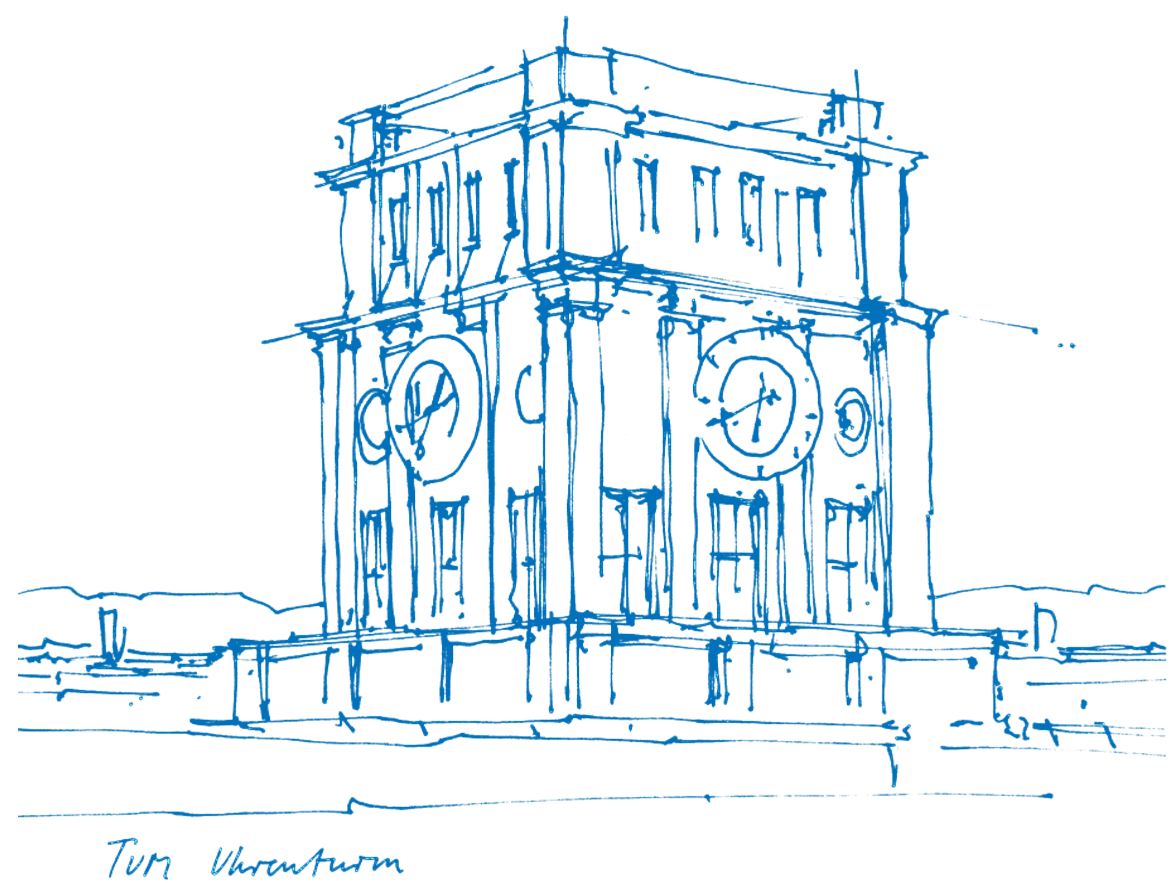
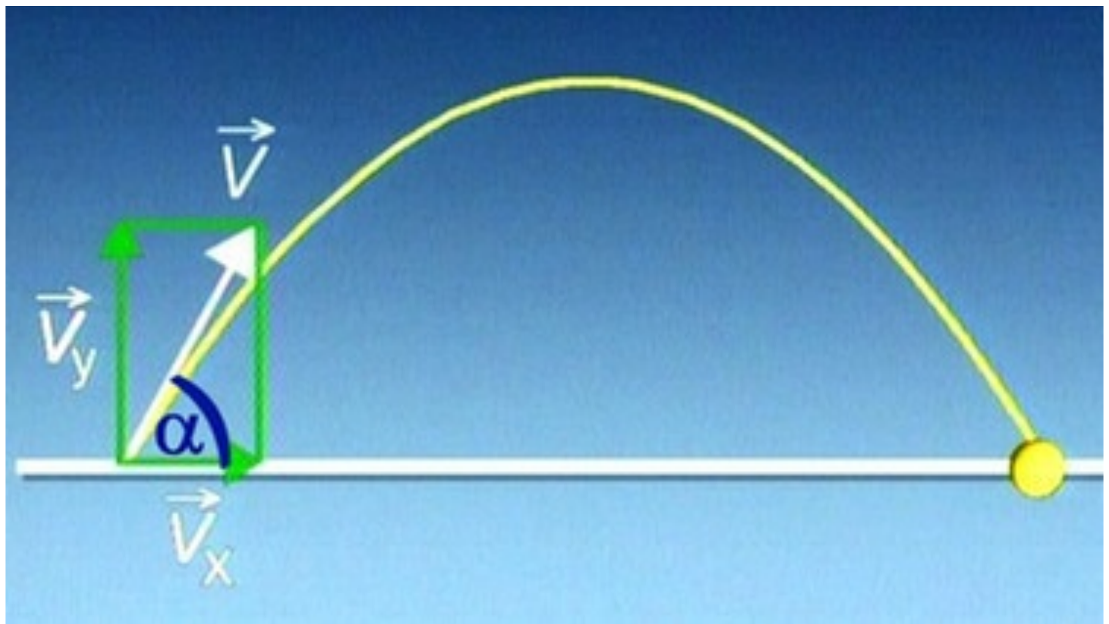
Physically constrained networks

Korbinian Abstreiter
Technical University of Munich
Department of Informatics
Computer Vision Group
Munich, 19th July 2019



I. Motivation

Hand-crafted models



https://www.br.de/telekolleg/faecher/physik/trimester2/telekolleg-physik-05-schwerkraft-100.html

- Mathematical description
- Based on theory (e.g. Laws of Physics)
- High effort to design model
- Domain knowledge required

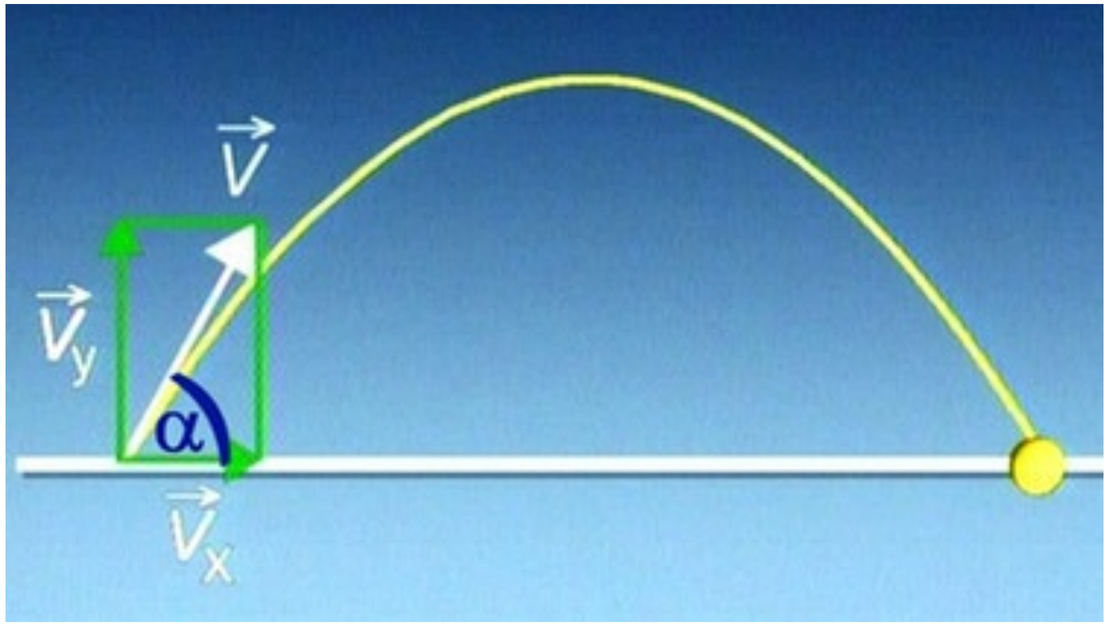
Deep learning models

https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRB_oBff0NlA087mkfhKQ9dujNPhQYjNOEoo3drYeuclw7Ov0HRfQ

- No mathematical description
- Based on data
- Predictions often violate known theoretical rules

I. Motivation

Hand-crafted models



https://www.br.de/telekolleg/faecher/physik/trimester2/telekolleg-physik-05-schwerkraft-100.html

- Mathematical description
- Based on theory (e.g. Laws of Physics)
- High effort to design model
- Domain knowledge required

Deep learning models

https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRB_oBff0NIA087mkfhKQ9dujNPhQYjNOEoo3drYeuclw7Ov0HRfQ

- No mathematical description
- Based on data
- Predictions often violate known theoretical rules

→ **Combine advantages using physically constrained networks**

I. Motivation

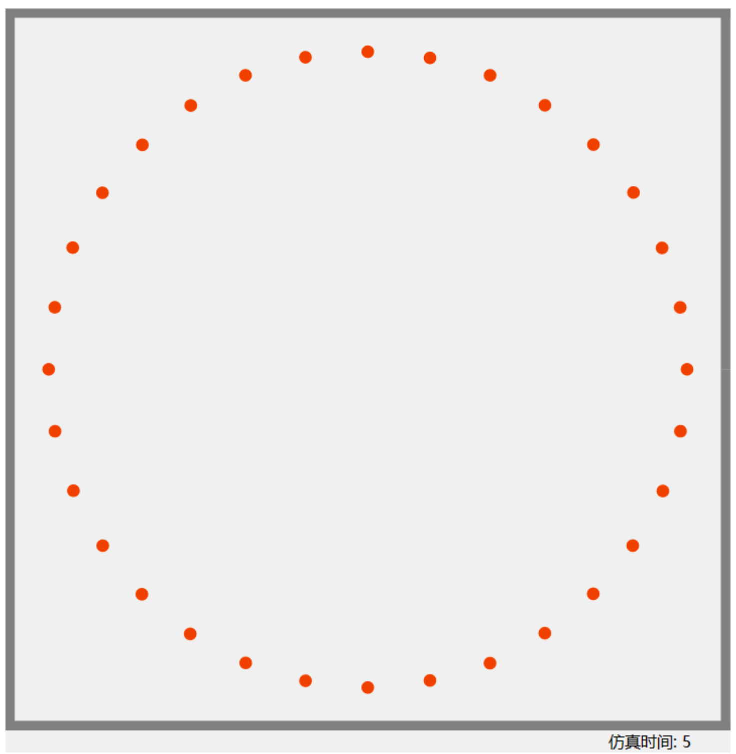
Pedestrian Trajectory Prediction

http://pedynamic.com/circle-antipode-experiments/

I. Motivation

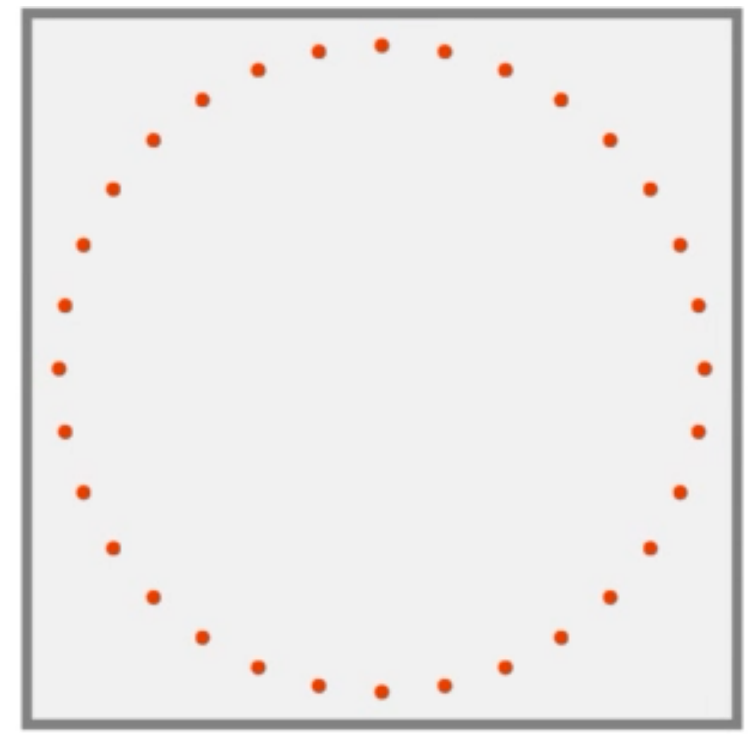
Pedestrian Trajectory Prediction

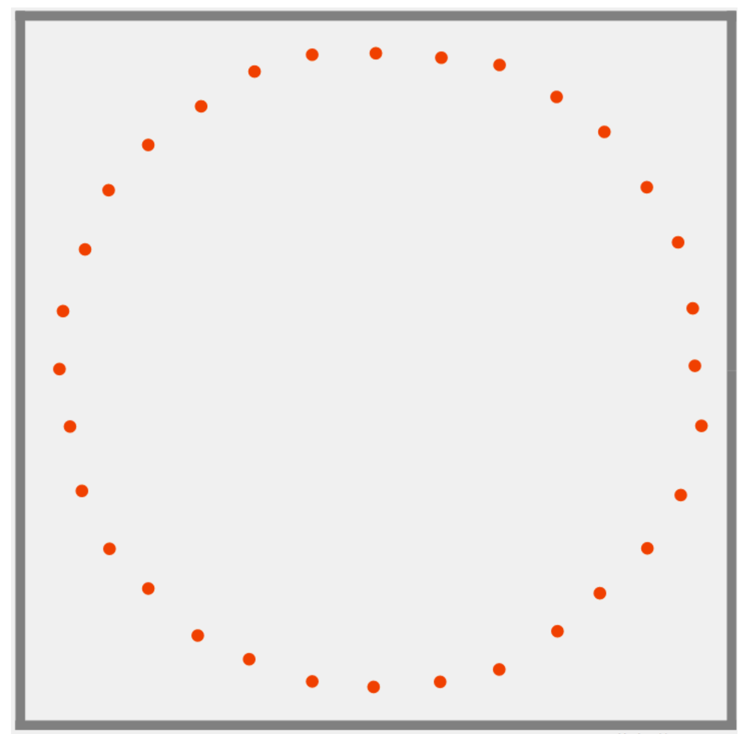
Hand-crafted



http://pedynamic.com/circle-antipode-simulation/

Deep learning





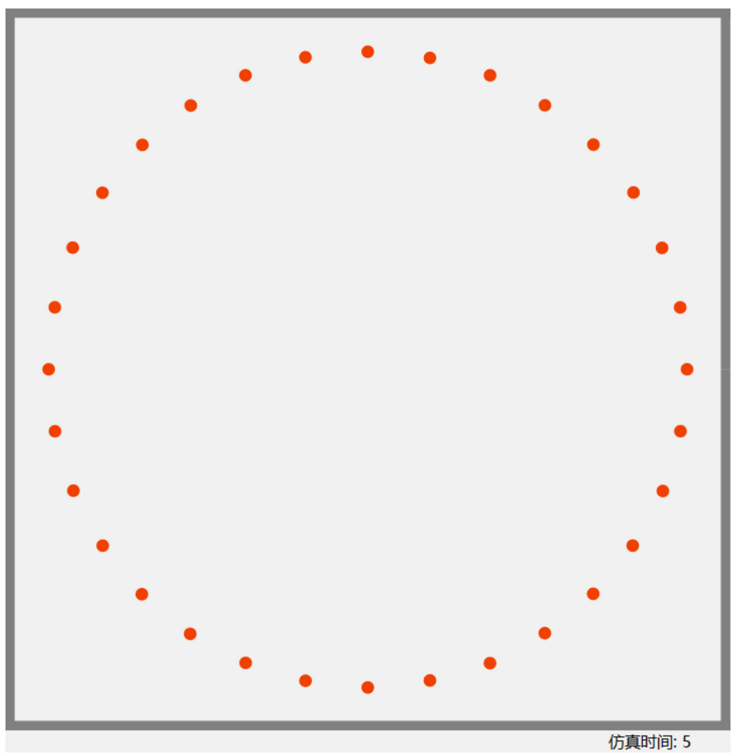
http://pedynamic.com/circle-antipode-simulation/

Physically constrained network

I. Motivation

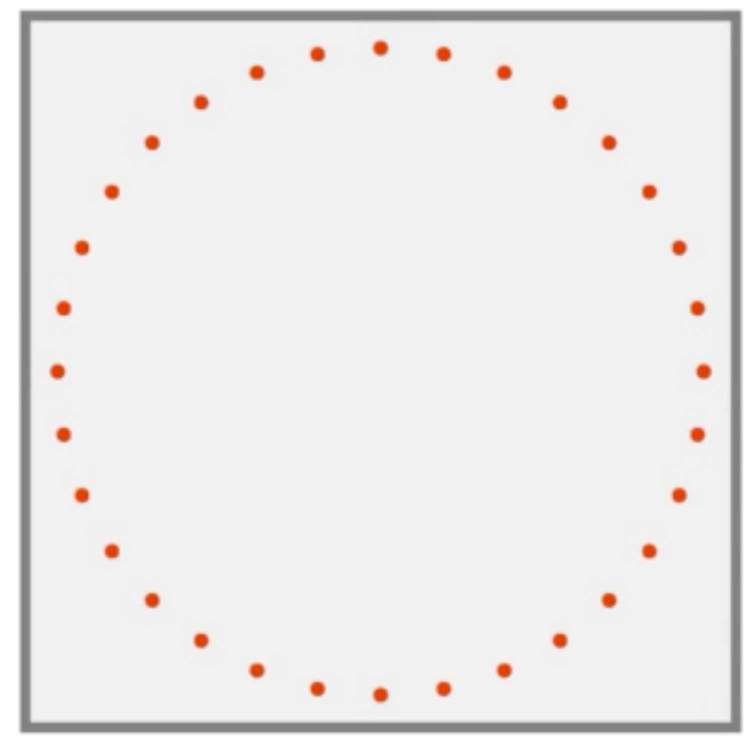
Pedestrian Trajectory Prediction

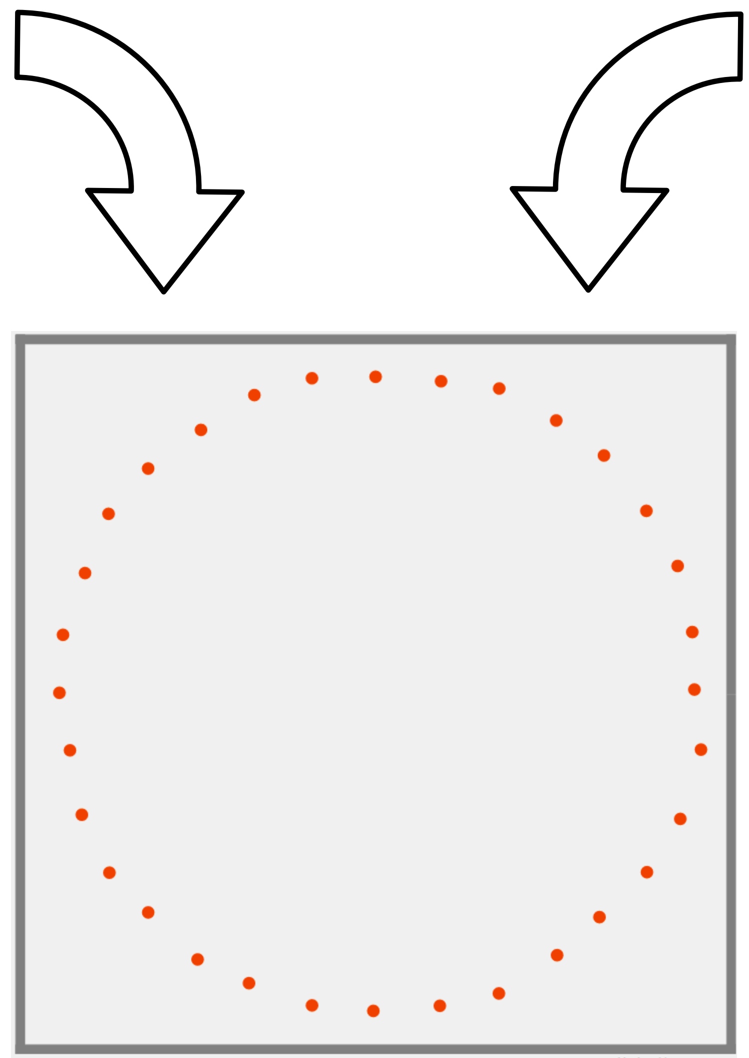
Hand-crafted



http://pedynamic.com/circle-antipode-simulation/

Deep learning





http://pedynamic.com/circle-antipode-simulation/

Physically constrained network

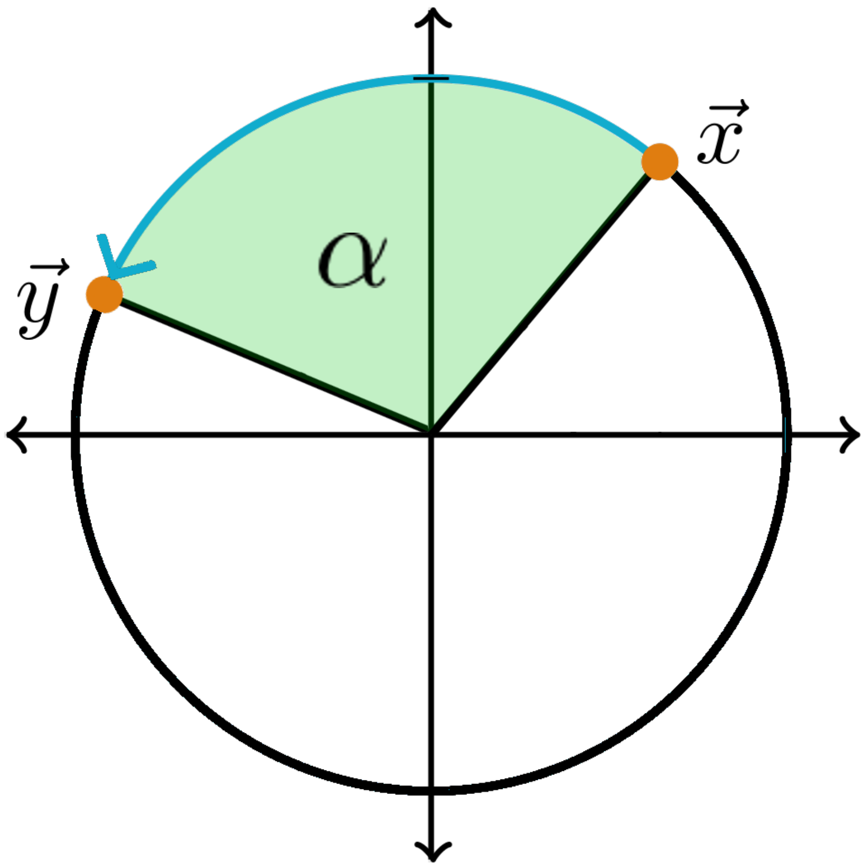
Outline

I. Motivation

II. Constrained optimisation

III. Experiment

IV. Results



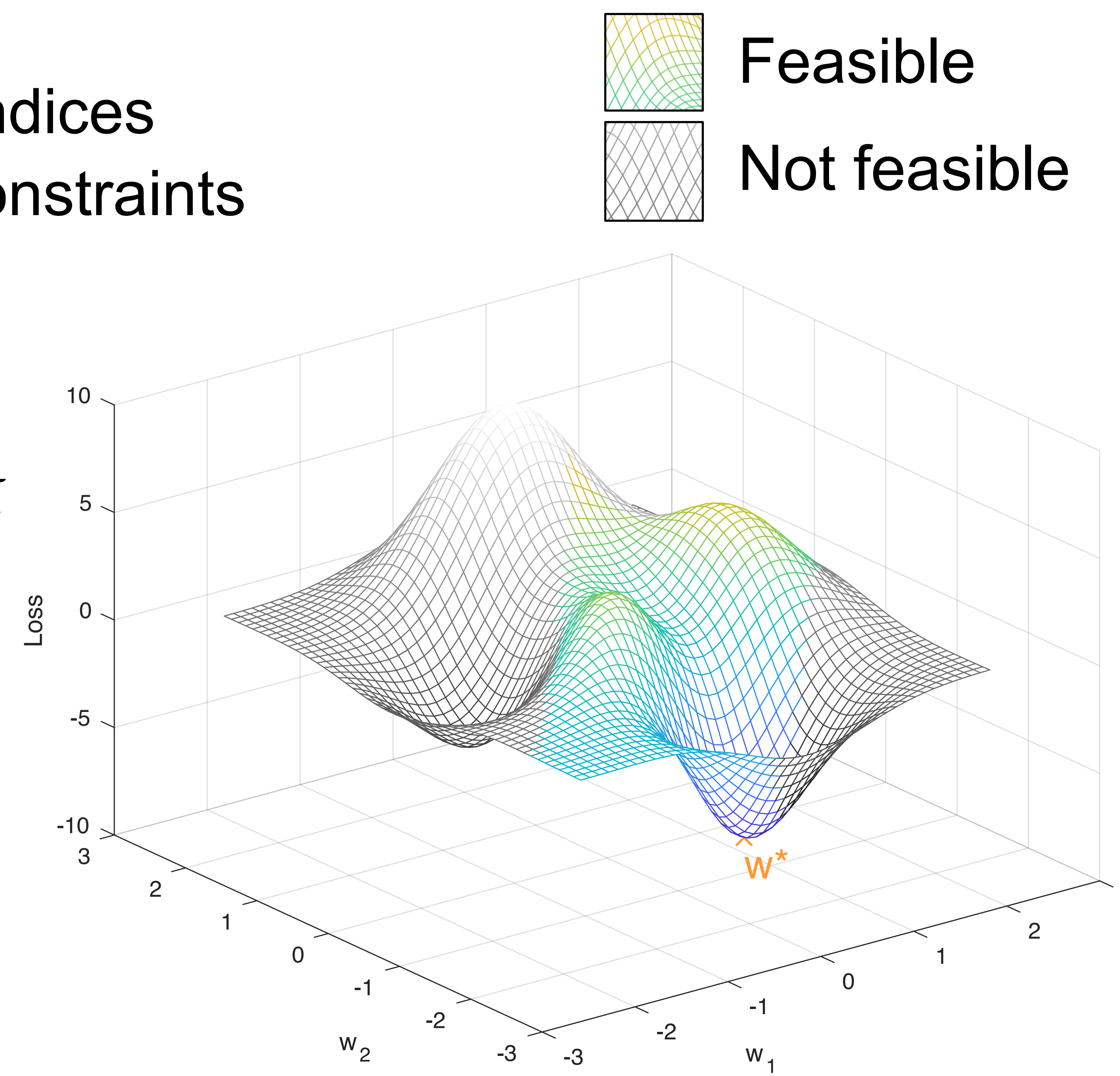
II. Constrained optimisation

Training a constrained network

$\mathcal{L} : \mathbb{R}^n \to \mathbb{R}$: Loss function
$\mathcal{I}$: Set of inequality constraint indices
$c_i : \mathbb{R}^n \to \mathbb{R}, i \in \mathcal{I}$: Inequality constraints

$$w^* = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w)$$

$$\text{subject to} \quad c_i(w) \geq 0, \; i \in \mathcal{I}$$



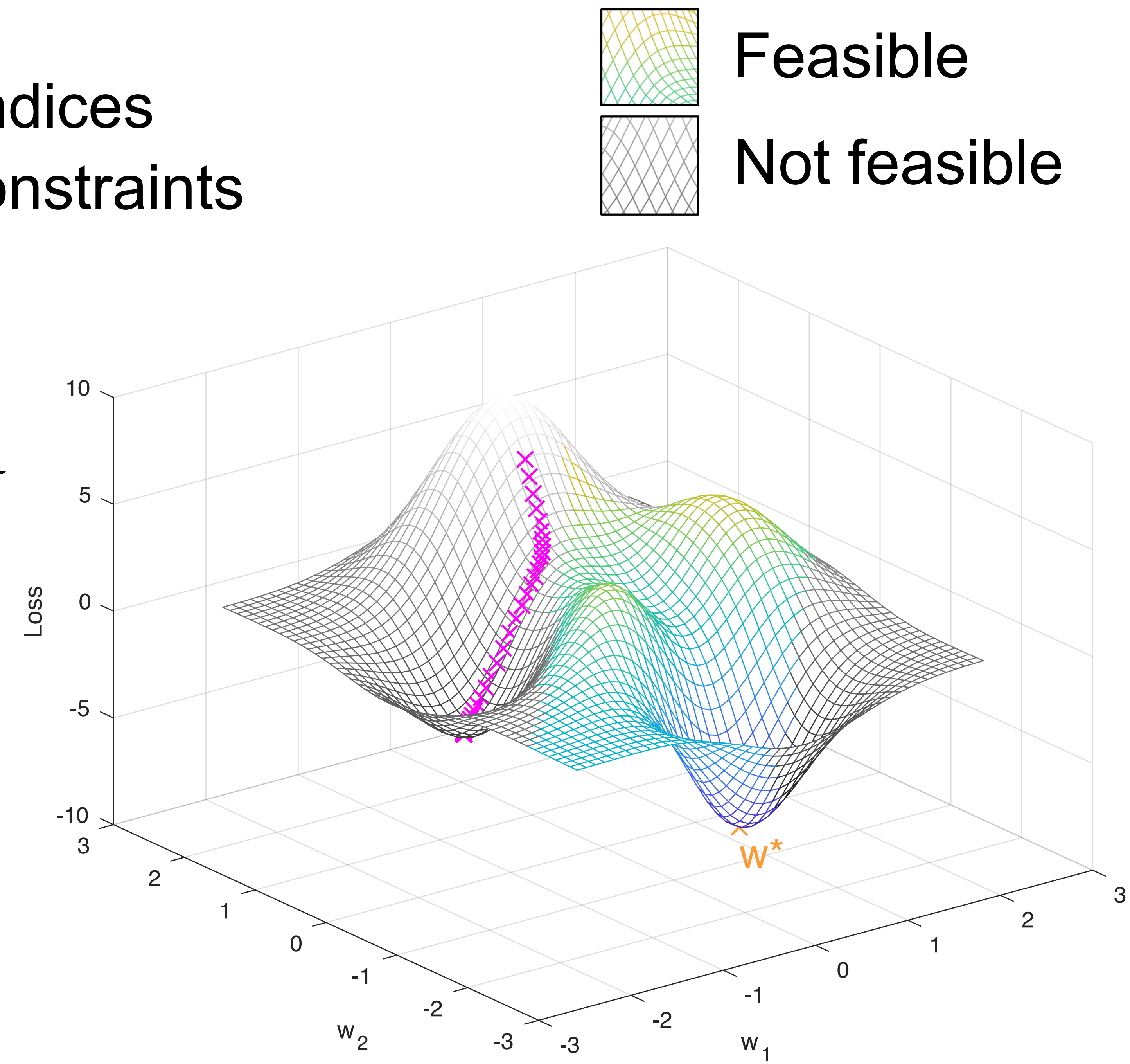
II. Constrained optimisation

Training a constrained network

$\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}$: Loss function
$\mathcal{I}$: Set of inequality constraint indices
$c_i : \mathbb{R}^n \rightarrow \mathbb{R}, i \in \mathcal{I}$: Inequality constraints

$$w^* = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w)$$

subject to $\quad c_i(w) \geq 0, \; i \in \mathcal{I}$



II. Constrained optimisation

Penalty Method

$$w^* = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w)$$
$$\text{subject to} \quad c_i(w) \geq 0, \; i \in \mathcal{I}$$

Constraint violations: $c_i^-(x) = \min(0, c_i(x))$

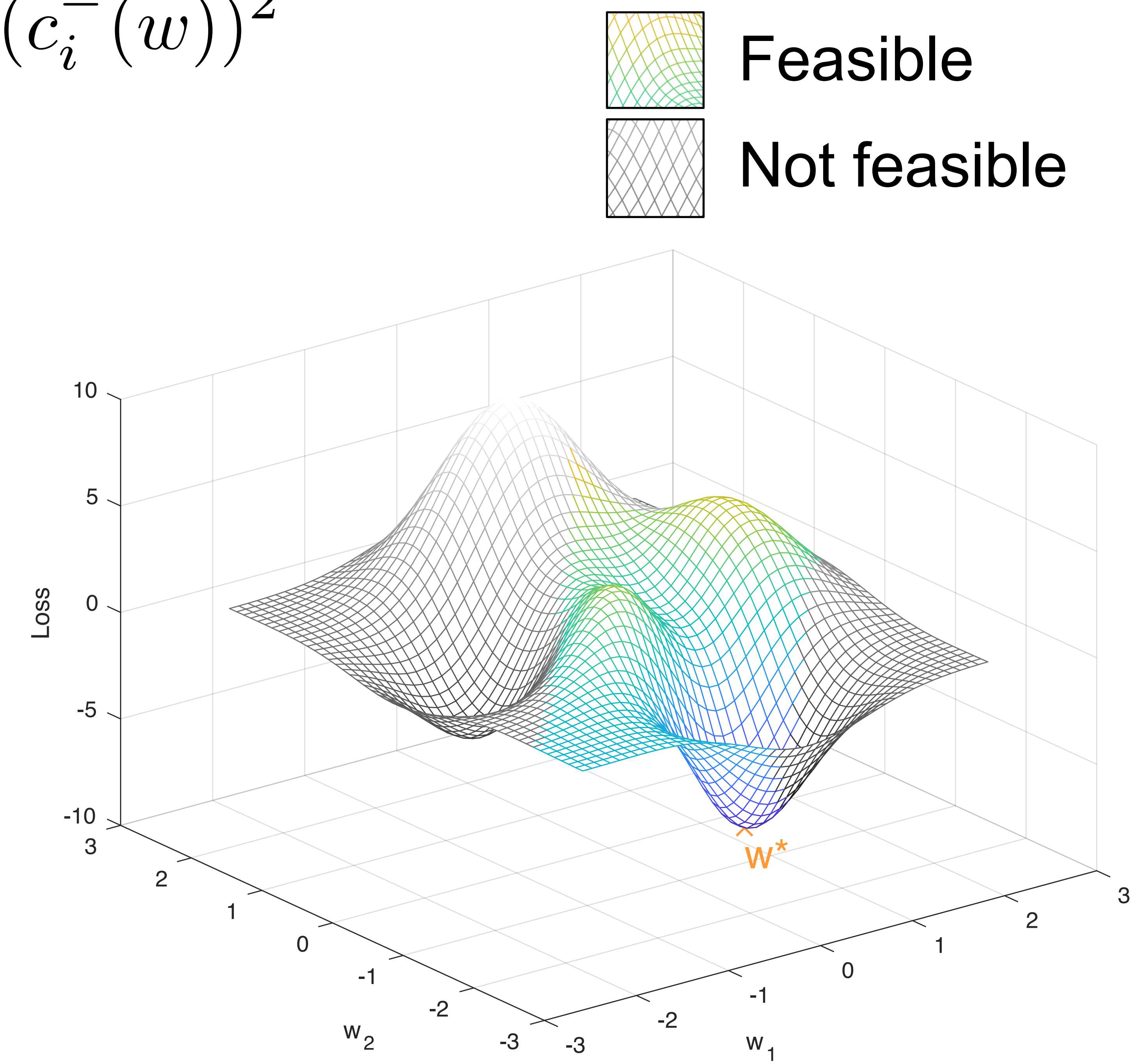
Penalty weight: $\rho > 0$

$$\overline{w} = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w) + \rho \sum_{i \in \mathcal{I}} (c_i^-(w))^2$$

II. Constrained optimisation

Penalty Method

$$\overline{w} = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w) + \rho \sum_{i \in \mathcal{I}} (c_i^-(w))^2$$

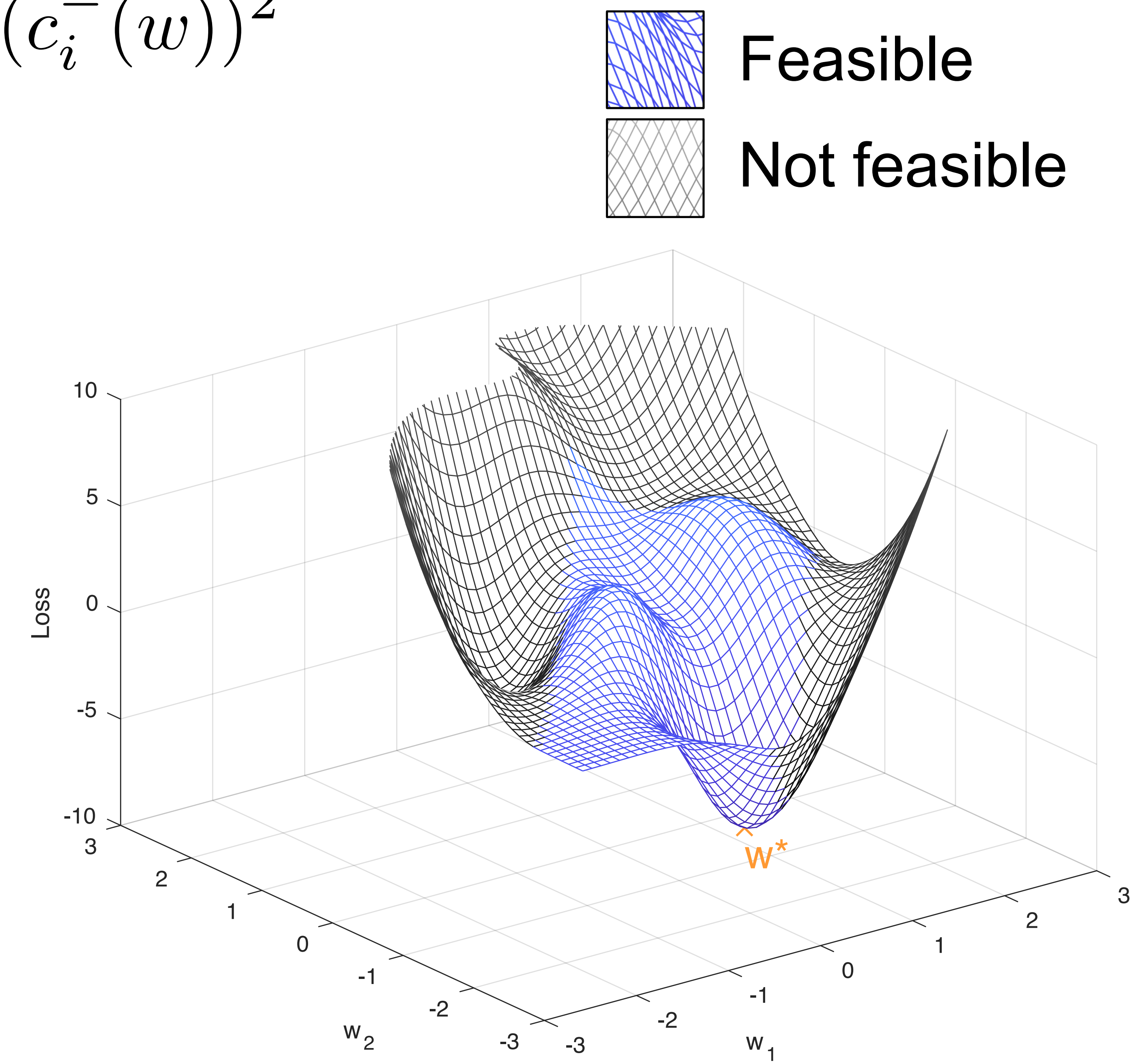


II. Constrained optimisation

Penalty Method

$$\overline{w} = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w) + \rho \sum_{i \in \mathcal{I}} (c_i^-(w))^2$$

$$\rho = 3$$

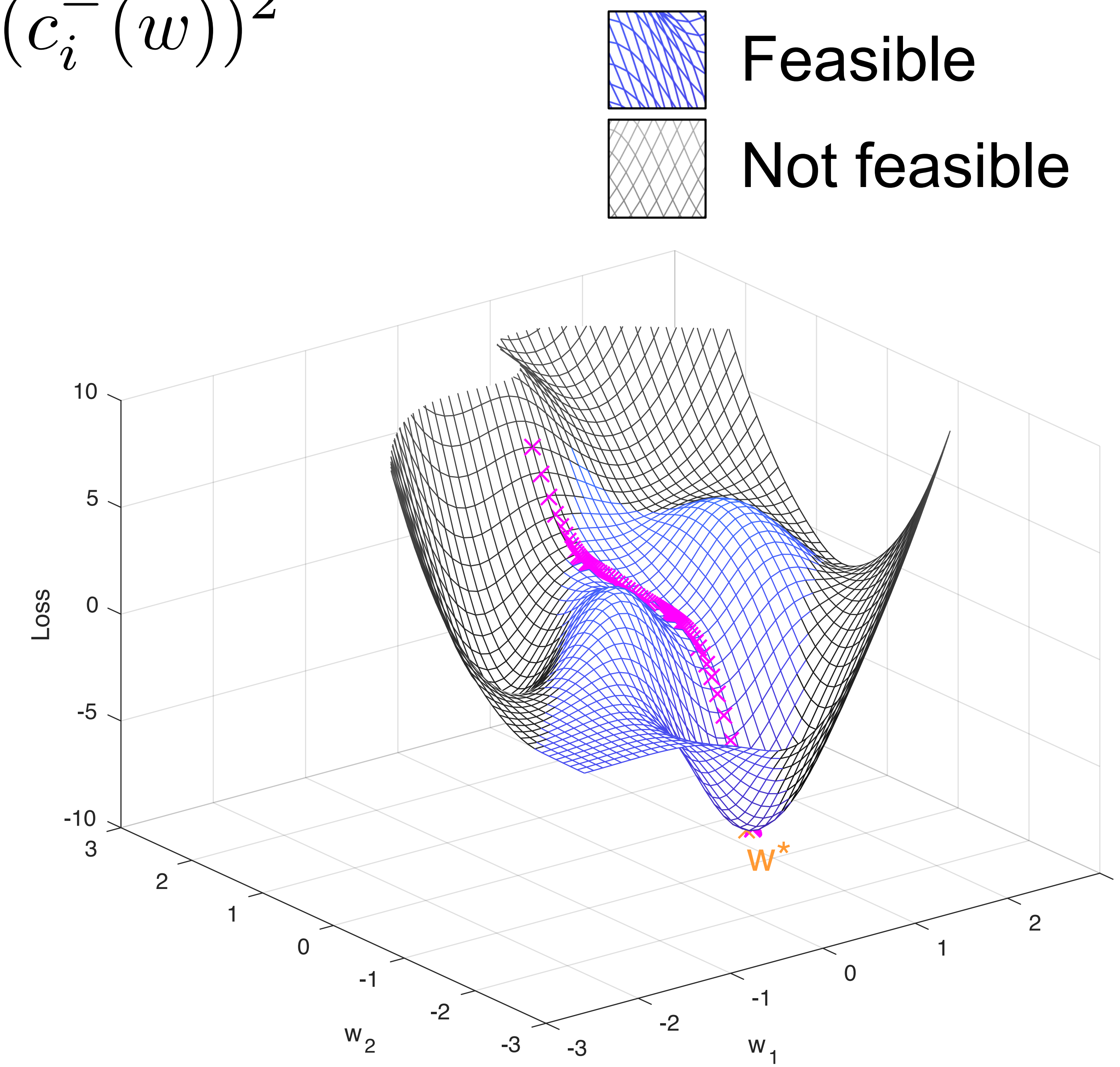


II. Constrained optimisation

Penalty Method

$$\overline{w} = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w) + \rho \sum_{i \in \mathcal{I}} (c_i^-(w))^2$$

$$\rho = 3$$

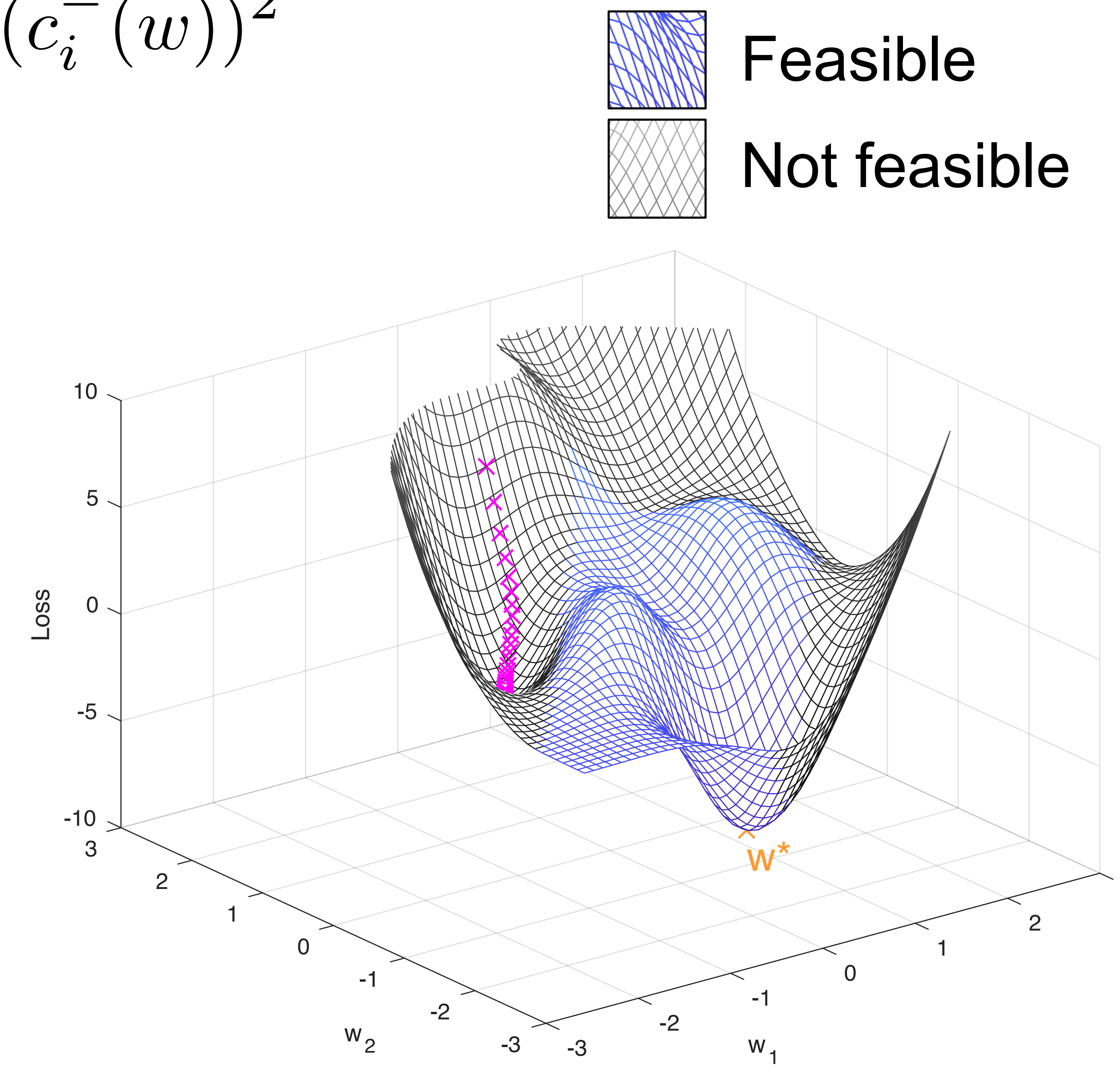


II. Constrained optimisation

Penalty Method

$$\overline{w} = \operatorname*{argmin}_{w \in \mathbb{R}^n} \quad \mathcal{L}(w) + \rho \sum_{i \in \mathcal{I}} (c_i^-(w))^2$$

$$\rho = 3$$



Requirement for feasible solution:

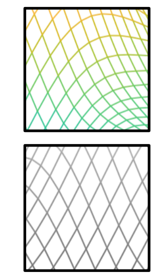
$\rho \to \infty$

Ill-conditioned

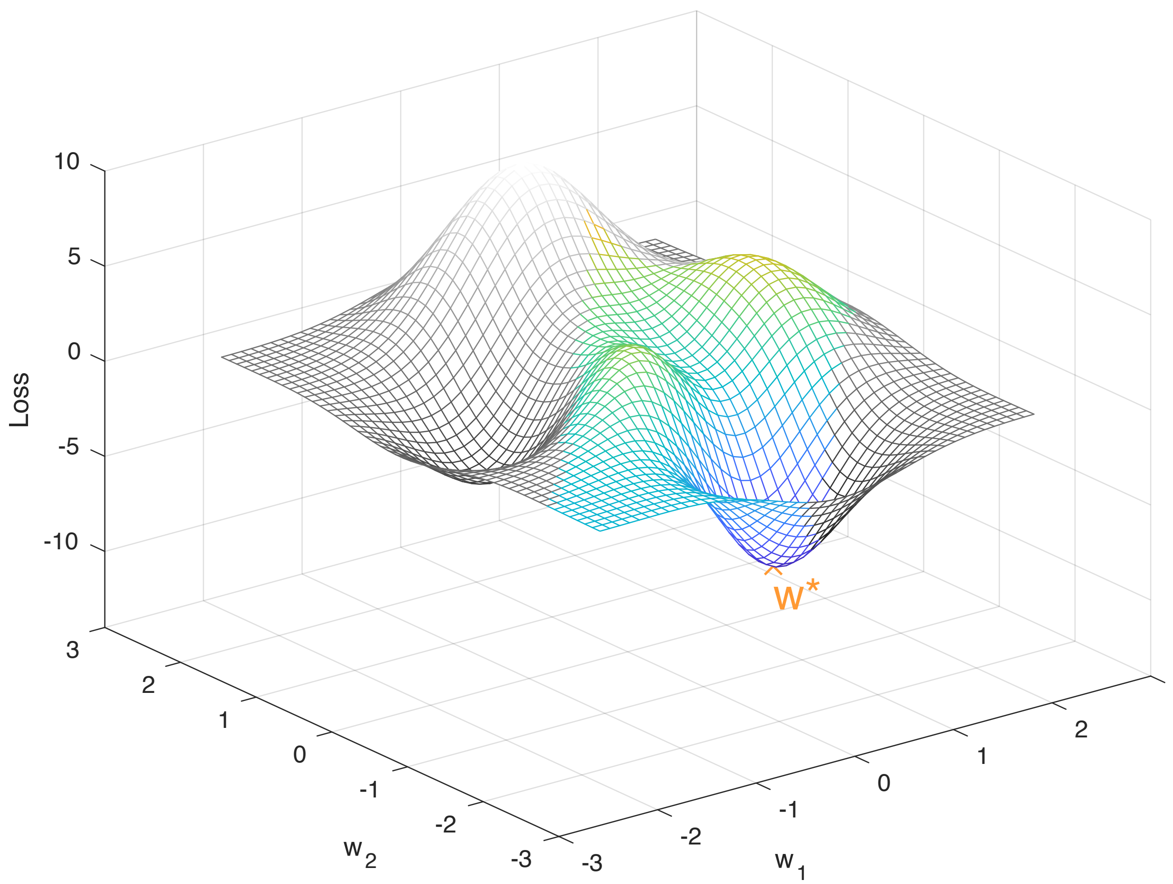
II. Constrained optimisation

Lagrangian Relaxation

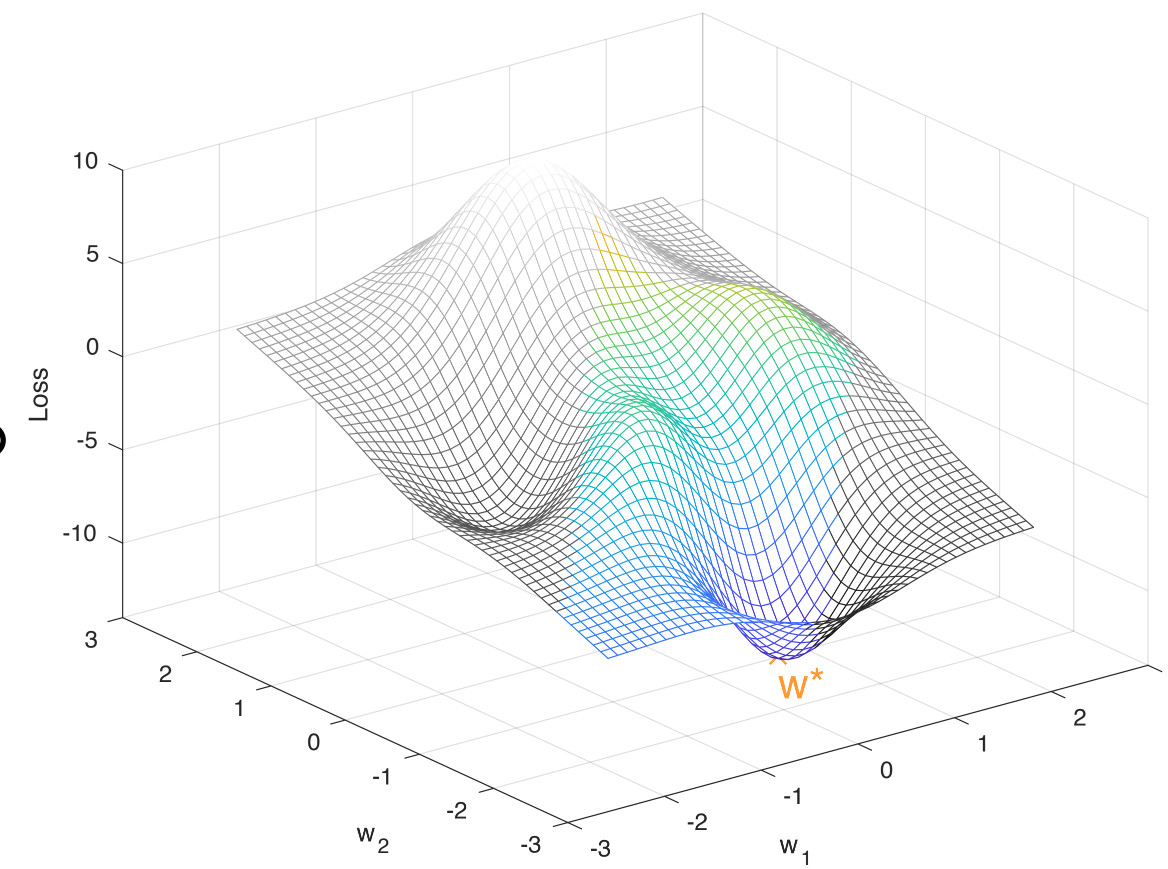
$\lambda_i, i \in \mathcal{I}$: Lagrangian multipliers


Feasible
Not feasible

$$\mathcal{L}'(w, \lambda) = \mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i c_i(w), \qquad \lambda \geq 0$$



$\mathcal{L}(w)$



$\mathcal{L}'(w, \lambda), \ \lambda = (1, 1, 1)$

II. Constrained optimisation

Lagrangian Relaxation

$\lambda_i, i \in \mathcal{I}$: Lagrangian multipliers

$$\mathcal{L}'(w, \lambda) = \mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i c_i(w), \qquad \lambda \geq 0$$

Lagrangian Dual: $\max\limits_{\lambda} \min\limits_{w \in \mathbb{R}^n} \mathcal{L}'(w, \lambda)$

⇨ **Augmented Lagrangian Method**

II. Constrained optimisation

Augmented Lagrangian Method

$w^* = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w)$ subject to $\quad c_i(w) = 0, \; i \in \mathcal{E}$	$\mathcal{E}$: Set of equality constraint indices $c_i : \mathbb{R}^n \to \mathbb{R}, \; i \in \mathcal{E}$: Equality constraints

For $k = 0, 1, \ldots$

$$w^k = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \left(\mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i^k \, c_i(w) + \frac{1}{2} \mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2 \right)$$

$$\lambda_i^{k+1} = \lambda_i^k - \mu_k \, c_i(w^k) \qquad \forall i \in \mathcal{E}$$

II. Constrained optimisation

Augmented Lagrangian Method

$w^* = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w)$ subject to $\quad c_i(w) = 0, \; i \in \mathcal{E}$	$\mathcal{E}$: Set of equality constraint indices $c_i : \mathbb{R}^n \to \mathbb{R}, \; i \in \mathcal{E}$: Equality constraints

$\lambda_i^k, i \in \mathcal{E}$: Lagrangian multiplier estimates

$\{\mu_k\}_{k=0}^{k=\infty}$: Increasing series of penalty weights

For $k = 0, 1, \ldots$

$$w^k = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \left(\mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i^k \, c_i(w) + \frac{1}{2} \mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2 \right)$$

$$\lambda_i^{k+1} = \lambda_i^k - \mu_k \, c_i(w^k) \qquad \forall i \in \mathcal{E}$$

II. Constrained optimisation

Augmented Lagrangian Method

$$w^* = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w)$$

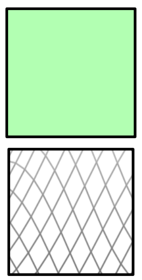
$$\text{subject to} \quad c_i(w) = 0, \ i \in \mathcal{E}$$

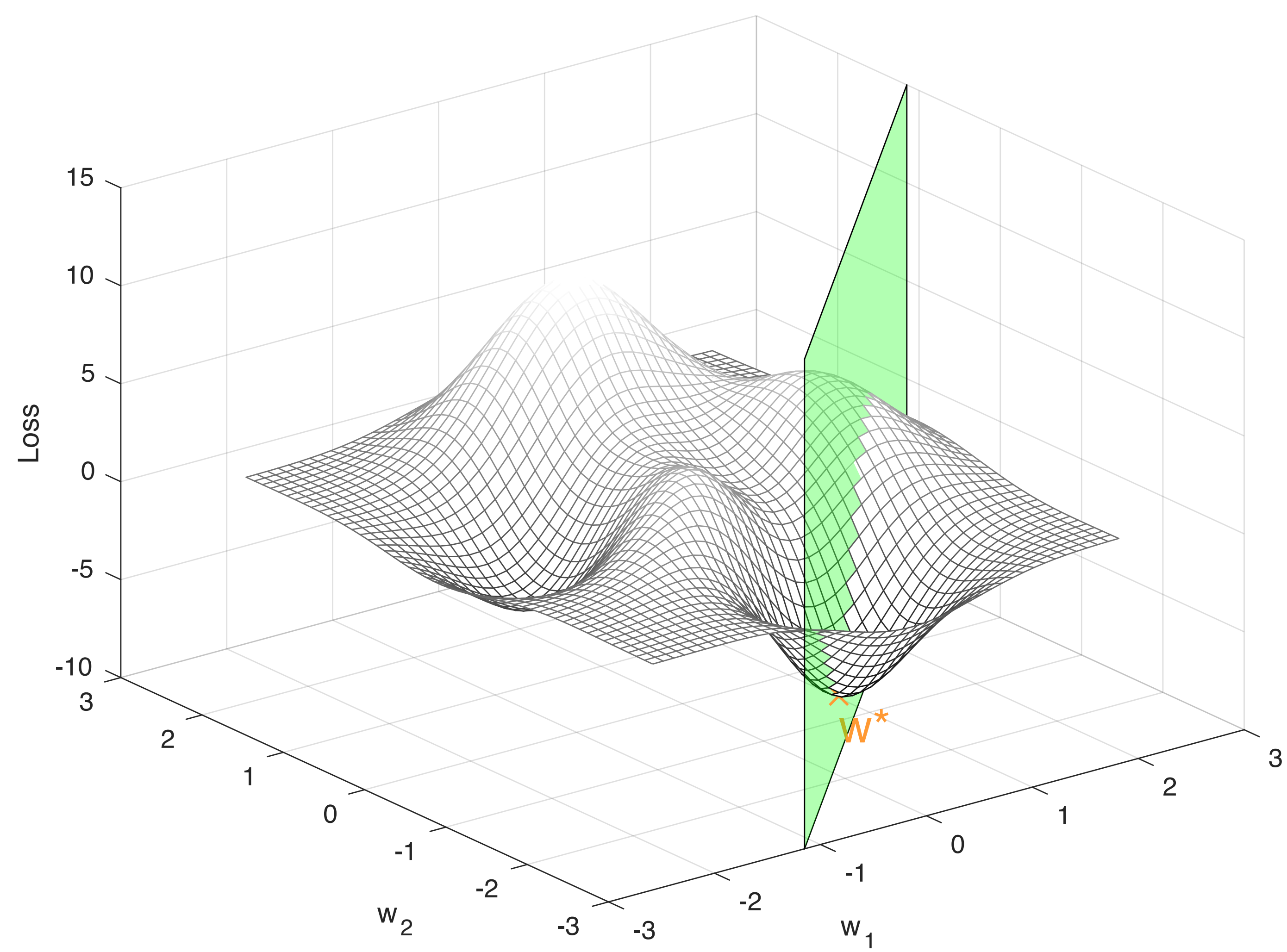
For $k = 0, 1, \ldots$

$$w^k = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \ \Big(\mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i^k \, c_i(w) + \frac{1}{2} \mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2 \Big)$$

$$\lambda_i^{k+1} = \lambda_i^k - \mu_k \, c_i(w^k) \qquad \forall i \in \mathcal{E}$$

k	0	1	2	3	4
μ_k					
λ_1^k					
$c_1(w^k)$					





II. Constrained optimisation

Augmented Lagrangian Method

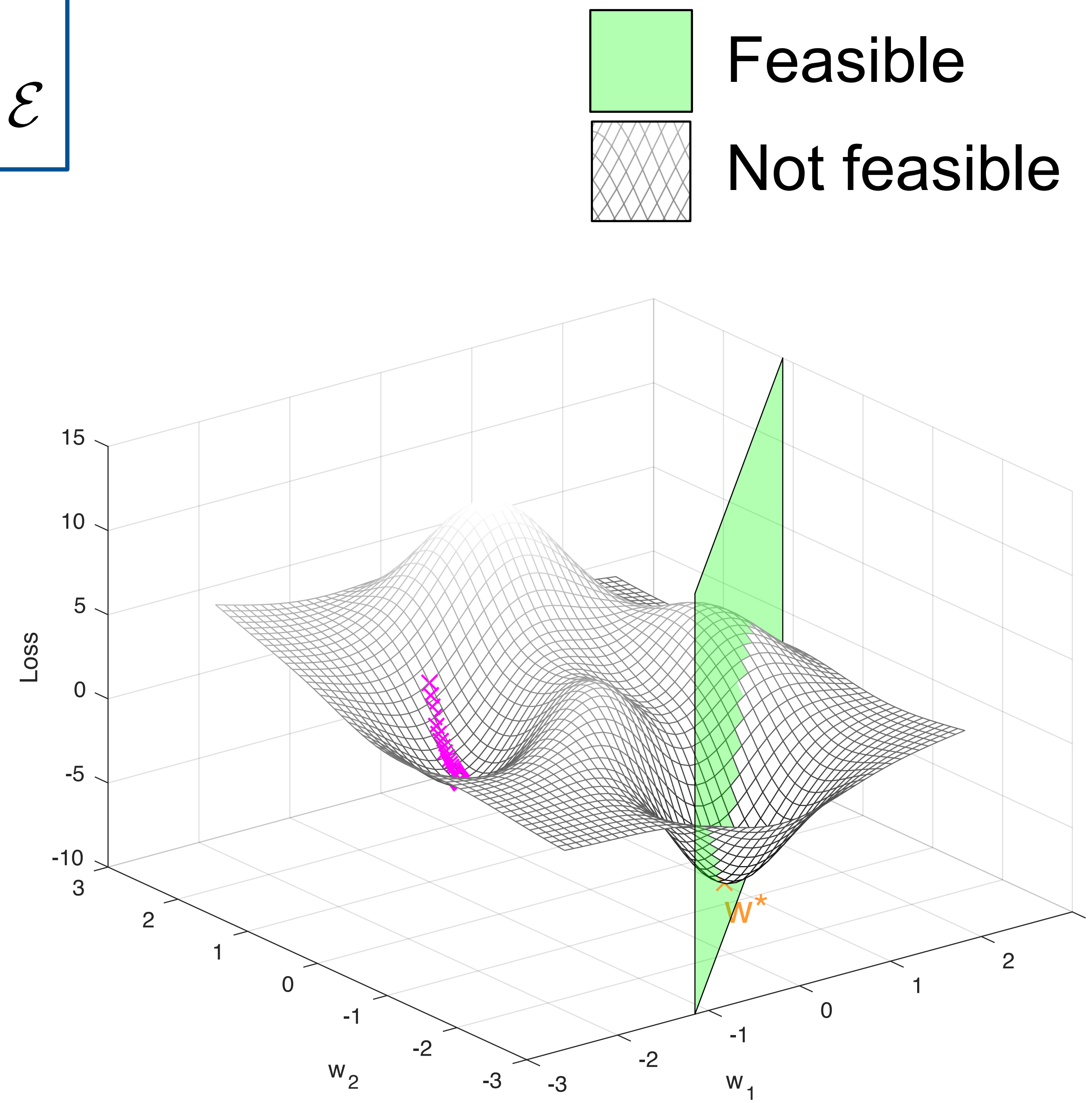
$$w^* = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w)$$
$$\text{subject to} \quad c_i(w) = 0, \;\; i \in \mathcal{E}$$

For $k = 0, 1, \ldots$

$$w^k = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \Big(\mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i^k \, c_i(w) + \frac{1}{2}\mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2\Big)$$

$$\lambda_i^{k+1} = \lambda_i^k - \mu_k \, c_i(w^k) \qquad \forall i \in \mathcal{E}$$

k	0	1	2	3	4
μ_k	0,125				
λ_1^k	0				
$c_1(w^k)$	-3,30				



II. Constrained optimisation

Augmented Lagrangian Method

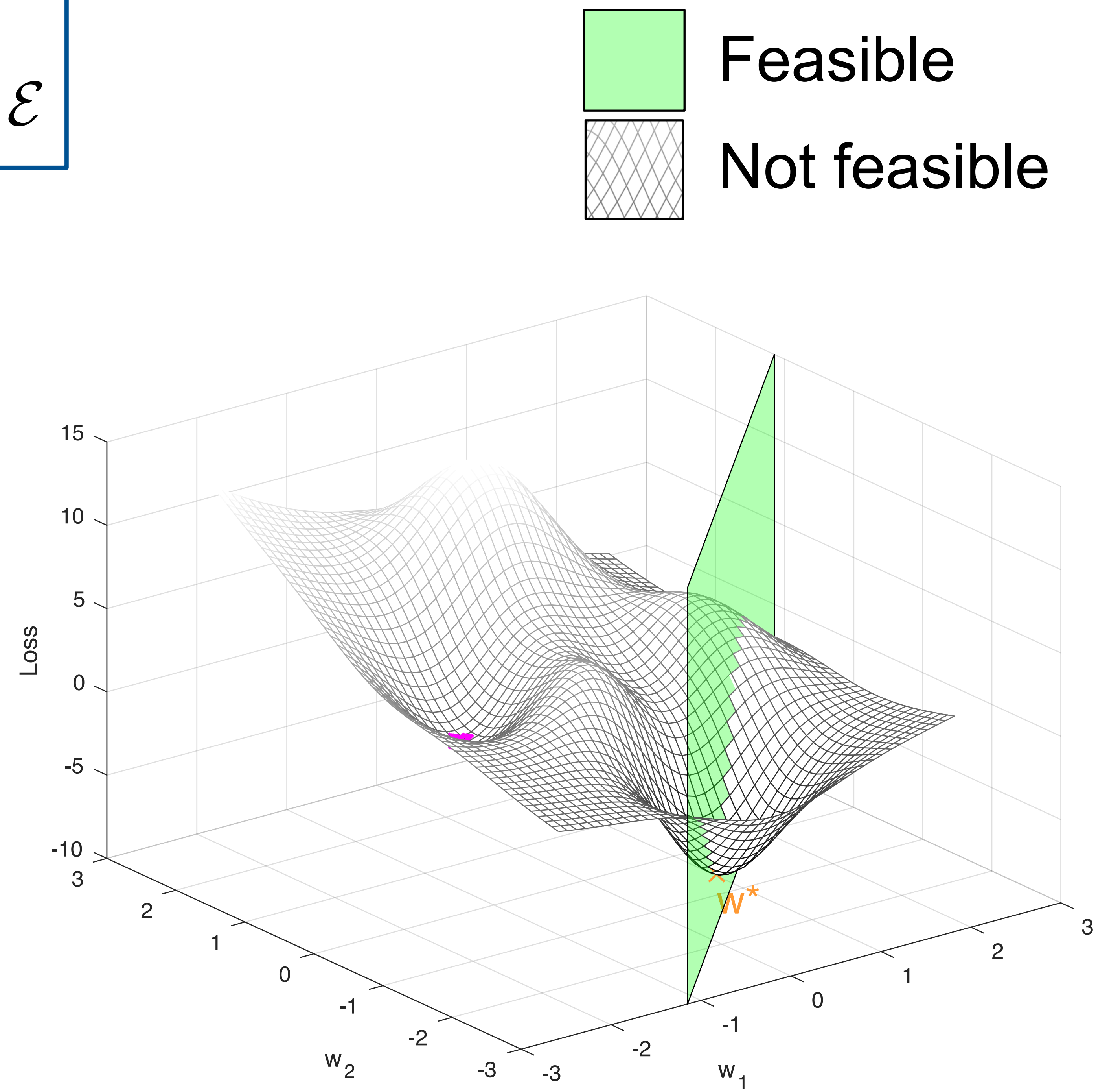
$$w^* = \operatorname*{argmin}_{w \in \mathbb{R}^n} \quad \mathcal{L}(w)$$
$$\text{subject to} \quad c_i(w) = 0, \ i \in \mathcal{E}$$

For $k = 0, 1, \ldots$

$$w^k = \operatorname*{argmin}_{w \in \mathbb{R}^n} \Big(\mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i^k \, c_i(w) + \frac{1}{2}\mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2\Big)$$

$$\lambda_i^{k+1} = \lambda_i^k - \mu_k \, c_i(w^k) \qquad \forall i \in \mathcal{E}$$

k	0	1	2	3	4
μ_k	0,125	0,25			
λ_1^k	0	0,41			
$c_1(w^k)$	-3,30	-3,19			



II. Constrained optimisation

Augmented Lagrangian Method

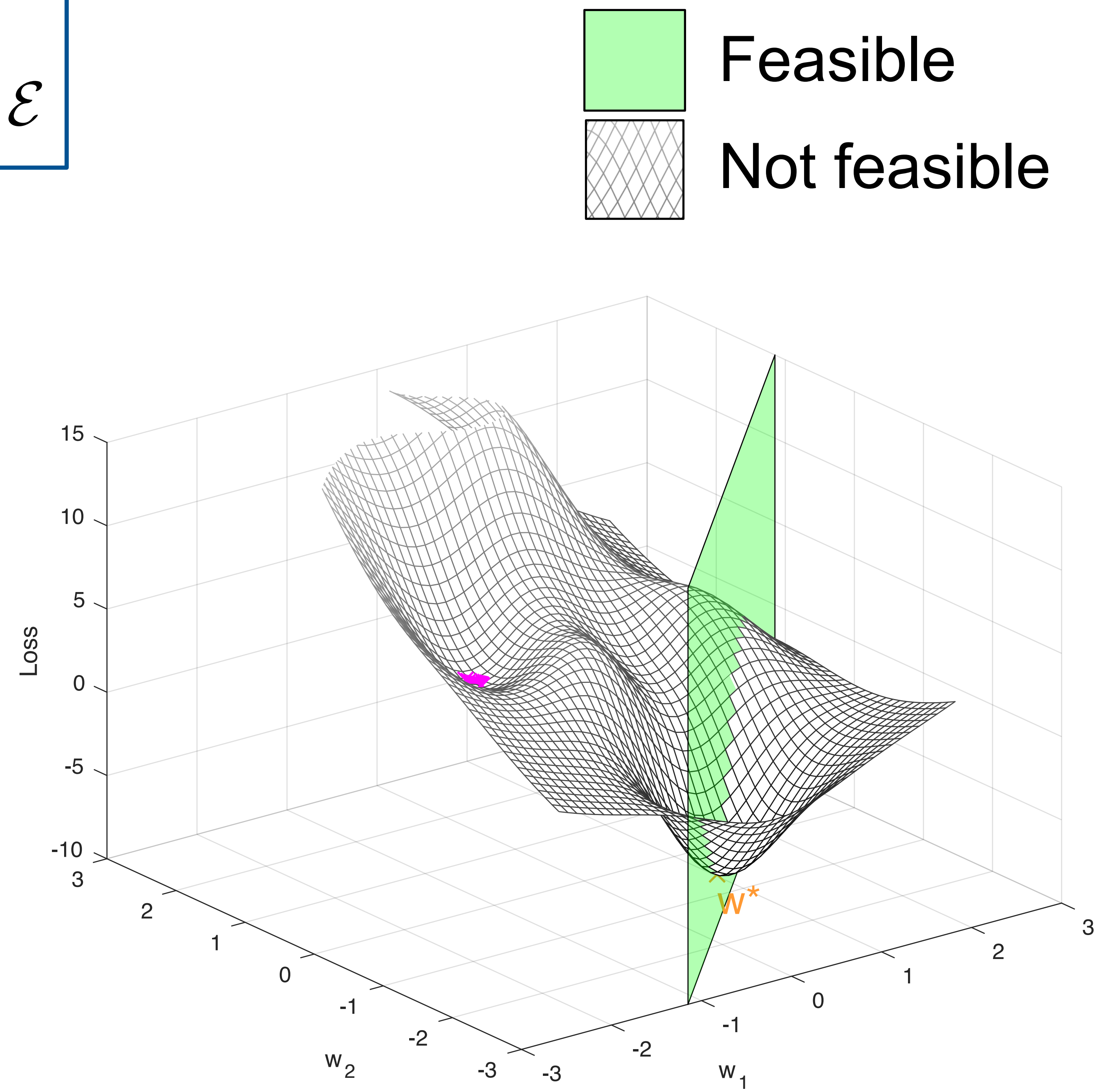
$$
\begin{aligned}
w^* = \; & \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w) \\
& \text{subject to} \quad c_i(w) = 0, \;\; i \in \mathcal{E}
\end{aligned}
$$

For $k = 0, 1, \ldots$

$$
\begin{aligned}
w^k = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \; \Big(\mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i^k \, c_i(w) \\
+ \frac{1}{2} \mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2 \Big)
\end{aligned}
$$

$$
\lambda_i^{k+1} = \lambda_i^k - \mu_k \, c_i(w^k) \qquad \forall i \in \mathcal{E}
$$

k	0	1	2	3	4
μ_k	0,125	0,25	0,5		
λ_1^k	0	0,41	1,21		
$c_1(w^k)$	-3,30	-3,19	-2,98		

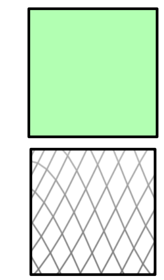


II. Constrained optimisation

Augmented Lagrangian Method

$$w^* = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w)$$

$$\text{subject to} \quad c_i(w) = 0, \;\; i \in \mathcal{E}$$

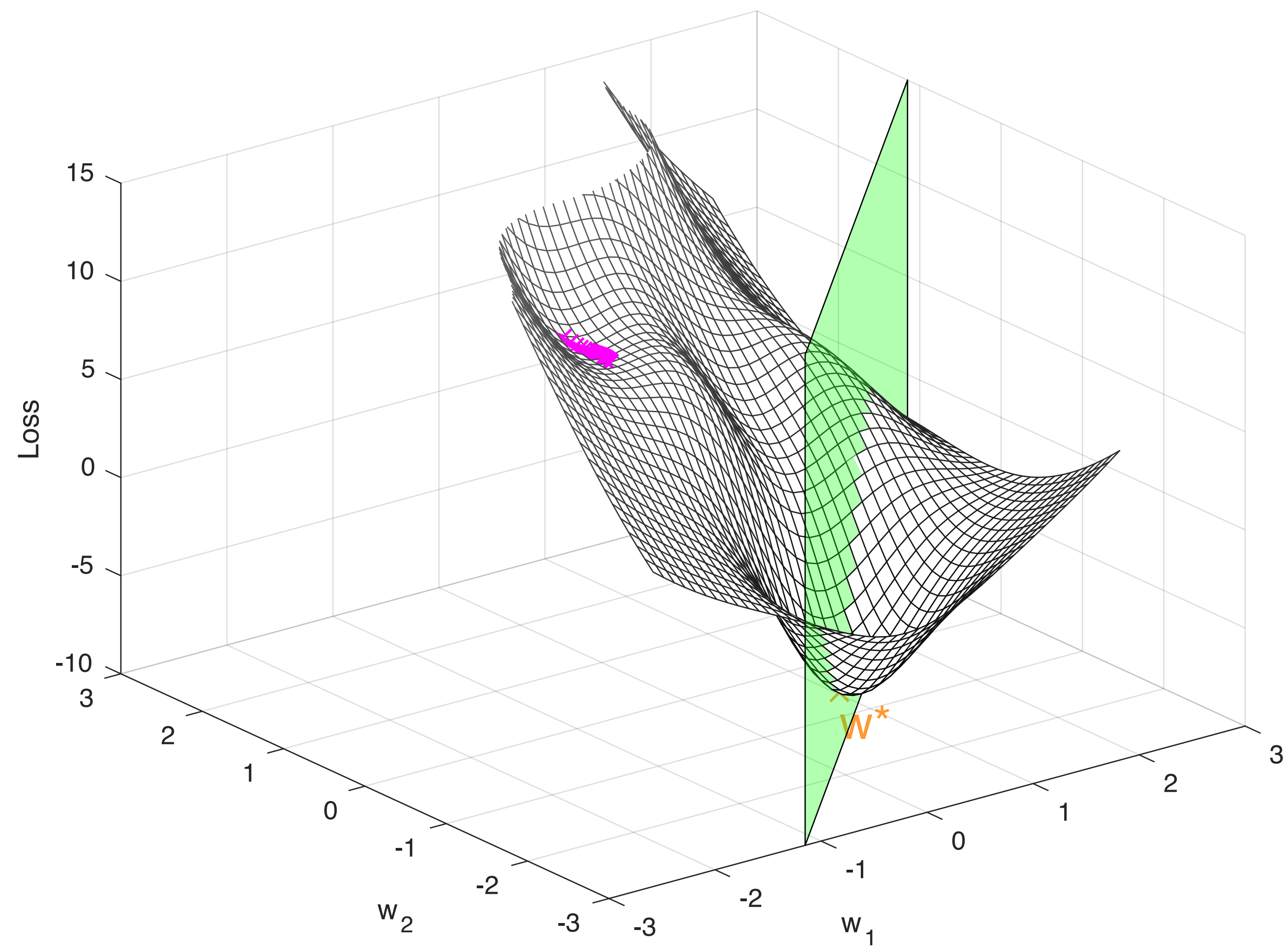


For $k = 0, 1, \ldots$

$$w^k = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \left(\mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i^k \, c_i(w) + \frac{1}{2} \mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2 \right)$$

$$\lambda_i^{k+1} = \lambda_i^k - \mu_k \, c_i(w^k) \qquad \forall i \in \mathcal{E}$$

k	0	1	2	3	4
μ_k	0,125	0,25	0,5	1	
λ_1^k	0	0,41	1,21	2,70	
$c_1(w^k)$	-3,30	-3,19	-2,98	-2,45	



II. Constrained optimisation

Augmented Lagrangian Method

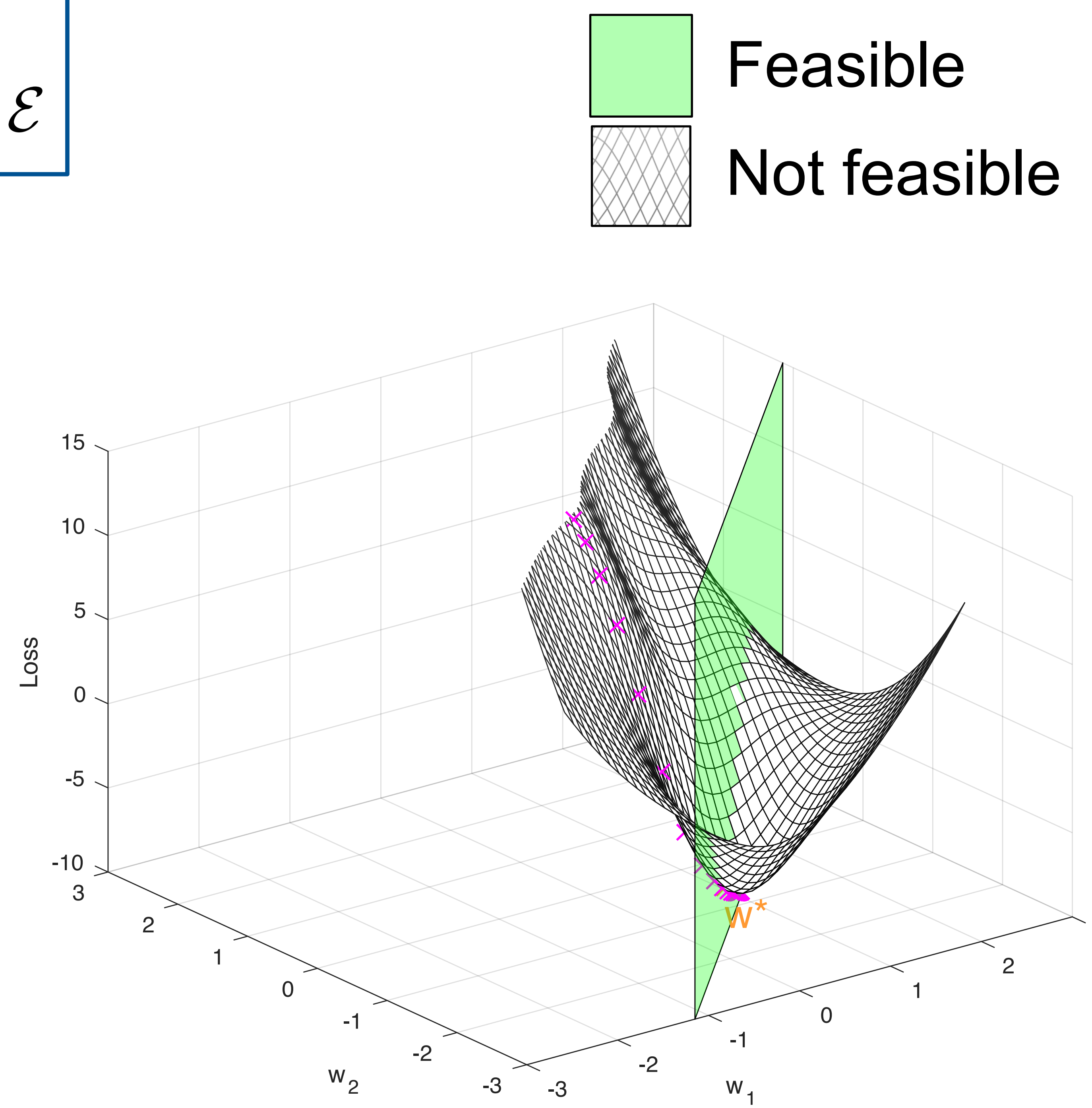
$$\begin{aligned} w^* = \; & \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} && \mathcal{L}(w) \\ & \text{subject to} && c_i(w) = 0, \;\; i \in \mathcal{E} \end{aligned}$$

For $k = 0, 1, \ldots$

$$w^k = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \Big(\mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i^k \, c_i(w) + \frac{1}{2} \mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2 \Big)$$

$$\lambda_i^{k+1} = \lambda_i^k - \mu_k \, c_i(w^k) \qquad \forall i \in \mathcal{E}$$

k	0	1	2	3	4
μ_k	0,125	0,25	0,5	1	2
λ_1^k	0	0,41	1,21	2,70	5,15
$c_1(w^k)$	-3,30	-3,19	-2,98	-2,45	-0,20



II. Constrained optimisation

Augmented Lagrangian Method

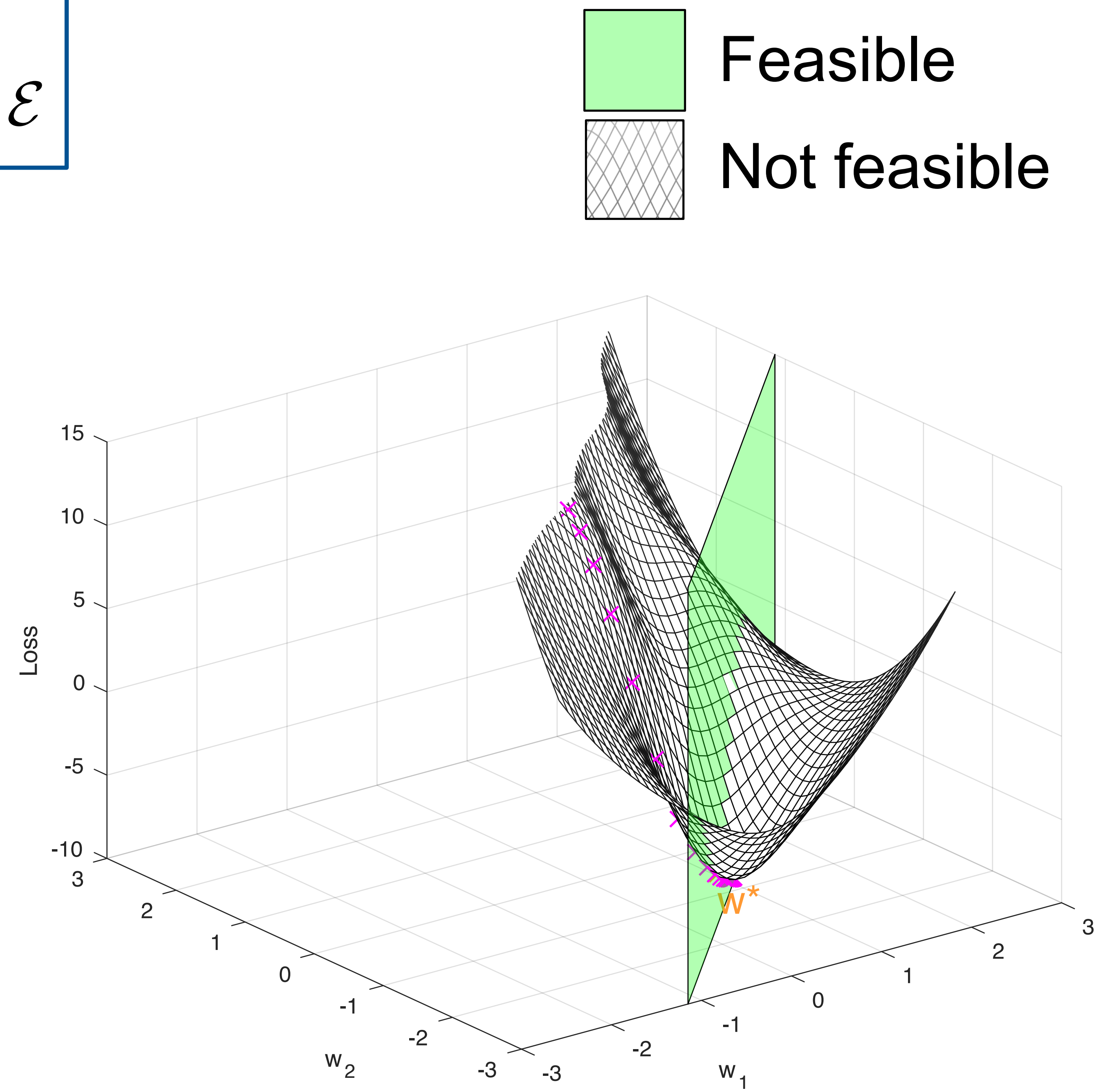
$$
\begin{aligned}
w^* = \quad & \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \quad \mathcal{L}(w) \\
& \text{subject to} \quad c_i(w) = 0, \;\; i \in \mathcal{E}
\end{aligned}
$$

For $k = 0, 1, \ldots$

$$
\begin{aligned}
w^k = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \; \Big(\mathcal{L}(w) - \sum_{i \in \mathcal{I}} \lambda_i^k \, c_i(w) \\
+ \frac{1}{2} \mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2 \Big)
\end{aligned}
$$

$$
\lambda_i^{k+1} = \lambda_i^k - \mu_k \, c_i(w^k) \qquad \forall i \in \mathcal{E}
$$

Convergence already with limited μ_k.



III. Experiment

Choosing a suitable experiment

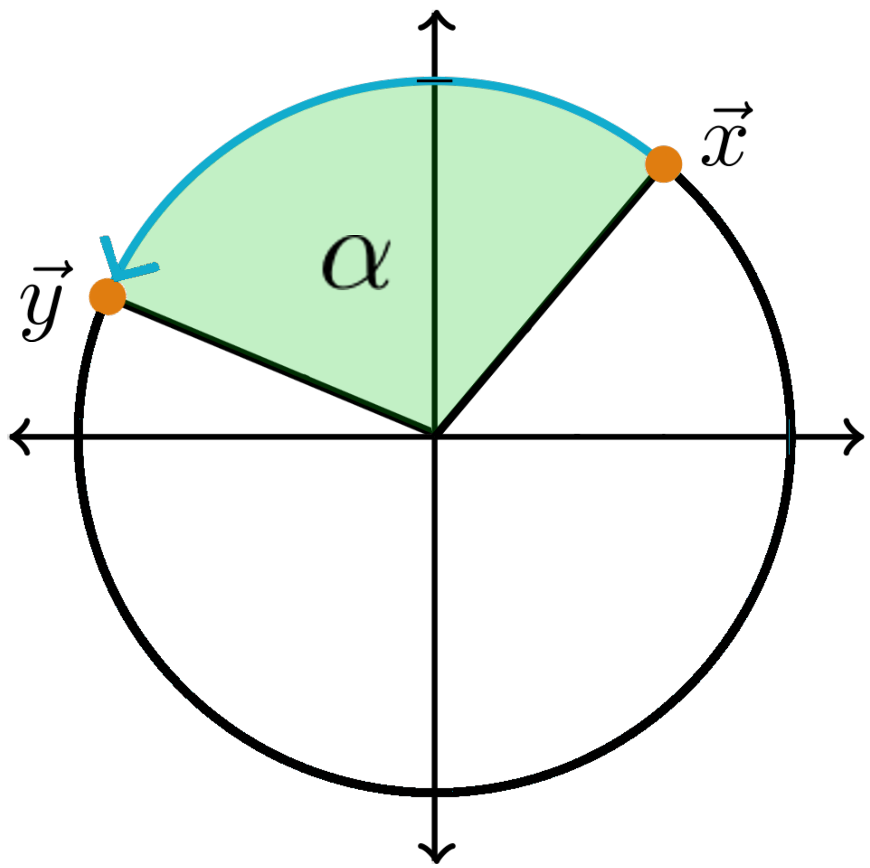
Goals:

- Predictions aligning with physical constraints
- Decrease required amount of learning data

⇨ **Rotation**

III. Experiment

Rotation



$$rot_{2D}(\vec{x},\alpha) = \underbrace{\begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}}_{R(\alpha)} \vec{x}$$

III. Experiment

Physical constraints

Determinant constraint:

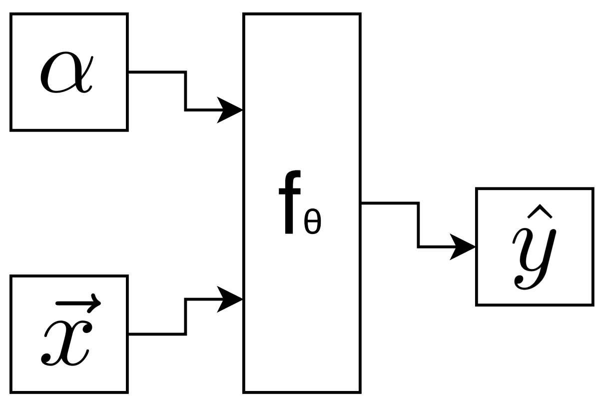
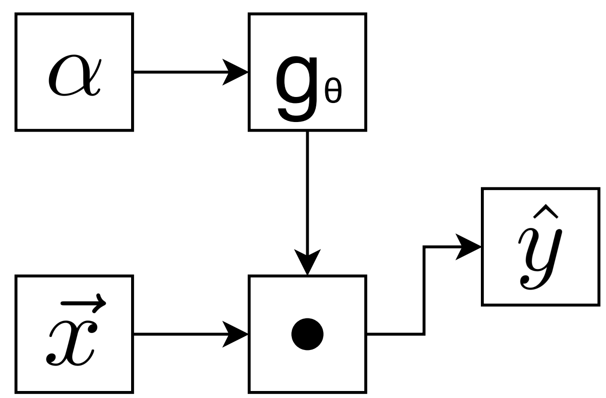
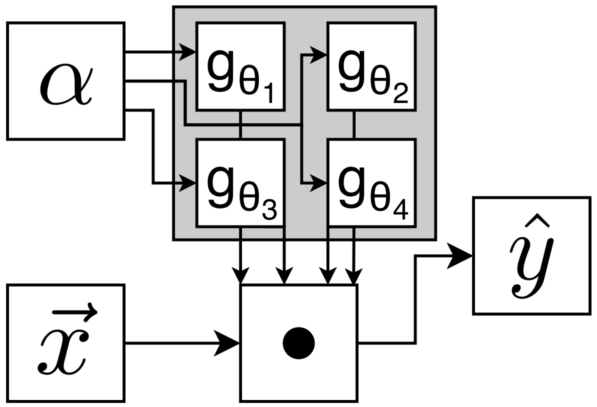
$$\det(R(\alpha)) = 1, \qquad \forall \alpha \in [-\pi, \pi]$$

Norm constraint:

$$||R(\alpha)\vec{x}||_2 = ||\vec{x}||_2, \qquad \forall \alpha \in [-\pi, \pi], \vec{x} \in \mathbb{R}^2$$

III. Experiment

Network architectures

	Point prediction	**Matrix prediction**	**Matrix entry prediction**
	Model 1	Model 2	Model 3
Structure			
Network function	$f_\theta : \mathbb{R}^3 \to \mathbb{R}^2$	$g_\theta : \mathbb{R} \to \mathbb{R}^{2\times 2}$	$g_{\theta_i} : \mathbb{R} \to \mathbb{R}, i \in \{0, 1, 2, 3\}$
Layer sizes	$3 \to 16 \to 16 \to 16 \to 2$	$1 \to 100 \to 4$	$[1 \to 50 \to 1] \times 4$
Activation	Tanh	Sigmoid	Sigmoid
Parameters	642	604	604

III. Experiment

Comparison criteria

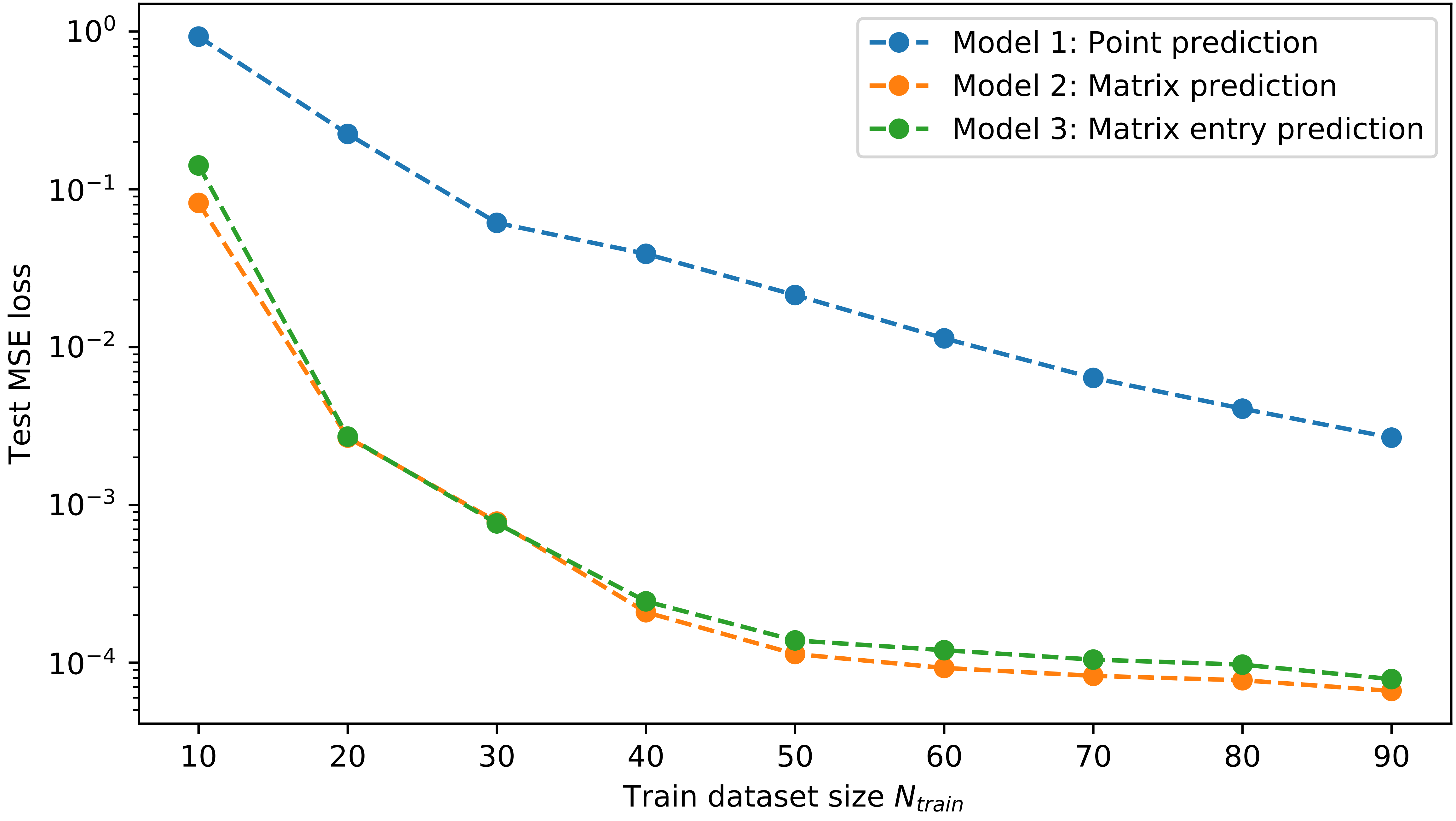
- Available data is currently the greatest limitation for deep learning
- Hyperparameters can be tuned

⇨ **Evaluate performance depending on amount of available training data**

IV. Results

Model baseline performance

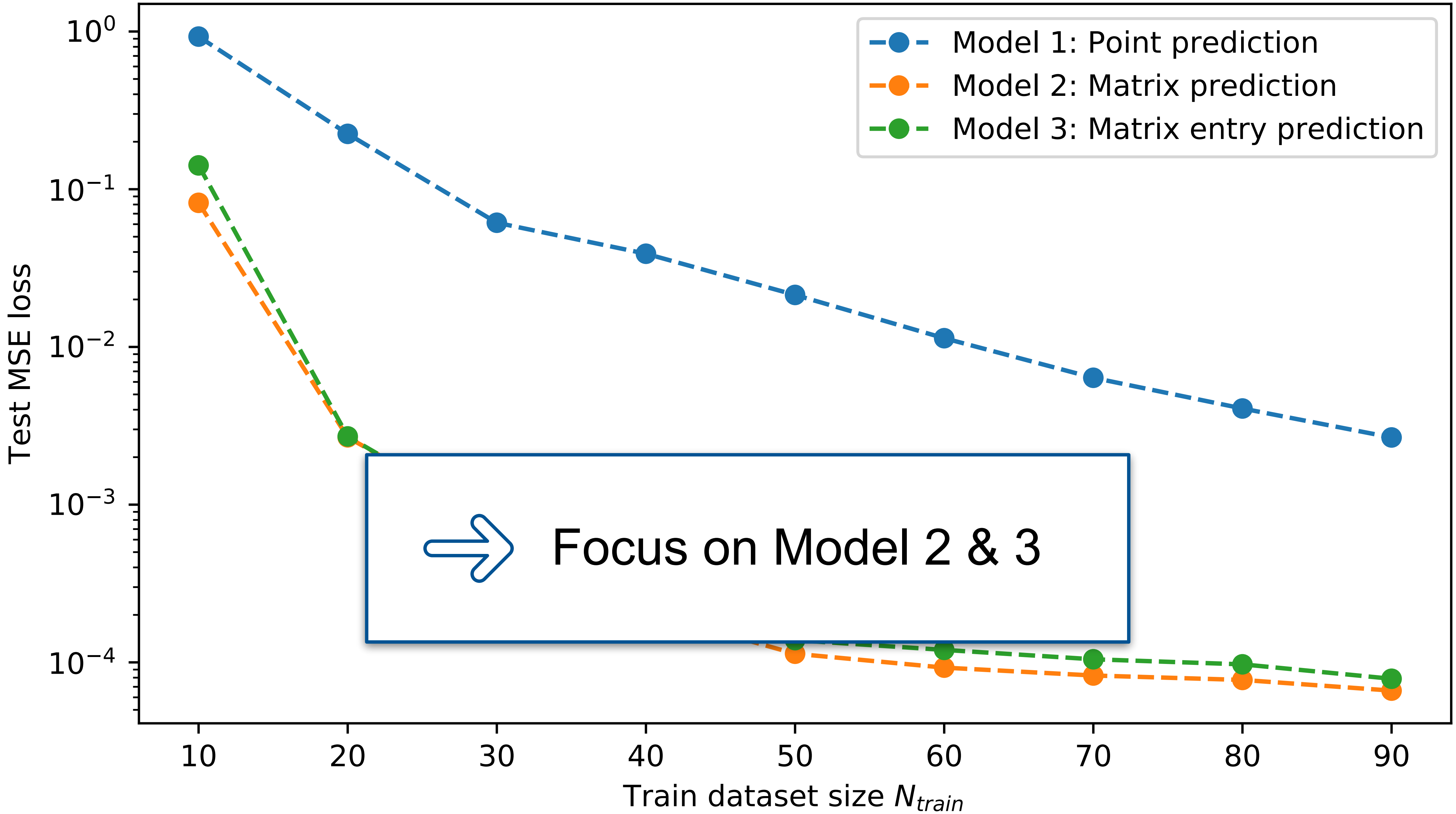
$$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y)$$



IV. Results

Model baseline performance

$$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y)$$

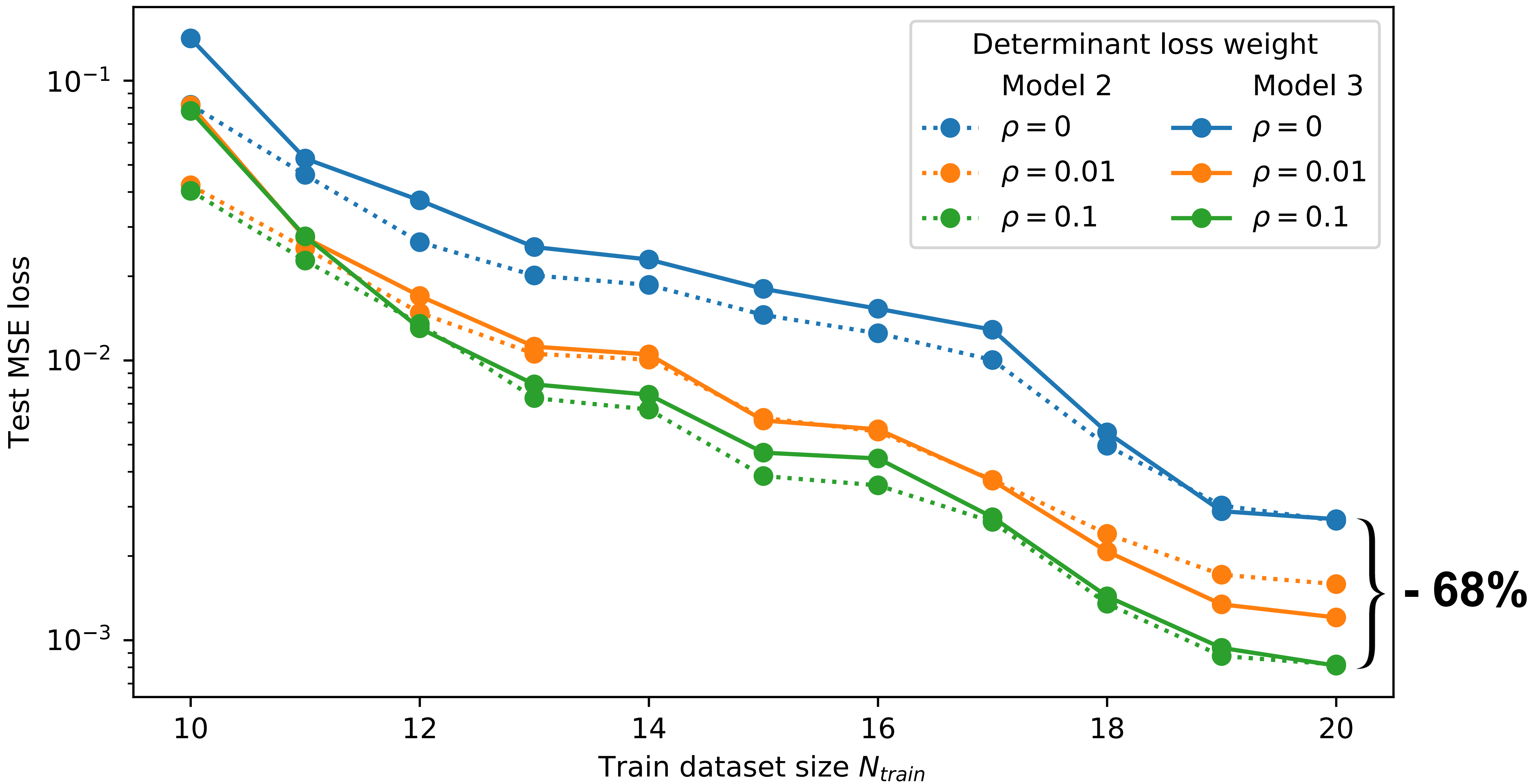


IV. Results

Penalty Method - Determinant constraint

$$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y) + \rho(\det(\hat{R}_w) - 1)^2$$

$\hat{R}_w$: Predicted rotation matrices

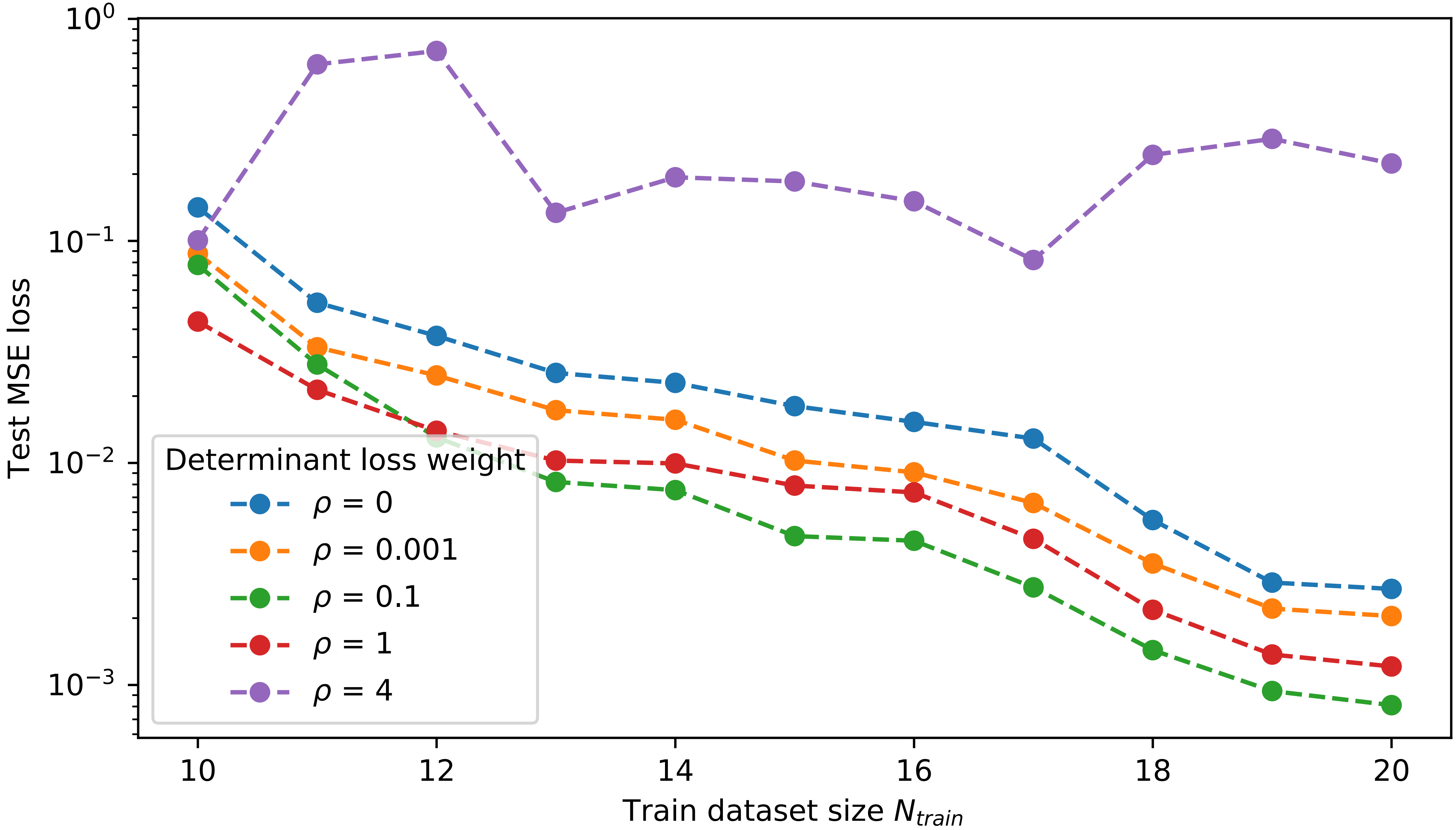


IV. Results

Penalty Method - Determinant constraint - Model 3

$$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y) + \rho(\det(\hat{R}_w) - 1)^2$$

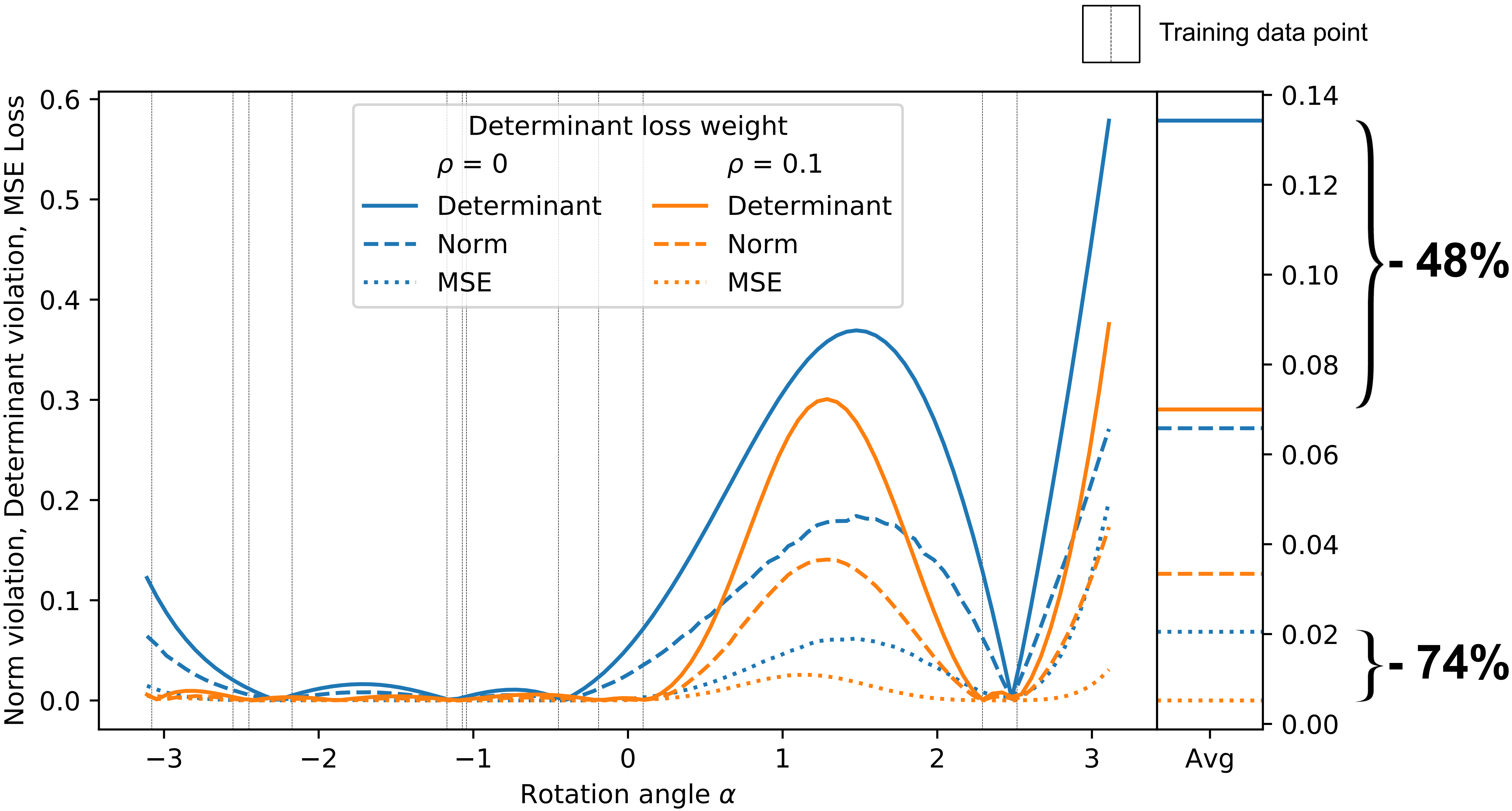
$\hat{R}_w$: Predicted rotation matrices



IV. Results

Penalty Method - Determinant constraint - Model 3

Result of Model 3 trained on 12 training data points

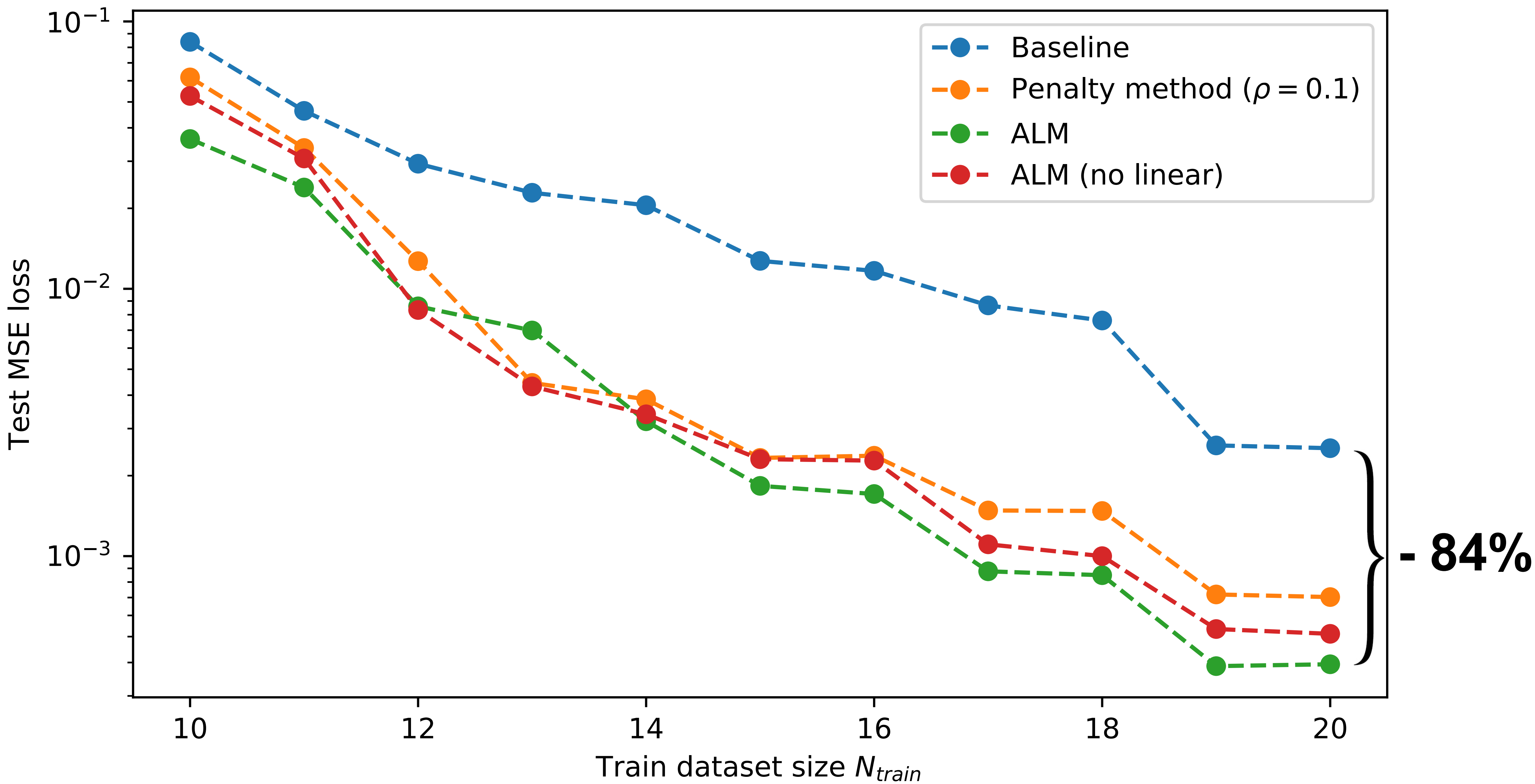


IV. Results

Augmented Lagrangian Method - Determinant constraint

Model 3

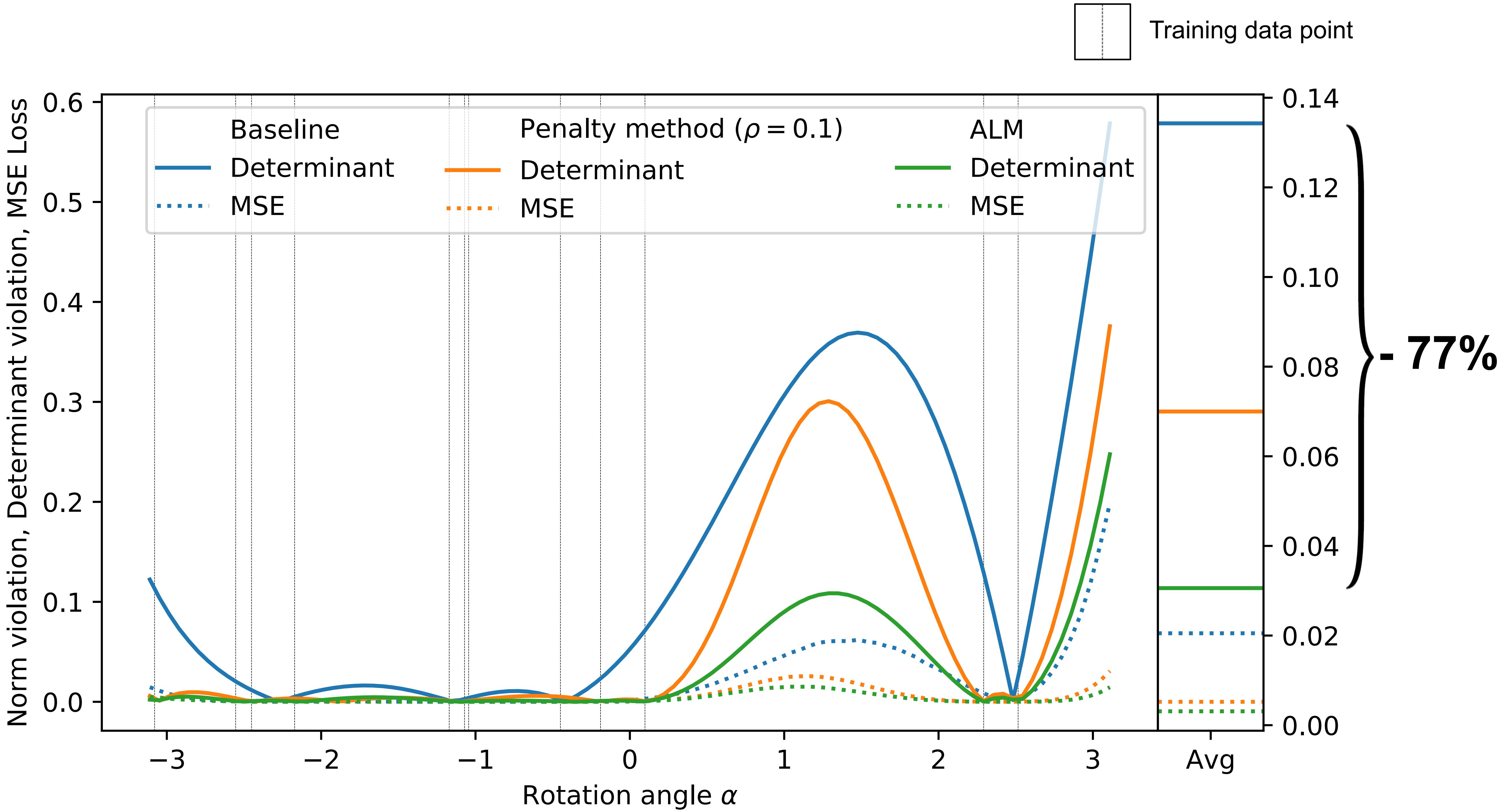
Baseline	$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y)$
Penalty Method	$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y) + \rho(\det(\hat{R}_w) - 1)^2$
ALM	$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y) - \lambda^k(\det(\hat{R}_w) - 1) + \frac{1}{2}\mu_k(\det(\hat{R}_w) - 1)^2$
ALM (no linear)	$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y) + \frac{1}{2}\mu_k(\det(\hat{R}_w) - 1)^2$



IV. Results

Augmented Lagrangian Method - Determinant constraint

Result of Model 3 trained on 12 training data points



IV. Results

Failed experiments

- Norm constraint

- Physical projection:
  - Scale rotation matrices with positive determinants
  - Project predictions to unit circle

IV. Results

Conclusion

Incorporation of physical constraints:
- Improves performance
- Predictions align better with physical constraints
- Less training data required

Penalty Method:
- Easy to implement
- Robust to choice of weight

Augmented Lagrangian Method:
- Many training iterations required
- Significant performance improvements

IV. Results

Outlook

- Validation of improvements on real world applications (e.g. Facial Keypoint Detection)
- Unsupervised physical training
- Check robustness to unbalanced datasets and noise

IV. Results

Github

Results and code to reproduce the experiments can be found at
https://github.com/KorbinianBstrtr/BAThesis_abstreik

References

- Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second edition, 2006.
- Jack Yurkiewicz. Constrained optimization and lagrange multiplier methods, by d. p. bertsekas, academic press, new york, 1982, 395 pp. price: $65.00. *Networks*, 15:138–140, 1985.
- Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*, pages 215–272. Cambridge University Press, New York, NY, USA, 2004.
- Yuxin Chen. Lecture notes in ele 538b: Large-scale optimization for data science, Spring 2018.

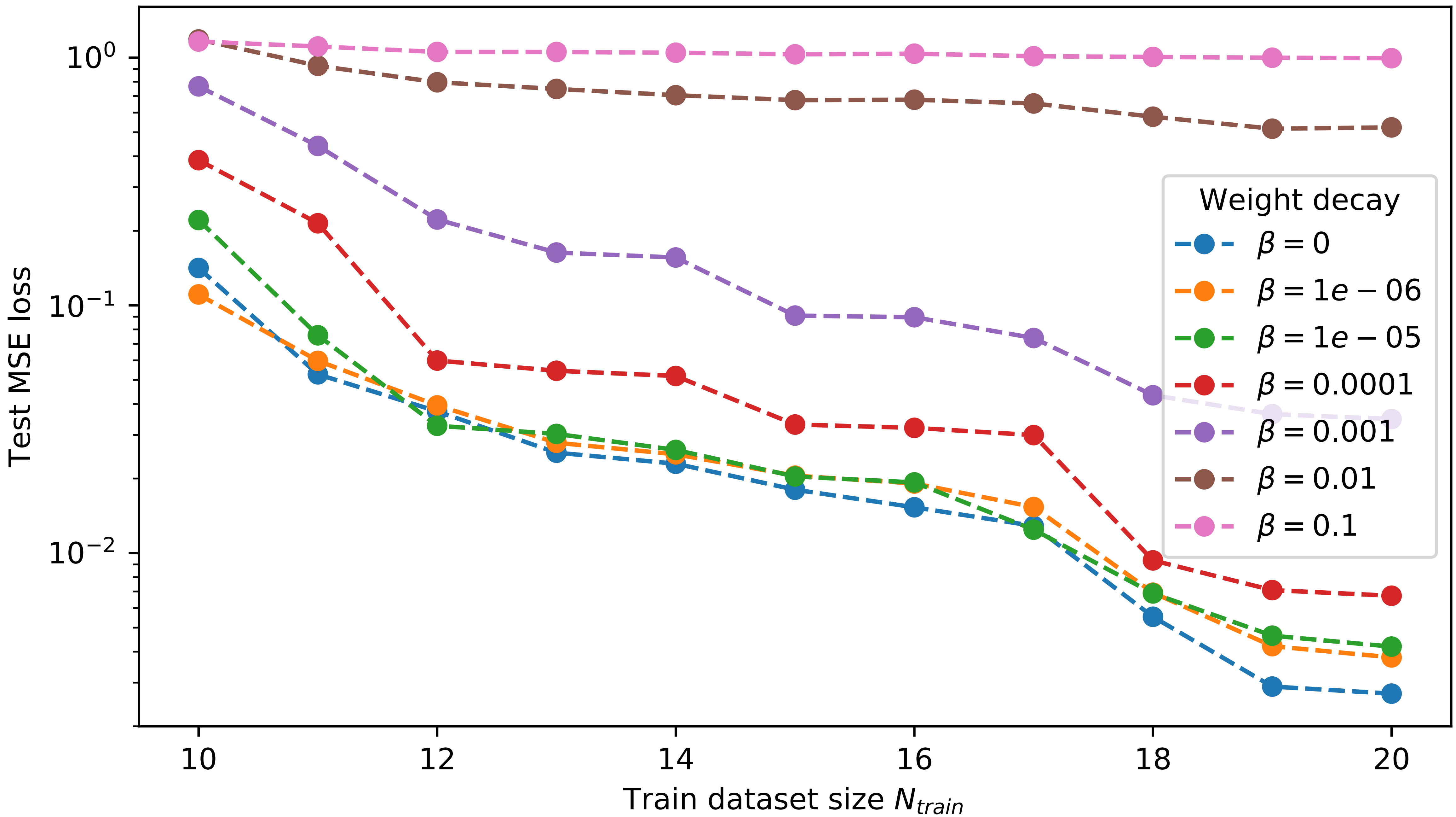
Images taken from:

- https://www.br.de/telekolleg/faecher/physik/trimester2/schiefer-wurf-102~_v-img__16__9__m_-4423061158a17f4152aef84861ed0243214ae6e7.jpg?version=e496f
- https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRB_oBff0NlA087mkfhKQ9dujNPhQYjNOEoo3drYeuclw7Ov0HRfQ
- http://pedynamic.com/circle-antipode-experiments/
- http://pedynamic.com/circle-antipode-simulation/

V. Appendix

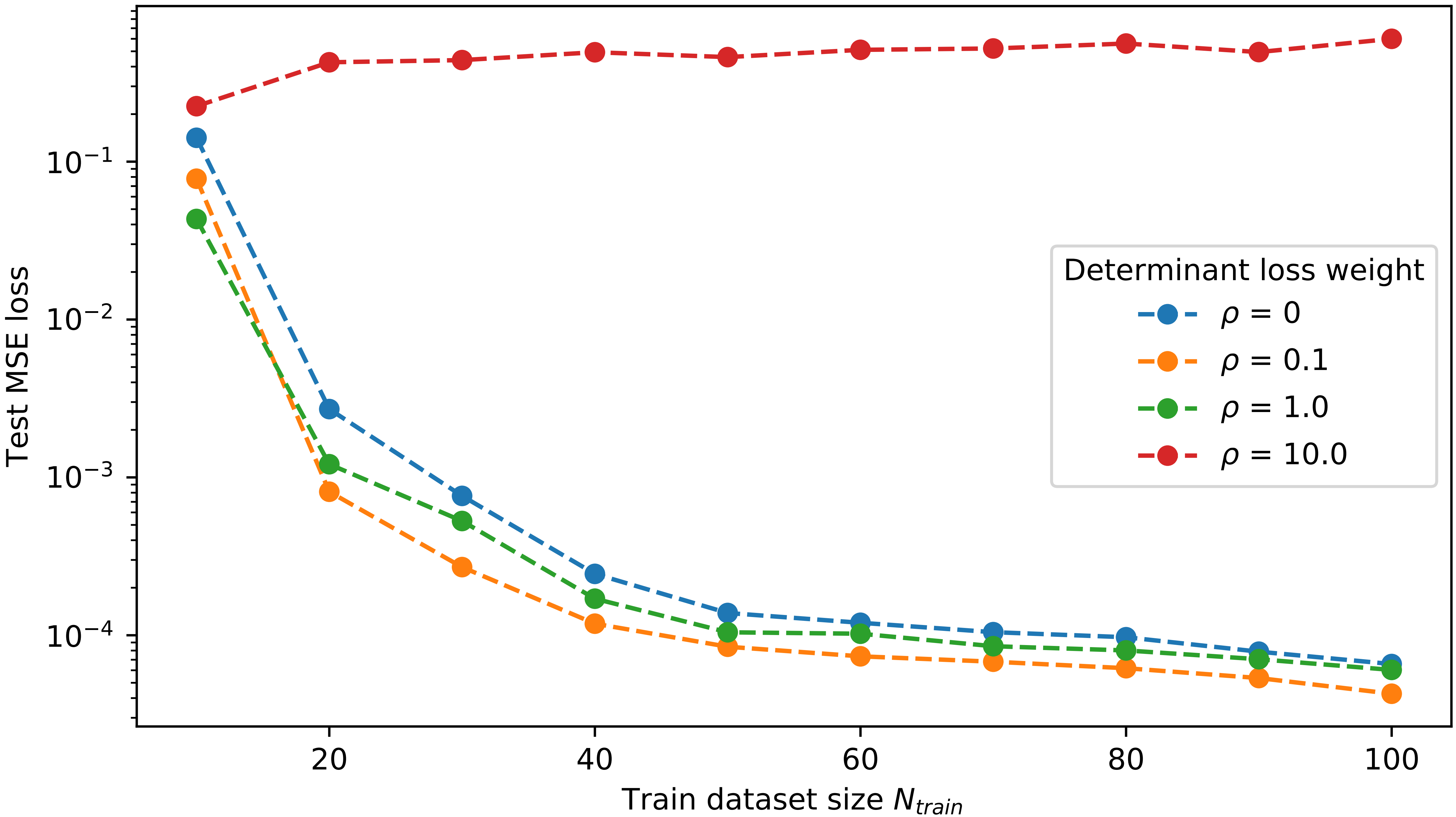
Model 3 with regularisation

$$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y) + \frac{\beta}{2}\sum_i w_i^2$$



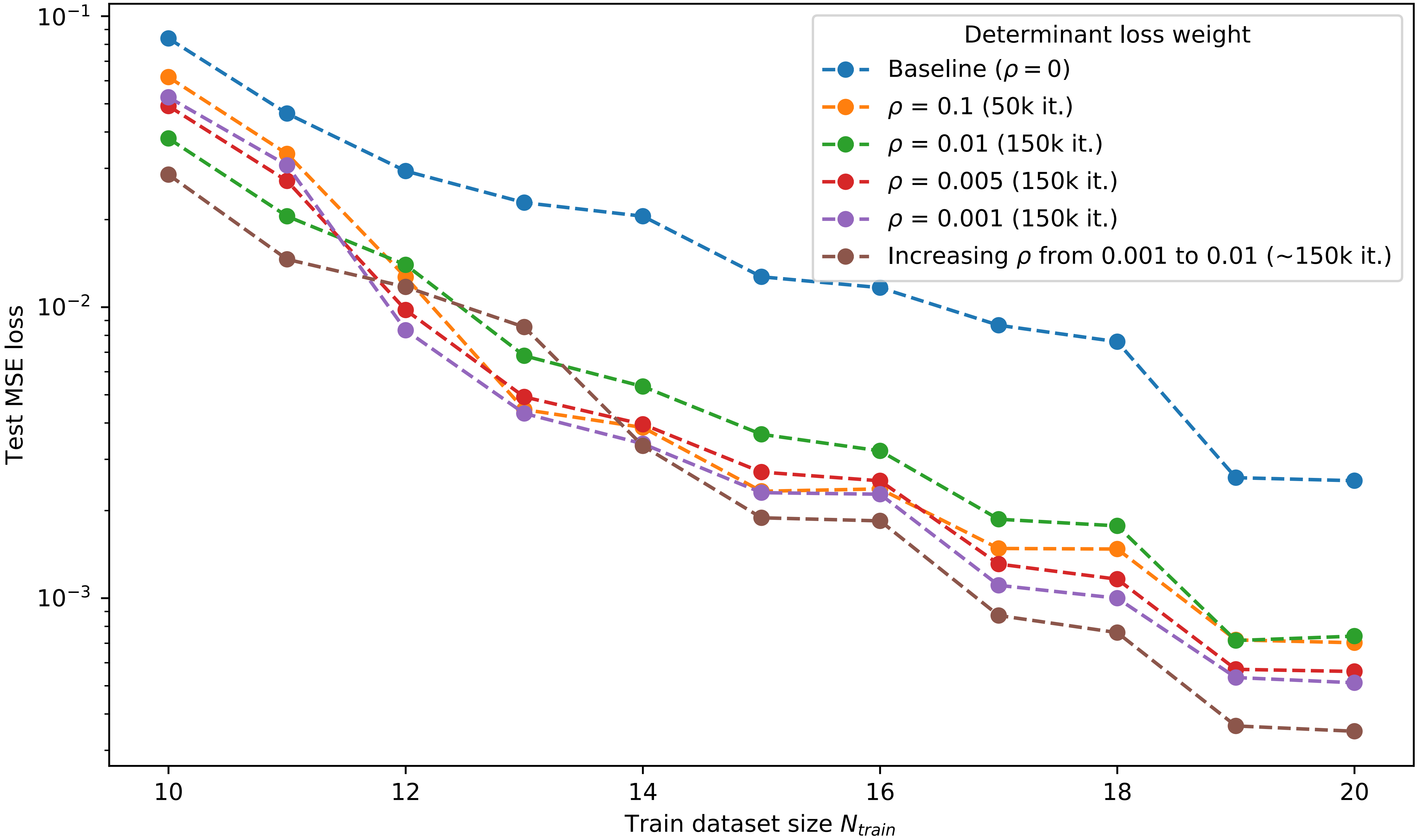
V. Appendix

Penalty Method Model 3



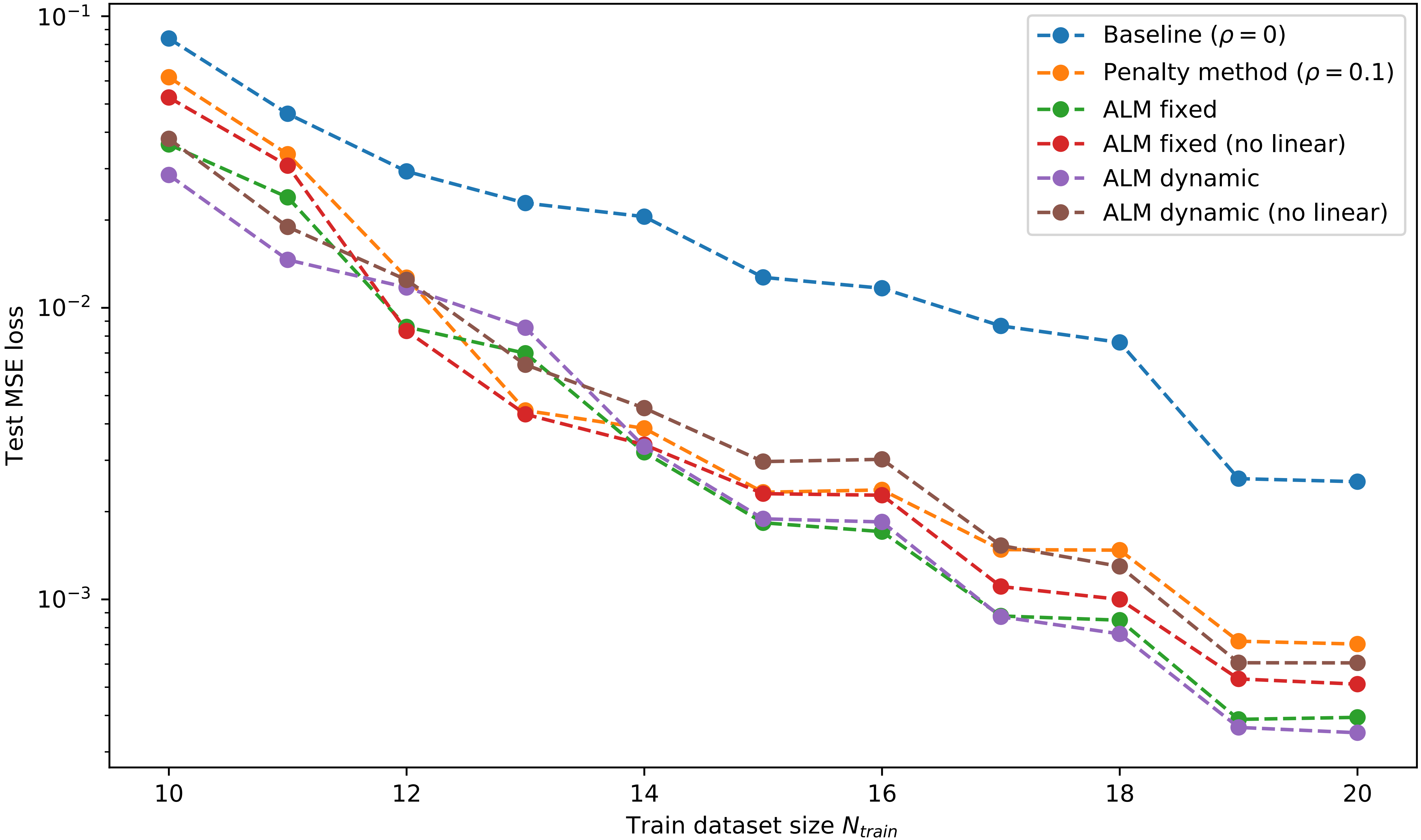
V. Appendix

Penalty Method Model 3 - Increasing weight



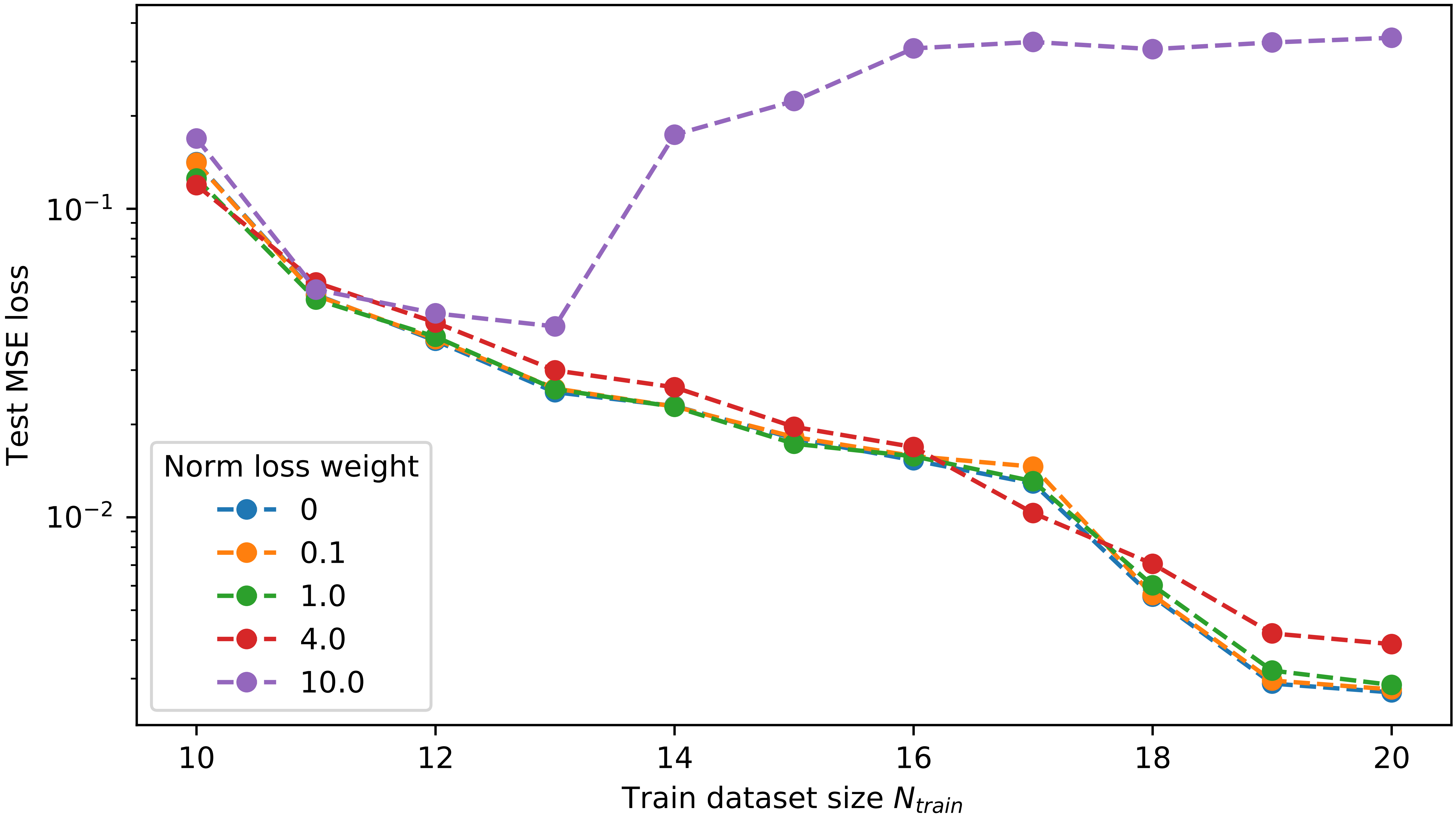
V. Appendix

ALM Model 3 - Fixed and Dynamic



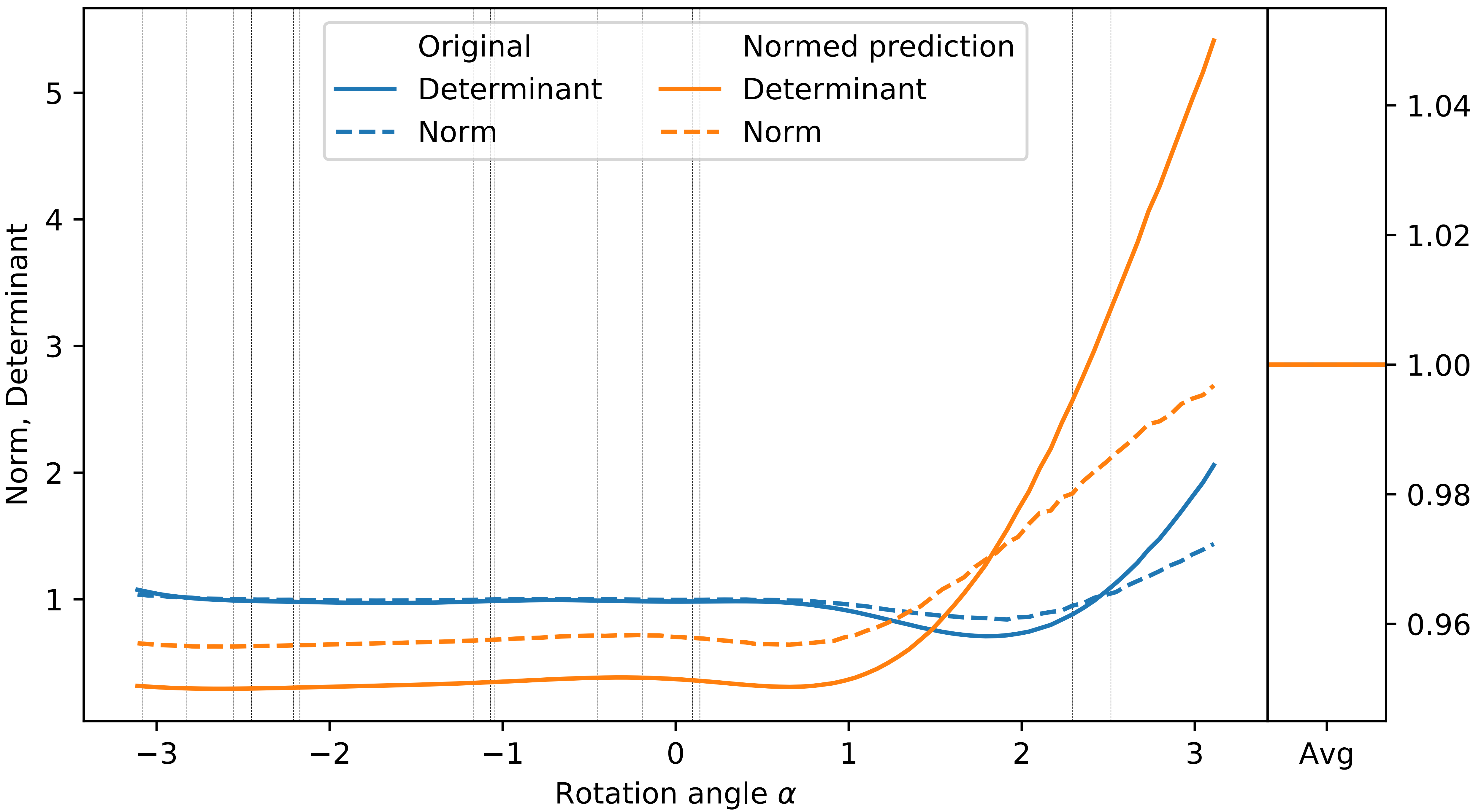
V. Appendix

Penalty Method Model 3 - Norm constraint



V. Appendix

Normed prediction Model 3



V. Appendix

Penalty Method Model 3 - 3 dimensional rotation