Lecture 14

Global Newton Methods

October 22, 2008

Outline

- Global Newton Methods
  - Path Search Algorithm

Local vs Global

- Locally convergent algorithms require initial iterate "sufficiently" close to "unknown" zero of the function under consideration

- When this requirement is not met, the algorithm can fail to converge to a "zero" (well-known for nondifferentiable equations)

- Globally convergent algorithms allow arbitrary starting point

- We embark in a study of such methods for nondifferentiable equations

$$G(x) = 0 \qquad \text{for } x \in X$$

Merit Functions

- Merit Functions play a key role in designing globally convergent algorithms for constrained equation systems

$$G(x) = 0 \qquad \text{for } x \in X$$

- Using them, we can recast the problem as a constrained minimization problem

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & x \in X \end{array}$$

- The merit function $\theta : X \to \mathbb{R}_+$ is such that

$$\theta(x) = 0 \quad \text{if and only if} \quad G(x) = 0$$

- Some natural merit functions (norm-based)

$$\theta(x) = \|G(x)\|^2$$

$$\theta(x) = \|G(x)\|_p \qquad \text{with } p \geq 1 \quad \text{or} \quad p = \infty$$

$$\theta(x) = \|G(x)\|_D$$

where $\|\cdot\|_D$ is the norm induced by a symmetric positive definite matrix D:

$$\|y\|_D = y^T D y$$

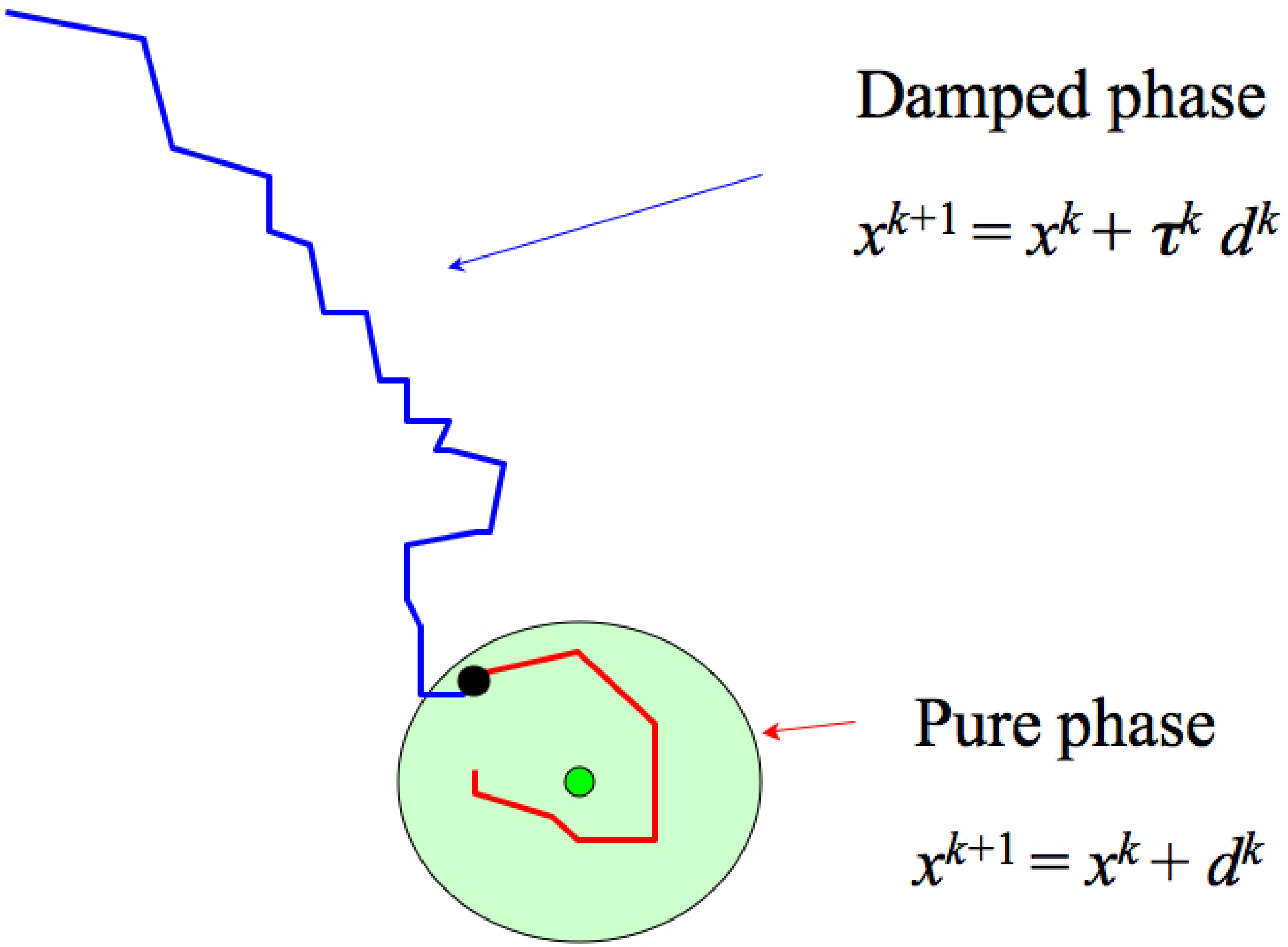
- For VIs and CPs, there are additional suitable merit functions

Path Search Algorithm for System of Equations

- In the differentiable case, a common approach to globalize the convergence of local algorithms is via line search on a certain merit function
- In the nondifferentiable case, this is not the most natural approach
- To motivate the path search algorithm, we revisit the classical **global** Newton method for continuously differentiable system of equations

$$G(x) = 0, \qquad G : \mathbb{R}^n \to \mathbb{R}^n, \qquad G \in C^1$$

- The classical global Newton method has two phases:
  - Damped phase: from start until "good neighborhood" is entered
  - Pure phase: local within the neighborhood
- The classic global method is also known as **damped Newton method**


Damped phase
$x^{k+1} = x^k + \tau^k d^k$
Pure phase
$x^{k+1} = x^k + d^k$

Damped Newton Method: Differentiable Case

- Let x^k be a current iterate

- When x^k is not in "good neighborhood" (close to a solution), the direction d^k is computed by solving

$$G(x^k) + JG(x^k)d = 0,$$

and new iterate is defined by **searching the segment** $[x^k, x^k + d^k]$

  - If the point $x^k + d^k$ has "sufficient descent", we set

$$x^{k+1} = x^k + d^k$$

- Otherwise, we backtrack from $x^k + d^k$ by searching the line segment (Newton's path) $p^k(\tau)$, given by

$$p^k(\tau) = x^k + \tau d^k \qquad \tau \in [0, 1]$$

- Since $G \in C^1$ and $d^k = -JG(x^k)^{-1}G(x^k)$, we have

$$G(p^k(\tau)) = (1 - \tau)\, G(x^k) + o(\tau) \qquad \text{for } \tau \text{ small}$$

- Thus, we are guaranteed to find τ^k such that $\|G(p^k(\tau^k))\|$ is sufficiently smaller than $\|G(x^k)\|$
- The next iterate is $x^{k+1} = p^k(\tau^k)$.

▶ The determination of τ_k (stepsize) is the **damping process**

Nondifferentiable Equations

- Due to nondifferentiability
  (1) There is no longer a single natural model for Newton direction selection
  (2) It is not clear that a point $p^k(\tau)$ with "sufficient" decrease exists on the path $p^k(\tau) = x^k + \tau d^k$ for $\tau \in [0, 1]$
- We know that issue (1) is addressed by the concept of a family of "nonsingular Newton approximations"
- An extension of the "nonsingular Newton approximation" concept exists that can handle the issue (2):

uniform approximations

Nonsingular Uniform Newton Approximation

Definition Let $\Omega \subset \mathbb{R}^n$ be an open set and $\bar{x} \in \Omega$. Let $G : \Omega \to \mathbb{R}^m$ be locally Lipschitz. We say that G has a **nonsingular uniform Newton approximation at $\bar{x}$**, if

1. there exists a Newton approximation family $\mathcal{A}$ on Ω

2. there exist scalars ϵ and L' such that for all $x \in \Omega$, the approximation $\mathcal{A}(x)$ is uniformly locally Lipschitz homeomorphism with ϵ and modulus L' on Ω.

- This definition strengthens conditions defining Newton approximation and nonsingular Newton approximation
- Loosely speaking, local properties of $A(x, \cdot)$ over $U \subseteq \Omega$ are forced to be "uniform" over Ω

- Formally, **condition (1)** requires that the family $\mathcal{A}(x)$ of approximations is such that

  (a) $A(x, 0) = 0$ for every $A(x, \cdot) \in \mathcal{A}(x)$ and all $x \in \Omega$

  (b) For some function $\Delta : \mathbb{R}_{++} \to \mathbb{R}_{+}$ with $\Delta(t) \to 0$ as $t \downarrow 0$, we have
  $$\frac{\|G(x) + A(x, x' - x) - G(x')\|}{\|x - x'\|} \leq \Delta(\|x - x'\|)$$
  for all $A(x, \cdot) \in \mathcal{A}(x)$ and all $x, x' \in \Omega$ with $x' \neq x$

- Condition (2) means that the scalars $\epsilon_{\mathcal{A}}$ and $L'_{\mathcal{A}}$ in the definition of the family $\mathcal{A}$ of locally Lipschitz homeomorphisms are required to be independent of $\mathcal{A}$

- Formally, **condition (2)** states that for each $x \in \Omega$ and every $A(x, \cdot) \in \mathcal{A}(x)$,
  (a) There are two open sets U_x and V_x such that $B(0, \epsilon) \subseteq U_x$ and $B(0, \epsilon) \subseteq V_x$
  (b) $A(x, \cdot)$ is Lipschitz homeomorphism mapping of U_x onto V_x with L' being the Lipschitz constant for the inverse of $A(x, \cdot) \mid_{U_x}$ (the map restricted to U_x)

We refer to $\mathcal{A}$ as a **nonsingular uniform Newton approximation** of G at $\bar{x}$.

- We say that G has a nonsingular uniform **strong** Newton approximation when condition (1) is strengthened to

(1') there exists a strong Newton approximation family $\mathcal{A}$ on Ω

- This corresponds to replacing condition (b) with the following:

(b') For scalar $\tilde{L}$ and some function $\Delta : \mathbb{R}_{++} \to \mathbb{R}_{+}$ with $\Delta(t) \to 0$ as $t \downarrow 0$, we have

$$\frac{\|G(x) + A(x, x' - x) - G(x')\|}{\|x - x'\|^2} \leq \tilde{L}$$

for all $A(x, \cdot) \in \mathcal{A}(x)$ and all $x, x' \in \Omega$ with $x' \neq x$

Global Homeomorphism

- The existence of nonsingular uniform Newton approximation at every $\bar{x} \in G$ implies that G is a global homeomorphism from $\mathbb{R}^n$ to itself
- This result is an extension of a classical theorem (Hadamard)
- We use this result in establishing global convergence of the Newton's method

Theorem 8.1.4 Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be continuous, and let G have a nonsingular uniform Newton approximation on $\mathbb{R}^n$. Then,

$$G \textbf{ is bijective on } \mathbb{R}^n$$

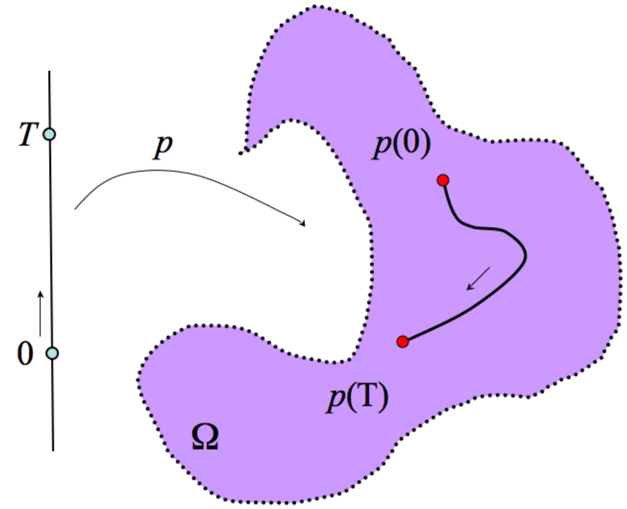
and, hence, G is a homeomorphism from $\mathbb{R}^n$ onto itself.

Paths

For an open set $\Omega \subseteq \mathbb{R}^n$, a *path* in Ω is a continuous scalar function

$$p : [0, \bar{\tau}] \to \Omega$$

for some $\bar{\tau} > 0$

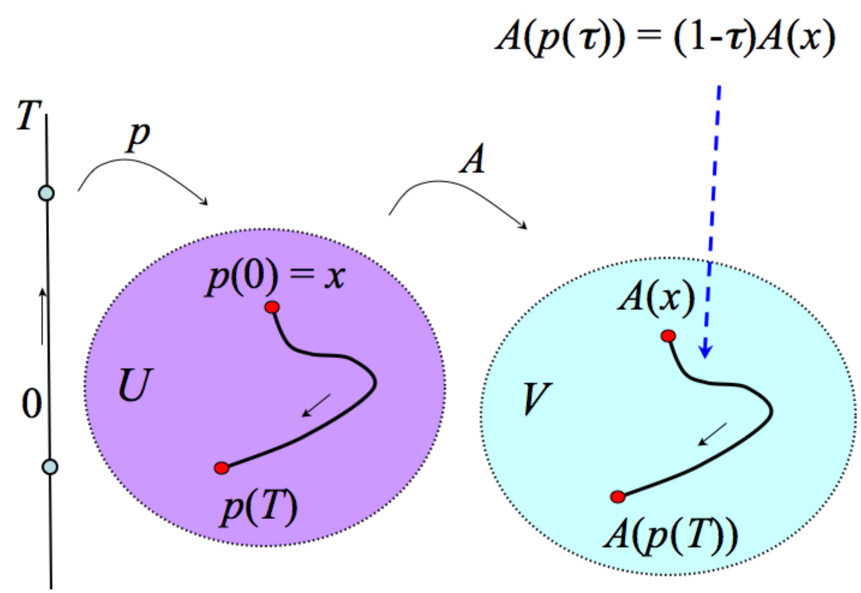


- In the global Newton algorithm we use special paths
- These are identified in the following proposition

Existence of a Special Path

Proposition 8.1.6 Let $U, V \subseteq \mathbb{R}^n$ be open sets, and let $A : U \to V$ be a Lipschitz homeomorphism. Let $\bar{x} \in U$ be such that $A(\bar{x}) \neq 0$ and $A(\bar{x}) + B(0, \epsilon) \subset V$ for some $\epsilon > 0$. Then, for $\bar{\tau} = \min\{1, \epsilon/\|A(\bar{x})\|\}$, there is a unique path $p : [0, \bar{\tau}] \to U$ such that

$$p(0) = \bar{x}, \qquad A(p(\tau)) = (1 - \tau)\, A(\bar{x}) \quad \text{for all } \tau \in [0, \bar{\tau}]$$



- In particular, the path p is given by

$$p(\tau) = A^{-1}((1-\tau)A(x)) \qquad \text{for all } \tau \in [0, \bar{\tau}].$$

- Furthermore p is Lipschitz continuous on $[0, \bar{\tau}]$ and A is local homeomorphism near $p(\bar{\tau})$
- Note that the specific choice of $\bar{\tau}$ corresponds to (silent requirement) the path $A(p(\cdot))$ being contained within the set $A(\bar{x}) + B(0, \epsilon)$, i.e.,

$$\|A(p(\tau)) - A(\bar{x})\| \le \epsilon \qquad \text{for all } \tau \in [0, \bar{\tau}]$$

- We use this proposition later on for the homeomorphism $G(x^k) + A(x^k, \cdot - x^k)$ with $A(x^k, \cdot) \in \mathcal{A}(x^k)$, where $\mathcal{A}(x^k)$ is a nonsingular uniform Newton approximation at an iterate x^k of interest
  - In this case, $\bar{x} = x^k$ and $\bar{\tau} = \min\{1, \epsilon/\|G(x^k)\|\}$
  - The scalar ϵ is related to the common ball radius playing a role in homeomorphic property of mappings $A(x, \cdot - x)$ in part (2) of the definition a nonsingular uniform Newton approximation

Path Newton Method

Step 0 Select initial vector x^0 and $\gamma \in (0,1)$. Set $k = 0$.

Step 1 If $G(x^k) = 0$, then stop.

Step 2 Select an approximation $A(x^k, \cdot) \in \mathcal{A}(x^k)$ and **the corresponding path** p^k with the domain defined by the scalar $\bar{\tau}_k = \min\{1, \epsilon/\|G(x^k)\|\}$ [cf. comments after Prop. 8.1.6].

Step 2(i) Set $i = 0$.

Step 2(ii) If

$$\|G(p^k(\bar{\tau}_k/2^i))\| \le \left(1 - \frac{\gamma\,\bar{\tau}_k}{2^i}\right)\|G(x^k)\|,$$

set $t_k = \frac{\bar{\tau}_k}{2^i}$, $i_k = i$, and go to Step 3. Otherwise, set $i =: i + 1$ and go to Step 2(ii).

Step 3 Set $x^{k+1} = p^k(t_k)$ and $k =: k + 1$, and go to Step 1.

- The steps 2(i) and 2(ii) are referred to as "backtracking" procedure
- The stepsize selection in 2(i)–(ii) is known as **Armijo stepsize rule**
- Instead of $\frac{1}{2}$ as the "backtracking" factor, any $\rho \in (0,1)$ can be used
  - In this case the test in Step 2(ii) is replaced with

$$\|G(p^k(\rho^i \bar{\tau}_k))\| \leq \left(1 - \gamma\, \rho^i\, \bar{\tau}_k\right) \|G(x^k)\|,$$

- The only difference from the classical method is the computation of d^k including a special path selection
- In view of Prop. 8.1.6, the path p^k exists with well defined $\bar{\tau}_k$
- A question remains on correctness of the algorithm

Does the loop at Step 2(ii) terminate in a finite number of steps

Correctness

Lemma 8.1.7. Let $\Omega \subseteq \mathbb{R}^n$ be an open set containing the vector $\bar{x}$. Let $G : \Omega \to \mathbb{R}^n$ be Lipschitz on Ω. Assume that $G(\bar{x}) \neq 0$ and G has a nonsingular Newton approximation at $\bar{x}$. Then, for each $A(x, \cdot) \in \mathcal{A}(\bar{x})$, there exists a unique Lipschitz continuous path $p : [0, \bar{\tau}] \to \Omega$ with largest $\bar{\tau} \leq 1$ such that

(a) $p(0) = \bar{x}$

(b) For every $\tau \in [0, \bar{\tau}]$, there holds

$$G(\bar{x}) + A(\bar{x}, p(\tau) - \bar{x}) = (1 - \tau)G(\bar{x})$$

(c) $G(\bar{x}) + A(\bar{x}, \cdot - \bar{x})$ is local homeomorphism near any point on the path

(d) For every $\gamma \in (0, 1)$, there is $\tau' \in (0, \bar{\tau})$ such that

$$\|G(p(\tau))\| \leq (1 - \gamma\tau)\,\|G(\bar{x})\| \qquad \text{for all } \tau \in [0, \tau']$$

Proof. (a)–(c) The results follow from Proposition 8.1.6, as applied to mapping $G(\bar{x}) + A(\bar{x}, \cdot - \bar{x})$.

(d) Suppose such τ' does not exist. Then, there exists a sequence $\{\tau_k\}$ converging to 0 and such that

$$\|G(p(\tau_k))\| > (1 - \gamma\tau_k)\,\|G(\bar{x})\| \qquad \text{for all } k$$

Since G has a Newton approximation at $\bar{x}$ and $p(0) = \bar{x}$, we have for all k,

$$\|G(p(\tau_k)) - G(\bar{x}) - A(\bar{x}, p(\tau_k) - \bar{x})\| \leq \|p(\tau_k) - \bar{x}\|\ \Delta(\|p(\tau_k) - \bar{x}\|)$$

By the path property in part (b), we obtain

$$\|G(p(\tau_k)) - (1 - \tau_k)G(\bar{x})\| \leq \|p(\tau_k) - \bar{x}\|\ \Delta(\|p(\tau_k) - \bar{x}\|)$$

The path p is Lipschitz continuous, so that for a scalar $L_p > 0$ and all k, we have

$$\|G(p(\tau_k)) - (1-\tau_k)G(\bar{x})\| \le \tau_k \, L_p \, \Delta(\|p(\tau_k) - \bar{x}\|)$$

Hence,

$$\lim_{k\to\infty} \frac{\|G(p(\tau_k)) - (1-\tau_k)G(\bar{x})\|}{\tau_k} = 0,$$

implying that

$$\|G(p(\tau_k))\| \le (1-\tau_k) \, \|G(\bar{x})\| + o(\tau_k)$$

By using the relation $\|G(p(\tau_k))\| > (1-\gamma\tau_k) \, \|G(\bar{x})\|$, we obtain

$$(1-\gamma) \, \tau_k \, \|G(\bar{x})\| \le o(\tau_k)$$

Because $\gamma \in (0,1)$, it follows that

$$\tau_k \, \|G(\bar{x})\| \le o(\tau_k) \qquad \text{for all } k.$$

Therefore

$$\|G(\bar{x})\| \leq \lim_{k\to\infty} \frac{o(\tau_k)}{\tau_k} = 0$$

- a contradiction, since $G(\bar{x}) \neq 0$.

Convergence of the Path Newton Method

Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz on $\mathbb{R}^n$, and let G have a **nonsingular uniform Newton approximation** $\mathcal{A}$ on $\mathbb{R}^n$. Then, the following statements are true:

(1) The map G has a unique zero.

(2) For any x^0, the sequence $\{x^k\}$ generated by the Path Newton method converges Q-superlinearly to the zero of G.

(3) If $\mathcal{A}$ is **nonsingular uniform strong Newton approximation** on $\mathbb{R}^n$, then $\{x^k\}$ converges **Q-quadratically**.

Proof

(1) By Theorem 8.1.4, G is a (global) homeomorphism from $\mathbb{R}^n$ onto itself. Thus, G has a unique zero.

(2) Note that in view of Steps 2(ii) and 3, we have

$$\|G(x^{k+1})\| \le (1 - \gamma\, t_k)\ \|G(x^k)\|$$

Thus, the sequence $\{\|G(x^k)\|\}$ is nonincreasing and therefore, it has a limit value. If we had proven that $t_k \ge \xi$ for all k, the preceding relation would imply that for all k,

$$\|G(x^{k+1})\| \le (1 - \gamma\, \xi)\, \|G(x^k)\| \le \cdots \le (1 - \gamma\, \xi)^{k+1} \|G(x^0)\|.$$

Hence, we would immediately have

$$\lim_{k\to\infty} \|G(x^k)\| = 0.$$

The rest of the proof is about showing that $t_k \ge \xi$ for some ξ and for all k. To arrive at a contradiction, suppose it is not bounded away from zero. Then, the **integers** i_k **are unbounded** and by the definition of i_k at Step 2(ii), we have

$$\|G(p^k(\bar{\tau}_k/2^{i_k-1}))\| > \left(1 - \frac{\gamma\,\bar{\tau}_k}{2^{i_k-1}}\right)\|G(x^k)\| \qquad \text{for all } k. \tag{1}$$

On the other hand, by the properties of $\Delta(t)$, we can choose $t^* > 0$ small enough so that

$$\Delta(t) \le \frac{1-\gamma}{L'} \qquad \text{for all } t \in (0, t^*]. \tag{2}$$

By the definition of the path p^k, we have for all $\tau \leq \bar{\tau}_k$ [cf. Lemma 8.1.7(b) with $\bar{x} = x^k$]

$$G(x^k) + A(x^k, p^k(\tau) - x^k) = (1 - \tau)G(x^k), \tag{3}$$

implying that

$$A(x^k, p^k(\tau) - x^k) = -\tau G(x^k).$$

By the (uniform) Lipschitz continuity of the inverse of $A(x^k, \cdot)$ and the fact $A(x^k, p^k(0) - x^k) = 0$, it follows that for all $\tau \leq \bar{\tau}_k$,

$$\|p^k(\tau) - x^k\| \leq \tau\, L'\, \|G(x^k)\| \leq \tau\, L'\, \|G(x^0)\|, \tag{4}$$

where the last inequality follows from nonincreasing property of $\|G(x^k)\|$. Choosing $\tau \leq \frac{t^*}{L'\,\|G(x^0)\|}$, we have $\|p^k(\tau) - x^k\| \leq t^*$, implying by Eq. (2)

that

$$\Delta(\|p^k(\tau) - x^k\|) \le \frac{1-\gamma}{L'}.$$

By using this relation and the first inequality of Eq. (4), we obtain for $\tau \le \min\left\{\bar{\tau}_k, \frac{t^*}{L'\,\|G(x^0)\|}\right\}$,

$$\|p^k(\tau) - x^k\|\;\Delta(\|p^k(\tau) - x^k\|) \le \tau\, L'\,\|G(x^k)\|\,\frac{1-\gamma}{L'} = \tau\,(1-\gamma)\,\|G(x^k)\|. \tag{5}$$

By the Taylor property of $\mathcal{A}$ [cf. (b) of the approximation definition)], we have

$$\begin{aligned} \|G(p^k(\tau))\| &\le \|A(x^k, p^k(\tau) - x^k) + G(x^k)\| \\ &\quad + \|p^k(\tau) - x^k\|\;\Delta(\|p^k(\tau) - x^k\|) \\ &\le (1-\tau)\,\|G(x^k)\| + \tau\,(1-\gamma)\,\|G(x^k)\|, \end{aligned}$$

where the last inequality follows by the path property (3) and relation (5). Hence, for $\tau \leq \min\left\{\bar{\tau}_k, \frac{t^*}{L' \|G(x^0)\|}\right\}$,

$$\|G(p^k(\tau))\| \leq (1 - \gamma\,\tau)\|G(x^k)\|. \tag{6}$$

Recall that the integers i_k are unbounded, so that for k large enough, we have $\bar{\tau}_k/2^{i_k-1} \leq \min\left\{\bar{\tau}_k, \frac{t^*}{L' \|G(x^0)\|}\right\}$. Thus, the relation (6) holds with $\tau = \bar{\tau}_k/2^{i_k-1}$. This, however contradicts relation (1). Hence, the stepsizes t_k are bounded away from zero, which completes the proof.

Convergence rate of (2) and (3). Note that by Step 2, we have $\bar{\tau}_k = \min\{1, \epsilon/\|G(x^k\|\}$. Since $\|G(x^k)\| \to 0$, it follows that eventually, $\bar{\tau}_k = 1$ for all k large enough. Similar to the derivation of relation (6), we can see that this relation also holds for $\tau \leq \min\left\{1, \frac{t^*}{L' \|G(x^k)\|}\right\}$. As $\|G(x^k)\| \to 0$, it follows that the quantities defining the interval for τ

approach infinity. Thus, relation (6) will eventually hold with $\tau = 1$, and hence, $t_k = 1$ at Step 2(ii).

Therefore, after some finite number of iterations, the Path Method will reduce to the pure Newton method (local method), and the convergence rate results follow by Theorem 7.2.5.