both in terms of efficieny and robustness. This result is very satisfactory as Newton- Krylov methods have been proven to be very efficient methods to solve systems of nonlinear equations.

4.2. Behavior in presence of noise

In practice the evaluation of systems of nonlinear equations often returns a result that is affected by noise, in particular if the evaluation is the outcome of simulator runs. For example Bierlaire and Crittin (forthcoming) describe such a problem in the context of transportation applications. Therefore, we

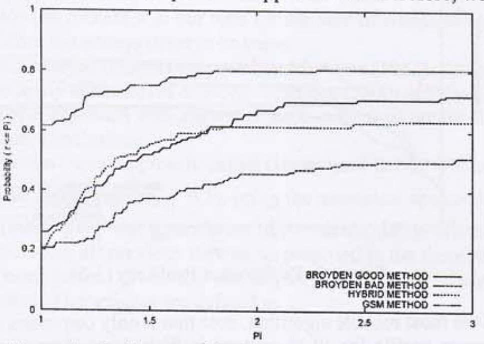


Figure 4 : Performance Profile on (1,3) with linesearch

conlude this section by an empirical analysis of the behavior of our method in the presence of noise in the function. Indeed, we speculate that the use of a larger sample of iterates within a least squares framework smooths the impact of noise on the method.

We consider a random function described by:

$$G(x) = F_s(x) + \varphi(x) \tag{50}$$

Where F_s $R^n \rightarrow R^n$ is deterministic and $\varphi(x)$ is a random perturbation. We want to identify x such that $F_s(x) = 0$, but we are not able to compute $F_s(x)$ accurately.

We consider two types of random noise:

1. Similarly to Choi and Kelley (2000), we first assume that the noise dereases near the solution, more preisely:

$$\phi(x) \sim N(0, \alpha^2 \| x - x^* \|^2) \text{ and } G(x_0) = F_s(x_0) = 0 \tag{51}$$

In this case, the noise is named *proportional.*

2. We then assume that the noise is constant, more preisely:

$$\varphi(x) \sim N(0, \alpha^2). \tag{52}$$

In this case, the noise is named *absolute*.

We have selected two problems where the behavior of BGM and GSM in their undamped version are almost similar in the deterministic case. Please

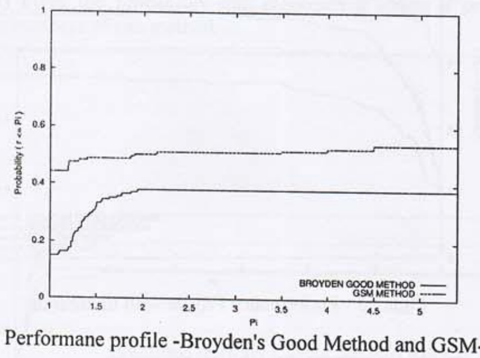


Figure 5: Performane profile -Broyden's Good Method and GSM-

note that we do not perform tests using the damped quasi-Newton framework as the underlying globalization strategy makes use of finite difierenes, whih is not compatible with the stohastiity present in the problems considered in this subsetion. For each function and each type of noise the results are presented for 4 levels of stohastiity, i. e. for four difierent values of the parameter defined in equations (51) and (52). We plot the relative nonlinear residual, that is $\|G(x_k)\|/\|G(x_0)\|$, against the number of function evaluations.

First we consider a problem given by Spediato and Huang (1997) and fully desrcibed in Section 6.4 in the Appendix. The results obtained with the proportional noise are presented in Figure 8. Figure 8(a) illustrates the deterministic case, with $\varphi(x) = 0$, where BGM is slightly better than GSM. When a noise with small variance ($\alpha = 0.001$, Figure 8(b)) is present, GSM decreases the value of the residual pretty quickly, while the descent rate of BGM is muh slower. When the variance of the noise inrceases ($\alpha = 0.01$ in Figure 8(), and $\alpha = 1$ in Figure 8(d)), the BGM is trapped in higher values of the residual, while GSM achieves a significant decrease. The results obtained with the absolute noise are presented in Figure 9. The values of α are the same as above. The behavior of the two methods is almost the same as for the proportional noise. GSM reaches a lower level than BGM of the residual for small ($\alpha = 0.001$, Figure 9(b)) and medium ($\alpha =$

0.01, Figure 9(b)) variances. When the variance is higher ($\alpha = 1$, Figure 9(d)) none of the two methods is able to significantly derease the relative residual.

The same tests have been accomplished with the Extended Rosenbrock Function given by Dennis and Schnabel (1996) and fully described in Secion 6.5 in the Appendix. Figure 10 reports the behavior of GSM and BGM applied to

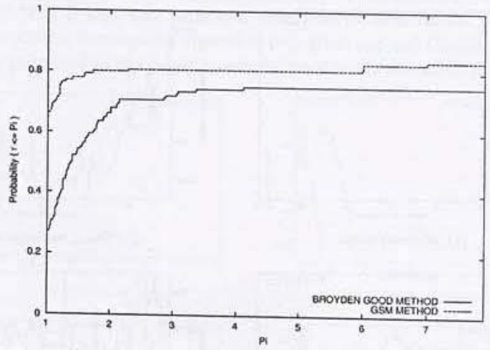


Figure 6: Performane profile with linesearch -Broyden's Good Method and GSM-

this problem perturbated with a proportional noise. Figure 10(a) reports the relative residual of the smooth system ($\alpha = 0$). In the presence of the small noise ($\alpha = 0.0001$, Figure 10(b)) both methods converge but BGM needs more than twice the number of iterations needed by GSM. When the noise increases ($\alpha = 0.01$, Figure 10()) BGM is totally disrcupted and diverges, while GSM still converges in less than 20 iterations. With the higher value of the noise ($\alpha = 1$, Figure 10(a)) both methods are stalled, but GSM achieves lower values for thebrelative residual. Figure 11 reports the behavior of GSM and BGM applied to this problem perturbated with absolute noise. Again Figure 11(a) reports the relative residual of the smooth system ($\alpha = 0$). For small ($\alpha = 0.0001$, Figure 11(b)) and medium ($\alpha = 0.01$, Figure 11(b)) value of the noise both methods reach the same value of relative residual with GSM using clearly less evaluations of F than BGM. With a larger noise ($\alpha = 1$, Figure 11(b)), as for the proportional case, BGM is stalled at a higher value than GSM.

We have performed the same analysis on other problems, and observed a similar behavior, that is a systematially better robustness of GSM compared to the lassiBGM when solving a noisy system of equations.

In summary, our method is more robust than BGM in the sense that it an solve noisy problems that BGM cannot. When both fail, GSM exhibits better decreases, whih may be advantageous in practice.

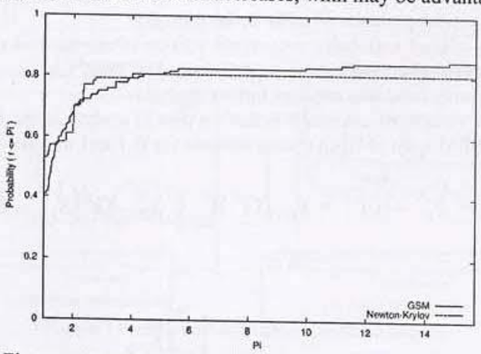


Figure 7: Performane profile -GSM and Newton-Krylo-

4.3. Large-scale problems

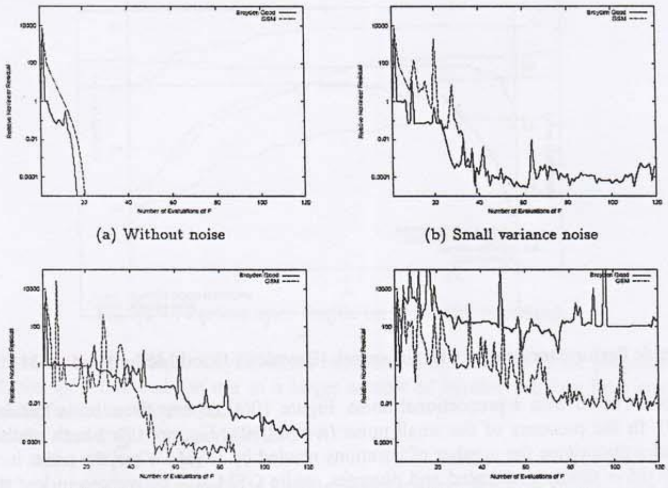
The main drawback of our approach is the relatively high cost in numerical linear algebra. Therefore it is particularly appropriate for medium-sale problems where F is very expensive to compute. Bierlaire and Crittin (forthcoming) propose an instance of this class of methods, designed to solve very large-scale systems of nonlinear equations without any assumption about the structure of the problem. The numerical experiments on standard large-scale problems show similar results: the algorithm outperforms classial large-scale quasi-Newton methods in terms of efficiency and robustness, its numerical performances are similar to the Newton-Krylov methods, and it is robust in presence of noise.

The complexity (both in time and memory) is linear in the size of the problem. Therefore, we were able to solve very large instanes of a problem given by Spediato and Huang (1997). The algorithm has been able to converge on a problem of size 2'000'000in four hours and 158 iterations.

We are strongly interested in globalizing the large-scale version of our method. However, it requires future research to adapt our linesearch framework and to get an efficcient globalization strategy in term of computational time.

5. Conclusion and perspectives

We have proposed a new class of generalized secant methods, based on the use of more than two iterates to identify the secant model. Contrarily to previous attempts for multi-iterate secant methods, the key ideas of this paper are (i) to use a least squares approah instead of an interpolation method to derive



(a) Without noise (b) Small variance noise

(c) Medium variance noise (d) Large variance noise

Figure 8: Behavior with proportional stochasticity

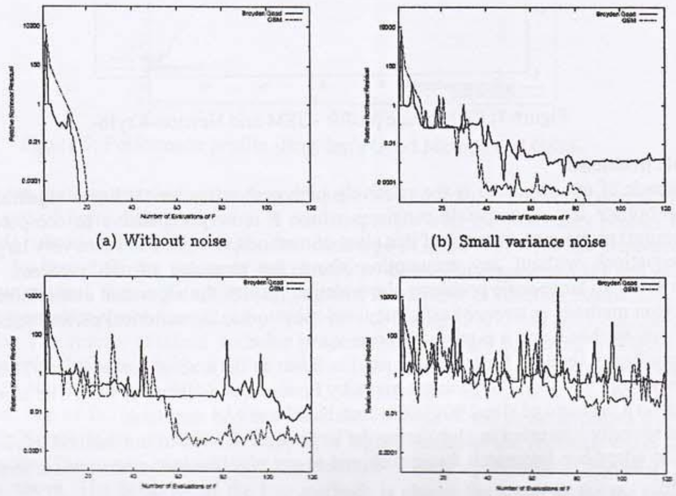
the secant model, and (ii) to explicitly control the numerical stability of the method.

A specific sub-class of this family of methods provides an update formula. We have proven the local convergene of an undamped quasi-Newton method based on this update formula. Moreover, we have performed extensive numerical experiments with several algorithms. The results show that our method produces signifiant improvement in term of robustness and number of function evaluations compared to classial methods. We have also shown that the globalization strategy presented in this paper signifiantly improves the robustness of quasi-Newton methods. Eventually, we have provided preliminary evidences that our method is more robust in the presence of noise in the function.

A theoretical analysis of a globally convergent version of our method must also be performed. We also conjeture that the local convergene rate is super-linear. And most importantly, the general behavior of the algorithm for solving noisy functions requires further analysis.

There are several variants of our methods that we plan to analyze in the future. Firstly, following Broyden's idea to derive BBM from (44), an update formula for B-1 k+1 an easily be derived in the ontext of our method:

$$B_{k+1}^{-1} = B_k^{-1} + \left(\Gamma\Gamma^T + Y_{K+1}\Omega^2 Y_{k+1}^T\right)^{-1} Y_{k+1}^T\, \Omega^2 \left(S_{k+1} - B_k^{-1} Y_{k+1}\right). \qquad (53)$$

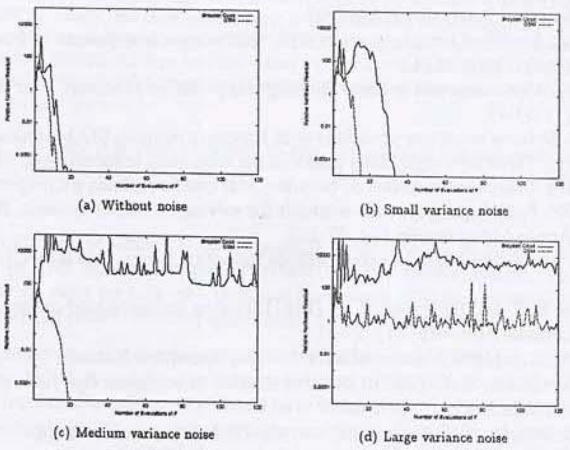


(a) Without noise (b) Small variance noise

(c) Medium variance noise (d) Large variance noise

From preliminary tests that we have performed, the "Good" and "Bad" versions of our method compare in a similar way as BGM and BBM. Seondly, non-update instanes of our lass of methods an be onsidered. In that case, the arbitrary matrix B0 k+1 in (10) may be diferent from Bk. Choosing a matrix independent from k allows to use iterative sheme designed to solve large- sale least squares. In that ase, hoosing a matrix independent from k would allow to apply Kalman filtering (Kalman, 1960) to incrementally solve (10) and, consequently, improve the numerical efficiency of the method. For large scale problems, an iterative scheme such as LSQR (Paige and Saunders, 1982) an be considered. LSQR an also improve the efficiency of Kalman filter for the inrcemental algorithm (see Bierlaire and Crittin, 2004).

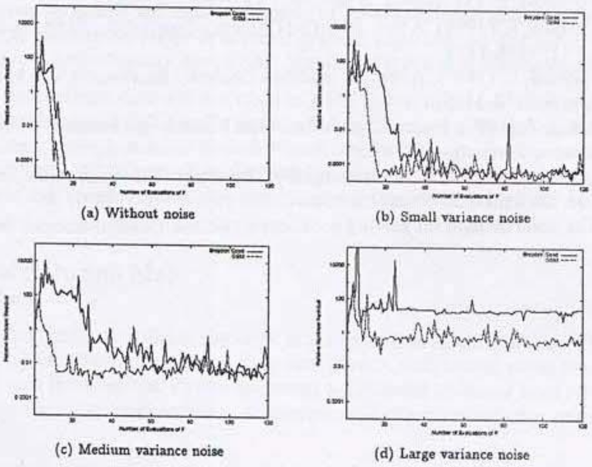
Finally, the ideas proposed in this paper can be tailored to optimization problems.



(a) Without noise (b) Small variance noise

(c) Medium variance noise (d) Large variance noise

Figure 10: Behavior with proportional stochasticity



(a) Without noise (b) Small variance noise

(c) Medium variance noise (d) Large variance noise

Figure 11: Behavior with absolute stochasticity

References

Bierlaire, M. and Crittin, F. (2004). An efficency algorithm for real-time estimation and predition of dynamiOD tables, Operations Researh 52(1): 116-127.

Bierlaire, M. and Crittin, F. (forthoming). Solving noisy large sale fixed point problems and systems of nonlinear equations, Transportation Siene.

Broyden, C. G. (1965). A lass of methods for solving nonlinear simultaneous equations, Mathematis of Computation 19: 577-593.

Broyden, C. G., Dennis, J. E. and Morfie, J. J. (1973). On the local and super-linear convergene of quasi-Newton methods, Journal of the Institute of Mathematis and its Appliations 12: 233-246.

Choi, T. D. and Kelley, C. (2000). Superlinear convergene and impliit filtering, SIAM Journal of Optimization 10.

Dennis, J. E. and Schnabel, R. B. (1996). Numerical methods for unonstrained optimization and nonlinear equations, Soiety for Industrial and Applied Mathematis (SIAM), Philadelphia, Pa.

Dolan, E. D. and Mor~e, J. J. (2002). Benhmarking optimization software with performane profiles, Mathematial Programming, Serie A 91: 201-213.

Eaton, J. W. (1997). GNU Otave: a high level interative language for numerical computations, www.octave.org.

Eyert, V. (1996). A comparative study on methods for convergene aeleration of iterative vector sequenes, Journal of computational physis 124: 271-285.
Ford, J. A. (1999). A survey of multi-step quasi-Newton methods, Proeedings of the International Conferene on Sientifi Computations, Beirut, Lebanon.
Ford, J. A. and Moghrabi, I. A. (1997). Alternating multi-step quasi-Newton methods for unonstrained optimization, Journal of Computational and Applied Mathematis 82: 105{116.
Gould, N. I. M., Leyfier, S. and Toint, P. L. (2005). A multidimensional filter algorithm for nonlinear equations and nonlineat least-squares, SIAM Journal on Optimization 15(1):17-38.
Gragg, W. and Stewart, G. (1976). A stable variant of the seant method for solving nonlinear equations, SIAM Journal on Numerical Analysis 13: 889{903.
Johnson, D. D. (1988). Modified Broyden's method for aelerating convergene in self-consistent alulations, Physial Review B 38(18): 12807-12813.
Kalman, R. E. (1960). A new approah to linear filtering and predition problems, J. of BasiEng., Trans. ASME, Series D 82(1): 33-45.
Kelley, C. T. (2003). Solving nonlinear equations with Newton's method, SIAM, Philadelphia, Pa.
Martinez, J. M. (1979). Three new algorithms based on the sequantial seant method, BIT 19: 236-243.
Martinez, J. M. (1982). Sobre dois metodos de broyden, Mat ematia Apliada e Computational 1.
Martinez, J. M. (2000). Pratial quasi-Newton methods for solving nonlinear systems, Journal of Computational and Applied Mathematis 124: 97-122.
Moghrabi, I. (1993). Multi-step quasi-Newton methods for optimization, PhD thesis, University of Essex, United Kingdom.
Mor~e, J. J., Garbow, B. S. and Hillstrom, K. E. (1981). Testing unconstrained optimization software, ACM Transations on Mathematial Software 7(1): 17-41.
Noedal, J. and Wright, S. J. (1999). Numerial optimization, Operations Researh, Springer Verlag, New-York
Ortega, J. M. and Rheinboldt, W. C. (1970). Iterative solution of nonlinear equations in several variables, AademiPress, New York.
Paige, C. C. and Saunders, M. A. (1982). LSQR : an algorithm for sparse linear equations and parse least squares, ACM Transations on Mathematial Software 8: 43-71.
Saad, Y. and Shultz, M. (1986). GMRES: A generalized minimal residual algorithm for solving nonsymmetrilinear systems, SIAM Journal on Numerial Analysis 7: 856-869.
Shnabel, R. B. and Eskow, E. (1991). A new modified Cholesky fatorization, SIAM Journal on Sientifiand Statistial Computing 11: 1136-1158.
Shnabel, R. B. and Eskow, E. (1999). A revised modified Cholesky fatorization, SIAM Journal on Optimization 9: 1135-1148.
Spediato, E. and Huang, Z. (1997). Numerial experiene with Newton-like methods for nonlinear algebraisystems, Computing 58: 69-99.
Vanderbilt, D. and Louie, S. G. (1984). Total energies of diamond (111) surfae reonstrutions by a linear ombination of atomiorbitals method, Physial Review B 30(10): 6118-6130.
Wolfe, P. (1959). The seant method for solving nonlinear equations, Commu-niations ACM 12: 12-13.