

$\theta =$

0.1

0.2

3.5

10

100

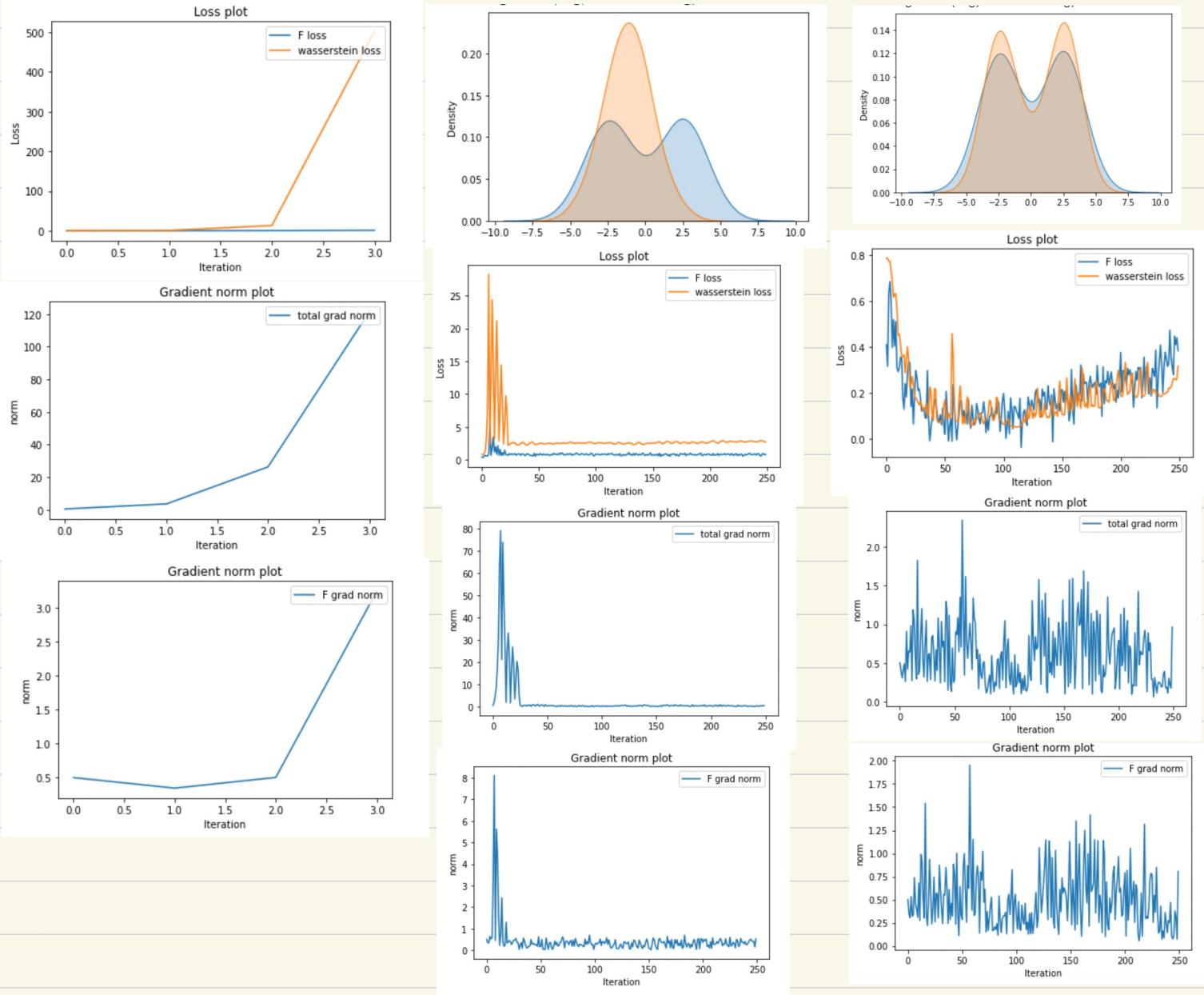
∞

SGD. $M=0.9$, $\eta=5e-2$, $n=100$, $(10, 25)$, $jko_t=$, ini $N(0, 2)$.

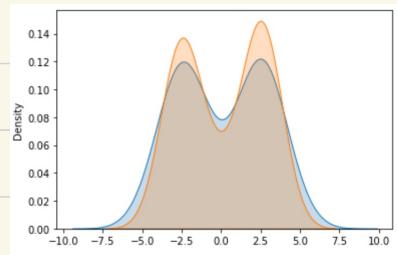
$jko_t=0.1$ diverges.

$jko_t=0.2$. $W_2=3.2$.

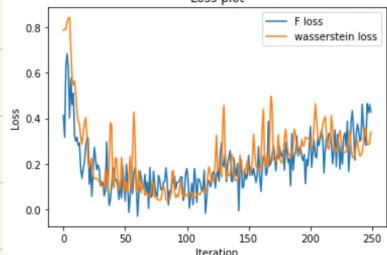
$jko_t=3.5$. $W_2=0.21$



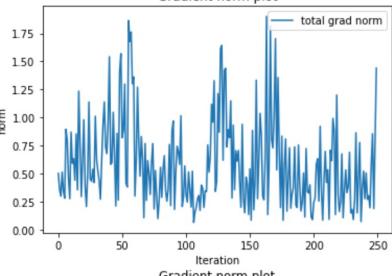
$\text{JK0-}t=10$, $W_2 = 0.223$.



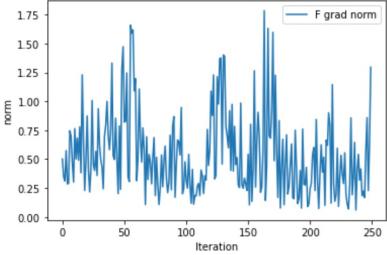
Loss plot



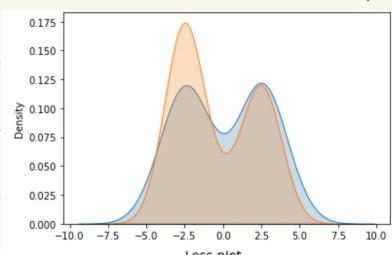
Gradient norm plot



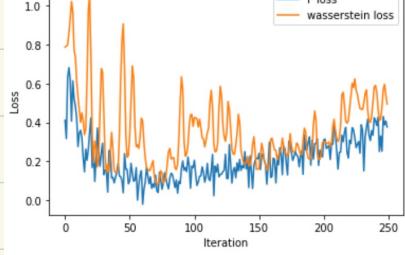
Gradient norm plot



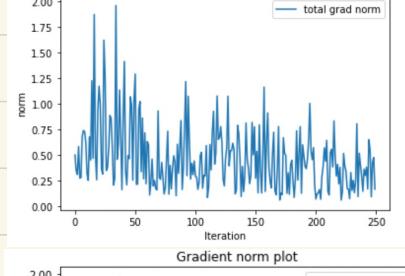
$\text{JK0-}t=100$, $W_2 = 0.879$



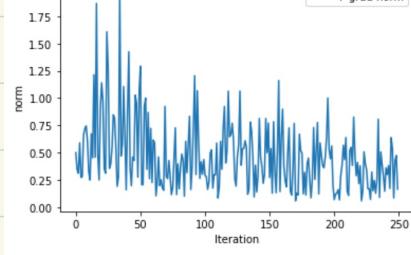
Loss plot



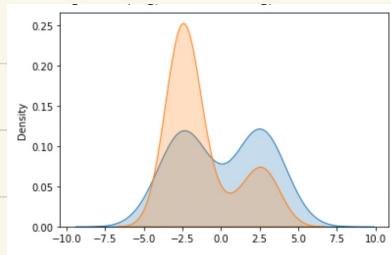
Gradient norm plot



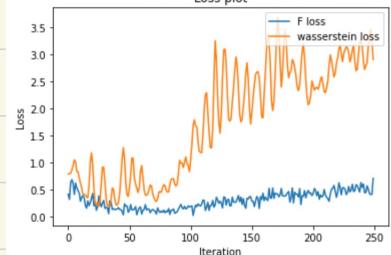
Gradient norm plot



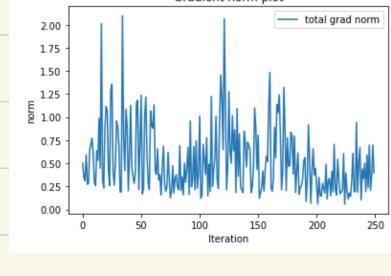
$\text{JK0-}t=\infty$, $W_2 = 3.923$.



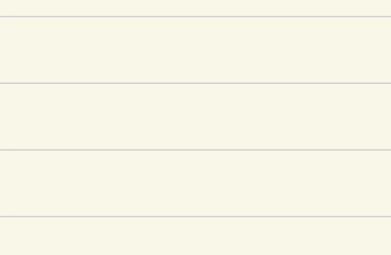
Loss plot



Gradient norm plot



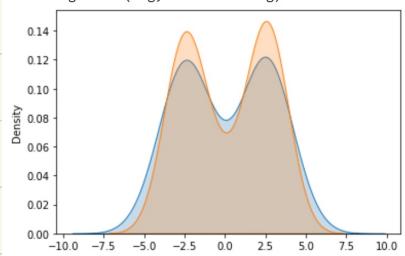
Gradient norm plot



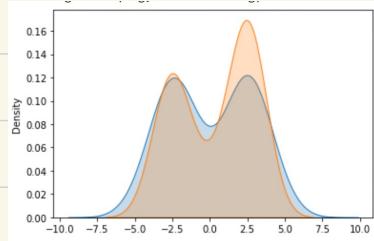
$N(0, 1)$

SGD. $M=0.9$, $\text{lr}=5e-2$, $n=100$, $(10, 25)$, $\text{jko_t}=3.5$, ini $N(0, 2)$.
WGAN.

$N(0, 2)$ $w_2 = 0.21$



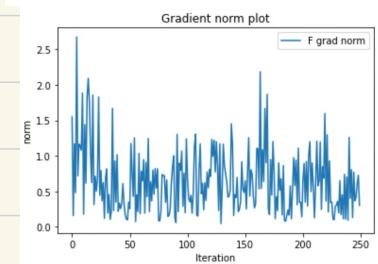
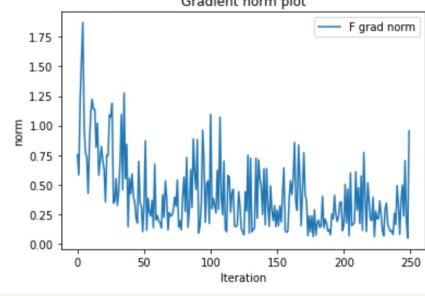
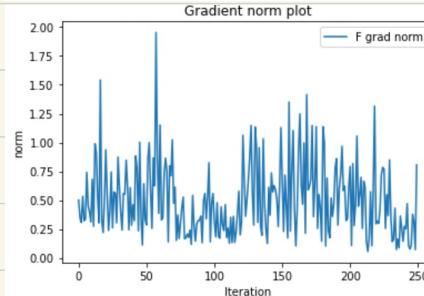
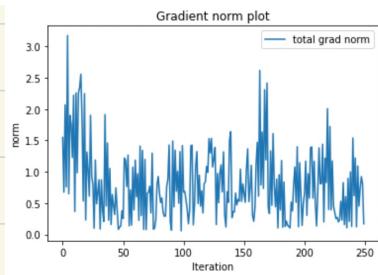
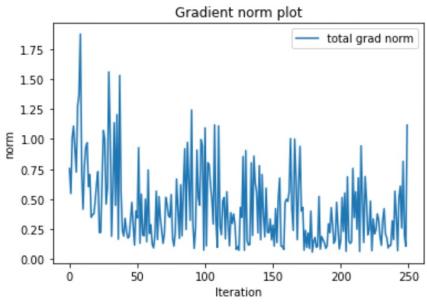
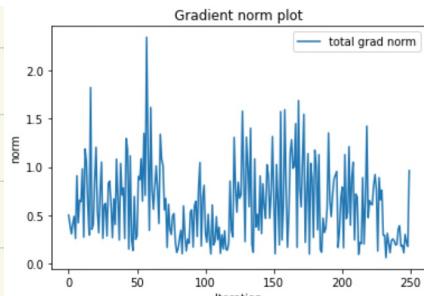
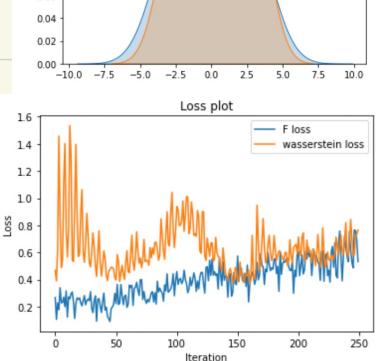
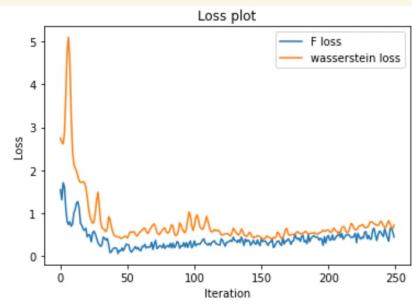
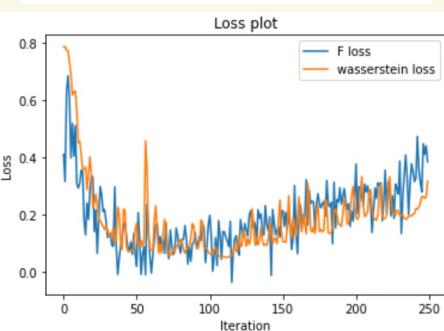
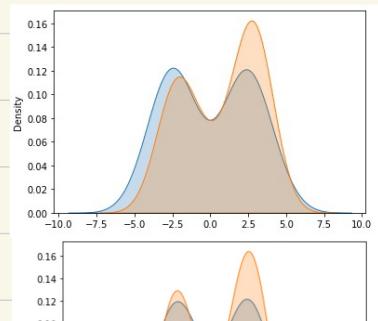
$N(0, 1)$. $w_2 = 0.581$



WGAN, $w_2 = 0.5807$.

WGAN GP result based on
10 samples + noise.

ini



$n=30$

ξ_0

SGD. $M=0.9$, $\text{lr}=5e-2$, $n=100$, $(10, 25)$, $\text{jko_t}=3.5$, ini $N(0, 2)$.

$n=1000$,

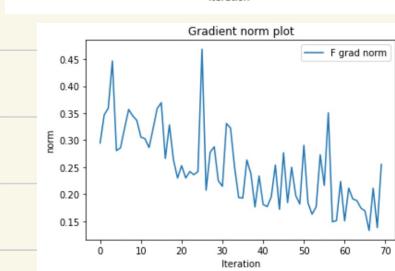
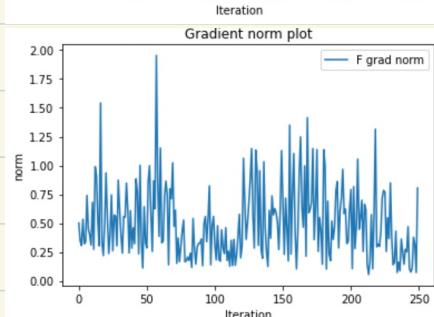
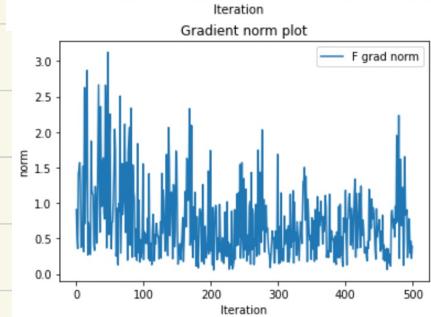
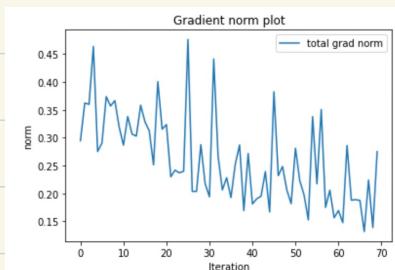
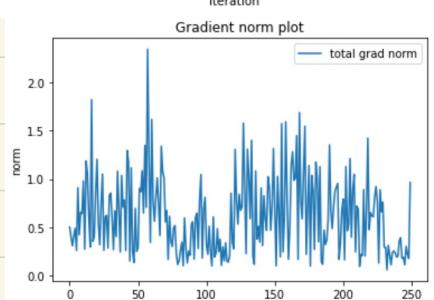
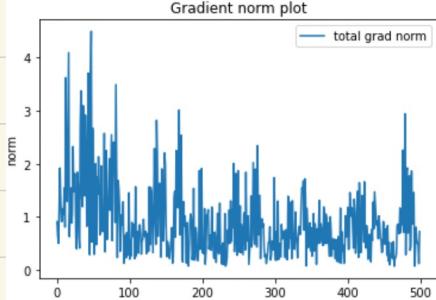
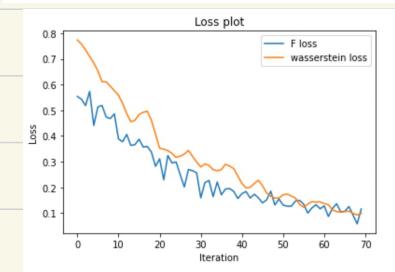
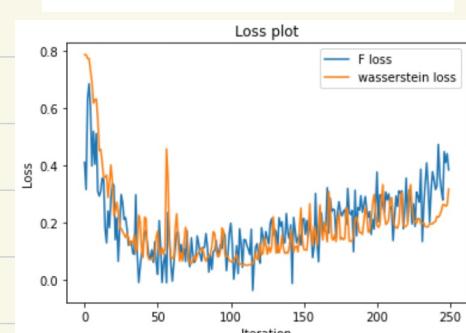
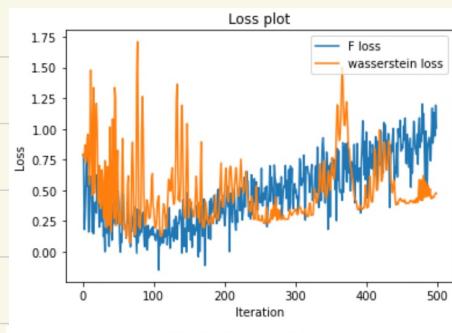
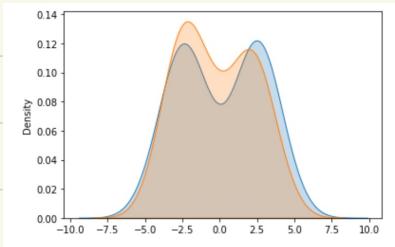
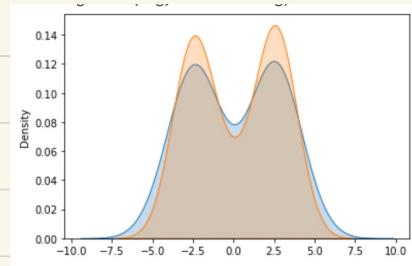
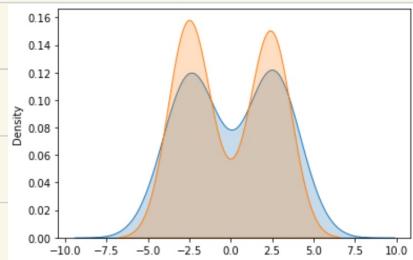
γ

$n=30, (10, 50), w_2=0.476$

$n=100, (10, 25), w_2=0.21$

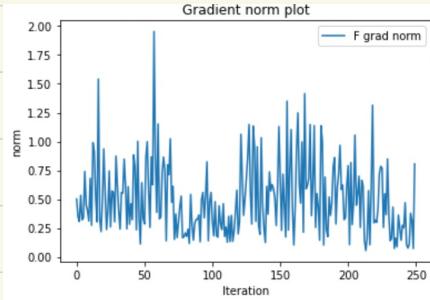
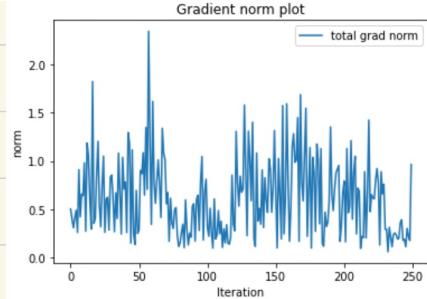
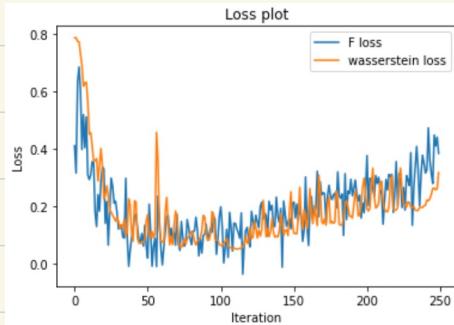
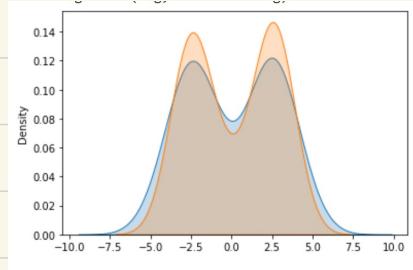
$n=1000,$

$(10, 7), W_2=0.255.$

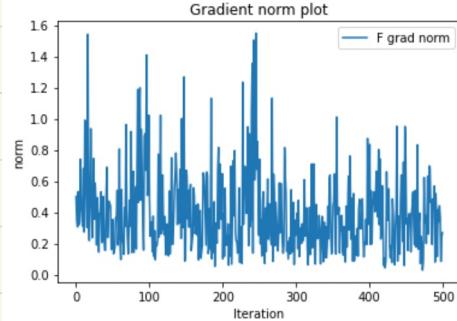
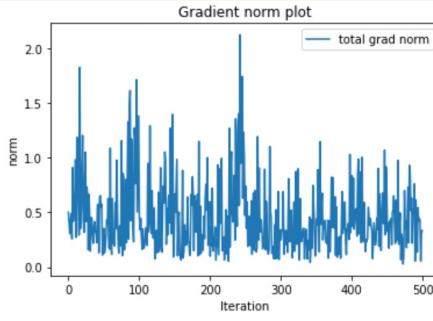
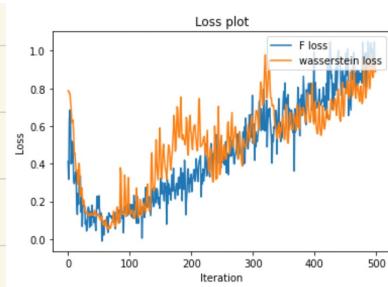
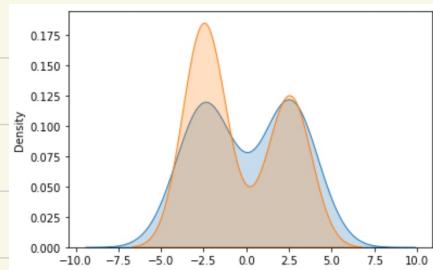


SGD. $M=0.9$, $\eta=5e-2$, $n=100$, $(10, 50)$, $jko_t=3.5$, ini $N(0, 2)$.

$n=100$, $(10, 25)$, $w_2=0.21$



$n=100$, $(10, 50)$, $w_2=1.322$.

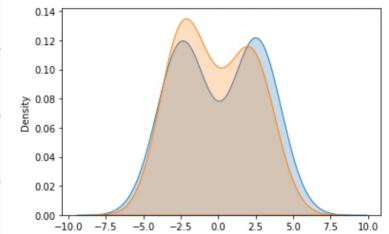


SGD. $M=0.9$, $\text{lr}=5e-2$, $n=1000$ (10, 75), $jko_t=3.5$, ini $N(0, 2)$.

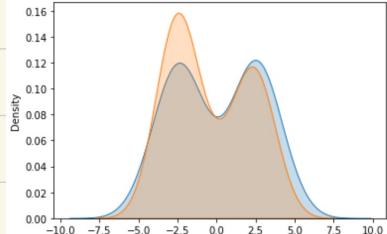
7

25

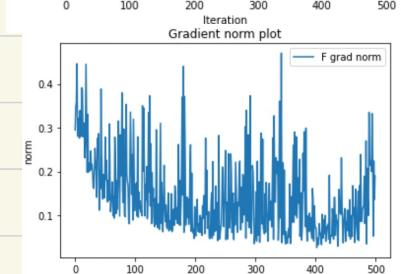
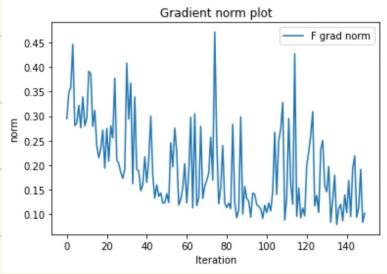
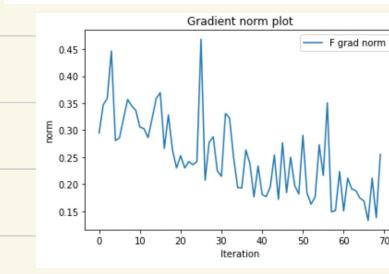
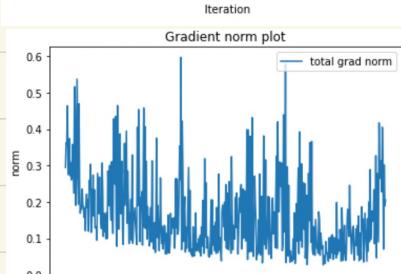
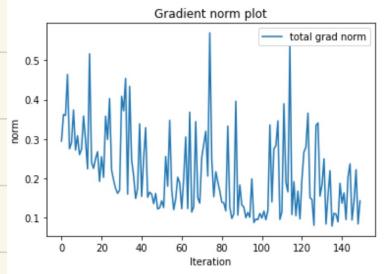
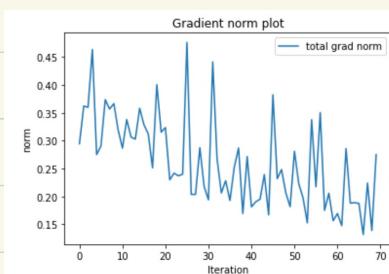
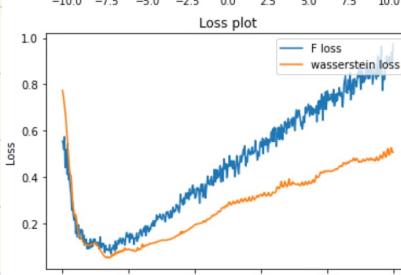
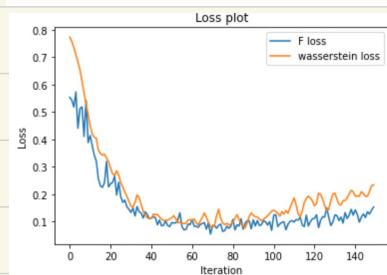
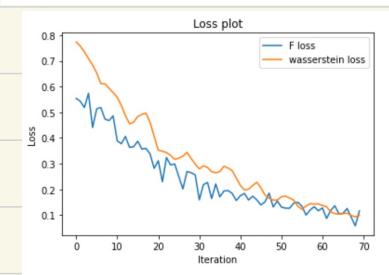
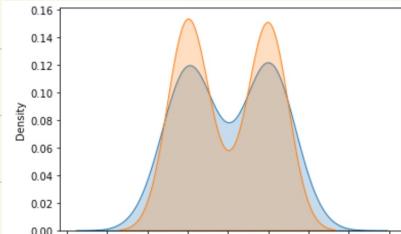
(10, 7). $W_2 = 0.255$.



(10, 15) $W_2 = 0.533$

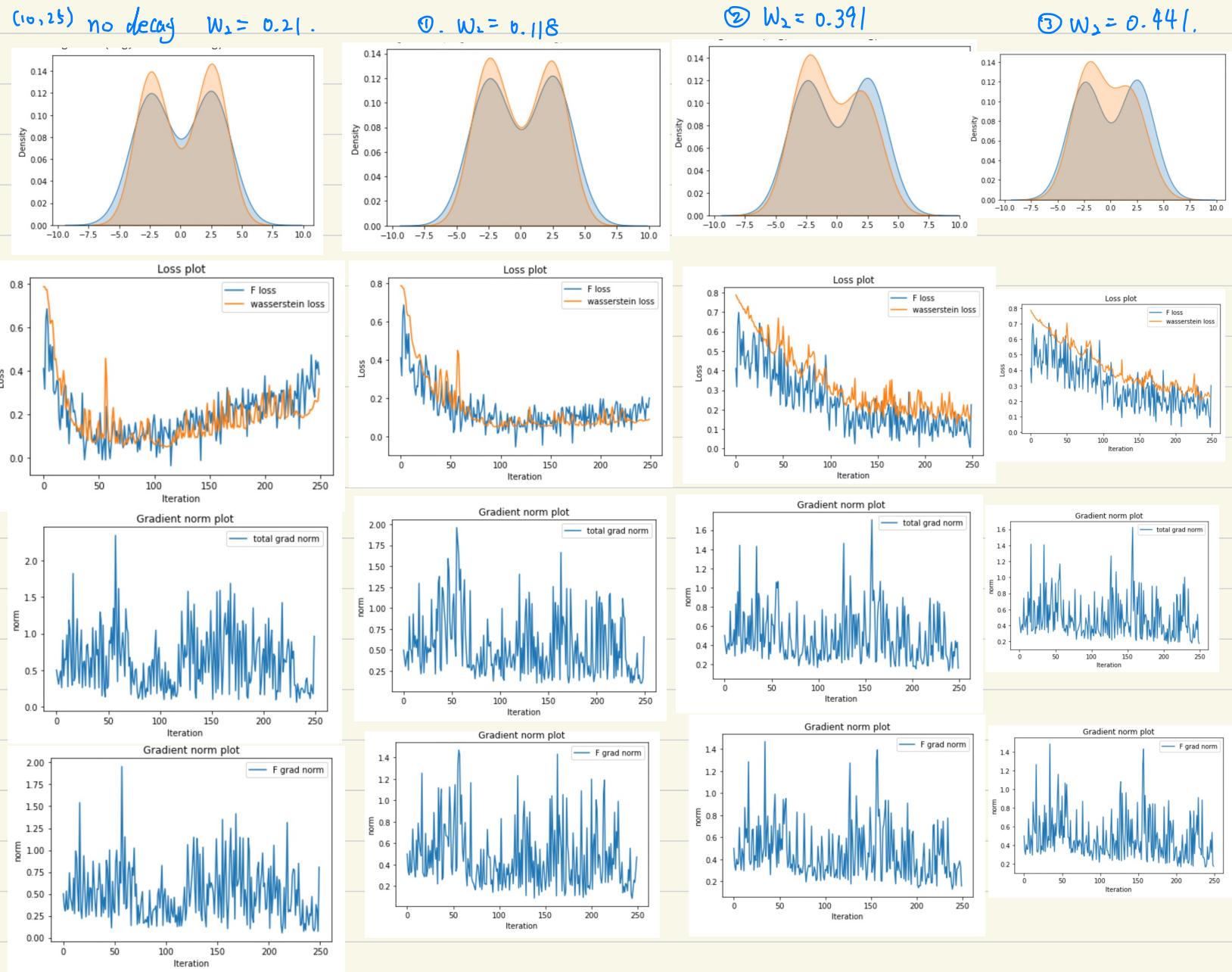


(10, 50), $W_2 = 0.426$.



- ① lr decay by 0.95 for each epoch. (inner).
 ② M decay ...
 ③ ① and ②.

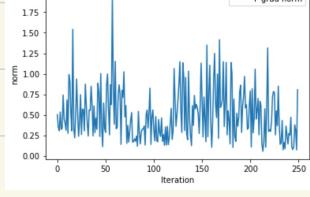
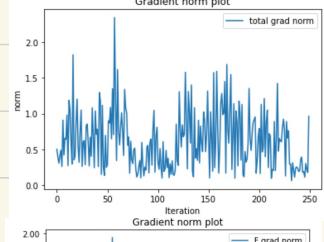
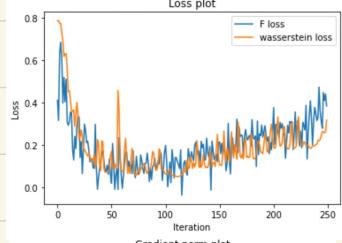
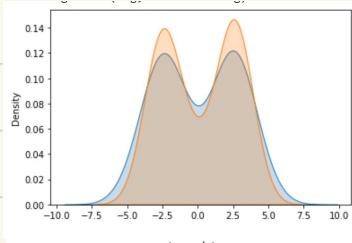
SGD. $M=0.9$, $\text{lr}=5e-2$, $n=100$, $(10, 25)$, $jko_t=3.5$, ini $N(0, 2)$.



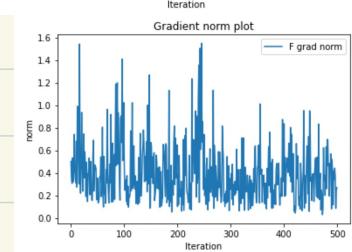
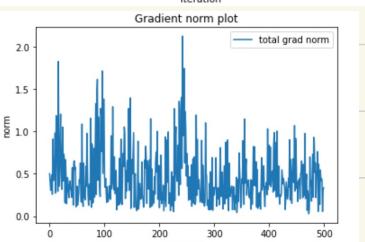
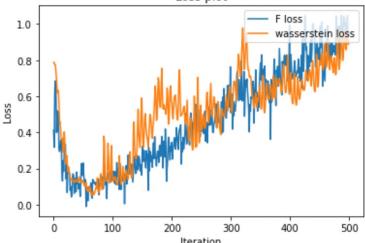
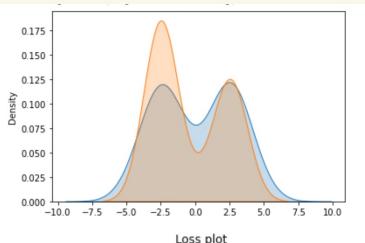
- ①. lr decay by 0.95 for each epoch. (inner).
 ②. M decay....
 ③. ① and ②.

SGD. $M=0.9$, $lr=5e-2$, $n=100$, $(10, 50)$, $jko_t=3.5$, ini $N(0, 2)$.

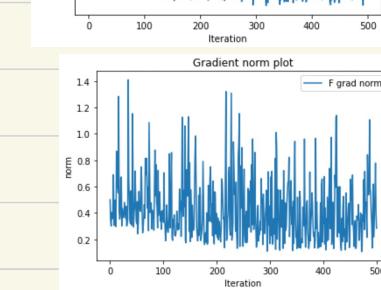
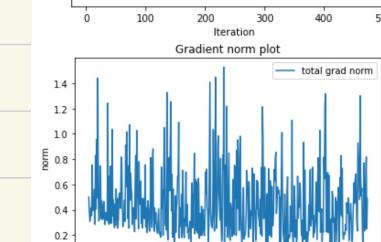
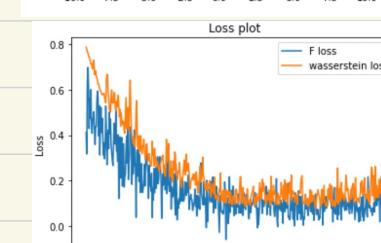
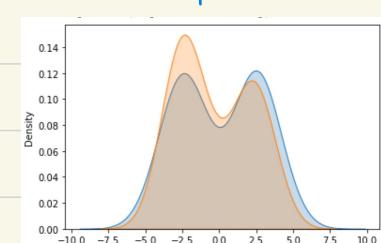
$(10, 25)$ no decay $W_1 = 0.21$.



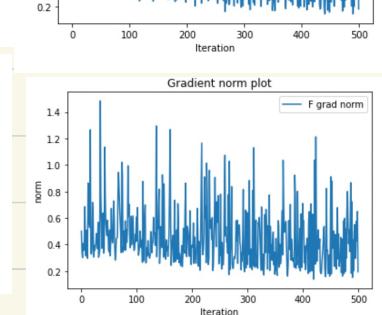
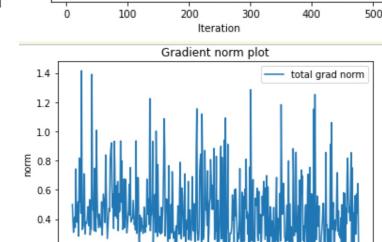
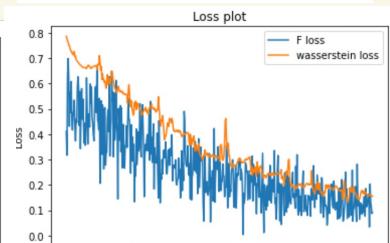
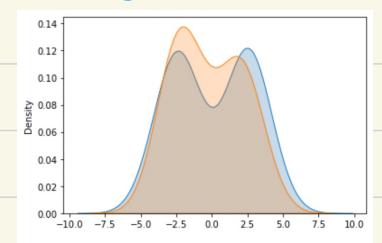
① $W_2 = 1.32$



② $W_2 = 0.39$



③ $W_2 = 0.334$



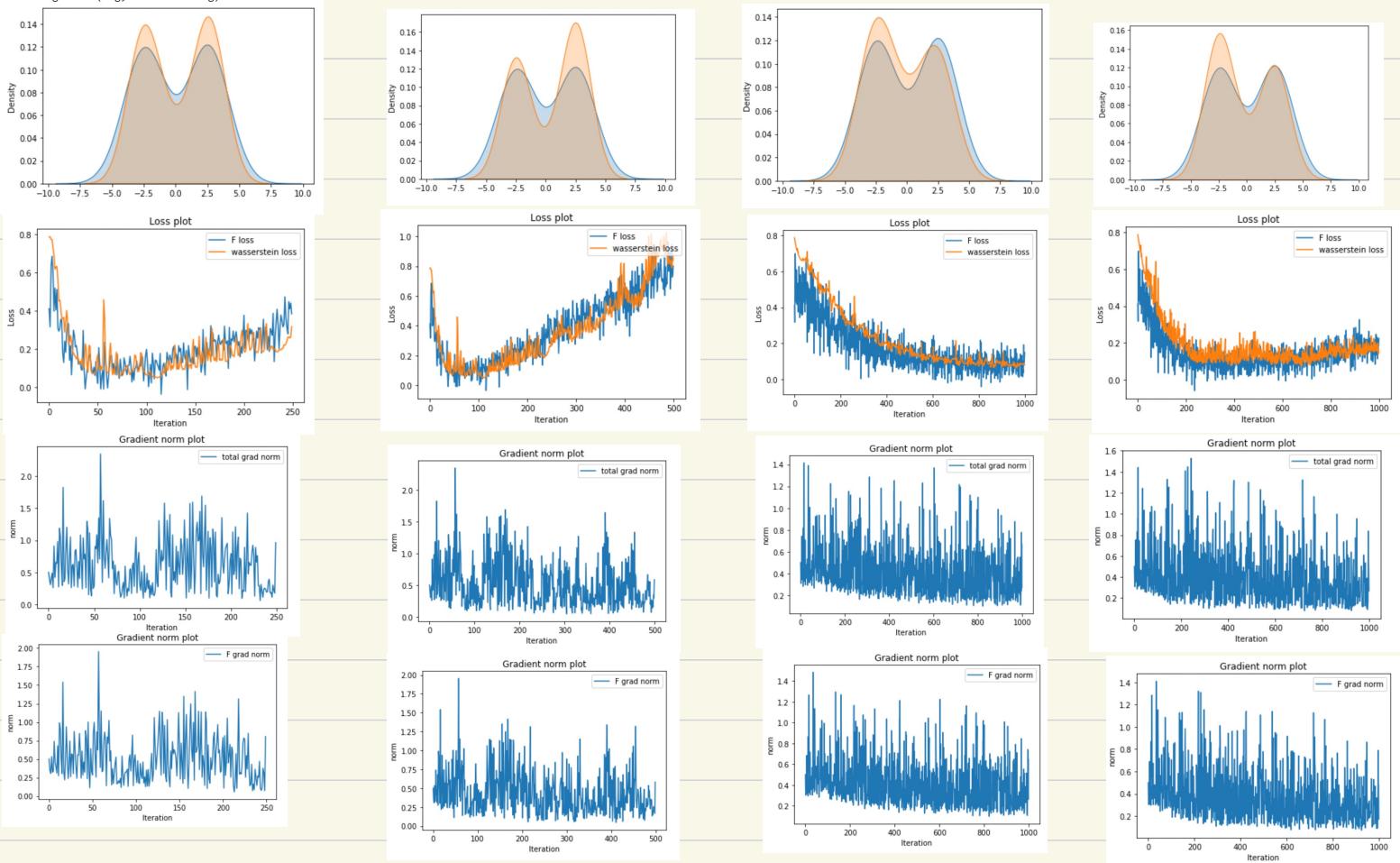
- ①. lr decay by 0.95 for each epoch. (inner).
 ②. M decay....
 ③. ① and ②.

SGD. $M=0.9$, $lr=5e-2$, $n=100$, $(20, 50)$, $jko_t=3.5$, ini $N(0, 2)$.

$(10, 25)$ no decay $W_1 = 0.21$. ① $(20, 25)$ $W_2 = 0.687$

② $W_2 = 0.252$.

③ $W_2 = 0.398$.



LBFGS.

in $N(0, 1)$, $\delta r = 0.1$, $\text{patient/e} = 1$, $(5, 50)$, $\text{history_size} = 10$, $\text{max_iter} = 4$.

$$W_2 = 0.487$$

