运筹与管理科学丛书 29

非线性方程组数值方法

范金燕　袁亚湘　著

科学出版社

国家科学技术学术著作出版基金资助出版

运筹与管理科学丛书 29

非线性方程组数值方法

范金燕 袁亚湘 著

科学出版社
北京

内 容 简 介

非线性方程组在国防、经济、工程、管理等许多领域有着广泛的应用. 本书系统介绍非线性方程组的数值方法和相关理论, 主要内容包括: 牛顿法、拟牛顿法、高斯–牛顿法、Levenberg-Marquardt 方法、信赖域方法、子空间方法、非线性最小二乘问题、特殊非线性矩阵方程等.

本书可作为运筹学、计算数学、应用数学专业研究生和高年级本科生的教材, 也可供从事数学的教师、科研人员及工程技术人员参考.

图书在版编目(CIP)数据

非线性方程组数值方法/范金燕, 袁亚湘著. —北京: 科学出版社, 2018.2
(运筹与管理科学丛书; 29)
ISBN 978-7-03-056605-8

Ⅰ. ①非… Ⅱ. ①范… ②袁… Ⅲ. ①非线性方程–方程组–数值方法 Ⅳ. ①O175

中国版本图书馆 CIP 数据核字(2018) 第 033264 号

责任编辑: 王丽平 / 责任校对: 邹慧卿
责任印制: 张 伟 / 封面设计: 陈 敬

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

北京建宏印刷有限公司 印刷
科学出版社发行 各地新华书店经销
*
2018 年 2 月第 一 版 开本: 720 × 1000 B5
2018 年 2 月第一次印刷 印张: 14 3/4
字数: 281 000

定价: 98.00 元

(如有印装质量问题, 我社负责调换)

《运筹与管理科学丛书》序

运筹学是运用数学方法来刻画、分析以及求解决策问题的科学. 运筹学的例子在我国古已有之, 春秋战国时期著名军事家孙膑为田忌赛马所设计的排序就是一个很好的代表. 运筹学的重要性同样在很早就被人们所认识, 汉高祖刘邦在称赞张良时就说道:“运筹帷幄之中, 决胜千里之外.”

运筹学作为一门学科兴起于第二次世界大战期间, 源于对军事行动的研究. 运筹学的英文名字 Operational Research 诞生于 1937 年. 运筹学发展迅速, 目前已有众多的分支, 如线性规划、非线性规划、整数规划、网络规划、图论、组合优化、非光滑优化、锥优化、多目标规划、动态规划、随机规划、决策分析、排队论、对策论、物流、风险管理等.

我国的运筹学研究始于 20 世纪 50 年代, 经过半个世纪的发展, 运筹学研究队伍已具相当大的规模. 运筹学的理论和方法在国防、经济、金融、工程、管理等许多重要领域有着广泛应用, 运筹学成果的应用也常常能带来巨大的经济和社会效益. 由于在我国经济快速增长的过程中涌现出了大量迫切需要解决的运筹学问题, 因而进一步提高我国运筹学的研究水平、促进运筹学成果的应用和转化、加快运筹学领域优秀青年人才的培养是我们当今面临的十分重要、光荣, 同时也是十分艰巨的任务. 我相信,《运筹与管理科学丛书》能在这些方面有所作为.

《运筹与管理科学丛书》可作为运筹学、管理科学、应用数学、系统科学、计算机科学等有关专业的高校师生、科研人员、工程技术人员的参考书, 同时也可作为相关专业的高年级本科生和研究生的教材或教学参考书. 希望该丛书能越办越好, 为我国运筹学和管理科学的发展做出贡献.

袁亚湘

2007 年 9 月

前　　言

非线性方程组的求解是数值代数和数值优化的重要问题, 也是科学计算和计算数学的核心问题, 它在国防、经济、工程、管理等许多领域有着广泛的应用. 研究快速有效地求解非线性方程组的数值方法具有重要的理论意义和实际价值.

本书系统介绍了求解非线性方程组的数值方法和相关理论, 其中大部分内容是近年来国内外关于非线性方程组的前沿研究成果. 本书主要内容包括：牛顿法、拟牛顿法、高斯–牛顿法、Levenberg-Marquardt 方法、信赖域方法、子空间方法、非线性最小二乘问题、特殊非线性矩阵方程等. 其中第 4 章的 Levenberg-Marquardt 方法在局部误差界下的收敛性质、第 5 章的一般罚函数意义下的信赖域方法和一类信赖域半径趋于零的信赖域方法、第 7 章的可分离非线性最小二乘、第 8 章的大规模问题的子空间方法以及第 10 章的 Kohn-Sham 方程的自洽场迭代等一系列结果都是作者或与合作者完成的研究结果.

本书可作为运筹学、计算数学、应用数学专业研究生和高年级本科生的教材，也可供从事数学的教师、科研人员及工程技术人员参考.

作者的研究成果长期得到国家自然科学基金委员会的资助, 在此表示感谢. 本书的出版还得到国家科学技术学术著作出版基金的支持, 作者表示诚挚的谢意.

范金燕　袁亚湘

2018 年 1 月

目　录

第 1 章 导　　论

1.1 问　　题

非线性方程组的一般形式为

$$F(x) = 0, \tag{1.1.1}$$

其中 $F(x) = (F_1(x), \cdots, F_m(x))^\mathsf{T}$, x 是 n 维实向量, $F_i(x)$ 是 x 的实值函数, 且至少有一个不是线性的, m 和 n 是正整数.

若存在 $x^* \in \mathbb{R}^n$, 使得 $F(x^*) = 0$, 则称 x^* 为非线性方程组 (1.1.1) 的解.

非线性方程组广泛存在于国防、经济、工程、管理等许多重要领域. 例如, 非线性有限元问题、非线性断裂问题、弹塑性问题、电路问题、电子系统计算以及经济与非线性规划问题等都可归结为非线性方程组问题. 只要包含有未知函数及其导函数的非线性项的微分方程, 无论是用差分方法还是有限元方法, 离散化后得到的方程组都是非线性方程组. 因而, 求解形如 (1.1.1) 的非线性方程组的问题, 已经引起了人们广泛的重视. 早在 20 世纪 70 年代以前, 在理论上和数值解法上都对它做了大量的研究和系统总结 [181]. 分别在 1983 年、1995 年和 1999 年出版的三本专著 [51], [131] 和 [282] 对非线性方程组的求解方法也有系统的介绍.

非线性方程组问题可以转化为非线性最小二乘问题

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|F(x)\|_2^2. \tag{1.1.2}$$

非线性方程组 (1.1.1) 有解当且仅当极小化问题 (1.1.2) 的极小值为零. 非线性最小二乘问题大量出现在工程技术和科学实验之中, 它在曲线拟合、参数估计、函数逼近等方面有着广泛的应用.

1.2 方 法 概 述

与线性方程组相比, 非线性方程组的求解问题无论在理论上还是在解法上都不如线性方程组成熟和有效. 例如, 非线性方程组是否有解, 有多少解, 理论上都没有很好地解决. 线性方程组有许多直接解法, 而对于非线性方程组, 除了少数极特殊情形, 非线性方程组 (1.1.1) 的计算方法都是数值迭代法. 即给出一个初始点 $x_1 \in \mathbb{R}^n$, 然后由算法产生迭代点 $x_k(k = 2, 3, \cdots)$, 希望某个点 x_k 是 (1.1.1) 的解或 (1.1.2) 的稳定点, 或者 $x_k(k = 1, 2, \cdots)$ 收敛到 (1.1.1) 的解或 (1.1.2) 的稳定点.

本书主要介绍非线性方程组的迭代方法, 包括牛顿法、拟牛顿法、高斯–牛顿法、Levenberg-Marquardt 方法、信赖域方法、子空间方法以及其他一些方法. 另外, 还将介绍非线性最小二乘和几类特殊非线性矩阵方程的迭代方法. 书中的大部分内容是作者近年来的研究成果.

对于迭代方法, 通常都要讨论以下三个基本问题: 第一个问题是, 迭代点列的适定性问题, 即要求迭代点列是有意义的. 例如对于牛顿法, Jacobi 矩阵必须是非奇异的. 第二个问题是, 迭代点列的全局收敛性问题. 第三个问题是, 迭代点列的局部收敛速度问题. 近年来, 对算法的复杂度分析, 即分析算法得到给定精度的解所需要的计算量, 越来越受到研究者的重视. 虽然算法复杂度不是本书的重点, 我们对算法复杂度结果给予了简要介绍, 如 4.1.5 小节和 4.4.2 小节.

本书利用线搜索技巧和信赖域技巧, 研究迭代方法的全局收敛性. 线搜索的基本思想是每次迭代沿着某个方向搜索下一个迭代点. 在第 k 次迭代产生一个搜索方向 d_k, 通常 d_k 是某个近似问题的解. 然后沿 d_k 搜索一个“好”点, 即计算步长 $\alpha_k > 0$, 使得点 $x_k + \alpha_k d_k$ 在一定意义下比当前迭代点 x_k 好. 取这个新点 $x_k + \alpha_k d_k$ 为下一个迭代点 x_{k+1}, 就完成了一次迭代. 线搜索方法的关键在于 d_k 的产生和 α_k 的选取. 搜索方向 d_k 的产生往往依赖于子问题的构造. 步长 α_k 的确定取决于某种一维线搜索技巧和某一价值函数. 我们认为价值函数之值愈小的点愈“好”. 对于非线性方程组, 价值函数通常是 (1.1.2) 中的目标函数.

信赖域的思想是每次迭代在一个区域内试图找到一个好的点. 在第 k 次迭代, 算法在一个区域上寻找一个试探步. 该区域称为信赖域, 通常是以当前迭代点为中心的一个小邻域. 试探步往往要求是某个子问题在信赖域的解. 试探步求出后利用价值函数来判断它是否可以被接受. 试探步的好坏还用来决定如何调节信赖域. 粗略地说, 如果试探步较好, 则信赖域半径不变或扩大; 否则将缩小.

在 Levenberg-Marquardt 方法中, 利用信赖域思想, 根据价值函数的实际下降量与预估下降量的比值来判断 Levenberg-Marquardt 步的好坏, 并决定如何调节方法的参数. 粗略地说, 如果试探步较好, 则参数不变或缩小; 否则将扩大.

传统的信赖域方法, 当迭代点列收敛时, 信赖域半径将大于某一正常数. 本书将着重讨论信赖域半径趋于零的信赖域方法. 特别地, 将讨论 Levenberg-Marquardt 方法和信赖域方法在局部误差界条件下的收敛性质.

近年来, 随着很多包含大量变量的实际问题的出现, 大规模非线性方程组受到了越来越多专家学者的关注. 每次迭代, 子空间方法在一个低维的子空间里构造一个小规模问题, 使其尽可能地逼近原大规模问题. 当子空间的维数远远小于原空间的维数时, 计算量和存储量都会大幅度降低.

1.3 收敛性与收敛速度

非线性方程组数值方法的收敛性与收敛速度是非线性方程组求解方法的基本问题. 给定一个算法和初始点 x_1, 其所产生的点列 $x_k(k=1,2,\cdots)$ 是否收敛到非线性方程组的一个解或解集? 如果收敛, 收敛速度快不快? 这些问题无疑是方法的使用者十分关心的问题. 下面给出有关收敛性与收敛速度的一些定义.

假设 X^* 为非线性方程组 (1.1.1) 的解集. 用

$$\operatorname{dist}(x,X^*)=\inf_{y\in X^*}\|x-y\|_2 \tag{1.3.1}$$

表示点 x 到 X^* 的欧氏距离.

定义 1.3.1 设 $x_k\in\mathbb{R}^n(k=1,2,\cdots)$ 是一无穷点列. 如果

$$\liminf_{k\to\infty}\operatorname{dist}(x_k,X^*)=0, \tag{1.3.2}$$

则称 $\{x_k\}$ 子收敛到 X^*; 如果

$$\lim_{k\to\infty}\operatorname{dist}(x_k,X^*)=0, \tag{1.3.3}$$

则称 $\{x_k\}$ 弱收敛到 X^*; 如果存在 $x^*\in X^*$, 使得

$$\lim_{k\to\infty}x_k=x^*, \tag{1.3.4}$$

则称 $\{x_k\}$ 强收敛到 X^* (或 $\{x_k\}$ 收敛到解 x^*).

由定义可知, 强收敛必是弱收敛, 弱收敛必是子收敛. 假设算法产生的点列 $\{x_k\}$ 子收敛到非线性方程组的解集 X^*. 在实际计算中, 由于集合 X^* 未知, 所以 $\operatorname{dist}(x_k,X^*)$ 是不可计算的, 因而, 实用的收敛性定义是将 $\operatorname{dist}(x_k,X^*)$ 替换成某一评价函数, 例如 $\|F(x_k)\|_2$ 或 $\|J(x_k)^\mathsf{T}F(x_k)\|_2$, 其中 $J(x_k)$ 是 $F(x)$ 在 x_k 处的导数矩阵.

为了讨论局部收敛速度, 假定 $x_k\to x^*$. 令 $\epsilon_k=\|x_k-x^*\|_2$. 显然, $\{x_k\}$ 收敛到 x^* 的速度就是 ϵ_k 趋于 0 的速度. 假设 $x_k\in\mathbb{R}^n(k=1,2,\cdots)$ 由某一算法所产生, 如果有 k 使得 $x_k=x^*$, 则算法必在第 k 步迭代终止. 所以, 在讨论算法收敛速度时, 我们可假设对所有的 k, 都有 $\epsilon_k>0$. 下列有关收敛速度的定义是由 Ortega 和 Rheinboldt[181] 给出的.

定义 1.3.2 对于 $p\in[1,+\infty)$, 称

$$Q_p=\limsup_{k\to\infty}\frac{\epsilon_{k+1}}{\epsilon_k^p} \tag{1.3.5}$$

为点列 $\{x_k\}$ 的商收敛因子, 也称为 Q 因子. 称

$$R_p = \begin{cases} \limsup\limits_{k\to\infty} \epsilon_k^{1/k}, & p = 1, \\ \limsup\limits_{k\to\infty} \epsilon_k^{1/p^k}, & p > 1 \end{cases} \tag{1.3.6}$$

为 $\{x_k\}$ 的根收敛因子, 简称 R 因子.

定义 1.3.3　称

$$O_Q = \inf\{p \mid p \in [1, +\infty), Q_p = +\infty\} \tag{1.3.7}$$

为点列 $\{x_k\}$ 的商收敛阶, 简称 Q 收敛阶.

$$O_R = \inf\{p \mid p \in [1, +\infty), R_p = 1\} \tag{1.3.8}$$

为点列 $\{x_k\}$ 的根收敛阶, 简称 R 收敛阶.

定义 1.3.4　如果 $Q_1 = 0$, 则称 $\{x_k\}$ 是 Q 超线性收敛到 x^*; 如果 $0 < Q_1 < 1$, 则称 $\{x_k\}$ 是 Q 线性收敛到 x^*; 如果 $Q_1 \geqslant 1$, 则称 $\{x_k\}$ 是 Q 次线性收敛到 x^*.

定义 1.3.5　如果 $R_1 = 0$, 则称 $\{x_k\}$ 是 R 超线性收敛到 x^*; 如果 $0 < R_1 < 1$, 则称 $\{x_k\}$ 是 R 线性收敛到 x^*; 如果 $R_1 = 1$, 则称 $\{x_k\}$ 是 R 次线性收敛到 x^*.

定义 1.3.6　如果 $Q_2 = 0$, 则称 $\{x_k\}$ 是 Q 超平方收敛到 x^*; 如果 $0 < Q_2 < +\infty$, 则称 $\{x_k\}$ 是 Q 平方收敛到 x^*; 如果 $Q_2 = +\infty$, 则称 $\{x_k\}$ 是 Q 次平方收敛到 x^*.

定义 1.3.7　如果 $R_2 = 0$, 则称 $\{x_k\}$ 是 R 超平方收敛到 x^*; 如果 $0 < R_2 < 1$, 则称 $\{x_k\}$ 是 R 平方收敛到 x^*; 如果 $R_2 \geqslant 1$, 则称 $\{x_k\}$ 是 R 次平方收敛到 x^*.

平方收敛也称为二次收敛. 类似可定义 Q 和 R 立方收敛. 从定义可知, $\{x_k\}$ Q 超线性收敛到 x^* 等价于

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|_2}{\|x_k - x^*\|_2} = 0. \tag{1.3.9}$$

式 (1.3.9) 与

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|_2}{\|x_k - x_{k+1}\|_2} = 0 \tag{1.3.10}$$

是等价的. 从 (1.3.10) 可知, 对于 Q 超线性收敛点列 $\{x_k\}$, 如果 $\|x_k - x_{k+1}\|_2$ 已相当小, 则 x_{k+1} 必定是 x^* 的一个很好的近似.

利用 Q 收敛阶和 R 收敛阶的定义, 有

推论 1.3.1　对于收敛点列 $x_k (k = 1, 2, \cdots)$, 它的 Q 收敛阶不大于其 R 收敛阶, 即

$$O_Q \leqslant O_R. \tag{1.3.11}$$

该推论的证明可见文献 [181]. 显然, R 收敛阶很高时有可能 Q 收敛阶很低. 但是有如下结果:

定理 1.3.1 设点列 $x_k(k=1,2,\cdots)$ 收敛到 x^*, 而且它的 R 收敛阶 $Q_R>1$. 则对任给常数 $\delta<O_R$, 存在点列 $\beta_k(k=1,2,\cdots)$, 使得

$$\beta_k \geqslant \|x_k - x^*\|_2 \tag{1.3.12}$$

对所有充分大的 k 都成立, 而且 β_k 收敛到 0 的 Q 收敛阶不小于 δ.

该定理的证明可见文献 [285]. 这个结果说明, R 收敛阶高的点列至少比某些特殊的 Q 收敛阶低的点列收敛得更快.

在以后的各章, 除特别说明外, 我们提到的线性收敛、超线性收敛以及二次收敛分别是指 Q 线性收敛、Q 超线性收敛以及 Q 二次收敛; 我们提到的向量范数 $\|\cdot\|$ 和矩阵范数 $\|\cdot\|$ 分别是指向量 2 范数 $\|\cdot\|_2$ 和矩阵 2 范数 $\|\cdot\|_2$.

第2章 牛 顿 法

牛顿法是求解非线性方程组的最基本方法. 目前使用的很多有效的迭代方法都是以牛顿法为基础并由它发展而来的. 本章我们讨论牛顿法和非精确牛顿法以及它们的收敛性质. 在本章, 我们假定方程的个数与变量个数相等, 即在 (1.1.1) 中 $m=n$.

2.1 牛 顿 法

牛顿法的基本思想是利用函数的线性展开, 在第 k 次迭代, 用线性方程组

$$L_k(x)=F_k+J_k(x-x_k)=0 \tag{2.1.1}$$

近似方程组 (1.1.1) , 从而得到第 $k+1$ 个近似解

$$x_{k+1}=x_k-J_k^{-1}F_k, \tag{2.1.2}$$

这里 $F_k=F(x_k), J_k=F'(x_k)$ 是 Jacobi 矩阵在 x_k 处的值. 它的几何意义是在 $n+1$ 维空间 (x,z), 利用 n 个超切平面

$$z=F_i(x_k)+\nabla F_i(x-x_k), \quad i=1,\cdots,n \tag{2.1.3}$$

与超平面 $z=0$ 的交点 $(x_{k+1},0)$ 作为超曲面 $z=F_i(x)$ 与超平面 $z=0$ 的交点的近似. 牛顿法要求 Jacobi 矩阵 J_k 非奇异, 且每次迭代要计算 J_k 的逆, 当 n 很大时会出现计算上的困难. 实际计算时, 一般求解下述 n 阶线性方程组

$$J_kd=-F_k \tag{2.1.4}$$

得到牛顿步 d_k.

算法 2.1.1 (牛顿法)

步 1 给出 $x_1\in\mathbb{R}^n$; $k:=1$.

步 2 如果 $F_k=0$, 则停; 求解线性方程组 (2.1.4), 得到牛顿步 d_k.

步 3 令 $x_{k+1}:=x_k+d_k$, $k:=k+1$, 转步 2.

牛顿法的优点是收敛快. 下面给出牛顿法的局部收敛性结果 [51].

定理 2.1.1 设 $F:\mathbb{R}^n\to\mathbb{R}^n$ 在开凸集 $D\subset\mathbb{R}^n$ 上连续可微, 存在 $x^*\in\mathbb{R}^n$ 及常数 $r,\kappa_{bij},\kappa_{lj}>0$ 使得 $N(x^*,r)\subset D, F(x^*)=0, J(x^*)$ 非奇异, $\|J(x^*)^{-1}\|\leqslant\kappa_{bij}$, 且有

$$\|J(x)-J(y)\| \leqslant \kappa_{lj}\|x-y\|, \quad \forall x,y \in N(x^*,r), \tag{2.1.5}$$

其中 $N(x^*,r)=\{x\in\mathbb{R}^n \mid \|x-x^*\|<r\}$, 则存在 $\epsilon>0$, 使得对任意 $x_1\in N(x^*,\epsilon)$, 由算法 2.1.1 产生的迭代点列 $\{x_k\}$ 收敛到 x^*, 且满足

$$\|x_{k+1}-x^*\| \leqslant \kappa_{bij}\kappa_{lj}\|x_k-x^*\|^2. \tag{2.1.6}$$

证明 令

$$\epsilon=\min\left\{r, \frac{1}{2\kappa_{bij}\kappa_{lj}}\right\}. \tag{2.1.7}$$

用数学归纳法证明, 对所有 k, 均有 $x_k\in N(x^*,\epsilon)$, J_k 非奇异, 且式 (2.1.6) 成立.

先证 J_1 非奇异. 因为 $\|x_1-x^*\|\leqslant\epsilon$, $J(x)$ 在 x^* 处 Lipschitz 连续, 所以由 (2.1.7) 可知

$$\begin{aligned}\|J(x^*)^{-1}(J_1-J(x^*))\| &\leqslant \|J(x^*)^{-1}\|\|J_1-J(x^*)\|\\ &\leqslant \kappa_{bij}\kappa_{lj}\|x_1-x^*\| \leqslant \kappa_{bij}\kappa_{lj}\epsilon \leqslant \frac{1}{2}.\end{aligned} \tag{2.1.8}$$

由矩阵逆扰动理论 (见文献 [51, Theorem 3.1.4]) 知 J_1 非奇异, 且

$$\|J_1^{-1}\| \leqslant \frac{\|J(x^*)^{-1}\|}{1-\|J(x^*)^{-1}(J_1-J(x^*))\|} \leqslant 2\|J(x^*)^{-1}\| \leqslant 2\kappa_{bij}. \tag{2.1.9}$$

因此, x_2 有定义, 且

$$\begin{aligned}x_2-x^* &= x_1-x^*-J_1^{-1}F_1\\ &= J_1^{-1}(F(x^*)-F_1-J_1(x^*-x_1)).\end{aligned} \tag{2.1.10}$$

从而由引理 (见文献 [51, Lemma 4.1.12]) 和 (2.1.9) 可知

$$\begin{aligned}\|x_2-x^*\| &\leqslant \|J_1^{-1}\|\|F(x^*)-F_1-J_1(x^*-x_1)\|\\ &\leqslant 2\kappa_{bij}\cdot\frac{\kappa_{lj}}{2}\|x_1-x^*\|^2\\ &= \kappa_{bij}\kappa_{lj}\|x_1-x^*\|^2.\end{aligned} \tag{2.1.11}$$

因为 $\|x_1-x^*\| \leqslant \dfrac{1}{2\kappa_{bij}\kappa_{lj}}$, 所以

$$\|x_2-x^*\| \leqslant \frac{1}{2}\|x_1-x^*\|, \tag{2.1.12}$$

故 $x_2\in N(x^*,\epsilon)$.

类似可证 (2.1.6) 和

$$\|x_{k+1} - x^*\| \leqslant \frac{1}{2}\|x_k - x^*\| \tag{2.1.13}$$

对所有 k 成立. 因此, $\{x_k\}$ 二次收敛到 x^*. □

定理 2.1.1 假设了非线性方程组 (1.1.1) 解的存在性. 如果 Jacobi 矩阵在解 x^* 处非奇异且在 x^* 的某个邻域内 Lipschitz 连续, 则牛顿法二次收敛.

Kantorovich [127] 给出了牛顿法的另一个著名的收敛性定理. Kantorovich 定理与定理 2.1.1 的主要区别是它没有假设非线性方程组 (1.1.1) 解的存在性. 如果 Jacobi 矩阵在初始点 x_1 处非奇异且在 x_1 的某邻域内 Lipschitz 连续, 初始牛顿步相对于 F 的非线性性充分小, 则非线性方程组 (1.1.1) 在 x_1 的某个邻域内解存在且唯一. 但在此条件下, 由牛顿法产生的迭代点列只具有 R 二次收敛性. 通常称定理 2.1.1 为牛顿法的局部收敛性定理, 称如下的 Kantorovich 定理为牛顿法的半局部收敛性定理.

定理 2.1.2 设 $F:\mathbb{R}^n \to \mathbb{R}^n$ 在开凸集 $D \subset \mathbb{R}^n$ 上二次连续可微. 如果存在 $x_1 \in D$ 及常数 B, K, η 使得 $F_1 \neq 0, F'(x_1)$ 非奇异且满足

$$\|J_1\| \leqslant B, \quad \|\nabla^2 F(x)\| \leqslant K, \quad \|J_1^{-1}F_1\| \leqslant \eta. \tag{2.1.14}$$

又如果 $\overline{N(x_1, t^*)} \subset D$, 其中 $t^* = \dfrac{2\eta}{1+\sqrt{1-2h}}, h = KB\eta \leqslant \dfrac{1}{2}$, 则由算法 2.1.1 产生的迭代点列 $\{x_k\}$ 均在闭球 $\overline{N(x_1, t^*)}$ 内, 并收敛到 $F(x) = 0$ 在邻域 $\overline{N(x_1, t^{**})} \cap D$ 内的唯一解 x^*, 其中 $t^{**} = \dfrac{1+\sqrt{1-2h}}{KB}$, 且有

$$\|x_k - x^*\| \leqslant \frac{2\eta_k}{1+\sqrt{1-2h_k}} \leqslant 2^{1-k}(2h)^{2^k-1}\eta, \tag{2.1.15}$$

其中 $\{\eta_k\}$ 和 $\{h_k\}$ 由如下的递推公式所定义:

$$B_0 = B, \quad \eta_0 = \eta, \quad h_0 = h, \tag{2.1.16}$$

$$B_k = \frac{B_{k-1}}{1-h_{k-1}}, \quad \eta_k = \frac{h_{k-1}\eta_{k-1}}{2-(1-h_{k-1})}, \quad h_k = KB_k\eta_k. \tag{2.1.17}$$

Kantorovich 定理的优点是无需假设 (1.1.1) 的解 x^* 的存在性和 $J(x^*)$ 的非奇异性. 事实上, 确实有方程组存在某个点满足定理 2.1.2 的假设条件, 但其在解处的 Jacobi 矩阵是奇异的. Kantorovich 定理的不足之处是迭代点列的 R 收敛性结果并不能显示点列的变化情况. Dennis[46] 证明了此时点列满足

$$\frac{1}{2}\|x_{k+1} - x_k\| \leqslant \|x_{k+1} - x^*\| \leqslant 2\|x_{k+1} - x_k\|. \tag{2.1.18}$$

许多学者基于 Kantorovich 定理对牛顿法进行了深入的研究. 关于 Kantorovich 定理的各种变形和推广, 以及牛顿迭代点列的误差估计、误差分析、计算复杂性和牛顿法的应用可参阅 [25, 51, 96, 131, 181].

牛顿法的优点是收敛快, 特别对于 F 的分量为仿射函数的方程组, 牛顿法只需一次迭代就可得到它的解. 另外牛顿法具有仿射不变性, 即对 F 作仿射变换 $G = AF$, 其中 A 为非奇异矩阵, 则

$$(G'(x))^{-1}G(x) = J(x)^{-1}A^{-1}AF(x) = J(x)^{-1}F(x). \tag{2.1.19}$$

这说明牛顿迭代点列在仿射变换下不变, 因而其收敛性和发散性不变.

牛顿法的缺点是, 对于许多问题, 它不一定全局收敛. 选取价值函数 $\phi(x) = \frac{1}{2}\|F(x)\|^2$. 根据无约束优化的全局收敛性策略, 一般要求迭代点列 $\{x_k\}$ 满足

$$\|F(x_k + d_k)\| \leqslant \|F(x_k)\|. \tag{2.1.20}$$

一个常用的全局收敛性策略是线搜索策略. 如果 $F_k \neq 0$, 则有

$$\nabla\phi_k^{\intercal} d_k = -F_k^{\intercal} J_k J_k^{-1} F_k = -F_k^{\intercal} F_k < 0. \tag{2.1.21}$$

上式表明牛顿方向 d_k 是 ϕ 在点 x_k 处的下降方向. 事实上, 牛顿方向 d_k 也是 F 在 x_k 处的线性模型

$$M_k(x_k + d) = F_k + J_k d \tag{2.1.22}$$

的根, 故 d_k 是二次模型

$$\begin{aligned} m_k(x_k + d) =& \frac{1}{2}\|M_k(x_k + d)\|^2 \\ =& \frac{1}{2}F_k^{\intercal} F_k + (J_k^{\intercal} F_k)^{\intercal} d + \frac{1}{2} d^{\intercal} J_k^{\intercal} J_k d \end{aligned} \tag{2.1.23}$$

的极小点. 因此可沿牛顿方向进行线搜索, 使得迭代点列

$$x_{k+1} = x_k - \alpha_k J_k^{-1} F_k, \quad k = 1, 2, \cdots \tag{2.1.24}$$

满足某种线搜索准则, 其中 $\alpha_k > 0$ 是一维搜索步长. 如果选取 α_k, 使得 ϕ_{k+1} 充分小于 ϕ_k, 则由线搜索牛顿法 (2.1.24) 产生的迭代点列 $\{x_k\}$ 的任一聚点 $\bar{x}^*$ 必满足 $F(\bar{x}^*) = 0$ 或 $J(\bar{x}^*)$ 奇异. 因此, 如果 $J(x)$ 对所有 $x \in \mathbb{R}^n$ 非奇异, 水平集 $L(x_1) = \{x \in \mathbb{R}^n \mid \phi(x) \leqslant \phi(x_1)\}$ 有界, 则 $\{x_k\}$ 收敛到非线性方程组 (1.1.1) 的解, 且当 k 充分大时, $\alpha_k \equiv 1$. 因此, 线搜索牛顿法仍具有二次收敛速度 [51].