2 Problems, algorithms, and solutions

2.1

(i) $\{1\}$,
(ii) $\emptyset$,
(iii) $\mathbb{R}$.

2.2 Suppose that there is another solution $x^{\star\star\star}$, say, to $g(x) = 0$ that is different to $x^\star = -3$ and $x^{\star\star} = 1$. There are three cases, depending on the relationship of $x^{\star\star\star}$ to the solutions $x^\star = -3$ and $x^{\star\star} = 1$. We consider each case in turn:

(i) $x^{\star\star\star} < -3$. Then:

$$\begin{aligned} g(x^{\star\star\star}) &= (x^{\star\star\star})^2 + 2x^{\star\star\star} - 3, \\ &= x^{\star\star\star}(x^{\star\star\star} + 2) - 3, \\ &> (-3)(-1) - 3, \text{ since } x^{\star\star\star} < -3, x^{\star\star\star} + 2 < -1, \\ &= 0. \end{aligned}$$

(ii) $-3 < x^{\star\star\star} < 1$. Then:

$$\begin{aligned} g(x^{\star\star\star}) &= (x^{\star\star\star})^2 + 2x^{\star\star\star} - 3, \\ &= (x^{\star\star\star} + 1)^2 - 4, \\ &< (2)^2 - 4, \text{ since } |x^{\star\star\star} + 1| < 2, \\ &= 0. \end{aligned}$$

(iii) $x^{\star\star\star} > 1$. Then:

$$\begin{aligned} g(x^{\star\star\star}) &= (x^{\star\star\star})^2 + 2x^{\star\star\star} - 3, \\ &= x^{\star\star\star}(x^{\star\star\star} + 2) - 3, \\ &> (1)(3) - 3, \text{ since } x^{\star\star\star} > 1, x^{\star\star\star} + 2 > 3, \\ &= 0. \end{aligned}$$

In each case, $g(x^{\star\star\star}) \neq 0$, so no such solution $x^{\star\star\star}$ exists that is different to $x^\star = -3$ and $x^{\star\star} = 1$.

2.3

(i) 1,

(ii) $\{2\}$.

2.4 Suppose that $\underline{f} \leq 1$. Then:

$$\begin{aligned} \underline{f} &\leq 1, \\ &\leq (x-2)^2 + 1, \forall x \in \mathbb{R}, \text{ since } (x-2)^2 \geq 0. \end{aligned}$$

So, $\underline{f}$ is a lower bound for the problem $\min_{x \in \mathbb{S}} f(x)$ according to Definition 2.2.

2.5 Suppose that $\underline{f} \leq f^{\star}$. Then:

$$\begin{aligned} \underline{f} &\leq f^{\star}, \\ &\leq f(x), \forall x \in \mathbb{S}, \end{aligned}$$

by definition of minimum. That is, $\underline{f}$ is a lower bound for $\min_{x \in \mathbb{S}} f(x)$ according to Definition 2.2.

2.6

Part	$x^{\star}$	$x^{\star\star}$	$x^{\star\star\star}$
(i) $h_1(x) \leq 0$ active?	Yes	Yes	No
(ii) $h_2(x) \leq 0$ active?	Yes	No	No
(iii) Active set?	$\{1,2\}$	$\{1\}$	$\emptyset$
(iv) Strictly feasible for $h_1(x) \leq 0$?	No	No	Yes
(v) Strictly feasible for $h_2(x) \leq 0$?	No	Yes	Yes
(vi) Strictly feasible for $h(x) \leq \mathbf{0}$?	No	No	Yes
(vii) On boundary of $\{x \in \mathbb{R}^2 \| h(x) \leq \mathbf{0}\}$?	Yes	Yes	No

2.7

(i) The contour set is defined by:

$$\begin{aligned} \mathbb{C}_f(\tilde{f}) &= \left\{x \in \mathbb{S} \,\middle|\, f(x) = \tilde{f}\right\}, \\ &= \left\{x \in \mathbb{S} \,\middle|\, (x_1)^2 + (x_2+1)^2 - 4 = \tilde{f}\right\}, \end{aligned}$$

which is the set of points $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ on the circle of radius $\sqrt{\tilde{f}+4}$ and center

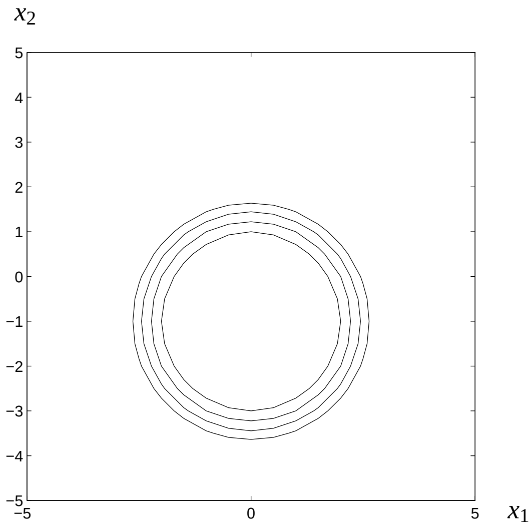

Fig. 1. Contour sets of function in Exercise 2.7, Part (i).

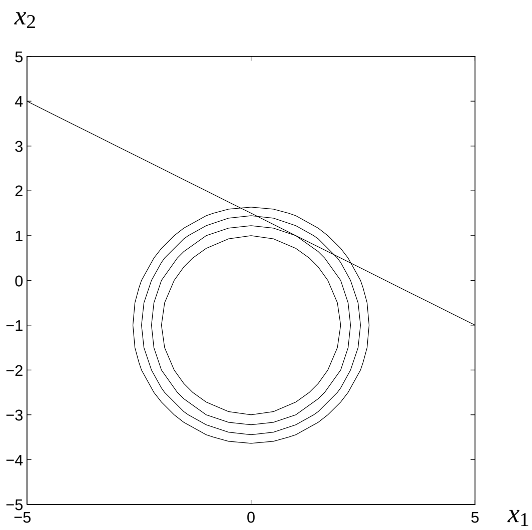

Fig. 2. Points satisfying $g(x) = \mathbf{0}$ and contour sets for Exercise 2.7, Part (ii).

$\begin{bmatrix} 0 \\ -1 \end{bmatrix}$. For $\tilde{f} = 0, 1, 2, 3$, the contour sets are the circles of radius $2, \sqrt{5}, \sqrt{6}$, and $\sqrt{7}$, respectively, all with center $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$. The contour sets are shown in Figure 1.

(ii) The set of points satisfying $g(x) = \mathbf{0}$ along with the contour sets of the function f are shown in Figure 2.

(iii) From Figure 2,

$$\begin{aligned}
\min_{x \in \mathbb{R}^2} \{ f(x) | x_1 + 2x_2 - 3 = 0 \} &= 1, \\
\operatorname{argmin}_{x \in \mathbb{R}^2} \{ f(x) | x_1 + 2x_2 - 3 = 0 \} &= \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.
\end{aligned}$$

2.8

(i) The variables are defined as follows:

- x_1 is the bandwidth for customer 1;
- x_2 is the bandwidth for customer 2; and
- x_3 is the bandwidth for customer 3.

Note that customer 3 only cares about getting data from point X to point Z, so there is only one variable associated with customer 3. We collect the three entries together into a vector $x \in \mathbb{R}^3$.

(ii) The objective is $f : \mathbb{R}^3 \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}^3_+, f(x) = \sum_{k=1}^{3} f_k(x_k).$$

(iii) With these definitions of variables, there are no equality constraints.

(iv) Both customer 1 and customer 3 traffic requires bandwidth on link a and the maximum bandwidth on this link is c_{a}. Both customer 2 and customer 3 traffic requires bandwidth on link b and the maximum bandwidth on this link is c_{b}. Since bandwidth must be positive and also since the objective function is only defined for non-negative values of bandwidth, we must also incorporate non-negativity constraints. Therefore, we can express the inequality constraints in the form $Cx \leq d$ with:

$$\begin{aligned}
C &= \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \\
d &= \begin{bmatrix} c_{\text{a}} \\ c_{\text{b}} \\ 0 \\ 0 \\ 0 \end{bmatrix}.
\end{aligned}$$

If the functions f_k were known to be strictly monotonically increasing then an alternative formulation would be to represent the capacity constraints as equalities since all capacity would be used at the optimum.

2.9

(i) Since $aB - Ab \neq 0$ we have unique solution:

$$\begin{bmatrix} X^\star \\ Y^\star \end{bmatrix} = \frac{1}{aB - Ab} \begin{bmatrix} B & -b \\ -A & a \end{bmatrix} \begin{bmatrix} c \\ C \end{bmatrix}.$$

(ii) The conditions are that:

$$\begin{aligned} aB - Ab &\neq 0, \\ \frac{1}{aB - Ab}(Bc - bC) &\geq 0, \\ \frac{1}{aB - Ab}(-Ac + aC) &\geq 0. \end{aligned}$$

(iii) If the conditions in the previous part are satisfied, then the solutions are the four possible values:

$$\begin{aligned} \begin{bmatrix} x^\star \\ y^\star \end{bmatrix} &= \begin{bmatrix} \sqrt{\frac{1}{aB-Ab}(Bc - bC)} \\ \sqrt{\frac{1}{aB-Ab}(-Ac + aC)} \end{bmatrix}, \begin{bmatrix} \sqrt{\frac{1}{aB-Ab}(Bc - bC)} \\ -\sqrt{\frac{1}{aB-Ab}(-Ac + aC)} \end{bmatrix}, \\ & \begin{bmatrix} -\sqrt{\frac{1}{aB-Ab}(Bc - bC)} \\ \sqrt{\frac{1}{aB-Ab}(-Ac + aC)} \end{bmatrix}, \begin{bmatrix} -\sqrt{\frac{1}{aB-Ab}(Bc - bC)} \\ -\sqrt{\frac{1}{aB-Ab}(-Ac + aC)} \end{bmatrix}. \end{aligned}$$

We could write this more compactly as:

$$\begin{bmatrix} x^\star \\ y^\star \end{bmatrix} = \begin{bmatrix} \pm\sqrt{\frac{1}{aB-Ab}(Bc - bC)} \\ \pm\sqrt{\frac{1}{aB-Ab}(-Ac + aC)} \end{bmatrix}.$$

However, if this notation is used then, to avoid confusion, it is important to explicitly note that all four combinations of plus and minus are valid solutions.

2.10 Suppose that an algorithm did exist for finding the minimum and all minimizers of unconstrained minimization problems $\min_{x \in \mathbb{R}^n} f(x)$, where $f : \mathbb{R}^n \to \mathbb{R}$ is an arbitrary partially differentiable function. Let $g : \mathbb{R} \to \mathbb{R}$ be an arbitrary polynomial in a single variable and define $f : \mathbb{R} \to \mathbb{R}$ by:

$$\forall x \in \mathbb{R}, f(x) = (g(x))^2.$$

The function f is partially differentiable since g is a polynomial, which is partially differentiable. Note that if $\min_{x\in\mathbb{R}^n} f(x) \neq 0$ then there is no solution to $g(x) = 0$. On the other hand, if $\min_{x\in\mathbb{R}^n} f(x) = 0$ and $x^\star \in \operatorname{argmin}_{x\in\mathbb{R}^n} f(x)$ then $g(x^\star) = 0$ and the minimizers of $\min_{x\in\mathbb{R}^n} f(x) = 0$ are the solutions of $g(x) = 0$. That is, the existence of a direct algorithm to minimize f and find all of its minimizers would also yield a direct algorithm to solve $g(x) = 0$. But no such direct algorithm to solve $g(x) = 0$ exists, so no such algorithm to minimize f and find all of its minimizers can exist.

2.11 Let $\|\bullet\|$ be the absolute value norm and let $\varepsilon > 0$ be given. We claim that $N = \lceil 1/\varepsilon \rceil$ will suffice in Definition 2.9, where $\lceil \bullet \rceil$ is the smallest integer that is greater than or equal to its argument. For, let $\nu \geq N$. Then

$$\begin{aligned}
\left\| x^{(\nu)} - 0 \right\| &= \left| \frac{1}{\nu+1} \right|, \\
&= \frac{1}{\nu+1}, \\
&\leq \frac{1}{N+1}, \\
&\leq \frac{1}{N}, \\
&\leq \varepsilon,
\end{aligned}$$

since $N \geq 1/\varepsilon$.

2.12

(i)

$$\begin{aligned}
\forall \nu \in \mathbb{Z}_+, \left\| x^{(\nu+1)} - x^\star \right\| &\leq C \left\| x^{(\nu)} - x^\star \right\|, \\
&\leq (C)^2 \left\| x^{(\nu-1)} - x^\star \right\|, \\
&\vdots \quad \vdots \\
&\leq (C)^{\nu+1} \left\| x^{(0)} - x^\star \right\|, \\
&\leq (C)^{\nu+1} \overline{\rho}.
\end{aligned}$$

So, $\left\| x^{(N)} - x^\star \right\| \leq (C)^N \overline{\rho} \leq \varepsilon \overline{\rho}$ if $(C)^N \leq \varepsilon$, which is true if:

$$N \ln(C) \leq \ln(\varepsilon),$$

or $N \geq \ln(\varepsilon)/\ln(C)$, noting that $0 < C < 1$. That is, $N = \lceil \ln(\varepsilon)/\ln(C) \rceil$.

(ii) We first note that:

$$\begin{aligned}
((\alpha)^2)^2 &= (\alpha)^4, \\
&= (\alpha)^{((2)^2)}, \\
(((\alpha)^2)^2)^2 &= ((\alpha)^{((2)^2)})^2, \\
&= (\alpha)^{((2)^3)}, \\
(\cdots(\alpha)^2)^2\cdots)^2 &= (\alpha)^{((2)^\nu)},
\end{aligned}$$

where there are ν exponentiations in total in the left-hand side of the last line. Therefore:

$$\begin{aligned}
\forall \nu \in \mathbb{Z}_+, \left\|x^{(\nu+1)} - x^\star\right\| &\leq C\left\|x^{(\nu)} - x^\star\right\|^2, \\
&\leq C(C)^2\left\|x^{(\nu-1)} - x^\star\right\|^4, \\
&= C(C)^2\left\|x^{(\nu-1)} - x^\star\right\|^{((2)^2)}, \\
&\leq C(C)^2(C)^{((2)^2)}\left\|x^{(\nu-2)} - x^\star\right\|^8, \\
&= C(C)^2(C)^{((2)^2)}\left\|x^{(\nu-2)} - x^\star\right\|^{((2)^3)}, \\
&\vdots \quad \vdots \\
&\leq C(C)^2(C)^{((2)^2)}\ldots(C)^{((2)^\nu)}\left\|x^{(0)} - x^\star\right\|^{((2)^{(\nu+1)})}, \\
&= (C)^{((2)^{\nu+1}-1)}\left\|x^{(0)} - x^\star\right\|^{((2)^{(\nu+1)})}, \\
&\leq (C)^{((2)^{\nu+1}-1)}(\overline{\rho})^{((2)^{(\nu+1)})}.
\end{aligned}$$

So, $\left\|x^{(N)} - x^\star\right\| \leq (C)^{((2)^N-1)}(\overline{\rho})^{((2)^N)} \leq \varepsilon\overline{\rho}$ if $(C)^{((2)^N-1)}(\overline{\rho})^{((2)^N-1)} \leq \varepsilon$, which is true if:

$$((2)^N - 1)\ln(C) + ((2)^N - 1)\ln(\overline{\rho}) \leq \ln(\varepsilon),$$

or $((2)^N - 1)(\ln(C) + \ln(\overline{\rho})) \leq \ln(\varepsilon)$. Since $\varepsilon < 1$ and $\ln(\varepsilon) < 0$ this inequality can only be true if $\ln(C) + \ln(\overline{\rho}) < 0$, which means that we must require $C\overline{\rho} < 1$. If this is true, then the condition becomes:

$$(2)^N - 1 \geq \frac{\ln(\varepsilon)}{\ln(C) + \ln(\overline{\rho})},$$

or:

$$N \geq \ln\left(\frac{\ln(\varepsilon)}{\ln(C) + \ln(\overline{\rho})} + 1\right) / \ln(2).$$

That is:

$$N = \left\lceil \ln\left(\frac{\ln(\varepsilon)}{\ln(C)+\ln(\overline{\rho})}+1\right) / \ln(2) \right\rceil .$$

(iii) We first note that:

$$\begin{aligned}
((\alpha)^R)^R &= (\alpha)^{((R)^2)}, \\
(((\alpha)^R)^R)^R &= ((\alpha)^{((R)^2)})^R, \\
&= (\alpha)^{((R)^3)}, \\
(\cdots(\alpha)^R)^R\cdots)^R &= (\alpha)^{((R)^\nu)},
\end{aligned}$$

where there are ν exponentiations in total in the left-hand side of the last line. Therefore:

$$\begin{aligned}
\forall \nu \in \mathbb{Z}_+, \left\|x^{(\nu+1)} - x^\star\right\| &\leq C\left\|x^{(\nu)} - x^\star\right\|^R, \\
&\leq C(C)^R\left\|x^{(\nu-1)} - x^\star\right\|^{((R)^2)}, \\
&\leq C(C)^R(C)^{((R)^2)}\left\|x^{(\nu-2)} - x^\star\right\|^{((R)^3)}, \\
&\vdots \quad \vdots \\
&\leq C(C)^R(C)^{((R)^2)}\dots(C)^{(R^\nu)}\left\|x^{(0)} - x^\star\right\|^{((R)^{(\nu+1)})}, \\
&= (C)^{\frac{((R)^{\nu+1}-1)}{(R-1)}}\left\|x^{(0)} - x^\star\right\|^{((R)^{(\nu+1)})}, \\
&\leq (C)^{\frac{((R)^{\nu+1}-1)}{(R-1)}}(\overline{\rho})^{((R)^{(\nu+1)})}.
\end{aligned}$$

So, $\left\|x^{(N)} - x^\star\right\| \leq (C)^{\frac{((R)^N-1)}{(R-1)}}(\overline{\rho})^{((R)^N)} \leq \varepsilon\overline{\rho}$ if $(C)^{\frac{((R)^N-1)}{(R-1)}}(\overline{\rho})^{((R)^N-1)} \leq \varepsilon$, which is true if:

$$\frac{((R)^N-1)}{(R-1)}\ln(C) + ((R)^N-1)\ln(\overline{\rho}) \leq \ln(\varepsilon),$$

or $((R)^N-1)(\frac{\ln(C)}{R-1}+\ln(\overline{\rho})) \leq \ln(\varepsilon)$. Since $\varepsilon < 1$ and $\ln(\varepsilon) < 0$ this inequality can only be true if $\ln(C)/(R-1)+\ln(\overline{\rho}) < 0$, which means that we must require $(C)^{1/(R-1)}\overline{\rho} < 1$. If this is true, then the condition becomes:

$$(R)^N - 1 \geq \frac{\ln(\varepsilon)}{\frac{\ln(C)}{R-1}+\ln(\overline{\rho})},$$

or:

$$N \geq \ln\left(\frac{\ln(\varepsilon)}{\frac{\ln(C)}{R-1}+\ln(\overline{\rho})}+1\right) / \ln(R).$$

That is:

$$N = \left\lceil \ln\left(\frac{\ln(\varepsilon)}{\frac{\ln(C)}{R-1} + \ln(\overline{\rho})} + 1\right) / \ln(R) \right\rceil.$$

2.13 We consider each sequence in turn and evaluate the corresponding limit.

(i) $\forall \nu \in \mathbb{Z}_+, x^{(\nu)} = 1/(\nu+1)$.

(a) $R = 0$:

$$\begin{aligned} \lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} &= \lim_{\nu\to\infty} \frac{\|1/(\nu+2)\|}{\|1/(\nu+1)\|^0}, \\ &= \lim_{\nu\to\infty} \|1/(\nu+2)\|, \\ &= 0. \end{aligned}$$

(b) $R = \frac{1}{2}$:

$$\begin{aligned} \lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} &= \lim_{\nu\to\infty} \frac{\|1/(\nu+2)\|}{\|1/(\nu+1)\|^{1/2}}, \\ &= \lim_{\nu\to\infty} \frac{\sqrt{\nu+1}}{\nu+2}, \\ &= \lim_{\nu\to\infty} \frac{\sqrt{\nu+1}}{\sqrt{\nu+2}} \frac{1}{\sqrt{\nu+2}}, \\ &\leq \lim_{\nu\to\infty} \frac{1}{\sqrt{\nu+2}}, \\ &= 0. \end{aligned}$$

(c) $R = 1$:

$$\begin{aligned} \lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} &= \lim_{\nu\to\infty} \frac{\|1/(\nu+2)\|}{\|1/(\nu+1)\|}, \\ &= \lim_{\nu\to\infty} \frac{\nu+1}{\nu+2}, \\ &= 1. \end{aligned}$$

(d) $R = 2$:

$$\begin{aligned} \lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} &= \lim_{\nu\to\infty} \frac{\|1/(\nu+2)\|}{\|1/(\nu+1)\|^2}, \\ &= \lim_{\nu\to\infty} \frac{(\nu+1)^2}{(\nu+2)}, \\ &= \infty. \end{aligned}$$

(ii) $\forall \nu \in \mathbb{Z}_+, x^{(\nu)} = (2)^{-\nu}$.

(a) $R = 0$:

$$\lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} = \lim_{\nu\to\infty} \frac{\|(2)^{-\nu-1}\|}{\|(2)^{-\nu}\|^0}, = \lim_{\nu\to\infty} \|(2)^{-\nu-1}\|, = 0.$$

(b) $R = \frac{1}{2}$:

$$\lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} = \lim_{\nu\to\infty} \frac{\|(2)^{-\nu-1}\|}{\|(2)^{-\nu}\|^{1/2}}, = \lim_{\nu\to\infty} (2)^{-(\nu/2)-1}, = 0.$$

(c) $R = 1$:

$$\lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} = \lim_{\nu\to\infty} \frac{\|(2)^{-\nu-1}\|}{\|(2)^{-\nu}\|}, = \lim_{\nu\to\infty} (2)^{-1}, = 1/2.$$

(d) $R = 2$:

$$\lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} = \lim_{\nu\to\infty} \frac{\|(2)^{-\nu-1}\|}{\|(2)^{-\nu}\|^2}, = \lim_{\nu\to\infty} (2)^{\nu-1}, = \infty.$$

(iii) $\forall \nu \in \mathbb{Z}_+, x^{(\nu)} = (2)^{-((2)^\nu)}$.

(a) $R = 0$:

$$\lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} = \lim_{\nu\to\infty} \frac{\left\|(2)^{-((2)^{\nu+1})}\right\|}{\|(2)^{-((2)^\nu)}\|^0}, = \lim_{\nu\to\infty} \left\|(2)^{-((2)^{\nu+1})}\right\|, = 0.$$

(b) $R = \frac{1}{2}$:

$$\begin{aligned}
\lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} &= \lim_{\nu\to\infty} \frac{\|(2)^{-((2)^{\nu+1})}\|}{\|(2)^{-((2)^{\nu})}\|^{1/2}}, \\
&= \lim_{\nu\to\infty} (2)^{-(3(2)^{\nu-1})}, \\
&= 0.
\end{aligned}$$

(c) $R = 1$:

$$\begin{aligned}
\lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} &= \lim_{\nu\to\infty} \frac{\|(2)^{-((2)^{\nu+1})}\|}{\|(2)^{-((2)^{\nu})}\|}, \\
&= \lim_{\nu\to\infty} (2)^{-((2)^{\nu})}, \\
&= 0.
\end{aligned}$$

(d) $R = 2$:

$$\begin{aligned}
\lim_{\nu\to\infty} \frac{\|x^{(\nu+1)}\|}{\|x^{(\nu)} - x^\star\|^R} &= \lim_{\nu\to\infty} \frac{\|(2)^{-((2)^{\nu+1})}\|}{\|(2)^{-((2)^{\nu})}\|^2}, \\
&= \lim_{\nu\to\infty} 1, \\
&= 1.
\end{aligned}$$

The limits are shown in the following table along with the rate of convergence.

$x^{(\nu)}$	$R=$ 0	1/2	1	2	Rate
$1/(\nu+1)$	0	0	1	∞	1 (but neither linear nor quadratic convergence)
$(2)^{-\nu}$	0	0	$\frac{1}{2}$	∞	1 (linear convergence)
$(2)^{-((2)^{\nu})}$	0	0	0	1	2 (quadratic convergence)

2.14 Let $x \in \mathbb{S}$ and assume that $f(x) \geq f(x^\star)$. Define $\phi : [0,1] \to \mathbb{R}$ by:

$$\forall t \in [0,1], \phi(t) = f(x^\star + t(x - x^\star)).$$

We have that:

$$\begin{aligned}
\forall t \in [0,1], \frac{d\phi}{dt}(t) &= \nabla f(x^\star + t(x - x^\star))^\dagger (x - x^\star), \\
\frac{d\phi}{dt}(0) &= \nabla f(x^\star)^\dagger (x - x^\star), \\
\frac{d^2\phi}{dt^2}(t) &= (x - x^\star)^\dagger \nabla^2 f(x^\star + t(x - x^\star))(x - x^\star).
\end{aligned}$$

To bound the rate of convergence and rate constant, we prove two inequalities. To prove the first inequality, note that:

$$\begin{aligned}
& f(x) - f(x^\star) \\
&= \phi(1) - \phi(0), \\
&= \int_{t=0}^{1} \frac{d\phi}{dt}(t)\, dt, \\
& \qquad \text{by the fundamental theorem of integral calculus applied to } \phi, \\
& \qquad\qquad \text{(see Theorem A.2 in Section A.4.4.1 of Appendix A)}, \\
&= \int_{t=0}^{1} \nabla f(x^\star + t(x - x^\star))^\dagger (x - x^\star)\, dt, \\
&\leq \int_{t=0}^{1} \|\nabla f(x^\star + t(x - x^\star))\| \, \|x - x^\star\| \, dt, \\
&\leq \int_{t=0}^{1} \overline{\kappa} \|x - x^\star\| \, dt, \text{ by assumption on } \nabla f, \\
&= \overline{\kappa} \|x - x^\star\|.
\end{aligned}$$

Since $f(x^{(\nu+1)}) \geq f(x^\star)$, we have that:

$$\begin{aligned}
\left\| f(x^{(\nu+1)}) - f(x^\star) \right\| &= f(x^{(\nu+1)}) - f(x^\star), \\
&\leq \overline{\kappa} \left\| x^{(\nu+1)} - x^\star \right\|.
\end{aligned}$$