E. Herberg, M. Marić, V. Stein
Heidelberg University

Exercise 1 - Solution

Date issued: 15th April 2024
Date due: 23rd April 2024

Homework Problem 1.1 (Optimality Condition Gap) 9 Points

Consider the optimization problem

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

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

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

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

Solution.

(i) We have

$$f'(x) = \begin{pmatrix} 4x_1 - 3x_2^2, & -6x_1x_2 + 4x_2^3 \end{pmatrix}, \quad f''(x) = \begin{pmatrix} 4 & -6x_2 \\ -6x_2 & -6x_1 + 12x_2^2 \end{pmatrix},$$

so for $x^* = (0,0)^\intercal$ we know that

$$f(x^*) = 0, \qquad f'(x^*) = (0,0), \qquad f''(x^*) = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}.$$

Since the Hessian $f''(x^*)$ is positive semidefinite with eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 0$, the first and second order optimality conditions in Theorems 2.1 and 2.2 are satisfied. The sufficient conditions in Theorem 2.3 however are not satisfied. (2 Points)

(ii) For arbitrary but fixed $d \in \mathbb{R}^2$, $d \neq 0$, we consider the line $\ell(t) = x^* + t\,d$ for $t \in \mathbb{R}$ and $x^* = (0,0)^\mathsf{T}$ to obtain

$$\begin{aligned} f(\ell(t)) &= \left(t\,d_1 - t^2\,d_2^2\right)\left(2t\,d_1 - t^2\,d_2^2\right) \\ &= \left(2t^2\,d_1^2 - 3t^3\,d_1\,d_2^2 + t^4\,d_2^4\right), \\ \frac{\mathrm{d}}{\mathrm{d}t} f(\ell(t)) &= 4td_1^2 - 9t^2\,d_1\,d_2^2 + 4t^3 d_2^4, \\ \frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\ell(t)) &= 4d_1^2 - 18t\,d_1\,d_2^2 + 12t^2 d_2^4. \end{aligned}$$

(2 Points)

and hence $\frac{\mathrm{d}}{\mathrm{d}t} f(\ell(t))|_{t=0} = 0$ as well as $\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\ell(t))|_{t=0} = 4d_1^2$. Accordingly, the sufficient conditions of first and second order for the function restricted to the ray are satisfied at $t = 0$, making $(0,0)$ a strict local minimizer (2 Points)

When $d_1 = 0$, then $f \circ \ell$ is of the form $f(\ell(t)) = t^4\,d_2^2 \geq 0 = f(\ell(0))$, making $t = 0$ an obvious (strict, global) minimizer of $f \circ \ell$. (1 Point)

(iii) Looking at the plot in homework problem 1.1 and at the function definition, we notice a bi-parabolic structure, i. e., we can make the Ansatz

$$f(x) = 2x_1^2 - 3x_1x_2^2 + x_2^4 \overset{!}{=} (ax_1 + bx_2^2)(cx_1 + dx_2^2)$$

to find that $bd = 1$ (e. g. $b = d = -1$) and accordingly $a + c = 3$ and $ac = 2$ (e. g. $a = 2$, $c = 1$), so that

$$f(x) = (2x_1 - x_2^2)(x_1 - x_2^2).$$

Accordingly, the parabolas $x_1 = 0.5x_2^2$ and $x_1 = x_2^2$ yield the zero-levelset of f and $x_1 = \alpha x_2^2$ for $\alpha \in (\frac{1}{2}, 1)$ yields negative values everywhere except for the origin.

We therefore consider the path $\gamma(t) = (3/4t^2, t)$ for $t \in \mathbb{R}$. Then,

$$\begin{aligned} f(\gamma(t)) &= \left(\frac{3}{4}t^2 - t^2\right)\left(\frac{3}{2}t^2 - t^2\right) \\ &= -\frac{1}{8}t^4 < 0 = f(x^*) \quad \text{für } t \neq 0, \end{aligned}$$

making it clear that x^* is not a local minimizer. (2 Points)

Conclusion: There is a gap between necessary and sufficient optimality conditions for the existence of minimizers. Even the surprisingly strong property in task (ii) is insufficient in addition to the first and second order necessary conditions.

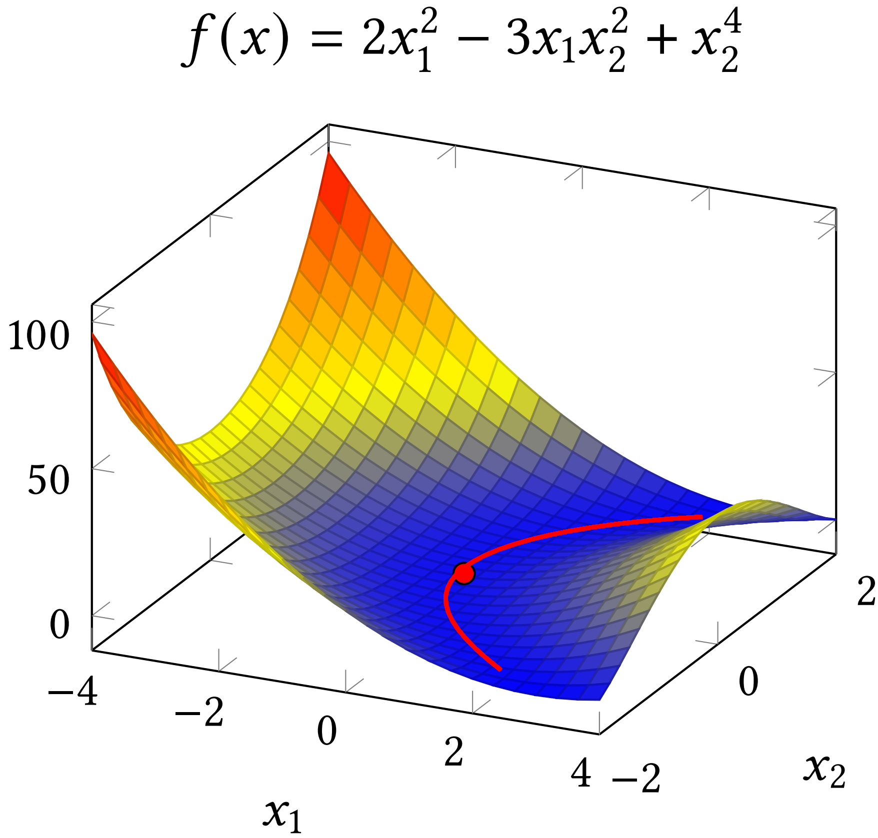

Figure 0.1: Plot of the function and a path with negative function values except for the origin.

Homework Problem 1.2 (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.

Solution.

Since f is differentiable at x, we know that its directional derivatives $f'(x, d)$ exist for every direction d and $f'(x, d) = f'(x)d$.

Now, let any $y \in \mathbb{R}^n$ be given. Due to the convexity, we know that for $t \in (0, 1)$

$$f(x + t(y - x)) \leq f(x) + t(f(y) - f(x)),$$

and accordingly,

$$0 = f'(x)(y - x) \xleftarrow{t \searrow 0,\, t<1} \frac{f(x + t(y - x)) - f(x)}{t} \leq f(y) - f(x),$$

implying that

$$f(x) \leq f(y) \; \forall y \in \mathbb{R}^n.$$

(2 Points)

Homework Problem 1.3 (Miscellaneous on Convergence Rates) 11 Points

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

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

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

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

(iv) (a) Let $\|\cdot\|_a, \|\cdot\|_b : \mathbb{R}^n \to \mathbb{R}$ 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. $\|\cdot\|_b$.

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

Solution.

(i) The condition that there exists $C > 0$ such that

$$\|x^{(k+1)} - x^*\|_M \leq C\,\|x^{(k)} - x^*\|_M^2 \quad \text{for all } k \in \mathbb{N}$$

does not imply convergence of the sequence as the constant C and the distance $\|x^{(k)} - x^*\|_M$ may remain greater or equal than 1, see, e. g., the alternating sequence $x^{(k)} = (-1)^k$ and $x^* = 0$ with $C = 1$. (1 Point)

(ii) When $x^{(k)}$ is Q-quadratically convergent to x^*, then there exists $C > 0$ such that

$$\|x^{(k+1)} - x^*\|_M \leq \underbrace{C\,\|x^{(k)} - x^*\|_M}_{=:\varepsilon^{(k)}}\,\|x^{(k)} - x^*\|_M \quad \text{for all } k \in \mathbb{N}$$

where $\varepsilon^{(k)} = C\,\|x^{(k)} - x^*\|_M$ is a null sequence by assumption.

When $x^{(k)}$ is Q-superlinearly convergent to x^*, then there exists a null sequence $\left(\varepsilon^{(k)}\right)$ such that for any $c \in (0, 1)$

$$\|x^{(k+1)} - x^*\|_M \leq \varepsilon^{(k)}\,\|x^{(k)} - x^*\|_M \leq c\,\|x^{(k)} - x^*\|_M \quad \text{for all } k \in \mathbb{N} \text{ sufficiently large}$$

due to the null sequence property.

When $x^{(k)}$ is Q-linearly convergent to x^*, then there exists $c \in (0, 1)$ such that

$$\|x^{(k+1)} - x^*\|_M \leq c\,\|x^{(k)} - x^*\|_M \leq c^k \|x^{(1)} - x^*\|_M \xrightarrow{k\to\infty} 0.$$

(1.5 Points)

(iii) (a) In each case, we can define the sequence $\varepsilon^{(k)} := \|x^{(k)} - x^*\|_M$ to trivially obtain that $\|x^{(k)} - x^*\|_M \leq \varepsilon^{(k)}$ with equality. Because, as we have seen, each convergent sequence with a rate converges to its limit and therefore the distance is a nullsequence, we know that $\varepsilon^{(k)}$ is a nullsequence.

When $x^{(k)}$ is Q-linearly convergent to x^* with constant $c \in (0, 1)$, then

$$\varepsilon^{(k+1)} = \|x^{(k+1)} - x^*\|_M \leq c\|x^{(k)} - x^*\|_M = c\varepsilon^{(k)} \quad \text{for all } k \in \mathbb{N} \text{ sufficiently large.}$$

When $x^{(k)}$ is Q-superlinearly convergent to x^* with nullsequence $\tilde{\varepsilon}^{(k)}$, then

$$\varepsilon^{(k+1)} = \|x^{(k+1)} - x^*\|_M \leq \tilde{\varepsilon}^{(k)} \|x^{(k)} - x^*\|_M = \tilde{\varepsilon}^{(k)} \varepsilon^{(k)} \quad \text{for all } k \in \mathbb{N}.$$

When $x^{(k)}$ is Q-quadratically convergent to x^* with constant $C > 0$, then

$$\varepsilon^{(k+1)} = \|x^{(k+1)} - x^*\|_M \leq C\|x^{(k)} - x^*\|_M^2 = C{\varepsilon^{(k)}}^2 \quad \text{for all } k \in \mathbb{N}.$$

(2 Points)

(b) Q-convergence forces the reduction in the distance to the minimizer relative to the previous distance in every iteration while R-convergence only requires a bound that behaves like that as a nullsequence, i. e., R-convergent sequences have the freedom of increasing the distance the limit an infinite number of times as long as the overall convergence remains fast.

Consider, e. g., the sequence

$$x^{(k)} := \begin{cases} c^k, & k \text{ even} \\ 0, & \text{else} \end{cases}$$

for $c \in (0, 1)$, which is clearly R-linearly convergent to 0 (it's bound is the "mother of all Q-linearly convergent sequences": c^k). It is not Q-linearly convergent, because it actually attains its limit in every other iterate. (1 Point)

(iv) (a) Let the norm equivalence $\underline{\alpha}\|\cdot\|_b \leq \|\cdot\|_a \leq \overline{\alpha}\|\cdot\|_b$ hold for two constants $\underline{\alpha}, \overline{\alpha} > 0$. Additionally, let $(x^{(k)})_{k\in\mathbb{N}_0}$, x^* be in $\mathbb{R}^n$.

- When $x^{(k)} \longrightarrow x^*$ superlinearly in $\|\cdot\|_a$, then there exist $\varepsilon^{(k)} \longrightarrow 0$, such that

  $$\underline{\alpha}\, \|x^{(k+1)} - x^*\|_b \leq \|x^{(k+1)} - x^*\|_a \leq \varepsilon^{(k)} \|x^{(k)} - x^*\|_a \leq \overline{\alpha}\, \varepsilon^{(k)} \|x^{(k)} - x^*\|_b$$

  and hence

  $$\|x^{(k+1)} - x^*\|_b \leq \frac{\overline{\alpha}}{\underline{\alpha}}\, \varepsilon^{(k)} \|x^{(k)} - x^*\|_b$$

  with $\frac{\overline{\alpha}}{\underline{\alpha}}\varepsilon^{(k)} \longrightarrow 0$. (1.5 Points)

- When $x^{(k)} \longrightarrow x^*$ Q-quadratically in $\|\cdot\|_a$, then the sequence converges to x^* w.r.t $\|\cdot\|_a$ and therefore also w.r.t. $\|\cdot\|_b$. Additionally, there is $C > 0$, such that

  $$\underline{\alpha}\|x^{(k+1)} - x^*\|_b \leq \|x^{(k+1)} - x^*\|_a \leq C\|x^{(k)} - x^*\|_a^2 \leq \overline{\alpha}^2 C\|x^{(k)} - x^*\|_b^2$$

and hence

$$\|x^{(k+1)} - x^*\|_b \leq \frac{\overline{\alpha}^2}{\underline{\alpha}} C \|x^{(k)} - x^*\|_b^2$$

with $\frac{\overline{\alpha}^2}{\underline{\alpha}} C > 0$.

(1.5 Points)

(b) Consider $\mathbb{R}^2$ with the euklidean Norm und a norm that has a scaling along one of the axes, i. e., for a parameter $\overline{\alpha} > 1$ the norms

$$\|x\|_a = \sqrt{x^\mathsf{T} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x} \quad \text{und} \quad \|x\|_b = \sqrt{x^\mathsf{T} \begin{pmatrix} \overline{\alpha}^2 & 0 \\ 0 & 1 \end{pmatrix} x}$$

Then

$$\|x\|_b \leq \|x\|_a \leq \overline{\alpha} \|x\|_b$$

for all $x \in \mathbb{R}^2$. We now consider the sequence that jumps between the axes, hitting the unfavourable relative scaling in every other iteration. For $c := \frac{1}{\overline{\overline{\alpha}}} \in (0, 1)$ define

$$x^{(k)} := \begin{cases} (c^k, 0) & \text{for k even} \\ (0, c^k) & \text{for k uneven} \end{cases}$$

It is clear that $\|x^{(k)}\|_a = c^k \longrightarrow 0$ and therefore

$$\|x^{(k+1)}\|_a = c^{k+1} = c\, c^k = c\, \|x^{(k)}\|_a$$

shows Q-linear convergence to $0 \in \mathbb{R}^2$ in the euclidean norm. For uneven $k \in \mathbb{N}_0$ however, we have that

$$\|x^{(k+1)}\|_b = \overline{\alpha} c^{k+1} = \overline{\alpha} c\, c^k = \overline{\alpha} c\, \|x^{(k)}\|_b = \|x^{(k)}\|_b.$$

Jumping to the scaled x_1-Achse (ungerades k auf gerades k) does not decrease the distance to the limit in the b-norm, so there can be no linear convergence in this norm. (The equivalence-constants scale the parameter c to 1.) (2 Points)

Any sequence $x^{(k)}$ that converges R-linearly in a norm $\|\cdot\|_a$ will also converge R-linearly in any equivalent norm $\|\cdot\|_b$, as we have

$$\|x^{(k)} - x^*\|_b \leq \overline{\alpha} \|x^{(k)} - x^*\|_a \leq \overline{\alpha} \varepsilon^{(k)}$$

for the corresponding Q-linearly convergent nullsequence $\varepsilon^{(k)}$. Of course, with the corresponding $c \in (0, 1)$, we know that

$$\left(\overline{\alpha}\varepsilon^{(k+1)}\right) \leq \overline{\alpha}c\varepsilon^{(k)} = c\left(\overline{\alpha}\varepsilon^{(k)}\right)$$

meaning that $\overline{\alpha}\varepsilon^{(k)}$ is also Q-linearly convergent to 0.

(0.5 Points)

Homework Problem 1.4

(Visualizing and Interpreting Convergence Rates) 7 Points

(i) For each of the following cases, give an example of a null sequence $\left(x^{(k)}\right)$ 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, but does not converge with higher order.

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

(iii) Plot the distance to the limit for the sequences from task (i) over the iterations in a standard and a y-semi-logarithmic plot. What do you observe?

Solution.

(i) Hint: When solving this problem, think about what kind of term your quotient should be first. E. g., for the superlinear case, you need the Quotient to be a null sequence.

(a) The sequence $\left(x^{(k)}\right) = \left(\frac{1}{k}\right)$ is obviously positive and converges to 0. However, because

$$1 > \frac{|x^{(k+1)} - 0|}{|x^{(k)} - 0|} = \frac{k}{k+1} \xrightarrow{k\to\infty} 1$$

the sequence can not converge Q-linearly, as the quotient is not bounded away from 1 uniformly (i. e. the quotient deteriorates and the convergence slows down progressively, as the sequence progresses). (1 Point)

(b) The sequences $\left(x^{(k)}\right) = \left(q^k\right)$ for $q \in (0, 1)$ are obviously positive and converges to 0. They are the prime example for this class of convergent sequences, as they satisfy the Q-linear convergence condition with equality in every iteration, and their constant coincides with their base. I. e., we have that

$$\frac{|x^{(k+1)} - 0|}{|x^{(k)} - 0|} = q \in (0, 1).$$

The sequence is not Q-superlinearly convergent as the quotient is a constant greater than zero, not a nullsequence as required. (1 Point)

(c) The sequence $\left(x^{(k)}\right) = \left(\frac{1}{k!}\right)$ is obviously positive and converges to 0. It converges Q-superlinearly, because

$$\frac{|x^{(k+1)} - 0|}{|x^{(k)} - 0|} = \frac{k!}{(k+1)!} = \frac{1}{k+1} \xrightarrow{k\to\infty} 0.$$

However, looking at the quotient for higher order convergence, i. e., for any $\alpha > 1$, we obtain that

$$\frac{|x^{(k+1)} - 0|}{|x^{(k)} - 0|^\alpha} = \frac{(k!)^\alpha}{(k+1)!} = \frac{(k!)^{(\alpha-1)}}{k+1} \xrightarrow{k\to\infty} \infty,$$

meaning that this sequences converges Q-superlinearly but not with any higher (exponential) order - especially not Q-quadratically. (1 Point)

(d) The sequences $\left(x^{(k)}\right) = \left(q^{\left(2^k\right)}\right)$ for $q \in (0, 1)$ are obviously positive and converge to 0. They converges Q-quadratically, because

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

They are the prime example for this class of convergent sequences, as they satisfy the Q-quadratic convergence condition with equality in every iteration, and their constant is 1. (1 Point)

Note: The sequences $\left(x^{(k)}\right) = \left(q^{\left(\alpha^k\right)}\right)$ for $q \in (0, 1)$ and $\alpha > 1$ are the prime examples for the convergence order α for the same reason.

(ii) Since the logarithm is a monotonically increasing function, the conditions in the definitions of Q-convergence rates can be equivalently transformed to the log-ed data, i. e., we investigate the conditions for the log-ed data. Note that when a sequence attains its limit, this transformation of course breaks down, as do the concepts of Q-convergence, as such sequences would need be stay constant after attaining their limit.

(a) In the Q-linearly convergent case, we have

$$\ln\left(\left|x^{(k+1)} - x^*\right|\right) \leq \ln\left(c\left|x^{(k)} - x^*\right|\right) = \ln(c) + \ln\left(\left|x^{(k)} - x^*\right|\right) \leq (k+1-k_0)\ln(c) + \ln\left(\left|x^{(k_0)} - x^*\right|\right)$$

for k sufficiently large (larger than k_0) I. e., for the log-ed data, we expect the data to show at least constant decrease (since $\ln(c) < 0$) in every iteration for sufficiently large k, so the semi-log plot will ultimately show a decreasing linear plot with slope $\ln(c)$, shifted up by the log-ed initial error when linear convergence starts up. (1 Point)

(b) In the Q-superlinearly convergent case, we have

$$\ln\left(\left|x^{(k+1)} - x^*\right|\right) \leq \ln\left(\varepsilon^{(k)}\left|x^{(k)} - x^*\right|\right) = \ln(\varepsilon^{(k)}) + \ln\left(\left|x^{(k)} - x^*\right|\right)$$

for a null-sequence $\varepsilon^{(k)} \in \mathbb{R}$, where $\ln(\varepsilon^{(k)}) \to -\infty$ shows that we can expect the decrease per step in the log-ed data to become arbitrarily large in the limit, the curve will bend down with increasing curvature. (1 Point)

(c) In the Q-quadratically convergent case, we have

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

for a $C > 0$. At the first glance, we should expect to see the same behavior as in the Q-superlinear case, because $\left|x^{(k)} - x^*\right|$ will play the role of the sequence $\varepsilon^{(k)}$, but we actually get some additional information on how fast the increasing decrease will increase (it will be the same as the magnitude of the log-ed sequence data).

Note that for a Q-quadratically convergent sequence, we can continue estimating the error using the definition iteratively to obtain

$$\left|x^{(k+1)} - x^*\right| \leq C\left|x^{(k)} - x^*\right|^2 \leq \cdots \leq C^{2^{k+1}-1}\left|x^{(0)} - x^*\right|^{2^{k+1}}$$

Accordingly, for the log-ed data, we have that

$$\ln\left(\left|x^{(k+1)} - x^*\right|\right) \leq 2^{k+1}\ln\left(C\left|x^{(0)} - x^*\right|\right) - \ln(C),$$

showing that we can expect a plot showing negative exponential growth. (1 Point)

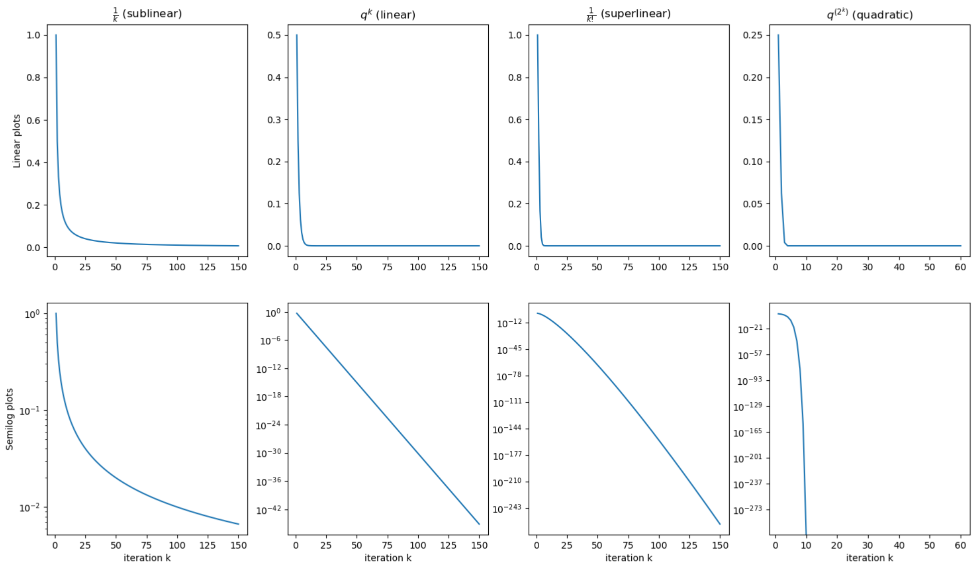

Figure 0.2: Convergence rates for various sequences in linear (top row) and y-semilog (bottom row) format.

(*iii*) See driver_ex_006_convergence_rate_visualization.py.

In the linear plots, we have no way of telling how fast a sequence is converging. In the semilog-plots, we can tell linear from superlinear convergence. Higher order convergence will always exhibit exponential behavior, that simply might be slower.

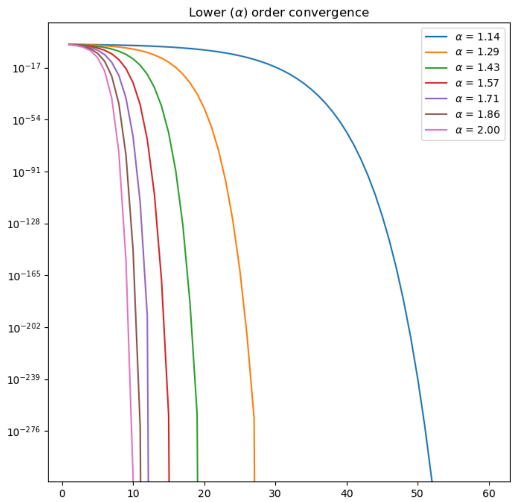

Figure 0.3: Convergence rate plots in y-semilog format for $q^{(\alpha^k)}$ with $q = 0.5$.

Please submit your solutions as a single pdf and an archive of programs via moodle.