Homework Problem 1.1 (Miscellaneous on Convergence Rates) 9.5 Points

(*i*) Explain why the definition of Q-quadratic convergence of a sequence requires the initial assumption that the sequence converges at all.

Otherwise, the sequence $x^{(k)} = (-1)^k$ with $C = 1$ would satisfy the requirements of Q-Quadratic convergence for $x^* = 0$

(*ii*) Show that Q-quadratic convergence implies Q-superlinear convergence which implies Q-linear convergence which implies convergence.

1. Q-quadratic $\Rightarrow$ Q-superlinear

Choose $\varepsilon^{(k)} := C \cdot \|x^{(k)} - x^*\|_M$. As $x^{(k)}$ converges to x^*, this is a null sequence. Then,

$$\|x^{(k+1)} - x^*\|_M \leq C \cdot \|x^{(k)} - x^*\|_M^2 \quad \text{(Q-quadratic convergence)}$$
$$= \varepsilon^{(k)} \|x^{(k)} - x^*\|_M$$

$\Rightarrow$ $x^{(k)}$ converges to x^* Q-superlinearly. $\square$

2. Q-superlinear $\Rightarrow$ Q-linear

As $\varepsilon^{(k)} \to 0$, there exists $c \in (0,1)$ s.t. $\exists N \in \mathbb{N}: \forall n \geq N: \varepsilon^{(k)} < c$ (from the definition of Q-superlinear convergence it is clear that $\varepsilon^{(k)} \geq 0 \quad \forall k \in \mathbb{N}$)

Thus, $\forall n \geq N$

$$\|x^{(n+1)} - x^*\|_M \leq \varepsilon^{(n)} \|x^{(n)} - x^*\|_M$$
$$\leq c \|x^{(n)} - x^*\|_M,$$

which is the definition of Q-linear convergence. $\square$

3. Q-linear $\Rightarrow$ convergence

Let $\varepsilon > 0$. Consider $C = \|x^{(0)} - x^*\|_M$.

As $c \in (0,1)$, we can find a $n \in \mathbb{N}$ s.t. $c^n \cdot C < \varepsilon$

$\Rightarrow$ $\|x^{(n)} - x^*\|_M \leq c^n \cdot \|x^{(0)} - x^*\|_M = c^n \cdot C < \varepsilon$ $\square$

(*iii*) (a) Show that the notions of Q-linear, Q-superlinear and Q-quadratic convergence of a sequence imply their respective R-convergence counterparts.

Q-linear $\Rightarrow$ R-linear:

Choose $\varepsilon^{(k)} = \|x^{(k)} - x^*\|_m$. $\exists c \in (0,1)$ s.t.

$$\|x^{(k+1)} - x^*\|_m \leq c \cdot \|x^{(k)} - x^*\|_m \quad \text{for large enough } k.$$

$$\Rightarrow \quad \varepsilon^{(k+1)} \leq c \cdot \varepsilon^{(k)}$$

$$\Rightarrow \quad |\varepsilon^{(k+1)} - 0| \leq c \cdot |\varepsilon^{(k)} - 0|$$

$\Rightarrow$ $\varepsilon^{(k)}$ converges to 0 Q-linearly w.r.t. $|\cdot|$

Also, $\|x^{(k)} - x^*\|_m = \varepsilon^{(k)}$ $\forall k \in \mathbb{N}$, i.e. $x^{(k)}$ converges R-linearly.

Q-superlinear $\Rightarrow$ R-superlinear

Analogously, choose $\varepsilon^{(k)} = \|x^{(k)} - x^*\|$.

It is enough to show that $\varepsilon^{(k)}$ converges to 0 Q-superlinearly

$\exists$ null sequence $(\delta^{(k)})$ s.t.

$$\|x^{(k+1)} - x^*\|_m \leq \delta^{(k)} \|x^{(k)} - x^*\|$$

$$\Rightarrow \quad \varepsilon^{(k+1)} \leq \delta^{(k)} \varepsilon^{(k)}$$

$$\Rightarrow \quad |\varepsilon^{(k+1)} - 0| \leq \delta^{(k)} |\varepsilon^{(k)} - 0| \qquad \square$$

Q-quadratically $\Rightarrow$ R-quadratically

Analogously, it is enough to show that $\varepsilon^{(k)} := \|x^{(k)} - x^*\|_m$ converges to 0 Q-quadratically:

$$\|x^{(k+1)} - x^*\|_m \leq C \|x^{(k)} - x^*\|^2 \Rightarrow \varepsilon^{(k+1)} \leq C \cdot \varepsilon^{(k)2} \Rightarrow |\varepsilon^{(k+1)} - 0| \leq C \cdot |\varepsilon^{(k)} - 0|^2 \quad \square$$

(b) Give an example that shows that R-convergence of a sequence generally does not imply Q-convergence.

Consider the sequence $x^{(k)} := 2^{(-1)^k - k}$. It doesn't converge Q-linearly to 0, as

$$|x^{(2k)} - 0| = 2^{1-2k} > 2^{-2k} = 2^{-1-2k+1} = |x^{(2k-1)} - 0| \quad \forall k.$$

It does, however, converge R-linearly: Let $\varepsilon^{(k)} := 2^{1-k}$.

Then $|x^k - 0| = 2^{(-1)^k - k} \leq 2^{1-k} = \varepsilon^{(k)}$ and for $c = \frac{1}{2}$,

$$|\varepsilon^{(k+1)} - 0| = 2^{1-(k+1)} = 2^{-k} = c \cdot 2^{1-k} = c \cdot |\varepsilon^{(k)} - 0|,$$

$\Rightarrow$ $\varepsilon^{(k)}$ converges Q-linearly to 0 and therefore $x^{(k)}$ converges Q-linearly to 0.

(iv) (a) Let $\|\cdot\|_a, \|\cdot\|_b \colon \mathbb{R}^n \to \mathbb{R}$ be equivalent norms. Show that Q-superlinear resp. Q-quadratic convergence of a sequence w.r.t. $\|\cdot\|_a$ implies Q-superlinear resp. Q-quadratic convergence w.r.t. $\|\cdot\|_b$.

Q-superlinear: $\exists$ null sequence $\varepsilon^{(k)}$ s.t.

$$\|x^{(k+1)} - x^*\|_a \leq \varepsilon^{(k)} \|x^{(k)} - x^*\|_a \quad \forall k.$$

As the norms are equivalent $\exists c: \|x\|_b \leq c \cdot \|x\|_a$ and $\exists C: \|x\|_a \leq C \cdot \|x\|_b$.
The sequence $\delta^{(k)} := c \cdot C \cdot \varepsilon^{(k)}$ is a null sequence, too.

$$\|x^{(k+1)} - x^*\|_b \leq c \cdot \|x^{(k+1)} - x^*\|_a \leq c \cdot \varepsilon^{(k)} \|x^{(k)} - x^*\|_a \leq c \cdot \varepsilon^{(k)} C \cdot \|x^{(k)} - x^*\|_b$$

$\Rightarrow$ x converges Q-superlinearly w.r.t. $\|\cdot\|_b$, too

Q-quadratic: $\exists D$ s.t. $\|x^{(k+1)} - x^*\|_a \leq D \cdot \|x^{(k)} - x^*\|_a^2, \quad \forall k.$

Analogously, $\exists c, C$: $\|x\|_b \leq c \cdot \|x\|_a$, $\|x\|_a \leq C \cdot \|x\|_b$

$$\Rightarrow \|x^{(k+1)} - x^*\|_b \leq c \cdot \|x^{(k+1)} - x^*\|_a \leq c \cdot D \cdot \|x^{(k)} - x^*\|_a^2 \leq c \cdot D \cdot C^2 \|x^{(k)} - x^*\|_b^2.$$

Also, as $x^{(k)} \to x^*$ w.r.t. $\|\cdot\|_a$, it also converges w.r.t. $\|\cdot\|_b$ according to Analysis 1.

(b) Give an example that shows that a similar statement can not hold for Q-linear convergence. Does it hold for R-linear convergence?

Consider the sequence
$$x^{(k)} = \begin{cases} 2^{-k} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \text{if } k \text{ even,} \\ 2^{-k+1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \text{else} \end{cases}$$

1. w.r.t. $\|\cdot\|_1$, it converges Q-linearly with $c = \frac{1}{2}$,

because
$$\|x^{(k)} - 0\|_1 = \begin{cases} (2^{-k} - 0) + (2^{-k} - 0) & \text{if } k \text{ even} \\ (2^{-k+1} - 0) + (0 - 0) \end{cases} = 2^{-k+1},$$

$$\Rightarrow \|x^{(k+1)} - 0\|_1 = 2^{-k} = \frac{1}{2} \cdot 2^{-k+1} = c \cdot \|x^{(k)} - 0\|_1$$

2. w.r.t. $\|\cdot\|_\infty$, it doesn't converge Q-linearly, because

$$\|x^{(2k+1)} - 0\|_\infty = 2^{-2k} = \|x^{(2k)} - 0\|_\infty \Rightarrow \nexists\, c \in (0,1).$$

3. Yes, R-linear convergence w.r.t. $\|\cdot\|_a$ $\Rightarrow$ R-linear convergence w.r.t. $\|\cdot\|_b$.

proof: Let $x^{(k)}$ converge R-linearly w.r.t. $\|\cdot\|_a$, i.e.

$$\|x^{(k)} - x^*\|_a \leq \varepsilon^{(k)} \quad \text{where } \exists d \in (0,1) \text{ s.t.}$$
$$|\varepsilon^{(k+1)} - 0| \leq d \cdot |\varepsilon^{(k)} - 0|$$

$$\|x^{(k)} - x^*\|_b \leq c \cdot \|x^{(k)} - x^*\|_a \leq c \cdot \varepsilon^{(k)}.$$

As $|c \cdot \varepsilon^{(k+1)} - 0| \leq d\, |c \cdot \varepsilon^{(k)} - 0|$, the sequence $c \cdot \varepsilon^{(k)}$ converges Q-linearly to 0 w.r.t. $|\cdot|$ $\Rightarrow$ $x^{(k)}$ converges R-linearly w.r.t. $\|\cdot\|_b$ $\square$

(*i*) For each of the following cases, give an example of a null sequence $(x^{(k)})$ in $(\mathbb{R}, |\cdot|)$ that

(a) converges, but does not converge Q-linearly,
(b) converges Q-linearly, but does not converge Q-superlinear,
(c) converges Q-superlinearly, but does not converge Q-quadratically,
(d) converges Q-quadratically.

(a) Consider the sequence $x^{(k)} := 2^{(-1)^k - k}$. It doesn't converge Q-linearly to 0, as

$$x^{(2k)} - 0| = 2^{1-2k} > 2^{-2k} = 2^{-1-2k+1} = |x^{(2k-1)} - 0| \quad \forall k,$$ but it obviously converges

(b) $x^{(k)} := 2^{-k}$. As $|x^{(k+1)} - 0| = 2^{-k-1} = \frac{1}{2} \cdot 2^{-k} = \frac{1}{2} \cdot |x^{(k)} - 0|$, there is no null sequence $\varepsilon^{(k)}$ s.t. $|x^{(k+1)} - 0| \leq \varepsilon_k \cdot |x^{(k)} - 0|$

(c) $x^{(k)} = 2^{-k^2}$. Then $|x^{(k+1)} - 0| = 2^{-(k+1)^2} = 2^{-(k^2+2k+1)} = 2^{-2k-1} \cdot 2^{-k^2} \leq \varepsilon^{(k)} \cdot |x^{(k)} - 0|$

for $\varepsilon^{(k)} = 2^{-k}$. However, $|x^{(k)} - 0|^2 = 2^{-2k^2}$

Let C s.t.
$$|x^{(k+1)} - 0| \leq C \cdot |x^{(k)} - 0|^2$$
$$\Leftrightarrow \quad 2^{-k^2-2k-1} \leq C \cdot 2^{-2k^2}$$
$$\Leftrightarrow \quad 2^{k^2-2k-1} \leq C$$
$$\Leftrightarrow \quad 2^{(k-1)^2} \leq 4C \quad \lightning$$

(d) $x^{(k)} = 2^{-2^k}$. For $C = 1$, we have

$$|x^{(k+1)} - 0| = 2^{-2^{k+1}} = 2^{-2^k \cdot 2} = \left(2^{-2^k}\right)^2 = |x^{(k)} - 0|^2$$

(*ii*) Explain what the Q-convergence rates of a sequence $x^k \to x^*$ will look like in a semi-logarithmic plot, i. e., when plotting the map $k \mapsto \ln|x\_k - x^*|$.

Es gilt: $\ln|x^{k+1} - x^*| - \ln|x^k - x^*| = \ln\left|\frac{x^{k+1} - x^*}{x^k - x^*}\right|$

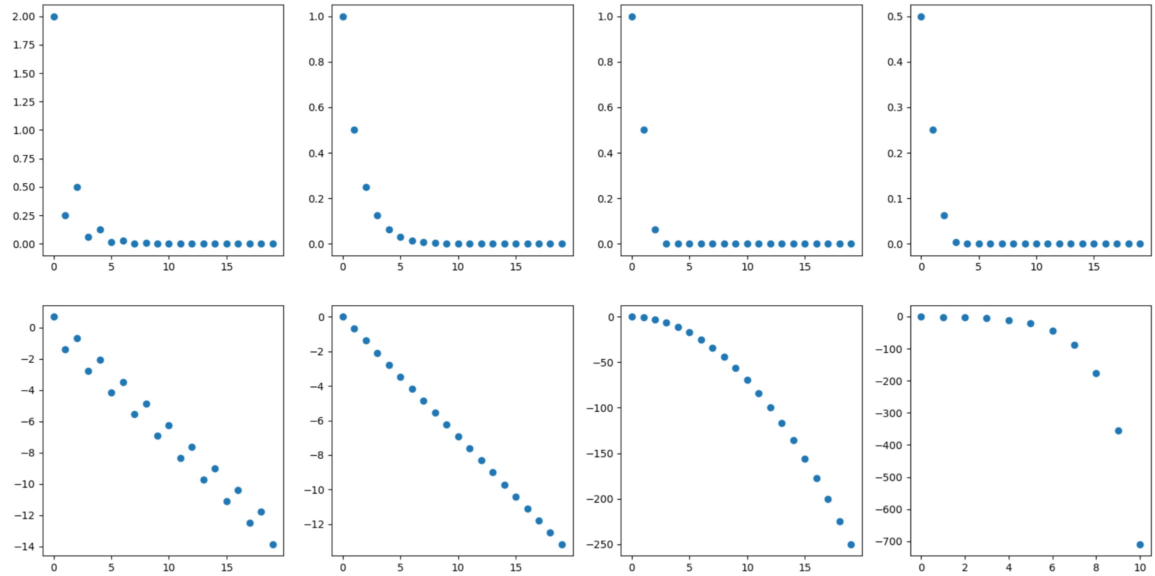
Q-linear: $\ln\left|\frac{x^{k+1} - x^*}{x^k - x^*}\right| \leq \ln c < 0$, as $c \in (0,1)$.

$\Rightarrow$ Q-linearly converging sequences are bounded by a decreasing linear function

Q-superlinear: $\ln\left|\frac{x^{k+1} - x^*}{x^k - x^*}\right| \leq \ln \varepsilon^{(k)}$. The distance has to decrease every step, so it can't be just linear; however, anything below linear like e.g. $-k^2$ is ok.

Q quadratic: $\ln|x^{k+1} - x^*| \leq 2 \cdot (\ln|x^k - x^*| + \ln c) \Rightarrow$ we expect x^k to be bounded above by an exponential function like 2^{-k}.

(c)


2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
1.0
0.8
0.6
0.4
0.2
0.0
0.5
0.4
0.3
0.2
0.1
0
5
10
15
−2
−4
−6
−8
−10
−12
−14
−50
−100
−150
−200
−250
−300
−400
−500
−600
−700
2
4
6
8

Homework Problem 1.3 (Optimality Condition Gap) 9 Points

Consider the optimization problem

$$\text{Minimize} \quad f(x) = (x_1 - x_2^2)(2x_1 - x_2^2) = 2x_1^2 - 3x_1x_2^2 + x_2^4 \quad \text{where } x \in \mathbb{R}^2.$$

(*i*) Show that the necessary optimality conditions of first and second order are satisfied at $(0,0)^\top$.

$$\nabla f(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \text{ first order O.C.}$$

$$\nabla f(x) = \begin{pmatrix} 4x_1 - 3x_2^2 \\ -6x_1x_2 + 4x_2^3 \end{pmatrix} \qquad Hf(0) = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{positive semidefinite}$$

$\Rightarrow$ second order O.C.

(*ii*) Show that $(0,0)^\top$ is a local minimizer for f along every straight line passing through $(0,0)$.

Every straight line through $(0,0)^\top$ described by $\vec{s}(t) = \begin{pmatrix} \lambda t \\ \mu t \end{pmatrix}$; $\lambda\mu \neq 0$

$$f(\vec{s}(t)) = 2\lambda^2 t^2 - 3\lambda\mu^2 t^3 + \mu^4 t^4$$

$$\left.\frac{\partial f(\vec{s}(t))}{\partial t}\right|_{t=0} = 0$$

$$\left.\frac{\partial^2 f(\vec{s}(t))}{\partial t}\right|_{t=0} = 2\lambda^2 > 0 \Rightarrow t=0, (0,0)^\top \text{ local minimum}$$

(*iii*) Show that $(0,0)^\top$ is not a local Minimizer of f on $\mathbb{R}^2$.

Assume $(0,0)$ was a local minimizer of f on $\mathbb{R}^2$. Then,
(in $\mathbb{R}$, every open set contains a ball)

w.l.o.g. $\exists \varepsilon < 1: \forall x$ s.t. $\|x\| < \varepsilon: 0 = f(0,0) \leq f(x)$

Consider $x^* := \begin{pmatrix} \frac{1}{5}\varepsilon^2 \\ \frac{1}{2}\varepsilon \end{pmatrix}$; $f(x^*) = \left(\frac{1}{5}\varepsilon^2 - \frac{1}{4}\varepsilon^2\right)\left(\frac{2}{5}\varepsilon^2 - \frac{1}{4}\varepsilon^2\right)$

$$= -\frac{1}{20}\varepsilon^2 \cdot \frac{3}{20}\varepsilon^2 = \frac{-3}{400}\varepsilon^2 < 0,$$

But: $\left\| \begin{pmatrix} \frac{1}{5}\varepsilon^2 \\ \frac{1}{4}\varepsilon \end{pmatrix} \right\| = \sqrt{\frac{1}{25}\varepsilon^4 + \frac{1}{4}\varepsilon^2} < \sqrt{\frac{1}{4}(\varepsilon^4 + \varepsilon^2)} = \frac{1}{2}\varepsilon\sqrt{1+\varepsilon^2} < \frac{1}{2}\varepsilon\sqrt{4} = \varepsilon$

Homework Problem 1.4 (First Order Conditions are Sufficient for Convex Functions) 2 Points

Let $f: \mathbb{R}^n \to \mathbb{R}$ be a convex function that is differentiable at $x \in \mathbb{R}^n$ with $f'(x) = 0$. Show that x is a global minimizer of f.

We know: $f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$

w.l.o.g. $x = 0$, $f(x) = 0$. Assume $\exists y \in \mathbb{R}^n$ s.t. $f(y) < 0$.

Then for $\alpha \in [0,1]$:

$$f(\alpha y) = f(\underset{0}{(1-\alpha) x} + \alpha y) < (1-\alpha) \underset{0}{f(x)} + \alpha f(y) = \alpha f(y) < 0$$

$$\Rightarrow \quad |f(\alpha y)| > \alpha |f(y)| \quad (1)$$

Let $\varepsilon = \frac{1}{2} \frac{|f(y)|}{\|y\|}$ (2)

with Taylor we find a δ s.t. $\forall \alpha$ with $\|\alpha y\| < \delta$:

$$\alpha |f(y)| \overset{(1)}{<} |f(\alpha y)| = |f(0) - f(\alpha y) - f'(0) \cdot \alpha y| < \varepsilon \|\alpha y\| \quad | : \alpha$$

$$\Leftrightarrow \quad |f(y)| < \varepsilon \|y\|$$

$\lightning$ because of (2).