CS730
Appendix B
Lipschitz Continuity, Rates of Convergence
& Related Results

1. Lipschitz continuity

Lipschitz continuity is a stronger type of continuity that implies uniform continuity and is extremely useful in analyzing convergence of algorithms.

1 Definition A function $g : R^n \to R^m$ is said to be Lipschitz continuous on $S \subset R^n$ if there exists a positive number K such that

$$\|g(y) - g(x)\| \leq K\|y - x\| \text{ for all } x,\ y \in S$$

where $\|\cdot\|$ denotes some norm. We say then $g \in LC_K(S)$.

If $f : R^n \to R$ and $\nabla f \in LC_K(S)$ we write $f \in LC^1_K(S)$.

Note that for $f : R^n \to R$:

$$\begin{array}{cccccc} f \in LC_K(S) & \Longrightarrow & \begin{array}{l} f \text{ is uniformly} \\ \text{continuous on } S \end{array} & \Longrightarrow & \begin{array}{c} f \text{ continuous} \\ \text{on } S \end{array} \\ & & \begin{array}{l} f \text{ is uniformly} \\ \text{continuous on } S \end{array} & \begin{array}{c} \Longleftarrow \\ (S \text{ compact}) \end{array} & \begin{array}{c} f \text{ continuous} \\ \text{on } S \end{array} \end{array}$$

Proof. (First implication)
$\forall \varepsilon > 0, \quad \exists \delta = \dfrac{\varepsilon}{K} : 0 < \|x - y\| < \delta = \dfrac{\varepsilon}{K} \Longrightarrow |f(y) - f(x)| \leqq K\|y - x\| < \varepsilon.$ □

The following provides a useful bound on the integral of a continuous function:

2 Lemma Let $g : [a, b] \subset R \to R^n$ be continuous on $[a, b]$. Then

$$\Big\| \int_a^b g(t)dt \Big\| \leqq \int_a^b \Big\|g(t)\Big\|dt$$

Proof. Since any norm is continuous both integrals exist and therefore for any $\varepsilon > 0$ there is a partition $a < t_0 <, \ldots, < t_p < b$ of $[a, b]$ such that

$$\Big\| \int_a^b g(t)dt - \sum_{i=1}^{p} g(t_i)(t_i - t_{i-1}) \Big\| < \varepsilon \tag{1}$$

and

$$\Big| \int_a^b \Big\|g(t)\Big\|dt - \sum_{i=1}^{p} \Big\|g(t_i)\Big\|(t_i - t_{i-1}) \Big| < \varepsilon \tag{2}$$

Hence

$$\begin{aligned}
\Big\| \int_a^b g(t)dt \Big\| &< \Big\| \sum_{i=1}^{p} g(t_i)(t_i - t_{i-1}) \Big\| + \varepsilon \quad \text{(By (1) and} \quad \|x\| - \|y\| \leqq \|x-y\|) \\
&\leqq \sum_{i=1}^{p} \Big\| g(t_i) \Big\| (t_i - t_{i-1}) + \varepsilon \\
&< \int_a^b \Big\| g(t) \Big\| dt + 2\varepsilon \quad \text{(By (2) and} \quad \|x\| - \|y\| \leqq \|x-y\|)
\end{aligned}$$

Since ε was arbitrary, the desired inequality follows by letting $\varepsilon \to 0$. □

The following fundamental Lemma gives a bound on the error in estimating a function by its linearization.

3 Lemma (QBL: Quadratic Bound Lemma) Let $f : S \subset R^n \to R^m$ then:

$$\left\rangle \begin{array}{c} \|\nabla f(y) - \nabla f(x)\| \leqq K\|y-x\| \\ \forall x,\ y \in S \end{array} \right\rangle \Longrightarrow \left\langle \begin{array}{c} \|f(y) - f(x) - \nabla f(x)(y-x)\| \leqq \dfrac{K}{2}\|y-x\|^2 \\ \forall x,\ y \in S \end{array} \right.$$

where $||\cdot||$ denotes the 2-norm.

Proof. Define $g(t) = f\Big(x + t(y-x)\Big),\ x,\ y \in S,$ and $t \in [0,1]$. Then

$$g'(t) := \frac{dg(t)}{dt} = \nabla f\Big(x + t(y-x)\Big)(y-x)$$

$$f(y) - f(x) = \int_0^1 g'(t)dt = \int_0^1 \Big(\nabla f\Big(x + t(y-x)\Big) - \nabla f(x)\Big)(y-x)dt + \nabla f(x)(y-x)$$

Hence

$$\begin{aligned}
\|f(y) - f(x) - \nabla f(x)(y-x)\| &= \Big\| \int_0^1 \Big(\nabla f\Big((x + t(y-x)\Big) - \nabla f(x)\Big)(y-x)dt \Big\| \\
&\leqq \int_0^1 \Big\| \nabla f\Big(x + t(y-x)\Big) - \nabla f(x) \Big\| \ \Big\| y - x \Big\| dt \\
&\qquad\qquad\qquad \text{(By Lemma 2)} \\
&\leqq \int_0^1 Kt \Big\| y - x \Big\|^2 dt \qquad \text{(By } f \in LC_K^1(S)) \\
&= \frac{K}{2} \Big\| y - x \Big\|^2 .
\end{aligned}$$

□

4 Equivalence of Lipschitz continuity inequality to other inequalities

Let $f : R^n \to R$ be once or twice differentiable as needed and let $k > 0$ and $K > 0$. Then:

$$1\text{a} \Longrightarrow 2\text{a} \Longleftrightarrow 4\text{a} \Longleftrightarrow 3\text{a} \Longleftarrow 1\text{a},$$

$$(3a \wedge 3b) \Longrightarrow 1\text{a},$$

and:

$$1\text{b} \Longleftarrow 2\text{b} \Longleftrightarrow 4\text{b} \Longleftrightarrow 3\text{b} \Longrightarrow 1\text{b},$$

where

$$
\begin{array}{lrcl}
\text{1a)} & & & \|\nabla f(y) - \nabla f(x)\| \leqq K\|y-x\| \\
\text{1b)} & k\|y-x\| & \leqq & \|\nabla f(y)\| - \nabla f(x)\| \\
\text{2a)} & & & \Big(\nabla f(y) - \nabla f(x)\Big)(y-x) \leqq K\|y-x\|^2 \\
\text{2b)} & k\|y-x\|^2 & \leqq & \Big(\nabla f(y) - \nabla f(x)\Big)(y-x) \\
\text{3a)} & & & y\nabla^2 f(x) y \leqq K\|y\|^2 \\
\text{3b)} & k\|y\|^2 & \leqq & y\nabla^2 f(x) y \\
\text{4a)} & & & f(y) - f(x) - \nabla f(x)(y-x) \leqq \dfrac{k}{2}\|y-x\|^2 \\
\text{4b)} & \dfrac{k}{2}\|y-x\|^2 & \leqq & f(y) - f(x) - \nabla f(x)(y-x)
\end{array}
$$

Here x, y are arbitrary points in R^n and all norms are 2-norms.

5 Theorem Let $f : R^n \to R$ be convex and differentiable on R^n, with Lipschitz continuous gradient with constant $K > 0$. Then

$$\Big(\nabla f(y) - \nabla f(x)\Big)(y-x) \geqq \frac{1}{K}\Big\|\nabla f(y) - \nabla f(x)\Big\|^2$$

Proof. We give the proof due to Golshtein & Tretyakov for the twice continuously differentiable case (See B. T. Polyak: "Introduction to optimization: Optimization Software, NY 1987, p. 24). For the once differentiable case see: J.-B. Baillon & G. Haddad. "Quelques properiétés des opérateurs angle-bornés et n-cycliquement monotones", Israel J. of Math. 26, 1977, 137-150, Corollary 10.

$$\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f\Big(x + t(y-x)\Big)dt(y-x) = \nabla f(x) + A(y-x),$$

where A is symmetric, positive semidefiinite by convexity of f (see B.1.4). Hence

$$\Big(\nabla f(y) - \nabla f(x)\Big)(y-x) = (y-x)A(y-x) \geqq \|A\|^{-1}\|A(y-x)\| \geqq K^{-1}\|\nabla f(y) - \nabla f(y)\|^2,$$

where the first inequality follows from Lemma 6 below and the second inequlaity from B.1.4 and $\|A\| \leqq K$. ∎

6 Lemma For symmetric positive semidefinite matrix A

$$zAz \geqq \|A\|^{-1}\|Az\|^2$$

Proof. z solves: $\min\{\|A\| \cdot zAz - zA^2z \,\Big|\, zz = 1\}$

$\Rightarrow \|A\|Az - A^2z - \lambda z = 0,\ zz = 1 \Rightarrow (\|A\|A - A^2)z = \lambda z,\ zz = 1$

$\Rightarrow z(\|A\|A - A^2)z = \lambda zz = \lambda \Rightarrow \|A\|zAz - \|Az\|^2 \geqq 0,$

where the last inequality follows from $\lambda = \text{eig}\,(\|A\|A - (\|A\|^2) \geqq 0$. ∎

2. Convergence of monotone sequences

1 Theorem If the nonincreasing sequence of real numbers $\{f_i\}$ has an accumulation point $\bar{f}$ then $\{f_i\}$ converges for $\bar{f}$.

Proof. We first prove that $\{f_i\}$ is bounded below. Let $\{f_{i_j}\} \to \bar{f}$. If $\{f_i\}$ is unbounded below then there exist $\bar{i}$, $\bar{j}$ for which we have

$$\bar{f} > f_{\bar{i}}$$

$$f_{\bar{i}} > f_{i_{\bar{j}}}$$

$$f_{i_{\bar{j}}} \geqq \bar{f}$$

which is a contradiction. Hence $\{f_i\}$ is bounded below and it converges. It must converge to $\bar{f}$ because for a convergent sequence all accumulation points are identical to the limit of the sequence. ∎

3. Forcing functions

1 Def. A function $\sigma : R_+ \to R_+$ where $R_+ = \{x | x \in R,\ x \geqq 0\}$ is a forcing function if: $\sigma(0) = 0$, $\sigma(\xi) > 0$ for $\xi > 0$, and

$$\lim_{i\to\infty} \sigma(\xi_i) = 0 \Longrightarrow \lim_{i\to\infty} \xi_i = 0.$$

2 Lemma If $\sigma_1(\xi)$ and $\sigma_2(\xi)$ are forcing functions then

$$\begin{aligned}
&\text{(a)} \quad \sigma(\xi) := \min\{\sigma_1(\xi),\ \sigma_2(\xi)\}\\
&\text{(b)} \quad \sigma(\xi) := \sigma_1(\xi) \cdot \sigma_2(\xi)\\
&\text{(c)} \quad \sigma(\xi) := \sigma_1\Big(\sigma_2(\xi)\Big)\\
&\text{(d)} \quad \sigma(\xi) := \sigma_1(\xi) + \sigma_2(\xi),
\end{aligned}$$

are all forcing functions

3 Examples $\sigma(\xi) = \xi$, $\sigma(\xi) = \xi^2$, $\sigma(\xi) = \sinh \xi$ are forcing functions.

4. Convergence rates

The authoritative reference on convergence rates is J. Ortega & W. Rheinboldt: "Iterative Solution of Nonlinear Equations in Several Variables", Academic Press, New York 1970, which is, however, unnecessarily complicated on this subject! All essential convergence rate results are given in the table on the next page, Table 1, and are based on the two following well-known criteria for convergence & divergence of series of positive terms e_i:

$$s_k = \sum_{i=1}^{k} e_i,\ e_i > 0$$

Ratio or Quotient Test	Root Test
$Q = \lim_{n\to\infty} \frac{e_{n+1}}{e_n}$	$R = \lim_{n\to\infty} (e_n)^{\frac{1}{n}}$
$0 \leqq Q < 1 \Rightarrow \{s_k\}$ converges	$0 \leqq R < 1 \Rightarrow \{s_k\}$ converges
$\Rightarrow \lim_{i\to\infty} e_i = \lim_{i\to\infty} (S_i - S_{i-1}) = 0$	$\Rightarrow \lim_{i\to\infty} e_i = 0$
$Q > 1 \Rightarrow \{s_k\}$ diverges	$R > 1 \Rightarrow \{s_k\}$ diverges

Thus, the <u>size</u> of Q (or R) is a criterion for <u>convergence</u>, while the <u>rate of change of</u>

$$Q_i = \frac{e_{i+1}}{e_i}, \quad \text{or} \quad R_i = \left(\frac{e_i}{\mu}\right)^{\frac{1}{i}} \qquad \text{for some } \ \mu > 0,$$

is a measure of the <u>rate of convergence</u>.

Rate of Convergence	(a)Quotient Rates		(b)Root Rates	
Superlinear of order r	$\exists r > 1, k > 0, \hat{i}$: $\frac{e_{i+1}}{e_i^r} \leqq k, \forall i \geqq \hat{i}$	3 $\Longrightarrow$ $k^{\frac{1}{r-1}} e_{\hat{i}} < 1$	$\exists r > 1, \gamma \in (0,1), \mu > 0, \bar{i}$: $e_i \leqq , \mu\gamma^{r^i} \quad \forall i \geqq \bar{i}$	$\Longrightarrow$ $\{e_i\} \to 0$
	$\Downarrow e_{\hat{i}} < 1$		$\Downarrow$	
Superlinear	$\exists \{k_i\} \to 0$: $\frac{e_{i+1}}{e_i} \leqq k_i$	2 $\Longrightarrow$	$\forall \gamma > 0 \ \ \exists \mu > 0, \bar{i}$: $e_i \leqq \mu\gamma^i \quad \forall i \geqq \bar{i}$ (γ may be arbitrarily close to 0)	$\Longrightarrow$ $\{e_i\} \to 0$
	$\Downarrow$		$\Downarrow$	
Linear	$\exists k \in (0,1), \hat{i}$: $\frac{e_{i+1}}{e_i} \leqq k \ \ \forall i \geqq \hat{i}$ (Ratio Test of Conv.)	1 $\Longrightarrow$	$\exists \gamma(0,1), \mu > 0$: $e_i \leqq \mu\gamma^i \quad \forall i \geqq \bar{i}$ (Root Test of Conv.)	$\Longrightarrow$ $\{e_i\} \to 0$
Sublinear Assume: $\{e_i\} \to 0$	$\exists \{k_i\} \to 1, \hat{i}$: $k_i \leqq \frac{e_{i+1}}{e_i} \quad \forall i \geqq \hat{i}$	4 $\Longrightarrow$	$\forall \gamma \in (0,1), \exists \mu > 0$: $\mu\gamma^i \leqq e_i \ \ \forall i \geqq \bar{i}$ (γ may be arbitrarily close to 1)	

Table 1 Rates of Convergence for the Sequence of Positive Numbers $\{e_i\}$. Typically e_i is the error $\|x_i - \bar{x}\|$, where $\bar{x}$ is a solution.

Proof of implications 1, 2, 3 and 4 of Table 1:

1. <u>$\Rightarrow$:</u> $e_{\hat{i}+1} \leqq k e_{\hat{i}}, e_{\hat{i}+2} \leqq k^2 e_{\hat{i}}, \ldots, e_i = e_{\hat{i}+i-\hat{i}} \leqq k^{i-\hat{i}} e_{\hat{i}} = (k^{-\hat{i}} e_{\hat{i}}) k^i$

2. <u>$\Rightarrow$:</u> $e_i \leqq k_{i-1} k_{i-2}, \ldots, k_0 e_0$

   For any $\gamma > 0$ pick $\bar{i}$ s.t. $k_i < \gamma$ for $i \geqq \bar{i}$ and define $\mu = \gamma^{-\bar{i}}(k_{\bar{i}-1} k_{\bar{i}-2}, \ldots, k_0) e_0$. Then for $i \geqq \bar{i}$:

$$\begin{aligned} e_i &\leqq (k_{i-1}, \ldots, k_{\bar{i}})(k_{\bar{i}-1}, \ldots, k_0) e_0 \\ &< (\gamma^{i-\bar{i}})(\gamma^{\bar{i}} \mu) = \mu\gamma^i \end{aligned}$$

3. <u>$\Rightarrow$:</u> $e_{i+1} \leqq \left(k^{\frac{1}{1-r}}\right)^{1-r} e_i^r$

   Let $t := k^{\frac{1}{1-r}}$
   Then $e_{i+1} \leqq t^{1-r} e_i^r$ or $\frac{e_{i+1}}{t} \leqq \left(\frac{e_i}{t}\right)^r$ for $i \geqq \hat{i}$

Let $\varepsilon_i := e_i/t$

Then $\varepsilon_{i+1} \leqq \varepsilon_i^r$ for $i \geqq \bar{i} \geqq \hat{i}$

$\varepsilon_{\bar{i}+1} \leqq \varepsilon_{\bar{i}}^r,\ \varepsilon_{\bar{i}+2} \leqq \varepsilon_{\bar{i}+1}^r \leqq \varepsilon^{r^2}, \ldots, \varepsilon_i = \varepsilon_{\bar{i}+i-\bar{i}} \leqq \varepsilon_{\bar{i}}^{r^{i-\bar{i}}} = \left(\varepsilon_{\bar{i}}^{r^{-\bar{i}}}\right)^{r^i}$

So

$$e_i \leqq t\left(\left(\frac{e_{\bar{i}}}{t}\right)^{r^{-\bar{i}}}\right)^{r^i} \qquad \left(t = k^{\frac{1}{1-r}}\right)$$

For a given $\gamma > 0$, pick $\bar{i}$ large enough such that $\left(\frac{e_{\bar{i}}}{k^{\frac{1}{1-r}}}\right)^{r^{-\bar{i}}} < \gamma$ and thus

$$e_i \leqq k^{\frac{1}{1-r}} \gamma^{r^i} \qquad \text{for } i \geqq \hat{i}$$

4. $\underline{\Rightarrow:}$ $e_i \leqq k_{i-1}k_{i-2}, \ldots, k_0 e_0$

   For any $\gamma \in (0,1)$ pick $\bar{i}$ s.t. $k_i \geqq \gamma$ for $i \geqq \bar{i}$ and define $\mu = \gamma^{-\bar{i}}(k_{\bar{i}-1} \ldots, k_0)e_0$. Then

$$\begin{aligned} e_i &\geqq (k_{i-1} \ldots, k_{\bar{i}})(k_{\bar{i}-1}, \ldots, k_0)e_0 \\ &\geqq (\gamma^{i-\bar{i}})(\mu\gamma^{-\bar{i}}) = \mu\gamma^i \end{aligned}$$

5 Linear convergence theorem (LCT)

Let $F\colon R^n \to R$, let $\{x_i\} \subset R^n$, let $S = \{x | f(x) \leqq f(x_0)\}$ and consider the following assumptions:

(i) (Standard assumption): $f \in LC^1_K(S)$

(ii) (Uniform strict convexity): $\big(\nabla f(y) - \nabla f(x)\big)(y-x) \geqq k\|y-x\|^2$, $\forall x, y \in S$ for some $k > 0$, or equivalently: $f(y) - f(x) - \nabla f(x)(y-x) \geqq \dfrac{k}{2}\|y-x\|^2$, $\forall x, y \in S^1$ for some $k > 0$.

(iii) (Algorithm consequence): $f(x_i) - f(x_{i+1}) \geqq \alpha\|\nabla f(x_i)\|^2$ for some $\alpha > 0, i = 0, 1, 2, \ldots$ where $\|\cdot\|$ denotes the 2-norm. Then

   (a) (i), (iii) $\Rightarrow$ Each accumulation point $\bar{x}$ of $\{x_i\}$ is stationary, that is $\nabla f(\bar{x}) = 0$.

   (b) (i), (ii), (iii) $\Rightarrow$ $\{x_i\}$ converges to the unique solution $\bar{x}$ of $\min\limits_{x \in R^n} f(x)$ at the linear root rate:

$$\|x_i - \bar{x}\| \leqq \left(\frac{2}{k}\Big(f(x_0) - f(\bar{x})\Big)\right)^{\frac{1}{2}} \left(\left(1 - \frac{2\alpha k^2}{K}\right)^{\frac{1}{2}}\right)^i$$

Proof.

(a) Let $\{x_{i_j}\} \Rightarrow \bar{x}$. Then

$$f_{ij} - f_{i_{j+1}} \geqq f_{i_j} - f_{i_j+1} \geqq \alpha\|\nabla f_{i_j}\|^2 \geqq 0$$

Taking the limit as $j \to \infty$ and invoking the continuity of ∇f we get: $0 = f(\bar{x}) - f(\bar{x}) \geqq \alpha\|\nabla f(\bar{x})\|^2 \geqq 0$.

(b) By (ii) S is compact and f is strictly convex. Hence $\{x_i\}$ must have at least one accumulation point $\bar{x}$ which minimizes $f(x)$ in R^n. By the strict convexity of f and part (a) above every accumulation point must be the same and hence [Buck, p. 47, Prob. 7] the sequence $\{x_i\}$ converges to $\bar{x}$ which is the unique solution of $\min\limits_{x\in R^n} f(x)$.

We establish now the linear root rate convergence of $\{x_i\}$.

$$\|\nabla f_i\|\cdot\|x_i-\bar{x}\| = \|\nabla f_i - \nabla f(\bar{x})\|\ \|x_i-\bar{x}\| \geqq \Big(\nabla f_i - \nabla f(\bar{x})\Big)(x_i-\bar{x}) \geqq k\|x_i-\bar{x}\|^2 \text{ (By (ii))}$$

Hence (iii) becomes

$$f_i - f_{i+1} \geqq \alpha k^2\|x_i-\bar{x}\|^2,$$

which upon using the quadratic bound lemma B.1.3 becomes:

$$f_i - f_{i+1} \geqq \frac{2\alpha}{K}k^2\Big(f_i - f(\bar{x})\Big),$$

which is equivalent to

$$\left(1-\frac{2\alpha k^2}{K}\right)\Big(f_i - f(\bar{x})\Big) \geqq f_{i+1} - f(\bar{x})$$

By induction we have then

$$\Big(f_i - f(\bar{x})\Big) \leqq \left(1-\frac{2\alpha k^2}{K}\right)^i\Big(f_0 - f(\bar{x})\Big) \qquad (*)$$

But from (ii) we have

$$\frac{k}{2}\|x_i-\bar{x}\|^2 + \nabla f(\bar{x})(x_i-\bar{x}) \leqq f_i - f(\bar{x})$$

or, since $\nabla f(\bar{x}) = 0$,

$$\|x_i-\bar{x}\| \leqq \left(\frac{2}{k}\Big(f_i - f(\bar{x})\Big)\right)^{\frac{1}{2}}.$$

Using this last inequality in (*) gives

$$\|x_i-\bar{x}\| \leqq \left(\left(1-\frac{2\alpha k^2}{K}\right)^{\frac{1}{2}}\right)^i\left(\frac{2}{k}\Big(f_i - f(\bar{x})\Big)\right)^{\frac{1}{2}}$$

□

Note: $\frac{k}{K}\to 0 \Leftrightarrow \frac{K}{k}\to\infty \Leftrightarrow$ condition number of $\nabla^2 f\to\infty \Rightarrow$ Poor rate of convergence.

6 Point of Attraction & Related Results

1 Ostrowski's point of attraction & convergence rate theorem

Assume that $G\colon R^n\to R^n$ is differentiable at the fixed point $\bar{x} = G(\bar{x})$, that $\nabla G(\bar{x})$ is symmetric and the positive spectral radius $\rho = \rho\Big(\nabla G(\bar{x})\Big) < 1$. Then $\bar{x}$ is a point of attration of the iteration

$$x_{i+1} = G(x_i).$$

That is for some $\delta > 0$:

$$\left.\begin{array}{r} x_0 \in S := \{x \big| \|x-\bar{x}\| < \delta\},\ x_{i+1} = G(x_i) \\ i = 0,1,\ldots \end{array}\right\rangle \Longrightarrow \left\langle \begin{array}{c} x_i \in S,\ i = 0,1,\ldots \\ x_i \to \bar{x} \end{array}\right.$$

Furthermore for any $\nu > 1$ such that $\nu\rho < 1$ we have that:

$$\frac{\|x_i - \bar{x}\|}{\|x_{i-1} - \bar{x}\|} \leqq \nu\rho < 1 \qquad \text{(Linear quotient rate)}$$

and

$$\|x_i - \bar{x}\| \leqq (\nu\rho)^i \|x_0 - \bar{x}\| \qquad \text{(Linear root rate)}$$

Proof. Fix $\nu > 1$ s.t. $\nu\rho < 1$. Since G is differentiable at $\bar{x}$, there exists $\delta > 0$ s.t.:

$$\|G(x) - G(\bar{x}) - \nabla G(\bar{x})(x-\bar{x})\| \leqq (\nu - 1)\rho\|x - \bar{x}\|,$$

whenever $x \in S = \{x \big| \|x - \bar{x}\| < \delta\}$. Therefore for any $x \in S$

$$\begin{aligned} \|G(x) - G(\bar{x})\| &\leqq \|G(x) - G(\bar{x}) - \nabla G(\bar{x})(x-\bar{x})\| + \|\nabla G(\bar{x})(x-\bar{x})\| \\ &\leqq \Big((\nu-1)\rho + \|\nabla G(\bar{x})\|\Big)\|x-\bar{x}\| = \nu\rho\|x-\bar{x}\| \end{aligned}$$

So if $x_0 \in S$ then

$$\|x_1 - \bar{x}\| \leqq \|G(x_0) - G(\bar{x}\| \leqq \nu\rho\|x_0 - \bar{x}\|$$

and so $x_i \in S$ and by induction:

$$\|x_i - \bar{x}\| \leqq \nu\rho\|x_{i-1} - \bar{x}\| \leqq \ \ldots \ \leqq (\nu\rho)^i\|x_0 - \bar{x}\| \leqq (\nu\rho)^i\delta \leqq \delta$$

Hence $x_i \in S,\ x_i \to \bar{x}$ and the two convergence rate results hold. □

Note 1: If $\nabla G(\bar{x})$ is nonsymmetric replace in the above ρ by $\sigma = \left[\rho\Big(\nabla G(\bar{x})^T\nabla G(\bar{x})\Big)\right]^{\frac{1}{2}}$

Note 2: The linear root rate implies that $\|x_i - \bar{x}\| \leqq (\bar{\nu}\rho)^i$ for $i \geqq \bar{i}$ for some $\bar{\nu},\ \bar{i}$ where: $1/\gamma\rho \leqq \bar{\nu} < 1/\rho$ for some $\gamma s.t. : 1 < \gamma < 1/\nu\rho$.

2 Differentiability of the Newton mapping

Let $F: R^n \to R^n$ be differentiable at $\bar{x}$ where $F(\bar{x}) = 0$ and let $C: R^n \to R^{n\times n}$ be continuous at $\bar{x}$ and nonsingular. Then the Newton mapping

$$G(x) := x - C(x)^{-1}F(x)$$

is well defined in a neighborhood for $\bar{x}$ and is differentiable at $\bar{x}$ with

$$\nabla G(\bar{x}) = I - C(\bar{x})^{-1}\nabla F(\bar{x}) \qquad (*)$$

Proof. We first show that $C(x)$ is nonsingular in a neighborhood of $\bar{x}$. Let $\beta = \|C(\bar{x})^{-1}\|$ and $0 < \varepsilon < (2\beta)^{-1}$. By the continuity of C at $\bar{x}$ there exists a $\delta > 0$ s.t.

$$\|C(x) - C(\bar{x})\| \leqq \varepsilon,$$

whenever $x \in S := \{x \Big| \|x - \bar{x}0 \leqq \delta\}$. Hence by the Banach Perturbation Lemma, $C(x)$ is nonsingular and

$$\|C(x)^{-1}\| \leqq \frac{\beta}{1 - \beta\varepsilon} < 2\beta \ \text{ for } x \in S.$$

Thus G is well defined for all $x \in S$. Now since F is differentiable at $\bar{x}$ we may assume that δ is chosen small enough so that

$$\|F(x) - F(\bar{x}) - \nabla F'(\bar{x})(x - \bar{x})\| \leqq \varepsilon\|x - \bar{x}\| \ \text{ for } x \in S.$$

Since $\bar{x} = G(\bar{x})$ we have for all $x \in S$ that:

$$\begin{aligned}
\|G(x) \quad - &\quad G(\bar{x}) - \Big(I - C(\bar{x})^{-1}\nabla F(\bar{x})\Big)(x - \bar{x})\| \\
= &\quad \|C(\bar{x})^{-1}\nabla F(\bar{x})(x - \bar{x}) - C(x)^{-1}F(x)\| \\
\leqq &\quad \|C(x)^{-1}\Big(F(x) - F(\bar{x}) - \nabla F(\bar{x})(x - \bar{x})\Big)\| \\
&\quad + \|C(x)^{-1}\Big(C(\bar{x}) - C(x)\Big)C(\bar{x})^{-1}\nabla F(\bar{x})(x - \bar{x}\| \\
\leqq &\quad \Big(2\beta\varepsilon + 2\beta^2\varepsilon\|\nabla F(\bar{x})\|\Big)\|x - \bar{x}\|
\end{aligned}$$

Since ε is arbitrary and $\|\nabla F(\bar{x})\|$ and β are constant, the above inequality shows that G is differentiable at $\bar{x}$ and that (*) holds. $\square$