国外数学名著系列
（影印版）15

Peter Deuflhard

Newton Methods for Nonlinear Problems

Affine Invariance and Adaptive Algorithms

非线性问题的牛顿法

仿射不变性和自适应算法

科学出版社
www.sciencep.com

图字:01-2005-6752

Peter Deuflhard: Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms

图书在版编目(CIP)数据

非线性问题的牛顿法:仿射不变性和自适应算法=Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms/(德)D. 彼得(Peter, D)著. —影印版. —北京:科学出版社,2006

(国外数学名著系列)

ISBN 7-03-016685-X

Ⅰ. 非… Ⅱ. 德… Ⅲ. 非线性-牛顿法-研究生-教材-英文 Ⅳ. 0242.23

中国版本图书馆 CIP 数据核字(2005)第 154393 号

科学出版社 出版
北京东黄城根北街16号
邮政编码:100717
http://www.sciencep.com
中国科学院印刷厂 印刷
科学出版社发行 各地新华书店经销

*

2006 年 1 月第 一 版 开本:B5(720×1000)
2006 年 1 月第一次印刷 印张:27 1/2
印数:1—2 000 字数:519 000

定价:68.00 元

(如有印装质量问题,我社负责调换(科印))

《国外数学名著系列》(影印版)专家委员会

(按姓氏笔画排序)

丁伟岳 王 元 石钟慈 冯克勤 严加安 李邦河
李大潜 张伟平 张继平 杨 乐 姜伯驹 郭 雷

项目策划

向安全 林 鹏 王春香 吕 虹 范庆奎 王 璐

执行编辑

范庆奎

《国外数学名著系列》(影印版)序

要使我国的数学事业更好地发展起来，需要数学家淡泊名利并付出更艰苦地努力。另一方面，我们也要从客观上为数学家创造更有利的发展数学事业的外部环境，这主要是加强对数学事业的支持与投资力度，使数学家有较好的工作与生活条件，其中也包括改善与加强数学的出版工作。

从出版方面来讲，除了较好较快地出版我们自己的成果外，引进国外的先进出版物无疑也是十分重要与必不可少的。从数学来说，施普林格（Springer）出版社至今仍然是世界上最具权威的出版社。科学出版社影印一批他们出版的好的新书，使我国广大数学家能以较低的价格购买，特别是在边远地区工作的数学家能普遍见到这些书，无疑是对推动我国数学的科研与教学十分有益的事。

这次科学出版社购买了版权，一次影印了 23 本施普林格出版社出版的数学书，就是一件好事，也是值得继续做下去的事情。大体上分一下，这 23 本书中，包括基础数学书 5 本，应用数学书 6 本与计算数学书 12 本，其中有些书也具有交叉性质。这些书都是很新的，2000 年以后出版的占绝大部分，共计 16 本，其余的也是 1990 年以后出版的。这些书可以使读者较快地了解数学某方面的前沿，例如基础数学中的数论、代数与拓扑三本，都是由该领域大数学家编著的“数学百科全书”的分册。对从事这方面研究的数学家了解该领域的前沿与全貌很有帮助。按照学科的特点，基础数学类的书以“经典”为主，应用和计算数学类的书以“前沿”为主。这些书的作者多数是国际知名的大数学家，例如《拓扑学》一书的作者诺维科夫是俄罗斯科学院的院士，曾获“菲尔兹奖”和“沃尔夫数学奖”。这些大数学家的著作无疑将会对我国的科研人员起到非常好的指导作用。

当然，23 本书只能涵盖数学的一部分，所以，这项工作还应该继续做下去。更进一步，有些读者面较广的好书还应该翻译成中文出版，使之有更大的读者群。

总之，我对科学出版社影印施普林格出版社的部分数学著作这一举措表示热烈的支持，并盼望这一工作取得更大的成绩。

王 元

2005 年 12 月 3 日

Preface

In 1970, my former academic teacher Roland Bulirsch gave an exercise to his students, which indicated the fascinating invariance of the *ordinary* Newton method under general affine transformation. To my surprise, however, nearly all *global* Newton algorithms used damping or continuation strategies based on residual norms, which evidently lacked affine invariance. Even worse, nearly all convergence theorems appeared to be phrased in not affine invariant terms, among them the classical Newton-Kantorovich and Newton-Mysovskikh theorem. In fact, in those days it was common understanding among numerical analysts that convergence theorems were only expected to give qualitative insight, but not too much of quantitative advice for application, apart from toy problems.

This situation left me deeply unsatisfied, from the point of view of both mathematical aesthetics and algorithm design. Indeed, since my first academic steps, my scientific guideline has been and still is that 'good' mathematical theory should have a palpable influence on the construction of algorithms, while 'good' algorithms should be as firmly as possible backed by a transparently underlying mathematical theory. Only on such a basis, algorithms will be efficient enough to cope with the enormous difficulties of real life problems.

In 1972, I started to work along this line by constructing global Newton algorithms with affine invariant damping strategies [59]. Early companions on this road were Hans-Georg Bock, Gerhard Heindl, and Tetsuro Yamamoto. Since then, the tree of affine invariance has grown lustily, spreading out in many branches of Newton-type methods. So the plan of a comprehensive treatise on the subject arose naturally. Florian Potra, Ekkehard Sachs, and Andreas Griewank gave highly valuable detailed advice. Around 1992, a manuscript on the subject with a comparable working title had already swollen to 300 pages and been distributed among quite a number of colleagues who used it in their lectures or as a basis for their research. Clearly, these colleagues put screws on me to 'finish' that manuscript.

However, shortly after, new relevant aspects came up. In 1993, my former coworker Andreas Hohmann introduced *affine contravariance* in his PhD thesis [119] as a further coherent concept, especially useful in the context of inexact Newton methods with GMRES as inner iterative solver. From then on, the former 'affine invariance' had to be renamed, more precisely, as *affine covariance*. Once the door had been opened, two more concepts arose: in 1996, myself and Martin Weiser formulated *affine conjugacy* for convex optimization [84]; a few years later, I found *affine similarity* to be important for steady state problems in dynamical systems. As a consequence, I decided to rewrite the whole manuscript from scratch, with these four affine invariance concepts representing the columns of a structural matrix, whose rows are the various Newton and Gauss-Newton methods. A presentation of details of the contents is postponed to the next section.

This book has two faces: the first one is that of a *textbook* addressing itself to graduate students of mathematics and computational sciences, the second one is that of a *research monograph* addressing itself to numerical analysists and computational scientists working on the subject.

As a *textbook*, selected chapters may be useful in classes on Numerical Analysis, Nonlinear Optimization, Numerical ODEs, or Numerical PDEs. The presentation is striving for structural simplicity, but not at the expense of precision. It contains a lot of theorems and proofs, from affine invariant versions of the classical Newton-Kantorovich and Newton-Mysovskikh theorem (with proofs simpler than the traditional ones) up to new convergence theorems that are the basis for advanced algorithms in large scale scientific computing. I confess that I did not work out all details of all proofs, if they were folklore or if their structure appeared repeatedly. More elaboration on this aspect would have unduly blown up the volume without adding enough value for the construction of algorithms. However, I definitely made sure that each section is self-contained to a reasonable extent. At the end of each chapter, exercises are included. Web addresses for related software are given.

As a *research monograph*, the presentation (a) quite often goes into the depth covering a large amount of otherwise unpublished material, (b) is open in many directions of possible future research, some of which are explicitly indicated in the text. Even though the experienced reader will have no difficulties in identifying further open topics, let me mention a few of them: There is no complete coverage of all possible combinations of local and global, exact and inexact Newton or Gauss-Newton methods in connection with continuation methods—let alone of all their affine invariant realizations; in other words, the above structural matrix is far from being full. Moreover, apart from convex optimization and constrained nonlinear least squares problems, general optimization and optimal control is left out. Also not included are recent results on interior point methods as well as inverse problems in L^2, even though affine invariance has just started to play a role in these fields.

Generally speaking, finite dimensional problems and techniques dominate the material presented here—however, with the declared intent that the finite dimensional presentation should filter out promising paths into the infinite dimensional part of the mathematical world. This intent is exemplified in several sections, such as

- Section 6.2 on ODE initial value problems, where stiff problems are analyzed via a simplified Newton iteration in function space—replacing the Picard iteration, which appears to be suitable only for nonstiff problems,
- Section 7.4.2 on ODE boundary value problems, where an adaptive multilevel collocation method is worked out on the basis of an inexact Newton method in function space,
- Section 8.1 on asymptotic mesh independence, where finite and infinite dimensional Newton sequences are synoptically compared, and
- Section 8.3 on elliptic PDE boundary value problems, where inexact Newton multilevel finite element methods are presented in detail.

The *algorithmic paradigm*, given in Section 1.2.3 and used all over the whole book, will certainly be useful in a much wider context, far beyond Newton methods.

Unfortunately, after having finished this book, I will probably lose all my scientific friends, since I missed to quote exactly that part of their work that should have been quoted by all means. I cannot but apologize in advance, hoping that some of them will maintain their friendship nevertheless. In fact, as the literature on Newton methods is virtually unlimited, I decided to not even attempt to screen or pretend to have screened all the relevant literature, but to restrict the references essentially to those books and papers that are either intimately tied to affine invariance or have otherwise been taken as direct input for the presentation herein. Even with this restriction the list is still quite long.

At this point it is my pleasure to thank all those coworkers at ZIB, who have particularly helped me with the preparation of this book. My first thanks go to Rainer Roitzsch, without whose high motivation and deep TEX knowledge this book could never have appeared. My immediate next thanks go to Erlinda Körnig and Sigrid Wacker for their always friendly cooperation over the long time that the manuscript has grown. Moreover, I am grateful to Ulrich Nowak, Andreas Hohmann, Martin Weiser, and Anton Schiela for their intensive computational assistance and invaluable help in improving the quality of the manuscript.

Nearly last, but certainly not least, I wish to thank Harry Yserentant, Christian Lubich, Matthias Heinkenschloss, and a number of anonymous reviewers for valuable comments on a former draft. My final thanks go to Martin Peters from Springer for his enduring support.

Berlin, February 2004

Peter Deuflhard

Table of Contents

Outline of Contents

This book is divided into eight chapters, a reference list, a software list, and an index. After an elementary introduction in Chapter 1, it splits into two parts: Part I, Chapter 2 to Chapter 5, on finite dimensional Newton methods for *algebraic equations*, and Part II, Chapter 6 to Chapter 8, on extensions to ordinary and partial *differential equations*. Exercises are added at the end of each chapter.

Chapter 1. This introductory chapter starts from the historical root, Newton's method for scalar equations (Section 1.1). The method can be derived either *algebraically*, which leads to *local* Newton methods only (presented in Chapter 2), or *geometrically*, which leads to *global* Newton methods via the concept of the Newton path (see Chapter 3).

The next Section 1.2 contains the *key to the basic understanding of this monograph.* First, four affine invariance classes are worked out, which represent the four basic strands of this treatise:

- *affine covariance*, which leads to *error* norm controlled algorithms,
- *affine contravariance*, which leads to *residual* norm controlled algorithms,
- *affine conjugacy*, which leads to *energy* norm controlled algorithms, and
- *affine similarity*, which may lead to *time* step controlled algorithms.

Second, the affine invariant local estimation of affine invariant Lipschitz constants is set as the central *paradigm* for the construction of adaptive Newton algorithms.

In Section 1.3, we give a roadmap of the large variety of Newton-type methods—essentially fixing terms to be used throughout the book such as ordinary and simplified Newton method, Newton-like methods, inexact Newton methods, quasi-Newton methods, Gauss-Newton methods, quasilinearization, or inexact Newton multilevel methods. In Section 1.4, we briefly collect details about iterative linear solvers to be used as inner iterations within finite dimensional inexact Newton algorithms; each affine invariance class is linked with a special class of inner iterations. In view of function space oriented inexact Newton algorithms, we also revisit linear multigrid methods. Throughout this section, we emphasize the role of adaptive error control.

PART I. The following Chapters 2 to 5 deal with *finite dimensional* Newton methods for algebraic equations.

Chapter 2. This chapter deals with *local* Newton methods for the numerical solution of systems of nonlinear equations with finite, possibly large dimension. The term 'local' refers to the situation that 'sufficiently good' initial guesses of the solution are assumed to be at hand. Special attention is paid to the issue of how to recognize, whether a given initial guess x^0 is 'sufficiently good'. Different affine invariant formulations give different answers to this question, in theoretical terms as well as by virtue of the algorithmic paradigm of Section 1.2.3. Problems of this structure are called 'mildly nonlinear'; their computational complexity can be bounded a-priori in units of the computational complexity of the corresponding linearized system.

As it turns out, different affine invariant Lipschitz conditions, which have been introduced in Section 1.2.2, lead to different characterizations of local convergence domains in terms of error oriented norms, residual norms, or energy norms, which, in turn, give rise to corresponding variants of Newton algorithms. We give three different, strictly affine invariant convergence analyses for the cases of affine covariant (error oriented) Newton methods (Section 2.1), affine contravariant (residual based) Newton methods (Section 2.2), and affine conjugate Newton methods for convex optimization (Section 2.3). Details are worked out for ordinary Newton algorithms, simplified Newton algorithms, and inexact Newton algorithms—synoptically for each of the three affine invariance classes. Moreover, affine covariance is naturally associated with Broyden's 'good' quasi-Newton method, whereas affine contravariance corresponds to Broyden's 'bad' quasi-Newton method.

Affine invariant *globalization*, which means global extension of the convergence domains of local Newton methods in the affine invariant frame, is possible along several lines:

- global Newton methods with damping strategy—see Chapter 3,
- parameter continuation methods—see Chapter 5,
- pseudo-transient continuation methods—see Section 6.4.

Chapter 3. This chapter deals with *global* Newton methods for systems of nonlinear equations with finite, possibly large dimension. The term 'global' refers to the situation that here, in contrast to the preceding chapter, 'sufficiently good' initial guesses of the solution are no longer assumed. Problems of this structure are called 'highly nonlinear'; their computational complexity depends on topological details of Newton paths associated with the nonlinear mapping and can typically not be bounded a-priori.

In Section 3.1 we survey globalization concepts such as

- steepest descent methods,
- trust region methods,
- the Levenberg-Marquardt method, and
- the Newton method with damping strategy.

In Section 3.1.4, a rather general geometric approach is taken: the idea is to derive a globalization concept without a pre-occupation to any iterative method, just starting from the requirement of affine covariance as a 'first principle'. Surprisingly, this general approach leads to a topological derivation of Newton's method with damping strategy via Newton paths.

In order to accept or reject a new iterate, *monotonicity tests* are applied. We study different such tests, according to different affine invariance requirements:

- the most popular *residual* monotonicity test, which is related to affine contravariance (Section 3.2),
- the error oriented so-called *natural* monotonicity test, which is related to affine covariance (Section 3.3), and
- the convex functional test as the natural requirement in convex optimization, which reflects affine conjugacy (Section 3.4).

For each of these three affine invariance classes, *adaptive trust region strategies* are designed in view of an efficient choice of damping factors in Newton's method. They are all based on the *paradigm* of Section 1.2.3. On a theoretical basis, details of algorithmic realization in combination with either *direct* or *iterative* linear solvers are worked out. As it turns out, an efficient determination of the steplength factor in global inexact Newton methods is intimately linked with the accuracy matching for affine invariant combinations of inner and outer iteration.

Chapter 4. This chapter deals with both *local* and *global Gauss-Newton* methods for *nonlinear least squares* problems in finite dimension—a method, which attacks the solution of the nonlinear least squares problem by solving a sequence of linear least squares problems. Affine invariance of both theory and algorithms will once again play a role, here restricted to *affine contravariance* and *affine covariance*. The theoretical treatment requires considerably more sophistication than in the simpler case of Newton methods for nonlinear equations.

In order to lay some basis, unconstrained and equality constrained *linear* least squares problems are first discussed in Section 4.1, introducing the useful calculus of generalized inverses. In Section 4.2, an affine contravariant convergence analysis of Gauss-Newton methods is given and worked out in the direction of *residual* based algorithms. Local convergence turns out to be only guaranteed for 'small residual' problems, which can be characterized in theoretical and algorithmic terms. Local and global convergence analysis as well as adaptive trust region strategies rely on some *projected residual* monotonicity test. Both *unconstrained* and *separable* nonlinear least squares problems are treated.

In the following Section 4.3, local convergence of *error* oriented Gauss-Newton methods is studied in affine covariant terms; again, Gauss-Newton methods are seen to exhibit guaranteed convergence only for a restricted problem class, named 'adequate' nonlinear least squares problems, since they are seen to be adequate in terms of the underlying statistical problem formulation. The globalization of these methods is done via the construction of two topological paths: the local and the global Gauss-Newton path. In the special case of nonlinear equations, the two paths coincide to one path, the Newton path. On this theoretical basis, adaptive trust region strategies (including rank strategies) combined with a natural extension of the *natural* monotonicity test are presented in detail for *unconstrained*, for *separable*, and—in contrast to the residual based approach—also for nonlinearly *constrained* nonlinear least squares problems. Finally, in Section 4.4, we study *underdetermined* nonlinear systems. In this case, a *geodetic Gauss-Newton path* exists generically and can be exploited to construct a quasi-Gauss-Newton algorithm and a corresponding adaptive trust region method.

Chapter 5. This chapter discusses the numerical solution of parameter dependent systems of nonlinear equations, which is the basis for parameter studies in systems analysis and systems design as well as for the globalization of local Newton methods. The key concept behind the approach is the (possible) existence of a *homotopy path* with respect to the selected parameter. In order to follow such a path, we here advocate *discrete continuation methods*, which consist of two essential parts:

- a *prediction* method, which, from given points on the homotopy path, produces some 'new' point assumed to be 'sufficiently close' to the homotopy path,
- an iterative *correction* method, which, from a given starting point close to, but not on the homotopy path, supplies some point on the homotopy path.

For the prediction step, *classical* or *tangent continuation* are the canonical choices. Needless to say that, for the iterative correction steps, we here concentrate on local Newton and (underdetermined) Gauss-Newton methods. Since the homotopy path is a mathematical object in the domain space of the nonlinear mapping, we only present the *affine covariant* approach.

In Section 5.1, we derive an adaptive *Newton continuation* algorithm with the ordinary Newton method as correction; this algorithm terminates locally in the presence of critical points including turning points. In order to follow the path beyond turning points, a *quasi-Gauss-Newton continuation* algorithm is worked out in Section 5.2, based on the preceding Section 4.4. This algorithm still terminates in the neighborhood of any higher order critical point. In order to overcome such points as well, we exemplify a scheme to construct *augmented systems*, whose solutions are just selected critical points of higher order—see Section 5.3. This scheme is an appropriate combination of Lyapunov-Schmidt reduction and topological universal unfolding. Details of numerical realization are only worked out for the computation of diagrams including simple bifurcation points.

PART II. The following Chapters 6 to 8 deal predominantly with *infinite dimensional*, i.e., function space oriented Newton methods. The selected topics are stiff initial value problems for ordinary differential equations (ODEs) and boundary value problems for ordinary and partial differential equations (PDEs).

Chapter 6. This chapter deals with *stiff* initial value problems for ODEs. The discretization of such problems is known to involve the solution of nonlinear systems per each discretization step—in one way or the other.

In Section 6.1, the contractivity theory for linear ODEs is revisited in terms of *affine similarity*. Based on an affine similar convergence theory for a simplified Newton method in *function space*, a *nonlinear contractivity* theory for stiff ODE problems is derived in Section 6.2, which is quite different from the theory given in usual textbooks on the topic. The key idea is to replace the Picard iteration in function space, known as a tool to show uniqueness in nonstiff initial value problems, by a simplified Newton iteration in function space to characterize stiff initial value problems. From this point of view, *linearly implicit* one-step methods appear as direct realizations of the simplified Newton iteration in function space. In Section 6.3, exactly the same theoretical characterization is shown to apply also to *implicit* one-step methods, which require the solution of a nonlinear system by some finite dimensional Newton-type method at each discretization step.

Finally, in a deliberately longer Section 6.4, we discuss *pseudo-transient continuation* algorithms, whereby steady state problems are solved via stiff integration. This type of algorithm is particularly useful, when the Jacobian matrix is singular due to hidden dynamical invariants (such as mass conservation). The (nearly) affine similar theoretical characterization permits the derivation of an *adaptive (pseudo-)time step strategy* and an accuracy matching strategy for a residual based inexact variant of the algorithm.

Chapter 7. In this chapter, we consider nonlinear two-point boundary value problems for ODEs. The presentation and notation is closely related to Chapter 8 in the textbook [71]. Algorithms for the solution of such problems can be grouped into two approaches: *initial value* methods such as multiple shooting and *global discretization* methods such as collocation. Historically, affine covariant Newton methods have first been applied to this problem class—with significant success.