Special Topics in Continuous Optimization and Optimal Control

Local and Global convergence of Newton methods

Mitodru Niyogi   Sumet Khumphairan

Faculty of Mathematics and Computer Science, Heidelberg University

December 7, 2019

Overview

1. Introduction
2. Background
3. Local Conv.
4. Failures: NM
5. Theorems
6. Global Conv.
7. Globalization Schemes
8. Conclusion

Introduction

Problem:

- Finding x^* such that $F(x^*) = 0$
- $F : \mathbb{R}^n \to \mathbb{R}^n$ (highly) nonlinear
- Important problem in continuous optimization and optimal control

Newton's method:

- Iterative method
- Solve linearized problem
- Many variants

Globalization strategies:

- Newton's method only locally convergent
- Hope to globalize convergence

Lipschitz Condition

- Given g: [a,b] $\rightarrow \mathbb{R}$ is called Lipschitz continuous with constant $\lambda > 0$ (denoted g ϵ Lip_λ[a,b]) if $\exists\lambda > 0$ such that $|g(x) - g(y)| \leq \lambda|x - y|$ for all x,y ϵ [a,b]

Contraction map

- g: [a,b] $\rightarrow \mathbb{R}$ is called contraction map if g ϵ Lip_λ[a,b] with $\lambda < 1$

Convex set

- A set C is convex if, for any x,y ϵ C and θ ϵ $\mathbb{R}$ with $0 \leq \theta \leq 1$, $\theta x + (1 - \theta)y\epsilon$ C

Convex function

- A function f: $\mathbb{R} \rightarrow \mathbb{R}$ is convex if its domain (denoted D(f)) is a convex set and if, for all x,y ϵ D(f) and θ ϵ $\mathbb{R}$ with $0 \leq \theta \leq 1$,
- $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$

Summary: Newton Method

- Fast (i.e. quadratic) local rate of convergence
- Scale-invariant w.r.t linear transformations of the variables
- Search direction p^k is not well defined if $\nabla^2 \mathrm{f}(\mathrm{x}^\mathrm{k})$ is singular, p^k is not a descent if $\nabla^2 \mathrm{f}(\mathrm{x}^\mathrm{k})$ is not positive definite
- Minimum points x^k can be attracted to sadle points or local maxima of f
- Very small neighbourhood of local convergence, Newton's method is not globally convergent
- Line search, trust region

1. Convergence properties of Newton methods

What is convergence?

- Convergence means **approaching a limit** as the argument of the function increases or decreases or as the number of terms in the series increases.
- Types of convergence: local and global
- **When does it converge locally?** When the initial approximation is already **close enough to the solution**, then the succesive aproximations of the iterative method guranteed to converge to a solution locally.
- Iterative methods for nonlinear equations and their systems, such as Newton's method are usually only locally convergent.

Rate of convergence

- $\{x^k\} \subset \mathbb{R}^n, x^* \epsilon \mathbb{R}^n, \{x^k\} \to x^*$ as k $\to \infty$
- $\{x^k\} \to x^k$ with rate r if $\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|^r} = c < \infty$, for sufficiently large k
- r=1: linear convergence (c < 1)
- r=2: quadratic convergence
- superlinear convergence: $\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|^r} \to 0$ as k $\to \infty$

Local Convergence Theorem

- Single dimension: If
- $\frac{\partial f}{\partial x}$ bounded away from zero
- $\frac{\partial^2 f}{\partial x^2}$ bounded
- Then Newton's method converges given a sufficiently close initial guess (and convergence is quadratic)

Multidimensional

- If $\|J_F^{-1}(x^k)\| \leq \beta$ (inverse is bounded)
- $\|J_F(x) - J_F(y)\| \leq l\|x - y\|$ (Derivative is Lipschitz continuous)

Example 1

$$f(x) = x^2 - 1 = 0, \quad find \ \ x \ \ (x^* = 1)$$

$$\frac{df}{dx}(x^k) = 2x^k$$

$$2x^k(x^{k+1} - x^k) = -\left(\left(x^k\right)^2 - 1\right)$$

$$2x^k(x^{k+1} - x^*) + 2x^k(x^* - x^k) = -\left(\left(x^k\right)^2 - \left(x^*\right)^2\right)$$

$$or \ \ (x^{k+1} - x^*) = \frac{1}{2x^k}(x^k - x^*)^2$$

- We see that the convergence is quadratic

Example 2

$$f(x) = x^2 = 0, \quad x^* = 0$$

$$\frac{df}{dx}(x^k) = 2x^k$$

$$\Rightarrow \; 2x^k(x^{k+1} - 0) = (x^k - 0)^2$$

$$x^{k+1} - 0 = \frac{1}{2}\left(x^k - 0\right) \qquad \text{for } x^k \neq x^* = 0$$

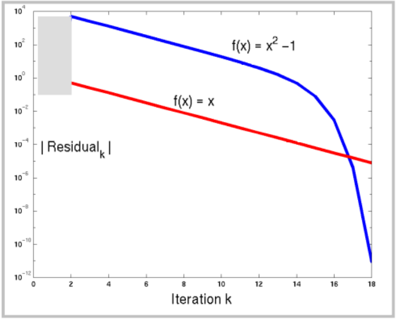
$$or \; (x_{k+1} - x^*) = \frac{1}{2}(x_k - x^*)$$

- Note: $\frac{\partial f}{\partial x}^{-1}$ not bounded away from zero
- We see the convergence is linear

Plot

Newton-Raphson Method – Convergence

Example 1,2



December 6, 2019 <number> courtesy Alessandra Nardi UCB

Convergence algorithm

Newton-Raphson Method – Convergence

$x^0 =$ Initial Guess, $k = 0$

Repeat {

$$\frac{\partial f(x^k)}{\partial x}(x^{k+1} - x^k) = -f(x^k)$$

$$k = k + 1$$

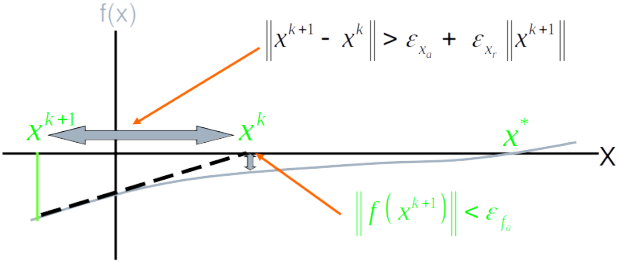
} Until ?

$\left\|x^{k+1} - x^k\right\| < threshold$? $\left\|f(x^{k+1})\right\| < threshold$?

Newton-Raphson Method – Convergence

Convergence Check

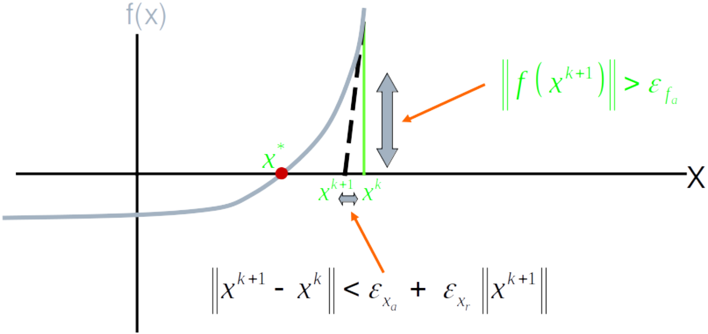
Need a "delta-x" check to avoid false convergence



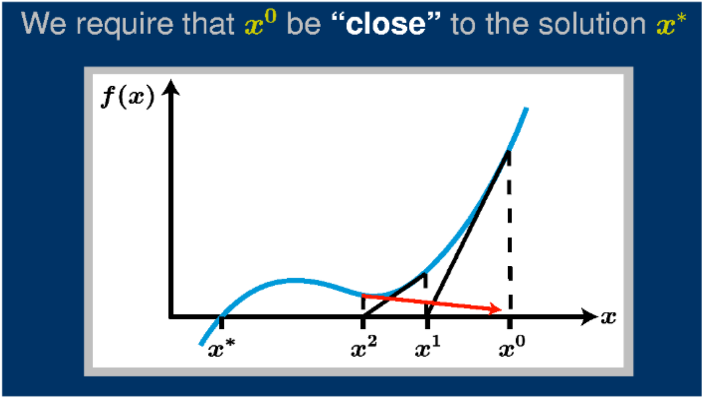
Newton-Raphson Method – Convergence

Convergence Check

Also need an "$f(x)$" check to avoid false convergence



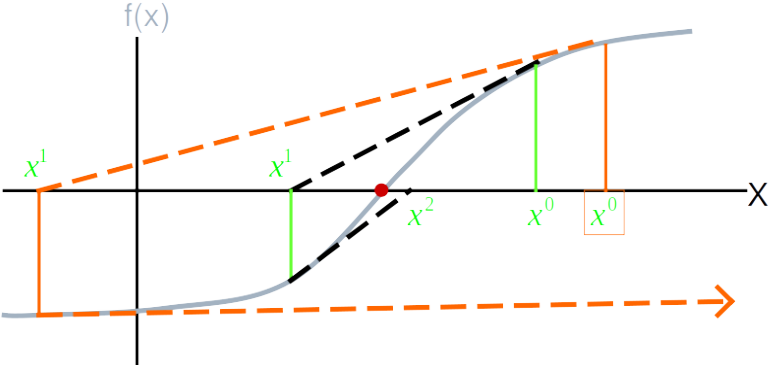
Newton-Raphson Method – Convergence



Newton-Raphson Method – Convergence

Local Convergence

Convergence Depends on a Good Initial Guess

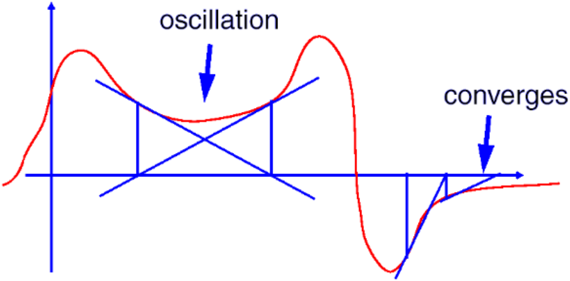


Newton-Raphson Method – Convergence

Local Convergence

Convergence Depends on a Good Initial Guess

Example:



Advantages of Newton Method:

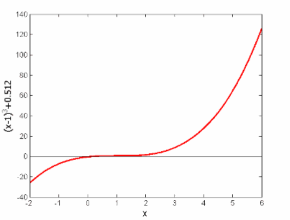
- Quadratic converges
- Requires only one guess
- This is very fast if we are close to a solution
- Doubles the correct digits in each iteration!

Drawbacks of Newton methods:

- **Divergence at inflection points**: If the initial guess or an iteration value of the root that is close to the inflection point of the function.

Table 1 Divergence near inflection point.

Iteration Number	x_i
0	5.0000
1	3.6560
2	2.7465
3	2.1084
4	1.6000
5	0.92589
6	−30.119
7	−19.746
18	0.2000



Divergence at inflection point for $f(x) = (x-1)^3 + 0.512 = 0$

26

Figure: Divergence at inflection points

Drawbacks: Division by zero

- $f(x) = x^3 - 0.03x^2 + 2.4 \times 10^{-6} = 0$

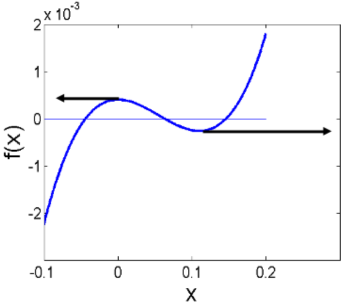


Figure: Pitfall of division by zero or near a zero number

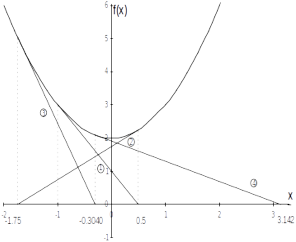
- For $x_0 = 0 or x_0 = 0.02$ the denominator will be zero

Drawbacks: Oscillations

- Oscillations near local maxima and minimum
- Results may oscillate abut the local maximum or minimum without converging on a root but converging on the local maximum or minimum.
- Leads to division by a number close to zero and may diverge.
- $f(x) = x^2 + 2 = 0$ has no real roots

Table 2 Oscillations near local maxima and mimima

Iteration Number	x_i	$f(x_i)$	$\lvert\in_a\rvert\%$
0	–1.0000	3.00	
1	0.5	2.25	300.00
2	–1.75	5.063	128.571
3	–0.30357	2.092	476.47
4	3.1423	11.874	109.66
5	1.2529	3.570	150.80
6	–0.17166	2.029	829.88
7	5.7395	34.942	102.99
8	2.6955	9.266	112.93
9	0.97678	2.954	175.96



<number>

Figure: Oscillations around local minima for $f(x) = x^2 + 2$

Drawbacks: Root jumping

- In some cases, where the function f(x) is oscillating and has a number of roots, one may choose an initial guess close to a root.
- However, the guesses may jump and converge to some other root.

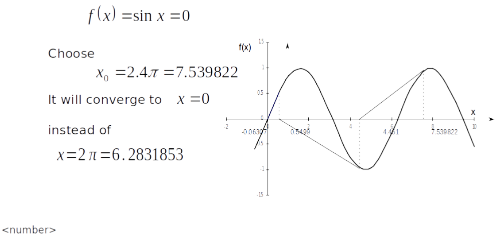


Figure: Root jumping from intended location of root for f(x)=$\sin$ x

Theorem

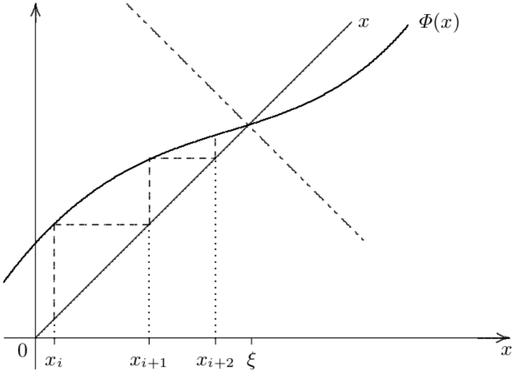
- Theorem. Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be an iterative function with fixed point ξ. U(ξ) is a neighbourhood of ξ, a number p $\geq$ 1 and a constant C $\geq$ 0 (with C $\leq$ 1 if p = 1) so that for all x ϵ U(ξ)

$$\|\Phi(x) - \xi\| \leq C\|x - \xi\|^p \tag{1}$$

- Then there is a neighbourhood (subset) V(ξ) $\subset$ U(ξ) of ξ so that for all starting points x_0 the iteration method defined by Φ geneerates the iteration steps x_i ϵ V(ξ) $\forall$ i $\geq$ 0 that converges to ξ at least with order p

Example

- E=$\mathbb{R}$, Φ is differentiable in a neighbourhood U(ξ) If $0 < \Phi'(\xi) < 1$, then the convergence will be linear (first order), x_i will converge monotonically to ξ



-

General Convergence Theorem

- Let the function Φ: E $\rightarrow$ E, E= $\mathbb{R}^n$, have a fixed point ξ: Φ $(\xi) = \xi$
- Further let $S_r(\xi)$:={ z ||| z - ξ|| < r} be a neighbourhood of ξ such that Φ is a contractive mapping in S_r (ξ) that is

$$\|\Phi(x) - \Phi(y)\| \leq K\|x - y\| \quad (2)$$

0 $\leq$ K < 1 for all x,y ϵ $S_r(\xi)$. Then for any x_0 ϵ $S_r(\xi)$, the generated sequence $x_{i+1} = \Phi(x_i)$, i =0,1,.., has the following properties

- $x_i \epsilon$ $S_r(\xi)$ $\forall$ i=0,1,..,
- $\|x_{i+1} - \xi\| \leq K\|x_i - \xi\| \leq K^{i+1}\|x_0 - \xi\|$ i.e., $\{x_i\}$ converges at least linearly to ξ

Banach Fixed Point Theorem

- Let $\Phi : E \to E$, $E = \mathbb{R}^n$ be an iterative fucnction, $x_0 \epsilon E$ be a starting point, and $x_{i+1} = \Phi(x_i)$, i=0,1, ... Further,let a neighbourhood $S_r(x_0) = \{x | \|x - x_0\| <$ r$\}$ of x_0 and a constant K where 0 < K < 1, exist such that
  - $\|\Phi(x) - \Phi(y)\| \leq K\|x - y\|$ for all x,y ϵ $\overline{S_r(x_0)} := \{x | \|x - x_0\| \leq r\}$
  - $\|x_1 - x_0\| = \|\Phi(x_0) - x_0\| \leq$ (1 - K)r < r

  Then it follows that
  - $x_i \ \epsilon \ S_r(x_0) \forall$ i = 0, 1, ...,
  - Φ has exactly one fixed point ξ, $\Phi(\xi) = \xi$, in $\overline{S_r(x_0)}$ and $\lim_{i\to\infty} x_i = \xi$,

    $\|x_{i+1} - \xi\| \leq K\|x_i - \xi\|$, as well as $\|x_1 - \xi\| \leq \frac{K^i}{1-K}\|x_1 - x_0\|$

Proof of quadratic convergence

Theorem. Assume that f is twice continuously differentiable on an open interval (a, b) and that there exists $x^* \in (a, b)$ with $f'(x^*) \neq 0$. Define Newton's method by the sequence

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \qquad k = 1, 2, \dots$$

Assume also that x_k converges to x^* as $k \to \infty$. Then, for k sufficiently large,

$$|x_{k+1} - x^*| \leq M|x_k - x^*|^2 \qquad \text{if } M > \frac{|f''(x^*)|}{2|f'(x^*)|}.$$

Thus, x_k converges to x^* *quadratically* (A&G, p. 52).

Proof. Let $e_k = x_k - x^*$, so that $x_k - e_k = x^*$. By Taylor's Theorem (A&G, Chap. 1, p. 5), setting $x = x_k$ and $h = -e_k$, we have

$$f(x_k - e_k) = f(x_k) - e_k f'(x_k) + \frac{(e_k)^2}{2} f''(\xi_k)$$

for some ξ_k between x_k and x^*. Since $x_k - e_k = x^*$ and $f(x^*) = 0$, we have

$$0 = f(x_k) - (x_k - x^*) f'(x_k) + \frac{(e_k)^2}{2} f''(\xi_k).$$

Since the derivative of f is continuous with $f'(x^*) \neq 0$, we have $f'(x_k) \neq 0$ as long as x_k is close enough to x^*. So we can divide by $f'(x_k)$ to give

$$0 = \frac{f(x_k)}{f'(x_k)} - (x_k - x^*) + \frac{(e_k)^2 f''(\xi_k)}{2f'(x_k)},$$

which, by the definition of Newton's method, gives

$$x_{k+1} - x^* = \frac{(e_k)^2 f''(\xi_k)}{2f'(x_k)}.$$

So

$$|x_{k+1} - x^*| \leq \frac{|f''(\xi_k)|}{2|f'(x_k)|} |x_k - x^*|^2.$$

By continuity, $f'(x_k)$ converges to $f'(x^*)$ and, since ξ_k is between x_k and x^*, ξ_k converges to x^* and hence $f''(\xi_k)$ converges to $f''(x^*)$, so, for large enough k,

$$|x_{k+1} - x^*| \leq M|x_k - x^*|^2 \qquad \text{if } M > \frac{|f''(x^*)|}{2|f'(x^*)|}.$$

In fact, it can be shown without assuming that x_k converges to x^*, that there exists $\delta > 0$ such that, if $|x_0 - x^*| \leq \delta$, then x_k converges to x^*, and hence from the above argument that the convergence rate is quadratic, but this requires a more complicated argument by induction.

1. Minimization Problems
2. Global Convergence Definition
3. Globalization Schemes
4. Conclusions

Minimization Problems

We consider the following minimization problem for a real function $h : \mathbb{R}^n \to \mathbb{R}$ of n variables

$$\min_x h(x).$$

Definition of Global Convergence

- The iterative method is called locally convergent with $V(\overline{x})$, a neigborhood of $\overline{x}$, if it generates, for all starting points $x_0 \in V(\overline{x})$, a sequence $\{x_k\}$ that converges to $\overline{x}$.

[Note that $\overline{x}$ is a minimum point for the function h.]

Definition of Global Convergence

- The iterative method is called locally convergent with $V(\overline{x})$, a neigborhood of $\overline{x}$, if it generates, for all starting points $x_0 \in V(\overline{x})$, a sequence $\{x_k\}$ that converges to $\overline{x}$.
- It is called globally convergent, if in addition $V(\overline{x}) = \mathbb{R}^n$.

[Note that $\overline{x}$ is a minimum point for the function h.]

Important Lemma

- Let $h : \mathbb{R}^n \to \mathbb{R}$ be a function which has a continuous derivative $Dh(x)$ for all $x \in V(\overline{x})$.

[Note that $D(\gamma, x) := \{s \in \mathbb{R}^n | \|s\| = 1, Dh(x)s \geq \gamma \|Dh(x)\|\}$ and $Dh(x) = \nabla h(x)^T = (\frac{\partial h(x)}{\partial x^1}, \cdots, \frac{\partial h(x)}{\partial x^n})$.]

Important Lemma

- Let $h : \mathbb{R}^n \to \mathbb{R}$ be a function which has a continuous derivative $Dh(x)$ for all $x \in V(\overline{x})$.
- Suppose further that $Dh(\overline{x}) \neq 0$, and let $1 \geq \gamma > 0$.

[Note that $D(\gamma, x) := \{s \in \mathbb{R}^n | \|s\| = 1, Dh(x)s \geq \gamma \|Dh(x)\|\}$ and $Dh(x) = \nabla h(x)^T = (\frac{\partial h(x)}{\partial x^1}, \cdots , \frac{\partial h(x)}{\partial x^n})$.]

Important Lemma

- Let $h : \mathbb{R}^n \to \mathbb{R}$ be a function which has a continuous derivative $Dh(x)$ for all $x \in V(\overline{x})$.
- Suppose further that $Dh(\overline{x}) \neq 0$, and let $1 \geq \gamma > 0$.

Then there is a neighborhood $U(\overline{x}) \subseteq V(\overline{x})$ of $\overline{x}$ and a number $\lambda > 0$ such that

$$h(x - \mu s) \leq h(x) - \frac{\mu\gamma}{4} \|Dh(\overline{x})\|$$

for all $x \in U(\overline{x}), s \in D(\gamma, x)$ and $0 \leq \mu \leq \lambda$.
[Note that $D(\gamma, x) := \{s \in \mathbb{R}^n | \|s\| = 1, Dh(x)s \geq \gamma \|Dh(x)\|\}$ and $Dh(x) = \nabla h(x)^T = \left(\frac{\partial h(x)}{\partial x^1}, \cdots, \frac{\partial h(x)}{\partial x^n} \right)$.]

Globalization Schemes

- Modified Newton Method with Exact Line Search

Globalization Schemes

- Modified Newton Method with Exact Line Search
- Modified Newton Method with Inexact Line Search

Globalization Schemes

- Modified Newton Method with Exact Line Search
- Modified Newton Method with Inexact Line Search
- Quasi-Newton Method

Globalization Schemes

- Modified Newton Method with Exact Line Search
- Modified Newton Method with Inexact Line Search
- Quasi-Newton Method
  - BFGS
  - Oren-Luenberger

Modified Newton Method with Exact Line Search

- Choose a starting point $x_0 \in \mathbb{R}^n$.
  Choose numbers $\gamma_k \leq 1, \sigma_k, k = 0, 1, .., $ with $\inf_k \gamma_k > 0, \inf_k \sigma_k > 0.$

Modified Newton Method with Exact Line Search

- Choose a starting point $x_0 \in \mathbb{R}^n$.
  Choose numbers $\gamma_k \leq 1, \sigma_k, k = 0, 1, .., $ with $\inf_k \gamma_k > 0, \inf_k \sigma_k > 0$.
- For all $k = 0, 1, ...,$ update

$$x_{k+1} := x_k - \lambda_k s_k$$

where $s_k \in D(\gamma_k, x_k)$, and $\lambda_k \in [0, \sigma_k \|Dh(x_k)\|]$ is such that

$$h(x_{k+1}) = \min_{\mu}\{h(x_k - \mu s_k) : 0 \leq \mu \leq \sigma_k \|Dh(x_k)\|\}.$$

Global Convergence Results for Modified Newton Method with Exact Line Search

- $K := \{x | h(x) \leq h(x_0)\}$ is compact, and
- h is continuously differentiable in some open set containing K.

Global Convergence Results for Modified Newton Method with Exact Line Search

- $K := \{x | h(x) \leq h(x_0)\}$ is compact, and
- h is continuously differentiable in some open set containing K.

Then for any sequence $\{x_k\}$ defined by this method:

Global Convergence Results for Modified Newton Method with Exact Line Search

- $K := \{x|h(x) \leq h(x_0)\}$ is compact, and
- h is continuously differentiable in some open set containing K.

Then for any sequence $\{x_k\}$ defined by this method:

- $x_k \in K$ for all $k = 0, 1, ...$ The sequence $\{x_k\}$ has at least one accumulation point $\overline{x}$ in K.
- Each accumulation point $\overline{x}$ of $\{x_k\}$ is a stationary point of h:

$$Dh(\overline{x}) = 0.$$

Proof for the Method with Exact Line Search: 1

From the definition of the sequence $\{x_k\}$ we have that the sequence $\{h(x_k)\}$ is monotone, i.e., $h(x_0) \geq h(x_1) \geq \cdots$. Hence, $x_k \in K$ for all k. K is compact; therefore, the sequence $\{x_k\}$ has at least one accumulation point $\overline{x} \in K$.

Proof for the Method with Exact Line Search: 2-(1)

Assume that $\overline{x}$ is an accumulation point of $\{x_k\}$ but is not a stationary point of h:

$$Dh(\overline{x}) \neq 0. \tag{3}$$

WLOG, let $\lim_{k\to\infty} x_k = \overline{x}$.
According to the important lemma, there is a neighborhood $U(\overline{x})$ and a number $\lambda \geq 0$ satisfying

$$h(x - \mu s) \leq h(x) - \mu \frac{\gamma}{4} \|Dh(\overline{x})\| \tag{4}$$

for all $x \in U(\overline{x}), s \in D(\gamma, x)$, and $0 \leq \mu \leq \lambda$.
Since $\lim_{k\to\infty} x_k = \overline{x}$, the continuity of $Dh(x)$ together with (3) implies there is k_0 such that for all $k \geq k_0, x_k \in U(\overline{x})$ and $\|Dh(x_k)\| \geq \frac{1}{2}\|Dh(\overline{x})\|$.

Proof for the Method with Exact Line Search: 2-(2)

Let $\Lambda := \min\{\lambda, \frac{1}{2}\sigma\|Dh(\overline{x})\|, \epsilon := \Lambda\frac{\gamma}{4}\|Dh(\overline{x})\|\}$.
Since $\Lambda \leq \lambda, x_k \in U(\overline{x}), s_k \in D(\gamma_k, x_k)$, (4) implies that
$h(x_{k+1}) \leq h(x_k) - \Lambda\frac{\gamma}{4}\|Dh(\overline{x})\| = h(x_k) - \epsilon$ for all $k \geq k_0$.
This means that $\lim_{k\to\infty} h(x_k) = -\infty$, which contradicts the fact that $h(x_k) \geq h(\overline{x})$ for all k. Hence, $\overline{x}$ is a stationary point of h.

Modified Newton Method with Inexact Line Search

- Choose a starting point $x_0 \in \mathbb{R}^n$.
  Choose numbers $\gamma_k \leq 1, \sigma_k, k = 0, 1, .., $ with $\inf_k \gamma_k > 0, \inf_k \sigma_k > 0$.
- For all $k = 0, 1, ...,$ obtain x_{k+1} from x_k as follows:
  - Select $s_k \in D(\gamma_k, x_k)$.
    Define $\rho_k := \sigma_k \|Dh(x_k)\|$, $h_k(\mu) := h(x_k - \mu s_k)$.
    Then, determine the smallest integer $j \geq 0$ such that
    $h_k(\rho_k 2^{-j}) \leq h_k(0) - \rho_k 2^{-j} \frac{\gamma_k}{4} \|Dh(x_k)\|$.
  - Determine $\bar{i} \in \{0, 1, ..., j\}$ such that $h_k(\rho_k 2^{-\bar{i}})$ is minimum and let
    $\lambda_k := \rho_k 2^{-\bar{i}}$.
    Then, update $x_{k+1} := x_k - \lambda_k s_k$.

[Note that $h(x_{k+1}) = \min_{1 \leq i \leq j} h_k(\rho_k 2^{-i})$.]

Global Convergence Results for Modified Newton Method with Inexact Line Search

- $K := \{x|h(x) \leq h(x_0)\}$ is compact, and
- h is continuously differentiable in some open set containing K.

Then for any sequence $\{x_k\}$ defined by this method:

- $x_k \in K$ for all $k = 0, 1, \ldots$. The sequence $\{x_k\}$ has at least one accumulation point $\overline{x}$ in K.
- Each accumulation point $\overline{x}$ of $\{x_k\}$ is a stationary point of h:

$$Dh(\overline{x}) = 0.$$

Proof for the Method with Inexact Line Search: 1

From the definition of the sequence $\{x_k\}$ we have that the sequence $\{h(x_k)\}$ is monotone, i.e., $h(x_0) \geq h(x_1) \geq \cdots$. Hence, $x_k \in K$ for all k. K is compact; therefore, the sequence $\{x_k\}$ has at least one accumulation point $\overline{x} \in K$.

Proof for the Method with Inexact Line Search: 2

Again, we will prove the second result by a contradiction, which is similar to the previous proof in the section of exact line seach. Assume that $\overline{x}$ is an accumulation point of a sequence $\{x_k\}$ but not a stationary point of h, i.e.,

$$Dh(\overline{x}) \neq 0.$$

By the important lemma and the hypotheses of the global convergence results, we can show that there is an $\epsilon \geq 0$ for which

$$h(x_{k+1}) \leq h(x_k) - \epsilon$$

for all $k > k_0$. This contradicts the fact that $h(x_k) \geq h(\overline{x})$ for all k. Therefore, $\overline{x}$ is a stationary point of h.

Quasi-Newton Methods

- Choose a starting point $x_0 \in \mathbb{R}^n$ and an $n \times n$ positive definite matrix H_0. Set $g_0 := g(x_0)$.
- For $k = 0, 1, \dots$ obtain x_{k+1}, H_{k+1} from x_k, H_k as follows:
  - if $g_k = 0$, stop: x_k is a stationary point for h. Otherwise
  - compute $s_k := H_k g_k (\approx H(x_k)^{-1} g_k)$.
  - Update $x_{k+1} = x_k - \lambda_k s_k$ by means of a minimization

$$h(x_{k+1}) \approx \min\{h(x_k - \lambda s_k) | \lambda \geq 0\},$$

$$g_{k+1} := g(x_{k+1}), p_k := x_{k+1} - x_k, q_k := g_{k+1} - g_k.$$

  - Choose suitable parameters $\gamma_k > 0, \theta_k \geq 0$, and compute $H_{k+1} = \psi(\theta_k, \gamma_k H_k, p_k, q_k)$ where

$$\psi(\theta, H, p, q) := H + (1 + \theta \frac{q^T H q}{p^T q}) \frac{p p^T}{p^T q}$$

$$- \frac{(1-\theta)}{q^T H q} H q \cdot q^T H - \frac{\theta}{p^T q} (p q^T H + H q p^T).$$

Global Convergence Results for Quasi-Newton Method (BFGS)

- $H(\overline{x})$ is positive definite.

[Note that $H(x) := \left(\frac{\partial^2 h(x)}{\partial x^i \partial x^k} \right)_{i,k=1,\ldots,n}$.]

Global Convergence Results for Quasi-Newton Method (BFGS)

- $H(\overline{x})$ is positive definite.
- $H(x)$ is Lipschitz continuous at $x = \overline{x}$.

[Note that $H(x) := \left(\frac{\partial^2 h(x)}{\partial x^i \partial x^k} \right)_{i,k=1,\ldots,n}$.]

Global Convergence Results for Quasi-Newton Method (BFGS)

- $H(\overline{x})$ is positive definite.
- $H(x)$ is Lipschitz continuous at $x = \overline{x}$.
- Given constants $0 < c_1 < c_2 < 1, c_1 \leq \frac{1}{2}$, $x_{k+1} = x_k - \lambda_k s_k$ is chosen so that

$$h(x_{k+1}) \leq h(x_k) - c_1 \lambda_k g_k^T s_k,$$

$$g_{k+1}^T s_k \leq c_2 g_k^T s_k.$$

[Note that $H(x) := \left(\frac{\partial^2 h(x)}{\partial x^i \partial x^k} \right)_{i,k=1,\ldots,n}$.]

Global Convergence Results for Quasi-Newton Method (BFGS)

- $H(\overline{x})$ is positive definite.
- $H(x)$ is Lipschitz continuous at $x = \overline{x}$.
- Given constants $0 < c_1 < c_2 < 1, c_1 \leq \frac{1}{2}$, $x_{k+1} = x_k - \lambda_k s_k$ is chosen so that

$$h(x_{k+1}) \leq h(x_k) - c_1 \lambda_k g_k^T s_k,$$

$$g_{k+1}^T s_k \leq c_2 g_k^T s_k.$$

Powell (1975) was able to show that $\exists V(\overline{x}) \subseteq U(\overline{x})$ such that the BFGS method is superlinearly convergent for all positive definite matrices H_0 and all $x_0 \in V(\overline{x})$. [Note that $H(x) := \left(\frac{\partial^2 h(x)}{\partial x^i \partial x^k} \right)_{i,k=1,\ldots,n}$.]

Global Convergence Results for Quasi-Newton Method (A Subclass of Oren-Luenberger)

- $H(\overline{x})$ is positive definite.
- $H(x)$ is Lipschitz continuous at $x = \overline{x}$.

Global Convergence Results for Quasi-Newton Method (A Subclass of Oren-Luenberger)

- $H(\overline{x})$ is positive definite.
- $H(x)$ is Lipschitz continuous at $x = \overline{x}$.
- $x_{k+1} = x_k - \lambda_k s_k$ is chosen so that

$$\lambda_k = \min\{\lambda \geq 0 | g(x_k - \lambda s_k)^T s_k = \mu_k g_k^T s_k\}, |\mu_k| < 1.$$

Global Convergence Results for Quasi-Newton Method (A Subclass of Oren-Luenberger)

- $H(\overline{x})$ is positive definite.
- $H(x)$ is Lipschitz continuous at $x = \overline{x}$.
- $x_{k+1} = x_k - \lambda_k s_k$ is chosen so that

$$\lambda_k = \min\{\lambda \geq 0 | g(x_k - \lambda s_k)^T s_k = \mu_k g_k^T s_k\}, |\mu_k| < 1.$$

- The line search is asymptotically exact, i.e., for large enough k

$$|\mu_k| \leq c\|g_k\|.$$

Global Convergence Results for Quasi-Newton Method (A Subclass of Oren-Luenberger)

- $H(\overline{x})$ is positive definite.
- $H(x)$ is Lipschitz continuous at $x = \overline{x}$.
- $x_{k+1} = x_k - \lambda_k s_k$ is chosen so that

$$\lambda_k = \min\{\lambda \geq 0 | g(x_k - \lambda s_k)^T s_k = \mu_k g_k^T s_k\}, |\mu_k| < 1.$$

- The line search is asymptotically exact, i.e., for large enough k

$$|\mu_k| \leq c\|g_k\|.$$

It can be shown Stoer (1977) that for all $k \geq 0$

$$\lim_k x_k = \overline{x}$$

$$\|x_{k+n} - \overline{x}\| \leq \gamma\|x_k - \overline{x}\|^2$$

for all positive definite initial matrices H_0 and for $\|x_0 - \overline{x}\|$ small enough

Conclusions for Local- and Global- Properties of Newton Methods

- If a starting point x_0 is chosen sufficiently close to the optimum point $\overline{x}$, under certain assumptions the sequence $\{x_n\}$ generated by Newton's method is at least locally convergent.

Conclusions for Local- and Global- Properties of Newton Methods

- If a starting point x_0 is chosen sufficiently close to the optimum point $\overline{x}$, under certain assumptions the sequence $\{x_n\}$ generated by Newton's method is at least locally convergent.
- Under certain conditions, the global convergence of Newton's method can be obtained by finding parameters λ_k and search directions s_k for the following update, $x_{k+1} = x_k - \lambda_k s_k$.

Reference

- Stoer, Josef, and Roland Bulirsch. Introduction to numerical analysis. Vol. 12. Springer Science & Business Media, 2013.
- Holistic Numerical Methods, accessed 15 November 2019, http://numericalmethods.eng.usf.edu.
- CSE 245, University of Califoria, San Diego, accessed 21 November 2019, http://cseweb.ucsd.edu/classes/wi15/cse245-a/.

Thank you very much for your attention.

Do you have any question?