CSC 860 / MATH 795: Stochastic Optimization & Simulation Methodology
Fall 2022 Assignment 2

Mariia Aleksandrovych, Yaffa Atkins, Atoosa Lotfi

EXERCISE 2.4. Show Theorem 2.6 for the case of decreasing gain size, using the following steps. First consider the recursion (2.18) (bias term zero):

$$\theta_{n+1} = \theta_n + \epsilon_n G(\theta_n), \quad \theta \in \mathbb{R}^d.$$

Let $\vartheta^\epsilon(\cdot)$ denote the interpolation process in Definition 2.9. Assume that $\sum_n \epsilon_n = \infty, \epsilon_n \to 0$ and let $x_n(t) = \vartheta^\epsilon(t_n + t), t \geq 0$, where $t_n = \sum_{k=1}^n \epsilon_k$ as before.

(a) Write the telescopic sum for $x_n(t+s) - x_n(t)$, and express this through an integral approximation.

(b) Show that $\{x_n\}$ is equicontinuous in the extended sense

(c) Use the Theorem of Ascoli-Arzelà to prove that, when $n \to \infty$, x_n converges (in the sup norm) to the solution of the ODE:

$$\frac{dx(t)}{dt} = G(x(t)). \tag{2.35}$$

(d) Argue like in the proof of Theorem 2.5 and extend your result to the biased version

$$\theta_{n+1} = \theta_n + \epsilon_n(G(\theta_n) + \beta_n(\theta_n)), \quad \theta \in \mathbb{R}^d,$$

where you assume that $\sum_n \epsilon_n \|\beta_n(\theta)\| < \infty$.

2.4)

a) Write the telescopic sum for $x_n(t+s)-x_n(t)$ and integral.

$$\theta_{n+1}^{\varepsilon}-\theta_n^{\varepsilon}=\sum_{i=n}^{n+m-1}(\theta_{i+1}^{\varepsilon}-\theta_i^{\varepsilon})=\sum_{i=1}^{n+m-1}\underbrace{e_i}\,G(\theta_i^{\varepsilon})$$

because ε_i is decreasing means $\lim_{i\to\infty}\varepsilon_i=0$

which, using $x_{\varepsilon_i}(t)=\theta_{m(t)}^{\varepsilon_i}$ and let $m(t+s)=t+s$ is equivalent to

$$x_{e_i}(t+s)-x_{\varepsilon_i}(t)=\sum_{i=m(t)}^{m(t+s)-1}\varepsilon_i\,G(\theta_i^{\varepsilon}) \qquad 1)$$

Although ε_i is decreasing, but x_{ε_i} is piecewise constant, and $G(x_{\varepsilon_i})$ is also piecewise constant so $\{t_n=n\underbrace{\varepsilon_i}, n\in\mathbb{N}\}$

each step is decreasing $t_{n+1}\leq t_n=n\varepsilon_i$

so integral on $[t, t+s]$ of $G(x_{\varepsilon_i})$ is sum;

$$\int_t^{t+s}G(x_{e_i}(u))\,du=\sum_{i=m(t)}^{m(t+s)-1}e_i\,G(\theta_i^{\varepsilon})+\rho(e_i) \qquad 2)$$

where $\rho(e_i)$ is the error in approximation and we predict by decreasing e_i, $\lim_{i\to\infty}\rho(e_i)\to 0$. (error decrease).

$$x_{e_i}(t+s)-x_{e_i}(t)=\int_t^{t+s}G(x_{e_i}(u))\,du-\rho(e_i)$$

t and s are both multiples of e_i, for simplicity didn't index them as t_i and s_i.

b). proof of equicontinuity of $x_{\epsilon_i}(t_n)$ in extended Case.

Consider interval $[r, q]$, for $0 \le r < q < \infty$, then the same in

1) Contains $m(q)-m(r)-1$ terms, for ϵ_i sufficiently small $m(r) \ge r/\epsilon_i$ and $m(q) \le q/\epsilon_i$ so the number of terms is bounded by $(q-r)/\epsilon_i$, because ϵ_i is decreasing then the number of terms in each interval increase.

Now we use Lipschitz and boundedness of G, if G is bounded by $\bar{G}$ then for small ϵ_i;

$$\| x_{\epsilon_i}(q) - x_{\epsilon_i}(r) \| = \left\| \sum_{i=m(r)}^{m(q)-1} \epsilon_i G(\theta_i^{\epsilon}) \right\| \le \epsilon_i \bar{G} \frac{(q-r)}{\epsilon_i} = \bar{G}(q-r)$$

G is Lipschitz then: $\| G(\theta_{n+1}^{\epsilon_i}) - G(\theta_n^{\epsilon_i}) \| \le L \| \theta_{n+1}^{\epsilon_i} - \theta_n^{\epsilon_i} \| = \epsilon_n L \| G(\theta_n^{\epsilon_i}) \|$; by triangle inequality:

$\| G(\theta_{n+1}^{\epsilon_i}) \| \le \| G(\theta_n^{\epsilon_i}) \| (1 + \epsilon_n L)$ because ϵ_n is decaying this one become smaller as well.

and by induction $\| G(\theta_{n+m}) \| \le \| G(\theta_n) \| (1+\epsilon_n L)^m$ for all m. by using $m = m(q)-1$, $m(t) = \lfloor t/\epsilon_n \rfloor$

because ϵ_n is decreasing $m(t)$ is increasing in each interval.

$$\| G(\theta_m) \| \le \| G(\theta_0) \| (1+\epsilon_n L)^{m(q)-1} \le \| G(\theta_0) \| (1+\epsilon_n L)^{q/\epsilon_n},$$

$m \le m(q)-1$, by using $\lim_{n \to \infty} \epsilon_n \to 0$, $(1+\epsilon_n L)^{q/\epsilon_n}$ is monotone towards $\lim_{\epsilon_n \to 0} (1+\epsilon_n L)^{\epsilon_n / L} = e^{L/q}$, so

as $\epsilon_n \to 0$, $\bar{G} = \| G(\theta_0) \| e^{L/q}$, for all $G(\theta_n)$ and $n \le m(q)$

θ_0 is independent of ϵ, and is fixed at θ_0 then $\bar{G} = \| G(\theta_0) \| e^{LT}$

$\Rightarrow | x(q) - x(r) | \le \bar{G}(q-r)$, $\epsilon_n \le \epsilon_0$ because ϵ_n is decaying for ϵ_n sufficiently small, $\forall \eta > 0$, and $T > 0$, it follows

$\| x_{\epsilon_i}(q) - x_{\epsilon_i}(r) \| \le \bar{\epsilon}$ where $|q-r| \le \bar{\epsilon}/\bar{G}$, $\sup_{0 \le q, r,\ |q-r| \le \bar{\epsilon}/\bar{G},\ q \in T} | x(q) - x(r) | \le \bar{\epsilon}$

c). use Ascoli_Arzela, and limit of interpolation is solution of ODE.
in part b, prooved that $\{x_{\varepsilon_i}(\cdot)\}$ is equicontinuous:
by Ascoli-Arzela any infinite subsequence of it is convergence.
if $\lim_{r\to\infty} \varepsilon_r \to 0 \Rightarrow \bar{x}(\cdot) = \lim_{r\to\infty} x_{\varepsilon_r}(\cdot)$ in the sup norm and contin

$$\lim_{r\to\infty} \left(x_{\varepsilon_r}(t+s) - x_{\varepsilon_r}(t) \right) \overset{a}{=} \lim \int_t^{t+s} G(x_{\varepsilon_r}(u))\,du \overset{b}{=} \int_t^{t+s} \lim G(x_{\varepsilon_r}(u))\,du$$

$$\overset{c}{=} \int_t^{t+s} G(\bar{x}(u))\,du$$, from 2) $\|\rho(\varepsilon_i)\|_{i\to\infty} \le 2\bar{\varepsilon}\bar{G}$ and Lebesque

Convergence; and becaus G is Continuous. (a, b, c), Therefor for any
Small $s>0$, $\lim \bar{x}_{\varepsilon_r}(\cdot)$ satisfies: $\dfrac{\bar{x}(t+s)}{s} = \frac{1}{s}\int_t^{t+s} \bar{G}_{\varepsilon_i}(\bar{x}(u))\,du$ 3)
by Continuity of $\bar{G}$ on $[t, t+s]$ taking the limit as $s\to 0$,
3) Converges to $G(\bar{x}(u))$ means any subsequent limit of $\{x_{\varepsilon_r}(\cdot)\}$ solves ODE.

d). if $B_{\varepsilon_i}(\theta_i^{\varepsilon}) \neq 0$, the equation 1) Turns to:

$$\int_t^{t+s} G(x_{\varepsilon_i}(u))\,du = \sum_{i=m(t)}^{m(t+s)-1} \varepsilon_i G(\theta_i^{\varepsilon}) + \sum_{i=m(t)}^{m(t+s)-1} \varepsilon_i B_{\varepsilon_i}(\theta_i^{\varepsilon}) + \rho(\varepsilon_i)$$

by using $\sum_n \|\varepsilon_i\| \|B_n\| < \infty$, $\sum_{m(r)}^{m(q)-1} \varepsilon_i B_{\varepsilon}(\theta_i) = (q-r)\,O(\varepsilon)$.
this can be added to $\rho(\varepsilon_i)$.(error).
and by using same proof as part c if $\lim_{\varepsilon_i\to 0} B(\theta) = B(\theta) \neq 0$
under appropriate smoothness, assumption on $B(\theta)$,
using Azcoli-Arzela, Any subsequence Converges to solution
of ODE for $\{x(t): t \geq 0\}$, by $dx(t)/dt = G(x(t)) + B(x(t))$
is shifts the ODE.

EXERCISE 2.6. Consider again the problem of Example 1.1 of the surfer at the beach who wishes to rescue a drowning victim. We wish to minimise $J(\theta)$, but the stationary points are only given in the implicit equation (1.7). Consider the gradient search method:

$$\theta_{n+1} = \theta_n - \epsilon J'(\theta_n).$$

(a) Show that as $\epsilon \to 0$ the interpolation processes converge to the ODE:

$$\frac{dx(t)}{dt} = \frac{\sin(\alpha_2(x(t))}{v_2} - \frac{\sin(\alpha_1(x(t))}{v_1}.$$

(b) Show that θ^* is stable and argue that a solution $x(t)$ of the above ODE must then satisfy $\lim_{t\to\infty} x(t) = \theta^*$.

(c) Program the procedure and plot the results, using $a = 2, b = 5, d = 10, v_1 = 3, v_2 = 1$, and $\epsilon = 0.05$. Hint: the derivative can be re written as:

$$J'(\theta) = \frac{1}{v_1}\frac{\theta}{\sqrt{\theta^2 + a^2}} - \frac{1}{v_2}\frac{d-\theta}{\sqrt{(d-\theta)^2 + b^2}}.$$

a) The total travel time:

$$J(\theta) = \frac{a}{v_1}\sec(\alpha_1(\theta)) + \frac{b}{v_2}\sec(\alpha_2(\theta))$$

Then the $J'(\theta)$ is $J'(\theta) = \frac{a}{v_1}\frac{\tan(\alpha_1(\theta))}{\cos(\alpha_1(\theta))}\frac{d\alpha_1}{d\theta} + \frac{b}{v_2}\frac{\tan(\alpha_2(\theta))}{\cos(\alpha_2(\theta))}\frac{d\alpha_2}{d\theta}$

Then using identities:

$$\tan(\alpha_1(\theta)) = \frac{\theta}{a} \qquad (1)$$

$$\tan(\alpha_2(\theta)) = \frac{d-\theta}{b} \qquad (2)$$

Form eq. (1)

$$\frac{\tan(\alpha_1(\theta))}{\theta} = \frac{1}{a} = const \;\Rightarrow \frac{d}{d\theta}\left(\frac{\tan(\alpha_1(\theta))}{\theta}\right) = 0 \;\Rightarrow \frac{1}{\cos^2(\alpha_1(\theta))}\frac{d\alpha_1}{d\theta} = \frac{\tan(\alpha_1(\theta))}{\theta} \Rightarrow$$

$$\frac{d\alpha_1}{d\theta} = \frac{1}{\theta}\tan(\alpha_1(\theta))\cos^2(\alpha_1(\theta)) \Rightarrow \frac{d\alpha_1}{d\theta} = \frac{1}{\theta}\sin(\alpha_1(\theta))\cos(\alpha_1(\theta))$$

From eq.(2)

$$\frac{\tan(\alpha_2(\theta))}{d-\theta} = \frac{1}{b} = const \;\Rightarrow \frac{d}{d\theta}\left(\frac{\tan(\alpha_2(\theta))}{d-\theta}\right) = 0 \;\Rightarrow \frac{1}{\cos^2(\alpha_2(\theta))}\frac{d\alpha_2}{d\theta} = -\frac{\tan(\alpha_2(\theta))}{d-\theta} \Rightarrow$$

$$\frac{d\alpha_2}{d\theta} = -\frac{1}{d-\theta}\tan(\alpha_2(\theta))\cos^2(\alpha_2(\theta)) \Rightarrow \frac{d\alpha_2}{d\theta} = -\frac{1}{d-\theta}\sin(\alpha_2(\theta))\cos(\alpha_2(\theta))$$

$$\begin{cases} \tan\left(\alpha_1(\theta)\right) = \dfrac{\theta}{a} \\ \dfrac{d\alpha_1}{d\theta} = \dfrac{1}{\theta}\sin\left(\alpha_1(\theta)\right)\cos\left(\alpha_1(\theta)\right) \\ \tan\left(\alpha_2(\theta)\right) = \dfrac{d-\theta}{b} \\ \dfrac{d\alpha_2}{d\theta} = -\dfrac{1}{d-\theta}\sin\left(\alpha_2(\theta)\right)\cos\left(\alpha_2(\theta)\right) \end{cases}$$

Substitute to the $J'(\theta)$ we have

$$\begin{aligned} J'(\theta) &= \frac{a}{v_1}\frac{\tan\left(\alpha_1(\theta)\right)}{\cos\left(\alpha_1(\theta)\right)}\frac{d\alpha_1}{d\theta} + \frac{b}{v_2}\frac{\tan\left(\alpha_2(\theta)\right)}{\cos\left(\alpha_2(\theta)\right)}\frac{d\alpha_2}{d\theta} = \\ &= \frac{\cancel{a}}{v_1}\frac{\cancel{\theta}}{\cancel{a}}\frac{1}{\cancel{\cos\left(\alpha_1(\theta)\right)}}\frac{1}{\cancel{\theta}}\sin\left(\alpha_1(\theta)\right)\cancel{\cos\left(\alpha_1(\theta)\right)} - \frac{\cancel{b}}{v_2}\frac{\cancel{d-\theta}}{\cancel{b}}\frac{1}{\cancel{\cos\left(\alpha_2(\theta)\right)}}\frac{1}{d-\theta}\sin\left(\alpha_2(\theta)\right)\cancel{\cos\left(\alpha_2(\theta)\right)} = \\ &= \frac{\sin\left(\alpha_1(\theta)\right)}{v_1} - \frac{\sin\left(\alpha_2(\theta)\right)}{v_2} \end{aligned}$$

Then the gradient search method becomes:

$$\theta_{n+1} = \theta_n - \varepsilon J'(\theta_n)$$
$$\frac{\theta_{n+1}-\theta_n}{\varepsilon} = -J'(\theta_n)$$
$$\frac{\theta_{n+1}-\theta_n}{\varepsilon} = -\frac{\sin\left(\alpha_1(\theta_n)\right)}{v_1} + \frac{\sin\left(\alpha_2(\theta_n)\right)}{v_2}$$

Rename $\theta_n \equiv x(t_n)$ so

$$\frac{x(t_{n+1})-x(t_n)}{\varepsilon} = -\frac{\sin\left(\alpha_1\left(x(t_n)\right)\right)}{v_1} + \frac{\sin\left(\alpha_2\left(x(t_n)\right)\right)}{v_2}$$

Next, we consider our interpolation process on an interval $(0,T)$, then each point is equally spaced with distance ε, such that $t_n = \varepsilon n,$ and $t_{n+1} = t_n + \varepsilon,$ for $0 \le n \le T/\varepsilon$

As $\varepsilon \to 0$ $T/\varepsilon \to \infty$

$$\lim_{\varepsilon\to 0}\frac{x(t_{n+1})-x(t_n)}{\varepsilon} \approx \frac{dx(t)}{dt}$$ and

$$\lim_{\varepsilon\to 0}\left(-\frac{\sin\left(\alpha_1\left(x(t_n)\right)\right)}{v_1} + \frac{\sin\left(\alpha_2\left(x(t_n)\right)\right)}{v_2}\right) \approx \frac{\sin\left(\alpha_2\left(x(t)\right)\right)}{v_2} - \frac{\sin\left(\alpha_1\left(x(t)\right)\right)}{v_1}$$

So $\frac{dx(t)}{dt} = \frac{\sin\left(\alpha_2\left(x(t)\right)\right)}{v_2} - \frac{\sin\left(\alpha_1\left(x(t)\right)\right)}{v_1}$

b) Minimum for the $J(\theta)$ is well defined: there is no $v \in R^d$ such that the $J(tv)$ is strictly monotone decreasing at $t \rightarrow \infty$.

$$J\left(\theta^*\right) = J^* = \min_{\theta} J(\theta)$$

Then, $x(t)$ solutions to ODE

$$\frac{dx}{dt} = G\left(x(t)\right) = \frac{\sin\left(\alpha_2\left(x(t)\right)\right)}{v_2} - \frac{\sin\left(\alpha_1\left(x(t)\right)\right)}{v_1}$$

With $x(0) = x_0 \neq S$, where S is set of stationary points.

Then, we construct Lyapunov function such that $V(x_0) = 0;\ V\left(x(t)\right) > 0 \text{ for } x(t) \neq x_0$

$$V(t) = J\left(x(t)\right) - J^*$$

For all trajectories on in a S: $x(t) \notin S$

$$\frac{dV}{dt} = \nabla J\left(x(t)\right)\frac{dx(t)}{dt} = \nabla J\left(x(t)\right)G\left(x(t)\right) = \left(\frac{\sin\left(\alpha_1(\theta)\right)}{v_1} - \frac{\sin\left(\alpha_2(\theta)\right)}{v_2}\right)\left(\frac{\sin\left(\alpha_2\left(x(t)\right)\right)}{v_2} - \frac{\sin\left(\alpha_1\left(x(t)\right)\right)}{v_1}\right) =$$

$$= -\left(\frac{\sin\left(\alpha_2\left(x(t)\right)\right)}{v_2} - \frac{\sin\left(\alpha_1\left(x(t)\right)\right)}{v_1}\right) \leq 0$$

Thus x_0 i.e. θ^* is stable point.

$$\left|V\left(x(t)\right)\right| = \left|J\left(x(t)\right) - J* = \frac{a}{v_1}\sec\left(\alpha_1\left(x(t)\right)\right) + \frac{b}{v_2}\sec\left(\alpha_2\left(x(t)\right)\right) - J*\right| \rightarrow \infty \text{ whenever as } t \rightarrow \infty$$

$V\left(x(t)\right)$ is bounded,

We have that the $\lim_{t\rightarrow\infty} x(t) \rightarrow \infty$

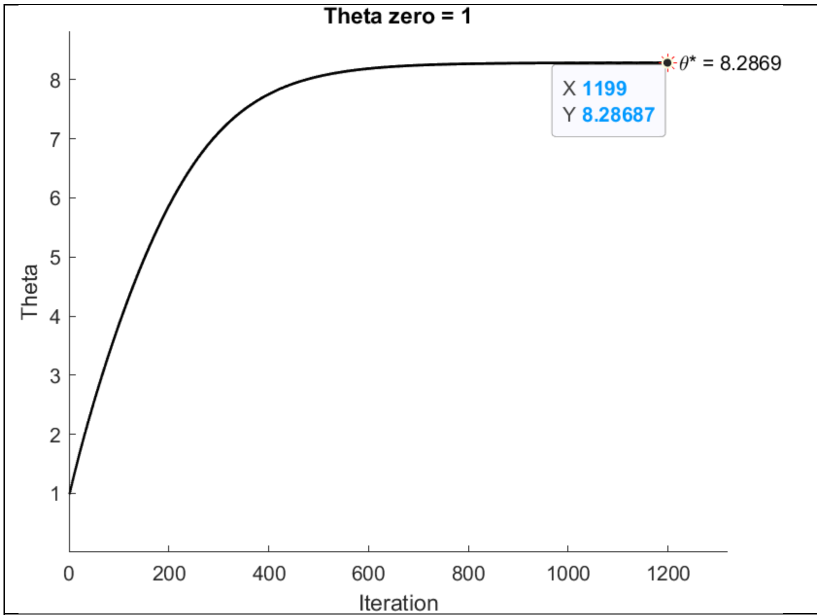
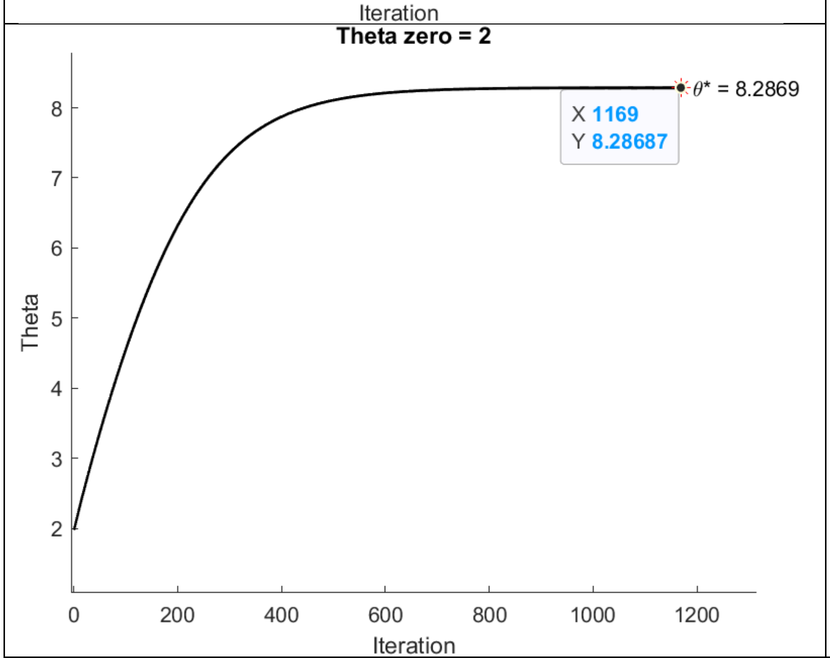
c). Code is at the end of this answer.

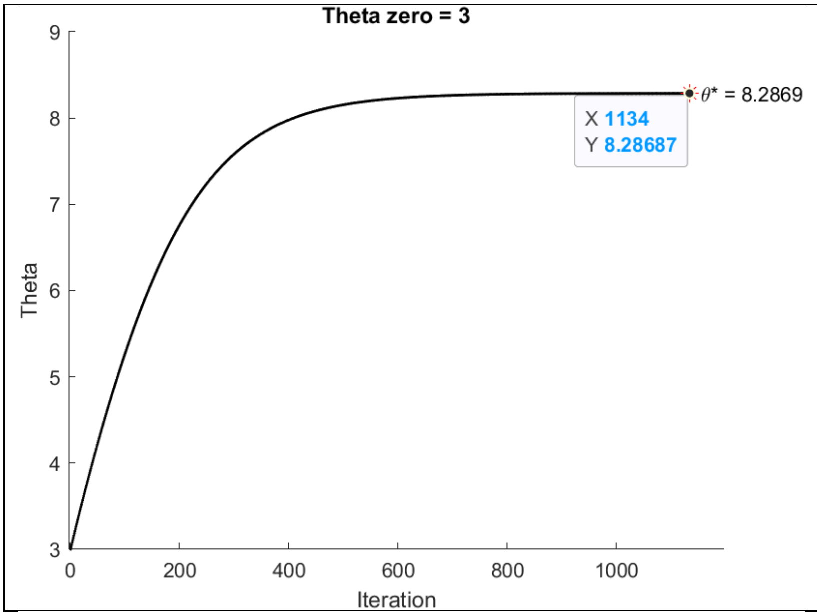
The distance on the shore between the surfer and a drowning person is d = 10. From this information we know, we conclude that the optimal crossing point to the water will be between 0 and 10. If we think further, it is obvious that 0 and 10 are also not optimal crossing points.

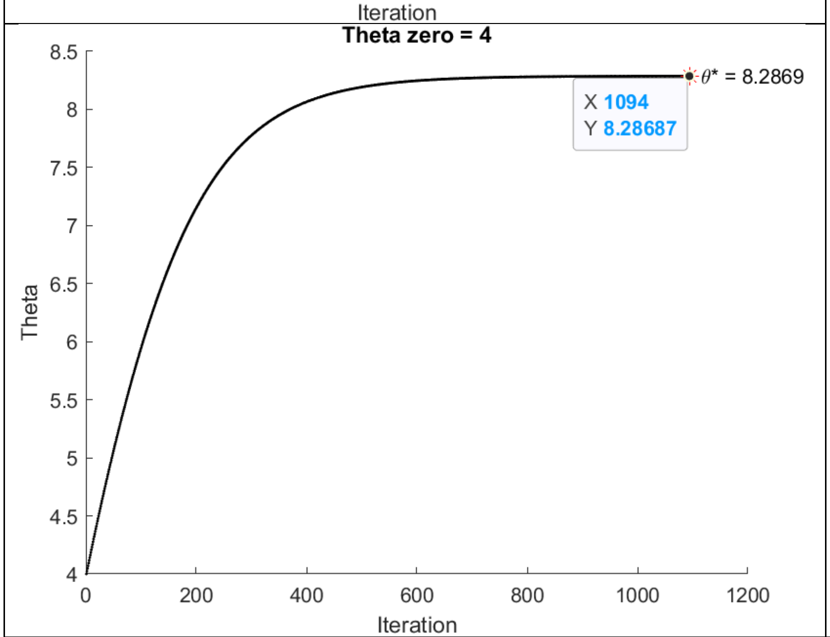
Hence, we tested crossing points form 1 till 9.

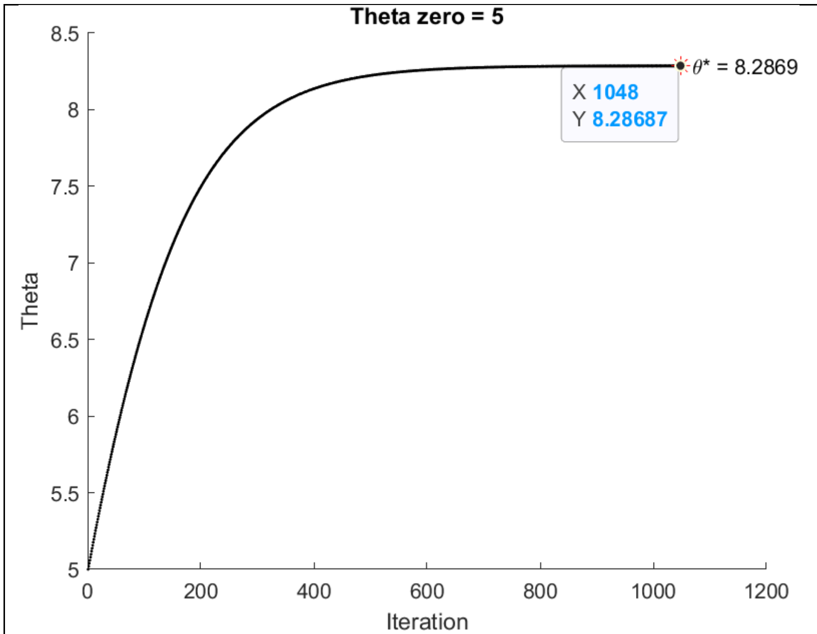
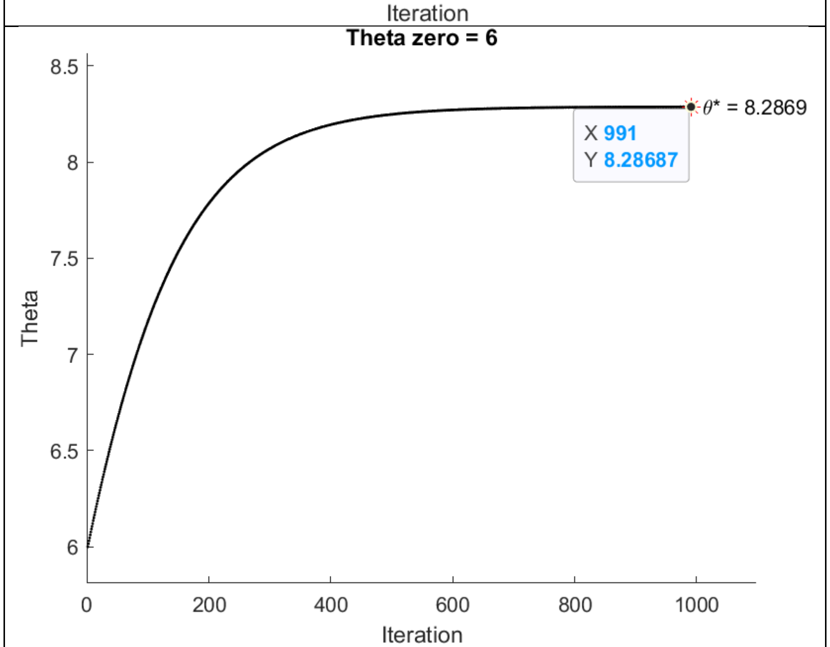
After performing iterations, the optimal point for crossing a shoreline for a surfer is 8.3.

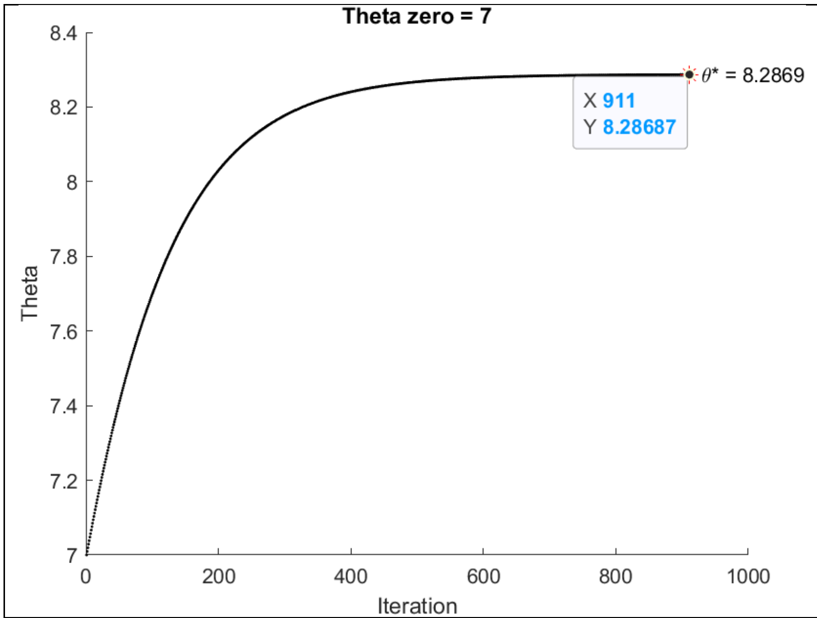
The plots:

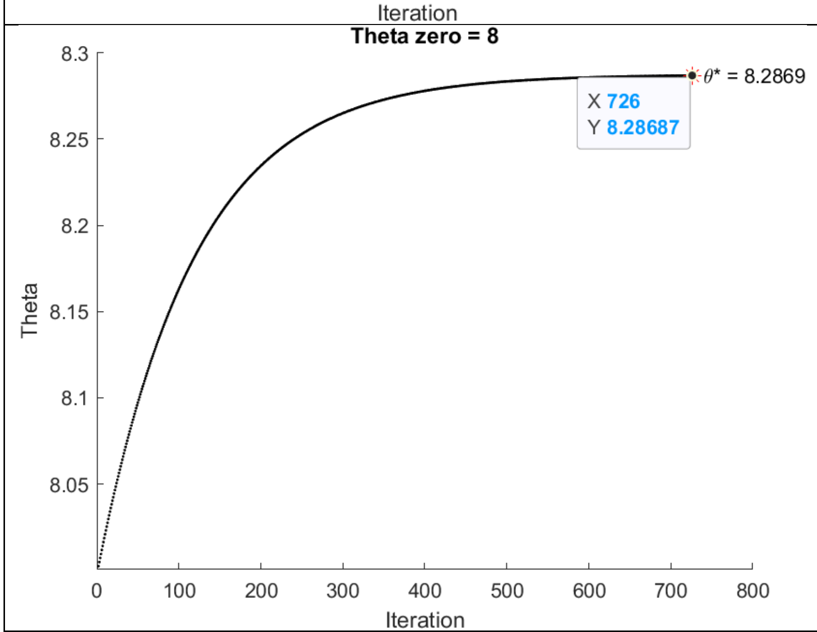
<table>
<tr>
<td>
</td>
<td rowspan="2">The algorithm converges to the optimal point 8.28 with around 1200 iterations for the starting point $\theta_0 = 1$. As the starting point is closer to the optimal point, it takes less iteration to achieve it. The graphs illustrate it clearly.</td>
</tr>
<tr>
<td>
</td>
</tr>
</table>

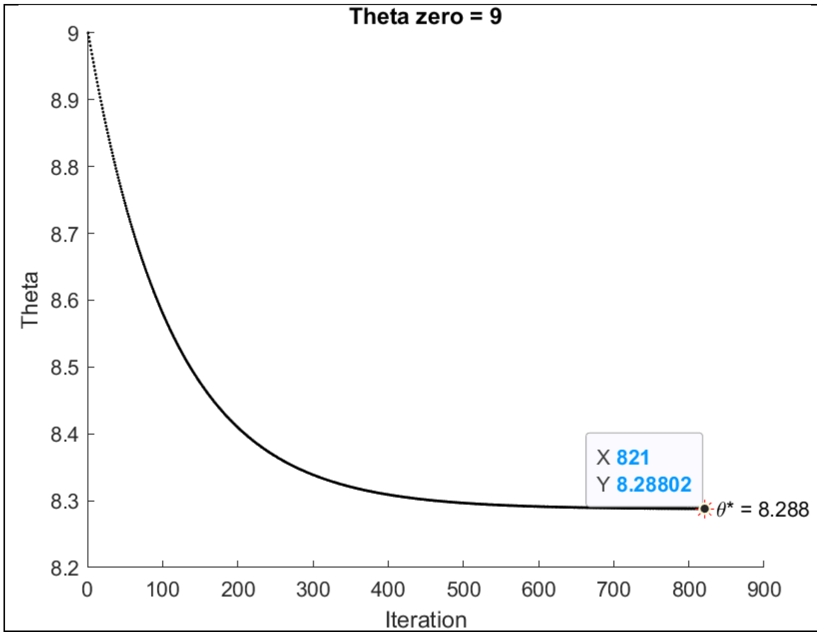

Theta zero = 3
Theta
Iteration
θ^* = 8.2869
X 1134
Y 8.28687


Theta zero = 4
Theta
Iteration
θ^* = 8.2869
X 1094
Y 8.28687

 Theta zero = 5 θ* = 8.2869 X 1048 Y 8.28687 Theta Iteration 8.5 8 7.5 7 6.5 6 5.5 5 0 200 400 600 800 1000 1200	
 Theta zero = 6 θ* = 8.2869 X 991 Y 8.28687 Theta Iteration 8.5 8 7.5 7 6.5 6 0 200 400 600 800 1000	


Theta zero = 7
θ^* = 8.2869
X 911
Y 8.28687
Theta
Iteration
8.4
8.2
8
7.8
7.6
7.4
7.2
7
0
200
400
600
800
1000

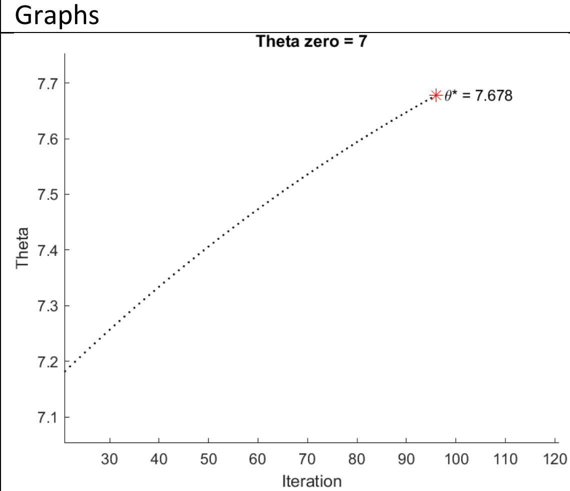

Theta zero = 8
θ^* = 8.2869
X 726
Y 8.28687
Theta
Iteration
8.3
8.25
8.2
8.15
8.1
8.05
0
100
200
300
400
500
600
700
800

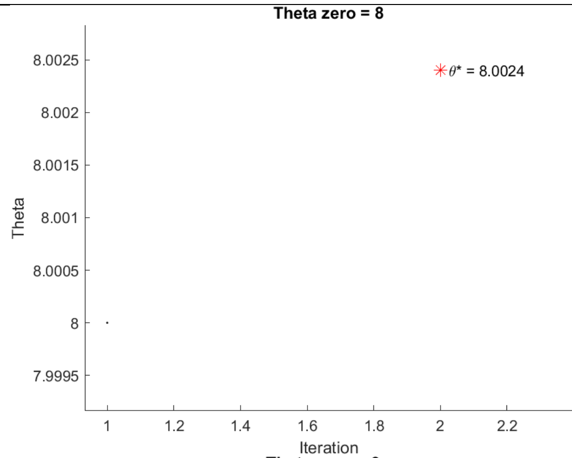


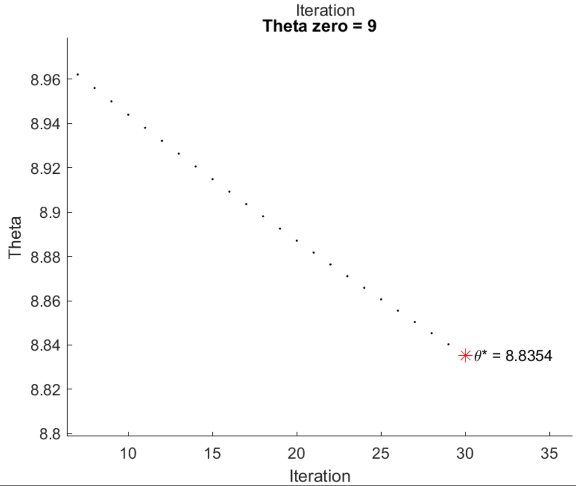
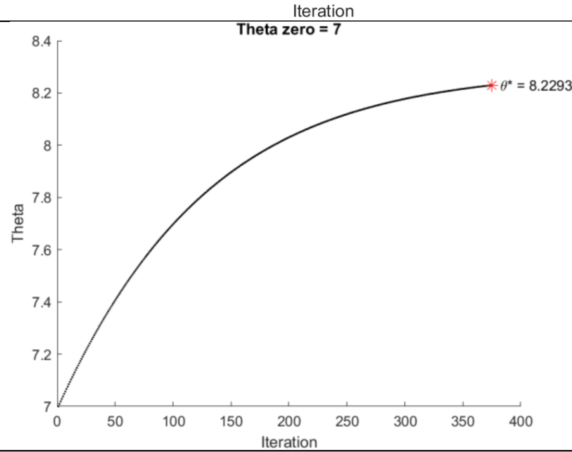
Stopping rule for iteration was chosen the next way:

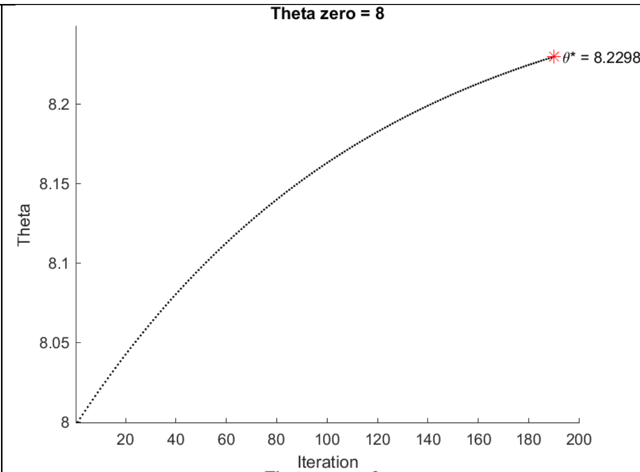
$\theta_{n+1} - \theta_n < 5 \cdot 10^{-6}$ then WHILE loop stops.

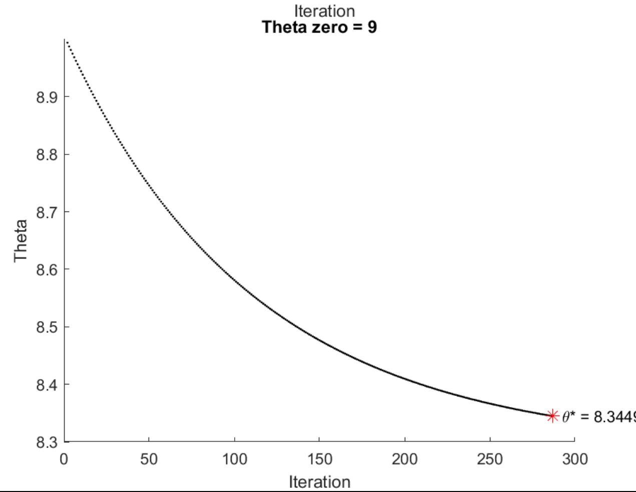
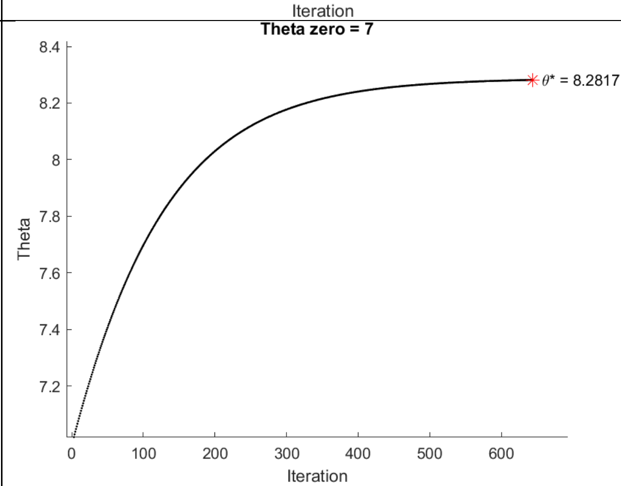
We tried a few different stopping criteria $\theta_{n+1} - \theta_n = residue$:

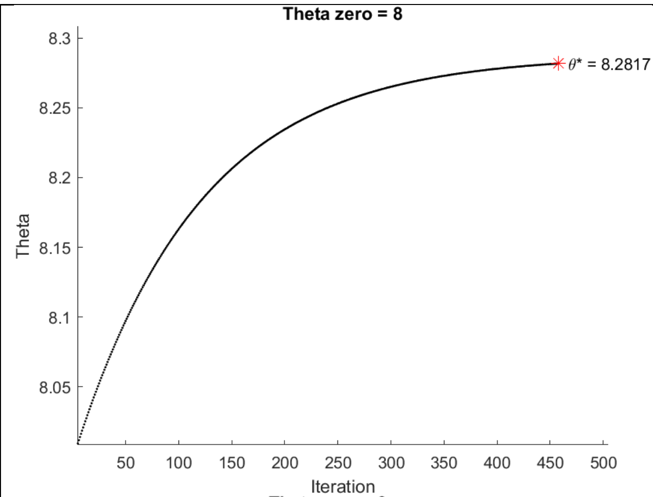
Stopping criteria	Graphs	Conclusions
$residue < 5 \cdot 10^{-3}$		It tooks less iterations, but it does not converge to the optimal point. If we below 8.3 the iteration does not reach it and stops at 7.678. if we start at 8, the algorithm will stop at second iteration on $\theta = 8.0024$. If we start above optimal point, it stops at 30 iterations with $\theta = 8.8354$.

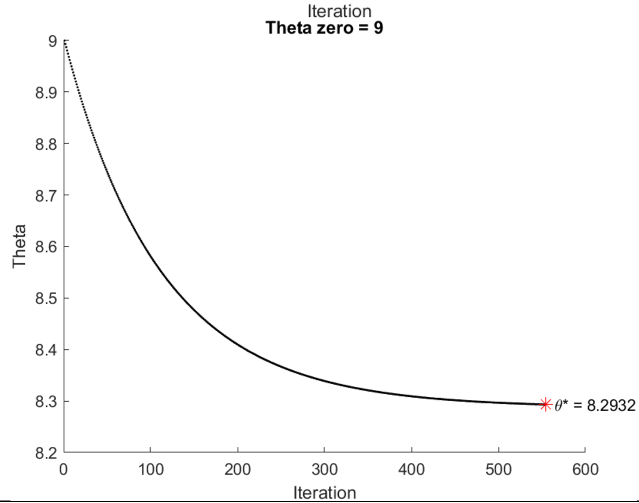
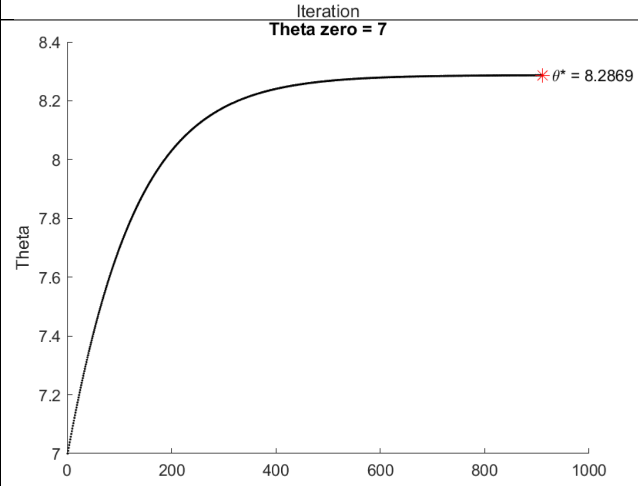
<table>
<tr>
<td></td>
<td>

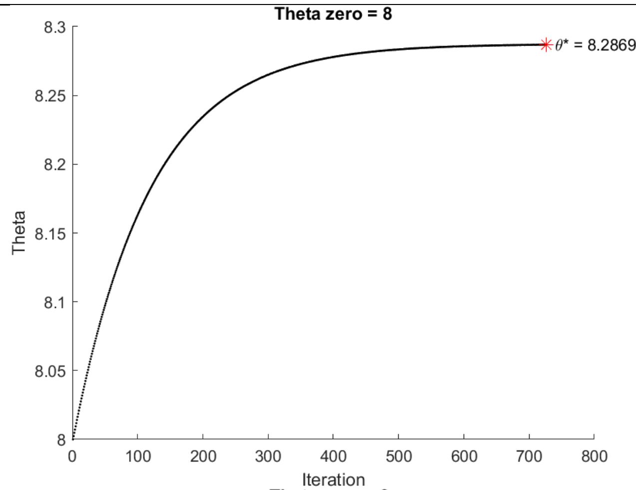

</td>
<td></td>
</tr>
<tr>
<td>$residue < 5 \cdot 10^{-4}$</td>
<td>
</td>
<td>We have the same situation but the points where the algorithm stops closer to optimal point.</td>
</tr>
</table>

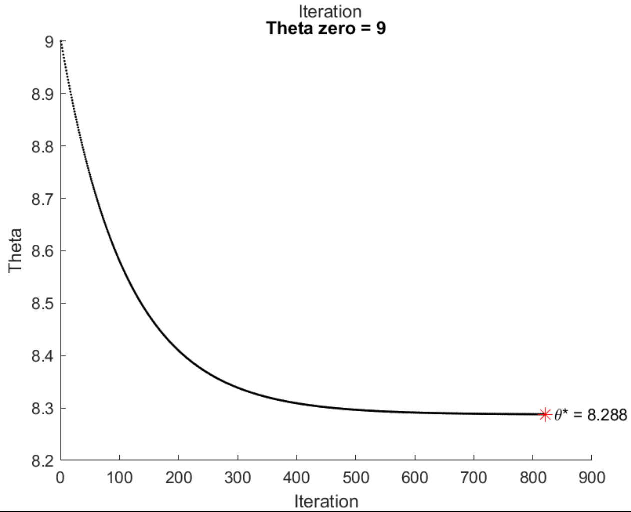
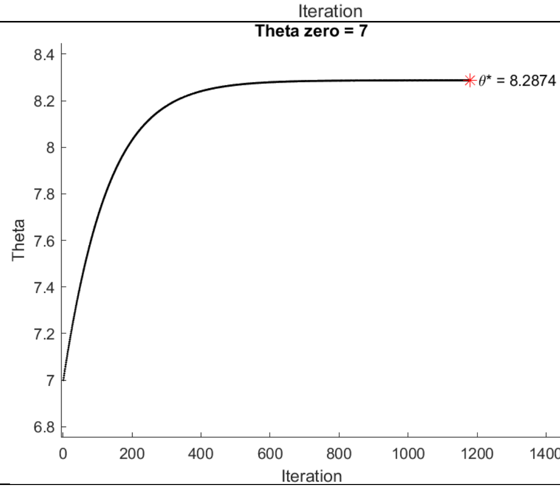
<table>
<tr>
<td></td>
<td>

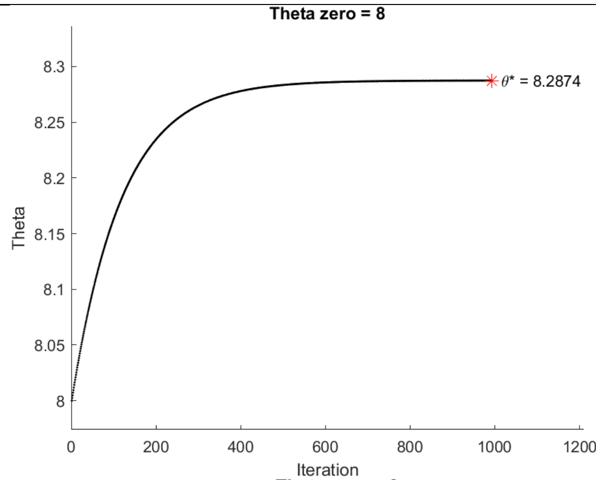

</td>
<td></td>
</tr>
<tr>
<td>$residue < 5 \cdot 10^{-5}$</td>
<td>
</td>
<td>The algorithm is closer to the optimal point. Thus, this choice will be enough for the problem. However, if we have a requirement of particular accuracy, we should reach the residue value when the algorithm reaches the same optimal point disregarding the beginning value.</td>
</tr>
</table>

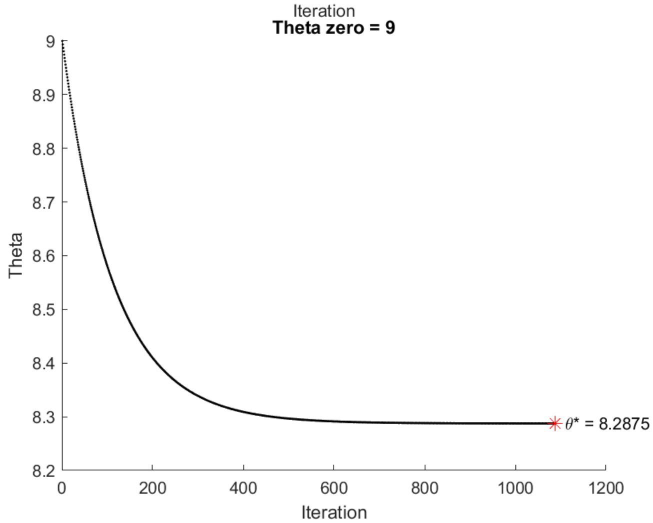
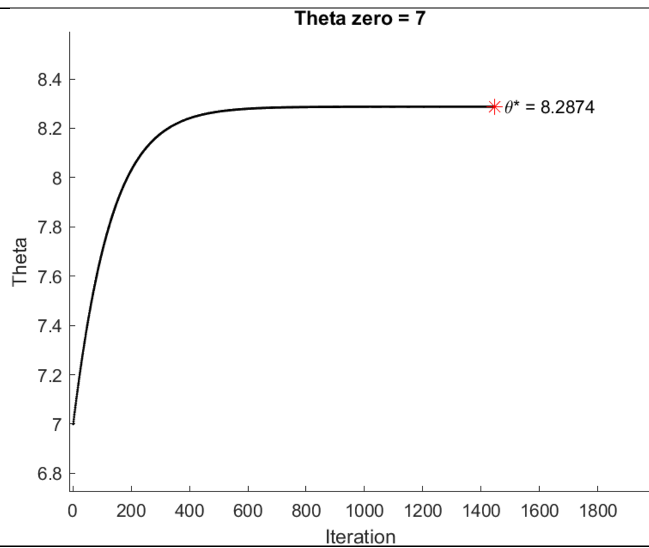
<table>
<tr>
<td></td>
<td>

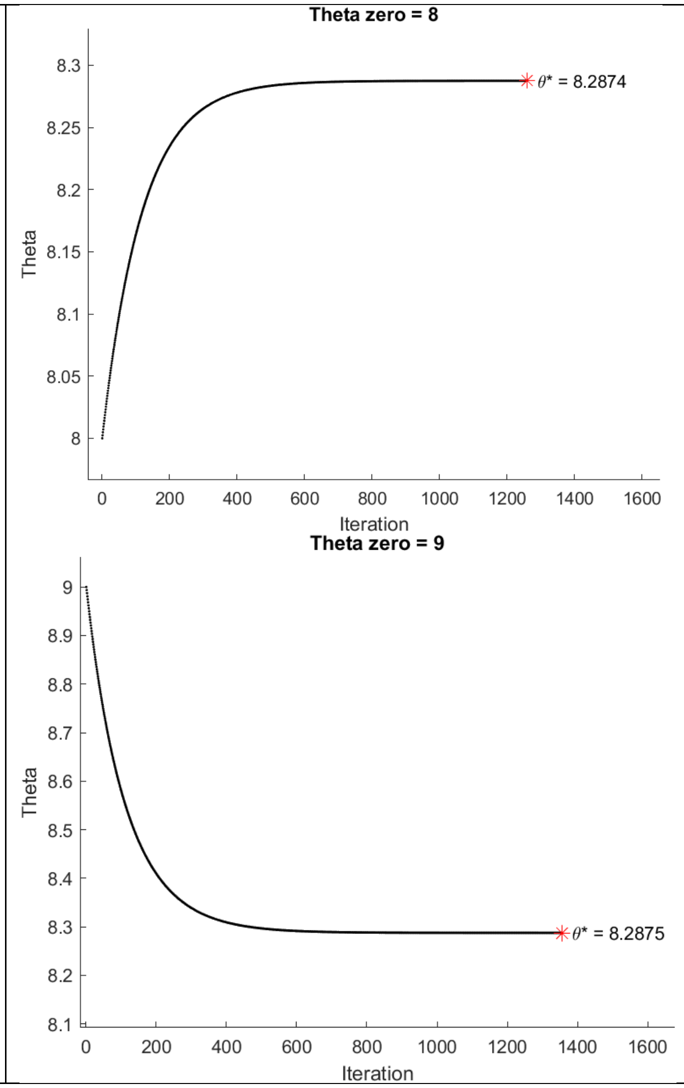

</td>
<td></td>
</tr>
<tr>
<td>$residue < 5 \cdot 10^{-6}$</td>
<td>
</td>
<td>This one choice, and more bellow, illustrate that these stopping criteria are more than enough for our problem.</td>
</tr>
</table>

<table>
<tr>
<td></td>
<td>


</td>
<td></td>
</tr>
<tr>
<td>$residue < 5 \cdot 10^{-7}$</td>
<td>
</td>
<td></td>
</tr>
</table>

<table>
<tr><td></td><td>


</td><td></td></tr>
<tr><td>$residue < 5 \cdot 10^{-8}$</td><td>
</td><td></td></tr>
</table>




```
%% Exersice 2.6 Matlab R2021a
%parameters
clear theta;
clear iter;
a = 2; b = 5; d = 10; v1 = 3; v2 = 1;
epsilon = 0.05;
% optimal point will be between 0 and 10.
% so we will test starting point i.e. theta zero from the 0 till 10
% loop for the different theta zero
for n = 1 : 1 : 9
    k = 1;
    %clear theta; clear iter;
    theta = []; iter = [];
    theta0 = n;
    theta(1) = theta0;
    ress = 10;
    while abs(ress)> 5e-6
        ress = epsilon * (1/v1 * (theta(k)/sqrt(theta(k)^2 + a^2)) - 1/v2 * ((d -
theta(k))/sqrt((d - theta(k))^2 + b^2)));
        theta (k+1) = theta (k) - epsilon * (1/v1 * (theta(k)/sqrt(theta(k)^2 + a^2)) - 1/v2 *
((d - theta(k))/sqrt((d - theta(k))^2 + b^2)));
```

```
        iter(k) = k;
        k = k + 1;
    end
    iter(end+1)= k;
    % plots of a value theta on each iteration
    figure(n); clf
    x  = iter;
    y = theta;
    x_end = iter(end);
    y_end = theta(end);
    scatter(x, y, 2,'k','filled'), xlabel('Iteration'), ylabel('Theta'), title(['Theta zero = ',
num2str(n)])
    hold on
    scatter (x_end, y_end, 80,'r','*'),  text(x_end ,y_end,['  \theta* = ' num2str(y_end)])
    hold off
end
```

Exercise 2.10. Let $J(\theta)$ be the temperature in the lecture room and denote by θ how far to the right the heat-regulator is turned. For adjustment θ of the regulator let $J(\theta)$ denote the resulting temperature in the room. For the following you may assume that $J(\theta)$ is a smooth function. Suppose you want to reach temperature α.

Consider the target vector field

$$G(\theta) = \alpha - J(\theta) \tag{2.37}$$

and argue that

(a) the problem

$$\min_{\theta} \frac{1}{2}(\alpha - J(\theta))^2 \tag{2.38}$$

is well posed,

(b) $G(\theta)$ in (2.37) is coercive for (2.38)

(c) the algorithm

$$\theta_{n+1} = \theta_n + \frac{1}{n+1}(\alpha - J(\theta_n))$$

converges for n towards infinity to the solution θ^* of (2.38). Mention explicitly the theorems that you use and the reason why the corresponding assumptions are satisfied.

a) Argue that $\min_{\theta}\left(G(\theta)^2\right)$ is well posed

(Ignoring ½ since it is a constant and does not change anything mathematically or in the proofs, squared is important because it creates a global min when the two values are closest rather than where they are furthest apart. We're trying to minimize the difference rather than minimize the value)

The solution set is not empty. Since we are talking about a physical dial, we can assume the dial is continuous and strictly monotonic. This means every point on the dial is unique, so each point on the dial references a different temperature, and each of those temperatures exist. Therefor the set of solutions is not empty. There is either one unique point on the dial where $G(\theta)$ tends to zero, and we approach the solution, the desired temperature, or the desired temperature α is not available on the dial and not reachable so as we reach the highest or lowest possible α, whichever is closest to $J(\theta)$, $G(\theta)$ is minimized as much as possible, and that value is the solution. either $J(\theta)$ goes through α or it doesn't and if it doesn't, the closest possible value is the solution.

The set of solutions for $\min_{\theta}\left(G(\theta)^2\right)$ contains only KKT points. KKT points are stationary points that satisfy their given constraints as seen in the formulas below.

$$\nabla_\theta \mathcal{L}(\theta^*, \lambda^*, \eta^*) = 0$$
$$\nabla_\lambda \mathcal{L}(\theta^*, \lambda^*, \eta^*) = g(\theta^*)^\top \leq 0, \lambda^* \geq 0, \text{ and } \forall i : \lambda_i^* g_i(\theta^*) = 0$$
$$\nabla_\eta \mathcal{L}(\theta^*, \lambda^*, \eta^*) = h(\theta^*)^\top = 0;$$

The set of solutions contains one solution which is a global minimum since it is $\min_{\theta}\left(G(\theta)^2\right)$ so it is a stationary point. The constraints here is $J(\theta) \leq$ the highest temperature on the dial or $J(\theta) \geq$ the lowest temperature on the dial. Since we are bound by a physical dial, all possible $J(\theta)$ will fall within the range of those constraints.

The set of KKT points are confined to a bounded set since θ is bound, so $J(\theta)$ is bound. The dial is bound to the range of the thermostat and can only go so far in each direction. Thermostats will max out above boiling and floor at below freezing.

There is no direction that is strictly decreasing an infinite number of times because on a physical dial you are unable to keep moving one degree hotter or colder forever, eventually you will reach the limit since θ is bound. Once you reach the global min or max value on the dial, you either need to turn around or stay where you are, you cannot keep going.

b) argue that $G(\theta)$ is coercive for $\min_{\theta}\left(G(\theta)^2\right)$

Conditions for coercive:

- *The vector field G is either Lipschitz continuous or bounded,*

α is a constant so it is bound and θ is bound to the physical dial so $J(\theta)$ is bound to the possible temperatures on the dial. Since all components of $G(\theta)$ are bound, $G(\theta)$ must be bound.

- **[Uniqueness of solution of ODE]** *G is locally Lipschitz and the solution $\{x(t) : t \geq 0\}$ to the initial value problem* $$\frac{dx(t)}{dt} = G(x(t)), \quad t \geq 0, \quad \text{with } x(0) = \theta_0,$$ *does not blow up in finite time,*

$G(\theta)$ is locally Lipschitz because within any local range of θ, say Δ is the width of that range, $G(\theta)/\Delta$ cannot vary more than the total possible range of $G(\theta)$. Additionally, $J(\theta)$ doesn't blow up in finite time because $J(\theta)$ can never be above boiling or below freezing, our options are limited. Something bounds will not blow up.

- *all asymptotically stable points of the ODE (2.27) are KKT points of the NLP.*

Ascensiontide does not apply here since θ is bounded so we never tend towards infinity.

c) Argue that $$\theta_{n+1} = \theta_n + \frac{1}{n+1}(\alpha - J(\theta_n))$$ converges for n towards infinity to the solution θ^*

We start at some given θ and keep moving towards the gradient $G(\theta) = (\alpha - J(\theta))$, this value keeps getting smaller and smaller and we get closer and closer to the solution θ^*. We are moving in the direction that shrinks the change. when α is bigger than $J(\theta)$, were moving towards a hotter temperature and when α is smaller than $J(\theta)$, we are moving towards a colder temperature. Since we are moving towards the global minimum of $G(\theta)$, $J(\theta)$ is getting closer to α so $(\alpha - J(\theta))$ keeps decreasing and since n is increasing with each iteration $\frac{1}{n+1}$ is also decreasing, so θ n+1 keeps slowly decreasing towards θ^*. If α is reachable, if it is an available temperature on the dial, we get infinitely closer to it and the algorithm converges as n tends towards infinity. If α is not available on the thermostat, then we stop once we reach the min or max $J(\theta)$, the highest or lowest option on the dial.

The theorem used to solve this problem is Theorem 2.8 which states that:

given the conditions $\sum_n \varepsilon_n = \infty$, $\varepsilon_n \to 0$ and $\sum_n \epsilon_n \|\beta_n(\theta_n)\| < \infty.$

given the algorithm $\theta_{n+1} = \theta_n + \epsilon_n (G(\theta_n) + \beta_n(\theta_n)).$ with limit point if it exists $\theta^* = \lim_{n\to\infty} \theta_n.$

If G is coercive for the well-posed NLP, then θ^* is a KKT point of the NLP.

Assumption $\sum_n \varepsilon_n = \infty$ is satisfied because although step size ε $\sum_n \epsilon_n \|\beta_n(\theta_n)\| < \infty.$ = $\frac{1}{n+1}$ is decreasing, infinitely adding decreasing values will still sum to infinity, $\varepsilon_n \to 0$ is satisfied because as n increases with each iteration, $\frac{1}{n+1}$ decreases so step size is decreasing. and because our bias is zero so 0* step size ε = 0 which is finite. As explained in section b of this exercise, $G(\theta)$ is coercive and as explained in section a of this exercise $\min_\theta \left(G(\theta)^2\right)$ is well posed so the solution θ^* is a KKT point of this NLP.

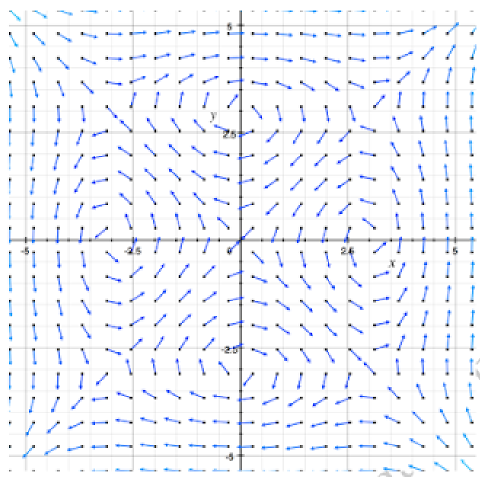


Figure 2.12: Gradient-field of $J(\theta)$ for Exercise 2.11.

EXERCISE 2.11. The gradient-field of a function $J(\theta)$ is shown in Figure 2.12.

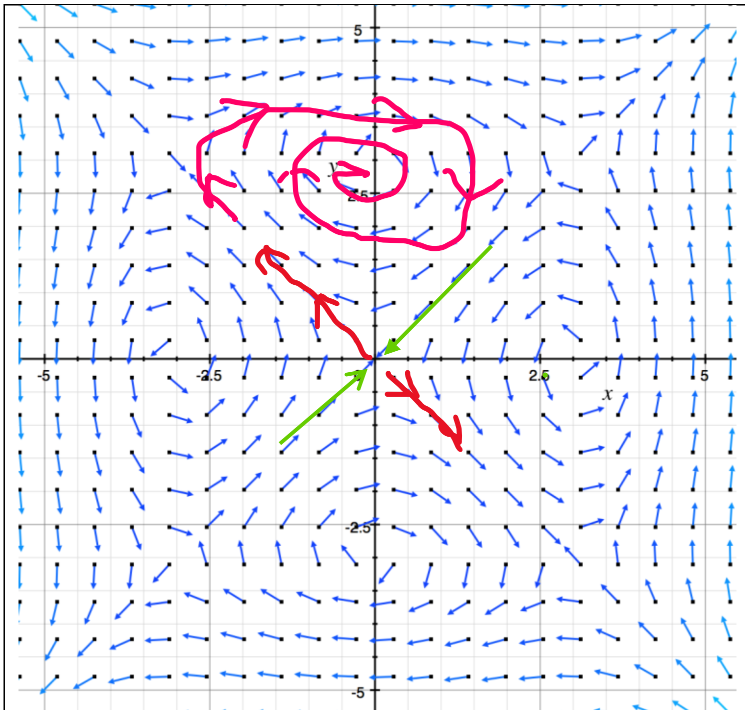
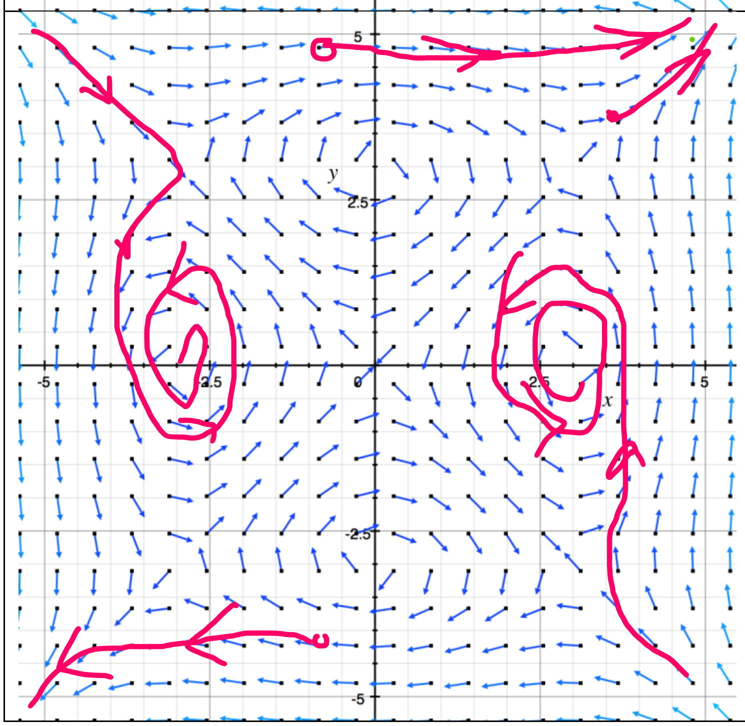
(a) Discuss with Figure 2.12 for the ODE

$$\frac{d}{dt}x(t) = -\nabla J(x(t))$$

the nature of point $(0,0)$ (stable, asymptotically stable, globally stable, or unstable).

(b) Judging from the figure, is this problem well-posed and is the vector-field coercive?

a) Point (0,0) from the Figure 2.12 is unstable. Here is why:

	(0,0) is not stable. If we start in the region box from -2.5 to 2.5 in both x and y axis, there are two trajectories that cross the origin. One of them is moving towards it, showing us that it is, it is or it could be could be stable point. Another one goes out of origin, shows us that in this trajectory it is unstable point. So, judging from picture we conclude that the (0,0) is saddle point. Meanwhile if we start from the other trajectories in this region we will ended up in the points (0,3), (0, -3), (-3,0), (3,0).
	(0,0) is not asymptotically stable. Asymptotically stable point because if we start from the infinity, or in our case form the outer box from -5 to 5 in both x and y axis, then we could not reach the origin. It is not globally asymptotically stable, because first of all it should be stable. But in this vector field it is not stable. **Conclusion, the point (0,0) is unstable**

b) The problem is not well-posed. To be well posed NLP, it should satisfy:

1. the set of solutions is not empty, contains only KKT points.

Our set of solution contains 4 KKT points (0,3), (0, -3), (-3,0), (3,0), one saddle point (0,0) in the box region from -2.5 to 2.5 in both x and y axis. Plus, we have two improper minimum $-\infty, +\infty$, when we start in the region of the outer box region from -5 to 5 in both x and y axis.

Our problem is ill-posed.

Vector field is not coercive:

1. Not all the points of the vector field are KKT.
2. If we look at the function on marginal as t -> inf then in all directions the vector field does not converge to a fixed point. This means, if you start inside you stay inside of the box region from -2.5 to 2.5 on both the x and y axis forever. If you stay outside, you stay in the outer box region from -5 to 5 on both the x and y axis ever and blow up in real-time.

If we look at the function on marginal as t -> inf then in all directions the vector field is not converge to a fixed point.