Optimisation CW

1. i) let $f(x_1, x_2) = |x_1 + x_2^2|$

$$f(x_1, \alpha x_1) = |x_1 + \alpha^2 x_1^2| = |x_1|\,|1 + \alpha^2 x_1| \geq \alpha^2 |x_1|^2 \rightarrow \infty \text{ as } |x_1| \rightarrow \infty$$

~~$= |1+\alpha^2|\,|x_1^2| = (1+\alpha^2)\,|x_1^2| = (1+\alpha^2)\,|x_1|^2 \rightarrow \infty$ as $|x_1| \rightarrow \infty$~~

also, $f(\alpha x_2, x_2) = |\alpha x_2 + x_2^2|$

~~$= (\alpha+1)|x_2|$~~

~~$= \alpha|x_2|\,|1 + x_2|$~~

$= |x_2|\,|\alpha + x_2|$ and $|x_2| \rightarrow \infty$

$\Rightarrow |\alpha + x_2| \rightarrow \infty$

so $f(\alpha x_2, x_2) \rightarrow \infty$.

f is coercive if $\lim_{\|x\| \rightarrow \infty} f(x) = \infty$.

take ~~the~~ for arbitrary $b \in \mathbb{R}$, $x = (-b^2, b)$

$$\lim_{b \rightarrow \infty} \|x\| = \lim_{b \rightarrow \infty} \sqrt{b^4 + b^2} \rightarrow \infty$$

however, $\lim_{b \rightarrow \infty} f(x) = \lim_{b \rightarrow \infty} |-b^2 + b^2| = \lim_{b \rightarrow \infty} 0 = 0$ so

f is not coercive.

ii) $f(x_1, x_2) = 4x_1^4 + x_2^2 - 4x_1^2 x_2 + 4$

$$= (2x_1^2 - x_2)^2 + 4 \geq 4$$

so f is bounded below by 4, and this global minimum is attained when $x_2 = 2x_1^2$ or $(t, 2b^2)\ \forall b \in \mathbb{R}$.

All these points are hence also local minima.

They are non-strict globally as not the unique minima.

~~$\nabla f = (-8x(2x^2 - y), -4x^2 + 2y)$~~

~~$\nabla^2 f = \begin{pmatrix} 48x^2 - 8y & -8x \end{pmatrix}$~~

~~$\nabla f = (-8x_1(2x_1^2 - x_2), -4x_1^2 + 2x_2)$~~

~~$\nabla^2 f = \begin{pmatrix} 48x_1^2 - 8x_2 & -8x_1 \\ -8x_1 & 2 \end{pmatrix}$~~

~~$$\nabla^2 f(x_1, 2x_1^2) = \begin{pmatrix} 48x_1^2 - 16x_1^2 & -8x_1 \\ -8x_1 & 2 \end{pmatrix}$$~~

~~$$= \begin{pmatrix} 32x_1^2 & -8x_1 \\ -8x_1 & 2 \end{pmatrix}$$~~

~~$\det(\nabla^2 f(x_1, x_2)) =$ ~~2~~ $64x_1^2 - 64x_1^2$~~

~~$= 0$~~

~~$\mathrm{tr}(\nabla^2 f) = 32x_1^2 + 2 > 0$~~

~~$\Rightarrow \nabla^2 f$ is ~~pos. s~~ positive semidefinite.~~

~~along $x_2 = 2x_1^2$~~

~~If $\nabla^2 f(x^*) > 0 \Rightarrow x^*$~~

f is also non-strict locally as for any ~~local~~ point x^* along $x_2 = 2x_1^2$, $(x_1, 2x_1^2)$, $(x_1 + \varepsilon, 2(x_1+\varepsilon)^2) = y^*$ where $y^* \in B(x^*, r)$, $x^* \neq y^*$ but $f(x^*) \not< f(y^*)$ as $f(x^*) = f(y^*) = 4$.

~~let $r \geq \sqrt{\varepsilon}$~~

Take the next point along the curve $x_2 = 2x_1^2$ and the value is still 4.

~~Finally, we conclude~~ Stationary points are at:

$\nabla f = (-8x_1(2x_1^2 - x_2), -4x_1^2 + 2x_2) = 0$

$\Rightarrow x_1 = 0$ ~~or~~ and ~~or~~ $x_2 = 2x_1^2$ or $x_2 = 2x_1^2$

$\Rightarrow x_2 = 2x_1^2$. All stationary points are along this line.

We have seen they are global minima that are locally and globally non-strict.

2. i) Using the recursion,

$$x_0 = \bar{x}$$

$$x_1 = a\bar{x} + du_1$$

$$x_2 = \cancel{a\bar{x}} \; a(a\bar{x} + du_1) + du_2$$

$$= a^2\bar{x} + adu_1 + du_2$$

$$\vdots$$

$$x^N = a^N\bar{x} + a^{N-1}du_1 + \dots + du_N$$

$$\underline{x} = \begin{pmatrix} x_0 \\ \vdots \\ x^N \end{pmatrix} = \underbrace{\begin{pmatrix} \bar{x} \\ a\bar{x} \\ a^2\bar{x} \\ \vdots \\ a^N\bar{x} \end{pmatrix}}_{-\underline{b}} + \underbrace{\begin{array}{c} \scriptstyle N \\ \begin{array}{c} \scriptstyle N+1 \\ \updownarrow \end{array} \begin{pmatrix} 0 & \dots & \dots & \dots & 0 \\ d & 0 & \dots & \dots & 0 \\ ad & d & 0 & \dots & 0 \\ a^2d & ad & d & 0 \dots & 0 \\ \vdots & & & & \\ a^{N-1}d & a^{N-2}d & \dots & & ad\,d \end{pmatrix} \end{array}}_{S} \underbrace{\begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}}_{\underline{u}}$$

$$\underline{x} = S\underline{u} - \underline{b} \qquad \text{where } \underline{b} = \begin{pmatrix} -\bar{x} \\ -a\bar{x} \\ \vdots \\ -a^N\bar{x} \end{pmatrix}$$

So $\min_{\underline{u} \in \mathbb{R}^N} \|\underline{x}\|_2^2 + \frac{\gamma}{2}\|\underline{u}\|_2^2, \; \gamma > 0$

$= \min_{\underline{u} \in \mathbb{R}^N} \|S\underline{u} - \underline{b}\|_2^2 + \frac{\gamma}{2}\|\underline{u}\|_2^2, \; \gamma > 0$

Let $\frac{\gamma}{2} = \lambda$, $R(\underline{u}) = \|\underline{u}\|_2^2$, then we have the RLS form $\min_{u} \|S\underline{u} - \underline{b}\|^2 + \lambda R(\underline{u})$, $R(\underline{u}) = \|D\underline{u}\|_2^2$ with $D = I_N$

The optimal solution is:

$$u^* = \underline{u}_{RLS} = (S^TS + \lambda I^T I)^{-1} S^T\underline{b}$$

$$= (S^TS + \tfrac{\gamma}{2}I)^{-1}S^T\underline{b}$$

$S^TS + \frac{\gamma}{2}I$ is invertible if $\text{Null}(S) \cap \text{Null}(I) = \{0\}$

but $\text{Null}(I) = \{0\} \Rightarrow \text{Null}(S) \cap \text{Null}(I) \subseteq \text{Null}(I) = \{0\}$

hence the solution u^* exists and is unique.

$$x_{\bar{x}}^{u^*} = S u^* - \underline{b}$$

If $\underline{u}$ solves unregularised LS, $\|x_{\bar{x}}^{\underline{u}}\|^2 \le \|x_{\bar{x}}^{u^*}\|^2$

Assume for contradiction, ~~$\|\underline{u}^*\| \le \|\underline{u}\|$~~, $\|\underline{u}^*\| > \|\underline{u}\|$,

then $\|x_{\bar{x}}^{\underline{u}}\|^2 + \frac{\gamma}{2}\|\underline{u}\|^2 \le \|x_{\bar{x}}^{u^*}\|^2 + \frac{\gamma}{2}\|\underline{u}^*\|^2$

which contradicts that $\underline{u}^*$ is optimal for RLS #

Hence, $\|\underline{u}^*\| \le \|\underline{u}\|$

ii/ In the control, ~~the~~ increasing gamma decreases the ~~initial~~ value of u_1 ~~while~~ and the curve slopes down ~~to~~ ~~the~~ less ~~steep~~ steeply to 0 at u_N for larger values of γ. $u_1 > 35$ when $\gamma = 0.001$ but $u_1 < 1$ when $\gamma = 1$. For the trajectory, all ~~plots~~ values of γ have ~~the~~ x_0 at 1, ~~x_N at 0~~ but ~~while~~ ~~larger~~ smaller values of γ slope down faster as i increases towards 0. ~~I~~ large γ, such as $\gamma = 1$ doesn't even slope down to 0 at ~~the~~ x_N, much less ~~s~~ steep. Increasing regularisation pushes $\underline{u}$ closer to 0, decreasing its cost in the cost function.

iii/ $\min_{\underline{u}\in\mathbb{R}^N} \|x_N^{\underline{u}}\|_2^2 + \frac{\gamma}{2}\|\underline{u}\|_2^2 - \delta\sum_{i=1}^{N}\log(u_{max} - u_i)$, $\gamma, \delta > 0$

Let $f(\underline{u}) = \|S\underline{u} - \underline{b}\|_2^2 + \frac{\gamma}{2}\|\underline{u}\|_2^2 - \delta\sum_{i=1}^{N}\log(u_{max} - u_i)$

$$2S^TS\underline{u} - 2S^T\underline{b} + \gamma\underline{u} + \delta\left(\frac{1}{u_{max}-u_1}, \dots \frac{1}{u_{max}-u_N}\right)^T$$

$$\nabla f(\underline{u}) = \cancel{2(S\underline{u} - \underline{b})S + \gamma\underline{u} + \delta\sum_{i=1}^{N}\frac{1}{u_{max}-u_i}}$$

Use Armijo backtracking.

for convergence

Let $\varepsilon = 0.01$ in the while loop, $\alpha = 0.5$, $\beta = 0.5$,

$t^k = s = 1$ initial stepsize, $d^k = -\nabla f(u^k)$

initial $\underline{u} = (1, \dots, 1)$ $\quad t^k = \beta t^k$

For the control, the ~~cur~~ front end for small i is flattened as the u_i are kept $\leq u_{max} = 8$, in comparison to the problem without the constraint but the rest of the curve ~~is~~ is similar in shape. This results in the curve for the state being slightly above the unconstrained curve for all u_i, though the curves meet at u_0 and u_N. ~~(This is because the cost function~~

w/

$$\min_{u \in \mathbb{R}^N} \|x_\pi^u\|_2^2 + \frac{\gamma_2}{2}\|u\|_2^2 + \cancel{\gamma_1\|u\|_1}\ \gamma_1 \sum_{i=1}^N L_\varepsilon(u_i), \quad \gamma_1, \gamma_2, \varepsilon > 0.$$

$$L_\varepsilon(u_i) = \begin{cases} \frac{1}{2}u_i^2 & \text{if } |u_i| \le \varepsilon \\ \varepsilon(|u_i| - \frac{1}{2}\varepsilon) & \text{otherwise.} \end{cases}$$

$$\frac{\partial}{\partial u} L_\varepsilon(u_i) = \begin{cases} u_i & \text{if } |u_i| \le \varepsilon \\ |\varepsilon| & \text{if } u_i > |\varepsilon| \end{cases}$$

Sketching $L_\varepsilon(u_i)$, we see it has similar shape to a $|x|$ graph, i.e. approximates $\|u\|_1 = |u_1| + \dots + |u_N|$, however is smooth (curved) at 0 hence is differentiable everywhere and has continuous derivative.* Hence, it preserves the linearity of $\|u\|_1$ for $u_i > |\varepsilon|$ and it is the linearity that allows to place emphasis on minimising u_i near 0 (i.e. promoting sparsity).

Also, for $|u_i| \le \varepsilon$, the $L_\varepsilon(u_i)$ function is similar to an L_2 norm behaviour. but behaves like L_1 norm outside.

Settings: initial $u = (1, \dots, 1)$, $\alpha = \beta = 0.5$, initial stepsize $= 1$
epsilon for convergence $= 0.001$, $d^k = -\nabla f(u^k)$

case i) with the L_1 norm only ~~...~~ compared to the L_2 norm only case reweights the ~~initial (large~~ control for small i (large values) much higher and the other controls much closer to 0. ~~This then~~ Hence, the curve goes sharply down and then suddenly flattens, sparsity is promoted. For the state, all values remain in the (0,1) bound but the curve drops sharply then flattens suddenly, and isn't smooth.

Proof $L_\varepsilon(u_i)$ is differentiable.

$$\lim_{h\to 0, h>0} \frac{L_\varepsilon(\varepsilon + h) - L(\varepsilon)}{h} = \frac{\varepsilon(\varepsilon + h - \frac{1}{2}\varepsilon) - \frac{1}{2}\varepsilon^2}{h}$$

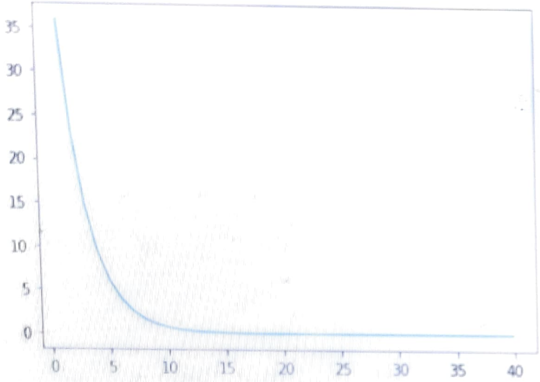
$$= \frac{\varepsilon h}{h} = \varepsilon \text{ equal to the limit as } h < 0.$$

so differentiable at $u_i = \varepsilon$.

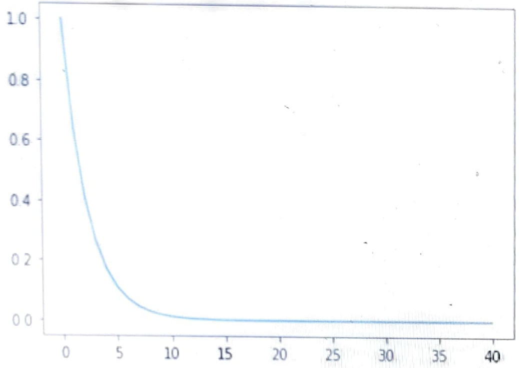
Similar for $u_i = -\varepsilon$, $\lim_{h\to 0, h<0} \frac{L(-\varepsilon + h) - L(-\varepsilon)}{h} = -\varepsilon$ equal to other direction.

~~At~~ Continuous at all other points so $L_\varepsilon(u_i)$ is differentiable.

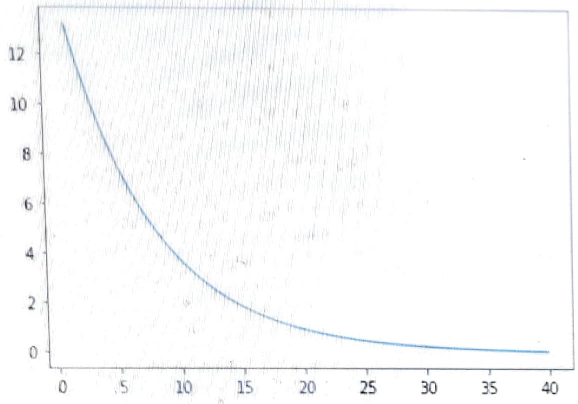
2.ii Control with gamma = 0.001



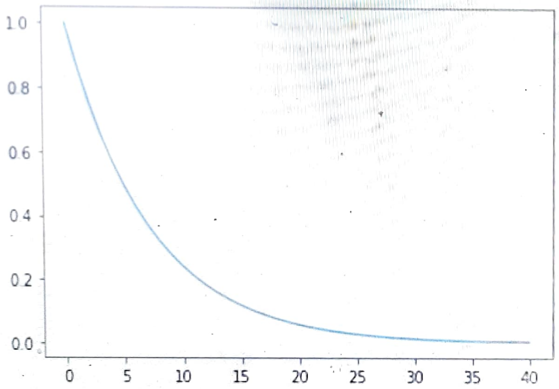
trajectory with gamma = 0.001



Control with gamma = 0.01



trajectory with gamma = 0.01

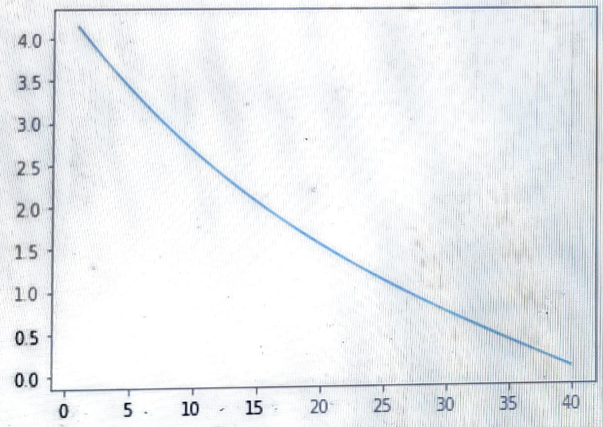


Text Predictions: On Accessibility: Investigate Focus

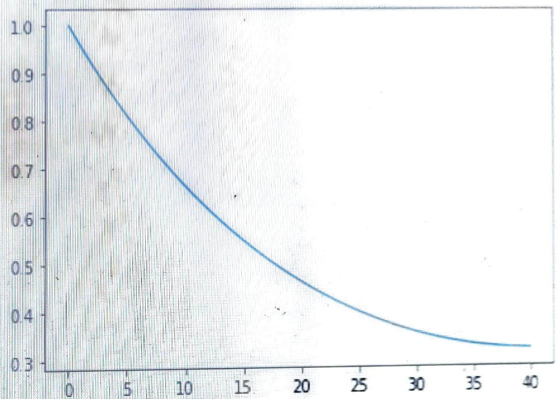
5°C Cloudy

sign Layout References Mailings Review View Help

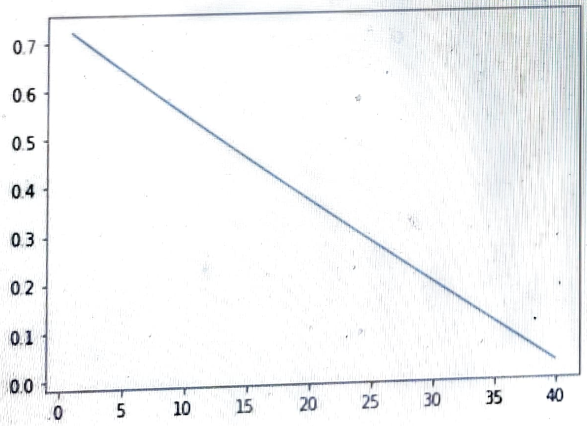
Control with gamma = 0.1



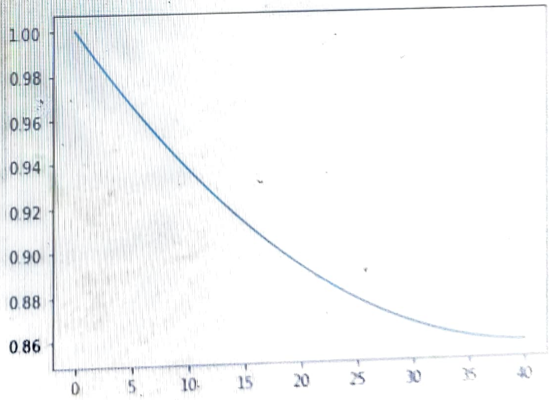
trajectory with gamma = 0.1

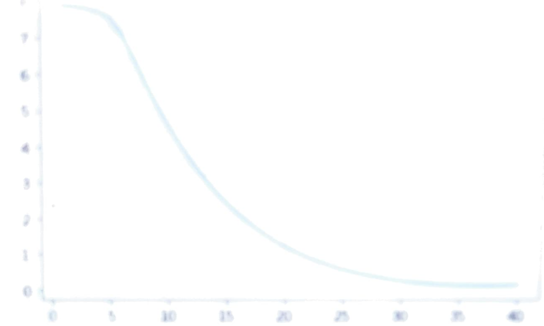


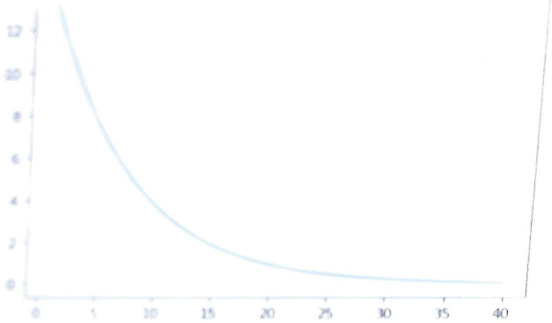
Control with gamma = 1



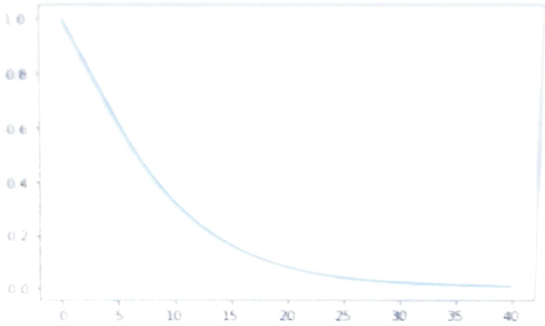
trajectory with gamma = 1



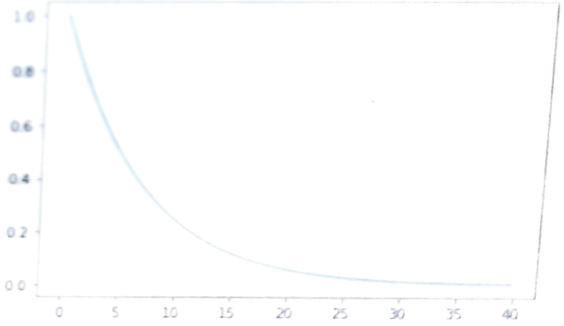


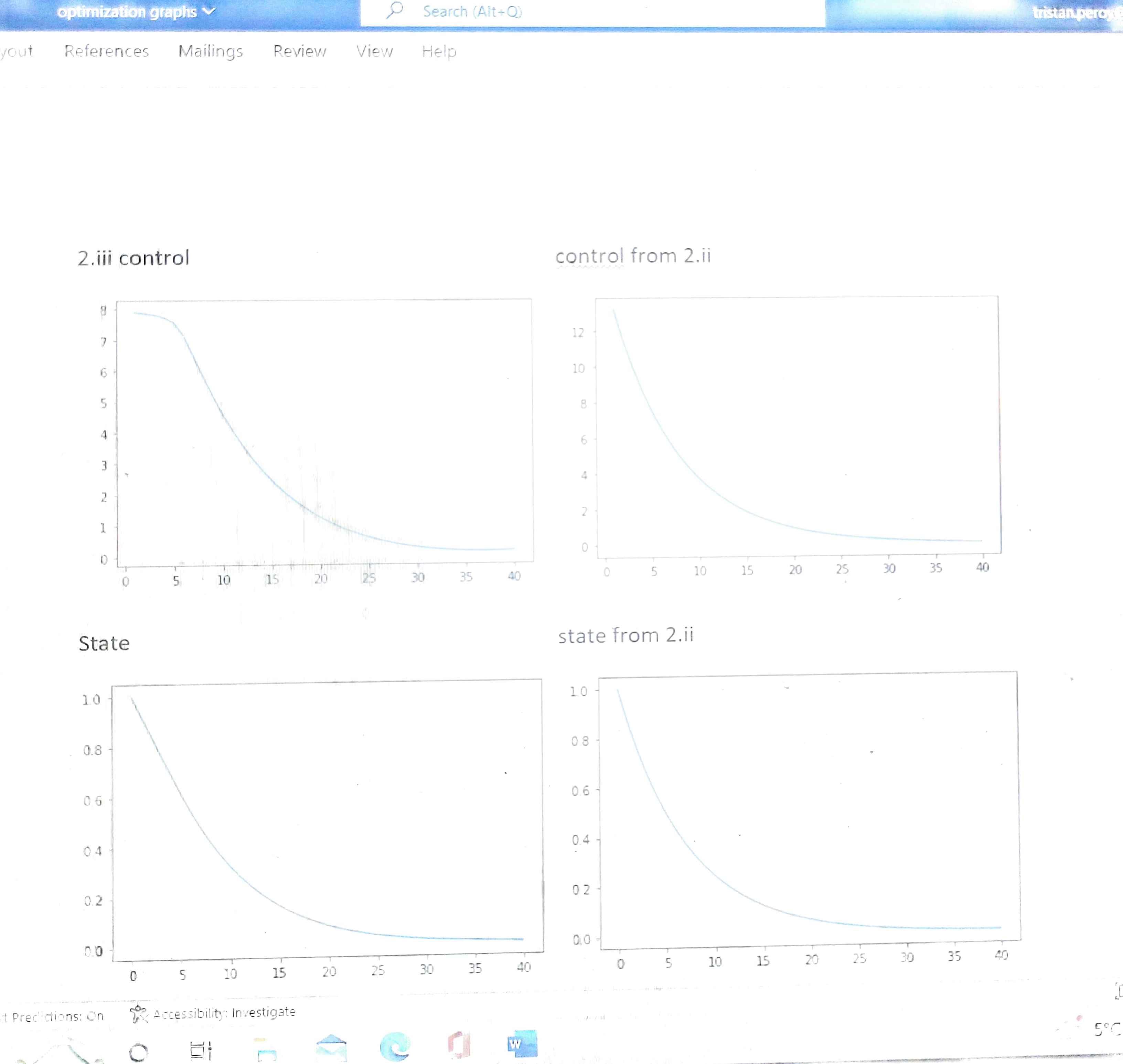


State



state *from 2.ii*





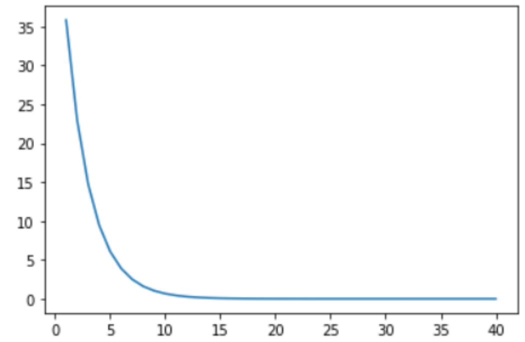
2.iii control

control from 2.ii

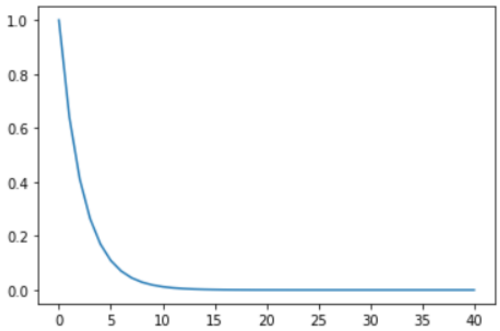
State

state from 2.ii

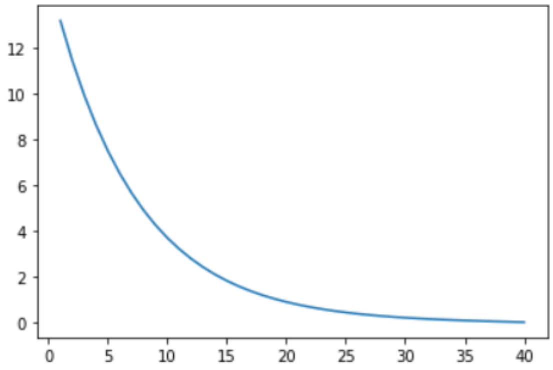
2.ii Control with gamma = 0.001



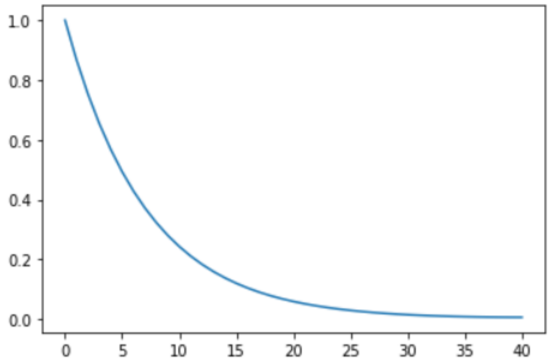
trajectory with gamma = 0.001



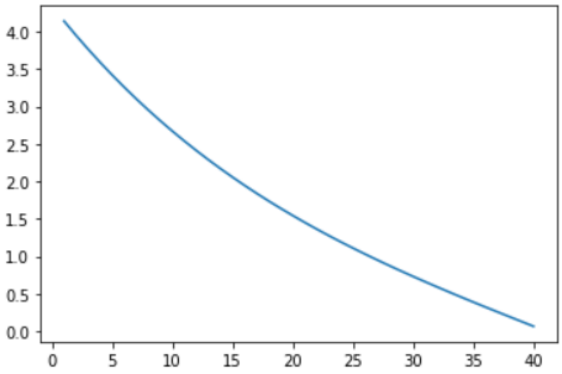
Control with gamma = 0.01



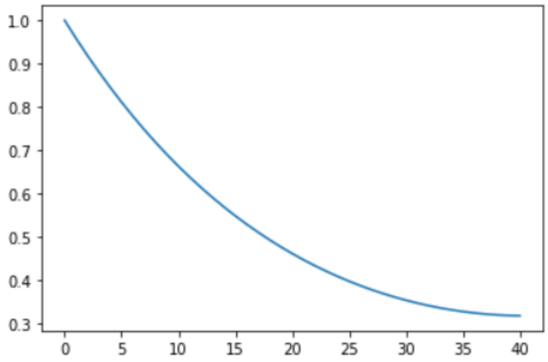
trajectory with gamma = 0.01



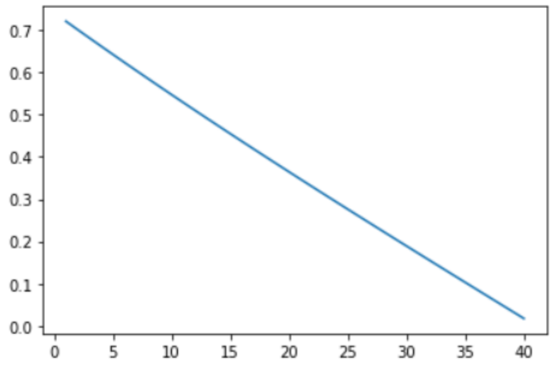
Control with gamma = 0.1



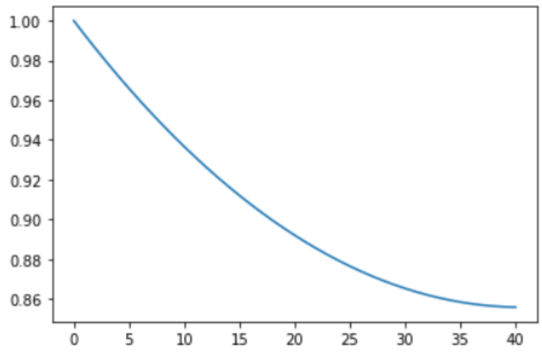
trajectory with gamma = 0.1



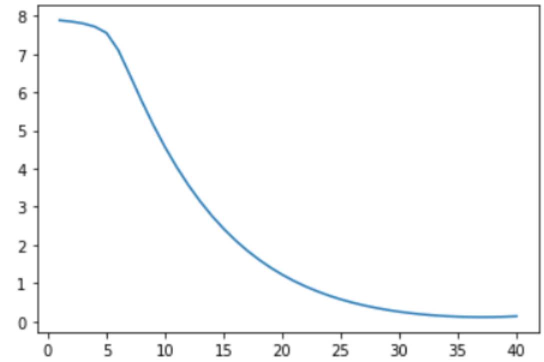
Control with gamma = 1



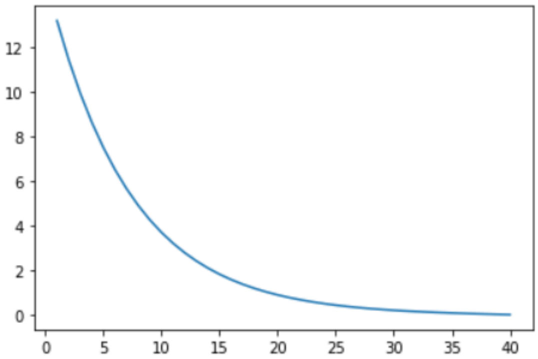
trajectory with gamma = 1



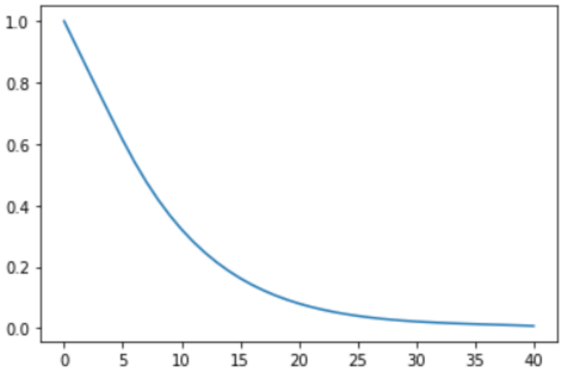
2.iii control



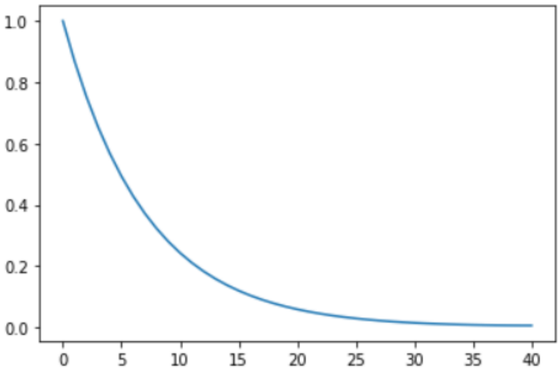
control from 2.ii



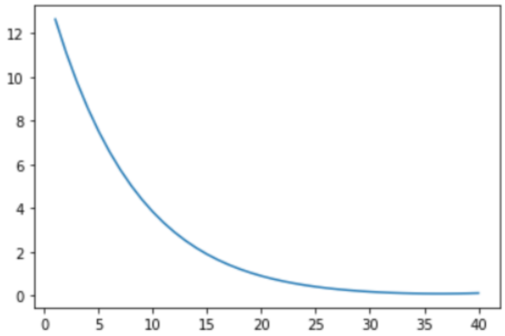
State



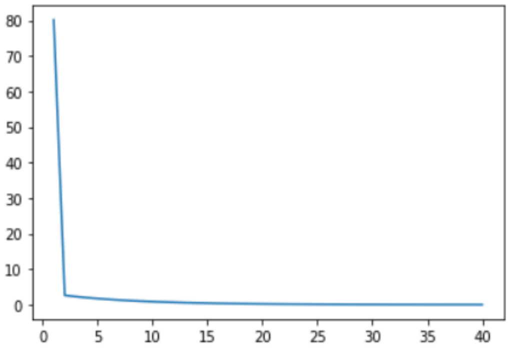
state from 2.ii



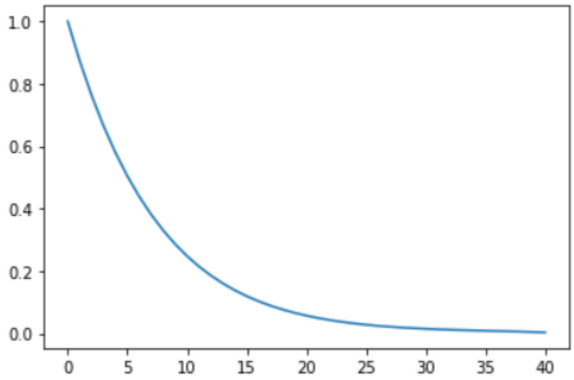
2.iv case 1 control



case 2 control



Case 1 state



case 2 state