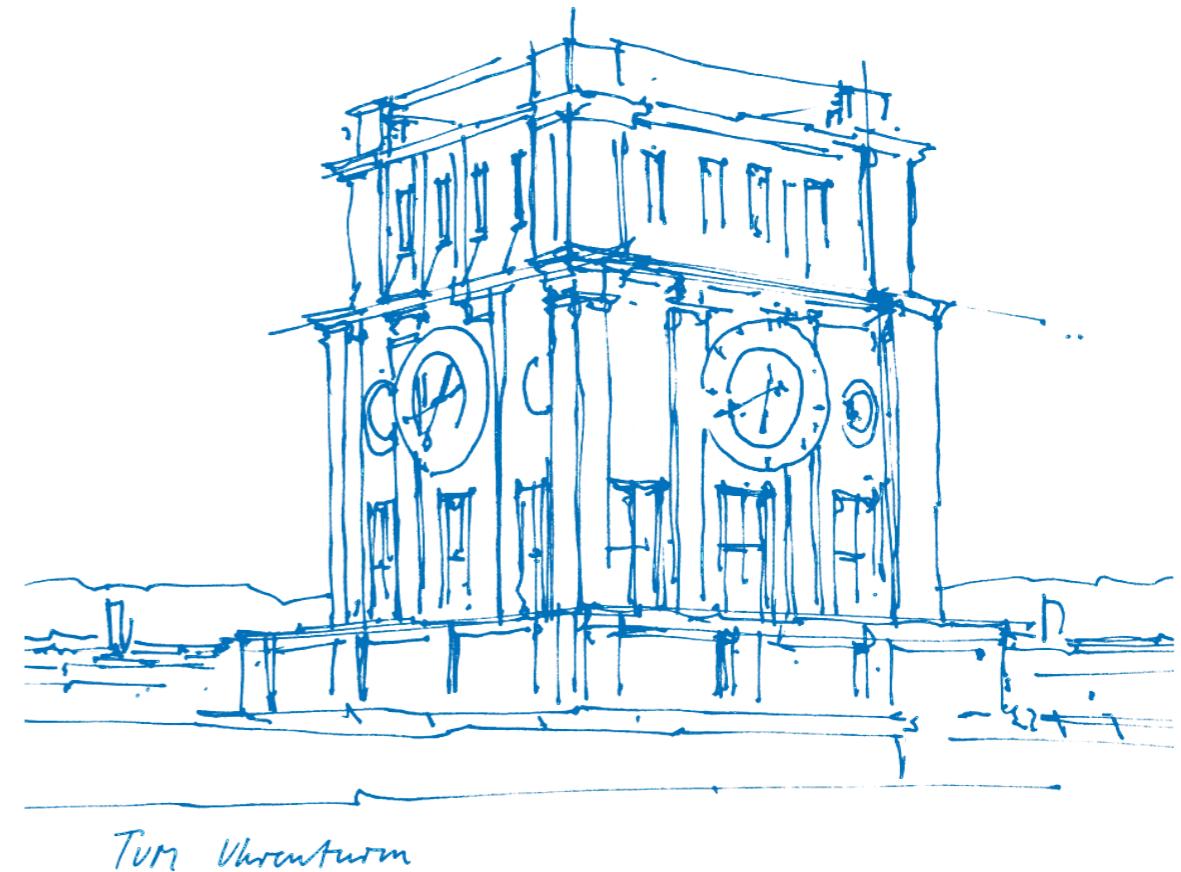


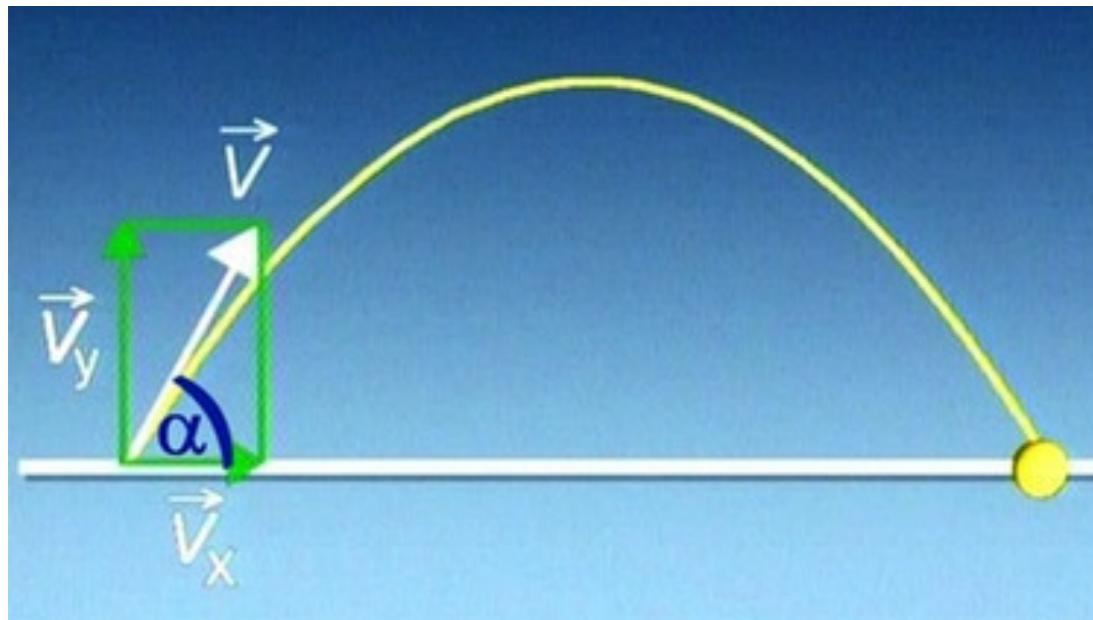
Physically constrained networks

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

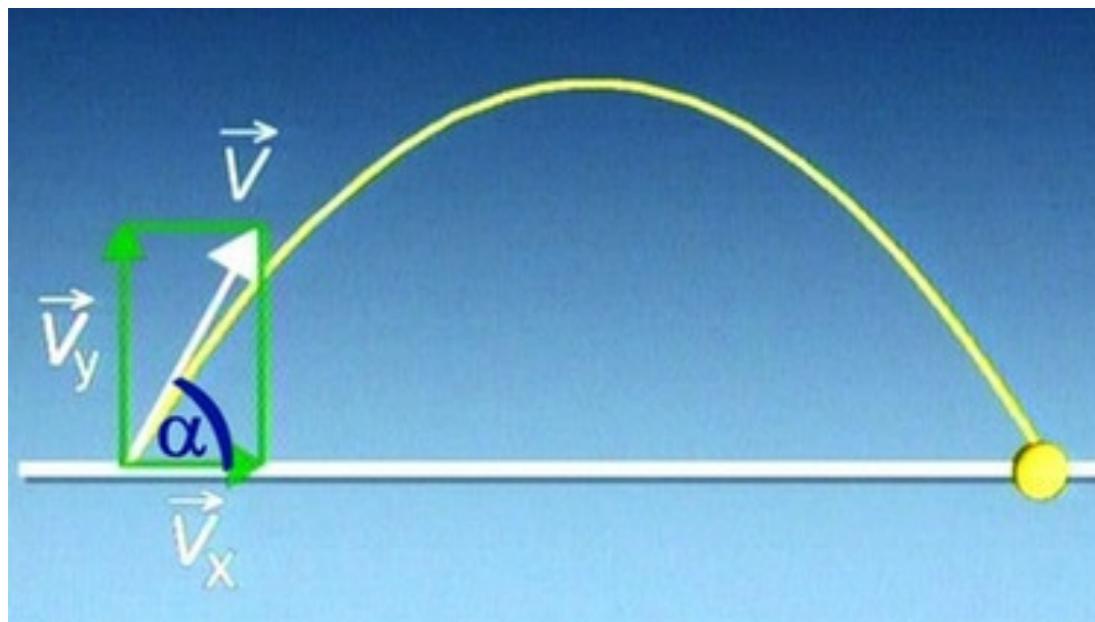


https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRB_oBff0NIA087mkfhKQ9dujNPhQYjNOEoo3drYeclw7Ov0HRfQ

- 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_oBff0NIA087mkfhKQ9dujNPhQYjNOEoo3drYeclw7Ov0HRfQ

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



Combine advantages using physically constrained networks

I. Motivation

Pedestrian Trajectory Prediction

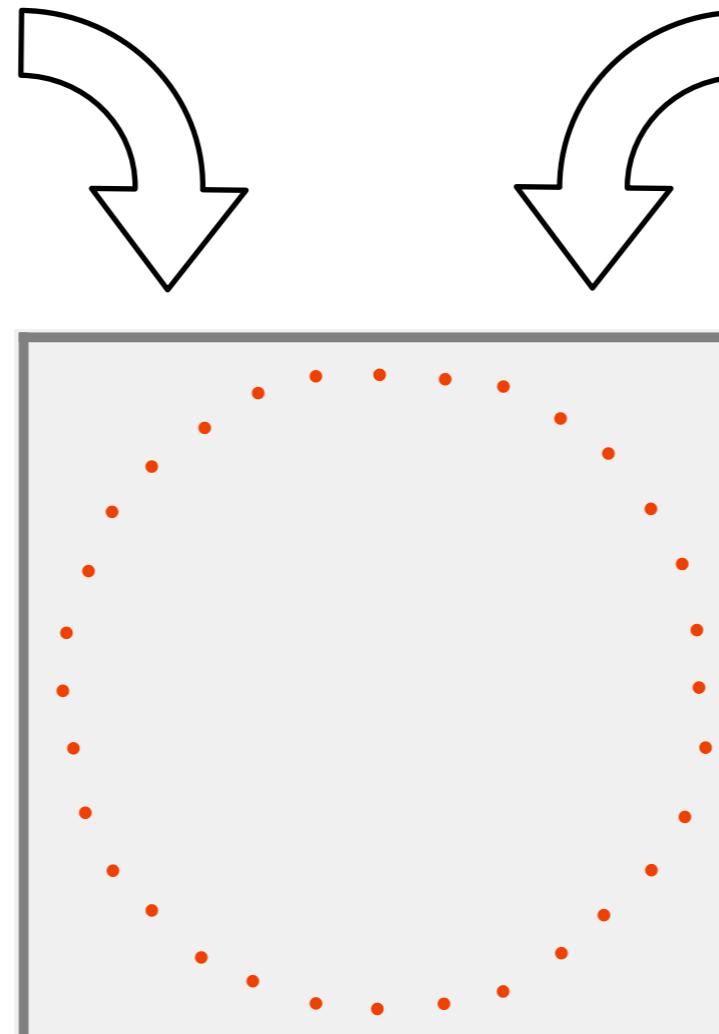
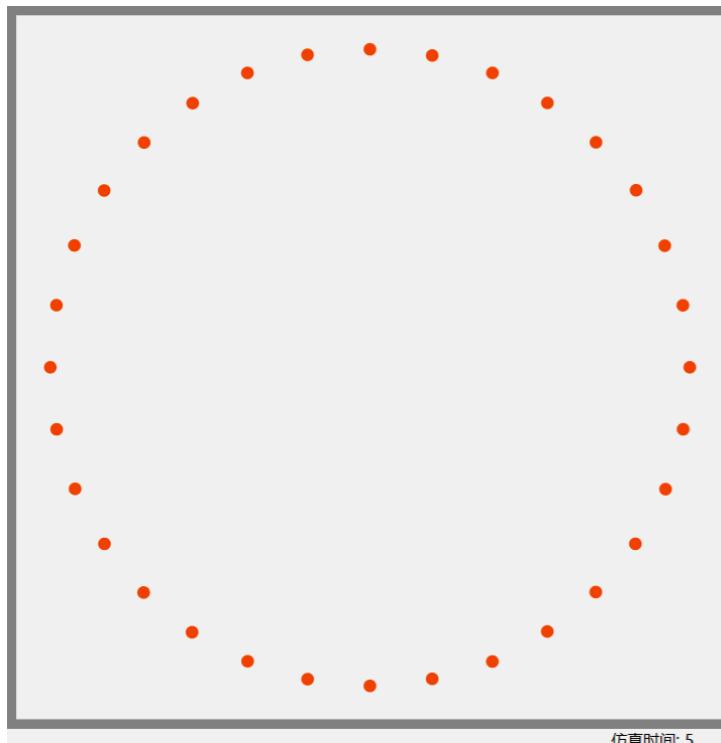


<http://pedynamic.com/circle-antipode-experiments/>

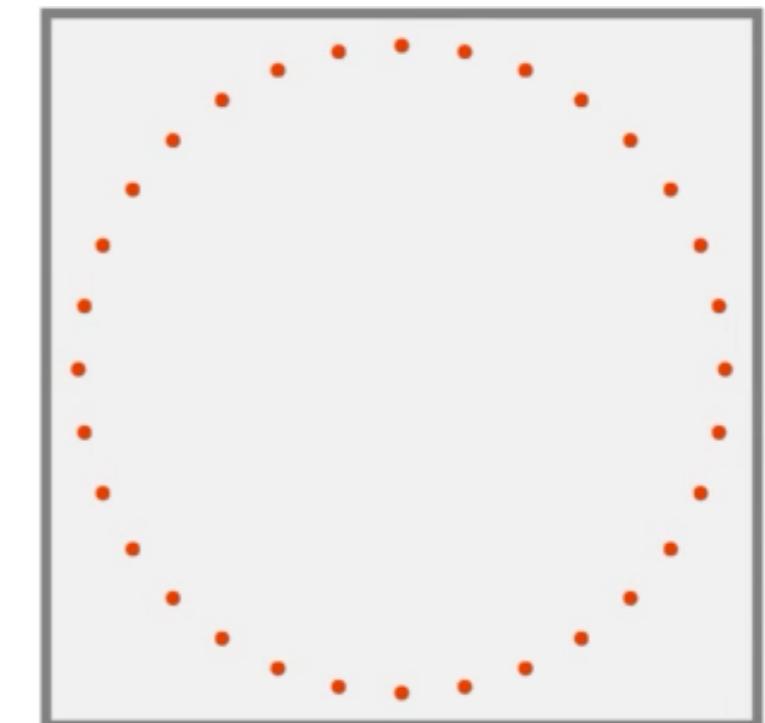
I. Motivation

Pedestrian Trajectory Prediction

Hand-crafted



Deep learning

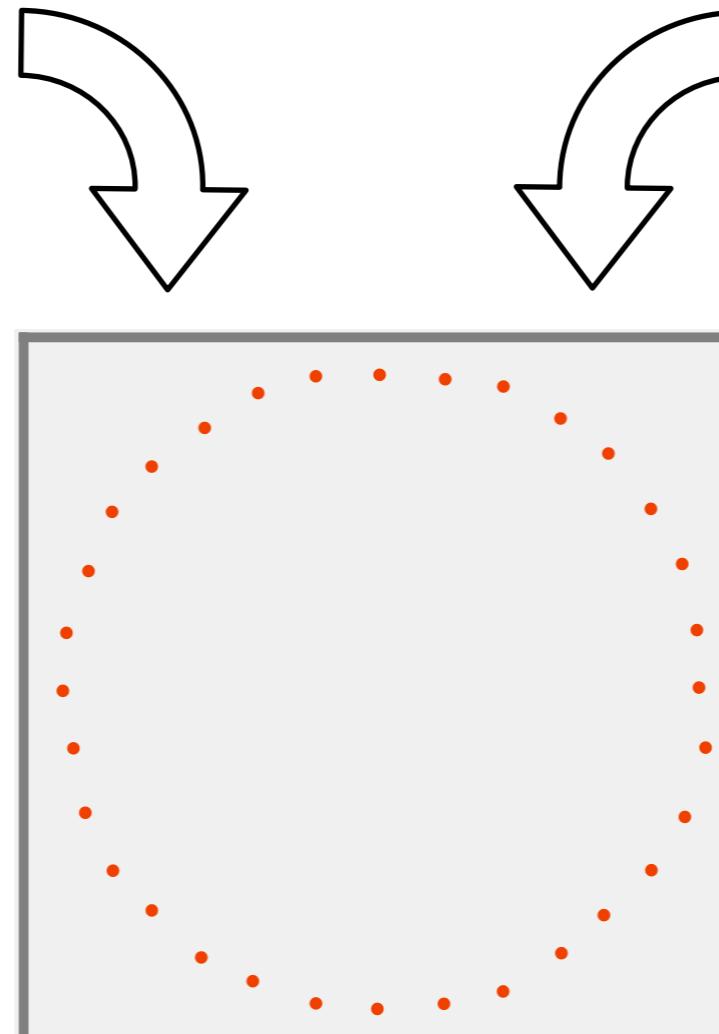
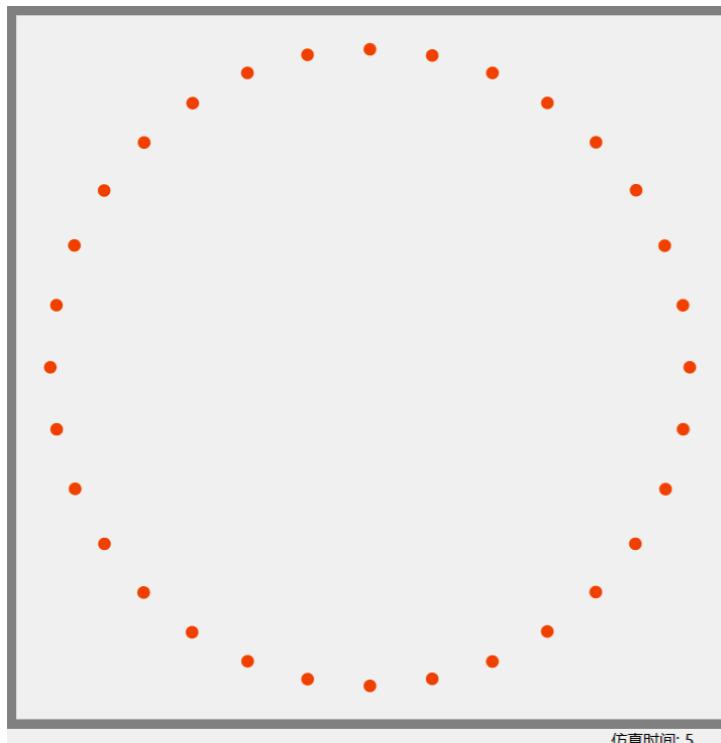


Physically constrained network

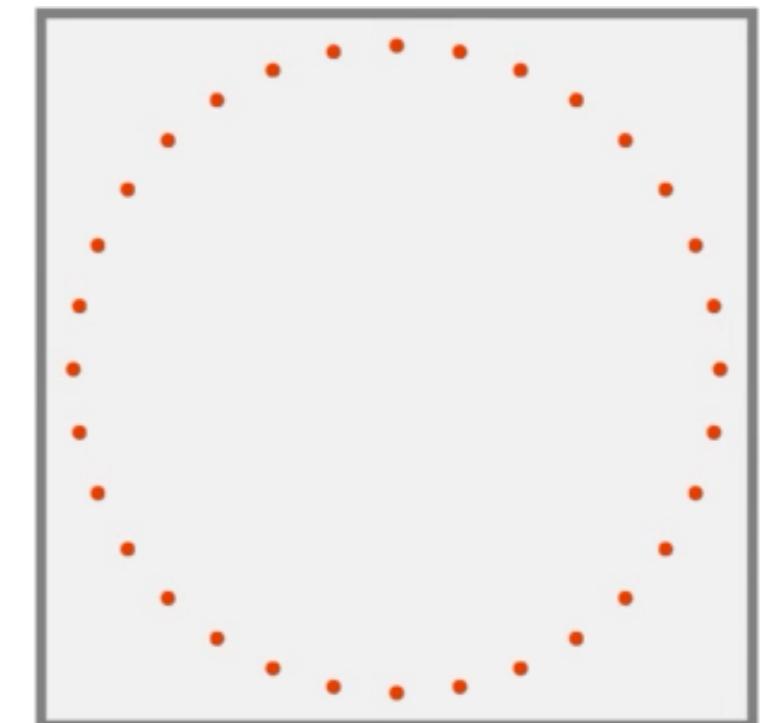
I. Motivation

Pedestrian Trajectory Prediction

Hand-crafted



Deep learning



Physically constrained network

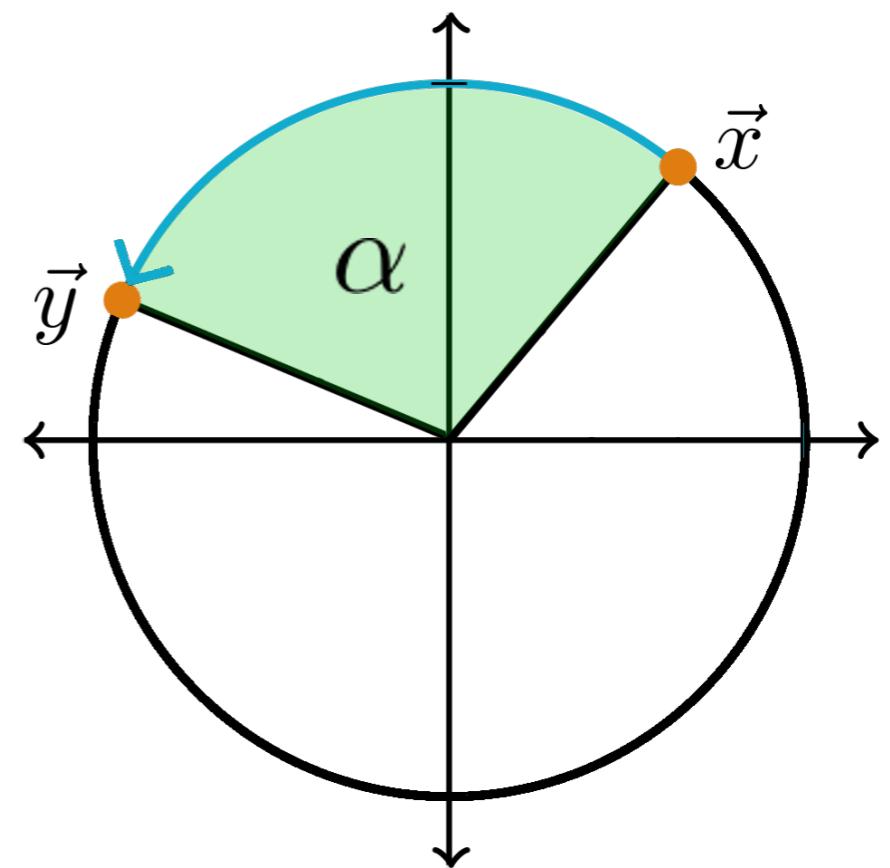
Outline

I. Motivation

II. Constrained optimisation

III. Experiment

IV. Results



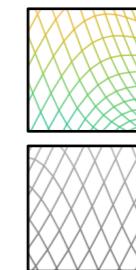
II. Constrained optimisation

Training a constrained network

$\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}$: Loss function

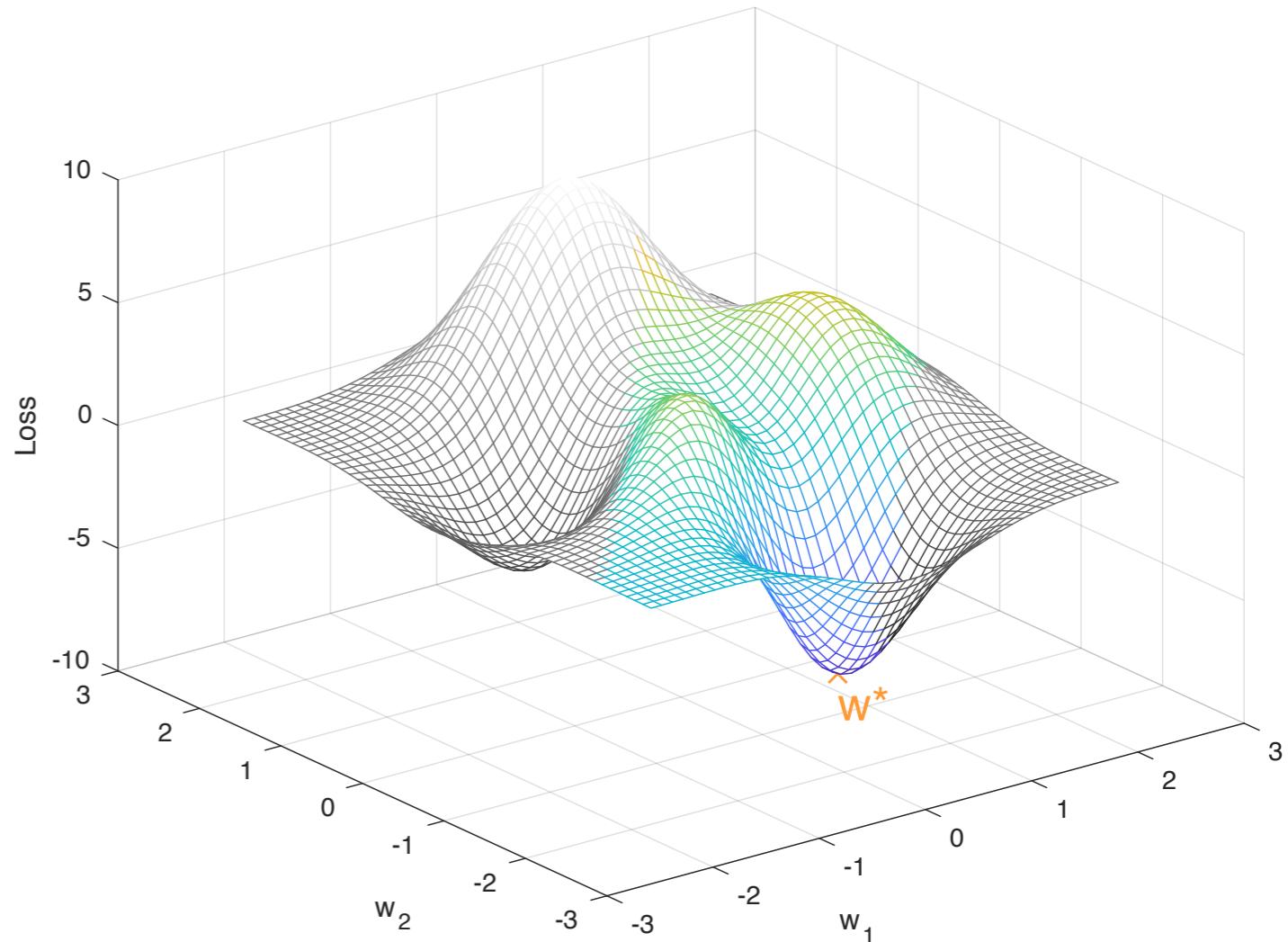
\mathcal{I} : Set of inequality constraint indices

$c_i : \mathbb{R}^n \rightarrow \mathbb{R}, i \in \mathcal{I}$: Inequality constraints



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

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



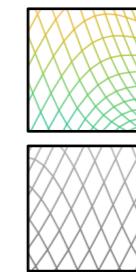
II. Constrained optimisation

Training a constrained network

$\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}$: Loss function

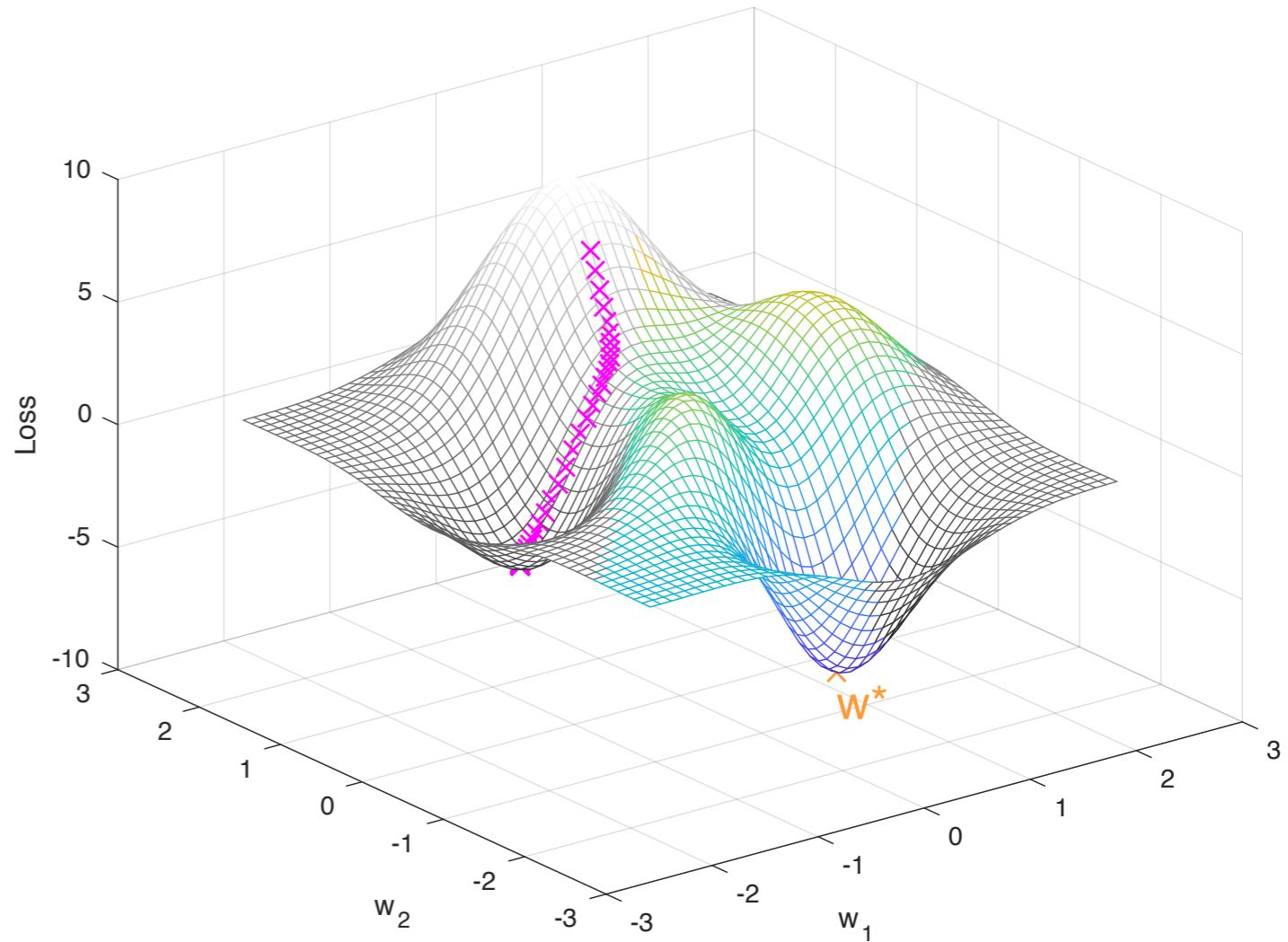
\mathcal{I} : Set of inequality constraint indices

$c_i : \mathbb{R}^n \rightarrow \mathbb{R}, i \in \mathcal{I}$: Inequality constraints



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

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



II. Constrained optimisation

Penalty Method

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

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

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

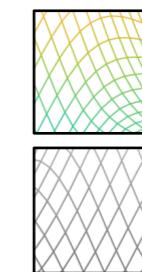
Penalty weight: $\rho > 0$

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

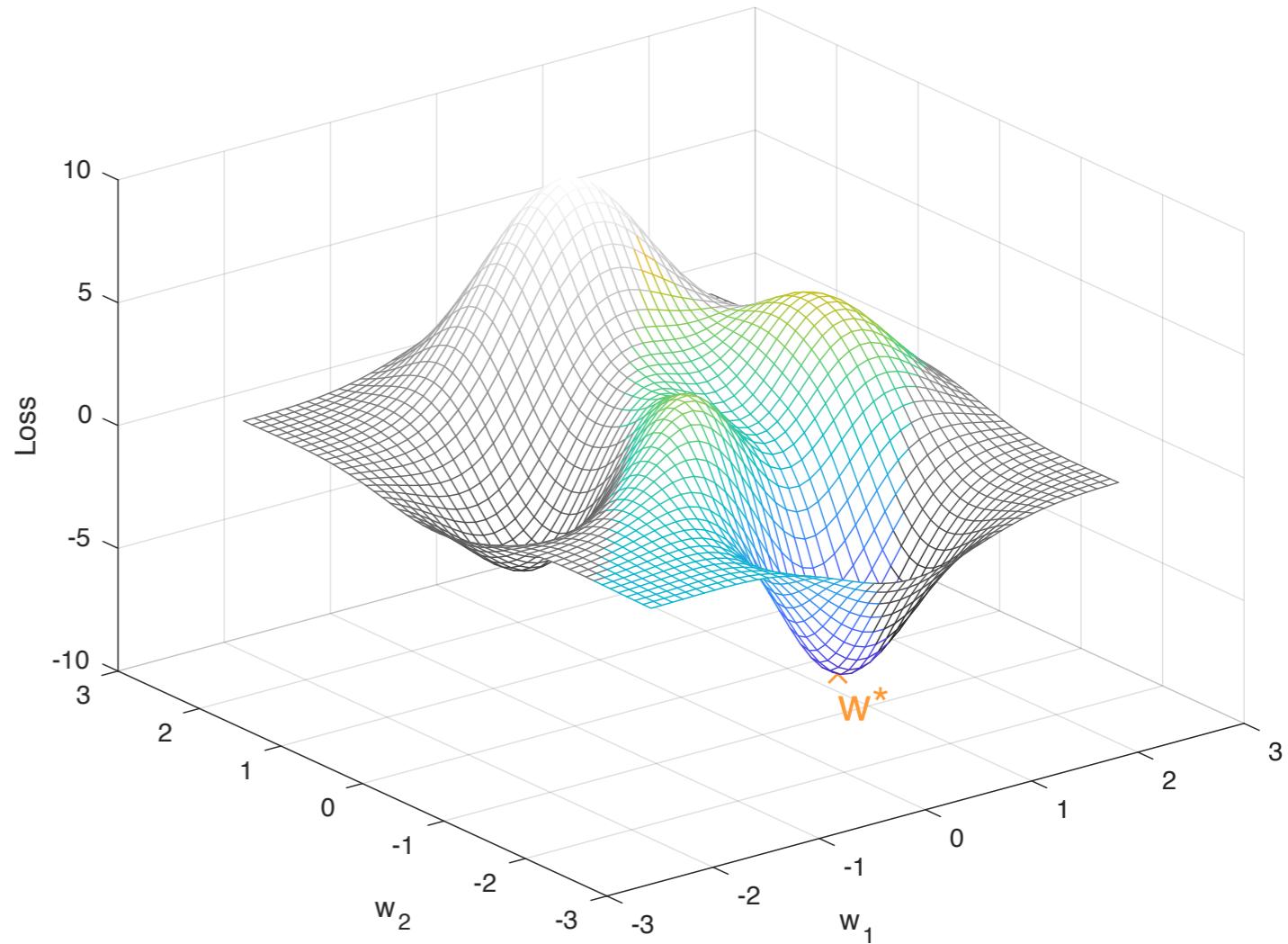
II. Constrained optimisation

Penalty Method

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



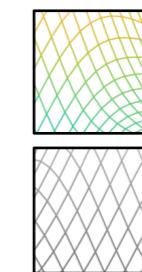
Feasible
Not feasible



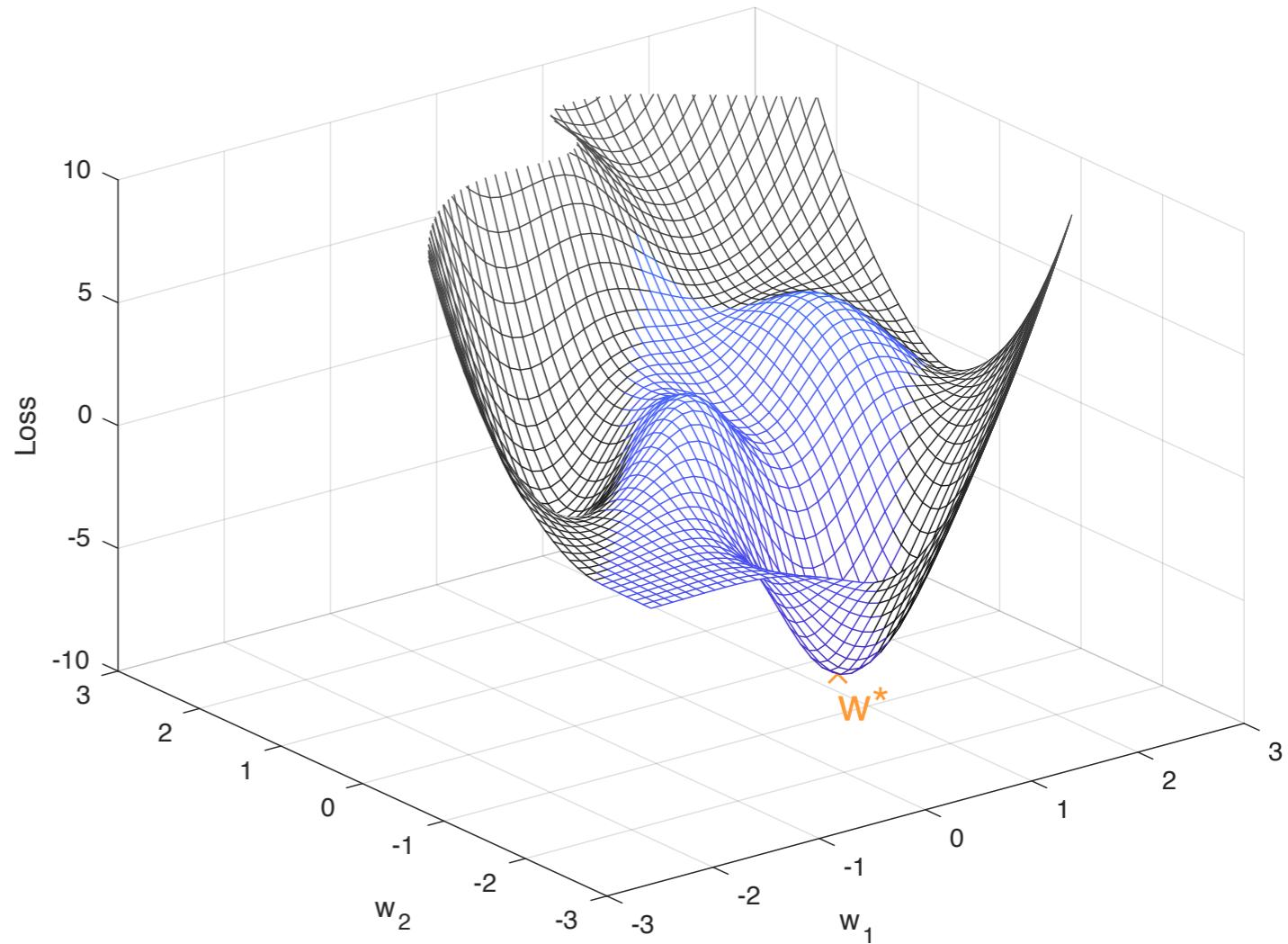
II. Constrained optimisation

Penalty Method

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



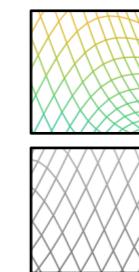
$$\rho = 3$$



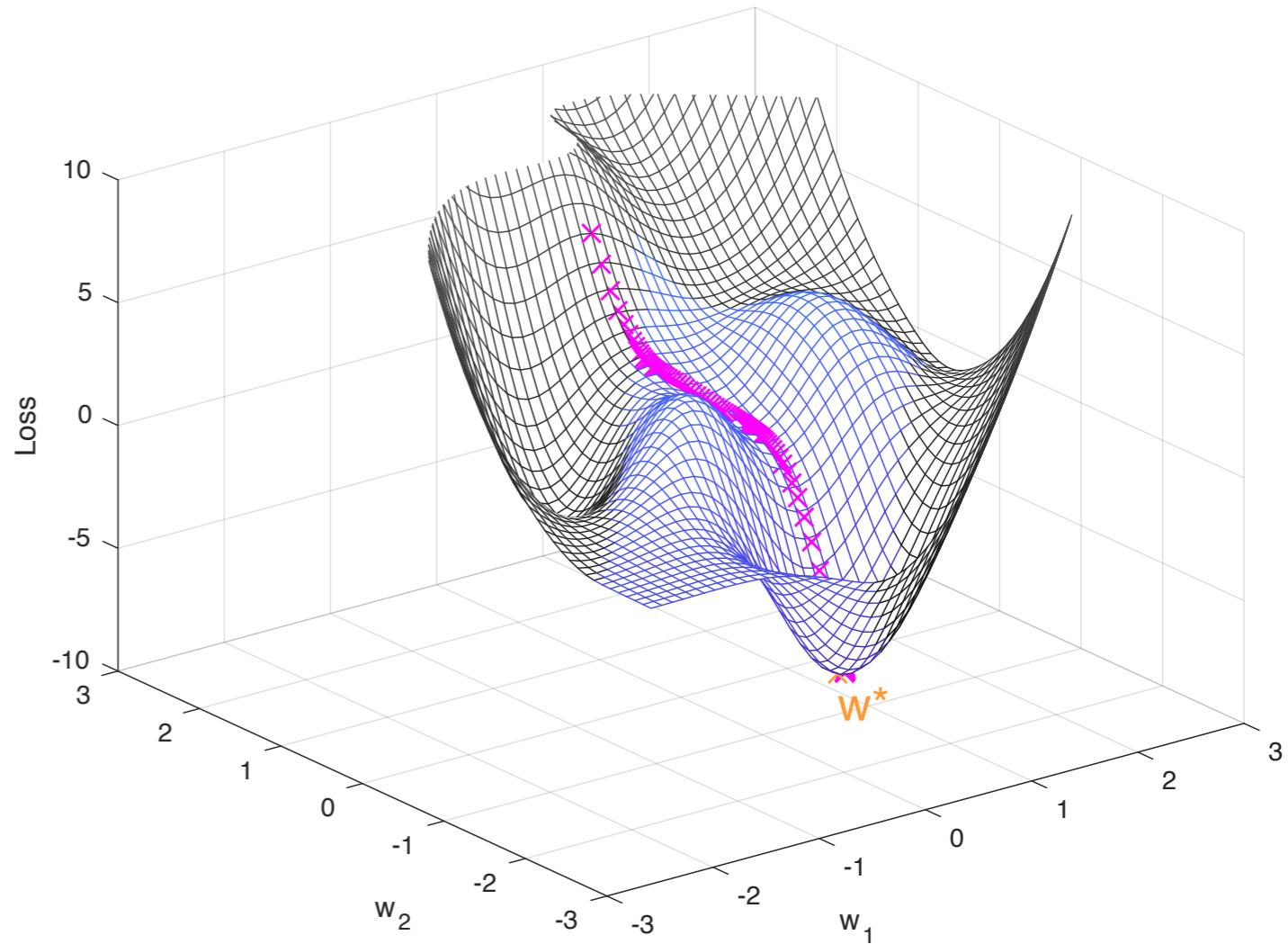
II. Constrained optimisation

Penalty Method

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



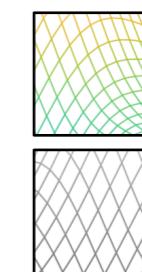
$$\rho = 3$$



II. Constrained optimisation

Penalty Method

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

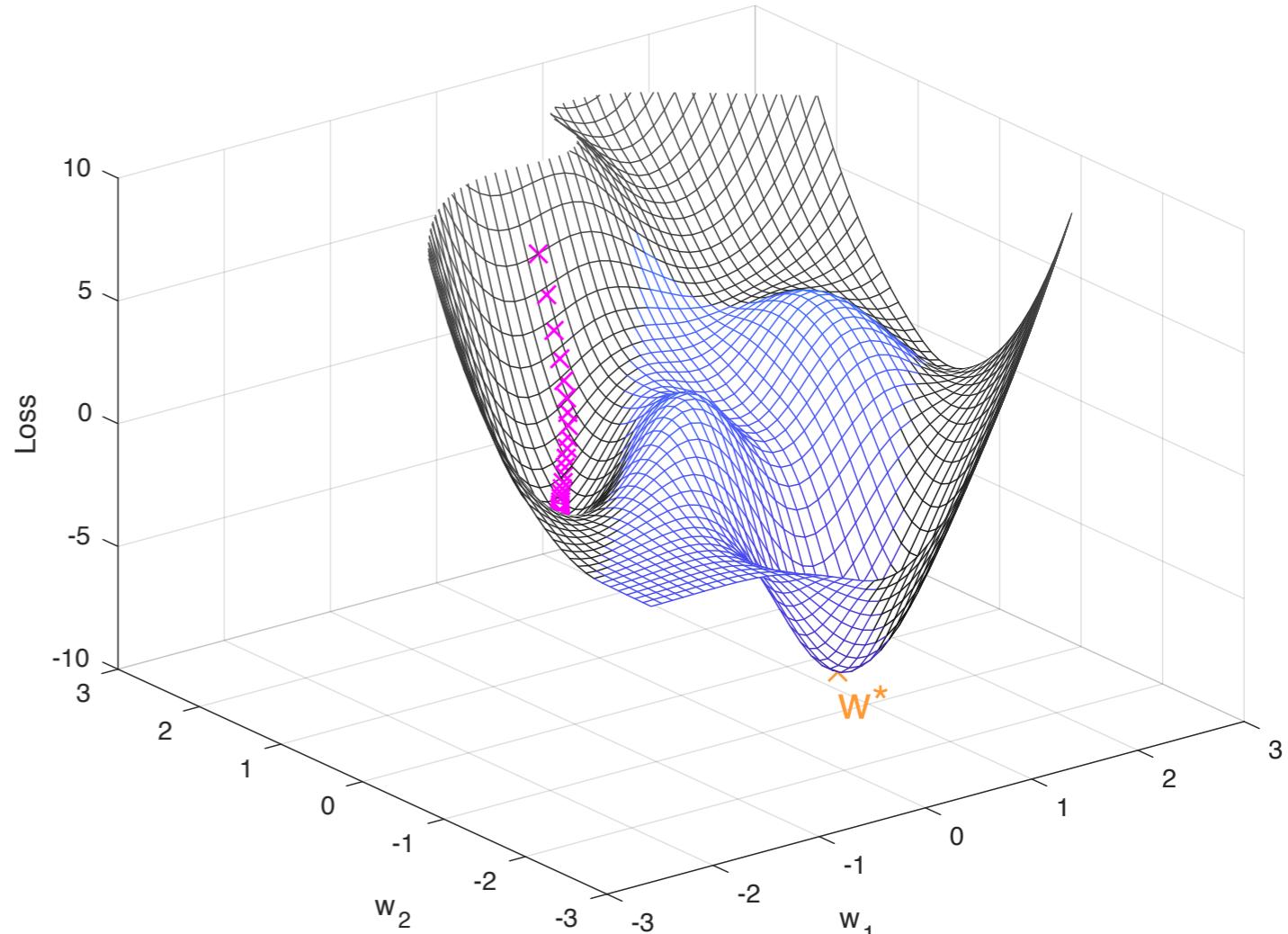


$$\rho = 3$$

Requirement for feasible solution:

$$\rho \rightarrow \infty$$

 Ill-conditioned

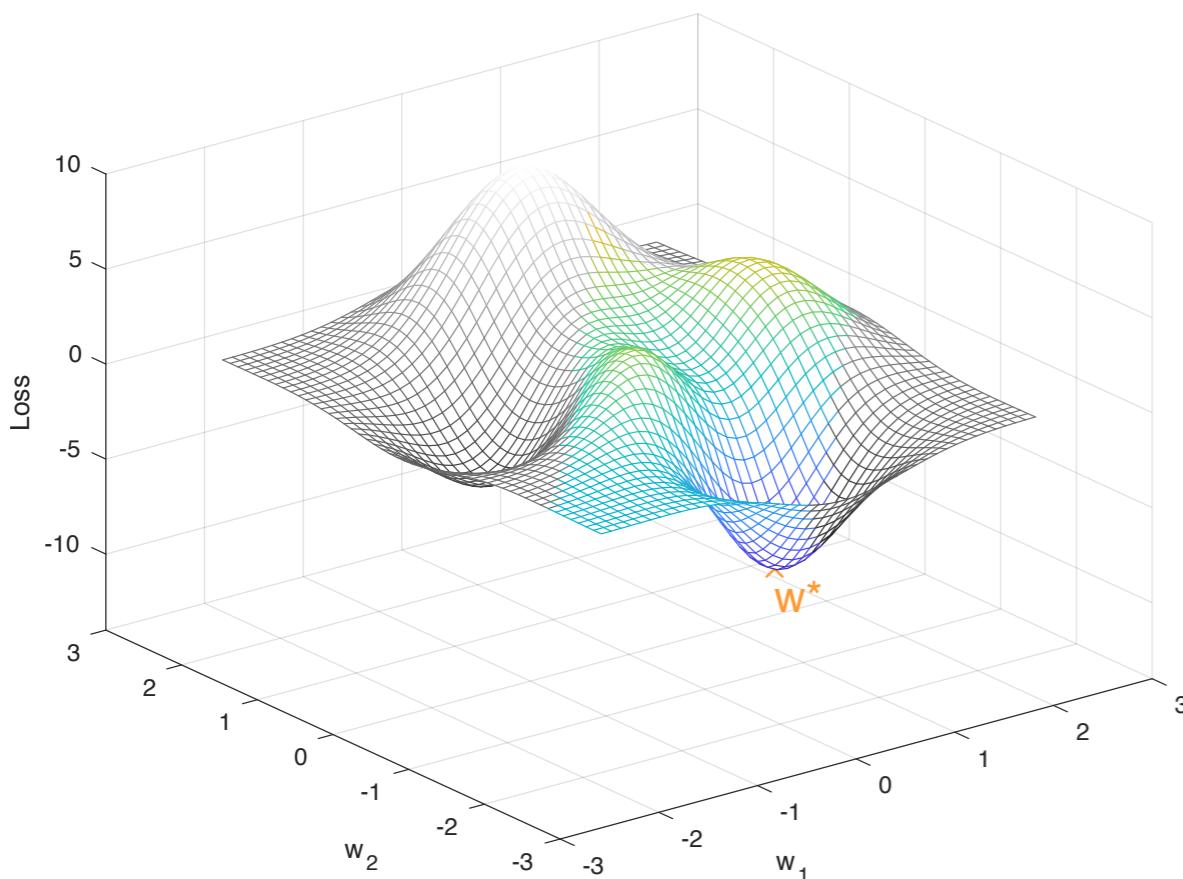
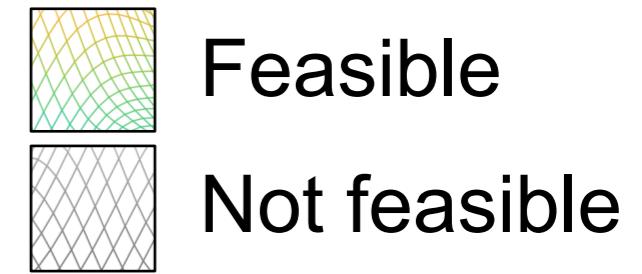


II. Constrained optimisation

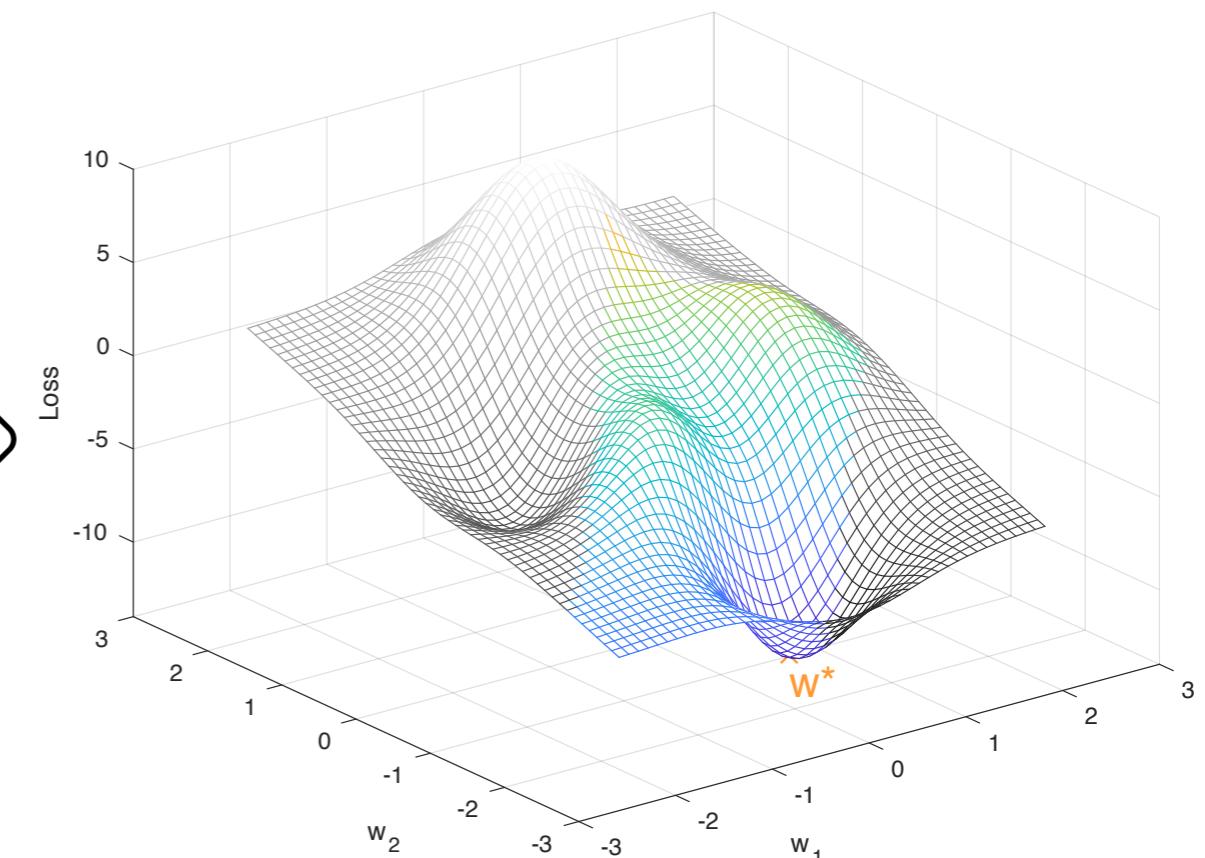
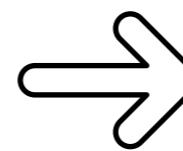
Lagrangian Relaxation

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

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



$$\mathcal{L}(w)$$



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

II. Constrained optimisation

Lagrangian Relaxation

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

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

$$\Rightarrow \mathcal{L}'(w^*, \lambda) \leq \mathcal{L}(w^*)$$

$$\Rightarrow (\min_{w \in \mathbb{R}^n} \mathcal{L}'(w, \lambda)) \leq \mathcal{L}(w^*)$$

$$\Rightarrow \text{Lagrangian Dual: } \max_{\lambda} \min_{w \in \mathbb{R}^n} \mathcal{L}'(w, \lambda)$$

II. Constrained optimisation

Augmented Lagrangian Method

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

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

\mathcal{E} : Set of equality constraint indices

$c_i : \mathbb{R}^n \rightarrow \mathbb{R}, i \in \mathcal{E}$: Equality constraints

II. Constrained optimisation

Augmented Lagrangian Method

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

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

\mathcal{E} : Set of equality constraint indices

$c_i : \mathbb{R}^n \rightarrow \mathbb{R}, \quad i \in \mathcal{E}$: Equality constraints

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

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

II. Constrained optimisation

Augmented Lagrangian Method

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

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

\mathcal{E} : Set of equality constraint indices

$c_i : \mathbb{R}^n \rightarrow \mathbb{R}, \quad i \in \mathcal{E}$: Equality constraints

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

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

For $k = 0, 1, \dots$

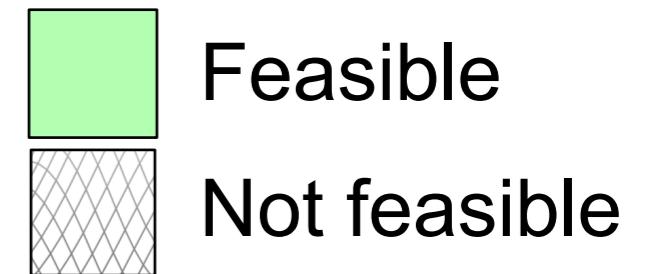
$$w^k = \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \left(\mathcal{L}(w) - \frac{1}{2} \sum_{i \in \mathcal{I}} \lambda_i^k c_i(w) + \mu_k \sum_{i \in \mathcal{I}} (c_i(w))^2 \right)$$

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

II. Constrained optimisation

Augmented Lagrangian Method

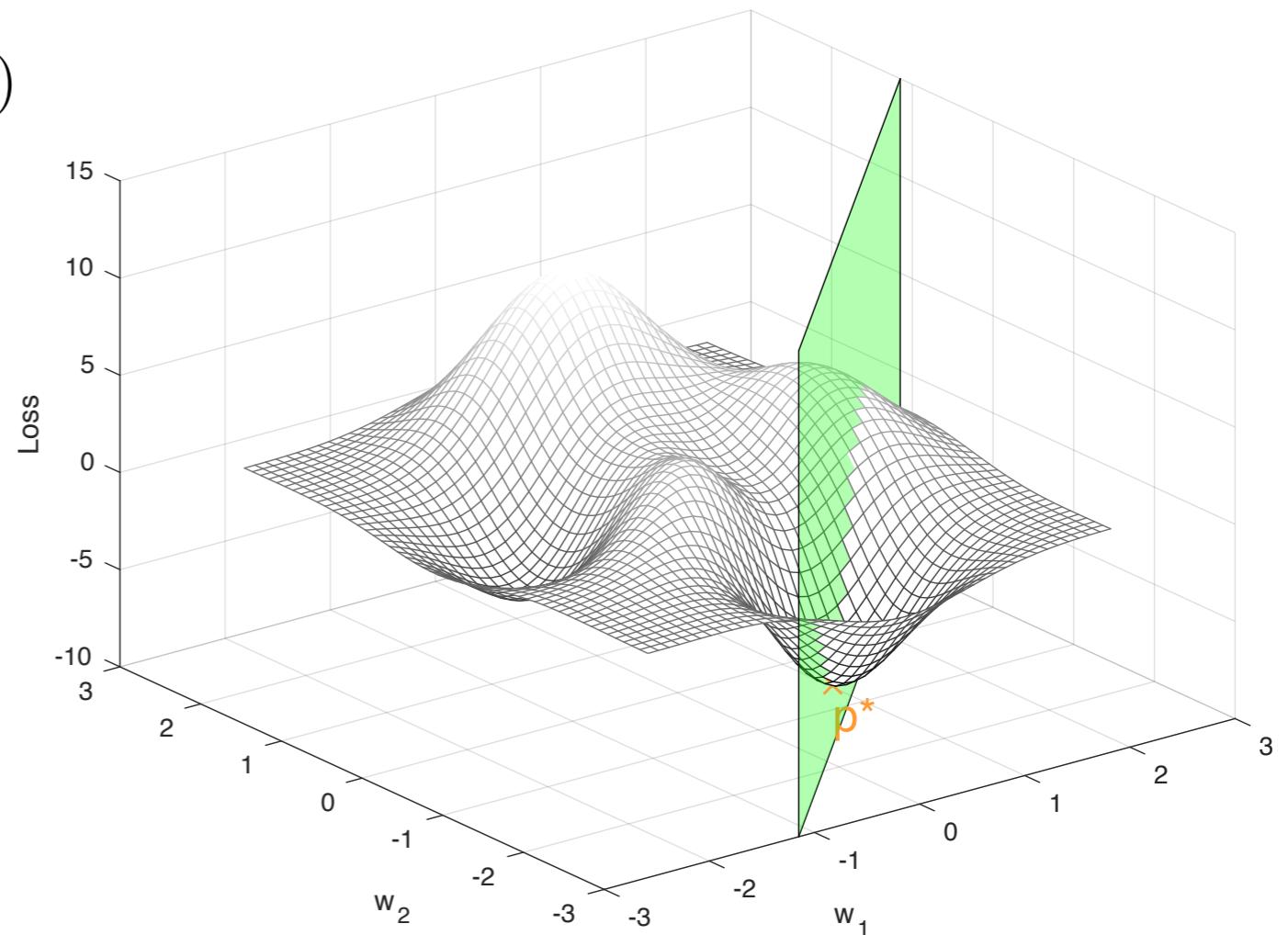
$$\boxed{\begin{aligned} w^* = \operatorname{argmin}_{w \in \mathbb{R}^n} \quad & \mathcal{L}(w) \\ \text{subject to} \quad & c_i(w) = 0, \quad i \in \mathcal{E} \end{aligned}}$$



For $k = 0, 1, \dots$

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

k	0	1	2	3	4	5
μ_k						
λ_1^k						
$c_1(w^k)$						

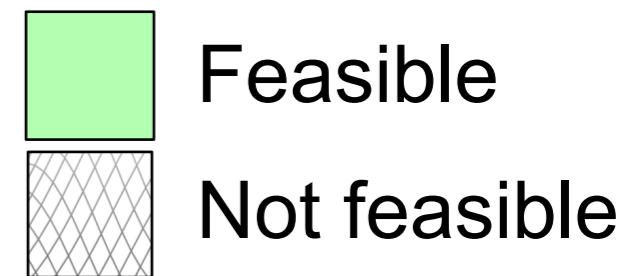


II. Constrained optimisation

Augmented Lagrangian Method

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

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

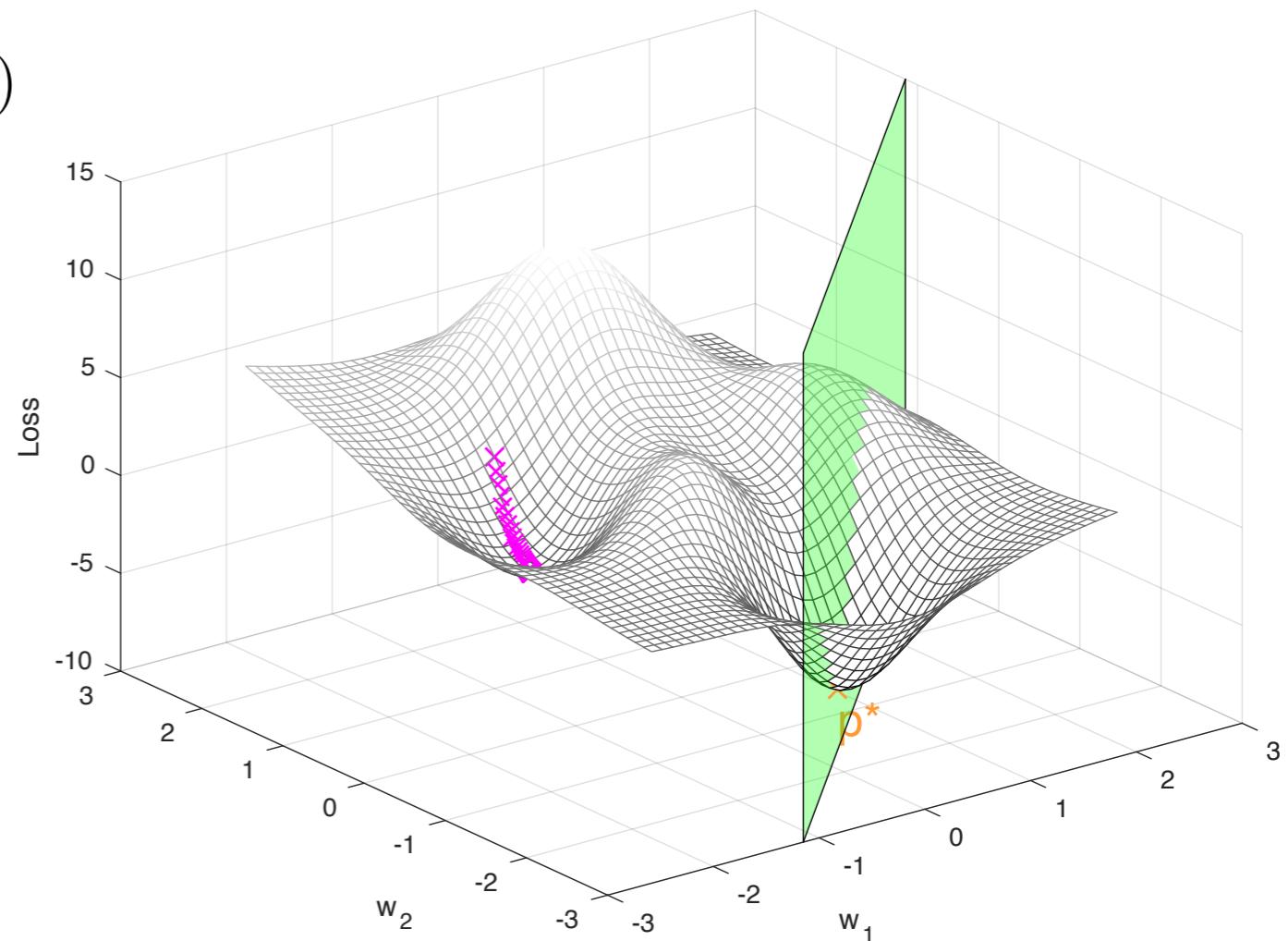


For $k = 0, 1, \dots$

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

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

k	0	1	2	3	4	5
μ_k	0,125					
λ_1^k	0					
$c_1(w^k)$	-3,30					

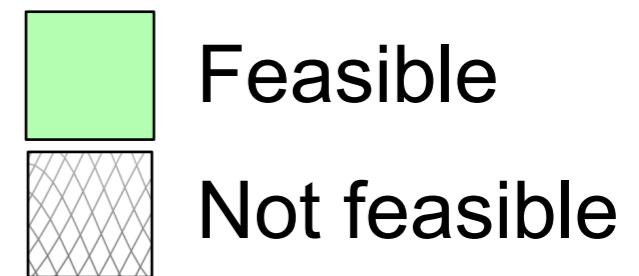


II. Constrained optimisation

Augmented Lagrangian Method

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

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

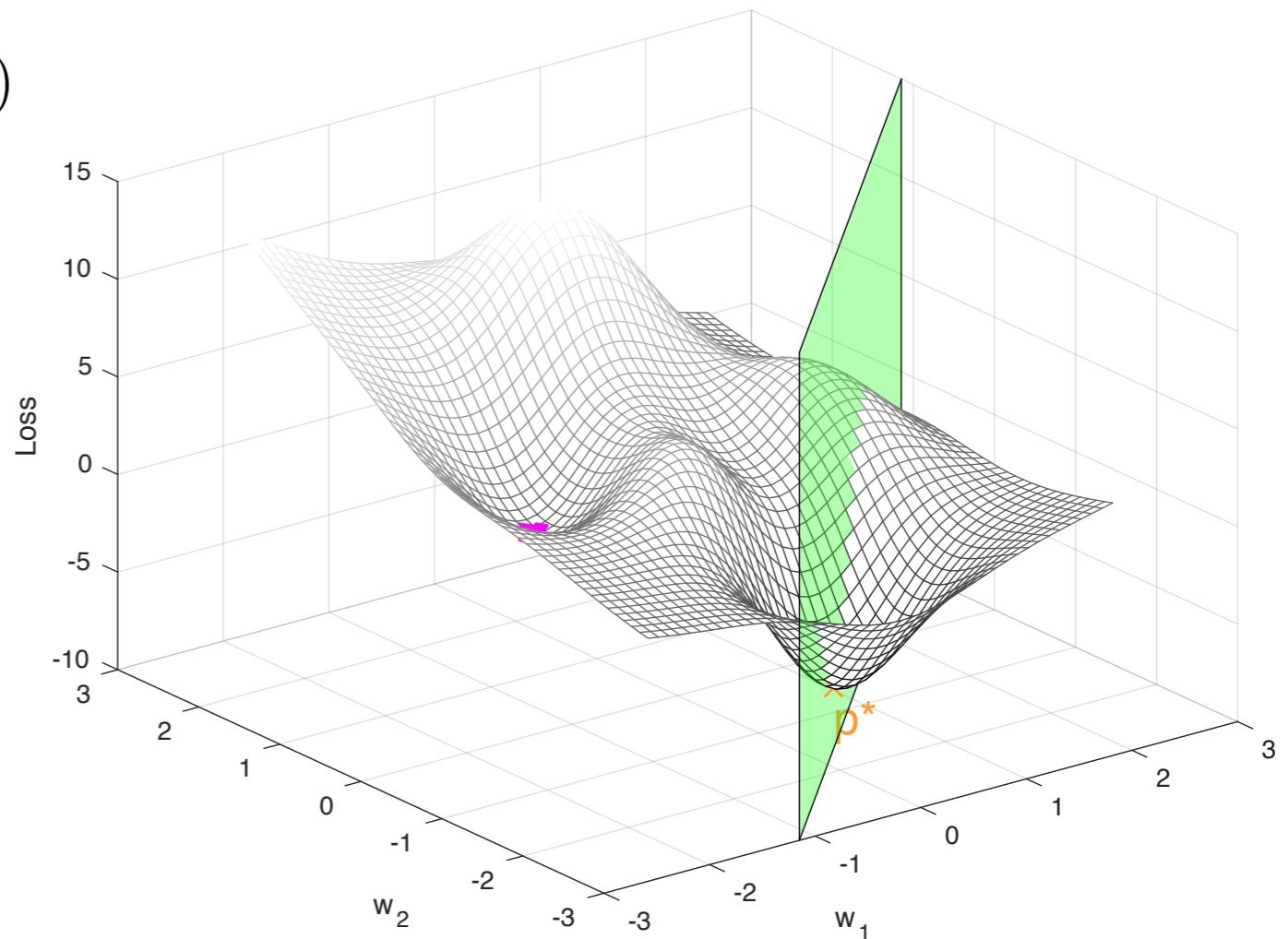


For $k = 0, 1, \dots$

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

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

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

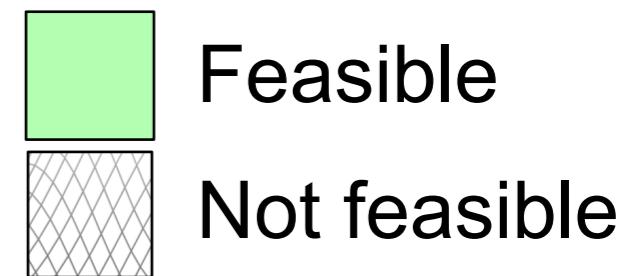


II. Constrained optimisation

Augmented Lagrangian Method

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

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

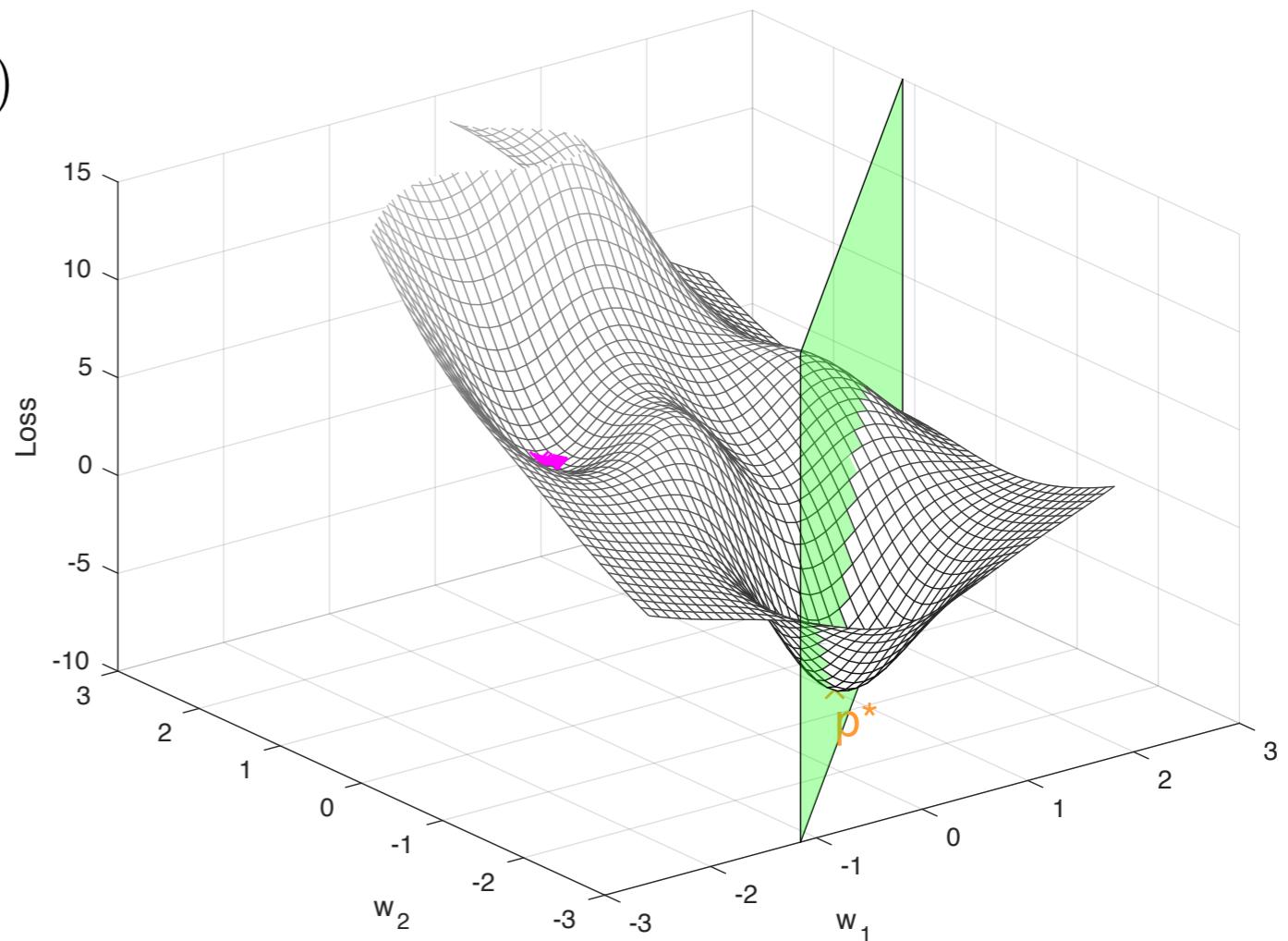


For $k = 0, 1, \dots$

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

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

k	0	1	2	3	4	5
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			

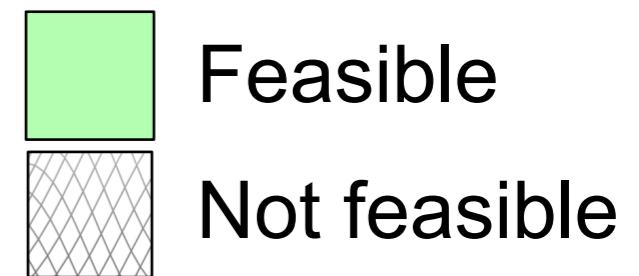


II. Constrained optimisation

Augmented Lagrangian Method

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

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

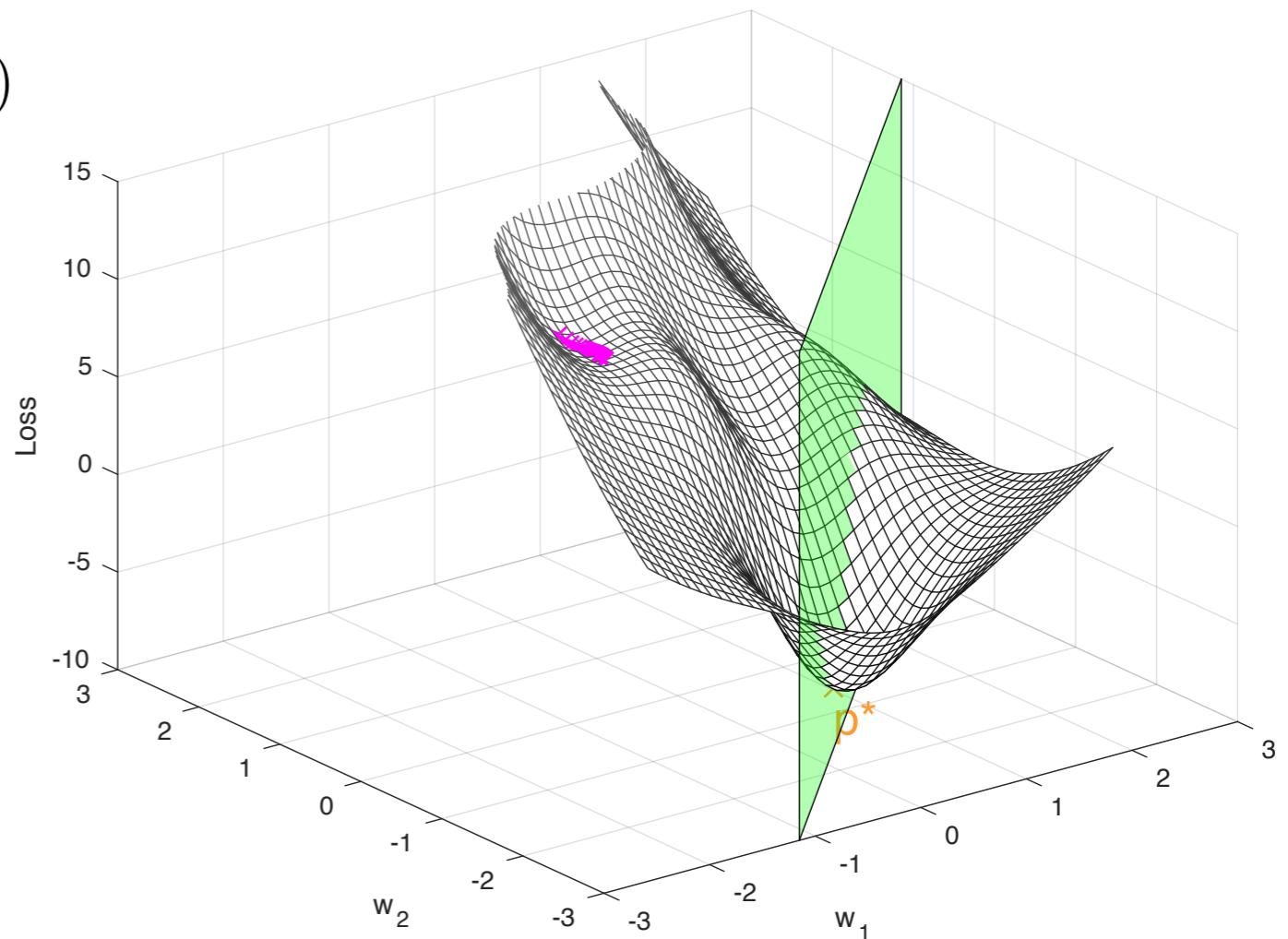


For $k = 0, 1, \dots$

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

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

k	0	1	2	3	4	5
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		

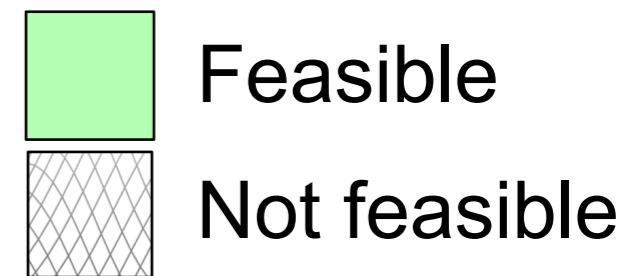


II. Constrained optimisation

Augmented Lagrangian Method

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

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

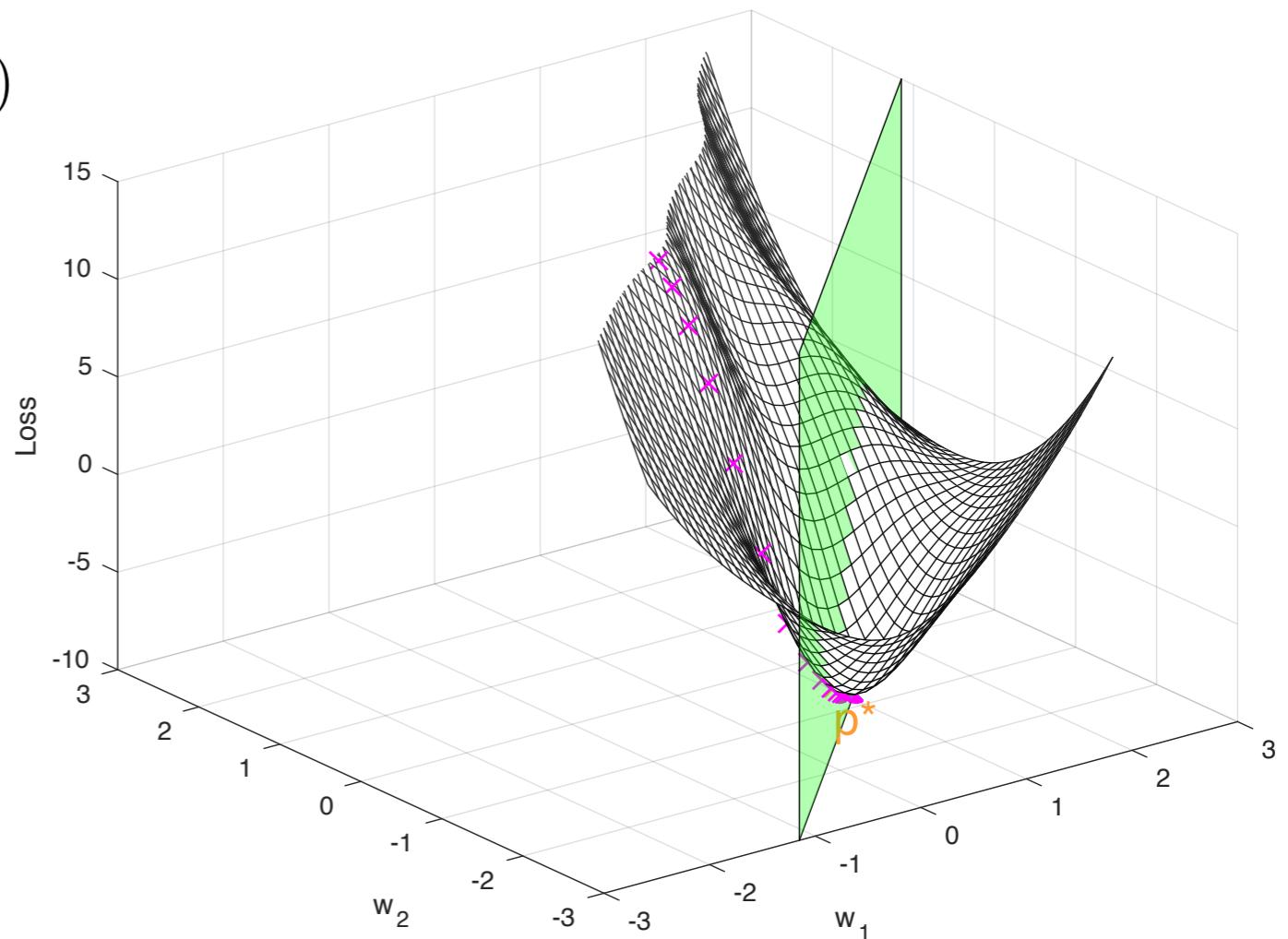


For $k = 0, 1, \dots$

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

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

k	0	1	2	3	4	5
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	

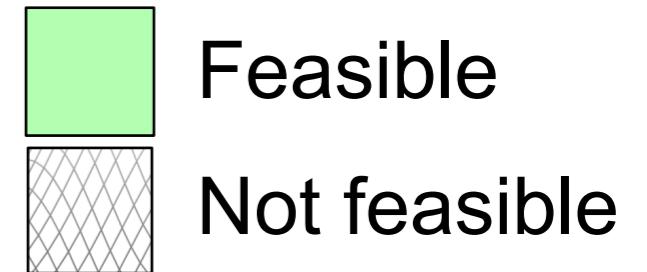


II. Constrained optimisation

Augmented Lagrangian Method

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

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

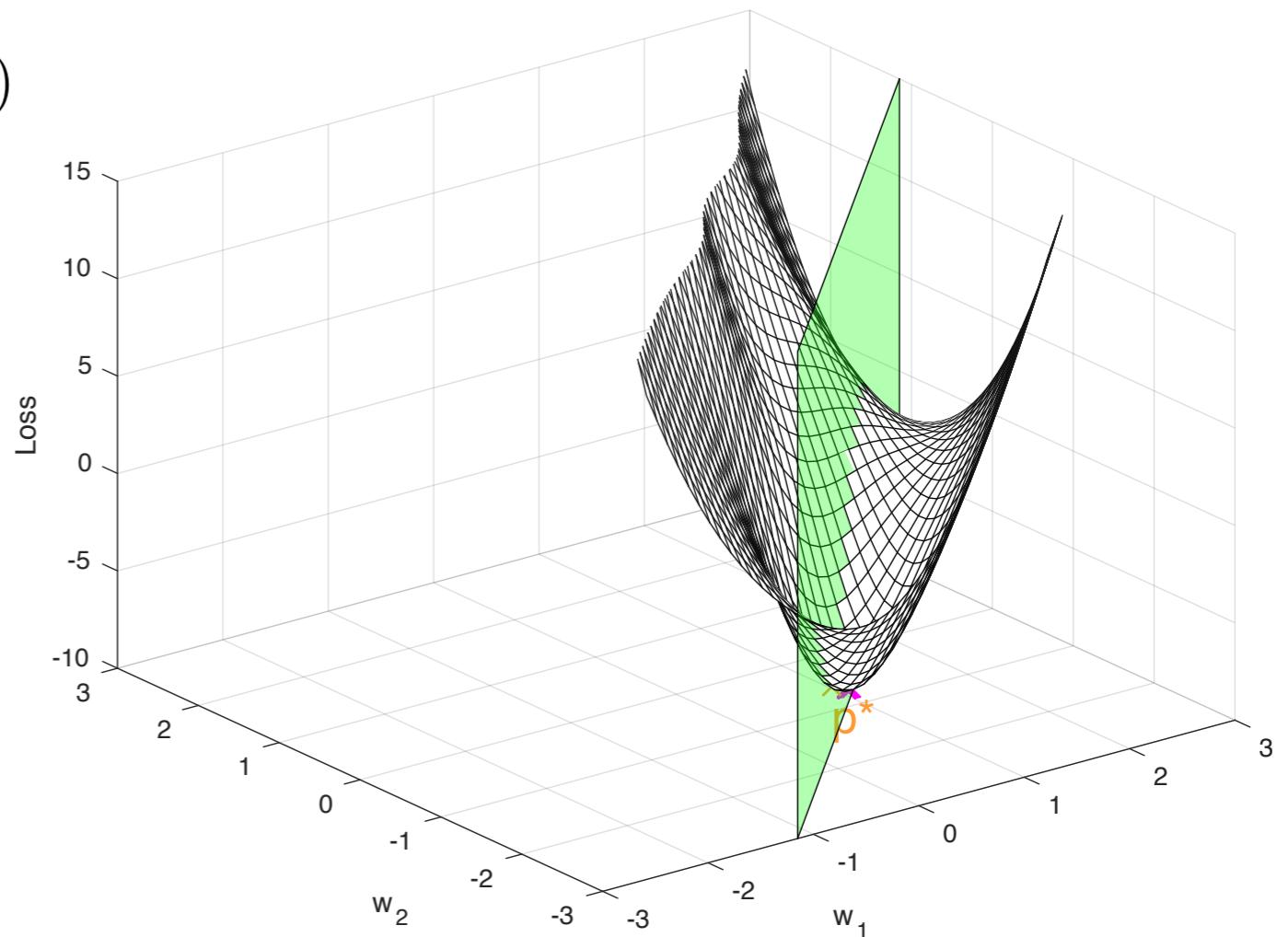


For $k = 0, 1, \dots$

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

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

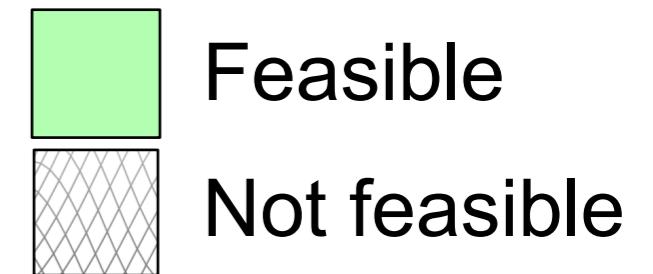
k	0	1	2	3	4	5
μ_k	0,125	0,25	0,5	1	2	4
λ_1^k	0	0,41	1,21	2,70	5,15	4,75
$c_1(w^k)$	-3,30	-3,19	-2,98	-2,45	-0,20	0,15



II. Constrained optimisation

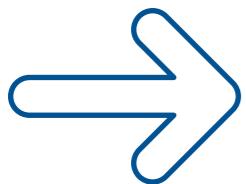
Augmented Lagrangian Method

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

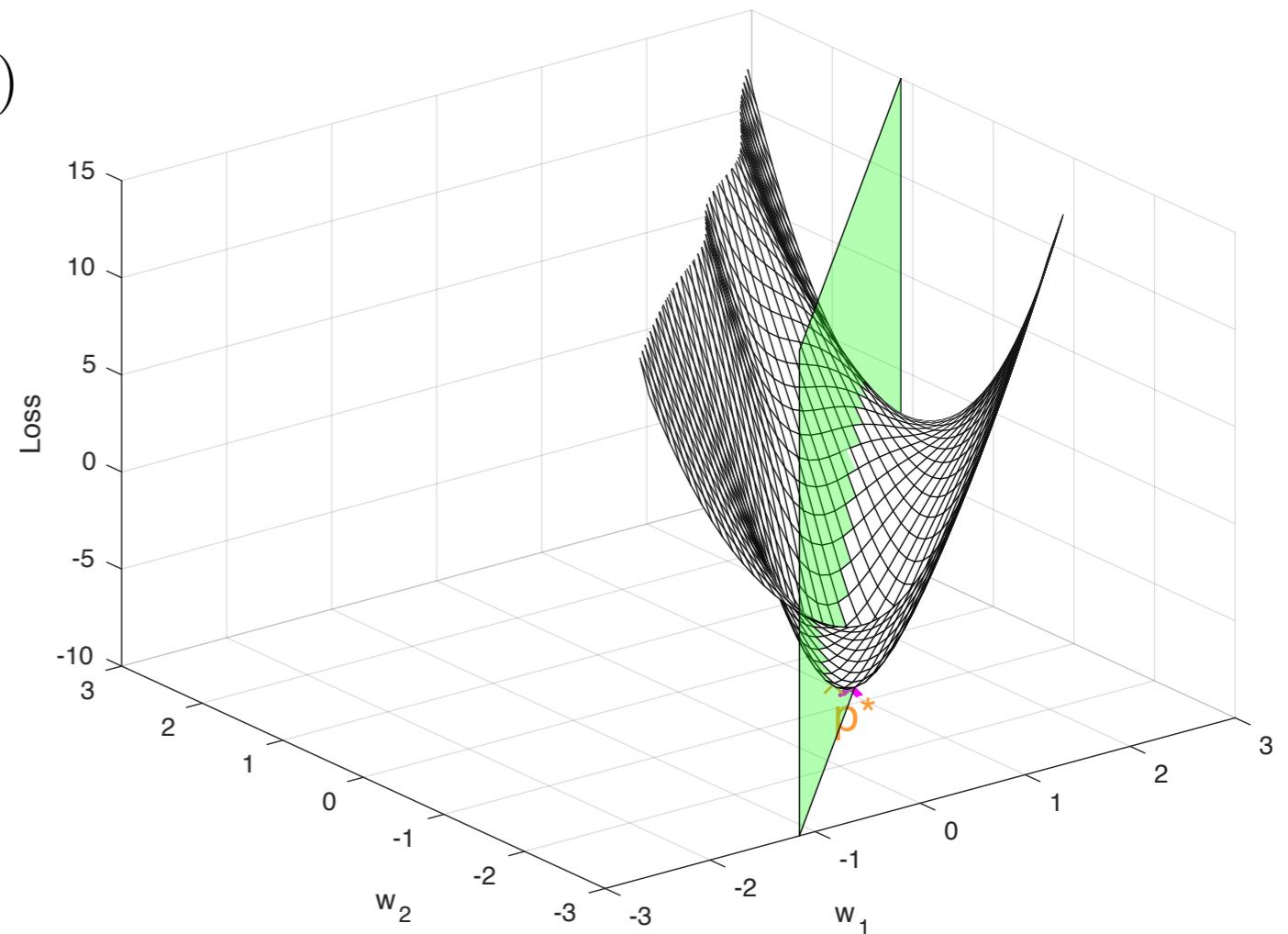


For $k = 0, 1, \dots$

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



Convergence already
with limited μ_k .



III. Experiment

Choosing a suitable experiment

Goals:

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

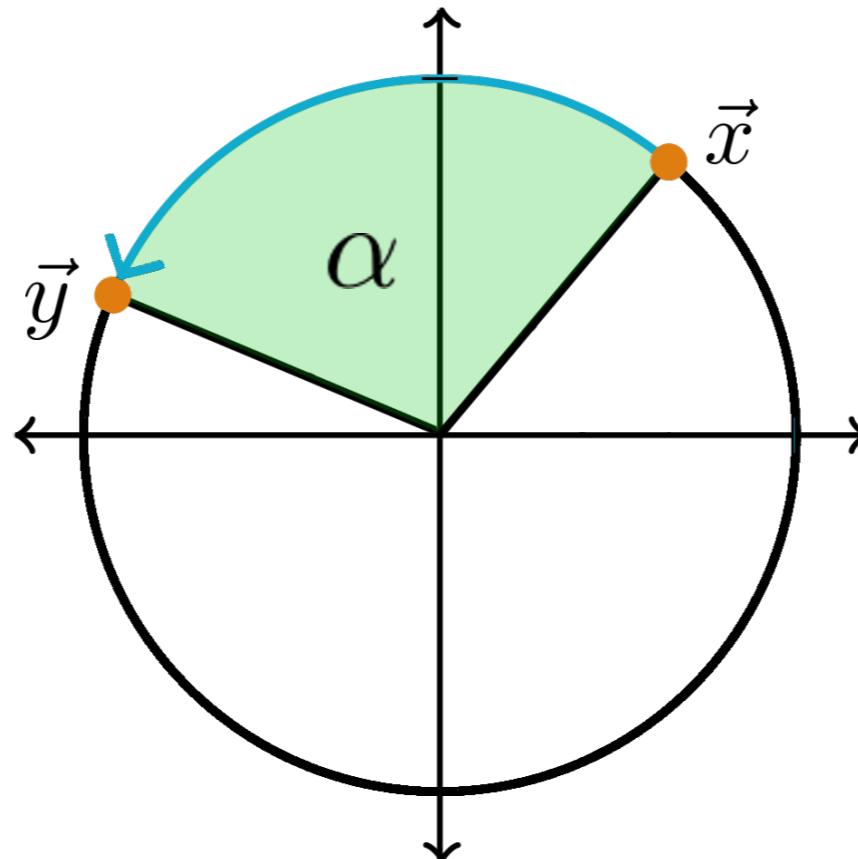
→ **Rotation**

Physical constraints:

- Determinant of rotation matrix equals one
- Norm preservation

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:

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

Norm constraint:

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

III. Experiment

Network architectures

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

III. Experiment

Comparison criteria

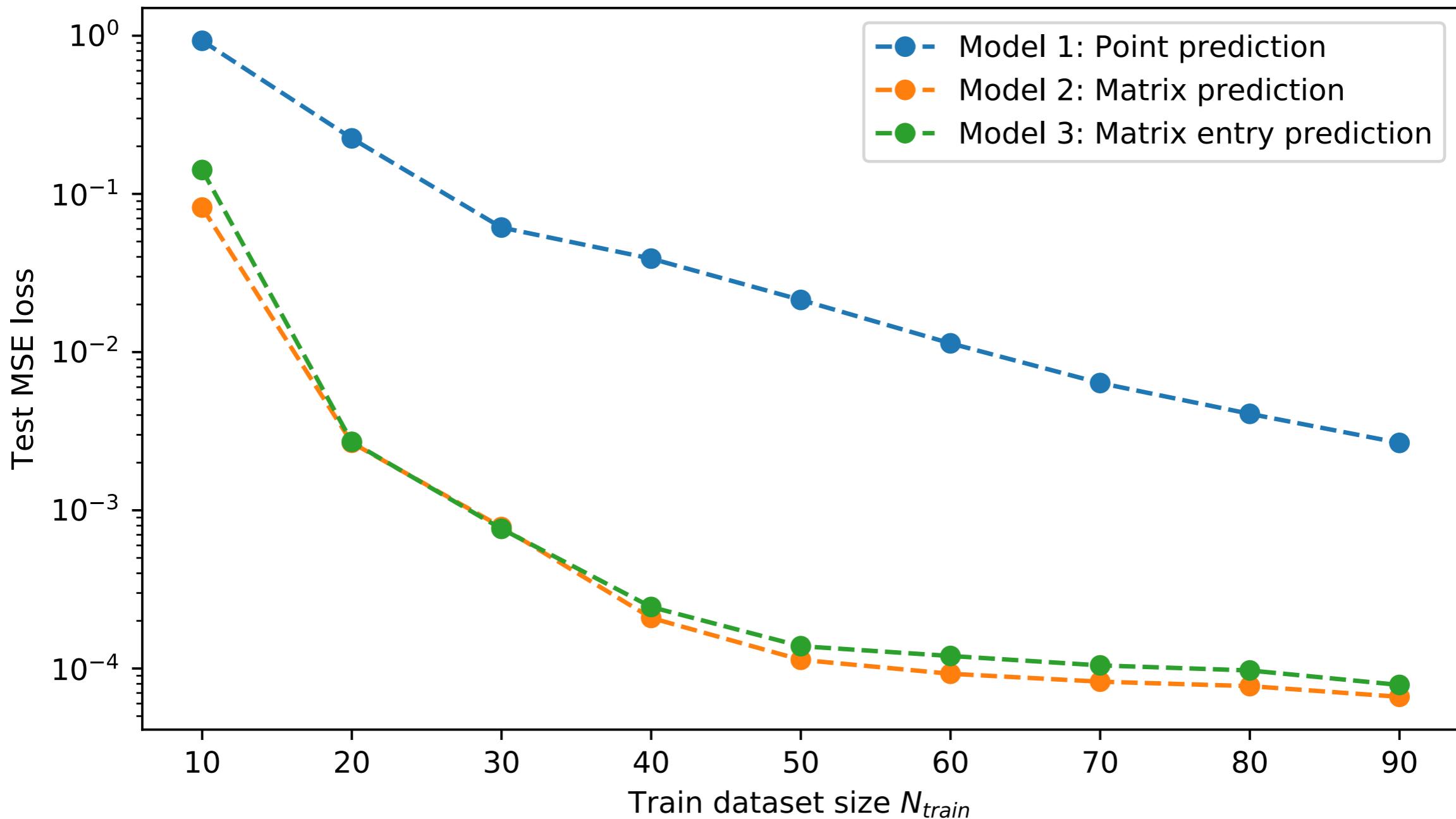
- Available data is currently the greatest limitation for deep learning
- Hyperparameters can be tuned

 **Compare performance w.r.t limited amount of available data**

IV. Results

Model baseline performance

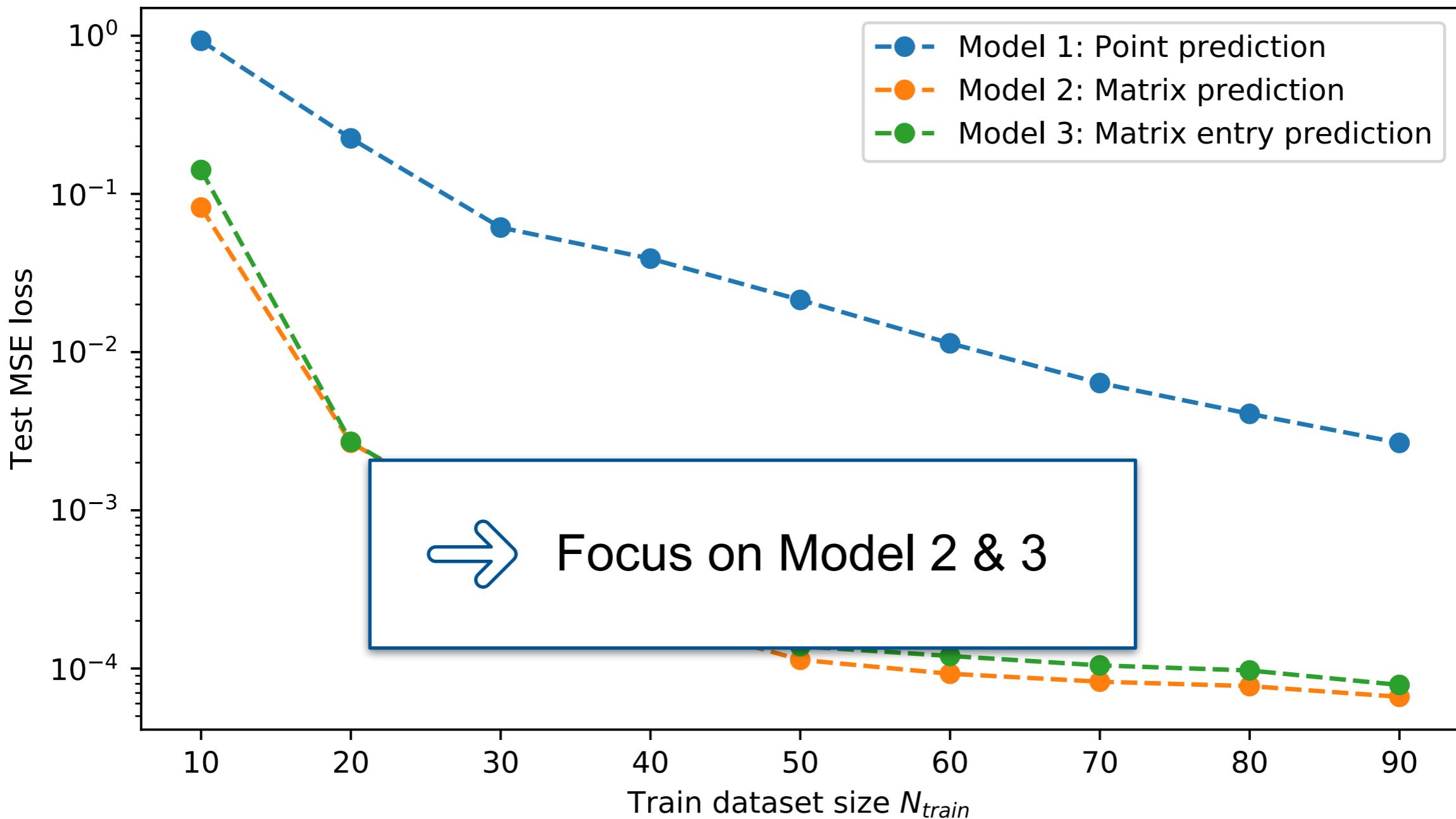
$$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y)$$



IV. Results

Model baseline performance

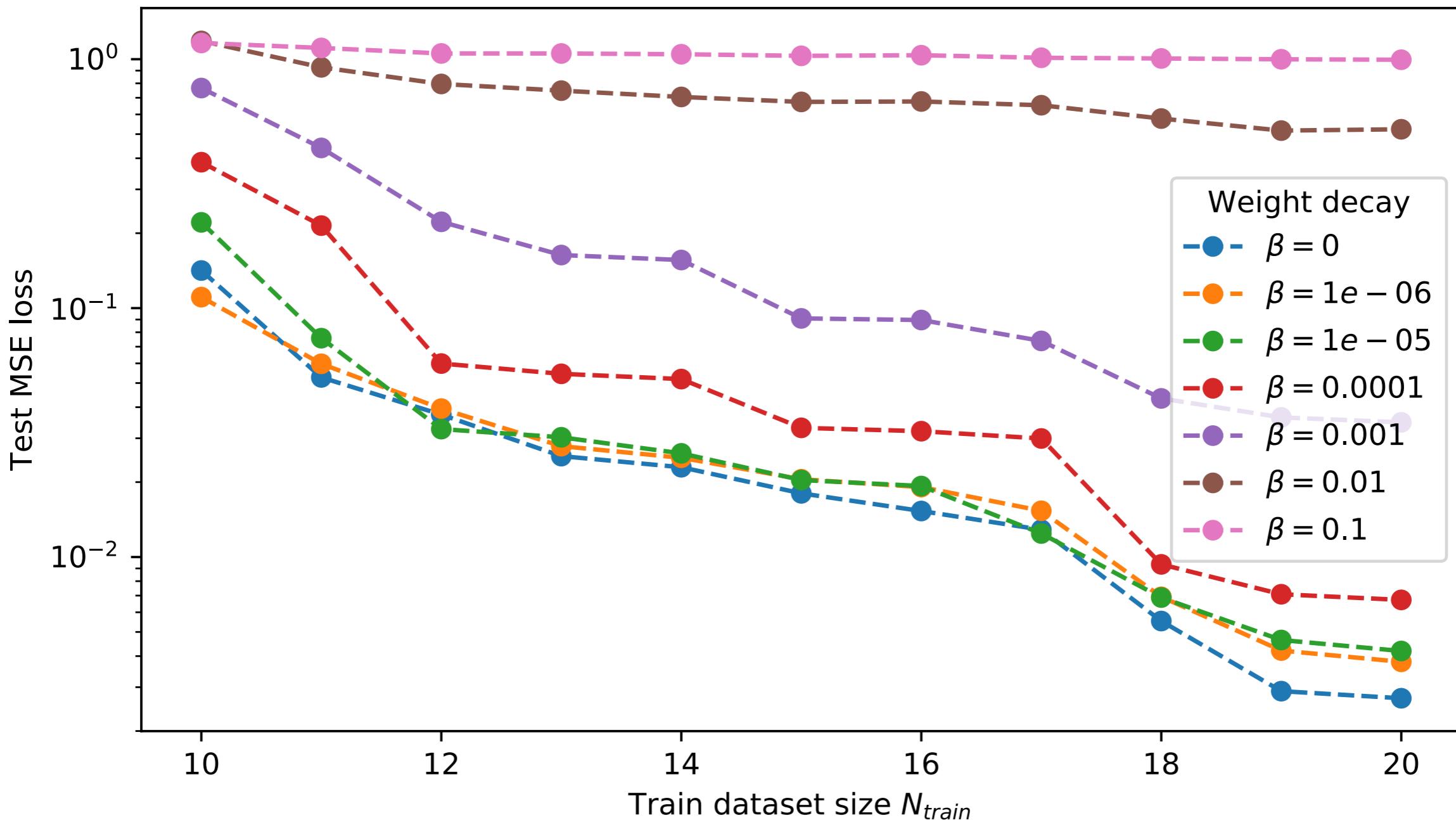
$$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y)$$



IV. Results

Model 3 with regularisation

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

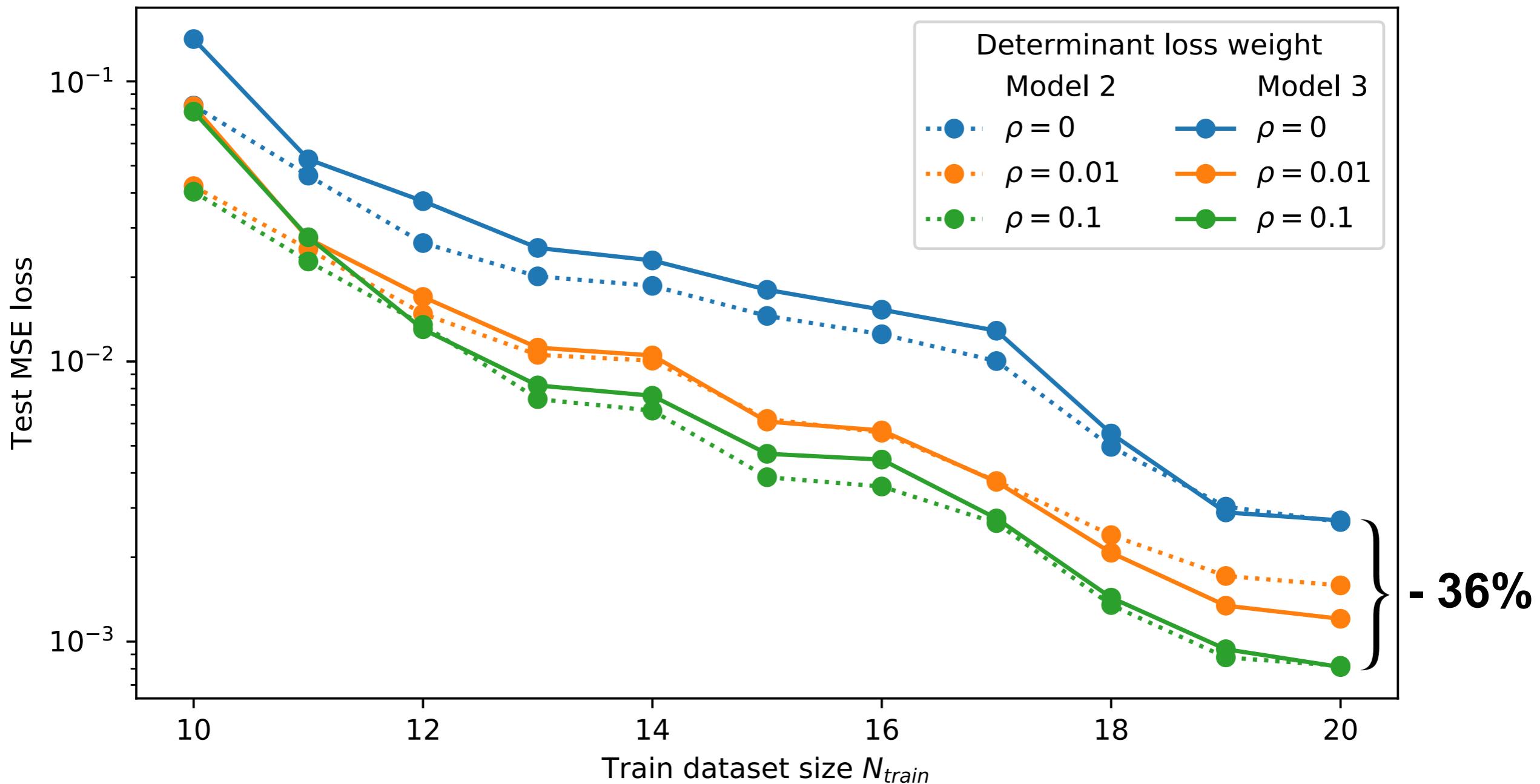


IV. Results

Penalty Method - Determinant constraint

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

\hat{R}_w : Predicted rotation matrices

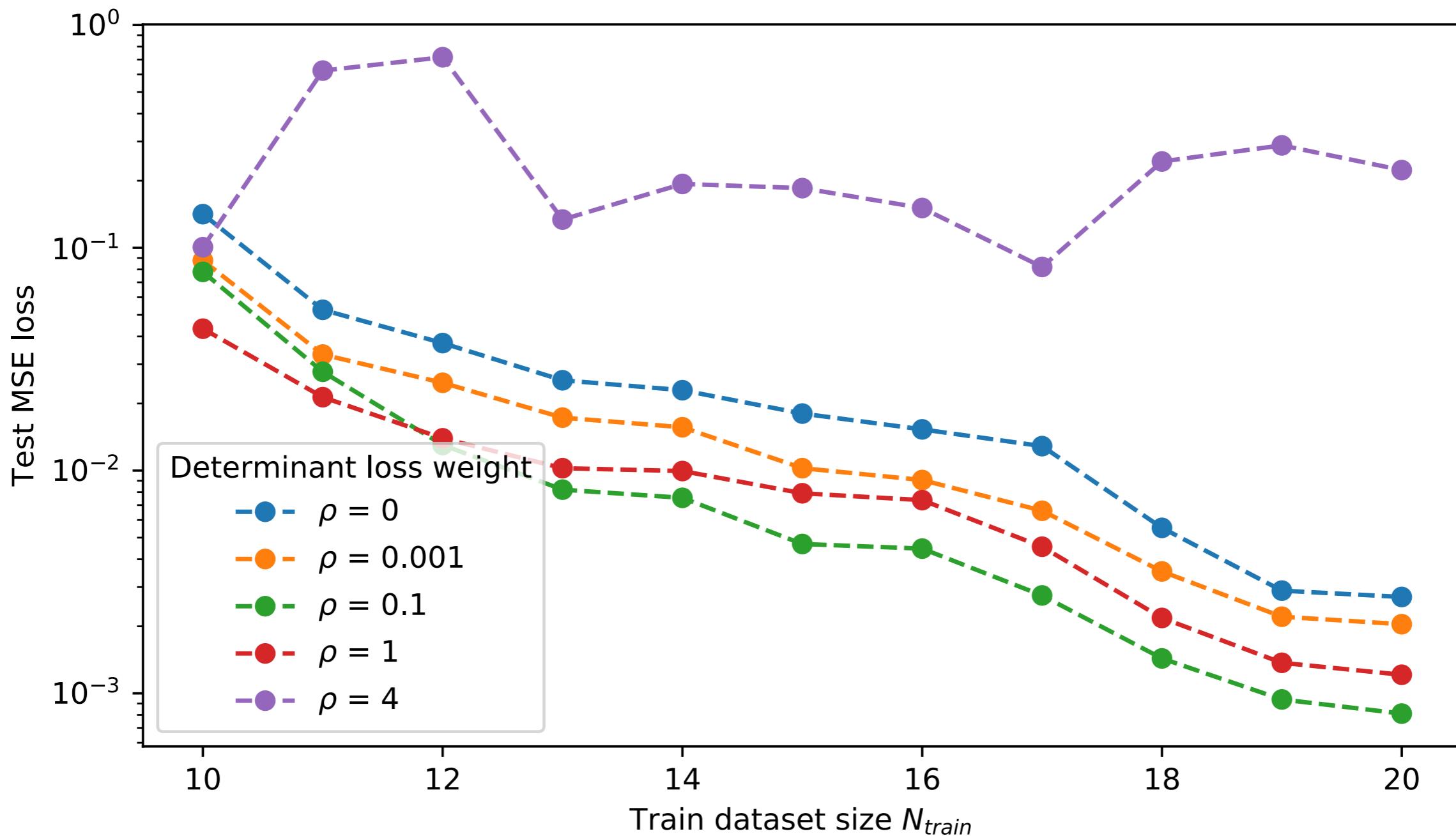


IV. Results

Penalty Method - Determinant constraint - Model 3

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

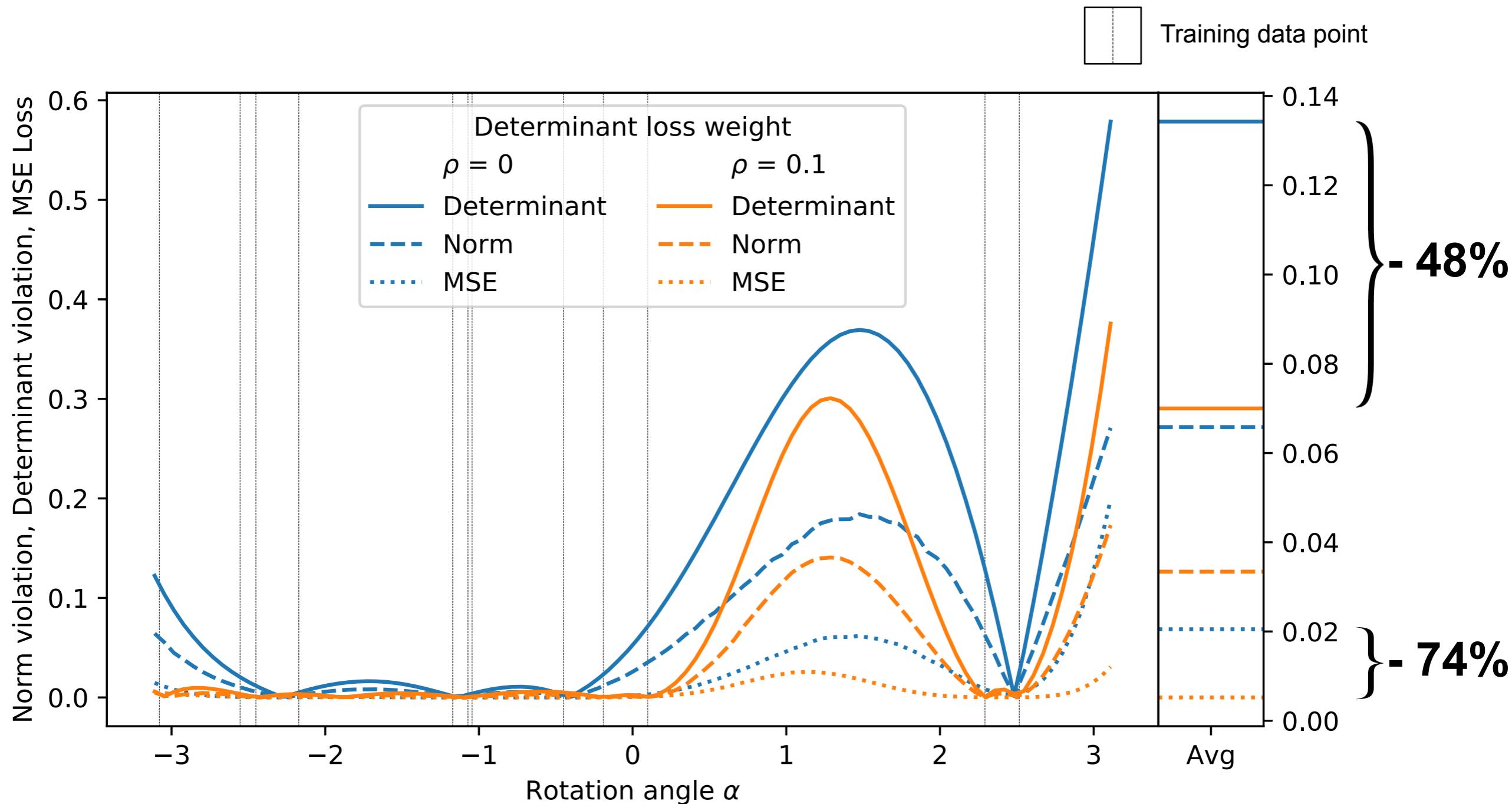
\hat{R}_w : Predicted rotation matrices



IV. Results

Penalty Method - Determinant constraint - Model 3

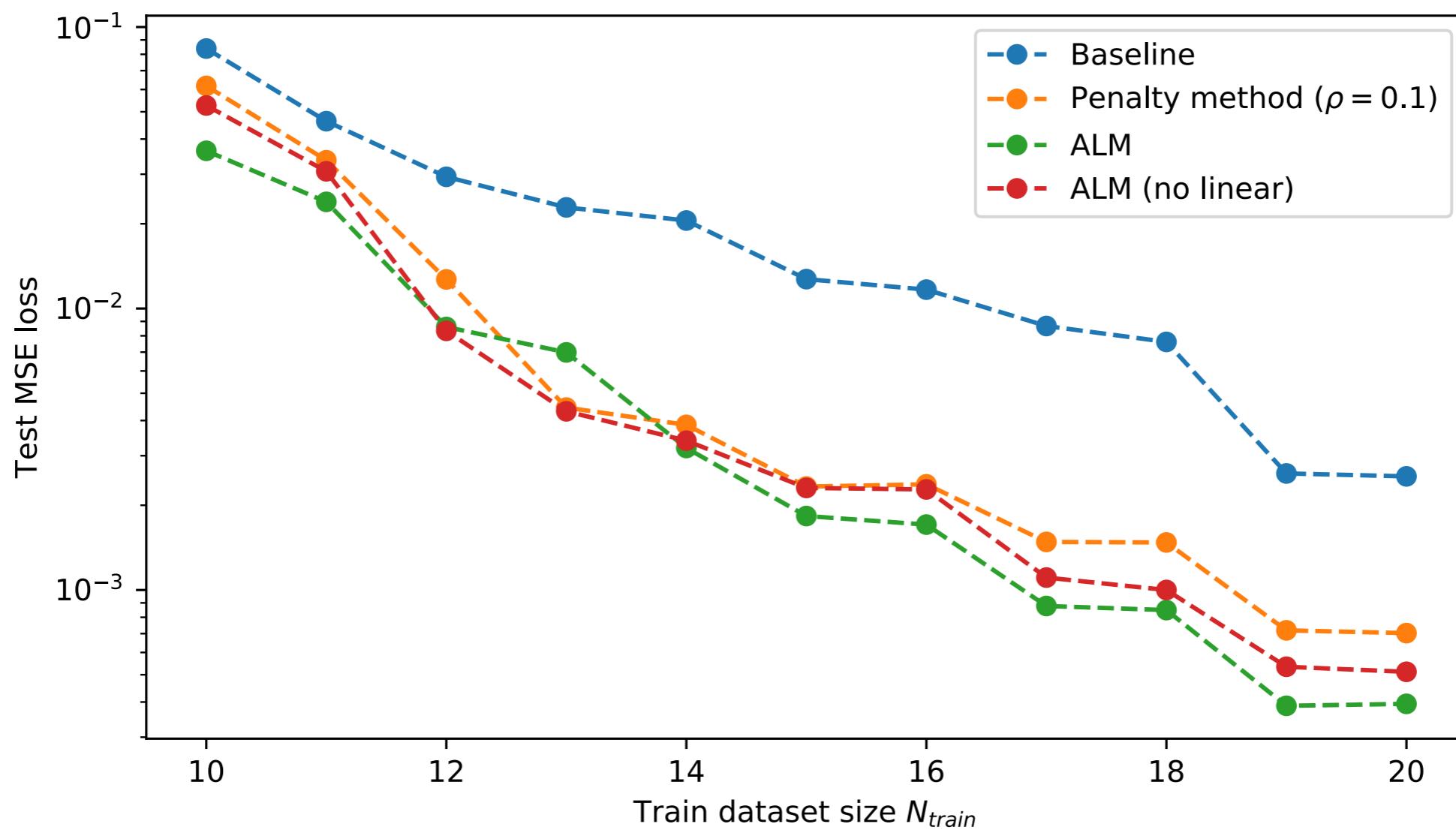
Result of Model 3 trained on 12 training data points



IV. Results

Augmented Lagrangian Method - Determinant constraint

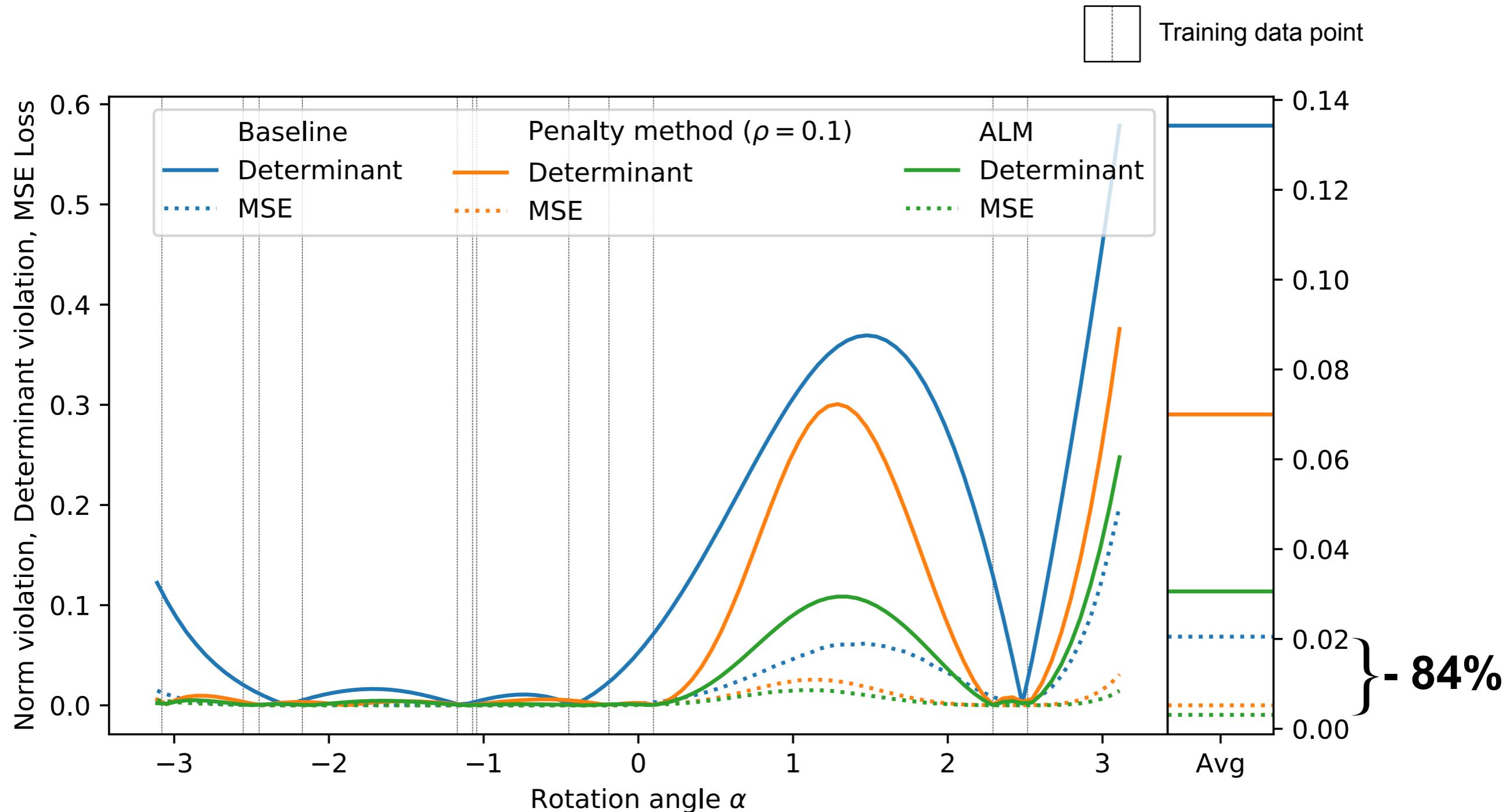
Baseline	$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y)$	Model 3
Penalty Method	$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y) + \rho \sum_{n=1}^{N_{train}} (\det(\hat{R}_{w_n}) - 1)^2$	
ALM	$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y) + \sum_{n=1}^{N_{train}} \lambda_n (\det(\hat{R}_{w_n}) - 1) + \mu_k \sum_{n=1}^{N_{train}} (\det(\hat{R}_{w_n}) - 1)^2$	
ALM (no linear)	$\mathcal{L}(w) = \text{MSE}(\hat{y}_w, y) + \mu_k \sum_{n=1}^{N_{train}} (\det(\hat{R}_{w_n}) - 1)^2$	



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:

- Difficult to find good parameters
- 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

References

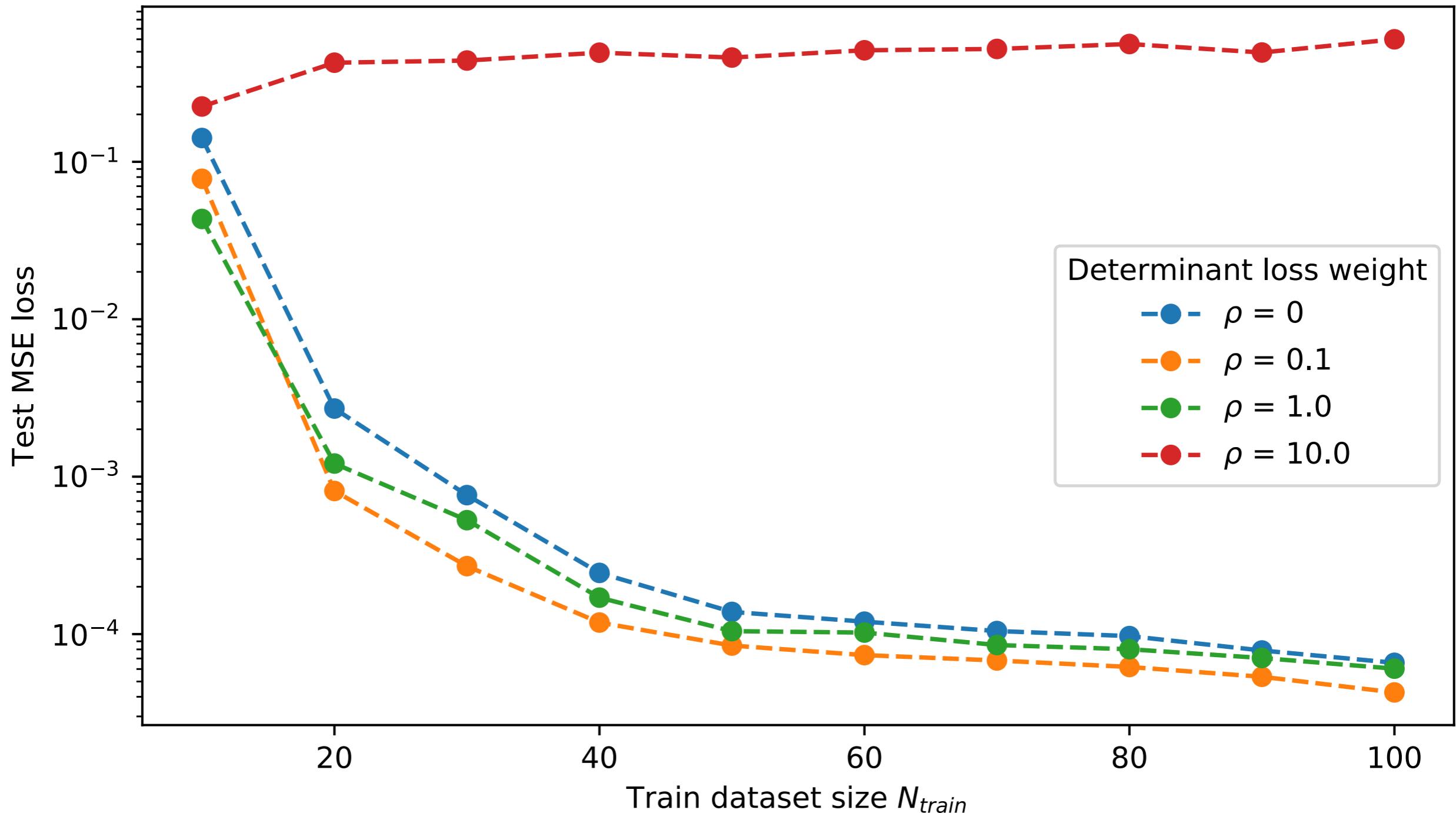
- Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second edition, 2006.
- Jack Yurkiewicz. Constrained optimization and lagrange multiplier methods, by d. p. bertsekas, academic press, new york, 1982, 395 pp. price: \$65.00. *Networks*, 15:138–140, 1985.
- 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_oBff0NIA087mkfhKQ9dujNPhQYjNOEoo3drYeulw7Ov0HRfQ
- <http://pedynamic.com/circle-antipode-experiments/>
- <http://pedynamic.com/circle-antipode-simulation/>

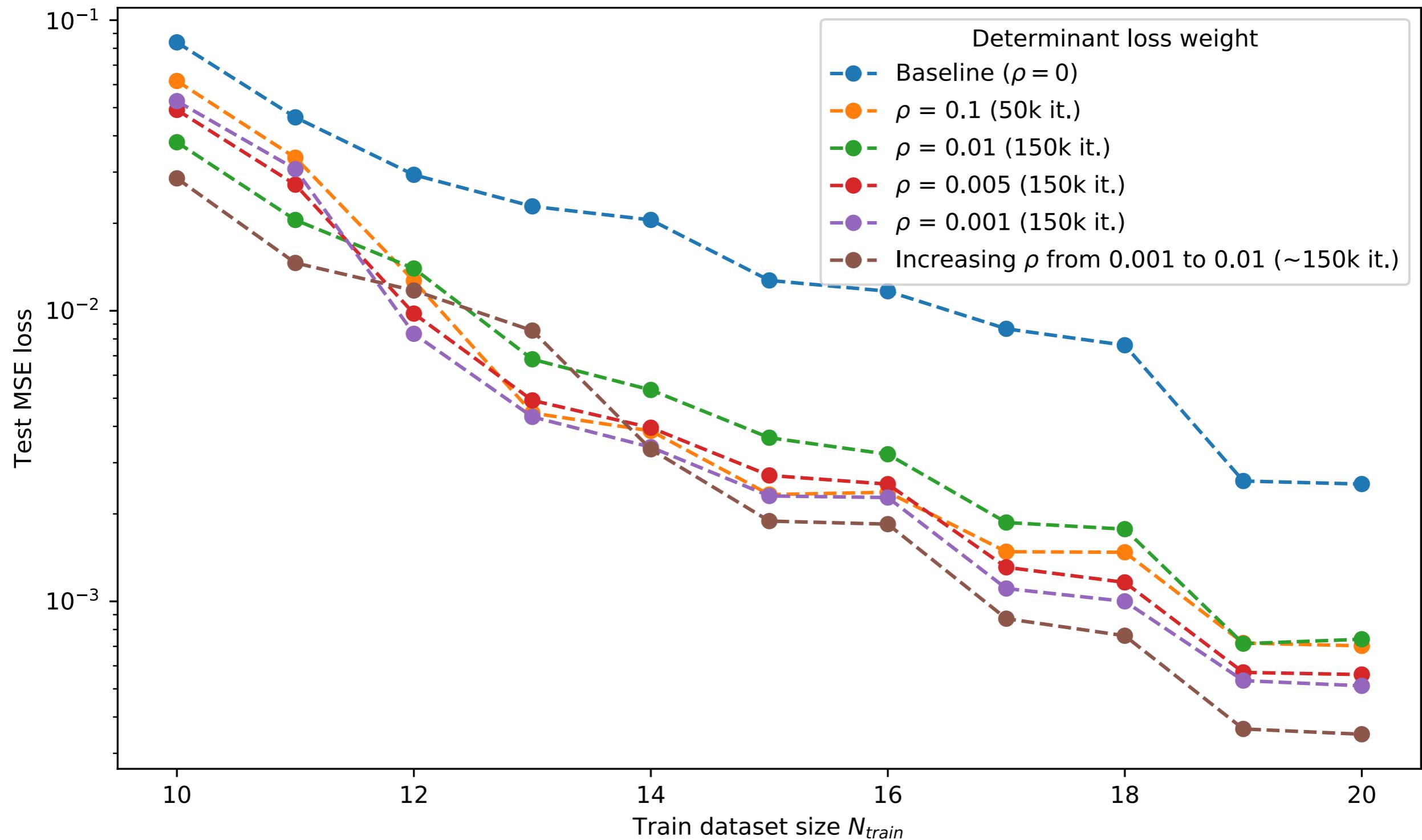
V. Appendix

Penalty Method Model 3



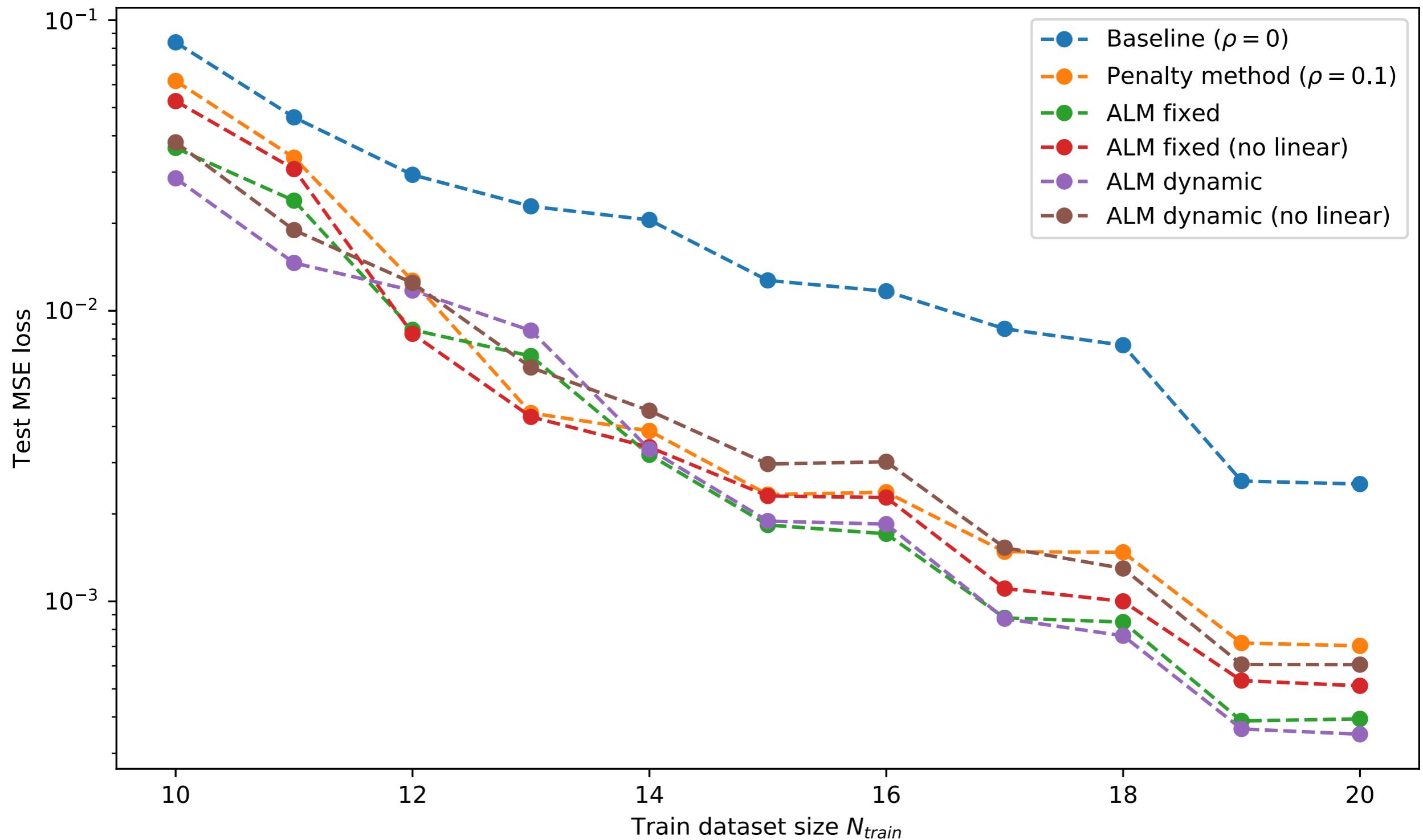
V. Appendix

Penalty Method Model 3 - Increasing weight



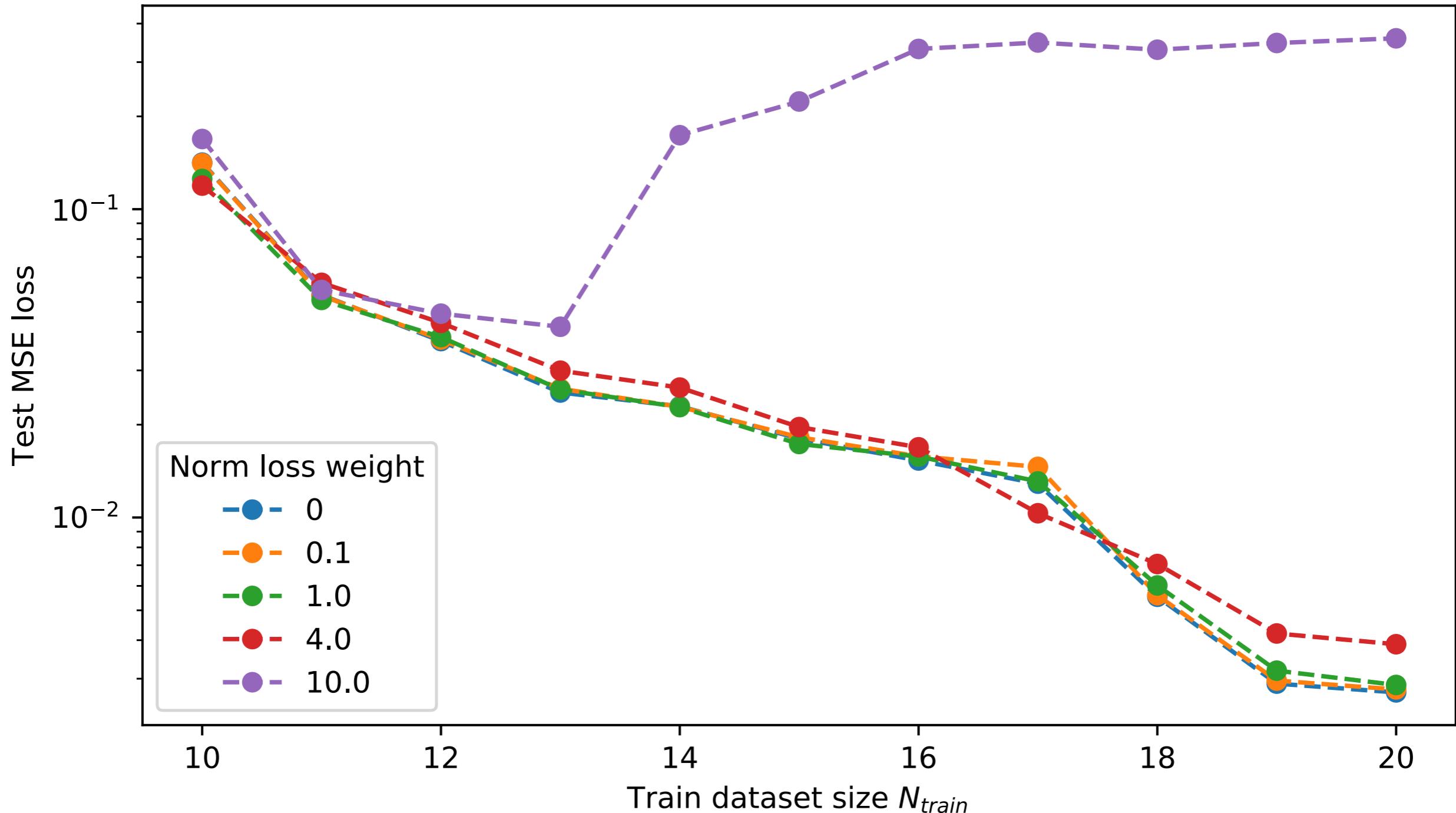
V. Appendix

ALM Model 3 - Fixed and Dynamic



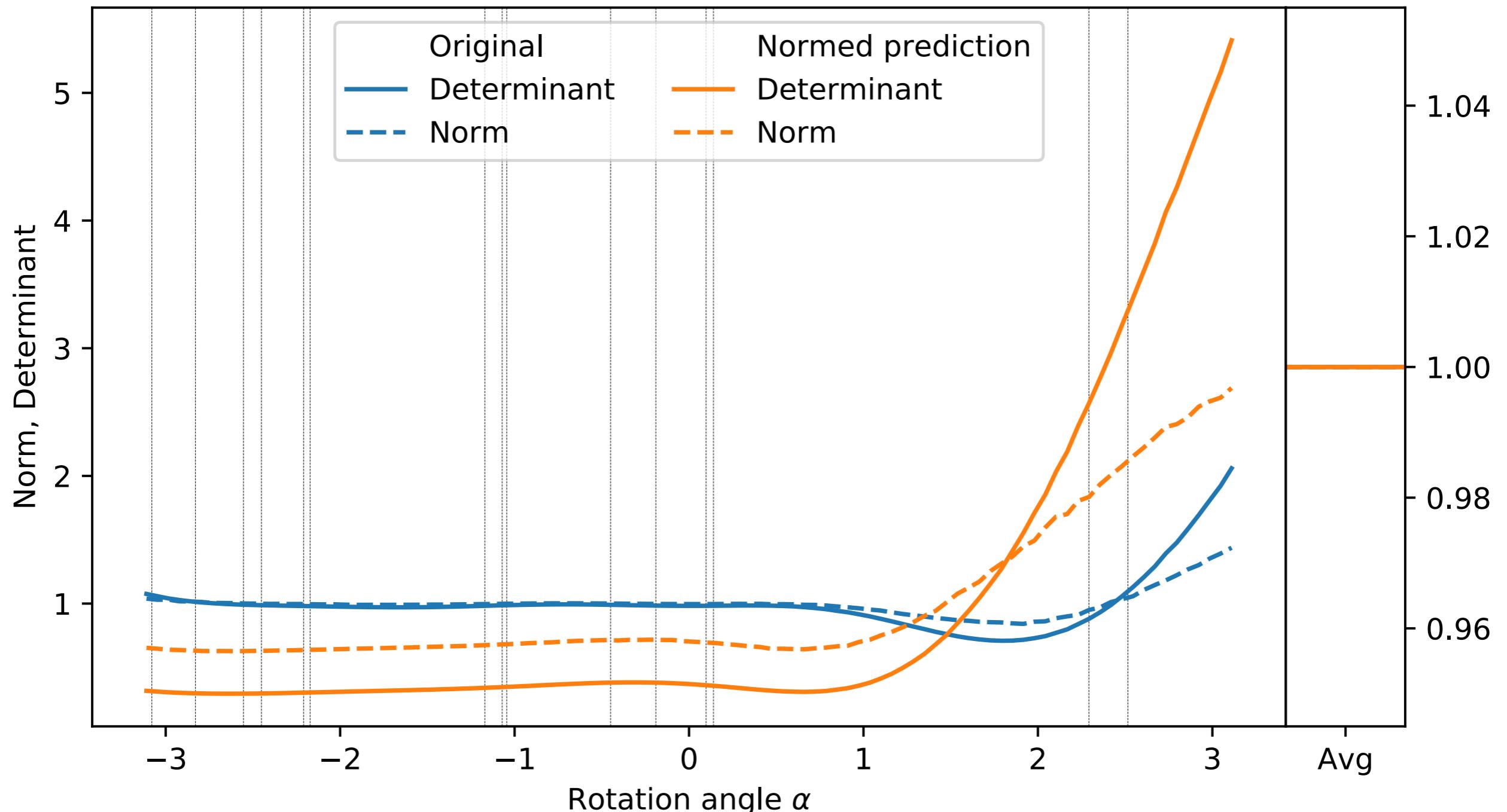
V. Appendix

Penalty Method Model 3 - Norm constraint



V. Appendix

Normed prediction Model 3



V. Appendix

Penalty Method Model 3 - 3 dimensional rotation

