Development of a Nonlinear Equations Solver with Superlinear Convergence at Regular Singularities.

DISSERTATION

zur Erlangung des akademischen Grades
doktor rerum naturalium.
(Dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von
Dipl.-Math. Suleiman Alabdallah

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekander Mathematisch-Naturwissenschaftlichen FakultätII:
Prof. Dr. Elmar Kulke

Gutachter:

1.Prof.Dr.Andreas Griewank.

2.Prof.Dr.Stephen J. Wright.

3.Prof.Dr.Natasa Krejic.

eingereicht am: 28.03.2014
Tag der mündlichen Prüfung: 04.07.2014

Development of a Nonlinear Equations Solver with Superlinear Convergence to Regular Singularities

Suleiman Alabdallah
Supervisor: Prof. Andreas Griewank

July 04, 2014

Acknowledgments

First, I would like to thank my supervisor Professor Dr. Andreas Griewank for his patience and dedication. His guidance helped me in all the time of research and writing of this thesis. This work have been impossible without his support.

In addition, I wish to thank Dr. Lutz Lehmann, who helped me by editing the code and brought my attention some properties of the polynomials and norms.

I want also to thank Dipl. Math. Trosten Bosse for the last day changes.

I have greatly benefited in the beginning from discussions with Dr. Adel Hamdi to understand the line-search procedure.

A special thanks to my big family in Syria for the their support, they provided me through my entire life. Words cannot express how grateful I am to my mother and father. Their prayer for me was what sustained me thus far.

Finally. My gratitude and love to my wife *Farida*, my children *Besher*, *Zain*, and the loved one *Amr*, without their patience and encouragement, I would not have finished this thesis.

Contents

List of Figures

List of Tables

Zusammenfassung

In dieser Arbeit präsentieren wir eine neue Art von Newton-Verfahren mit Liniensuche, basierend auf Interpolation im Bildbereich nach Wedin et al. [LW84]. Von dem resultierenden stabilisierten Newton-Algorithmus wird theoretisch und praktisch gezeigt, dass er effizient ist im Falle von nichtsingulären Lösungen. Darüber hinaus wird beobachtet, dass er eine superlineare Rate von Konvergenz bei einfachen Singularitäten erhält. Hingegen ist vom Newton-Verfahren ohne Liniensuche bekannt, dass es nur linear von fast allen Punkten in der Nähe einer singulären Lösung konvergiert. In Hinsicht auf Anwendungen auf Komplementaritätsprobleme betrachten wir auch Systeme, deren Jacobimatrix nicht differenzierbar sondern nur semismooth ist. Auch hier erreicht unser stabilisiertes und beschleunigtes Newton-Verfahren Superlinearität bei einfachen Singularitäten.

Abstract

In this thesis we present a new type of line-search for Newton's method, based on range space interpolation as suggested by Wedin et al. [LW84]. The resulting stabilized Newton algorithm is theoretically and practically shown to be efficient in the case of nonsingular roots. Moreover it is observed that it maintains a superlinear rate of convergence at simple singularities. Whereas Newton's method without line-search is known to converge only linearly from almost all points near the singular root. In view of applications to complementarity problems we also consider systems, whose Jacobian is not differentiable but only semismooth. Again, our stabilized and accelerated Newton's method achieves superlinearity at simple singularities.

Chapter 1

Introduction

Systems of nonlinear algebraic equations arise in many fields of applications with the number of variables ranging from just a handful to thousands or even millions. It is well understood that fast local convergence at a superlinear rate can only be achieved by Newton-like methods. The usual conditions for thus are that the Jacobian formed by the first partial derivatives is nonsingular at the nearest root and varies continuously in a surrounding neighborhood. In particular its Lipschitz continuity implies the quadratic convergence of Newton's method.

In this thesis we consider the situation where the nonsingularity condition is exactly or nearly violated. Such a situation arises at or near turning points with respect to a(nearly) critical parameter. It also occurs through so-called nonlinear complementarity problem functions that characterize complementarity. Then there is also the additional complication that the resulting Jacobian is Lipschitz continuous but not properly differentiable. In the smooth case, i.e., twice continuously scenario the behavior of Newton's method has been extensively studied in the seventies and eighties of the last century, see [Red78], [Red79], [DK80a], [Gri80b], [Gri80a], [GO81], [Gri85].

At so called regular singularities of order $k \in N$ one finds that Newton's method converges with the linear rate $\frac{k}{k+1}$ from most starting points near a singular root. More specifically the set of suitable starting points forms a starlike domain of density 1 with respect to the singular root. At irregular singularities Newton's method may converge at some other linear rate depending on the size of higher mixed derivatives or, if these are too large, it may even be repelled by the root, see [GO81].

Wright et al. [OW09] have recently considered equation systems arising from complementarity problems where the Jacobian may be both singular and nondifferentiable. Under the mild assumption that the Jacobian is semismooth at a regular order singularity, they showed that Newton's method still converges with the linear rate $\frac{1}{2}$ from within a starlike domain of positive density. They also suggested and tested the acceleration of Newton's method by nearly doubling every other step, when the halving pattern of the step size seem to have settled in. Variations of this two-step scheme had been analysed by Kelly et al. and Griewank in the smooth case. There is also a three-step scheme where every third step is exactly doubled. It achieves a three step quadratic rate under certain stronger assumptions on the first order singularity.

The main drawback of the acceleration schemes mentioned above is that the algorithm or the user has to decide somehow, that the convergence pattern indicates a singularity and then modify the step multipliers accordingly. It may then also turn out in the end that the solution is only nearly singular so that the decision needs to be reversed at some later stage.

Instead we pursue an acceleration approach that is based on the range space line-search of Wedin et al. [LW84]. It makes complete sense in the nonsingular case and automatically generates the two-step pattern in simple singular cases. At least that is our observation in the first order regular case, even when the Jacobian is only semi-smooth.

Unfortunately, we have not been able to establish this result analytically. In nearly singular situations, which arise in the vicinity of turning points we observe a significant acceleration in an intermediate phase before the method reduces to full-step Newton in the immediate vicinity of the root.

The thesis is organized as follows. In Chapter 2 we discuss the behavior of Newton's method at nonsingular solutions. We analyse the local and semilocal convergence in the case that the Jacobian F' is nonsingular at a root $x_* \in F^{-1}(0)$. Locally, quadratic convergence of Newton's method is obtained. Kantorovich theorem is used to obtain quadratic convergence under semilocal conditions.

The concept of regular singularity is introduced and the regularity condition is derived in the first section of Chapter 3. In Section 2 of Chapter 3 we present convergence results of Newton's method at singular roots when the function F is continuously differentiable to a sufficient order.

Then applications of NLEs reformulations of NCP are discussed where at a singularity the Jacobian F' is not Fréchet differentiable but strongly semismooth.

In Chapter 4 we introduce the acceleration of Newton's method at singular solutions. Here we discuss the two and three step method. Those techniques require three times differentiability of F and starting points whose errors remain in a wedge around the nullspace N of $F'(x_*)$.

The two and three point schemes require deciding that the problem is singular, but our parabolic line-search method, which is discussed in Chapter 5 can achieve 2-step superlinear convergence for simple singularity automatically i.e., without being told that a solution is singular. Experimentally we have shown that line-search is effective in both singular and nonsingular cases.

We also briefly discuss an extension based on cubic rather than quadratic interpolation, which did not fulfill our expectations.

Numerical results and applications to nonlinear complementary problem are considered in Chapter 6. To illustrate the behavior of the line-search, the so-called multiplier mountain is analysed. For large dimensions we tested Bratu problem.

We obtain Newton fractals for some problems from the literature and compare it with our line-search Newton's method.

Chapter 2

Newton-like Methods

In this section, we discuss briefly the behavior of Newton's method and analyse its local and semilocal convergence, in the case that the inverse of the derivative of the function exists. We start by discussing the behavior of Newton's method geometrically for scalar functions $f : \mathbb{R} \to \mathbb{R}$ and then we generalize it to multivariate functions $F : \mathbb{R}^n \to \mathbb{R}^n$. The distinguishing feature of Newton's method is that if $f'(x)$ is Lipschitz continuous in a neighborhood containing the root x_* and $f'(x_*)$ is nonzero, then the sequence of Newton iterates converges locally and quadratically to x_*. This means there exist $\delta > 0$ and $c \geq 0$ such that the sequence of Newton iterates $\{x_j\}$ produced by Newton's method obeys

$$\|x_{j+1} - x_*\| \leq c\,\|x_j - x_*\|^2, \quad if \quad \|x_0 - x_*\| \leq \delta.$$

Locally we prove quadratic convergence of Newton's method when we use the exact derivative $F' : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ of the function $F : \mathbb{R}^n \to \mathbb{R}^n$. If we approximate the derivative F' by some matrix then we will get linear or superlinear convergence of Newton-like methods.

As a semilocal convergence results we consider the Kantorovich theorem, which imposes conditions on the initial point to establish the existence of a nearby root, which is found by Newton's method. If those conditions are satisfied then a solution exists close to the initial point x_0 and the sequence converges quadratically to that point.

2.0.1 Idea of Newton's Method

Assume we have to solve a scalar equation

$$f(x) = 0.$$

with an appropriate guess x_0 of the unknown solution x_* at hand.

We will use the perturbation

$$\hat{x} = x - x_* \text{ and } \hat{x}_j = x_j - x_*, \quad j = 0, 1, 2, \ldots \tag{2.0.1}$$

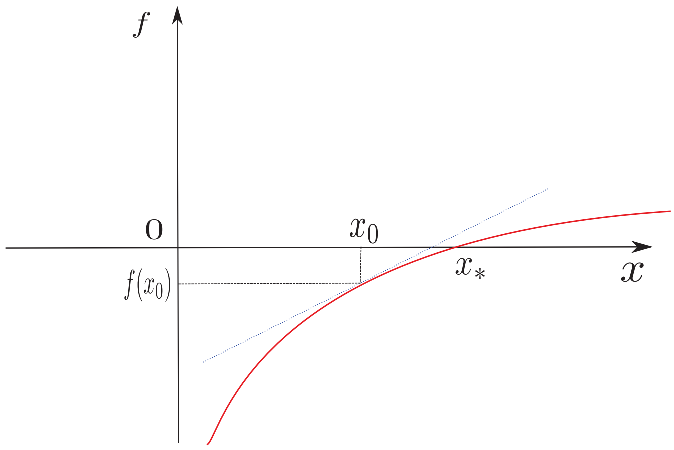


Figure 2.1: Newton's method for a scalar equation

which denote the difference between the sequence $\{x_j\}_{j\geq 0}$ and x_*.

By Taylor's expansion we have

$$0 = f(x_*) = f(x_0 - \hat{x}_0) = f(x_0) - f'(x_0)\hat{x}_0 + O(|\hat{x}_0|^2).$$

If we drop terms of order higher than linear in the perturbation, we arrive at the approximate equation

$$f'(x_0)\hat{x}_0 \approx f(x_0),$$

which, assuming $f'(x_0) \neq 0$, leads to the precise equation

$$x_1 - x_0 = -\frac{f(x_0)}{f'(x_0)},$$

for a first correction of the starting guess. From this, an iterative scheme is constructed by repetition

$$x_{j+1} = x_j - \frac{f(x_j)}{f'(x_j)}, \quad j = 0, 1, \ldots,$$

which is called Newton's method in the scalar case.

2.0.2 Geometric Interpolation of Newton's method

Looking at the Graph 2.1 of $f(x)$ any root can be interpreted as the intersection of this graph with the real axis. The graph of $f(x)$ is replaced by its tangent $p(x) = f(x_0) + f'(x_0)(x - x_0)$ in x_0 and the first iterate x_1 is considered as the intersection of the tangent with the real axis. Upon repeating this geometric process, the close-by solution point x_* can be approximated up to any desired accuracy. By geometric insight, the iterative process will converge globally for convex or concave f, which includes the case of arbitrarily bad initial guesses as well. The geometric derivation seems to be restricted to the scalar case. A careful examination of the subject in more than one dimension leads to a topological path called *Newton path* (see [Deu04]).

Now we consider to the general case of Newton's method.

2.1 Newton-like Methods for Nonlinear Equations

We will make standard assumptions on the function F for local convergence.

Definition 2.1.1 *(The standard assumptions)*

Let

$$F(x) = 0 \tag{2.1.1}$$

be a nonlinear equation, where $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$*, has a root* x_**. Let* F *be differentiable and non-singular for all* x *in a neighborhood* $D \subset \mathbb{R}^n$ *of* x_* *and let* F' *be Lipschitz continuous with Lipschitz constant* L*. i.e., for all* $x, y \in D$ *and the constant* $L > 0$

$$\|F'(x) - F'(y)\| \leq L\|x - y\|, \tag{2.1.2}$$

where $\|.\|$ *denote to Euclidean norms for vectors and the induced spectral norm for matrices.*

Suppose we have a starting guess $x_0 \in \mathbb{R}^n$ for an unknown solution x_* at hand. We are interested in solving the nonlinear equation (2.1.1) with Newton's method.

Definition 2.1.2 *Given* $F : D \subset \mathbb{R}^n \to \mathbb{R}^n$ *and* $x_0 \in \mathbb{R}^n$*, Newton's method computes iteratively the sequence*

$$x_{j+1} = x_j + d_j. \tag{2.1.3}$$

Where

$$d_j := d(x_j) := -F'(x_j)^{-1}F(x_j). \tag{2.1.4}$$

is called Newton direction.

There is a large variety of Newton-type methods, which will be named and briefly sketched here.

Definition 2.1.3 *(Exact Newton's method)*
Any of the finite dimensional Newton-type methods requires the numerical solution of the linear equations

$$F'(x_j)d_j = -F(x_j), \tag{2.1.5}$$

Whenever direct elimination methods (like Gaussian elimination, QR decomposition,...) are applicable, we speak of the exact Newton's methods.

Definition 2.1.4 *(Newton-like method)*
This type of Newton's method is characterized by the fact that, in finite dimension, the Jacobian matrices are either replaced by some fixed 'close by' $F'(z)$ *with* $z \approx x_0$*, or by some other approximation* $H_j \approx F'(x_j)$ *and one computes*

$$H_j d_j = -F(x_j), \quad x_{j+1} = x_j + d_j, \quad j = 1, 2, \ldots, \tag{2.1.6}$$

Definition 2.1.5 *(Simplified Newton's method)*
This variant of Newton's method is characterized by keeping the initial derivative throughout the whole iteration:

$$F'(x_0)d_j = -F(x_j), \quad x_{j+1} = x_j + d_j, \quad j = 1, 2, \dots, \tag{2.1.7}$$

This method is also called chord method.

Definition 2.1.6 *(Inexact Newton's method)*
For extremely large scale nonlinear problems the arising linear systems for the Newton corrections can no longer be solved directly ('exactly'), but must be solved iteratively ('inexactly') which gives the name inexact Newton's methods. The whole scheme then consists of an inner iteration (at Newton step j)

$$F'(x_j)d_j^q = -F(x_j) + r^q, \quad j = 1, 2, \dots,$$

$$x_{j+1}^q = x_j + d_j^q, \quad q = 1, 2, \dots, q_j^{max}$$

in terms of residuals r_j^q and an outer iteration where, given x_0, the iterates are defined as

$$x_{j+1}^q = x_{j+1} \quad for \quad q = q_j^{max} \quad j = 1, 2, \dots,$$

Inexact Newton's method are sometimes also called truncated Newton's method.

Although Newton's method is theoretically attractive, it may be difficult to use in practice. In fact, each step requires the solution of the linear system (2.1.5), even though the inverse $F'(x_j)^{-1}$ is rarely computed explicitly[OR70].

A necessary assumption for the solvability of the above linear problems (2.1.5) is that the Jacobians $F'(x)$ are invertible at all occurring arguments. For this reason, standard convergence theorems typically require a prior that the inverse $F'(x)^{-1}$ exists and is bounded

$$\|F'(x)^{-1}\| \leq \beta < \infty, \quad x \in D. \tag{2.1.8}$$

From a computational point of view, such a theoretical quantity β defined over the domain D seems to be hard to get, apart from rather simple situations. Sampling of local estimates like

$$\|F'(x_0)^{-1}\| \leq \beta_0, \tag{2.1.9}$$

seems to be preferable, but is still quite expensive.

The following lemma is often referred to as the Banach lemma.

Lemma 2.1.7 *(Banach Perturbation Lemma)*

Let A, $C \in \mathbb{R}^{n\times n}$. Assume that A is invertible and $\|A^{-1}\| \leq \alpha$. If $\|A - C\| \leq \beta$ and $\alpha\beta < 1$, then C is also invertible and $\|C^{-1}\| \leq \frac{\alpha}{1-\alpha\beta}$.

Proof: See Ortega and Rheinboldt [OR70]. □

In order to study the convergence properties of the above Newton iterations, some information on the variability of the Jacobian is needed, as already stated in the scalar equation case. The standard form to include this information is via a Lipschitz condition (2.1.2)

With this additional assumption, Banach perturbation Lemma 2.1.7 implies the existence of some upper bound β such that

$$\|F'(x)^{-1}\| \leq \beta \leq \frac{\beta_0}{1 - L\beta_0\|x - x_0\|}$$

for

$$\|x - x_0\| \leq \frac{1}{L\beta_0}, \quad x \in D.$$

Classical convergence theorems for Newton's method use certain combinations of these assumptions. The first classical convergence theorems for Newton's method is Newton-Kantorovich theorem (see [OR70], [Ort68]). It requires assumptions (2.1.9) and (2.1.2) to show existence and uniqueness of a solution x_* as well a quadratic convergence of Newton iterates from within a neighborhood characterized by a so-called Kantorovich quantity $h_0 := \|\hat{x}_0\|\beta_0 L < \frac{1}{2}$ and a corresponding convergence ball around x_0 with radius $\rho_0 \sim \frac{1}{L\beta_0}$.

The second classical convergence theorems for Newton's method is Newton-Mysovskikh theorem (see [OR70]) requires assumptions (2.1.8) and (2.1.2) to show uniqueness (not existence!) of a solution x_* and quadratic convergence within a neighborhood characterized by the slightly different quantity $h_0 := \|\hat{x}_0\|\beta L < 2$ and a corresponding convergence ball around x_0 with radius $\rho_0 \sim \frac{1}{L\beta}$, [Deu04].

2.1.1 Local Convergence Analysis

In this section we are going to analyse the local convergence properties of Newton's method. In this case we must choose x_0 to be sufficiently close to x_*. But first, we bring the following well-known propositions to mind.

We may express the fundamental theorem of calculus as follows.

Theorem 2.1.8 *Let F be differentiable in an open set $D \subset \mathbb{R}^n$ and let $x_* \in D$. Then for all $x \in D$ sufficiently near x_* we have*

$$F(x) - F(x_*) = \int_0^1 F'(x_* + t\,\hat{x})\hat{x}\,dt.$$

The next lemma is an important consequence of the standard Assumptions 2.1.1.

Lemma 2.1.9 *Assume that the standard Assumptions 2.1.1 hold, then there is $\delta > 0$ so that for all $x \in B_\delta$*

$$\|F'(x)\| \leq 2\|F'(x_*)\|, \tag{2.1.10}$$
$$\|F'(x)^{-1}\| \leq 2\|F'(x_*)^{-1}\|, \tag{2.1.11}$$
$$\|F'(x_*)^{-1}\|^{-1}\|\hat{x}\|/2 \leq \|F(x)\| \leq 2\|F'(x_*)\|\|\hat{x}\|. \tag{2.1.12}$$

Here B_δ denote the ball of radius δ about x_*, i.e.

$$B_\delta = \{x \ : \ \|\hat{x}\| < \delta\}.$$

Proof: See Kelley [Kel95]. $\square$

The convergence result on Newton's method follows from Lemma2.1.9.

Lemma 2.1.10 *Let the standard Assumptions 2.1.1 hold, then there are* $L_0 > 0$ *and* $\delta > 0$ *such that if* $x_c \in B_\delta$ *the Newton iterate from* x_c *given by* $x_+ = x_c - F'(x_c)^{-1}F(x_c)$ *satisfies*

$$\|\hat{x}_+\| \leq L_0\|\hat{x}_c\|^2 \tag{2.1.13}$$

Proof: Let $\delta > 0$ be small enough so that the conclusions of Lemma 2.1.9 hold. By Theorem 2.1.8 First we note that

$$F'(x_*)^{-1}F(x) = F'(x_*)^{-1}\int_0^1 F'(x_*+t\,\hat{x}_*)\,\hat{x}_*\,dt = \hat{x}_* - \int_0^1 [I - F'(x_*)^{-1}F'(x_*+t\,\hat{x}_*)]\,\hat{x}_*\,dt$$

then

$$\hat{x}_+ = \hat{x}_c - F'(x_c)^{-1}F(x_c) = F'(x_c)^{-1}\int_0^1 [F'(x_c) - F'(x_* + t\,\hat{x}_c)]\,\hat{x}_c\,dt.$$

By Lemma 2.1.9 and the Lipschitz continuity of F' yields

$$\|\hat{x}_+\| \leq 2\|F'(x_*)\|\|\hat{x}_c\|\int_0^1 (Lt\|\hat{x}_c\|)dt = L\|F'(x_*)^{-1}\|\|\hat{x}_c\|^2$$

This proves the relation (2.1.13) by taking $L_0 = L\,\|F'(x_*)^{-1}\|$. The proof of convergence of the full-step Newton iteration will be complete if we reduce δ if needed so that $L_0\delta < 1$. $\square$

Now we can prove the local convergence of Newton iteration given by (2.1.3).

Theorem 2.1.11 *(Local Convergence)*

Let the standard Assumptions 2.1.1 hold. Then there is δ *such that if* $x_0 \in B_\delta$ *the Newton iteration given by (2.1.3) converges q-quadratically to* x_*.

Proof: Let δ be small enough so that the conclusions of Lemma 2.1.10 hold. Reduce δ if needed so that $L_0\delta = \epsilon < 1$. Then if $j \geq 0$ and $x_j \in B_\delta$, then Lemma 2.1.10 implies that

$$\|\hat{x}_{j+1}\| \leq L_0\|\hat{x}_j\|^2 \leq \epsilon\|\hat{x}_j\| \leq \|\hat{x}_j\| \tag{2.1.14}$$

and hence $x_{j+1} \in B_{\epsilon\delta} \subset B_\delta$. Therefore if $x_{j+1} \in B_\delta$ we may continue the iteration. Since $x_0 \in B_\delta$ by assumption, the entire sequence $\{x_j\} \in B_\delta$. (2.1.14) then implies that $x_j \to x_*$ q-quadratically see [Kel95]. $\square$

We present now the linear convergence results for Newton-like methods in the next theorem.

Theorem 2.1.12 *(Linear Convergence [Kel95])*

Let the standard Assumptions 2.1.1 hold. Then there are $K_H > 0$, $\delta > 0$, *and* $\delta_1 > 0$ *such that if* $x_0 \in B_\delta$ *and the matrix* $H(x)$ *satisfies*

$$\|I - H(x)F'(x_*)\| = \sigma(x) \leq \delta_1, \tag{2.1.15}$$

for all $x \in B_\delta$ then the iteration $x_{j+1} = x_j - H(x_j)F(x_j)$ converges q-linearly to x_* and

$$\|\hat{x}_{j+1}\| \leq K_H[\sigma(x_j) + \|\hat{x}_j\|]\|\hat{x}_j\|.$$

Proof: *By (2.1.15) we have*

$$\|H(x)\| = \|H(x)F'(x_*)F'(x_*)^{-1}\| \leq \|H(x)F'(x_*)\|\|F'(x_*)^{-1}\| \tag{2.1.16}$$
$$\leq M_H = (1+\delta_1)\|F'(x_*)^{-1}\|. \tag{2.1.17}$$

Using the above equality and

$$\hat{x}_+ = \hat{x}_c - H(x_c)F(x_c) = \int_0^1 [I - H(x_c)F'(x_* + t\hat{x}_c)]\hat{x}_c dt$$
$$= [I - H(x_c)F'(x_*)]\hat{x}_c + H(x_c)\int_0^1 [F'(x_*) - F'(x_* + t\hat{x}_c)]\hat{x}_c dt.$$

we have

$$\|\hat{x}_+\| \leq \sigma(x_c)\|\hat{x}_c\| + \frac{M_H L}{2}\|\hat{x}_c\|^2.$$

This completes the proof with $K_H = 1 + \frac{M_H L}{2}$.

□

2.1.2 Semilocal Convergence Analysis

In the following section we are going to introduce Kantorovich's theorem which established semi-local convergence of Newton's method. Here we are free to choose x_0.

2.1.3 Newton-Kantorovich Theorem

Kantorovich's theorem asserts that the iterative method of Newton, applied to a general system of nonlinear equations $F(x) = 0$, converges to a solution x_*, provided the Jacobian of the system satisfies a Lipschitz condition near x_0 and its inverse at x_0 satisfies certain boundedness conditions. The theorem also gives computable error bounds for the iterates. The system of equations takes the form $F(x) = 0$. The Jacobian of F at $x_0 \in \mathbb{R}^n$ is the Fréchet derivative $F'(x_0)$. It is assumed that F is defined and has a Fréchet differential at each point of a given convex open set $D_0 \subset \mathbb{R}^n$. Kantorovich has given two basically different proofs of this result using recurrence relations [Kan48] or majority functions[LVK59].

Now we give a proof which is a modification of the second approach and is easier to understand and present [Ort68] .

Theorem 2.1.13 *(Newton-Kantorovich Theorem)*

Let F be defined as in (2.1.1) and $F : D \subset \mathbb{R}^n \to \mathbb{R}^n$ is F-differentiable on a convex set $D_0 \subset D$, and assume for an operator norm that (2.1.2) is satisfied *and* $\forall x, y \in D_0$. *Assume for some* $x_0 \in D_0$ *that* $\Gamma = F'(x_0)^{-1}$ *exists and that* $\alpha = L\beta\eta \leq \frac{1}{2}$ *where* $\|\Gamma\| \leq \beta, \quad \|\Gamma F(x_0)\| \leq \eta$. *Set*

$$t_* = (L\beta)^{-1}[1 - (1-2\alpha)^{\frac{1}{2}}], \qquad t^{**} = (L\beta)^{-1}[1 + (1-2\alpha)^{\frac{1}{2}}], \tag{2.1.18}$$

and suppose that $B(x_0, t_*) = \{x \mid \|x - x_*\| \leq t_*\} \subset D_0$. *Then the Newton iterates given by (2.1.3) are well-defined, remain in* $B(x_0, t_*)$, *and converge to a solution* $x_* \in B(x_0, t_*)$ *of* $F(x_*) = 0$ *which is unique in* $B(x_0, t^{**}) \cap D_0$. *Moreover, if* $\alpha < \frac{1}{2}$ *the order of convergence is quadratic.*

The proof will be an easy consequence of the following lemmas which serve to isolate the essential points.

Lemma 2.1.14 *Let* $\{y_j\}$ *be a sequence in* $\mathbb{R}^n$ *and* $\{t_j\}$ *a sequence of nonnegative real numbers such that*

$$\|y_{j+1} - y_j\| \leq (t_{j+1} - t_j) \quad j = 0, 1, \ldots, \tag{2.1.19}$$

and $t_j \to t_* < \infty$. *Then there exists a* $y_* \in \mathbb{R}^n$ *such that* $y_j \to y_*$ *and*

$$\|y_* - y_j\| \leq (t_* - t_j) \quad j = 0, 1, \ldots, \tag{2.1.20}$$

Proof: The proof is immediate from

$$\|y_{j+p} - y_j\| \leq \sum_{i=1}^{p} \|y_{j+i} - y_{j+i-1}\| \leq (t_{j+p} - t_j) \leq (t_* - t_j)$$

which shows that $\{y_j\}$ is a Cauchy sequence. $\square$

Definition 2.1.15 *We shall say that* $\{t_j\}$ *majorizes* $\{y_j\}$ *if (2.1.19) holds.*

In the following two lemmas the relevant assumptions of the theorem are assumed to hold.

Lemma 2.1.16 *For all* $x \in Q \cap D_0, \quad F'(x)^{-1}$ *is defined on all of* $\mathbb{R}^n$ *and*

$$\|F'(x)^{-1}\| \leq \frac{\beta}{1 - L\beta\|x - x_0\|}.$$

Where $Q = \{x \mid \|x - x_0\| < \frac{1}{L\beta}\}$. *If* x *and* $N(x) = x - F'(x)^{-1}F(x)$ *are in* Q, *then*

$$\|N(N(x)) - N(x)\| \leq \frac{1}{2} \; \frac{L\beta\|x - N(x)\|^2}{1 - L\beta\|x_0 - N(x)\|}.$$

Proof: The first statement follows from Banach Lemma 2.1.7. To prove second statement we note that, since $F(x) + F'(x)(N(x) - x) = 0$, we find

$$\begin{aligned}\|N(N(x)) - N(x)\| &= \|F'(N(x))^{-1}F(N(x))\| \\ &\leq \frac{\beta}{1 - L\beta\|x_0 - N(x)\|}\|F(N(x)) - F(x) - F'(x)(N(x) - x)\|\end{aligned}$$

and the result follows by use of the mean value Theorem [LVK59]:

$$\begin{aligned}\|F(y) - F(x) - F'(x)(y - x)\| &= \| \int_0^1 [F'(\theta y + (1 - \theta)x) - F'(x)](y - x)d\theta\| \\ &\leq \frac{L}{2}\|y - x\|^2.\end{aligned}$$

□

Lemma 2.1.17 *The Newton sequence $\{x_j\}$ is well-defined and is majorized by the sequence defined by*

$$t_{j+1} = t_j - \frac{(L\beta/2)t_j^2 - t_j + \eta}{L\beta t_j - 1}, \quad j = 0, 1, \ldots, \quad t_0 = 0.$$

Moreover, $t_j \to t_*$, where t_* is defined by (2.1.18).

Proof: See [Ort68], [LVK59]. □

The proof of Newton-Kantorovich theorem now is ready to do based on last three lemmas.

Proof: (Newton-Kantorovich Theorem 2.1.13)

The Lemmas 2.1.14 and 2.1.17 show that there exists an $x_j \in B(x_0, t_*)$ such that $x_j \to x_*$. That x_* is a solution follows in the usual way from

$$\begin{aligned}\|F(x_j)\| &= \|F'(x_j)(x_{j+1} - x_j)\| \\ &\leq [\ \|F(x_0)\| + \|F'(x_0) - F'(x_j)\|\ ]\ \|x_j - x_{j+1}\| \\ &\leq [\ \|F'(x_0)\| + Lt_*\ ]\ \|x_j - x_{j+1}\| \to 0\end{aligned}$$

and the continuity of F in $B(x_0, t_*)$. If $\alpha < \frac{1}{2}$,the roots t_* and t_{**} are distinct and the order of convergence of t_j to t_* is at least quadratic; hence, by (2.1.20) the order of convergence of x_j to x_* is at least quadratic. Finally, the uniqueness statement follows as in [LVK59] by consideration of the simplified Newton iteration(2.1.7).
□

We have presented in this chapter Newton's method for nonsingular roots, and could then prove the quadratic convergence of Newton's method, this conclusion is no longer possible when the Jacobian is singular at a solution $x_* \in F^{-1}(0)$. Then only linear convergence can be expected, as we will see in the following chapters.

Chapter 3

Newton's Method at Singular Points

In this section we will describe the behavior of Newton's method at singular solutions x_*, where $F'(x_*)$ is not invertible. For this case we are going to present the concept of regular singularity. To see the behavior of Newton's method at such points we start by characterizing a regular singularity, then we present the convergence results.

Assumptions 3.0.18 *For $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, suppose $x_* \in F^{-1}(0)$ is a singular solution of a nonlinear equation $F(x) = 0$ and F' is Lipschitz continuous for all $x \in \mathbb{R}^n$ and the constant $L > 0$. We define $N := \ker F'(x_*)$, and $N_\perp$ denote the orthogonal complement of N, such that $N \oplus N_\perp = \mathbb{R}^n$, and let $N^* := \ker F'(x_*)^\top$. Let $1 \leq m := \dim N$. We denote by P_N, $P_{N_\perp}$, and P_{N^*} the projection operators into N, $N_\perp$, and N^*, respectively, where $P_N : N \longrightarrow N_\perp$, $\quad P_{N^*} : N^* \longrightarrow N^*_\perp$, so that*

$$P_{N^*} F'(x_*) = F'(x_*) P_N = 0.$$

We set

$$P_{N_\perp} = I - P_N, \qquad P_{N^*_\perp} = I - P_{N^*}.$$

where $N^*_\perp$ is the orthogonal complement of N^*.

We will frequently write the elements of $\mathbb{R}^n$ in the form

$$x = x_* + \rho\, t, \quad \textit{with} \quad \rho = \|x - x_*\| \quad \textit{and} \quad t \in S. \tag{3.0.1}$$

Where

$$S \equiv \{t \in N : \ \|t\| = 1\}, \tag{3.0.2}$$

is the unit sphere of directions in $\mathbb{R}^n$.

The perturbation $\hat{x}$ will refer to the difference between the sequence $\{x_j\}_{j \geq 0}$ and the root x_*, we write

$$\hat{x}_j = x_j - x_*, \quad j = 0, 1, \ldots$$

The determinant of the Jacobian $F'(x)$ of the function F is

$$\delta(x) = \det(F'(x)).$$

The singular set is given by

$$\delta^{-1}(0) = \{x \in \mathbb{R}^n \,:\, \delta(x) = 0\}.$$

3.1 Characterization of Regular Singularities

The Newton iteration is invariant with respect to nonsingular affine transform on the domain of F and nonsingular linear transformation on the range of F respectively. Hence we may assume without loss of generality that

$$x_* = 0, \quad F'(x_*) = (I - P_{N^*}), \quad and \quad N^* = \mathbb{R}^m \times \{0\}^{n-m} \tag{3.1.1}$$

which implies in particular that the nullspace N is identical to N^*.

From the third assumption in (3.1.1), we have that the projection operator P_{N^*} is the identity on its range N^*, i.e. $P_{N^*}\, x = x$, $\forall x \in N^*$, and takes the form

$$P_{N^*} = \begin{pmatrix} I_{m\times m} & 0_{m\times n-m} \\ 0_{n-m\times m} & 0_{n-m\times n-m} \end{pmatrix},$$

where 0 represents the zero matrix and subscripts indicate dimensions. By substituting in the second assumption of (3.1.1), we obtain

$$F'(0) = I - P_{N^*} = \begin{pmatrix} 0_{m\times m} & 0_{m\times n-m} \\ 0_{n-m\times m} & I_{n-m\times n-m} \end{pmatrix}. \tag{3.1.2}$$

Definition 3.1.1 *(Order of Singularity)*

We define the order of the singularity at $x_* \in F^{-1}(0)$ as the index $k \geq 0$ for which F has a Lipschitz continuous $(1+k)th$ derivatives and

$$P_{N^*}F^{(k+1)}(x_*) \neq 0, \quad and \quad P_{N^*}F^{(q)}(x_*) = 0 \quad for \quad q \in [1,k],$$

and we will call "m" the defect of the singularity. In the scalar case $m = 1$ the index $k+1$ gives the algebraic multiplicity of the root x_*.

Let $(.)|_N$ and $(.)|_{N_\perp}$ denote the restriction maps for N and $N_\perp$ respectively. Using (3.1.1), the Jacobin F' can be partitioned as follows:

$$F'(x) = \begin{pmatrix} P_{N^*}F'(x)|_N & P_{N^*}F'(x)|_{N_\perp} \\ P_{N^*_\perp}F'(x)|_N & P_{N^*_\perp}F'(x)|_{N_\perp} \end{pmatrix} := \begin{pmatrix} B(x) & C(x) \\ D(x) & E(x) \end{pmatrix}. \tag{3.1.3}$$

In conformity with the partitioning in (3.1.2), the submatrices B, C, D, and E are $(k+1)th$ times differentiable and have dimensions $m \times m, m \times n-m, n-m \times m$, and $n-m \times n-m$, respectively.

To compute the determinant of a Jacobin F', we reduce it to block lower triangular matrix, by multiplying it from the left by the unitary block matrix

$$\begin{pmatrix} I & -C(x)E(x)^{-1} \\ 0 & I \end{pmatrix}.$$

The result is the block triangular form

$$\begin{pmatrix} B(x) - C(x)E(x)^{-1}D(x) & 0 \\ D(x) & E(x) \end{pmatrix},$$

whose determinant is equal to the product of the determinants of the diagonal blocks, i.e.

$$\delta(x) = \det\Big(B(x) - C(x)\,E(x)^{-1}\,D(x)\Big)\ \det E(x). \tag{3.1.4}$$

Let

$$G(x) := B(x) - C(x)\,E(x)^{-1}\,D(x) \in \mathbb{R}^{m\times m}, \tag{3.1.5}$$

and then by the assumption on $F'(x)$, the principal submatrix $E(x_*) = I$ must be nonsingular, so that in some neighborhood of x_* the determinant function takes the form

$$\delta(x) = \det(G(x))\det(E(x)) \approx \det(G(x)). \tag{3.1.6}$$

From the Assumption 3.1.1 we have $x_* = 0$, $x = \rho\, t$ and by applying Taylor's theorem to the entries of F' in (3.1.3) we find for B that

$$\begin{aligned} B(x) = P_{N^*}F'(x)|_N &= \sum_{j=0}^{k} P_{N^*}F^{(j+1)}(0)|_N(\rho t)^j(j!)^{-1} + O(\rho^{k+1}) \\ &= (k!)^{-1}P_{N^*}F^{(k+1)}(0)|_N(\rho t)^k + O(\rho^{k+1}), \\ &= (k!)^{-1}\rho^k\bar{B}(t) + O(\rho^{k+1}). \end{aligned} \tag{3.1.7}$$

Since

$$\bar{B}(t) := P_{N^*}F^{(k+1)}(x)|_N\, t^k. \tag{3.1.8}$$

In the same way we obtain for the other entries of F'

$$\begin{aligned} C(x) &= (k!)^{-1}\rho^k\bar{C}(t) + O(\rho^{k+1}) = O(\rho^k) \\ D(x) &= O(\rho), \quad \textit{and} \quad E(x) = I + O(\rho). \end{aligned} \tag{3.1.9}$$

For some $r > 0$, $E(x)$ is invertible for all $\rho < r$ and all $t \in S \subset \mathbb{R}^n$, with $E^{-1}(x) = I + O(\rho)$. Then the Jacobian $F'(x)$ is invertible if and only if the reduced Jacobian $G(x)$ defined by (3.1.5) is nonsingular. This claim follows from (3.1.10) by reducing r if necessary to apply (3.1.9), we have

$$G(x) = B(x) + O(\rho^{k+1}) = \frac{\rho^k}{k!}\bar{B}(t) + O(\rho^{k+1}).$$

At the singularity $x_* = 0$ we have the submatrices $C(x_*) = D(x_*) = 0$ and $E(x_*) = I$. Hence

$$
\begin{aligned}
\delta(x) &= \det(G(x)) \det(E(x)), \\
&= \det\left(B(x) + O(\rho^{k+1})\right)(I + O(\rho)), \\
&= \left((k!)^{-1}\rho^k\right)^m \det \bar{B}(t) + O(\rho^{km+1}).
\end{aligned}
\tag{3.1.10}
$$

If we differentiate the first relation in (3.1.10) we obtain the gradient

$$\nabla\delta(x) = \nabla \det\left(B(x) - C(x)\,E(x)^{-1}\,D(x)\right) \det E(x) + \left(B(x) - C(x)\,E(x)^{-1}\,D(x)\right) \nabla \det E(x).$$

This reduces at x_* to

$$\nabla\delta(x_*) = \nabla \det(B(x_*)) \det(E(x_*)) = \nabla \det(B(x_*)). \tag{3.1.11}$$

If

$$\nabla \det(B(x))|_{x=x_*} \neq 0, \tag{3.1.12}$$

it follows from the implicit function theorem that the singular set $\delta^{-1}(0)$ forms a smooth hypersurface in some neighborhood of x_*.

Remark 3.1.2 *We are mainly interested throughout this thesis in first order singularities with a one-dimensional nullspace i.e., $m = 1 = k$. Then there are vectors v, $u \in S$ spanning the nullspace N of $F'(x_*)$ and its transpose N^* of $F'(x_*)^\top$ respectively, which are both identical to the first Cartesian basis vector in $\mathbb{R}^n$ by the normalization* (3.1.1).

Remark 3.1.3 *After the above normalization, we can consider* $\det(\bar{B})$ *as $m \times m$ matrix depending on $t \in S \cap N \subset \mathbb{R}^n$.*

Definition 3.1.4 *(Regular Singularity)*

The singularity x_* is said to be regular if

$$\det(\bar{B}(t)) \neq 0 \quad for\, some \quad t \in S \cap N \subset \mathbb{R}^n. \tag{3.1.13}$$

This reduces for the case in Remark 3.1.2 to

$$\bar{B} = \nabla B(x_*)^\top v = \nabla\delta(x_*)^\top v = u^\top F''(x_*)vv \neq 0, \tag{3.1.14}$$

In which case we call the singularity simple.

Remark 3.1.5 *If* $\det(\bar{B}(t)) = 0$ *for all $t \in S \subset \mathbb{R}^n$, so that the regularity condition (3.1.13) is violated, then the singularity x_* is called irregular [GO83]. When $m = 1 = k$ then the null vector v given in (3.1.14) is orthogonal to* $\nabla\delta(x_*)$ *and thus tangential to the singular set* $\delta^{-1}(0)$.

Remark 3.1.6 *The strong regularity condition that* $\det(\bar{B}(t)) \neq 0$ *for all $t \in S \cap N \subset \mathbb{R}^n$ had been used by Reddien [Red78], Decker and Kelley [DK80a], which is called in this case strong regularity.*

Remark 3.1.7 *When $m = 1$, then the condition* $\det(\bar{B}(t)) \neq 0$ *for all $t \in S \cap N \subset \mathbb{R}^n$ is equivalent to the condition* $\det(\bar{B}(t)) \neq 0$ *for some $t \in S \cap N \subset \mathbb{R}^n$.*

Proposition 3.1.8 *Strong regularity, i.e.* $\det(\bar{B}(t)) \neq 0 \quad \mathit{for\,all} \quad t \in S \cap N \subset \mathbb{R}^n$ *implies that* m *is even or equal to one [GO81].*

Proof: Consider an $m \times m$ matrix representation of $\bar{B}(t)$. This is a homogeneous function of t. Now

$$\det(\bar{B}(t)) := (-1)^m \det \bar{B}(-t), \tag{3.1.15}$$

and t can be deformed continuously into $-t$ while remaining on the surface of the sphere $\|l\| = \|t\|$ in N, provided $m > 1$. Thus it follows from (3.1.15) that there exists an $l \in N$ with $\|l\| = \|t\|$, such that $\det(\bar{B}(l)) = 0$ provided m is odd > 1; so that the assumption excludes all odd cases. □

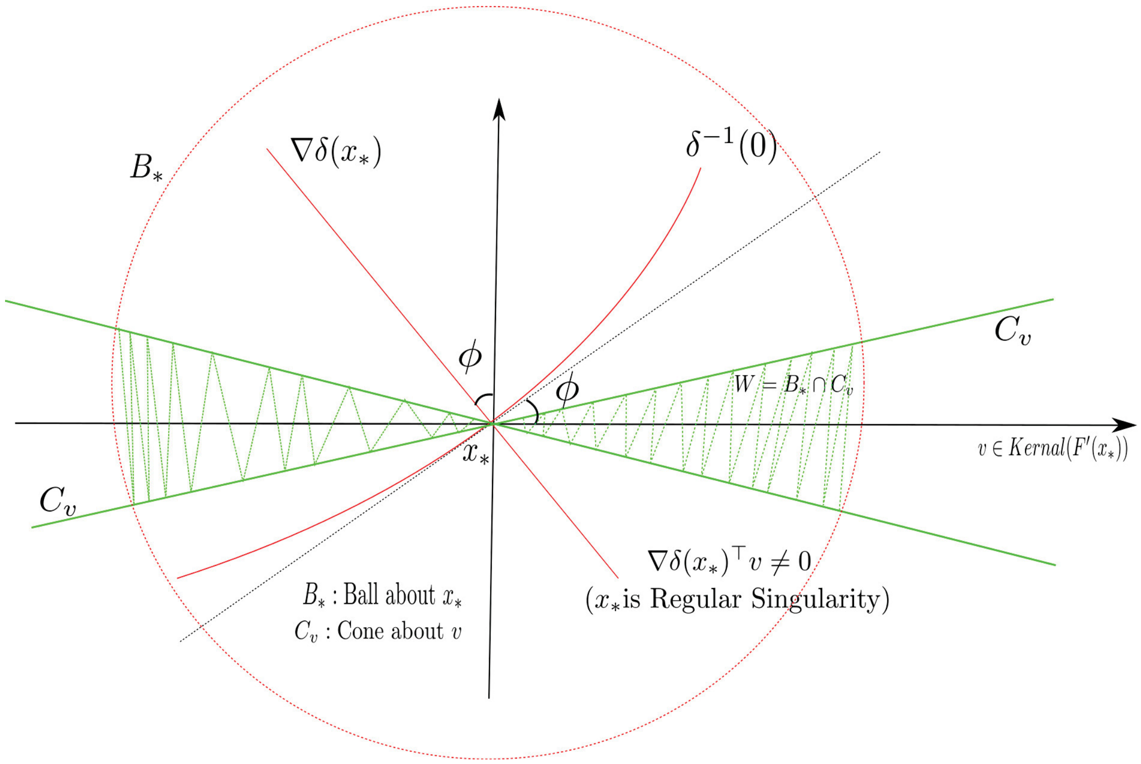


Figure 3.1: Singular set and set of good start points $W = B_* \cap C_v$ for $m = 1 = k$

Now we are going to define an open starlike domain $\mathcal{A}$ of initial points.

Definition 3.1.9 *Any open set $\mathcal{A} \subset \mathbb{R}^n$ is called starlike domain with respect to x_* if*

$$x \in \mathcal{A} \quad \Longrightarrow \quad \{(1-\lambda)x_* + \lambda x,\ 0 < \lambda < 1\} \in \mathcal{A}.$$

In other words $\mathcal{A}$ contains the open line segment between x_* and any one of its elements $x \in \mathcal{A}$.

Particular examples of starlike domains are balls, cones and their intersections called wedges, which were used by Reddien [Red78], [Red79] and Decker and Kelley [DK80a] as domains of convergence.

Definition 3.1.10 *A domain of convergence is any open set of initial points from where Newton's method converging to a particular root x_*.*

Definition 3.1.11 *The set of included directions is defined by*

$$\mathcal{A}_s = \{t \in S \ : \ \mathcal{A} \cap \{x_* + \rho\, t\}_{\rho>0} \neq \emptyset\}.$$

Definition 3.1.12 *A vector $t \in S \subset \mathbb{R}^n$ is called an excluded direction for $\mathcal{A}$ if*

$$x_* + \rho\, t \notin \mathcal{A} \quad \textit{for all} \quad \rho > 0.$$

Definition 3.1.13 *A starlike domain $\mathcal{A}$ has density 1 at x_* if the set of the excluded directions has measure zero in the unit sphere $S \subset \mathbb{R}^n$.*

3.2 Convergence Results at Smooth Singularity

The behavior of Newton's method when the Jacobian $F'(x)$ is singular at the solution x_* has been analysed extensively in the literature [Ral66],[Red78],[Red79], [DK80a],[DK80b], [Gri80b], [GO81], [DK82],[DKK83],[GO83],[KS83],[EP84], [DK85], [Gri85], [QS91], [Kan94].

Later on [KH99], [BKM03], [FP03], [SY05],[OW07], [LVZ06], [OW07], [OW09], [HMT09], [BK08], [BKL10],[LZ10], [LZ12b] and [LZ12a].

Several sufficient conditions for its convergence have been formulated. Under certain regularity and smoothness assumptions, the existence of special regions (cones, starlike regions) about the solution x_* has been proven. The Jacobian $F'(x)$ is nonsingular at all points of these regions except at x_*. If the initial iterate lies in such a region, then the Newton iterates will remain in this region and converge linearly to x_*.

In general there are two difficulties when one wants to prove the convergence for singular problems, First, there may be singular manifolds containing x_* on which $F'(x)$ is singular. Hence the starting points must be chosen from some region about x_* in which $F'(x)$ is invertible. Second, we must show that the subsequent iterates remain in some region of invertibility like $W_{\mu,\theta}$, which will be given by (3.2.4), [DKK83].

Rall in[Ral66] analysed the univariate case $n = 1$ in detail. He found that Newton's method will succeed and can be modified to retain quadratic convergence. He also sketched some results for the higher dimensional case, which were partly wrong. Cavanagh [CAV70] provides more details for Rall's ideas, however, he does not have the generality of [Red78], but he made the stringent assumption that $F'(x)$ is nonsingular in a deleted neighborhood of the root. Reddien [Red78] gave sufficient conditions for the convergence of the Newton iteration for the particular case in which $F'(x)$ has a smooth singular manifold of codimension one in a neighborhood of x_*.

The extension of these results to the case in which the null space of $F'(x_*)$ has dimension $m > 1$ has been considered by Decker and Kelley [DK80a], and Reddien [Red79]. Griewank and Osborne [GO81] succeeded in weakening the conditions under which the main results in [Red78] are derived. The original conditions cannot be satisfied when the dimension of the null space is an odd integer $m > 1$. Their results also represent a slight generalization over those in [Red79] (they are valid if some components of $F(x)$ are linear in x), and the derivation is significantly more direct.

Remark 3.2.1 *The regularity condition given in (3.1.13) are typically needed to prove convergence of Newton's methods to singular solutions, such conditions concern the behavior of certain directional derivatives of F' on the null spaces N and N^*.*

For a first order singularity $k = 1$, Griewank assumed that $F''(x)$ is Lipschitz continuous near the regular singularity solution x_* and satisfies the regularity condition given in (3.1.13).

Griewank's convergence analysis shows that the first Newton step takes the initial point x_0 from the original starlike domain $\mathcal{A}$ into a simpler starlike domain W defined by [(26) in [OW09], a wedge around a vector t in the null space N. The domain W is similar to the domains of convergence $W_{\mu,\theta}$ found in earlier works [Red78], [DK80a]. Linear convergence is then proved inside W. For F twice continuously differentiable, the convergence domain $\mathcal{A}$ is much larger than W and the set of excluded directions from $\mathcal{A}$ has zero measure.

Griewank [Gri80b] constructed an open starlike domain of initial points $\mathcal{A}$, from which Newton's method converges linearly to x_*. He provides in [Gri85] a comprehensive survey of the singularity results.

One can find in all those above studies that the quadratic convergence of the Newton iterates is degraded to linear.

Griewank in [Gri85] combined various results from [DK80a],[DK80b],[Gri80b] and [Gri80a] to obtain the following convergence theorem of Newton's method at regular singularities.

Theorem 3.2.2 *Let $F \in \mathcal{C}^{k+1}(\mathbb{R}^n)$ have a regular singularity of order $k > 0$ at $x_* \in F^{-1}(0)$. Then there exists a starlike domain $\mathcal{A}$ with density 1 at x_* such that any Newton sequence*

$$x_j = x_* + \rho_j t_j = x_{j-1} - F'(x_{j-1})^{-1} F(x_{j-1}), \quad j = 1, 2, \dots, \tag{3.2.1}$$

generated from some initial point $x_0 \in \mathcal{A}$ converges to x_* and satisfies:

(i) $\frac{\rho_{j+1}}{\rho_j} = \frac{k}{k+1} + O(\rho_j)$,

(ii) $t = \lim_{j\to\infty} t_j \in N \cap \mathcal{A}_s$.

(iii)

$$\theta_j \equiv \min\{cos^{-1}(s^\top t_j \,:\, s \in N \cap S)\} := \begin{cases} O(\rho_j^2) & \textit{if } k = 1 \\ O(\rho_j) & \textit{otherwise} \end{cases} \tag{3.2.2}$$

(iv) $\frac{F(x_j)}{\rho_j^2} = \frac{1}{2k^2} \nabla^2 F(x_*) tt + O(\rho_j)$.

(v) $\lim_{j\to\infty} \frac{\sigma_j}{\sigma_{j+1}} = (\frac{k}{k+1})^{-k} \in [2, e]$.

Where

$$\sigma(t) := \begin{cases} 0 & \text{if } \bar{B}(t) \text{ is singular} \\ \min(1, \ \|\bar{B}(t)^{-1}\|^{-1}) & \text{otherwise} \end{cases} \tag{3.2.3}$$

denotes the smallest singular value of $F'(x_j)$ and $O(\rho_j^i)$ may be any sequence that is majorized by a multiple of the sequence $\{\rho_j^i\}$.

Remark 3.2.3 *From the regular singularity Theorem 3.2.2 we conclude the following.*

- *Linear convergence (i) is almost certain whenever the initial point x_0 is sufficiently close to a regular singularity x_*.*

- *According to (ii) the iterates approach x_* along a regular direction $t \in N \cap \mathcal{A}_s$.*
- *The angle θ_j between the discrepancies $x_j - x_*$ and the nullspace N at x_* tends rapidly to zero.*
- *The last assertion (v) shows that the condition number of the Jacobian essentially doubles at each step. However, at least in the simple case, it appears that this gradual deterioration of the conditioning does not prevent Newton's method form attaining the maximal solution accuracy that one can reasonably expect.*

Remark 3.2.4 *As a consequence of (iv) and (i), the Newton iteration reduces eventually the residual $F(x_j)$ by the factor $(\frac{k}{k+1})^2$ at each step. This feature allows one to monitor the progress towards a singularity and to estimate its order.*

To see that, let look at the following simple example for $k = 1$:

$$F(x) = x^2 \implies x_{j+1} = \tfrac{1}{2}x_j \implies \|F(x_{j+1})\| = \tfrac{1}{4}\|F(x_j)\|$$

In [Gri80a] Griewank showed, that the chord method and other stabilized Newton's methods $x_{j+1} = x_j - H_j^{-1}F(x_j)$ with $\sup \|H_j^{-1}\| < \infty$ can only converge sublinearly in that

$$\lim_{j\to\infty} \|\hat{x}_j\|^{\frac{1}{j}} = 1.$$

Some of results in the above literature apply also to mappings between Banach or Hilbert spaces. In this thesis we will only consider the finite dimensional case.

Now we want to discuss some other convergence results of Newton's method at singular solutions due to [Red78], [Red79], [DK80a], [Gri80b], [Gri80a], [GO81], [Gri85]. Reddien in [Red78] indicates that the convergence region around x_* must have quite a special structure.

The singular set $\delta^{-1}(0)$ of $F'(x)$ near x_* may range from a single point to the union of several codimension one smooth manifold through x_*. Hence the nonsingularity of F' can be expected only in carefully selected regions about x_*. An added difficulty is that the Newton iterates must remain in the chosen region of invertibility of F'.

The following starlike set $W_{\mu,\theta}$ satisfies both above requirements.

Definition 3.2.5 *For $\mu,\ \theta > 0$ and $\hat{x} = x - x_*$, we define the region of acceptable initial iterates for Newton's method $W_{\mu,\theta}$ by*

$$W_{\mu,\theta} = \{x \in \mathbb{R}^n \,:\, 0 < \|\hat{x}\| \leq \mu,\ \|P_{N_\perp}(\hat{x})\| \leq \theta\,\|P_N(\hat{x})\|\}, \tag{3.2.4}$$

which is a wedge rooted at x_* around the null space N of $F'(x)$.

Remark 3.2.6 *We note that the wedge $W_{\mu,\theta}$ is the intersection of a ball and a cone. It is only convex if $m = 1$.*

Reddien in [Red79] gave two generalizations of the results in [Red78], he put conditions for convergence when the dimension of the nullspace is greater than one and when the solution is of higher order. He had defined the subset wedge

$W_{\mu,\theta,\psi}$, for ψ, μ, $\theta > 0$ of $W_{\mu,\theta}$ given by (3.2.4) and for L is a one dimensional subspace of N

$$W_{\mu,\theta,\psi} = \{x \in \mathbb{R}^n \,:\, x \in W_{\mu,\theta},\ \|(I - P_L)P_N(\hat{x})\| \leq \psi \ \|P_N(\hat{x})\|\}\,. \qquad (3.2.5)$$

Where P_L is any bounded projection onto the subspace L of N. For the case $m > 1$ he uses that $\dim L = 1$ and let Newton step start inside the new smaller wedge $W_{\mu,\theta,\psi}$.

Decker and Kelley in [DK80a], Griewank and Osborne [GO81] had extended the theorem of Reddien [Red78] to the case when $m < \infty$, this theorem gives sufficient conditions for convergence of Newton iterates for singular problems. It has been extended by Decker and Kelley in [DK80a].

We deduced from the above analysis that Newton's method converges Q-linearly if $F'(x_*)$ is singular. This situation is analysed and acceleration techniques are suggested by many authors, including Reddien [Red78], Decker and Kelley [DK80a], [DK80b], [DK82], Decker, Keller and Kelley [DKK83], Kelley and Suresh [KS83], Griewank and Osborne [GO83], Griewank [Gri85].

In summary, their papers show that when the Jacobian at the solution has a null space and the order of singularity is k with a certain regularity condition being satisfied, then from good starting points in W, Newton's method is locally Q-linearly convergent with asymptotic reducing ratio $\frac{k}{k+1}$. That means the Newton step is too short. For acceleration we need to extend the Newton's step by a step multiplier larger than k ideally converging to $k+1$. While Rall already noted that in the univariate case $n = 1 = m$ Newton's method can be accelerated even when the order $k > 1$, this seems rather difficult in the multivariate case. Hence, we consider only first order singularities with $k = 1$ from now on.

So far we have assumed that F is $(k+1)th$ times continuously differentiable. When F is the nonlinear equation formulation of a nonlinear complementarity problem according to the approach of Evtushenko [EP84], then lack of strict complementarity leads to F having a first order singularity without F' being Fréchet differentiable. This situation was carefully analysed by Wright et al. [OW09], who showed that $F'(x)$ is still strongly semismooth as defined below.

By the standard Assumptions 3.0.18, we note that F' is Lipschitz continuous but singular, so that the celebrated results of Qi and Sun [QS91] can and need not be invoked. Instead we will see that a weakened version of Griewank starlike convergence theorem applies.

The main difference in the results are that the domain of convergence $\mathcal{A}$ need no longer have density 1, i.e., there may be full-dimensional cones of excluded directions.

We now list various definitions relating to the smoothness of a function.

Definition 3.2.7 *(Directionally Differentiable)*

Let $G : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^{\tilde{n}}$ with Ω open, $x \in \Omega$ and $v \in \mathbb{R}^n$. If the limit

$$\lim_{t \downarrow 0} \frac{G(x + tv) - G(x)}{t}$$

exists in $\mathbb{R}^{\tilde{n}}$, then G has a directional derivative at x along v, we denote this limit by $G'(x; v)$. G is directionally differentiable at x if $G'(x; v)$ exists for every $v \neq 0$.

Definition 3.2.8 *(B-differentiable)*

G is B(ouligand)-differentiable at $x \in \Omega$ if G is Lipschitz continuous in a neighborhood of x and directionally differentiable at x.

Definition 3.2.9 *(Strongly Semismooth)*

Let $G : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^{\tilde{n}}$ with Ω open, G be locally Lipschitz continuous on Ω. G is strongly semismooth at $\bar{x} \in \Omega$ if G is directionally differentiable near $\bar{x}$ and

$$\lim \sup_{\bar{x} \neq x \to \bar{x}} \frac{\|G'(x; x - \bar{x}) - G'(\bar{x}; x - \bar{x})\|}{\|x - \bar{x}\|} < \infty.$$

Further, G is strongly semismooth on Ω if G is strongly semismooth at every $\bar{x} \in \Omega$.

Proposition 3.2.10 *If $F'(x)$ is strongly semismooth at x, then it is B-differentiable at x. Further, if $F'(x)$ is B-differentiable at x, then $F''(x; .)$ is Lipschitz continuous [QS91]. Hence, for F' strongly semismooth at x_*, there is some L_{x_*} such that*

$$\|(F')'(x_*; h_1) - (F')'(x_*; h_2)\| \leq L_{x_*} \|h_1 - h_2\|. \tag{3.2.6}$$

Proposition 3.2.11 *If $F'(x)$ is strongly semismooth at x_* and $\|x - x_*\|$ is sufficiently small, we have the following crucial estimate from equation (7.4.5) of [FP03]*

$$F'(x) = F'(x_*) + (F')'(x_*; x - x_*) + O(\|x - x_*\|^2). \tag{3.2.7}$$

Now Wright et al. weaken the smoothness assumptions of Griewank by replacing the second derivative of F by a directional derivative of F'. The assumptions follow:

Assumptions 3.2.12 *(Assumptions of Wright et al.,[OW09])*

For $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, let x_* be a singular solution, satisfying the regularity condition (3.1.13) with $\bar{B}(t) = B(x_*; t) = P_{N^*} F''(x_*; d)|_N$ the directional derivative for $t \in S \cap N \subset \mathbb{R}^n$ with $k = 1$ and $m \geq 1$ and $F'(x)$ is strongly semismooth at x_* but $F \notin \mathcal{C}^2(\mathbb{R}^n)$ i.e., F'' need not exist.

Wright et al. show that Griewank's convergence results [Gri80b], [Gri85] hold under this assumptions.

Theorem 3.2.13 *Suppose Assumptions 3.2.12 hold. Then there exists a starlike domain $\mathcal{A}$ about x_* with positive density such that, if Newton's method for $F(x)$ is initialized at any $x_0 \in \mathcal{A}$, the iterates converge linearly to x with asymptotic rate $\frac{1}{2}$. If the problem is converted to standard form (3.1.1) and $x_0 = \rho_0 t_0$, where $\rho_0 = \|x_0\| > 0$ and $t_0 \in S$, then the iterates converge inside a cone whose axis depends on t_0.*

Proof: See [OW07]. □

Remark 3.2.14 *Wright et al. use the properties (3.2.6) and (3.2.7) to show that F' is smooth enough for the steps in the proof to hold. Finally, they make an insignificant change to a constant required by the proof due to a loss of symmetry in $\mathcal{A}$. (Symmetry is lost in moving from derivatives to directional derivatives because directional derivatives are positively but not negatively homogeneous [OW09]). The proof in [Gri80b] also considers regularities larger than* 2*, for which higher derivatives are required. Here we are only interested in first order regularity.*

Chapter 4

Acceleration of Newton's Method at Singularities

When the sequence $\{x_j\}_{j\geq 0}$ produced by a Newton iterates converge to x_*, then the discrepancy $x_j - x_*$ lies mainly in the nullspace N of $F'(x_*)$. Acceleration schemes typically attempt to stay within a cone around the nullspace N while lengthening the Newton steps.

We will present some of those acceleration techniques proposed in the early 1980s.

Decker and Kelley [DK82] prove superlinear convergence for a scheme in which every second Newton step is essentially doubled in length along the subspace N.

Decker, Keller, and Kelley [DKK83] prove superlinear convergence when every third step is overrelaxeded, provided the third derivative of F at x_* satisfies a coercivity condition on the nullspace N of $F'(x_*)$.

Kelley and Suresh [KS83] require the third derivative of F at x_* bounded over the truncated cone about the nullspace N of $F'(x_*)$. Overrelaxing every other step by a factor approaching 2 results in superlinear convergence.

In summary, their techniques require regularity condition (3.1.13) at x_* . All those schemes require starting points whose error remain in a wedge around the subspace N, and all assume three times differentiability of F.

In contrast the acceleration technique due to Wright et al. [OW09] does not require the starting point x_0 to be in a cone about the subspace N, and requires only strong semismoothness of F' at x_*.

We will focus in this section on two acceleration schemes, that are closely related to Newton's method, since all later Newton steps cover roughly half of the distance to the root x_*. Firstly we discuss two point method [OW09], then a three-point method [Gri80a].The notations of last chapter are also used here.

Overrelaxation is known to improve the rate of convergence of Newton's method to a singular solution [Gri85]. The overrelaxed iterate is with η_j

$$x_{j+1} = x_j + \eta_j d_j, \quad F'(x_j)d_j = -F(x_j), \quad j = 0, 1, \ldots. \tag{4.0.1}$$

Where Newton direction d_j is defined in (4.0.1). On every other step the multiplier η_j is equal to 2 for the case of Kelley and Decker [DK82], and $\eta_j \in [1, 2)$ for the case considered by Wright et al. [OW09].

4.1 Two-Point Method

In first part of this section we study the two point method considered by Kelley and Decker [DK82]. They give sufficient conditions such that the modified Newton's method produces the subsequence $\{x_{2j}\}_{j\geq 0}$ of the sequence $\{x_j\}_{j\geq 0}$, which converges quadratically to the root x_* if $F'(x_*)$ is singular.

Their main results can be presented in the followed theorem

Theorem 4.1.1 *With the same notations as in the Theorem 3.2.2 assume that:*

1. $m \leq 2$.
2. $\bar{B}(t)$ *is nonsingular for all* $t \in S \cap N$.
3. *There is* $c_0 > 0$ *so that for all* $v \in N$, $\|G'''(x_*)vvv\| \geq c_0 \|v\|^3$.

Then, for μ *and* θ *sufficiently small, and* $x_0 \in W_{\mu,\theta}$ *the sequence* $\{x_j\}$ *given by*

$$\begin{aligned} x_{2j+1} &= x_{2j} + d_{2j}, \quad j \geq 0,, \\ x_{2j+2} &= x_{2j+1} + (I + P_N)d_{2j+1}, \end{aligned} \tag{4.1.1}$$

stays in $W_{\mu,\theta}$ *and converges to* x_*. *Moreover, there is an* $M > 0$ *so that for* $j \geq 0$

$$\|\hat{x}_{2j+2}\| \leq M \|\hat{x}_{2j}\|^2.$$

Proof: See [DK82]. □

This result is theoretically very interesting, but not really implementable since the nullspace N at x_* and thus the projection P_N onto it are generally not known. Therefore the overrelaxed step must be held back a bit, i.e., the step multipliers η_j must be less than 2.

If every step is overrelaxed, we can show that in general the condition $\eta < \frac{4}{3}$ must be satisfied to ensure convergence and, as a result, the rate of linear convergence is no faster than $\frac{1}{3}$. On special problem all steps may be overrelaxed by some $\eta < \frac{4}{3}$ as we observed in Table 6.11. Now, we focus on a technique in which overrelaxation occurs only on every second step; that is, standard Newton steps are interspersed with steps of the form (4.0.1) for some fixed, $\eta \in (1, 2]$. Broadly speaking, each pure Newton step refocuses the iterates along the null space. Kelley and Suresh prove superlinear convergence for this method when η is systematically increased to 2 as the iterates converge. However, their proof requires the third derivative of F evaluated at x_* to satisfy a coercivity condition and assumes a starting point x_0 that lies near a regular direction in the null space of $F'(x_*)$.

We state here the main result of Wright et al.[OW09].

The major assumptions are that regularity condition (3.1.13) holds at x_* and that $x_0 \in R_\eta$, where R_η is a starlike domain whose excluded directions are identical to those of $\mathcal{A}$ defined in Section 4 in [OW07] but whose rays may be shorter. The lengthy proof appears in full in [[OW07], Sect. 5].

Theorem 4.1.2 *Suppose Assumption 3.2.12 holds and let $\eta \in [1, 2)$. There exists a starlike domain $R_\eta \subseteq R$ about x_* such that if $x_0 \in R_\eta$ and with iterates defined by*

$$\begin{aligned} x_{2j+1} &= x_{2j} - F'(x_{2j})^{-1}F(x_{2j}), \\ x_{2j+2} &= x_{2j+1} - \eta F'(x_{2j+1})^{-1}F(x_{2j+1}), \end{aligned} \tag{4.1.2}$$

for $j = 0, 1, 2, \ldots$ then the iterates $\{x_j\}$ converge linearly to x_* and

$$\lim_{j\to\infty} \frac{\|x_{2j+2} - x_*\|}{\|x_{2j} - x_*\|} = \frac{1}{2}(1 - \frac{\eta}{2}).$$

Proof: See [OW07]. □

As an immediate consequence we see that if we gradually enlarge η towards 2 at just the right rate superlinear convergence can be achieved theoretically.

Corollary 4.1.3 *Under the assumptions of Theorem 4.1.2, the multiplier $\eta = \eta_j$ can gradually be pushed towards 2, so that*

$$\lim_{j\to\infty} \frac{\|x_{2j+2} - x_*\|}{\|x_{2j} - x_*\|} = 0,$$

i.e., we have two-step superlinear convergence.

The two point method with doubling every other step is not feasible in the multivariate case because $x^1_{j+1} = 2x_{j+1} - x_j = x_j - F'(x_j)^{-1}F(x_j)$ even though closer to x_*, may be much less suitable as a starting point for the next iteration. Nevertheless under certain conditions one can restore at least R-superlinear convergence

Now we study a technique in which overrelaxation occurs only on every third step, that is three-point method [Gri80a].

4.2 Three-Point Method

Now we present the idea of three point method, we start with the full step Newton iteration and then we double every third iteration. That produces a new procedure, which accelerates Newton iteration, its analysis and proof of convergence is due to Griewank [Gri80a]. We want to introduce the behavior of this method in this section.

The iterations of the method can be written as

$$\begin{aligned} y_j &= x_j - F'(x_j)^{-1}F(x_j), \\ z_j &= y_j - F'(y_j)^{-1}F(y_j), \quad j = 0, 1, \ldots, \\ x_{j+1} &= z_j - 2F'(z_j)^{-1}F(z_j), \end{aligned} \tag{4.2.1}$$

Theorem 4.2.1 (Second Order Three-Point Method) *Let $F \in C^{3,1}(\mathbb{R}^n, \mathbb{R}^n)$ have a strongly regular first order singularity at x_*. If*

$$P_{N^*}F'''(x)t^3 \neq 0, \quad \forall\, t \in N \cap S,$$

then there exists a constant $\bar{\rho}$ and $\hat{\theta}$ such that the three-point iteration x_j given by (4.2.1) converges Q-quadratically to x_* with

$$\theta\left(\frac{x_{j+1} - x_*}{\|x_{j+1} - x_*\|}\right) = O(\|x_j - x_*\|),$$

from all initial points $x_0$in the starlike domain

$$\hat{V} \equiv \{x_* + \rho t \,:\, t \in S,\, \theta(t) < \hat{\theta},\; 0 < \rho < \bar{\rho}\}.$$

Proof: See [Gri80a]. □

The two multi-point schemes obtain the same results as Newton's method in considerably fewer steps and produce some iterates that are much closer to the solution. However, these gains can only be realized if it is assumed or known that the Jacobian at the sought after solution is singular. If the Jacobian $F'(x_*)$ turns out to be only nearly singular, then the $2-$ and $3-$step overrelaxation methods converge with an $R-$order of $\sqrt{2}$ and $\sqrt[3]{4}$ respectively [Gri85].

The three point method like the two point method has some theoretical appeal but both must eventually break down due to rounding errors. In practice both methods reach about the same solution accuracy as Newton's method.

In summary, the two and three point methods require deciding that the problem is singular, while our line-search to be discussed in the next section detects that automatically.

Chapter 5

Parabolic Range Space Interpolation

We consider a nonlinear equation $F(x) = 0$, where $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, and our aim is to compute efficiently a solution x_*. For that we intend to apply Newton's method with a line-search procedure based on parabolic range space interpolation.

5.1 Line-search

The Newton's method can be made more powerful if we use line-search techniques. Before we describe these techniques, we need to define a merit function.

5.1.1 Merit Functions

A merit function is a scalar-valued function of a variable x which indicates whether a new iterate is better or worse than the current one, in the sense of making progress toward a root of F. In nonlinear equation solving, the merit function is usually obtained by combining the n components of the vector F in some way. The most widely used merit function is the sum of squares, defined by

$$f(x) = \frac{1}{2}\|F(x)\|^2.$$

Any root x_* of F obviously has $f(x_*) = 0$. Since $f(x) \geq 0$ for all x, each root if F is a global minimizer of f. However, local minimizers of f are not roots of the gradient F if f is strictly positive at the point in question. The derivative of f is given by

$$\nabla f(x) = g(x) = F'(x)^\top F(x).$$

Definition 5.1.1 *The vector d is a descent direction with respect to the function $f(x)$ at the point x if it satisfies the condition*

$$g(x)^\top d < 0.$$

5.1.2 Line-search Strategy

Each iteration of a descent method computes a search direction d_j and then decides how far to move along that direction. The iteration is given by

$$x_{j+1} = x_j + \eta_j d_j, \quad j = 1, 2, 3, \dots . \tag{5.1.1}$$

where the positive scalar η_j is called the step length. The success of a descent method depends on effective choices of both the direction d_j and the step length η_j. Most line-search algorithms require d_j to be a descent direction, because this property guarantees that the function f can be reduced along this direction.

In our case, the direction d_j is Newton's direction. This direction is defined and descent, whenever $\det(F'(x_j)) \neq 0$ since

$$g_j^\top d_j = {f'}_j^\top d_j = ({F'}_j^\top F_j)^\top(-F_j^{-1} F_j) = -\|F_j\|^2 < 0, \quad d_j = -F'(x_j)^{-1}F(x_j). \tag{5.1.2}$$

Definition 5.1.2 *A line-search which chooses η_* to minimize $f(x_j + \eta_j d_j)$ is said to be perfect or exact.*

Definition 5.1.3 *A weak or inexact line-search is one which accepts any value of η_j such that $f(x_j + \eta_j d_j) - f(x_j)$ is sufficiently negative.*

A perfect line-search gives the greatest possible reduction in f along the search direction.

5.1.3 Effective Line-search

Definition 5.1.4 *A line-search is called effective in a level set*

$$D_0 = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\} \tag{5.1.3}$$

for a Lipschitz continuously differentiable function f if it ensures the inequality

$$f(x_{j+1}) - f(x_j) \leq -\delta \|g(x_j)\|^2 \cos^2 \phi_j, \tag{5.1.4}$$

for $x_j, x_{j+1} \in D_0$, and some coefficient $\delta > 0$ that does not depend on j. Here ϕ_j represents the angle, between the search direction $d(x_j)$ and $-g(x_j)$, i.e.

$$\cos \phi_j = \frac{-g(x_j)^\top d_j}{\|g(x_j)\|\|d_j\|} \quad \text{with} \quad 0 \leq |\phi_j| \leq \frac{\pi}{2}. \tag{5.1.5}$$

Now we introduce the so-called Not-too-small and Not-too-large conditions, which guarantee the effectiveness of many line-search algorithms. Before that we define the reduction ratio:

Definition 5.1.5 *The reduction ratio $q : \mathbb{R} \to \mathbb{R}$ is given by*

$$q_j(\eta) := \frac{f(x_j + \eta d_j) - f(x_j)}{\eta g(x_j)^\top d_j} = \frac{f(x_j) - f(x_j + \eta d_j)}{\eta |g(x_j)^\top d_j|}.$$

It compares the actual change of f in the numerator with the expected reduction based on the linear Taylor expansion of f at x_j in direction d_j for step size η in the denominator.

Definition 5.1.6 Not-too-large *condition on $\eta_j > 0$*

(a) *Upper Goldstein test:*

$$q_j(\eta_j) \geq \sigma \in (0, \tfrac{1}{2}),$$

i.e., the actual reduction $f(x_j) - f(x_j + \eta_j d_j)$ should be at least σ times the expected linearized reduction.

Definition 5.1.7 Not-too-small *conditions*

(b_0) *Lower Goldstein test: $q_j(\eta_j) \leq (1 - \sigma)$.*

(b_1) *Lower Armijo test: $q_j(\hat{\eta}_j) \leq \sigma$ for some $\hat{\eta}_j \leq \beta\eta_j$ with constant $\beta > 1$, i.e., condition (a) fails for $\hat{\eta}_j$ that is not very much bigger than η_j.*

(b_2) *Weak Wolfe condition:*

$$g(x_j + \eta_j d_j)^\top d_j \geq \gamma g(x_j)^\top d_j \quad \text{with} \quad \gamma \in (0, 1). \tag{5.1.6}$$

(b_3) *Strong Wolfe condition: $|g(x_j + \eta_j d_j)^\top d_j| \leq \gamma |g(x_j)^\top d_j|$ with $\gamma \in (0, 1)$.*

($\gamma \approx 0$ forces almost exact line-searches).

Proposition 5.1.8 ([Noc06]) *Every line-search that implements a combination of the Not-too-large condition (a) with any one of the Not-too-small conditions (b_j), $\quad j = 0, 1, 2, 3$ is effective in the sense of Definition 5.1.4.*

5.1.4 Zoutendijk Condition

Definition 5.1.9 (Zoutendijk Condition) *The divergence property*

$$\sum_{j=0}^{\infty} \cos^2 \phi_j = \infty, \tag{5.1.7}$$

is called Zoutendijk condition.

The Zoutendijk condition ensures that the angle between d_j and $-g(x_j)$ tends sufficiently slowly to the angle $\frac{\pi}{2}$ or is bounded away from it.

Theorem 5.1.10 ([Noc06]) *Suppose that the descent method (5.1.1) is applied to the function f with starting point $x_0 \in D_0$. Furthermore, assume that the search directions d_j are descent directions and the step multipliers η_j are computed by an effective line-search. Then the property*

$$\sum_{j=0}^{\infty} \cos^2 \phi_j \|g(x_j)\|^2 < \infty,$$

holds.

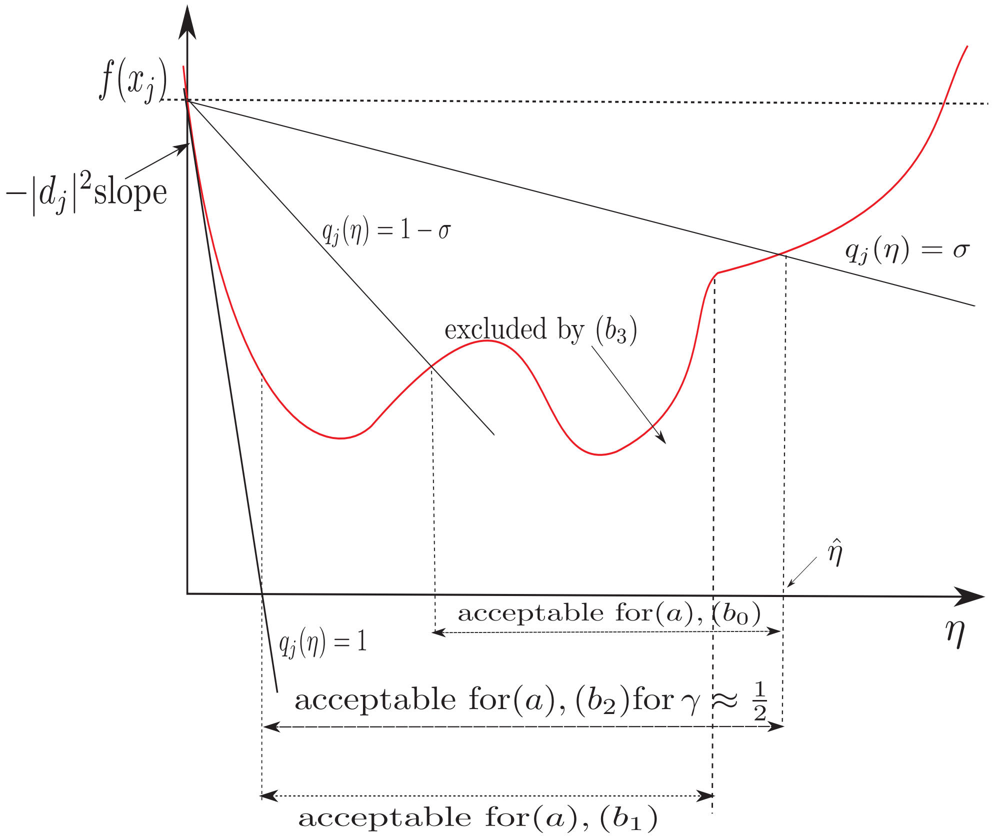


Figure 5.1: Line-search Conditions

Corollary 5.1.11 ([Noc06]) *Consider an algorithm to minimize the function f, in which there holds at each iteration the inequality (5.1.4). Then if the search directions are selected such that the Zoutendijk condition (5.1.7) holds, either the objective function tends to $-\infty$, or*

$$\inf \|g(x_j)\| = 0, \tag{5.1.8}$$

which means that there is at least one stationary cluster point.

Proof: If the decreasing sequence $\{f(x_j)\}$ is bounded from below by f^*, then from (5.1.4) we have

$$\delta \cos^2 \phi_j \|g(x_j)\|^2 \leq f(x_j) - f(x_{j+1}).$$

By summing the both sides of this inequality from j to ∞, where the coefficient $\delta > 0$ does not depend on j yields

$$\sum_{j=0}^{+\infty} \cos^2 \phi_j \|g(x_j)\|^2 \leq \frac{1}{\delta} \sum_{j=0}^{+\infty} [f(x_j) - f(x_{j+1})] \leq \frac{1}{\delta} [f(x_1) - f^*] < +\infty.$$

As a results, if $|g_j|$ were bounded from below, by $\gamma > 0$, then we would have

$$\gamma^2 \sum_{j=0}^{\infty} \cos^2 \phi_j < +\infty,$$

which is a contradiction to the Zoutendijk condition (5.1.7).

□

Corollary 5.1.12 *If the search directions d_j are chosen such that $|\cos(\phi_j)|$ is bounded away from zero we get*

$$g(x_j) \to 0,$$

which means that all cluster points are stationary.

In other words, we can be sure that the gradient norms $\|g(x_j)\|$ converge to zero, provided that the search directions d_j are never too close to orthogonality with the gradient.

The previous corollary applies to Newton direction provided F' stays nonsingular since

$$\|d_j\| = \|F_j'^{-1} F_j\| \leq \|F_j'^{-1}\| \|F_j\| \implies \frac{1}{\|d_j\|} \geq \frac{1}{\|F_j'^{-1}\| \|F_j\|}.$$

Denoting by $\kappa(A) = \|A\| \|A^{-1}\|$ the condition number of a square matrix A we obtain

$$\cos\phi_j = \frac{\|F(x_j)\|^2}{\|g(x_j)\| \|d_j\|} \geq \frac{1}{\|F'(x_j)^{-1}\| \|F'(x_j)\|} = \frac{1}{\kappa(F_j')} > 0.$$

5.1.5 Backtracking Line-search

When a search direction d_j is available, we would like to determine the step length η_j that is effective. The following backtracking line-search ensures that the conditions (a) and b_1 are satisfied.

Backtracking Line-search Algorithm

Initialize $\eta_0 = 1$.

Given $\sigma \in (0, \frac{1}{2}), r \in (0, 1), 0 < l < u < 1$,

1. While $q(\eta) \geq \sigma$, set $\eta = \frac{\eta}{r}$ with $l \leq r \leq u < 1$.
2. While $q(\eta) \leq \sigma$, set $\eta = \eta r$ with $l \leq r \leq u < 1$.

Termination of Backtracking Line-search

Lemma 5.1.13 *Backtracking line-search terminates with η satisfying (a) and (b_1).*

Proof: If the assertion (1) was executed infinitely often, we would have $f(x+\eta d) \leq f(x)$ with $\eta \to \infty$, which contradicts the assumed compactness of the level set. If the assertion (2) was executed infinitely often, we would have $\eta \to 0$ since it is reduced by $r \leq u < 1$ at each iteration. That would imply $q(\eta) \to 1$, which contradicts the test $q(\eta) \leq \sigma$ being violated infinitely often. □

In the next sub-section we are going to provide a suitable step multiplier η by a new line-search strategy based on parabolic range space interpolation.

5.2 Parabolic Range Space Interpolation

We present in this section a new type of line-search, which is based on the range space interpolation of Wedin, Lindstroem [LW84] to improve Newton's method. We prove that this line-search is efficient in the nonsingular case. Our aim is to apply this modified Newton's method for singular problems, since we found in numerical experiments that it gives better results for singular problems than the unmodified Newton method. Furthermore it is stable under small perturbations of singular problems. This line-search procedure does not require the computation or even the existence of any second derivatives of the original problem.

5.2.1 The Idea of the Method

For a given descent direction d at the point x, we aim to find the multiplier η which enables us to reduce the function $f(x) = \frac{1}{2}\|F(x)\|^2$ in the direction d.

In our line-search we approximate the function

$$\tilde{F}(\eta) \equiv F(x + \eta d), \tag{5.2.1}$$

by a Hermite polynomial $P(\eta)$

$$P(\eta) = \tilde{F}(\eta),$$

and compute the new trial multiplier

$$\eta_* = \arg\min \tfrac{1}{2}\|P(\eta)\|_2^2.$$

We always start with a current tentative step multiplier η_c, which enable us to go from a base point x to a current one $x+\eta_c d$, then perform a parabolic interpolation $P(\eta)$

$$P(\eta) = a_0 + a_1\eta + a_2\eta^2 = \tilde{F}(\eta), \quad a_i \in \mathbb{R}^n, i = 0, 1, 2. \tag{5.2.2}$$

It is based on the initial descent and the two function values of F at x and $x + \eta_c d$. Then we compute a minimizer η_* of the function

$$\psi(\eta) = \tfrac{1}{2}\|P(\eta)\|^2 = \tfrac{1}{2}\|\tilde{F}(\eta)\|^2 \approx \tfrac{1}{2}\|F(x + \eta d)\|^2, \tag{5.2.3}$$

so that $\psi'(\eta_*) = 0$. We accept η_c if it satisfies the line-search condition

$$\frac{2}{3} \le \frac{\eta_*}{\eta_c} \le \frac{9}{8}. \tag{5.2.4}$$

5.2.2 Computation of the Coefficients

The vector coefficients a_i, $i = 0, 1, 2$, of the parabola (5.2.2) are computed as follows.

The first coefficient a_0 can be obtained by substituting $\eta = 0$ in the relation (5.2.2)

$$a_0 = P(0) = \tilde{F}(0) = F(x). \tag{5.2.5}$$

Evaluating the tangent $P'(\eta)$ at $\eta = 0$ yields for the Newton direction d that

$$a_1 = P'(0) = \tilde{F}'(0) \equiv F'(x)^\top d = -F(x) \tag{5.2.6}$$

and, thus, we have

$$a_1 = -a_0 = -F(x).$$

Using $a_0 = F(x)$ and $a_1 = -a_0$, one can find the last coefficient a_2 by requiring $\tilde{F}(\eta) = P(\eta)$ at $\eta = \eta_c$, i.e.

$$\tilde{F}(\eta_c) = P(\eta_c) = a_0 + a_1\eta_c + a_2\eta_c^2 = F(x + \eta_c d),$$

which yields

$$a_2 = \tfrac{1}{\eta_c^2}\left[F(x + \eta_c d) - (1 - \eta_c)F(x)\right]. \tag{5.2.7}$$

Hence, the polynomial $P(\eta)$ (5.2.2) becomes

$$P(\eta) = \frac{\eta^2}{\eta_c^2}[F(x + \eta_c d) - F(x)(1 - \eta_c)] + F(x)(1 - \eta), \tag{5.2.8}$$

with variable coefficient $a_2 = a_2(\eta_c)$ depending on η_c. The latter coefficient can be shown to be bounded.

Lemma 5.2.1 *Provided $x + \eta_c d$ belongs to the compact level set, the coefficient a_2 given by* (5.2.7) *of the parabola $P(\eta) = a_2\eta^2 - a_0\eta + a_0$ is bounded above by $\|a_2\| \le \frac{L}{2}\|d\|^2$.*

Proof: We know from (5.2.7) that a_2 is given by

$$\eta_c^2 a_2 = F(x + \eta_c d) - F(x)(1 - \eta_c),$$

by taking the norm of a_2, and use the expression of Taylor's Theorem we get the following

$$\begin{aligned}
\|\eta_c^2 a_2\| &= \|F(x + \eta_c d) - F(x)(1 - \eta_c)\| \\
&= \left\| F(x) + \int_0^1 F'(x + t\eta_c d)\,\eta_c\, d\, dt - F(x)(1 - \eta_c) \right\|, \qquad \text{For } t \in (0, 1) \\
&= \left\| F(x) + \int_0^1 [F'(x + t\eta_c d) - F'(x)]d\,\eta_c\, dt + F'(x)d\eta_c - F(x)(1 - \eta_c) \right\| \\
&\le \int_0^1 \|[F'(x + t\eta_c d) - F'(x)]d\,\eta_c\|dt, \qquad \text{where } F'(x)d = -F. \\
&\le \tfrac{L}{2}\eta_c^2\|d\|^2,
\end{aligned}$$

yields that the coefficient a_2 is bounded

$$\|a_2\| \leq \frac{L}{2}\|d\|^2.$$

$\square$

We have from (5.2.3) that

$$\psi(\eta) = \frac{1}{2}\|P(\eta)\|^2$$

now we substitute $a_1 = -a_0$ and expanding this function yields

$$\psi(\eta) = \frac{1}{2}\|a_0\|^2 - \|a_0\|^2\eta + \frac{1}{2}(\|a_0\|^2 + 2a_0^\top a_2)\eta^2 - a_0^\top a_2\eta^3 + \frac{1}{2}\|a_2\|^2\eta^4. \qquad (5.2.9)$$

The derivative of $\psi(\eta)$ is

$$\psi'(\eta) = -\|a_0\|^2 + (\|a_0\|^2 + 2a_0^\top a_2)\eta - 3a_0^\top a_2\eta^2 + 2\|a_2\|^2\eta^3. \qquad (5.2.10)$$

Proposition 5.2.2 *The cubic function $\psi'(\eta)$ has a smallest root η_* inside the interval $(0, 2]$ and this root is a minimizer of ψ. Moreover, $\eta_* = \eta_*(\eta_c)$ is a continuous function from $\eta_c \in [0, \infty)$ to $(0, 2]$ with $\lim_{\eta_c \to \infty} \eta_*(\eta_c) = 0$.*

Proof: The polynomial $\psi(\eta)$ has at $\eta = 0$ a positive value $\frac{1}{2}\|a_0\|^2$ and a negative slope

$$\psi'(0) = -\|a_0\|^2 \leq 0.$$

Since $\psi(\eta)$ tends to infinity for $\eta \to 0$, there must exist a first local minimizer $\eta_* > 0$, which is the smallest positive root of $\psi'(\eta) = 0$. Moreover, we see that the derivative is nonnegative at $\eta = 2$

$$\psi'(2) = \|a_0\|^2 - 8a_0^\top a_2 + 16\|a_2\|^2 = (\|a_0\| - 4\|a_2\|)^2 \geq 0,$$

which means that there exists $\eta_* \in (0, 2]$ such that $\psi'(\eta_*) = 0$.

$\square$

Rather than computing η_* for given $\psi(\eta)$ by Cardano's formula we approximate it with full working accuracy by a stabilized Newton's method.

From now on we consider for fixed F, x and Newton direction d the root $\eta_*(\eta_c)$ as a continuous function of η_c.

If η_c is a fixed point such that $\eta_*(\eta_c) = \eta_c$ then the parabolic interpolation suggests that the current point $x + \eta_c d$ achieves the optimal reduction possible. Therefore we can terminate the line-search whenever the ratio $\frac{\eta_*}{\eta_c}$ is sufficiently close to 1. After some experimentation, we settled for the line-search condition (5.2.4). As we will see this condition implies effectiveness in the sense of Definition 5.1.4 on nonsingular regions. On the other hand it favors multipliers η_c that are bigger than 1 and may even come close to 2.

A typical situation is shown in Figure 5.2. We see that the curve $\eta_*(\eta_c)$ must intersect the diagonal $\eta_*(\eta_c) = \eta_c$ at least once due to continuity. Hence the line-search condition (5.2.4) will be satisfied for at least one interval of η_c values. To

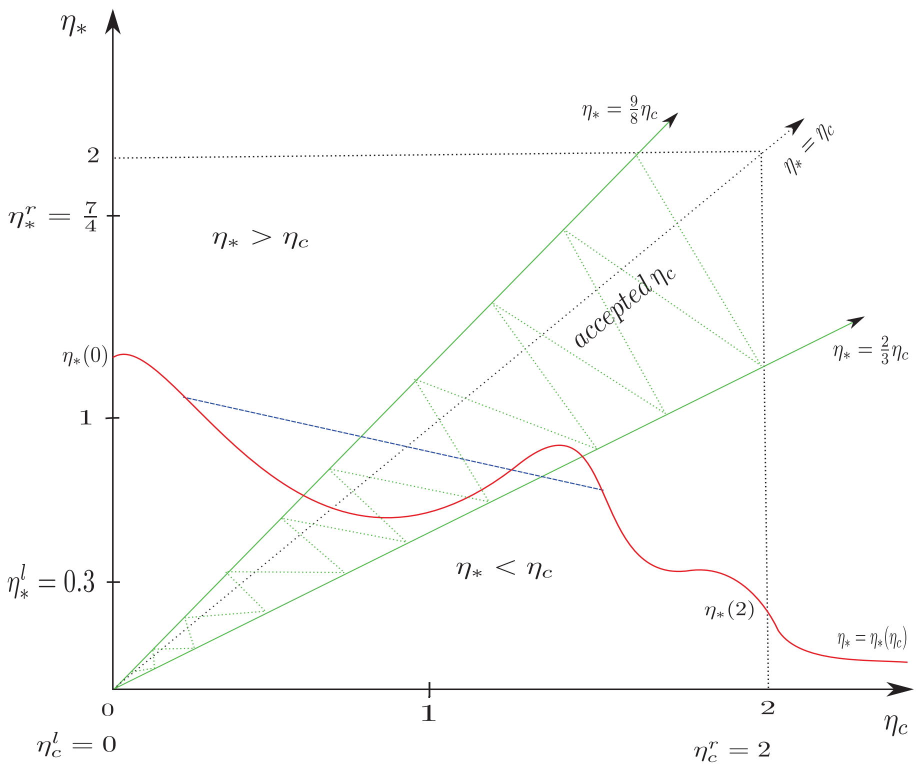


Figure 5.2: Upper and lower bound of PRSI Line-search

compute such acceptable values of η_c we use the following line-search algorithm. We assume the point x and the Newton direction d are given and denote the quadratic residual $\psi(\eta)$ formed at η_c by $\psi_c(\eta)$.

5.2.3 PRSI Line-search Algorithm

Initialize:

(I1) $\eta_c^1 = 1,\ \eta_*^1 = \arg\min \psi_c(\eta)$

(I2) if $(\frac{2}{3} < \eta_*^1 < \frac{4}{3})$ then accept $\eta_c = 1$ and exit.

(I3) $\eta_c = \eta_*^1,\ \eta_* = \arg\min \psi_c(\eta)$

(I4) $\eta_c^\ell = 10^{-5},\ \eta_*^\ell = \eta_*^1 + 1 - \frac{\eta_* - 1}{\eta_*^1 - 1}, \quad \eta_*^\ell = \max\{10^{-5}, \eta_*^\ell\}$

(I5) $\eta_c^r = 2,\ \eta_*^r = \eta_*^1 + 1 - \frac{\eta_* - 1}{\eta_*^1 - 1}, \quad \eta_*^r = \min\{2, \eta_*^r\}$

Loop:

(L1) If $\frac{8}{9} \leq \frac{\eta_*}{\eta_c} \leq \frac{9}{8}$ then accept η_c and exit.

(L2) If $(\eta_* > \eta_c)$

- then $\eta_c^\ell = \eta_c,\ \eta_*^\ell = \eta_*$ and $\eta_*^r = \frac{\eta_*^r + \eta_c^r}{2}$.
- else $\eta_c^r = \eta_c,\ \eta_*^r = \eta_*$ and $\eta_*^\ell = \frac{\eta_*^\ell + \eta_c^\ell}{2}$.

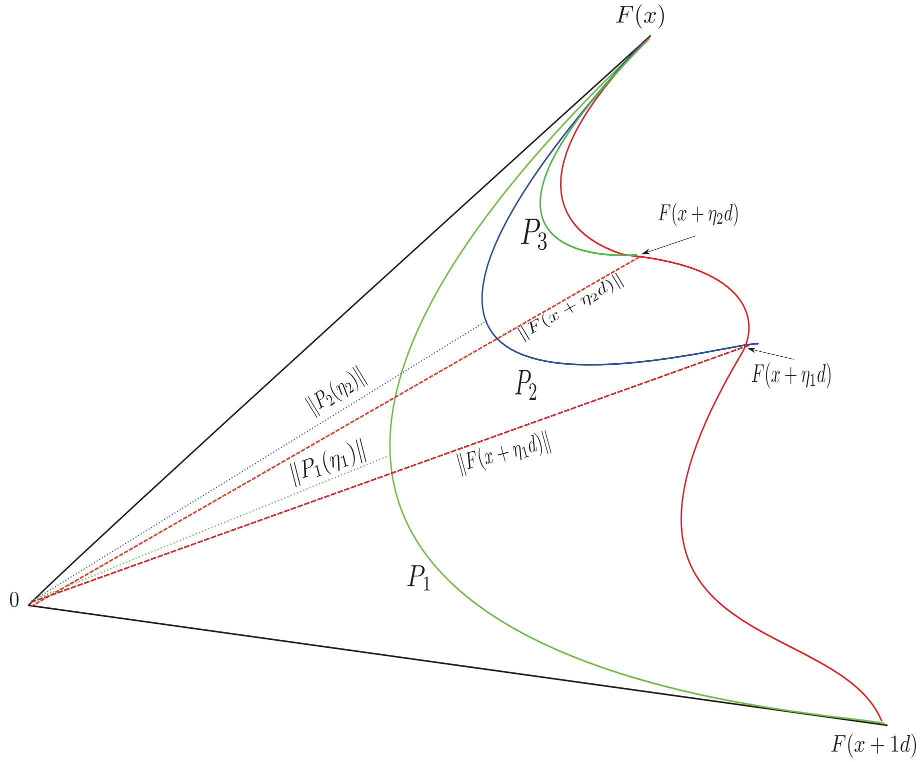


Figure 5.3: PRSI line-search procedure

(L3) Set $\eta_c = \dfrac{\eta_*^\ell \eta_c^r - \eta_*^r \eta_c^\ell}{(\eta_c^r - \eta_*^r) + (\eta_*^\ell - \eta_c^\ell)}$.

(L4) Compute $\eta_* = \arg\min \psi_c(\eta)$ and go to $(L1)$.

The line-search must terminate with a suitable step size η_c, since $\eta_*^r - \eta_*^\ell$ is monotonically reduced by a certain fraction.

The line-search condition is presented graphically in Figure 5.3.

Starting from an initial guess $\eta_0 = 1$, the method generates a sequence of step multipliers $\{\ \eta_j\}_{j=0}^{\infty}$ with the corresponding residuals $F(x+\eta_j d)$ and interpolants $P_j(\eta_j)$ in the range space at a fixed point x and direction d. The line-search procedure is illustrated in Figure 5.3.

In the first step, the function $F(x+\eta d)$ is approximated by the parabola $P_1(\eta)$, which interpolates $F(x)$ and $F(x+\eta_0 d) = F(x+d)$. Since the minimizer $\eta_* = \eta_1$ of $\|P_1(\eta)\|$ does not sufficiently coincide with a minimizer of $\|F(x+\eta d)\|$, i.e. the line-search condition $\frac{2}{3} \leq \frac{\eta_1}{\eta_c} \leq \frac{9}{8}$ failed, one can construct a better approximation $P_2(\eta)$, which now intersects $F(x)$ and $F(x+\eta_1 d)$. The process is repeated until the convergence criterion (5.2.4) is satisfied, as for example in the third step.

5.3 Properties of Line-search in Nonsingular Case

The main goal is, to show that our line-search is effective, and to prove thus it can produce a reduction of $\|F\|$ in the case of F' is invertible in a compact level set containing x_*. to this end, we need to prove some lemmas about the behavior of the derivative $\psi'(\eta)$ of the function $\psi(\eta)$ defined in (5.2.3).

The cubic function $\psi'(\eta)$ is given by (5.2.10) has some properties, which can give us good information's about the minimizer η_*. We present those properties next.

5.3.1 Lower Bound of η

In the next lemma we show that the cubic function has minimizer η_*, which is bounded away from zero. This ensures that the line-search is efficient.

Lemma 5.3.1 *Suppose that there exist a bounded level set*

$$D := \{x \in \mathbb{R}^n \mid \|F(x)\| \leq \|F(x_0)\|\},$$

such that the Jacobian $F'(x)$ is invertible for all $x \in D$ and Lipschitz continuous with Lipschitz constant L. Then the step multiplier η_c satisfying the line-search condition (5.2.4) is uniformly bounded away from zero for $x \in D$.

Proof: We want prove that the minimizer η_* of the cubic function ψ' is bounded below. We apply the minimality condition of $P(\eta)$ given by (5.2.2).

$$\psi'(\eta) =< P(\eta), P'(\eta) \;> = < \; \gamma\eta + a_0,\, 2a_2\eta - a_0 >= 0.$$

Where the coefficient $\gamma = a_2\eta - a_0$ is bounded above, because $a_0 = F(x)$ is bounded above by $\|a_0\| = \|F(x)\| \leq \|F(x_0)\|$ for all $x \in D$.

The norm of the coefficient a_2 is bounded above by $\frac{L}{2}\|d\|^2$ according to Lemma 5.2.1.

By applying the triangular inequality we find

$$\|\gamma\| = \|a_2\eta - a_0\| \leq \|a_2\eta\| + \|a_0\| \leq L\|d\|^2 + \|F(x_0)\|.$$

Then we have from parallelogram equality

$$\|\gamma\eta + 2a_2\eta\| = \|\gamma\eta - 2a_2\eta + 2a_0\|.$$

That yields

$$\|2a_0\| \leq \|\gamma\eta + 2a_2\eta\| + \|\gamma\eta - 2a_2\eta\|.$$

Implies

$$\|a_0\| \leq \;|\gamma\|\eta + 2\|2a_2\|\eta.$$

That gives

$$\|a_0\| \leq \|a_2\eta - a_0\|\eta + 2\|a_2\|\eta.$$

Which implies

$$\|a_0\| \leq \|a_2\|\eta^2 + \|a_0\|\eta + 2\|a_2\|\eta.$$

Which yields

$$\eta^2 + (\tfrac{\|a_0\|}{\|a_2\|} + 2)\eta - \tfrac{\|a_0\|}{\|a_2\|} \geq 0. \tag{5.3.1}$$

The solution η_* of the equation (5.3.1) has two cases

First $\eta_* \geq 1$ then η_* is bounded below by 1.

Second $\eta_* < 1$ then by solve the equation (5.3.1) we find $\eta_* \geq \frac{1}{1+3\|a_2\|/\|a_0\|} \equiv \bar{\eta} > 0$.

Hence η_* is bounded below

$$\eta_* \geq \min(1, \bar{\eta}) > 0. \tag{5.3.2}$$

Then from the line-search condition (5.2.4) yields

$$\eta_c \geq \tfrac{8}{9}\eta_* \geq \tfrac{8}{9}\min(1,\ \bar{\eta}). \tag{5.3.3}$$

The proof is complete. □

Lemma 5.3.2 *Let u, v be vectors such that there exist α and $\beta \in \mathbb{R}$ with*

$$< u + \alpha v,\ u + \beta v >= 0.$$

Then, it holds

$$\|u\| \leq max(|\alpha|, |\beta|)\|v\|.$$

Proof:

The claim follows by applying the Parallelogram equality

$$\|u + \frac{1}{2}(\alpha + \beta)v\|^2 = \|u + \frac{1}{2}(\alpha - \beta)v\|^2.$$

to the assumption via

$$\|u\| \leq \|u + \frac{1}{2}(\alpha + \beta)v\| + \|\frac{1}{2}(\alpha + \beta)v\|.$$

Using the triangle inequality, we get

$$\|u\| \leq \|\frac{1}{2}(\alpha - \beta)v\| + \|\frac{1}{2}(\alpha + \beta)v\|,$$

and, thus,

$$\|u\| \leq \frac{1}{2}(|\alpha + \beta| + |\alpha - \beta|)\|v\| \leq \max(|\alpha|, |\beta|)\|v\|.$$

by the properties of the norm. □

5.3.2 Reduction of $\|F\|$

The idea now is, we consider the line-search parameter η_c as good if the minimizer η_* of the interpolating polynomial is close by. In fact, 'close by' needs not be very close at all.

Proposition 5.3.3 *Under the same assumptions of Lemma 5.3.1 and if η_c satisfies the line-search condition (5.2.4) then it holds that*

$$\|F(x+\eta_c d)\| \leq (1-\frac{\eta_c}{3})\|F(x)\|.$$

This bound ensures a substantial reduction of the residual when $\eta_c \in (0,2]$.

Proof:

We have from the minimality condition of the function $P(\eta)$

$$2\eta_*\psi'(\eta_*) = 2\eta_* P(\eta_*)^\top P'(\eta_*) \;=\; 0 \;= < 2a_2\eta_* - 2a_0\eta_* + 2a_0 \,,\, 2a_2\eta_* - a_0\eta_* >, \tag{5.3.4}$$

from $a_2 = \frac{1}{\eta_c^2}[F_c - (1-\eta_c)F_0]$, where $a_0 = F_0$ we can isolate

$$a_2\eta_c^2 = F_c - (1-\eta_c)F_0.$$

We now define the quotient q with $q\eta_* = \eta_c$ and multiply the second component of (5.3.4) by $q\eta_c = q^2\eta_*$ to yield

$$\begin{aligned} q^2\eta_* P'(\eta_*) = q^2(2a_2\eta_*^2 - a_0\eta_*) &= 2F_c - 2(1-\eta_c)F_0 - q\eta_c F_0 \\ &= 2F_c - 2[1-(1-\frac{1}{2}q)\eta_c]F_0, \quad \eta_c = q\eta_* \\ &= 2F_c - [2-(2-q)\eta_c]F_0 \end{aligned}$$

and multiplying the first component of (5.3.4) by $2q^2$

$$\begin{aligned} 2q^2P(\eta_*) = q^2(2a_2\eta_*^2 + 2(1-\eta_*)a_0 &= 2a_2\eta_c^2 + (1-\eta_*)q^2F_0, \quad \eta_c^2 a_2 = F_c - (1-\eta_c)F_0 \\ &= 2(F_c - (1-\eta_c)F_0) + (1-\eta_*)q^2F_0 \\ &= 2F_c + (q-1)(q+1-\eta_c)F_0. \end{aligned}$$

Thus we find by replacing corresponding components in (5.3.4)

$$\langle 2F_c + 2(q-1)(q+1-\eta_c)F_0 \,,\, 2F_c - [2-(2-q)\eta_c]F_0\rangle = 0,$$

and applying Lemma 5.3.2 that

$$\|F_c\| \leq C(q,\eta_c)\|F_0\| \quad \text{with} \quad C(q,\eta_c) = \max(|q-1||q+1-\eta_c|, |1-(1-\frac{1}{2}q)\eta_c|).$$

The factor $C(q,\eta_c)$ is a convex function in η_c. The linear function through the values at $\eta_c = 0$ and $\eta_c = 2$ is an upper bound for $C(q,\eta_c)$ on $\eta_c \in [0,2]$. With

$|q-1| \leq \frac{1}{3}$ we obtain the value at $\eta_c = 0$ as $C(q,0) = \max(|q^2-1|, 1) = 1$. The value $\eta_c = 2$ is

$$C(q,2) = \max(|q-1|^2, |q-1|) \leq \frac{1}{3}.$$

The linear function defined by these values is $1 + \eta_c \frac{(\frac{1}{3}-1)}{2-0} = 1 - \frac{\eta_c}{3}$ and we get a descent in the function value of at least

$$\|F(x+\eta_c d)\| \leq (1 - \frac{\eta_c}{3})\|F(x)\|.$$

□

5.3.3 Behavior of Line-search near a Root

Lemma 5.3.4 *The parabolic range space interpolation line-search takes the full step $\eta_c = 1$ if the point x is sufficiently close to a nonsingular solution of $F(x) = 0$.*

Proof:

We have from the orthogonality condition of the function $P(\eta)$ that

$$< a_2\eta^2 - a_0\eta + a_0,\ a_2\eta^2 - \frac{\eta}{2}a_0 > = \ 0$$

and, thus,

$$\|2a_2\eta^2 + (1 - \frac{3}{2}\eta)a_0\| = \|(1 - \frac{\eta}{2})a_0\|.$$

The latter formula implies for all $t \in [-\frac{1}{2}, \infty)$ that

$$(1+t)\{|\frac{2t+1}{2}| - |\frac{2t-1}{2}|\}\|a_0\| \leq 2\|a_2\|$$

holds by using the transformation

$$\eta = \frac{1}{1+t}, \quad \text{for} \quad \eta \in (0,2] \tag{5.3.5}$$

and substituting

$$\|2a_2 + ((1+t)^2 - \frac{3}{2}(1+t))a_0\| = \|((1+t)^2 - \frac{1+t}{2})a_0\|.$$

This relation is equivalent to

$$(1+t)\min(2|t|, 1) \leq \frac{2\|a_2\|}{\|a_0\|},$$

which can be further simplified

$$(1+t)\min(2|t|,1) = \begin{cases} (1+t) > |t| & \text{if } t > \frac{1}{2} \\ |t| \geq |t| & \text{if } t \in [-\frac{1}{2}, 0] \\ (1+t)2|t| > |t| & \text{if } t \in (0, \frac{1}{2}). \end{cases}$$

As a consequent we can see that

$$|t| \leq \frac{2\|a_2\|}{\|a_0\|}. \tag{5.3.6}$$

The resubstitution $t = \dfrac{1}{\eta} - 1$ then implies

$$\frac{1}{1 - \frac{2\|a_2\|}{\|a_0\|}} \geq \eta \geq \frac{1}{1 + \frac{2\|a_2\|}{\|a_0\|}}. \tag{5.3.7}$$

Since $\|a_2\| \leq \frac{L}{2}\|d\|^2$ and

$$\|a_2\| \leq \frac{L}{2}\|d\|^2 = \frac{L}{2}\|F'(x)^{-1}F(x)\|^2 \leq \frac{L}{2}\|F'(x)^{-1}\|^2\|a_0\|^2$$

it follows that the denominator of the lower and upper bound of (5.3.7) tend to one as x approaches a nonsingular solution. Therefore, η gets 1 and the line-search condition (5.2.4) will be fulfilled. $\square$

5.4 Properties of Line-search in Singular Case

As we have detailed in Section 3.2 the unmodified Newton's method converges to regular singularities of order k with the linear rate $\frac{k}{k+1}$. More specifically the convergence Theorem 3.2.2 states that during the final approach the Newton steps tend to aim in the right direction towards the root but never go quite far enough. In the simple case $k = 1$ they stop short exactly half way to the root. Stabilization in the sense of reducing the step length below 1 by a conventional line-search or bounding the Jacobian inverse will only deteriorate the situation, quite likely resulting in a sublinear rate of convergence. In contrast it was established theoretically and confirmed numerically that nearly doubling every other Newton step can recover a superlinear rate of convergence when $k = 1$.

Before analyzing the behavior of the range space line-search, let us briefly consider the effect of a classical parabolic interpolation of the merit function $f(x) = \frac{1}{2}\|F(x)\|^2$ defined in Section 5.1.1 near a simple singularity. Given the current point x and the Newton step d, one would obtain by quadratic interpolation, matching the values at $\eta = 0$ and $\eta = 1$ and the slope at $\eta = 0$.

$$f(x + \eta d) \;\approx\; \|F(x)\|^2(\tfrac{1}{2} - \eta) + \tfrac{1}{2}\left[\|F(x + d)\|^2 + \|F(x)\|^2\right]\eta^2$$

The minimum of the right hand side is attained at

$$\eta_* \;=\; \frac{1}{1 + \|F(x + d)\|^2/\|F(x)\|^2} \;\leq\; 1,$$

so that we can never achieve an overrelaxed step. If $x - x_*$ is close to the nullspace of the Jacobian at x_* we know from the assertions (iv) and (i) of convergence Theorem 3.2.2 that the residual is approximately quarter so that $F(x + d) \approx \frac{1}{4}F(x)$, which means that $\eta_* \approx \frac{16}{17}$. Hence the line-search based on quadratic interpolation of the sum of squares residual $f(x)$ has almost no effect on the convergence of Newton's method to simple singularities. Things are dramatically different for the range space line-search.

When $\eta_c = 1$ then the range space interpolant given in (5.2.8) takes the simple form

$$P(\eta) \;=\; F(x)(1 - \eta) + F(x + d)\eta^2$$

Assuming without loss of generality that $\|F(x)\| = 1$ and abbreviating $\|F(x + d)\| = r$ and $\phi = \arccos(\frac{1}{r}F(x)^\top F(x + d))$ and substituting into (5.2.9), we obtain the residual

$$\psi(\eta) = \tfrac{1}{2} - \eta + (\tfrac{1}{2} + r\,\cos\phi)\eta^2 - r\,\cos(\phi)\eta^3 + \tfrac{1}{2}r^2\eta^4$$

and for the minimizer $\eta_* = \eta_*(r, \phi)$ the stationarity condition

$$0 \;=\; -1 + (1 + 2\,r\,\cos\phi)\eta_* - 3\,r\,\cos(\phi)\eta_*^2 + 2r^2\eta_*^3$$

Now suppose we keep η_* fixed and consider the corresponding contours with respect to the polar coordinates (r, ϕ) in the plane.

Lemma 5.4.1 *For $\eta_* \in (0, 2]$ the contours $\{(r, \phi) \;:\; \eta_*(r, \phi) = \eta_*\}$ are circles with the radius $\rho = \frac{1}{2\eta_*}(\frac{1}{\eta_*} - \frac{1}{2})$ and the centers at the points $\left[\frac{1}{2\eta_*}(\frac{3}{2} - \frac{1}{\eta_*}), 0\right]$ in Cartesian coordinates.*

Proof:

With $\bar{r} = r\eta_*$ the stationarity condition becomes

$$0 \; = \; -1 + (\eta_* + 2\,\bar{r}\,\cos(\phi)) - 3\,\bar{r}\,\cos(\phi)\eta_*^2 + 2\bar{r}^2\eta_*$$

and thus

$$\bar{r}^2 + \bar{r}(\tfrac{1}{\eta_*} - \tfrac{3}{2})\cos(\phi) = \tfrac{1}{2\eta_*} - \tfrac{1}{2}$$

this is one form of a circle equation. On both sides of the η axis we find apposite path of the circle of $\cos(\phi) = \pm 1$ so that

$$(\bar{r} \;\pm\; \tfrac{1}{2}(\tfrac{1}{\eta_*} - \tfrac{3}{2}))^2 = \tfrac{1}{2\eta_*} - \tfrac{1}{2} + \tfrac{1}{4}(\tfrac{1}{\eta_*} - \tfrac{3}{2})^2 = -\tfrac{1}{4\eta_*} + \tfrac{1}{16} + \tfrac{1}{4\eta_*^2} = (\tfrac{1}{2\eta_*} - \tfrac{1}{4})^2$$

which implies

$$\bar{r} \;=\; \mp\,(\tfrac{1}{2\eta_*} - \tfrac{3}{4}) + |\tfrac{1}{2\eta_*} - \tfrac{1}{4}|$$

We have $\eta_* \leq 2 \implies (\frac{1}{2\eta_*} - \frac{1}{4}) \geq 0 \implies |\frac{1}{2\eta_*} - \frac{1}{4}| = \frac{1}{2\eta_*} - \frac{1}{4}$ which yields

$$\bar{r} \;=\; \mp\,(\tfrac{1}{2\eta_*} - \tfrac{3}{4}) + \tfrac{1}{2\eta_*} - \tfrac{1}{4}$$

Now we have two cases

$$\bar{r} = +(\tfrac{1}{2\eta_*} - \tfrac{3}{4}) + \tfrac{1}{2\eta_*} - \tfrac{1}{4} \implies \bar{r} = \tfrac{1}{\eta_*} - 1 \implies r = \tfrac{1}{\eta_*^2} - \tfrac{1}{\eta_*}$$

$$\bar{r} = -\,(\tfrac{1}{2\eta_*} - \tfrac{3}{4}) + \tfrac{1}{2\eta_*} - \tfrac{1}{4} \implies \bar{r} \;=\; \tfrac{1}{2} \implies r \;=\; \tfrac{1}{2\eta_*}$$

That is

$$r \;\in\; \left\{\tfrac{1}{\eta_*} - \tfrac{1}{\eta_*^2}\;,\;\tfrac{1}{2\eta_*}\right\}. \tag{5.4.1}$$

By averaging and subtracting the lower and upper bounds on r one obtains the center $(\frac{3}{4\eta_*} - \frac{1}{2\eta_*^2})$ and the width $(\frac{1}{\eta_*^2} - \frac{1}{2\eta_*})\frac{1}{2}$respectively.

$\square$

Solving (5.4.1)for η_* in terms of r we obtain the spiked function

$$\eta_* \;=\; \begin{cases} 2\Big/(1+\sqrt{1-4r}) & if\ r \leq \frac{1}{4} \\ \frac{1}{2r} & if\ r \geq \frac{1}{4} \end{cases}$$

The situation is depicted in Figure 5.4. As one can see the spike occurs exactly at $r = \frac{1}{4}$ i.e., when the full Newton step leads to the quartering of the residual $F(x)$ to $F(x+d) = \frac{1}{4}F(x)$. Then and only then the suggested step multiplier η_* reaches its maximal value 2. To come even close to doubling the step the quartering must occur quite accurately. The lower bound on r for given η_* has vertical slope at $\eta_* = 2$. The center line in the middle between the lower and upper bound is drawn in red, the width labeled by ρ.

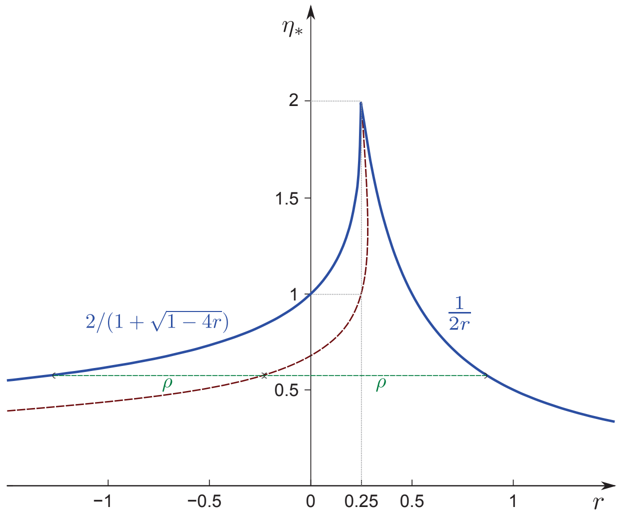


Figure 5.4: Step multiplier $\eta_* = \eta_*(r, 0)$.

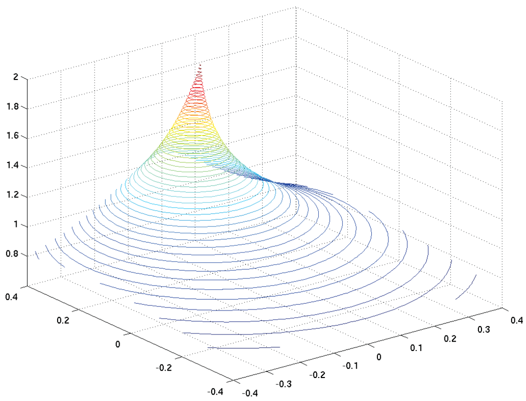


Figure 5.5: The level set of the Multiplier Mountain.

A three dimensional version, which we call the multiplier mountain is depicted in Figure 5.5 where one can clearly see the circular contours. The border line between extending and reducing the step multiplier is the circle with radius $\frac{1}{2}$ centered at the point $(\frac{1}{4}, 0)$.

In our numerical experiments to be reported later we observed that the line-search does indeed overrelaxed steps when the root looks singular in that $r \approx \frac{1}{4}$ and $\phi \approx 0$. Unfortunately, we have not been able to demonstrate that rigorously.

5.5 Cubic Extension

With the hope of also accelerating Newton's method at second order singularities we considered the generalization of the range space interpolation from quadratic to cubic. As additional information we used the directional derivative $F'(x+\eta_c)d$, which can be computed quite cheaply by forward differentiation or approximated by differencing. The interpolating polynomial is obtained via

$$P(\eta) = a_0(1-\eta) + a_2\eta^2 + a_3\eta^3 \approx F(x+\eta d),$$

and

$$P'(\eta) = -a_0 + 2a_2\eta + 3a_3\eta^2 \approx F'(x+\eta d)\, d,$$

by solving the two relations above and substituting $a_0 = F(x)$ and $a_1 = -F(x)$ we obtain

$$a_2 = \tfrac{1}{\eta_c^2}\left[3F(x+\eta_c d) - \eta_c F'(x+\eta_c d)d + (2\eta_c - 3)F(x)\right],$$

$$a_3 = \tfrac{1}{\eta_c^3}\left[-2F(x+\eta_c d) + \eta_c F'(x+\eta_c d)d + (2-\eta_c)F(x)\right],$$

Now the norm

$$\psi(\eta) = \tfrac{1}{2}\|P(\eta)\|_2^2,$$

is a polynomial of degree 6 and its derivative $\psi'(\eta)$ is quintic. Of its five roots three may be minimizers of $\psi(\eta)$. We have implemented this approach using a line-search criterion similar to (5.2.8).

Unfortunately we obtained a nonsignificant gain at second order singularities compared to parabolic interpolation and the two yield almost identical results at first order singularities (see Figure 6.3 and 6.4 for Bratu problem.)

Hence the extra conceptual and computational effort does not seem justified.

Chapter 6

Numerical Results

In this section we report the numerical results obtained for a set of test problems. We introduce the following notations:

- $p_j = \frac{\eta_{j+1}}{\eta_j}\frac{\|d_{j+1}\|}{\|d_j\|}$ is the reduction ratio, where d_j is the Newton direction and η_j is the step multiplier of line-search.
- ls is the number of iterations per line-search.
- The stopping criterion is $\|F(x)\| < 10^{-11}$.

6.1 Nonsingular case

We use as a test problem for the nonsingular case the function

$$F(x,y) = \begin{pmatrix} -2x+3y+4y^2+x^2+x^2y+x^3 \\ x-2y+y^2+3yx^2+xy^2+y^3 \end{pmatrix}. \tag{6.1.1}$$

One nonsingular root of this function is $(0,0)$. The Jacobian of F at $(0,0)$ is $\begin{pmatrix} -2 & 3 \\ 1 & -2 \end{pmatrix}$.

Table 6.1 shows the behavior of the step multiplier η_j, for $x_0 = (0.1, 0.03)$.

j	ls	η_j	x_j	$\|F(x_j)\|$	p_j
1	1	0.5642815848	(-0.05775922926,-0.03911510821)	0.02267079486	0.05741164477
2	1	1	(-0.01412698875,-0.007757523356)	0.005606032978	0.311966136
3	1	1	(-0.0009284100954,-0.0004937610213)	0.0003820122093	0.2803822273
4	1	1	(-4.365156722e-06,-2.304270042e-06)	1.833767385e-06	0.06947122653
5	1	1	(-9.651125467e-11,-5.091043654e-11)	4.063954752e-11	0.004716097878
6	1	1	(-4.71394463e-20,-2.486565021e-20)	1.985186508e-20	2.210656984e-05

Table 6.1: The behavior of η_j for $x_0 = (0.1, 0.03)$, the nonsingular case.

x_0	PRSI iterations	Newton iterations
(0.1, 0.03)	6	7
(-1, -0.01)	9	10
(0.0189, 0.0625)	6	7

Table 6.2: The Number of iterations of the pure Newton's method and using PRSI line-search, for different initial points, for the nonsingular solution $(0, 0)$.

In Table 6.2 we compare the full step Newton's method and parabolic range space interpolation Newton's method for different initial points. Here we find that the modified Newton iterations are always faster than the full step one when converging to the same solution.

Only some early steps are extended before the full-step convergence pattern settles in.

For this and all other problems, we have visualized the behavior of Newton's method and its variants by so-called fractals. In Subsection 6.5.1, the Pictures 6.7, 6.8, 6.9 are the fractals of the problem (6.1.1). As we can see there is a second, nonsingular root far away from (nearly) singular solutions at or near the origin.

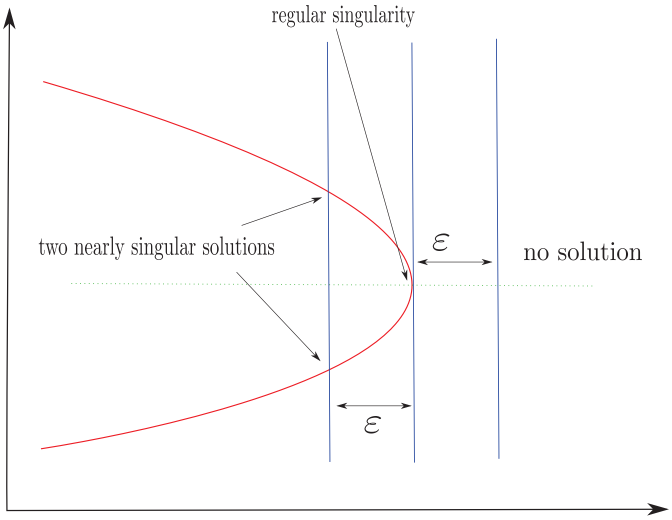


Figure 6.1: Perturbation of the Problem

6.2 (Nearly) Singular Case

The parameter depended function

$$F(x,y) = \begin{pmatrix} y + xy + y^2 + 0.1x^2 + 1.1x^3 + yx^2 \\ x^2 + y^2 + yx + 0.2x^3 + 1.2y^3 + xy^2 + \varepsilon \end{pmatrix}. \tag{6.2.1}$$

with $\varepsilon \in \mathbb{R}$, has a regular singularity at $(0,0)$ if $\varepsilon = 0$. The Jacobian of F at this root is $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. The root $(0,0)$ is a regular singularity because an elementary calculation yields that

$$\nabla \det(F')^\top v = -2 \neq 0.$$

Here $v = \binom{1}{0}$ is a nullvector of F' at the root $(0,0)$.

By variation of ε the above problem yields the following cases depicted in Figure 6.1.

1. If $\varepsilon < 0$ then we have two nearly singular solutions.
2. If $\varepsilon > 0$ then we have no solution but a nonzero minimum of $\|F\|$.
3. If $\varepsilon = 0$ then we have a regular singularity.

Tables 6.3 and 6.4 show the behavior of the step multiplier η_j for the small negative perturbation $\varepsilon = -10^{-5}$. Starting from $x_0 = (-1, 1)$ the x_j reach the solution $x_* = (-0.0031628, -9.6858e - 07)$ after 10 iterations. If the initial point is $x_0 = (1, 1.5)$ then x_j converge to the nearly singular solution $x_* = (0.0031618, -1.0312e{-}06)$ using 18 iterations.

Table 6.5 shows the behavior of the step multiplier η_j for the small positive perturbation $\varepsilon = 10^{-5} > 0$. If we start from $x_0 = (-0.5, -1.5)$ then the iterates converge to $x_* = (-0.69461, -1.0836)$ after 5 iterations.

j	ls	η_j	x_j	$\|F(x_j)\|$	p_j
1	1	1	(-0.89072, 0.59017)	0.44999	0.14138
2	1	1	(-0.44991, 0.083541)	0.15148	1.5833
3	1	1	(-0.13978, -0.10323)	0.091093	0.53908
4	1	1	(-0.11956, 0.018525)	0.020399	0.34093
5	1	1	(-0.054481, -0.0017575)	0.0033962	0.55232
6	1	1	(-0.027624, -4.6793e-05)	0.00075023	0.39476
7	2	1.7661	(-0.0034957, 1.5086e-05)	1.6351e-05	0.89661
8	1	1	(-0.0031779, -9.7355e-07)	9.6035e-08	0.013188
9	1	1	(-0.0031628, -9.6858e-07)	2.2873e-10	0.04747
10	1	1	(-0.0031628, -9.6858e-07)	1.3017e-15	0.002386

Table 6.3: The behavior of η_j for the initial point $x_0 = (-1, 1)$, for $\varepsilon = -10^{-5} < 0$. First solution $x_* = (-0.0031628, -9.6858e-07)$.

j	ls	η_j	x_j	$\|F(x_j)\|$	p_j
1	2	1.6188	(-0.028519, 0.72648)	1.5593	0.42898
2	6	0.10842	(0.25501, 0.59158)	1.4689	0.24398
3	6	0.021341	(0.35522, 0.53533)	1.4523	0.36599
4	6	0.0068093	(0.40813, 0.50319)	1.447	0.53875
5	6	0.004623	(0.4502, 0.47597)	1.4434	0.80935
6	6	0.0064532	(0.49844, 0.44236)	1.4385	1.1734
7	5	0.02614	(0.59683, 0.36373)	1.4212	2.1422
8	1	1	(0.86404, -0.2598)	0.77553	5.386
9	1	1	(0.46299, 0.026967)	0.30425	0.72678
10	2	1.402	(0.13202, 0.046842)	0.065944	0.67251
11	1	1	(0.082098, 0.0040746)	0.0092001	0.19825
12	1	1	(0.042285, 0.00042114)	0.0019424	0.6082
13	2	1.5089	(0.010815, -0.00014623)	0.00017121	0.78728
14	1	1	(0.0058423, -1.1218e-06)	2.4295e-05	0.15806
15	1	1	(0.0037775, -9.8392e-07)	4.3054e-06	0.41504
16	1	1	(0.003212, -1.0291e-06)	3.2244e-07	0.27385
17	1	1	(0.0031622, -1.0312e-06)	2.5042e-09	0.088145
18	1	1	(0.0031618, -1.0312e-06)	1.5589e-13	0.0078901

Table 6.4: The behavior of η_j for the initial point $x_0 = (1, 1.5)$, for $\varepsilon = -10^{-5} < 0$. Second solution $x_* = (0.0031618, -1.0312e-06)$.

j	ls	η_j	x_j	$\|F(x_j)\|$	p_j
1	1	1	(-0.92615, -1.1873)	0.49591	0.17619
2	2	1.4722	(-0.68943, -1.0816)	0.0069604	0.49046
3	1	1	(-0.69468, -1.0836)	9.0464e-05	0.021613
4	1	1	(-0.69461, -1.0836)	1.5294e-08	0.013084
5	1	1	(-0.69461, -1.0836)	3.6802e-16	0.0001719

Table 6.5: The behavior of η_j for given initial point, for $\varepsilon = 10^{-5} > 0$.

j	ls	η_j	x_j	$\|F(x_j)\|$	p_j
1	2	1.7797	(0.38586, -0.044846)	0.14619	0.27366
2	1	1	(0.18833, 0.014208)	0.048879	0.25113
3	2	1.7101	(0.032927, -0.0030635)	0.0031724	0.75837
4	1	1	(0.015888, -2.0202e-05)	0.00025306	0.1107
5	2	1.646	(0.0028233, 1.6312e-05)	1.8961e-05	0.75476
6	1	1	(0.0014161, 3.2979e-08)	2.02e-06	0.10772
7	2	1.8759	(8.8053e-05, -2.5976e-08)	2.6367e-08	0.94374
8	1	1	(4.402e-05, -6.6021e-13)	1.9474e-09	0.033155
9	2	1.9938	(1.358e-07, 7.4796e-13)	7.5003e-13	0.99664

Table 6.6: Singular case, The behavior of η_j for given initial point, for $\varepsilon = 0$.

x_0	j_{PRSI}	j_{FSN}
(1, 0.5)	9	20
(1, 1.5)	18	22
(-0.493259, -0.369245)	9	19
(1.57571, -0.61938)	10	20
(0.980752, 0.176084)	13	23

Table 6.7: Singular case, The number of iterations of full Newton step and PRSI line-search for different initial points, for $\varepsilon = 0$.

Table 6.6 shows the behavior of the step multiplier η_j in the singular case without perturbation, i.e., $\varepsilon = 0$. When one starts from $x_0 = (1, 0.5)$ then the regular singularity solution is $x_* = (0, 0)$. In this case the fractals are given in the Pictures 6.13, 6.14, 6.15.

In Table 6.7 we compare the full Newton step and Newton's method with PRSI line-search for different initial points in the case $\varepsilon = 0$ and for the regular singularity solution $(0, 0)$. As we can see from the table Newton's method with PRSI line-search needs approximately half of the iterations of the full step Newton's method.

As we can see for $\varepsilon < 0$ there is some early acceleration when the two close-by roots look like a single singular root. Then the full-step convergence settles in.

In the exactly singular case $\varepsilon = 0$ we see significant acceleration throughout the iteration. At the end it looks as through a single unit step interspersed with extended steps is sufficient to realign the the iterates along the nullvector.

In the case $\varepsilon > 0$ the iterations diverted to a nonsingular root far from the origin.

The fractals of the problem (6.2.1) for the case $\varepsilon = 0$ are given in Subsection 6.5.1.

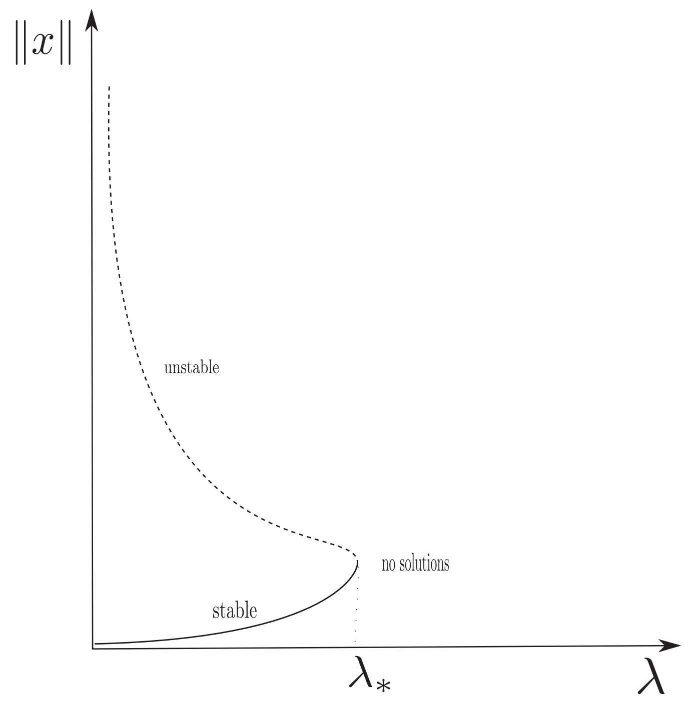


Figure 6.2: Fold bifurcation (turning point) at $\lambda = \lambda^*$, solid line represents stable solutions, dashed line indicates unstable solutions.

6.3 Bratu problem

We apply the new method for a large number of variables to solve a discretized Bratu problem on the unit square with an equidistant decomposition into $n = M^2$ parts in each direction, that is, a mesh width $h = \frac{1}{M}$.

The Bratu problem is the nonlinear partial differential equation:

$$\Delta u + \lambda e^u = 0, \text{ on } [0,1] \times [0,1], \quad u = u(x,y), \tag{6.3.1}$$

with boundary conditions

$$u(0,y) = u(1,y), \quad u(x,0) = \sin(2\pi x), \quad u(x,1) = w(x) = 2.2,$$

The Bratu problem has two solutions if $\lambda < \lambda^*$, no solutions if $\lambda > \lambda^*$ and a unique singular solution if $\lambda = \lambda^*$, as presented in Figure 6.2.

The Jacobian is very sparse with the typical structure for $2D$ partial differential equations. Therefore the arising linear systems can be solved efficiently for large M by special software UMFPACK [Dav].

In the Diagrams 6.3 and 6.4 we display along the vertical axis the number of iterations needed by Newton's method to converge as function of λ. Here the λ-axis is scaled logarithmically according to the function

$$-\log_{10}(\lambda^* - \lambda).$$

The exact value of λ^* depends on the grid number M. Hence, for each value of M we have a new value of λ^*, which was computed by nested bisection.

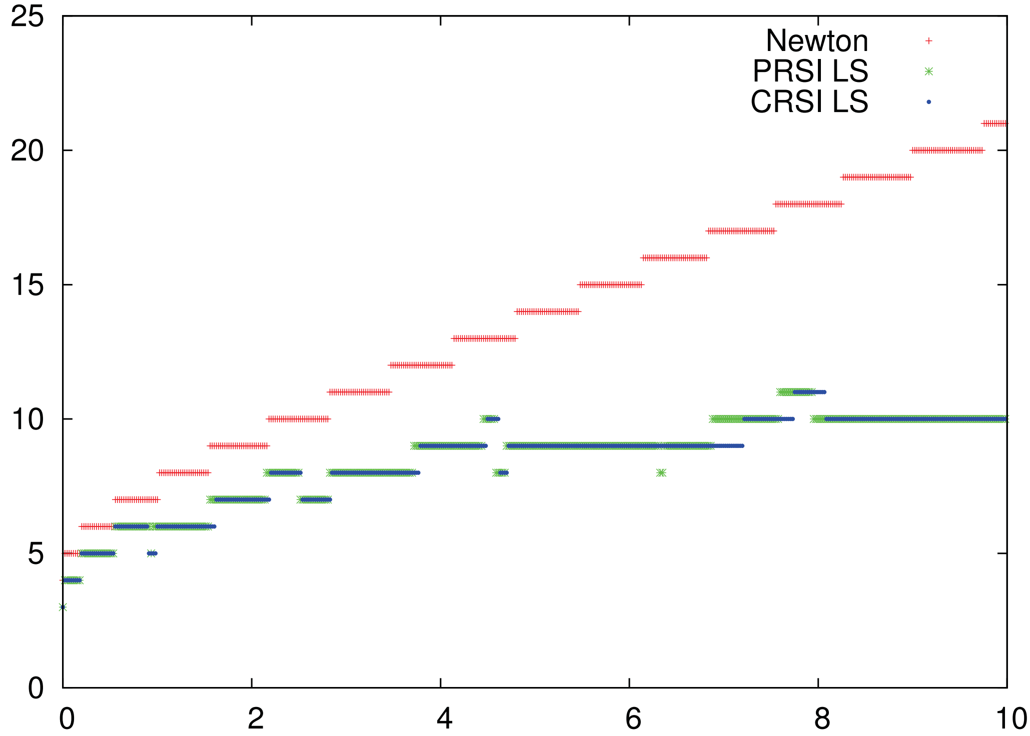


Figure 6.3: Comparing Newton's method without and with quadratic or cubic line-search on Bratu problem with $M = 12$, $n = 144$ variables, where $\lambda^* = 1.022057436608385$

The Picture 6.3, 6.4 show the behavior of the modified and unmodified Newton's method, where the red color refers to full-step Newton, green to the parabolic and blue to the cubic range space interpolation line-search.

We observe that Newton's method with full step needs more iterations than with either line-search. The iteration count of the full step method grows approximately linearly with the logarithm of the difference between λ^* and λ. In contrast for both line-search variant we note that the iteration counts stay approximately constant as λ approaches λ^*. The extra effort for performing the cubic interpolation compared to the quadratic one does not seem to pay off. Throughout we use $\|F\| \leq 10^{-11}\sqrt{n}$ as stopping criterion for the Bratu problem.

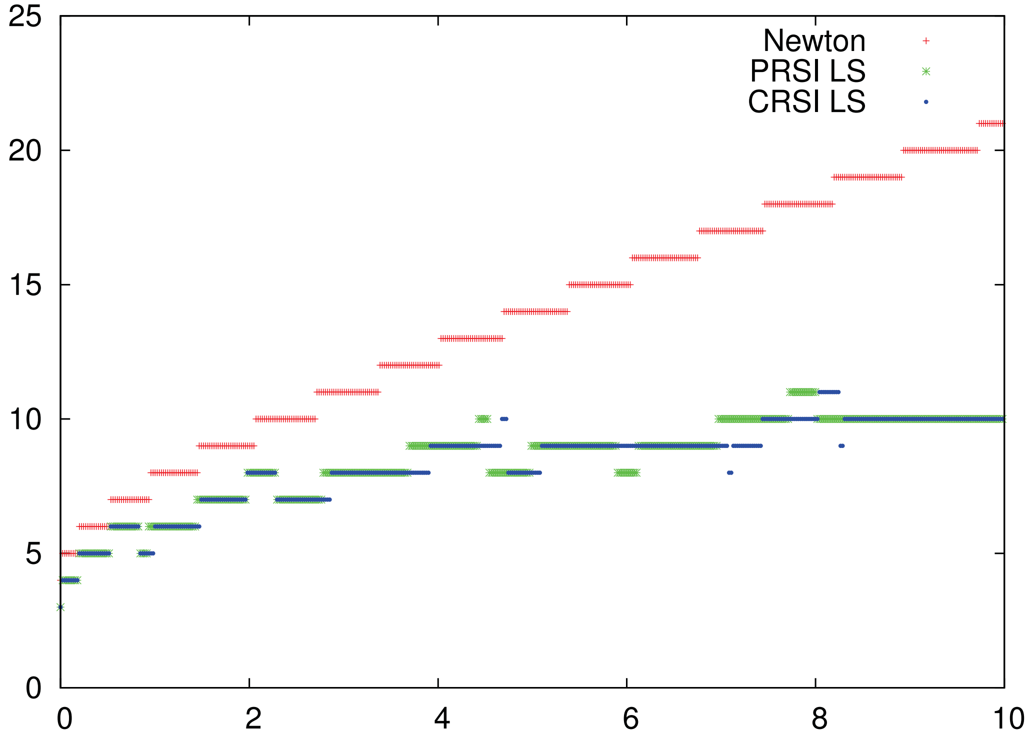


Figure 6.4: Comparing Newton's method without and with quadratic or cubic line-search on Bratu problem with $M = 40$, $n = 1600$ variables, where $\lambda^* = 1.025046903052621$

6.4 Complementarity Problems

We apply the new method to an equations reformulation of nonlinear complementarity problems, whose derivative is well defined and strongly semismooth at x_* if the function F defining the NCP is sufficiently smooth. Conditions on F are derived that ensure that the appropriate regularity conditions are satisfied for the nonlinear reformulation of the NCP at x_*.

Definition 6.4.1 *A nonlinear complementarity problem NCP is defined by a mapping $F(x) = (F_i(x))_{(i=1\dots n)} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$. A solution is an $x \in \mathbb{R}^n$ satisfying*

$$x \geq 0, \quad F(x) \geq 0, \quad x^\top F(x) = 0. \tag{6.4.1}$$

Definition 6.4.2 *A solution x is degenerate if both $F_i(x)$ and the i-th component x_i vanish for some index i, i.e. there is no strict complementarity.*

We discuss a nonlinear equations reformulation to solve the nonlinear complementarity problem NCP, which involves reformulating the problem as system of nonlinear equations. This involves constructing a function $\Psi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ with the property that zeros of Ψ correspond to solutions of the NCP. Such a function is called NCP-function [BM00].

Definition 6.4.3 *The function Ψ is defined by*

$$\Psi_i(x) = \psi(x_i, F_i(x)), \tag{6.4.2}$$

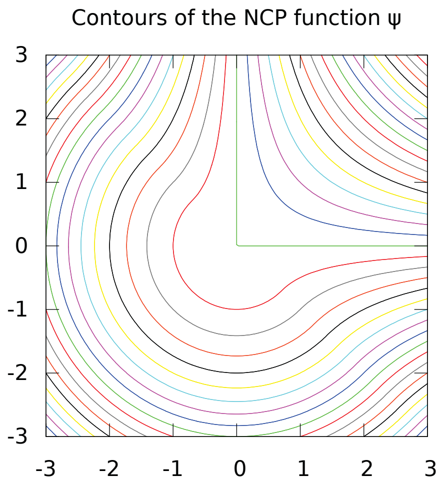


Figure 6.5: The Level curves of the Function ψ

where

$$\psi(a, b) := 2ab - (\min(0, a + b))^2. \tag{6.4.3}$$

This function has the property that

$$\psi(a, b) = 0 \quad \text{if and only if} \quad a \geq 0, \quad b \geq 0 \quad \text{and} \quad ab = 0. \tag{6.4.4}$$

The Jacobian of Ψ is

$$\Psi'(x) = (\psi_i')_{i=1}^n, \tag{6.4.5}$$

where

$$\psi_i' = \begin{cases} 2x_i F_i' + 2F_i e_i^\top & \text{if} \quad x_i + F_i \geq 0 \\ -2x_i e_i^\top - 2F_i F_i' & \text{if} \quad x_i + F_i \leq 0 \end{cases} \tag{6.4.6}$$

This reformulation is apparently due to Evtushenko [EP84]. One can easily see that the function Ψ has a singular Jacobian if there is no strict complementarity.

We apply the new method to find a solution for a nonlinear equations reformulation $\Psi(x) = 0$. Of particular interest are solution x_* at which due to lack of strict complementarity ψ_i' is singular and only semismooth.

We describe here some computational results from the simple NCP test problems of small dimensions given in [OW09]. A solution is any x satisfying

$$0 \leq x \perp F(x) \geq 0,$$

and we denote such x by x_*.

1. aff1

$$F(x) = \begin{pmatrix} x_1 + 2x_2 \\ x_2 - 1 \end{pmatrix} \tag{6.4.7}$$

2. DIS61 ([DIS03], Example 6.1)

$$F(x) = \begin{pmatrix} (x_1 - 1)^2 \\ x_1 + x_2 + x_2^2 - 1 \end{pmatrix} \tag{6.4.8}$$

3. quarquad

$$F(x) = \begin{pmatrix} -(1 - x_1)^4 + x_2 \\ 1 - x_2^2 \end{pmatrix} \tag{6.4.9}$$

4. affknot1

$$F(x) = \begin{pmatrix} x_2 - 1 \\ x_1 \end{pmatrix} \tag{6.4.10}$$

5. affknot2

$$F(x) = \begin{pmatrix} x_2 - 1 \\ x_1 + x_2 - 1 \end{pmatrix} \tag{6.4.11}$$

6. quadknot

$$F(x) = \begin{pmatrix} x_2 - 1 \\ x_1^2 \end{pmatrix} \tag{6.4.12}$$

7. munson4 (from MCPLIB [MCP])

$$F(x) = \begin{pmatrix} -(x_2 - 1)^2 \\ -(x_1 - 1)^2 \end{pmatrix} \tag{6.4.13}$$

8. ne-hard (from MCPLIB [MCP])

$$F(x) = \begin{pmatrix} \sin x_1 + x_1^2 \\ x_2^3 + x_1 x_3 \\ x_3^2 - 200 + x_1 x_2 \end{pmatrix} \tag{6.4.14}$$

9. doubleknot

$$F(x) = \begin{pmatrix} 1 - x_1 + x_2 + x_3 \\ x_1 - 1 \\ x_4 - 1 \\ 1 + x_3 - x_4 \end{pmatrix} \tag{6.4.15}$$

10. quad1

$$F(x) = \begin{pmatrix} x_1 - 1 \\ x_2^2 \end{pmatrix} \tag{6.4.16}$$

11. quad2

$$F(x) = \begin{pmatrix} x_1^2 \\ x_2 \end{pmatrix} \tag{6.4.17}$$

Problem	x_0	x_*
quarquad	(0.1, 0.9)	(0, 1)
affknot1	(0.9, 0.1)	(0, 1)
affknot2	(0.5, 0.5)	(0, 1)
quad2	(-1, -1)	(0, 0)
quad1	(0.9, 0.1)	(1, 0)
quadknot	(0.5, 0.5)	(0, 1)
munson4	(0, 0)	(1, 1)
DIS61	(1.5, -0.5)	(1, 0)
ne-hard	(10, 1, 10)	(0, 0, $\sqrt{200}$)
doubleknot	(0.5, 0.5, 0.5, 0.5)	(1, 0, 0, 1)

Table 6.8: Starting points and solutions for given problem.

				O&W Accelerated		PRSI line-search		
problem	n	dim N	FNS iters	total iters	$\eta = 1.9$	total iters	$\eta > 1$	Add. Eval.
quarquad	2	1	16	10	5	6	2	2
affknot1	2	1	20	10	7	2	1	1
affknot2	2	1	19	10	5	1	1	1
quad2	2	2	20	13	4	4	4	4
quad1	2	1	15	9	4	7	3	3
quadknot	2	2	18	8	5	4	4	5
munson4	2	2	19	12	4	6	3	3
DIS61	2	2	19	12	5	4	4	4
ne-hard	3	2	25	19	5	12	8	4
doubleknot	4	2	22	14	5	11	4	5

Table 6.9: Comparison of the number of iterations for the Full Newton Step, accelerated Newton's method of Wright et al., with $\eta = 1.9$ on accelerated steps, PRSI iterations with $0 < \eta \leq 2$.

j	ls	η_j	x_j	$\|F(x_j)\|$	p_j
1	2	1.6081	(0.017285, 0.92795)	0.14858	0.48477
2	1	1	(-0.0012705, 1.0061)	0.013231	0.05521
3	1	1	(-6.8204e-06, 1)	7.7658e-05	0.076725
4	1	1	(-2.2466e-10, 1)	2.7914e-09	0.0060037
5	1	1	(-2.7164e-19, 1)	1.0866e-18	3.6241e-05

Table 6.10: The behavior of the step multiplier η of PRSI line-search method for NCP with strict complementarity, e.g. the function aff1 (6.4.7) for initial point $(1, 2)$.

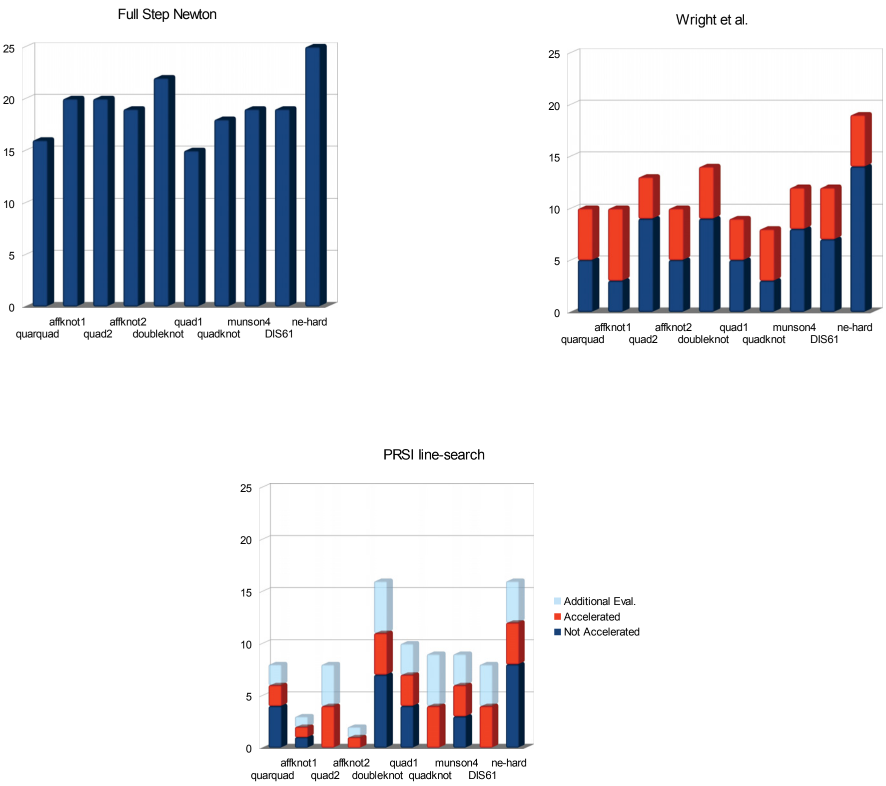


Figure 6.6: Graphical presentation of the data of the Table 6.9. Red represents accelerated steps, blue indicates unaccelerated steps, light blue accounts for the addtional pure function evaluations without Jacobian calculations and factorizations during the line-searches. The full height of the columns represents the total number of function evaluations, which can be directly compared to the histograms for the full-step Newton method and its accelerations according to Oberlin and Wright.

j	ls	η_j	x_j	$\|F(x_j)\|$	p_j
1	2	1.7361	(1.3056, 1.3056)	0.34476	0.32736
2	2	1.8304	(1.0552, 1.0552)	0.0090949	0.36051
3	2	1.9516	(1.0027, 1.0027)	2.0832e-05	0.20967
4	2	1.9973	(1, 1)	1.5076e-10	0.051492

Table 6.11: The behavior of step multiplier η of PRSI line-search method for the function munson4 (6.4.13) at the initial point $(2, 2)$.

Table 6.9 compares the number of iterations of the pure Newton's method, accelerated Newton's method of Wright et al. and PRSI line-search for the NCP test problems. It is clearly shown that our line-search is faster than the other two approaches.

Table 6.10, 6.11 show the behavior of the step multiplier η for the new method applied to two different NCP examples.

As we can see in Table 6.10 the Newton method with line-search behaves almost like full-step Newton on nonsingular problems. Note that strict complementarity also implies that Ψ is as smooth as F in some neighborhood of the solution. Singularity of Ψ' due to lack strict complementarity makes the line-search variant much more effective relative to the full-step version, especially if Ψ' vanishes completely so that $N = \mathbb{R}^n$. This is the case for problem munson4 (6.4.13) where as shown in Table 6.11 all steps are accelerated, leading to a reduction of the iteration count by the factor 3 compared to full-step Newton. On problems where N is properly contained in $\mathbb{R}^n$ on average only every step can be accelerated in order to stay close to the wedge shape domain of construction. On such problems the reduction in the iteration count is roughly by a factor of 2.

6.5 Newton's Fractals

6.5.1 Description of the Fractals

We consider the same problems above for the singular and nonsingular case and we want to plot this functions with respect to the number of iterations j in both cases unmodified Newton's method and parabolic range space interpolation Newton's method:

The plot represents a grid in the square $[-1, 1] \times [-1, 1]$. Each point of the grid was assigned the number of iterations j which Newton's method takes to reach a first point in a target set W around the solution in the center of the picture. The stopping criterion was chosen to be: $(|y| < \mu|x| : |x| < \rho)$, whose $\rho = 10^{-11}$, and μ is selected as $\|F(x)\| \leq \mu = 10^{-11}$.

6.5.2 Description of the Pictures

We have four different types of pictures.

1. Newton fractal.
2. Line search parameters.
3. Norm of function values.
4. Determinant of the Jacobian.

1. Newton Fractals: The singular root is always in the center of the square. The nullspace corresponds to the horizontal axis through the center. The singular set is a smooth curve that runs transversal to the nullspace, it is not drawn explicitly. The colors indicate the basins of attraction of various roots of the tested problem. The intensity of the color represents the number of steps needed to approximate the root with a specified tolerance. The darker color means that more iterations are needed. The gray color indicates divergence.
2. The step multiplier η : The color blue indicates that η is very small or even zero, which may be the case if the determinant is close to zero, and hence, possibly large steps. This can result in within a sensitivity of the iteration with respect to the starting point. The meaning of the other colors are as follows green: $\eta = 1$, yellow: $\eta = 1.5$ and red:$\eta = 2$. Flat color in the pictures indicate the line-search stops after one iteration.
3. The norm $\|F\|$ of function F: The picture shows the shapes of the level sets of the function. The color scale is adapted to size of the function.
4. Determinant of the Jacobian $\delta(x)$: Blue color presents the singular set $\delta^{-1}(0)$, the color yellow indicate when the determinant takes the values bounded away from zero.

We produce fractals for the following problem, where the solutions are always $(0, 0)$

1. Rall's shifted example, nonsingular [Ral66]

$$F_{07}(x,y) = \begin{pmatrix} x^2 - xy + y^2 + 2x \\ 3x^2 + 2xy + 8x + 4y \end{pmatrix}. \tag{6.5.1}$$

2.

$$F_{04}(x,y) = \begin{pmatrix} x^3(-y^2 + y + 2) + 2x^2y + 2xy^2(y^2 + y - 2) \\ y - x(y^2 - 2) \end{pmatrix}. \tag{6.5.2}$$

3. Reddien's example [Red78], regular singularity

$$F_{05}(x,y) = \begin{pmatrix} x + xy + y^2 \\ x^2 - 2x + y^2 \end{pmatrix}. \tag{6.5.3}$$

4. Kelley and Decker's example, regular singularity [DK80a]

$$F_{06}(x,y) = \begin{pmatrix} x^2 + 2y + y^2 \\ y + x^2 \end{pmatrix}. \tag{6.5.4}$$

5. Griewank's example for simple singularity $k = 1$, regular singularity [Gri80a]

$$F_{21}(x,y) = \begin{pmatrix} \frac{y^2}{2} \\ y - x \end{pmatrix}. \tag{6.5.5}$$

6. Griewank's example for $k = 2$ [Gri80a]

$$F_{22}(x,y) = \begin{pmatrix} \frac{y^2}{2} \\ y - \frac{x^2}{2} \end{pmatrix}. \tag{6.5.6}$$

7. Irregular singularity example

$$F_{03}(x,y) = \begin{pmatrix} -4xy + 4x^3 \\ 4y - 2x^2 \end{pmatrix}. \tag{6.5.7}$$

8. Wright et al. NCP example affknot2 (6.4.11)

$$F_{23} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \tag{6.5.8}$$

where

$$f_1(x,y) = \begin{cases} 2xy - 2x & \text{if} \quad x + y - 1 \geq 0 \\ -x^2 - y^2 + 2y - 1 & \text{otherwise} \end{cases}$$

and

$$f_2(x,y) = \begin{cases} 2xy + 2y^2 - 2y & \text{if} \quad x + 2y - 1 \geq 0 \\ -x^2 - 2y^2 + 2y + 2x - 2xy - 1 & \text{otherwise} \end{cases}$$

The function given in (6.2.1) has a singular root $(0, 0)$ for the case $\varepsilon = 0$ as it was shown in Section 6.2. In the first four Pictures (6.13), (6.14) we see that the singular root $(0, 0)$ is in the center of the pictures and is colored by green. There is other nonsingular root is colored by blue. In the right hand side of the Picture (6.13) we note the light green color of line-search method, which indicate that it needs fewer iterations to reach the root $(0, 0)$. The case is different for other methods in Pictures (6.14) and in the left hand side of the Picture (6.13), where we see the darker green for full-step Newton. The big green wedge in RHS of the Picture (6.13) indicate the large domain of starting point which enable the line-search method to converge to the singular root$(0, 0)$.

As we have seen in the Table (6.6) for this function, the parameter η takes 7 accelerated steps of the total number of iterations 11, i.e. $\eta > 1.69$. That is clear in the LHS of Picture (6.15), where the color indicates the value of η. Here we see the blue color which shows when $\eta \to 0$ near to the singular set as it is shown in the determinant pictures.

The case is different for the nonsingular roots of the problem (6.1.1) presented in Pictures (6.7), (6.8), (6.9). Here we see three roots. The root $(0, 0)$ is colored by green. In the other side we see the (nearly) singular root, which is colored by brown. For the root $(0, 0)$ we note there is no big difference in the color indication of the pictures. The other two methods in Picture (6.1.1) are approximately the same. As we have seen in Table 6.2. Here the line-search is a little bit faster than full-step Newton. While in the 3- and 2-step methods we see the green region is rather smaller. The parameter η in the Picture (6.9) is always 1, except for one step, it is 1.67807. Because of that the picture shows more green color because the iterations sequence takes often the full Newton step.

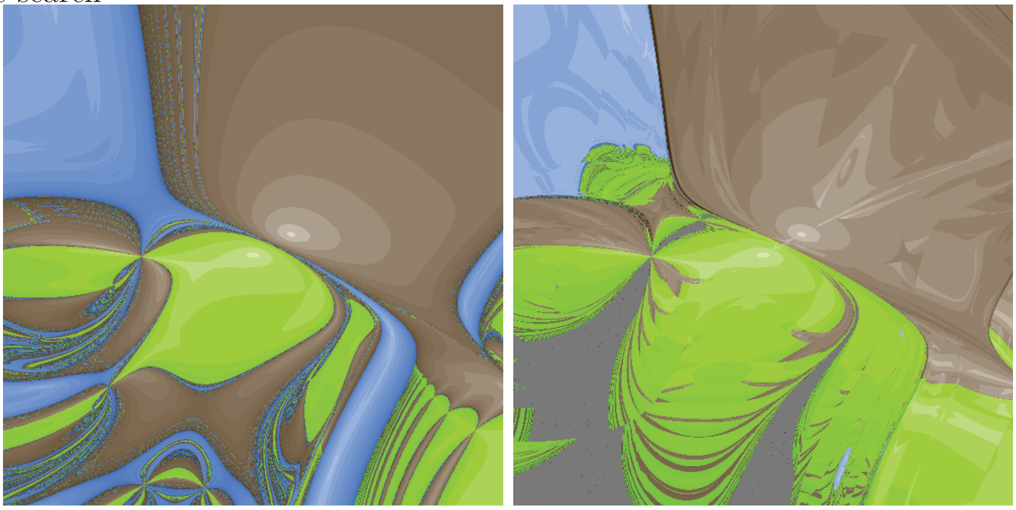
For the singularity of order 2 we consider the problem given in (6.5.6), its Pictures (6.28), (6.29), (6.29) show a solution in the center of the picture of PRSI line-search which is colored by green. The picture shows that the line-search could converge to the singular solution $(0, 0)$ with few iterations except the slow curve through the solution. While full-step Newton picture does not display the solution. The 2- and 3-step methods give chaos of jumped iterations, the solution also does not appear. In the determinant picture we see clearly the singular set. The line-search made good performance of convergence to singular solution. While the other methods failed. The parameter η through the singular set goes to zero and take the full-step near the orthogonal axis to the singular set.

We see the Pictures (6.31), (6.32), (6.33) of the irregular singularity example (6.5.7). The irregular singular root $(0, 0)$ is colored by green and centered the pictures. We see in the determinant picture the singular set is like a parabola. The slow curve appear orthogonally to the tangential of singular set. At the slow curve and its parallels we note that the line-search does not converge to the root. The plot of NCP example (6.5.8) of the function affknot2 (6.4.11) is given in Pictures (6.34), (6.35), (6.36). Here the singular root $(0, 1)$ of NCP is colored by blue and centered the picture, since we change here the radius of the pictures. we note that there is only one solution appears in the pictures. The light blue indicates the fewer iterations to reach the singular root $(0, 1)$. This is shown in the line-search picture. The Table 6.9 gives an example where the NCP needs only 6 iterations to reach the root. While the full-step Newton needs 19 iterations. The 2- and 3-step method pictures also show faster convergence than full-step Newton method, but the faster one of all is PRSI line-search. The parameter η is always bigger than 1.72 i.e., the iterations sequence in this case do not do any full-step Newton.

The other fractals of problems have generally the same behavior of above discussed examples. Since we showed the singular, nonsingular cases, irregular example and the NCP example. One can see their behavior in the following conclusion of the fractals.

Conclusion In the pictures we observe the improvement of line-search over full step Newton. In the full step picture the vast regions of darker green color signal a large number of iterations needed to approach the solution. In contrast the line-search picture has a large bright green region of fast convergence. Also divergence is more likely for the full step method and the performance of the 3-step and 2-step variants lies somewhere in the middle. As we already noted in general the line-search method needs about half as many iterations as the full step method.

Figure 6.7: Newton fractal for example (6.1.1), left with FSN, right with PRSI line-search



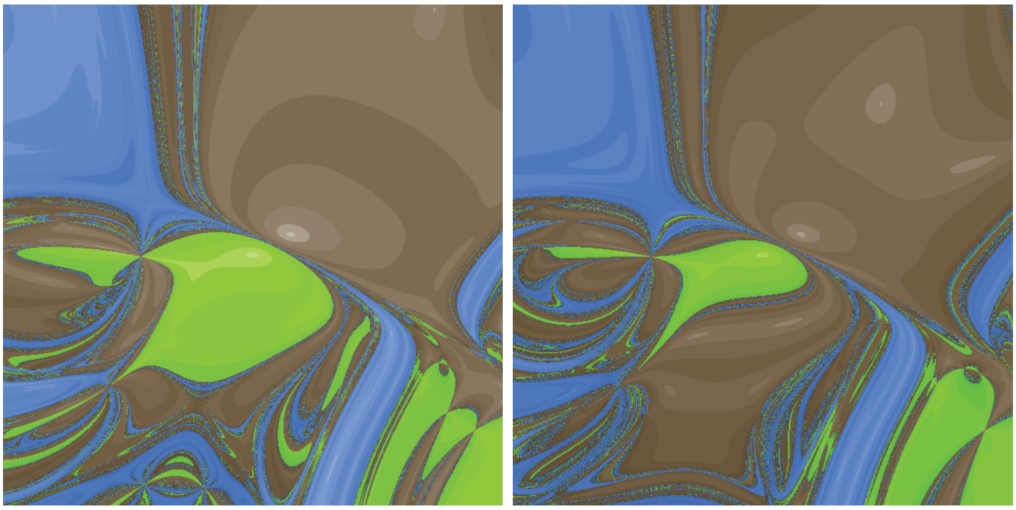


Figure 6.8: Newton fractal for example (6.1.1), left with 3-step method, right with 2-step method

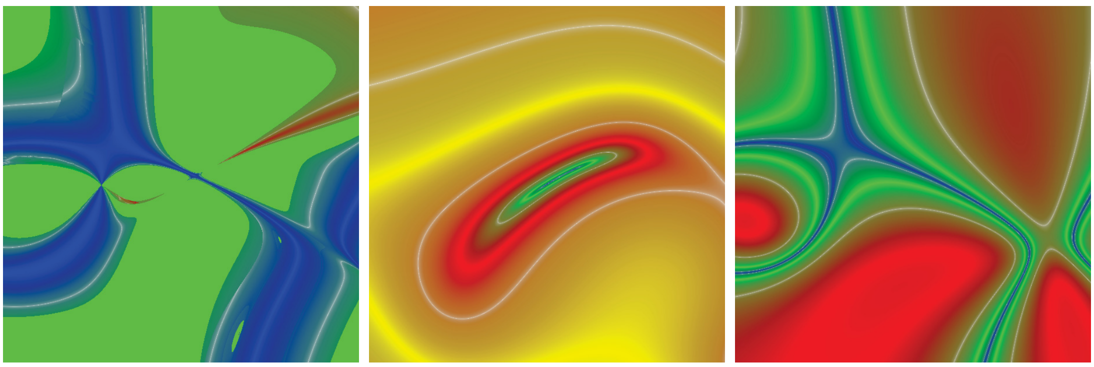
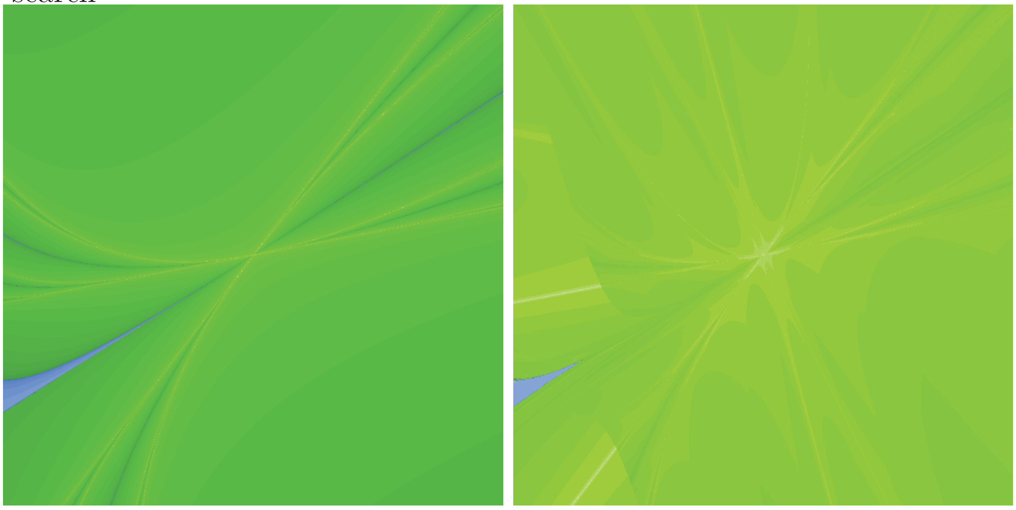


Figure 6.9: Newton fractal for example (6.1.1), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Figure 6.10: Newton fractal for example (6.5.1), left with FSN, right with PRSI line-search



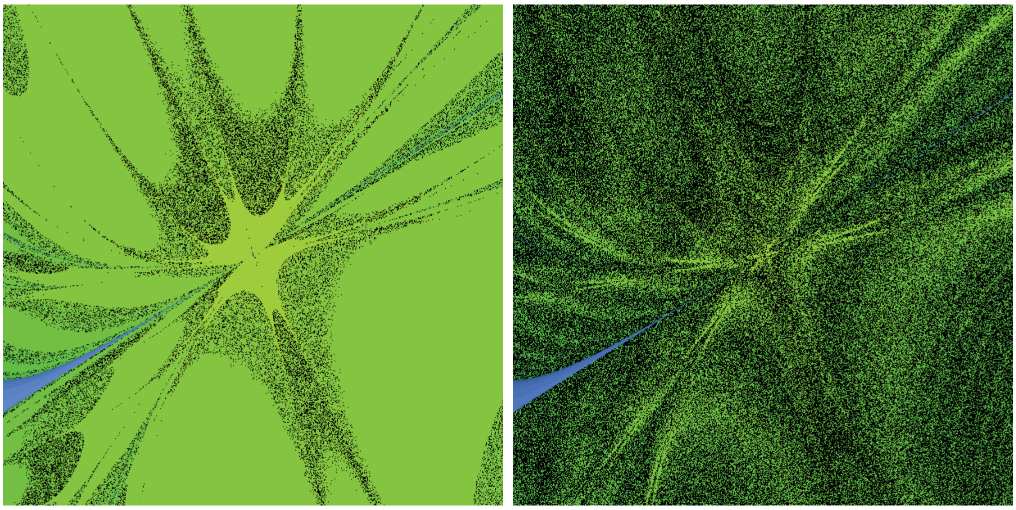


Figure 6.11: Newton fractal for example (6.5.1), left with 3-step method, right with 2-step method

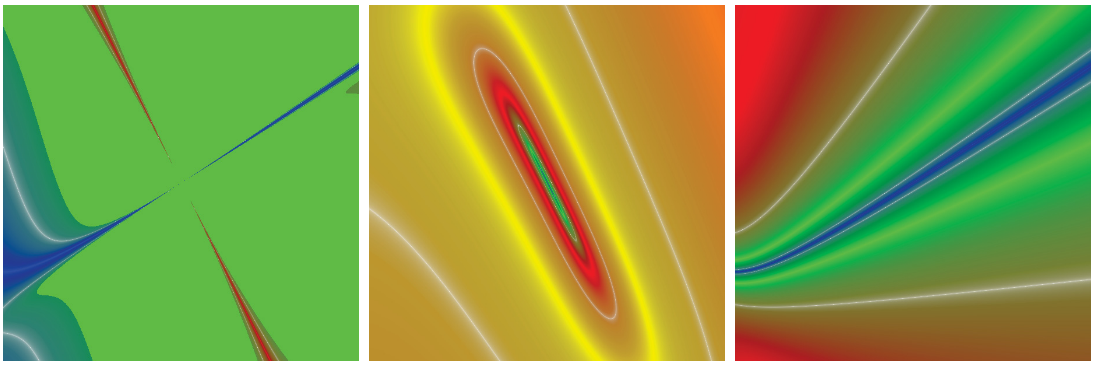
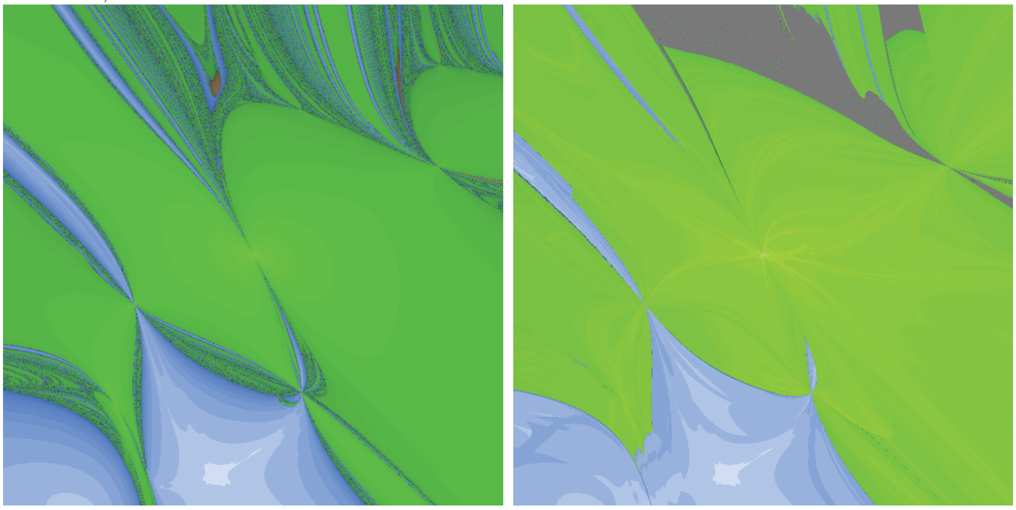


Figure 6.12: Newton fractal for example (6.5.1), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Figure 6.13: Newton fractal for example (6.2.1), left with FSN, right with PRSI line-search, for $\varepsilon = 0$



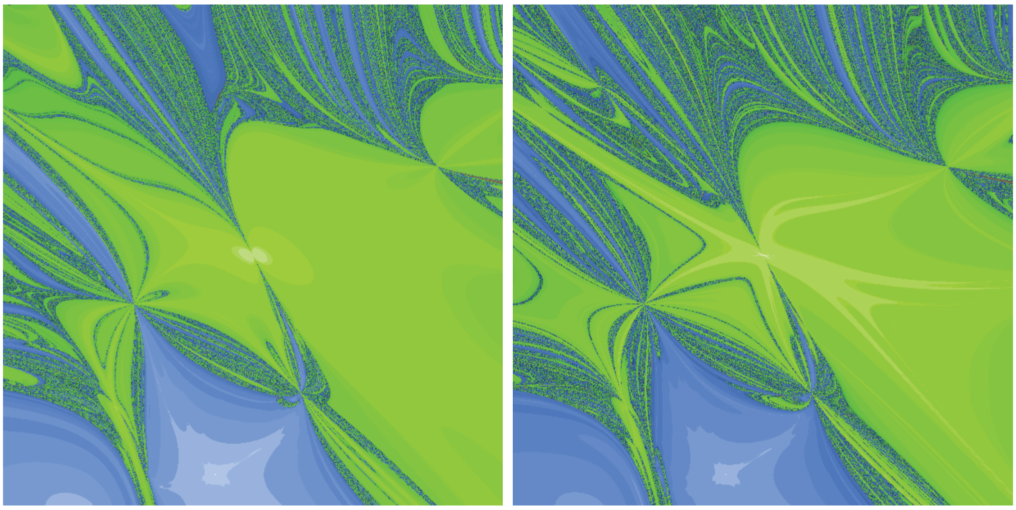


Figure 6.14: Newton fractal for example (6.2.1), left with 3-step method, right with 2-step method

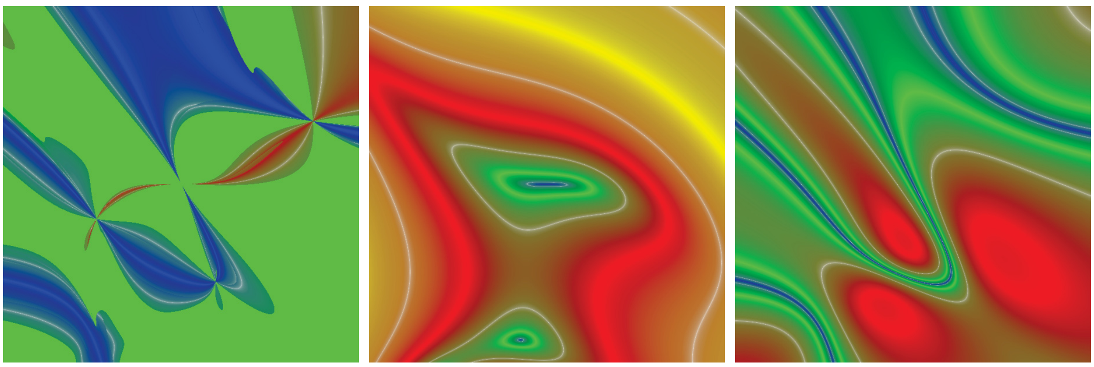
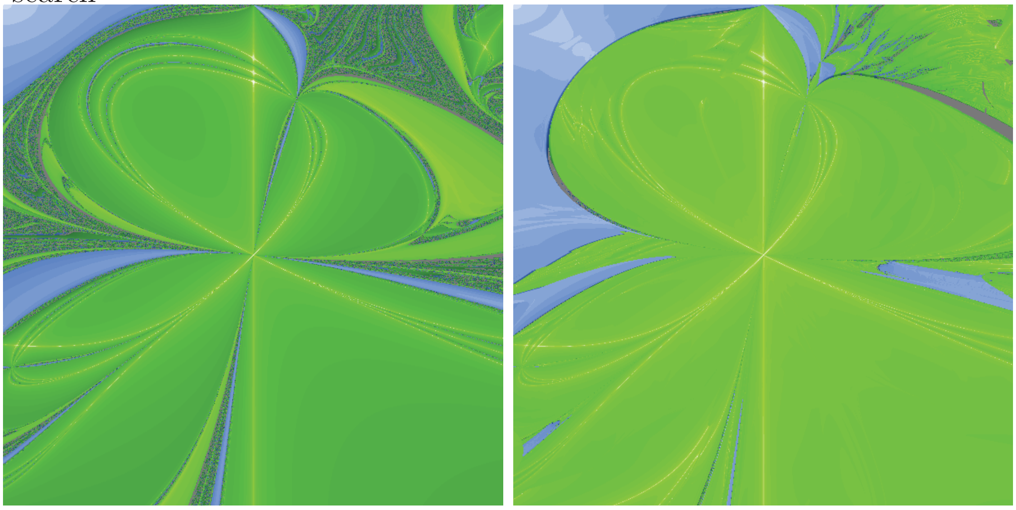


Figure 6.15: Newton fractal for example (6.2.1), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Figure 6.16: Newton fractal for example (6.5.2), left with FSN, right with PRSI line-search



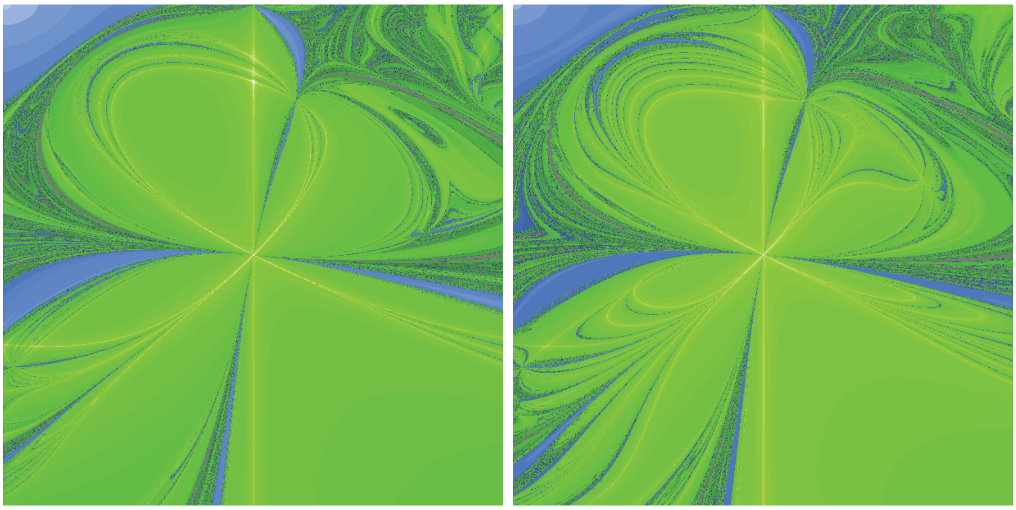


Figure 6.17: Newton fractal for example (6.5.2), left with 3-step method, right with 2-step method

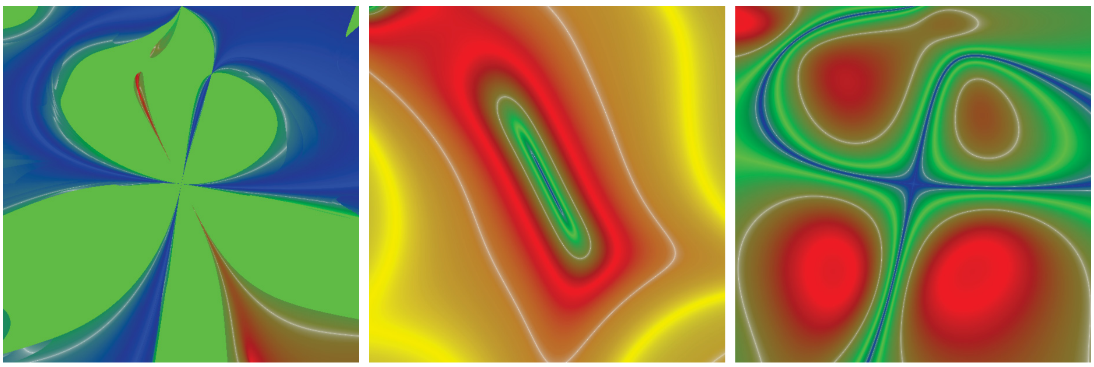


Figure 6.18: Newton fractal for example (6.5.2), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Figure 6.19: Newton fractal for example (6.5.3), left with FSN, right with PRSI line-search

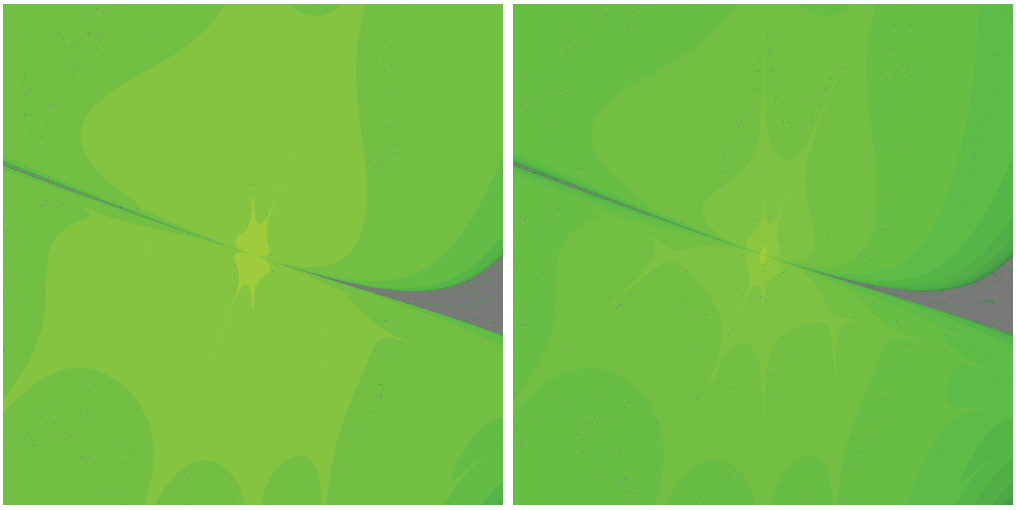


Figure 6.20: Newton fractal for example (6.5.3), left with 3-step method, right with 2-step method

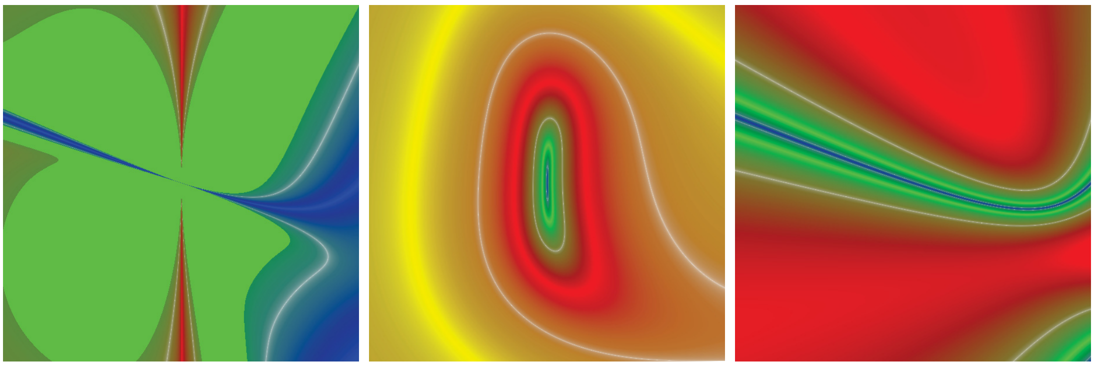


Figure 6.21: Newton fractal for example (6.5.3), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Figure 6.22: Newton fractal for example (6.5.4), left with FSN, right with PRSI line-search

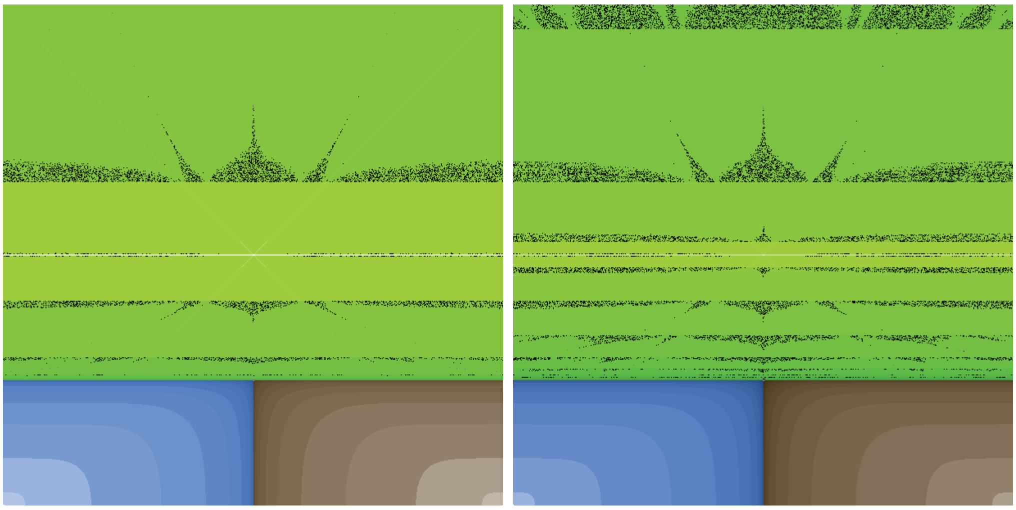


Figure 6.23: Newton fractal for example (6.5.4), left with 3-step method, right with 2-step method

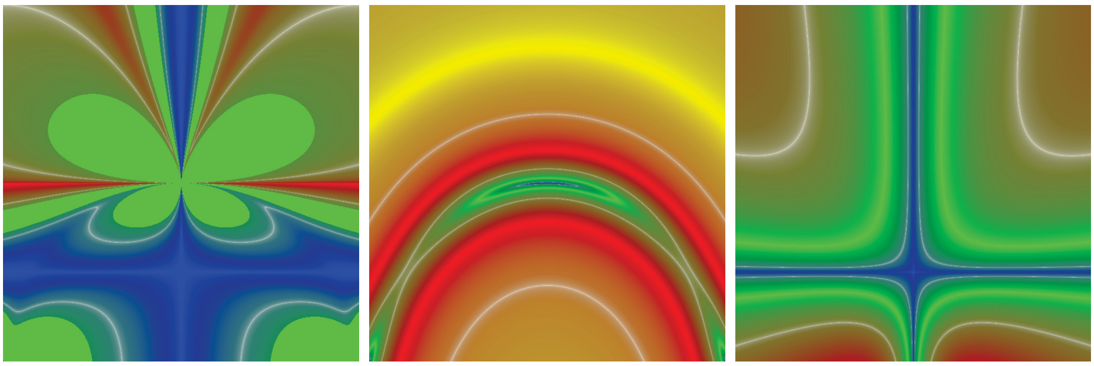


Figure 6.24: Newton fractal for example (6.5.4), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Figure 6.25: Newton fractal for example (6.5.5), left with FSN, right with PRSI line-search

Figure 6.26: Newton fractal for example (6.5.5), left with 3-step method, right with 2-step method

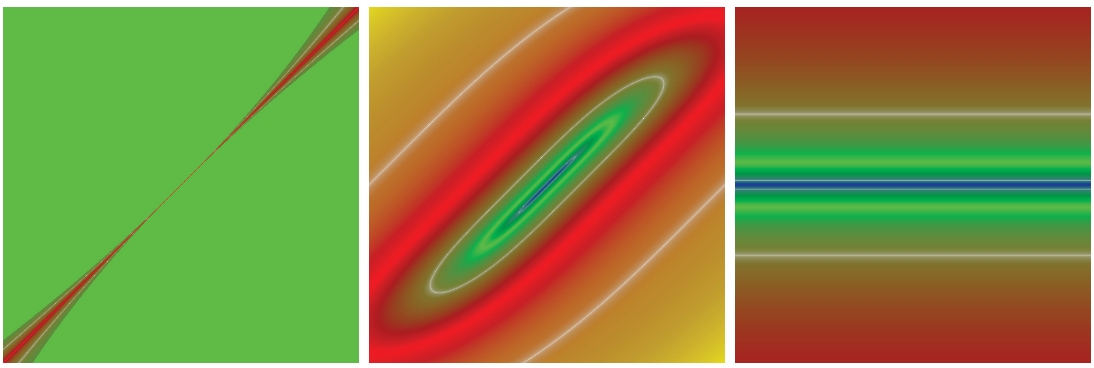
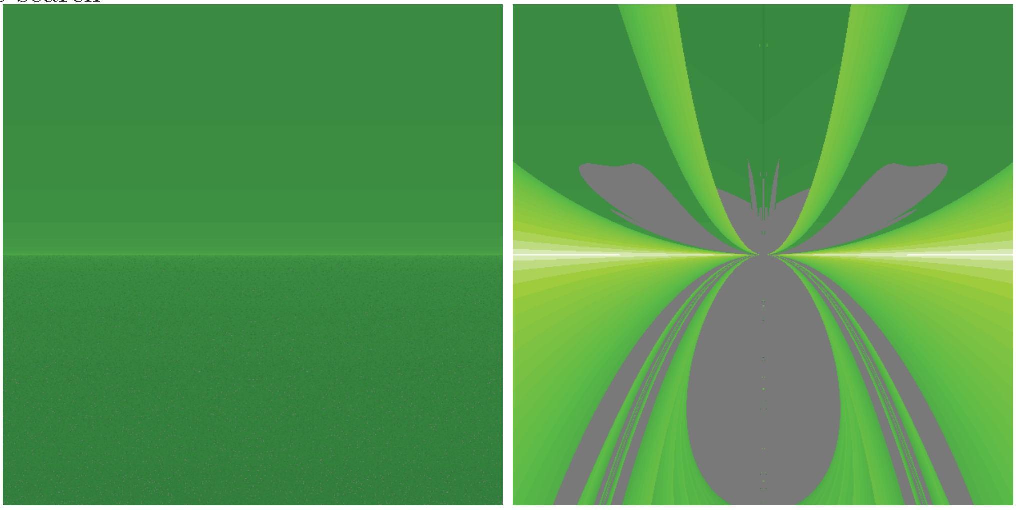


Figure 6.27: Newton fractal for example (6.5.5), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Figure 6.28: Newton fractal for example (6.5.6), left with FSN, right with PRSI line-search



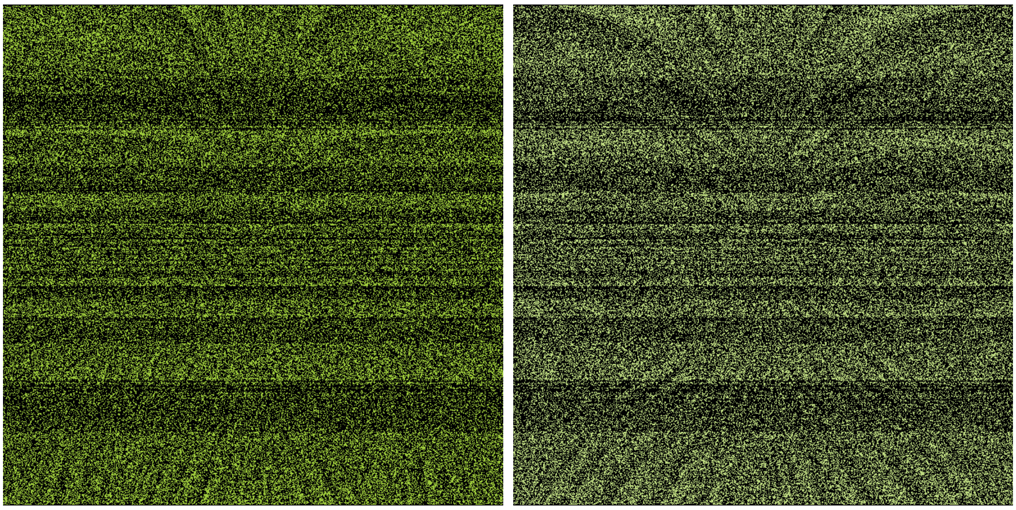


Figure 6.29: Newton fractal for example (6.5.6), left with 3-step method, right with 2-step method

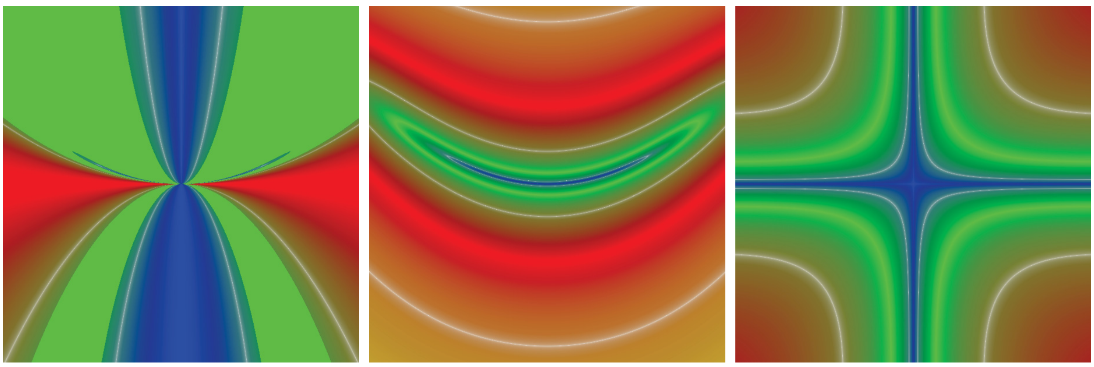


Figure 6.30: Newton fractal for example (6.5.6), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Figure 6.31: Newton fractal for example (6.5.7), left with FSN, right with PRSI line-search

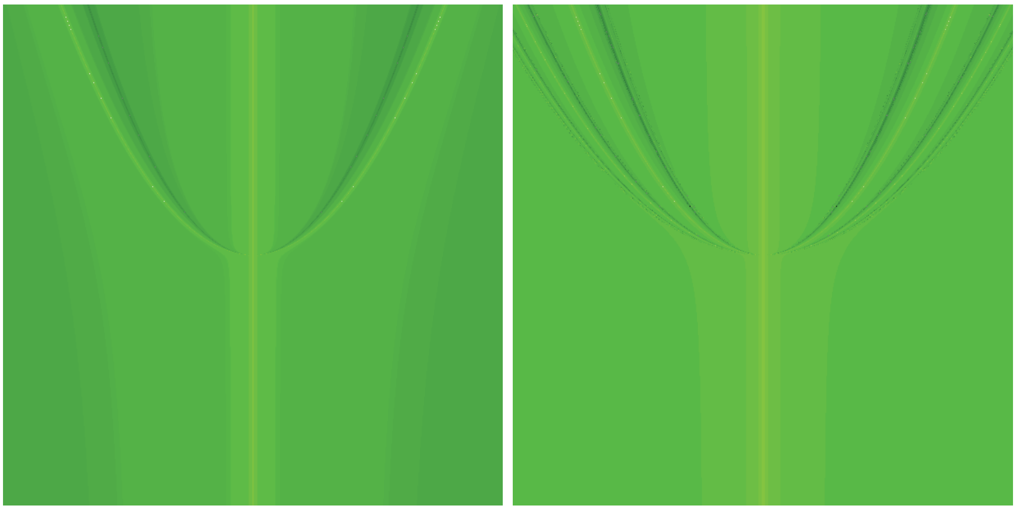
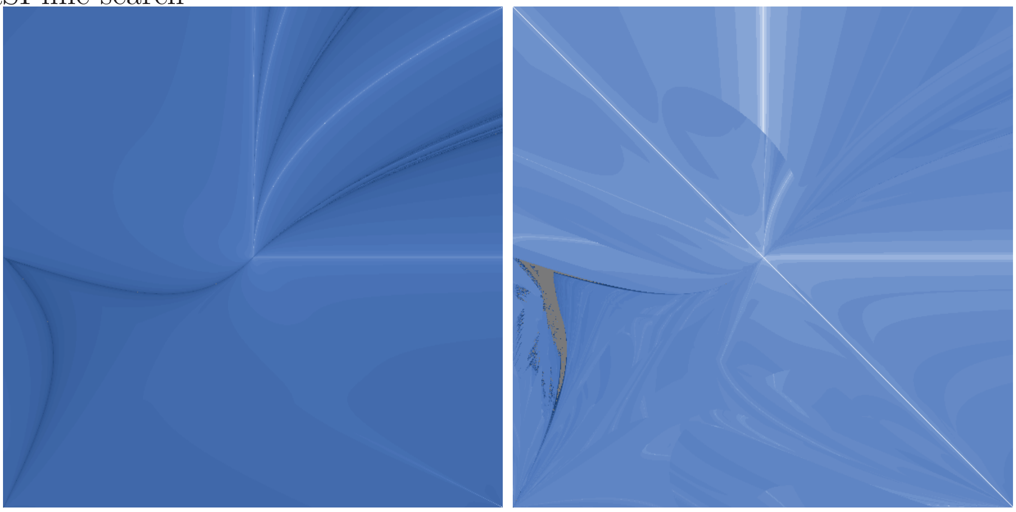


Figure 6.32: Newton fractal for example (6.5.7), left with 3-step method, right with 2-step method

Figure 6.33: Newton fractal for example (6.5.7), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Figure 6.34: Newton fractal for example NCP of (6.4.11), left with FSN, right with PRSI line-search



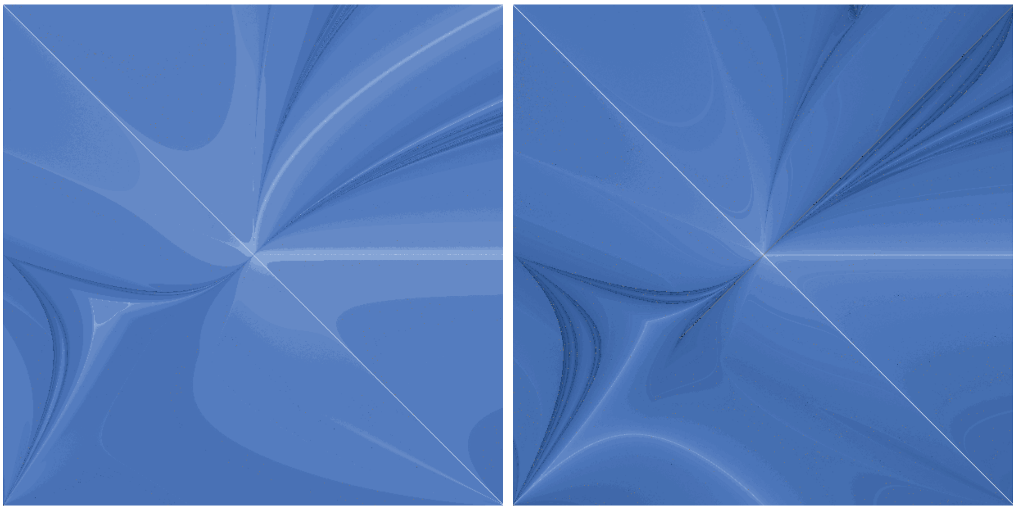


Figure 6.35: Newton fractal for example NCP of (6.4.11), left with 3-step method, right with 2-step method

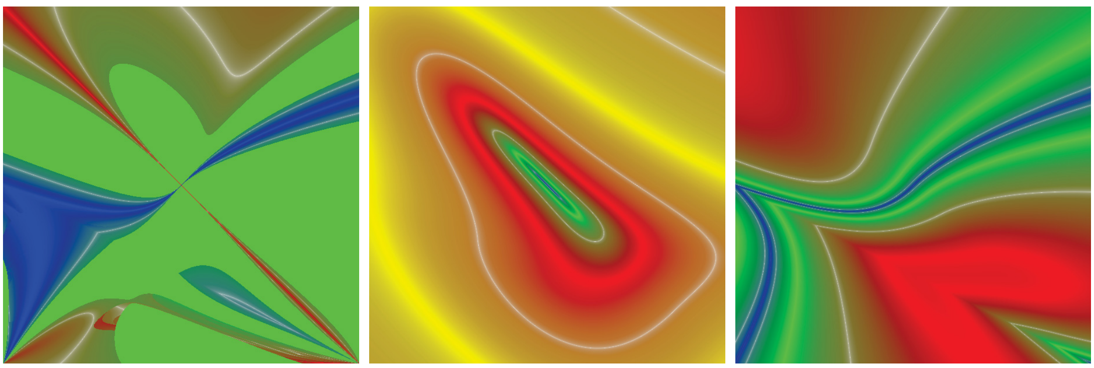


Figure 6.36: Newton fractal for example NCP of (6.4.11), left with line-search parameters, middle with norm of the function values , right with the Determinant $\delta(x)$

Chapter 7

Summary and Discussion

In this thesis we exam the modification of Newton's method by a line-search with the aim of stabilizing its performance at nearly singular root. We first reviewed convergence results for the full-step Newton method at nonsingular and singular solutions. While quadratic convergence is lost even for smooth nonlinear systems, linear convergence from within starlike domain of density 1 has been established provided the singular solution is regular in a certain sense. There are two key parameters, m the dimension of the nullspace and k of the order of the singularity. At regular singularities of order k the rate of convergence of full-step Newton is linear with the asymptotic ratio $k/k+1$. The exact pattern of convergence is displayed in Theorem 3.2.2.

Regular singularities with $m = 1 = k$ occur for example at quadratic turning points of parameter depended equations, e. g. the Bratu problem from combustion theory, which we use as test problem. First order singularities with m dimensional nullspace arise in complementarity systems whose solutions violate strict complementarity in m components. Here the system Jacobian is typically only directionally differentiable but semismooth. As shown by Oberlin et al. the convergence theory mention above still valid with minor modifications.

Based on the particular convergence pattern to singularities acceleration techniques have been developed in Chapter 4. Multi step quadratic or superlinear convergence can be recovered by the so-called 3- or 2-step variants of Newton's method. They require the detection of singularity and additional regularity conditions, which are difficult if not impossible to verify.

To overcome these short comings we have developed a line-search criterion that promises to implicitly detect and remedy singularity without effecting convergence in nonlinear cases. In particular, as illustrated by the multiplier mountain, we can expect the almost doubling of every other Newton step in the case of a regular first order singularity. This expectation was largely verified in our numerical experiments on problems with $k = 1$ and $1 \leq m < n$. The acceleration compared to full-step Newton's method usually reduced the number of steps needed to reach a certain solution accuracy by a factor of 2. In cases where $m = n$ the reduction factor was even close to 3 since all later steps could be accelerated as observed for example on some complementarity problems, see e. g. Table 6.11. In the vicinity the quadratic turning point of the Bratu problem nearly singular solutions were computed with the a number of iterations largely independent of the near-criticality.

No significant gains were observed when the line-search was based on cubic rather than parabolic interpolation of the residual path $F(x + \eta d)$.

In summary we conclude that the parabolic line-search is easy to implement, effective near regular first order singularity, and enhances global convergence properties for nonsingular and singular problems alike. The principal remaining challenge is an analytical proof of superlinear convergence in at least the simply singular case $k = 1 = m$.

Bibliography

[BK08] Sandra Buhmiler and Nataša Krejić. A new smoothing quasi-newton method for nonlinear complementarity problems. *Journal of Computational and Applied Mathematics*, 2008.

[BKL10] Sandra Buhmiler, Nataša Krejić, and Zorana Lužanin. Practical quasi-newton algorithms for singular nonlinear systems. *Numerical Algorithms*, 2010.

[BKM03] Ernesto G Birgin, Nataša Krejić, and José Mario Martínez. Globally convergent inexact quasi-newton methods for solving nonlinear systems. *Numerical algorithms*, 2003.

[BM00] Stephen C. Billups and Katta G. Murty. Complementarity problems. *Journal of Computational and Applied Mathematics*, 2000. Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.

[CAV70] R. C. CAVANAGH. *Difference equations and iterative processes*. PhD thesis, University of Maryland, 1970.

[Dav] Timothy A. Davis. Algorithm 832: Umfpack v4.3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Softw.*

[Deu04] Peter Deuflhard. *Newton methods for nonlinear problems*. Springer Series in Computational Mathematics. Springer-Verlag, 2004. Affine invariance and adaptive algorithms.

[DIS03] A.N. Daryina, A.F. Izmailov, and M.V. Solodov. *A Class of Active-set Newton Methods for Mixed Complementarity Problems*. Informes de matemática. Inst. de Matemática Pura e Aplicada, 2003.

[DK80a] D. W. Decker and C. T. Kelley. Newton's method at singular points i. *SIAM*, 1980.

[DK80b] D. W. Decker and C. T. Kelley. Newton's method at singular points ii. *SIAM*, 1980.

[DK82] D. W. Decker and C. T. Kelley. Convergence acceleration for newton's method at singular points. *SIAM*, 1982.

[DK85] D. W. Decker and C. T. Kelley. Expanded convergence domains for newton's method at nearly singular roots. *SIAM*, 1985.

[DKK83] D. W. Decker, H. B. Keller, and C. T. Kelley. Convergence rates for newton's method at singular points. *SIAM*, 1983.

[EP84] Y. G. Evtushenko and V. A. Purtov. Sufficient conditions for a minimum for nonlinear programming problems. *Soviet Math. Dokl. 30, 313–316*, 1984.

[FP03] F. Facchinei and J.S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.

[GO81] Andreas Griewank and M. R. Osborne. Newton's method for singular problems when the dimension of the null space is > 1. *SIAM*, 1981.

[GO83] A. Griewank and M. R. Osborne. Analysis of newton's method at irregular singularities. *SIAM*, 1983.

[Gri80a] Andreas Griewank. *Analysis and modifications of Newton's method at singularities*. Dissertation, Australian National University, 1980.

[Gri80b] Andreas Griewank. Starlike domains of convergence for newton method at singularities. *Springer*, 1980.

[Gri85] Andreas Griewank. On solving nonlinear equations with simple singularities or nearly singular solutions. *SIAM Journal on Numerical Analysis,Rev. 27*, 1985.

[HMT09] José L Hueso, Eulalia Martínez, and Juan R Torregrosa. Modified newton's method for systems of nonlinear equations with singular jacobian. *Journal of Computational and Applied Mathematics*, 2009.

[Kan48] L. V. Kantorovich. *Functional analysis and applied mathematics*. Uspehi Mat. Nauk, 1948.

[Kan94] Christian Kanzow. Some equation-based methods for the nonlinear complementarity problem. *Optimization Methods and Software*, 3(4):327–340, 1994.

[Kel95] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. SIAM, Philadelphia, 1995.

[KH99] Peter Kravanja and Ann Haegemans. A modification of newton's method for analytic mappings having multiple zeros. *Computing*, 1999.

[KS83] C. T. Kelley and R. Suresh. A new acceleration method for newton's method at singular points. *SIAM*, 1983.

[LVK59] G. P. Akilov L. V. Kantorovich. *Functional Analysis in Normed Spaces*. Translated by D. Brown, Pergamon, New York, 1964. Uspehi Mat. Nauk, Moscow, 1959.

[LVZ06] Anton Leykin, Jan Verschelde, and Ailing Zhao. Newton's method with deflation for isolated singularities of polynomial systems. *Theoretical Computer Science*, 2006.

[LW84] Per Lindström and Per-Åke Wedin. A new linesearch algorithm for nonlinear least squares problems. *Mathematical programming*, 1984.

[LZ10] Nan Li and Lihong Zhi. A symbolic-numeric algorithm for computing isolated singular solutions of nonlinear systems accurately: Case of breadth one. *Theoretical Computer Science*, 2010.

[LZ12a] Nan Li and Lihong Zhi. Computing the multiplicity structure of an isolated singular solution: Case of breadth one. *Journal of Symbolic Computation*, 2012.

[LZ12b] Nan Li and Lihong Zhi. Verified error bounds for isolated singular solutions of polynomial systems: Case of breadth one. *Theoretical Computer Science*, 2012.

[MCP] ftp://ftp.cs.wisc.edu/math-prog/mcplib.

[Noc06] S. Nocedal, J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2006.

[OR70] J. M. Ortega and Werner. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, INC, 1970.

[Ort68] J. M. Ortega. The newton-kantorovich theorem. *The American Mathematical Monthly*, 1968.

[OW07] Christina Oberlin and Stephen J. Wright. An accelerated Newton method for equations with semismooth jacobians and nonlinear complementarity problems. extended version. optimization technical report. *Math. Program.*, 2006, Rivesed Jan 2007.

[OW09] Christina Oberlin and Stephen J. Wright. An accelerated newton method for equations with semismooth jacobians and nonlinear complementarity problems. *Math. Program.*, 2009.

[QS91] L. Qi and J. Sun. A nonsmooth version of newton's method. 1991.

[Ral66] L. B. Rall. Convergence of the Newton process to multiple solutions. *Numerische Mathematik*, 1966.

[Red78] G. W. Reddien. On newton's method for singular problems. *SIAM*, 1978.

[Red79] G. W. Reddien. Newton's method and high order singularities. *Computers & Mathematics with Applications. An International Journal*, 1979.

[SY05] Yun-Qiu Shen and Tjalling J. Ypma. Newton's method for singular nonlinear equations using approximate left and right nullspaces of the jacobian. *Appl. Numer. Math.*, July 2005.