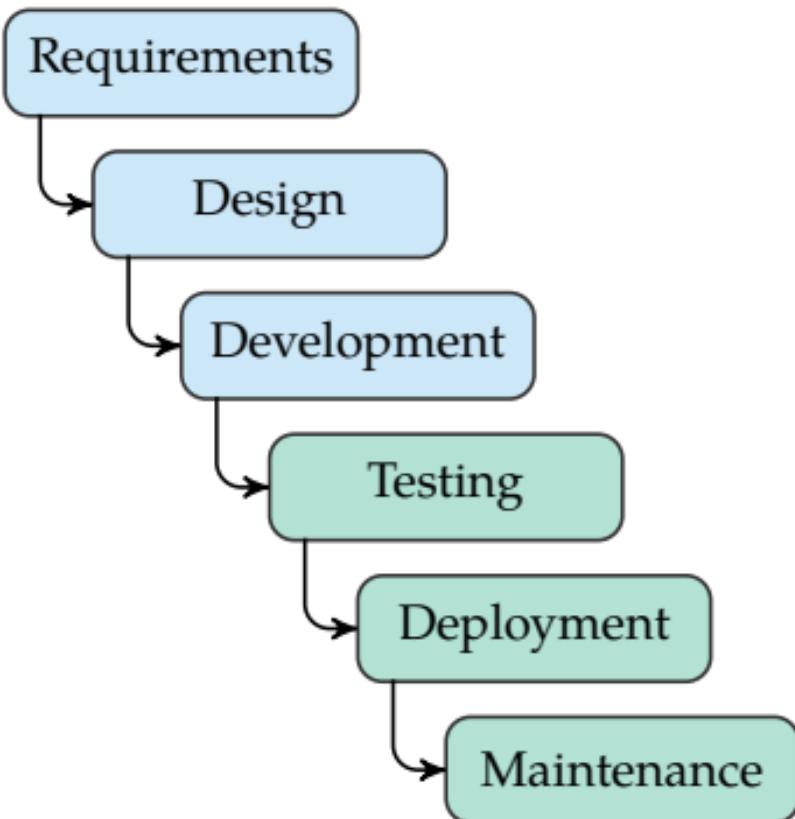


Define Requirements

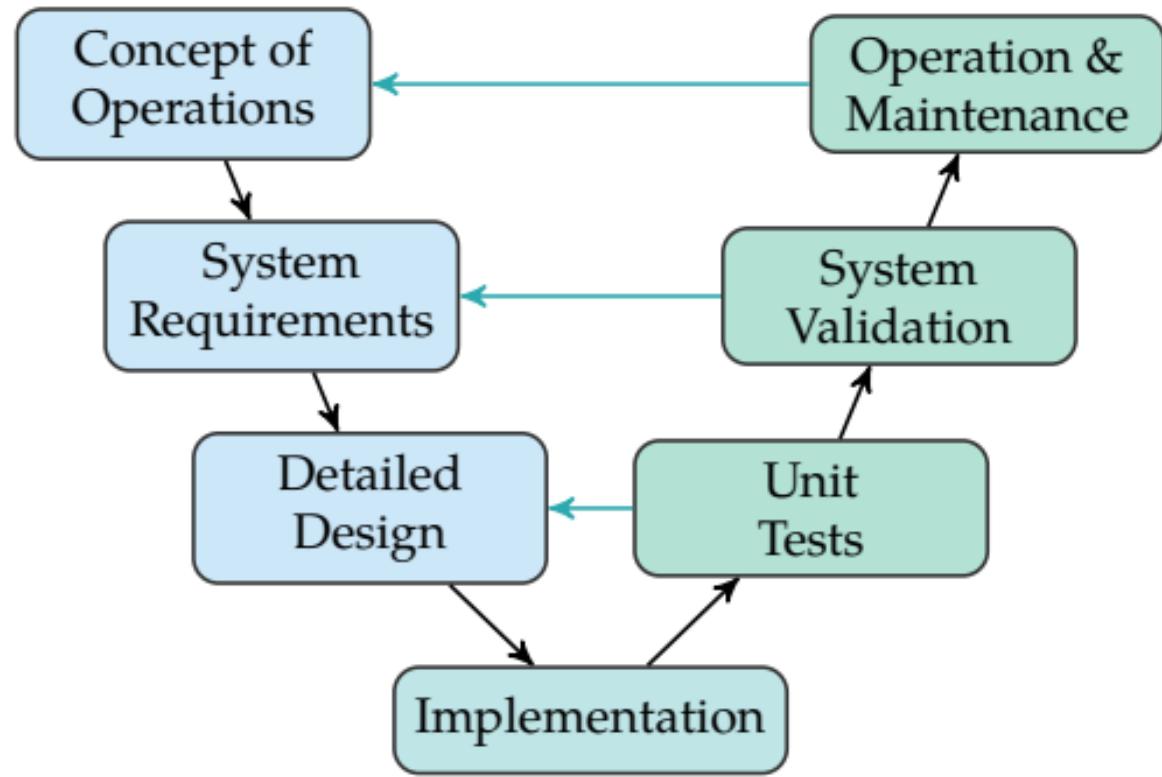
Design

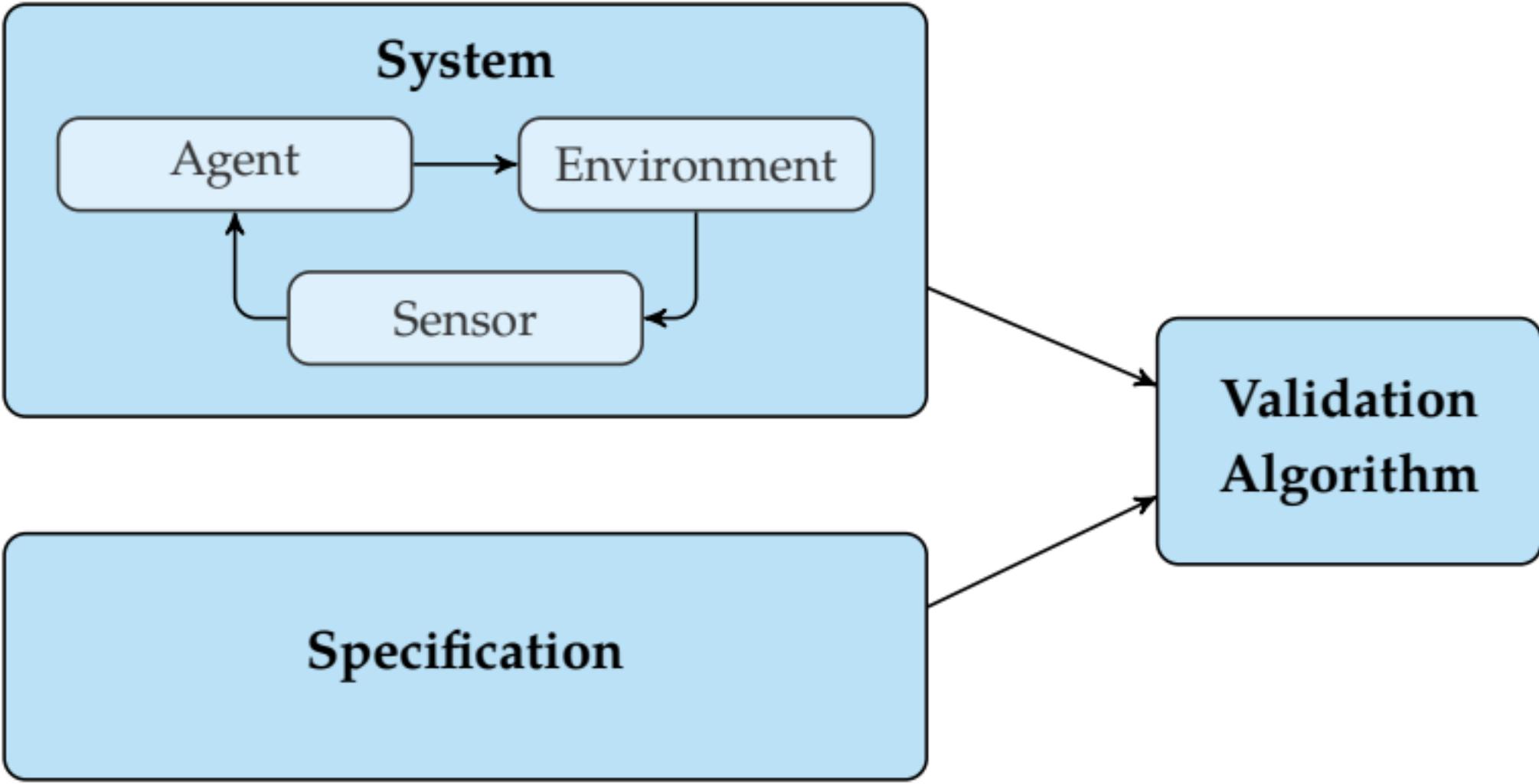
Validate

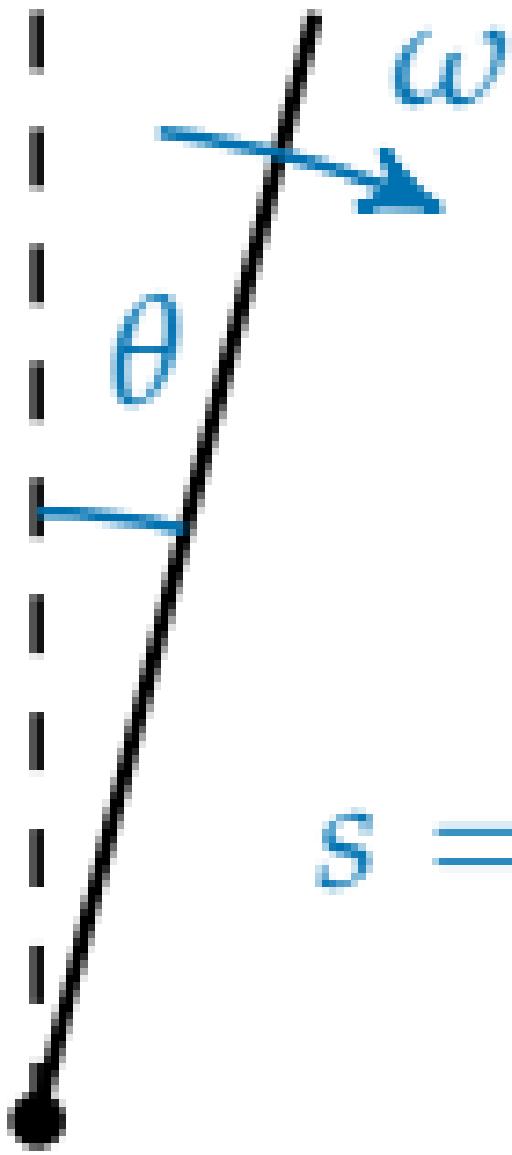
# Waterfall Model



# V Model

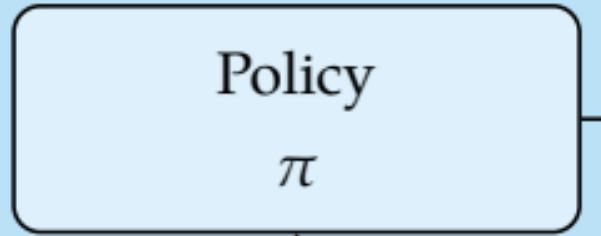






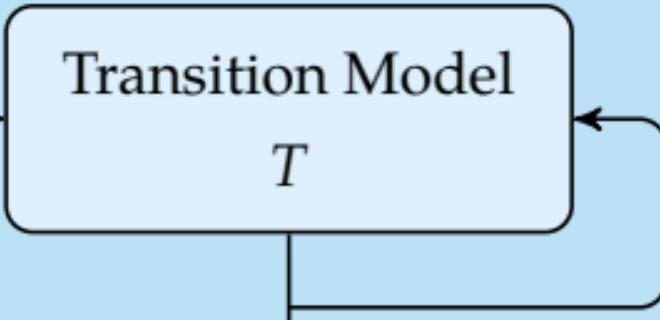
$$s = [\theta, \omega]$$

**Agent**

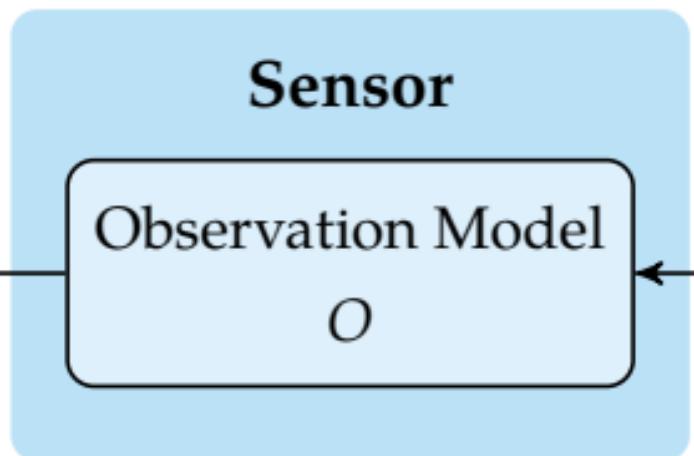


action  $a$

**Environment**

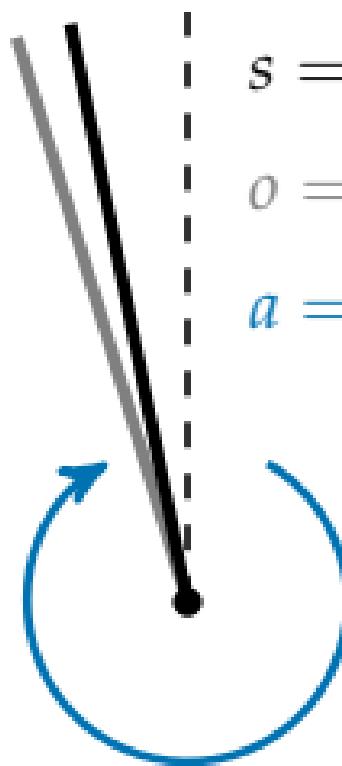


observation  $o$

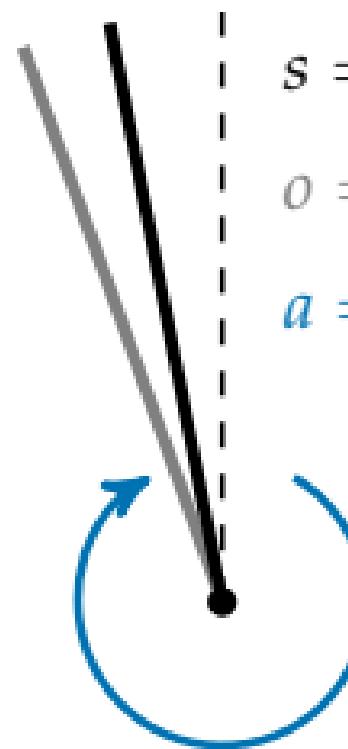


state  $s$

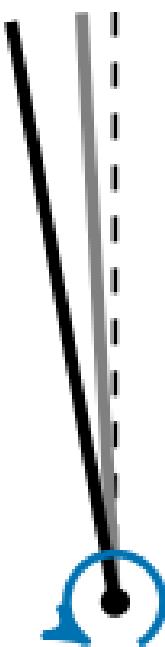
Step 1



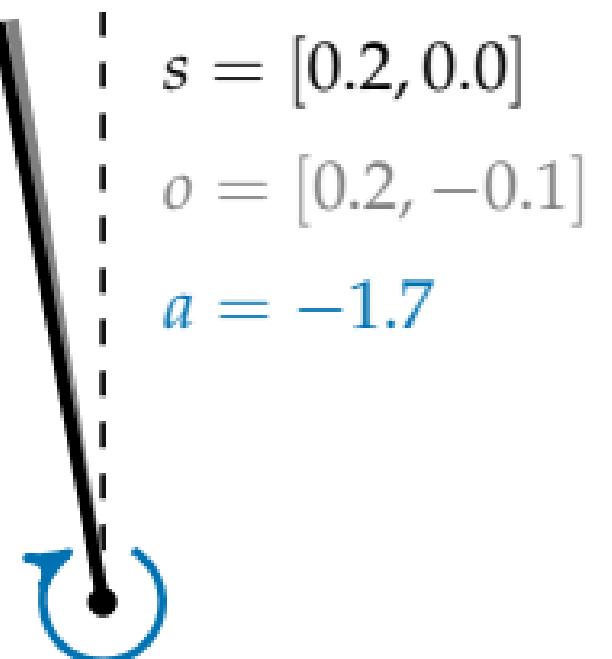
Step 2

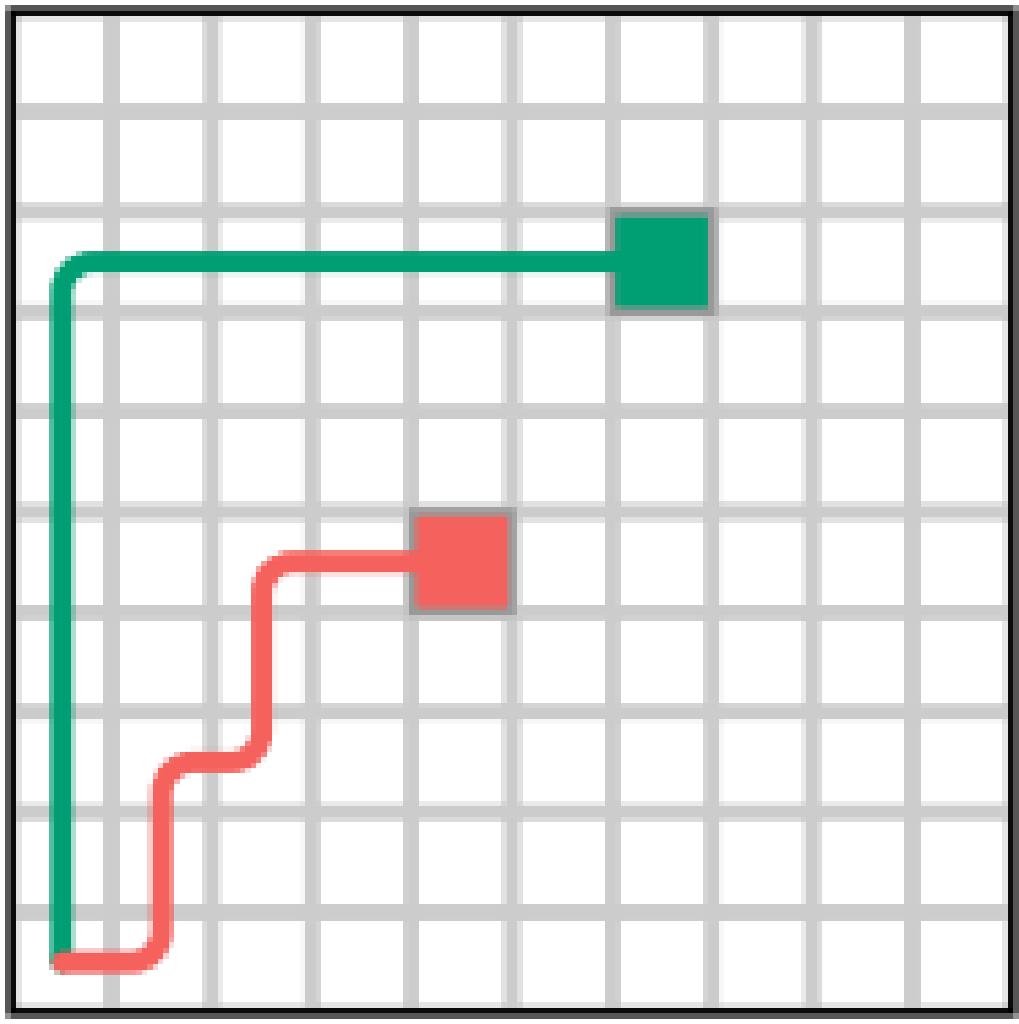


Step 3

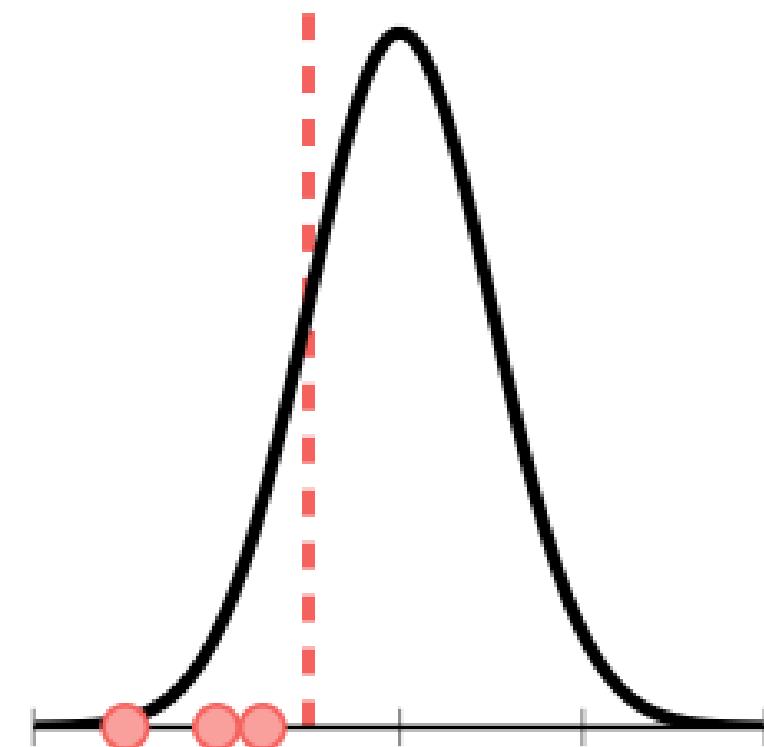


Step 4

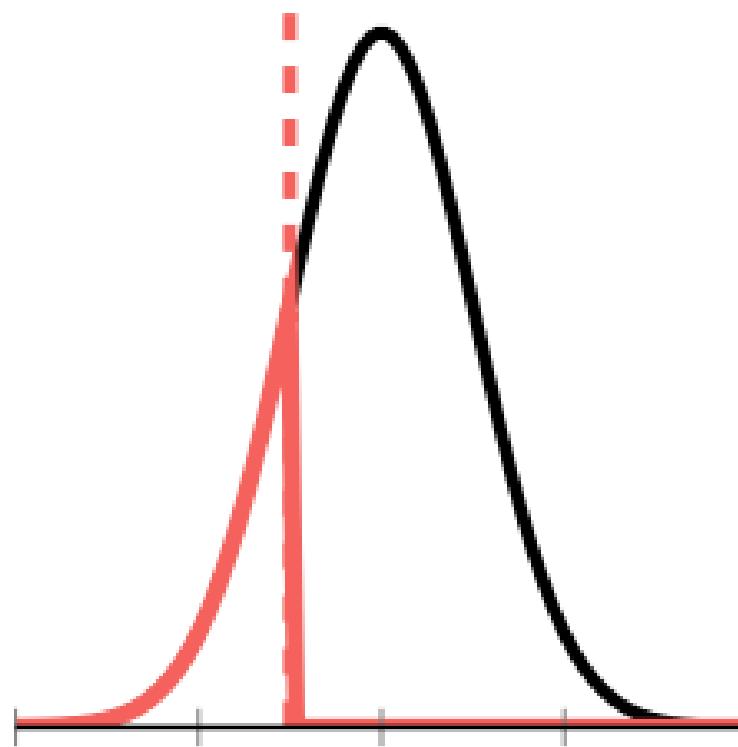




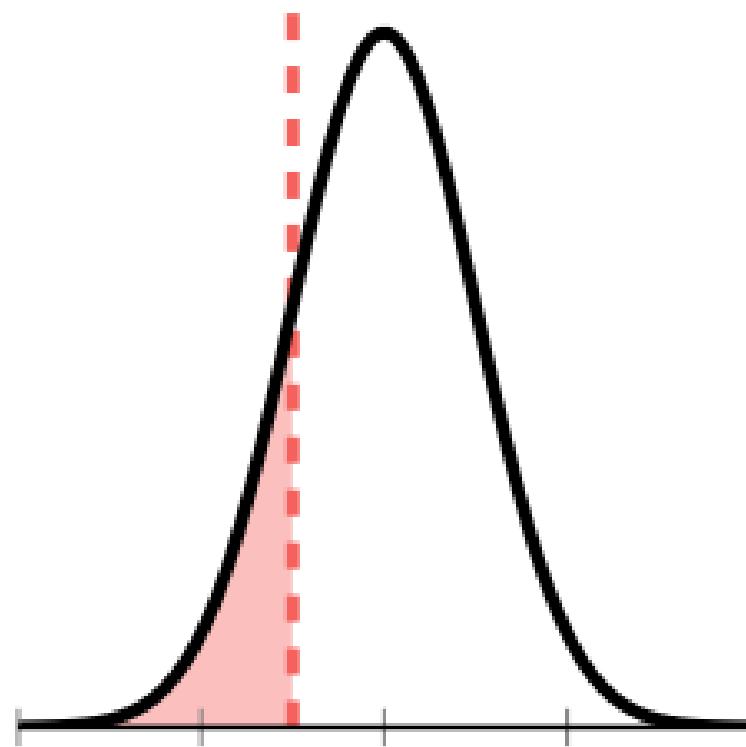
Falsification



Failure Distribution



Failure Probability

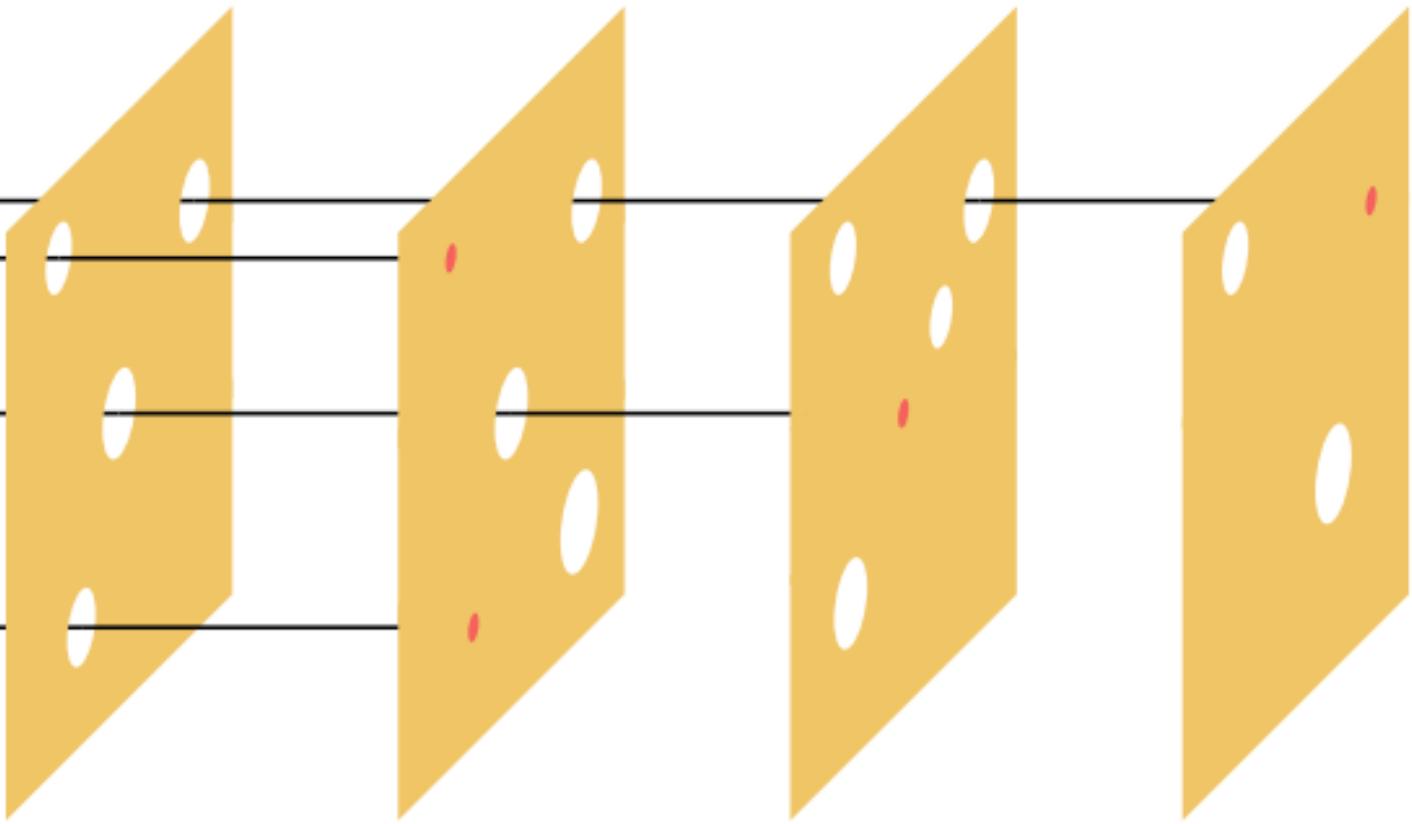


Failure  
Analysis

Formal  
Guarantees

Explanations

Runtime  
Assurances



$P(x)$ 

0.3

0.2

0.1

1

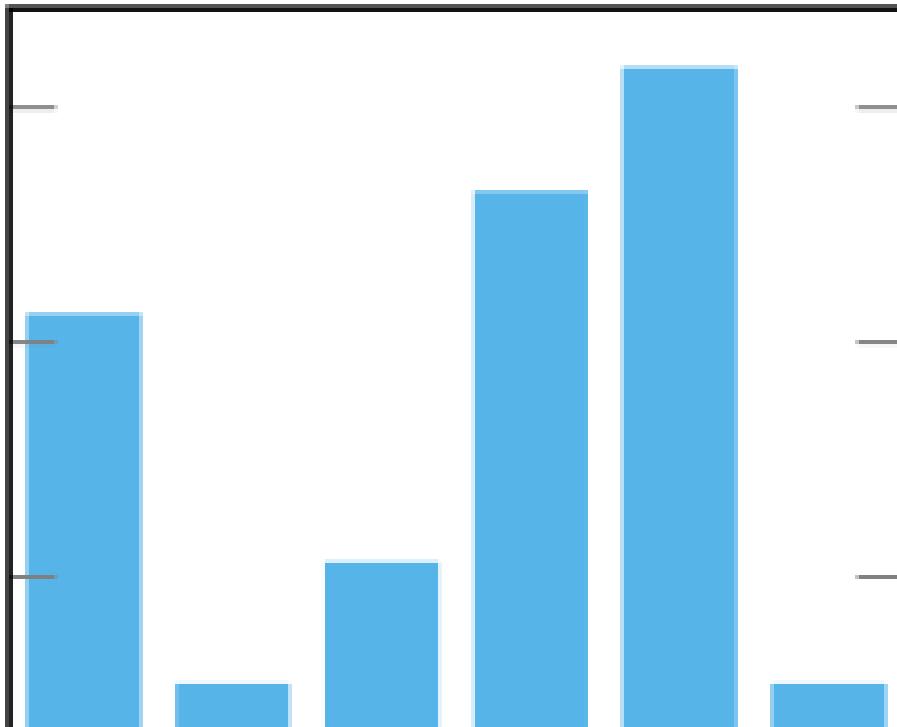
2

3

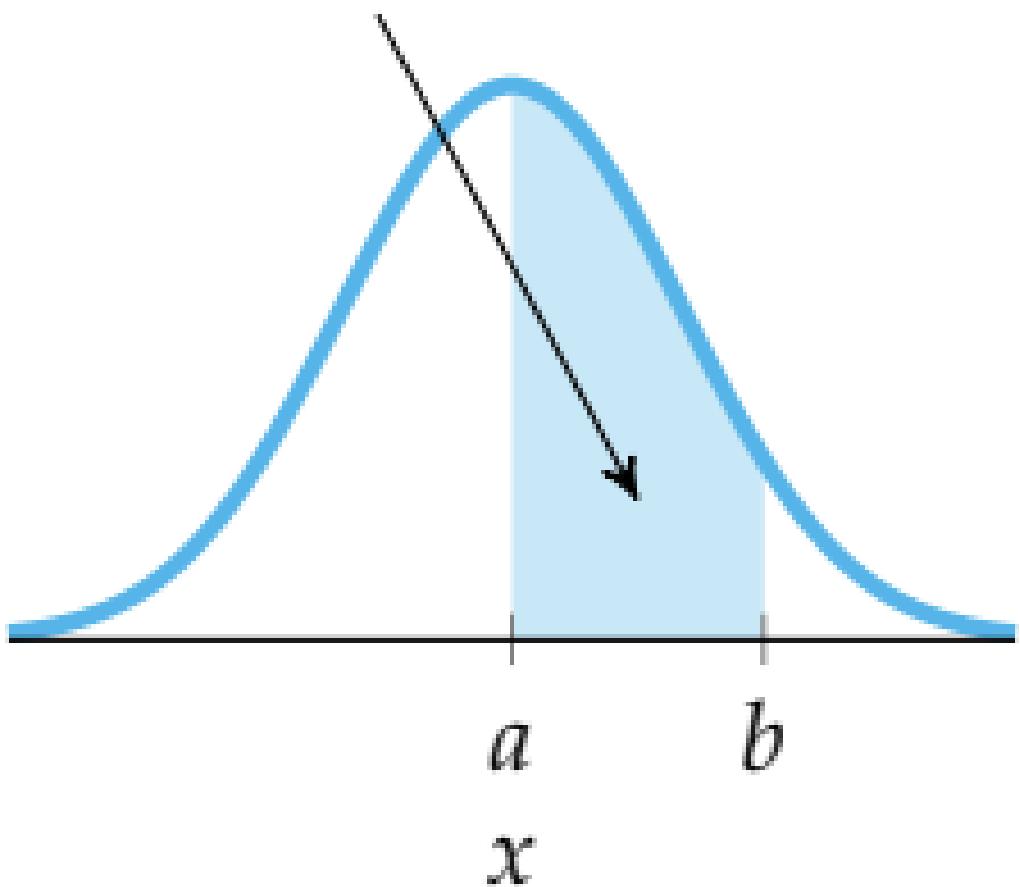
4

5

6

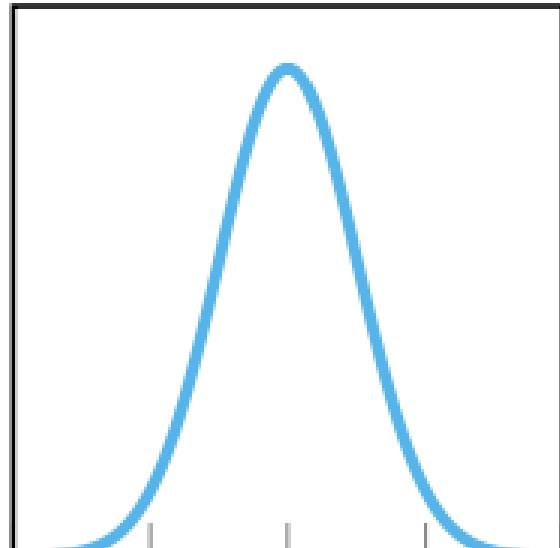
 $x$ 

$$P(a \leq x \leq b)$$



$$\mu = 0$$

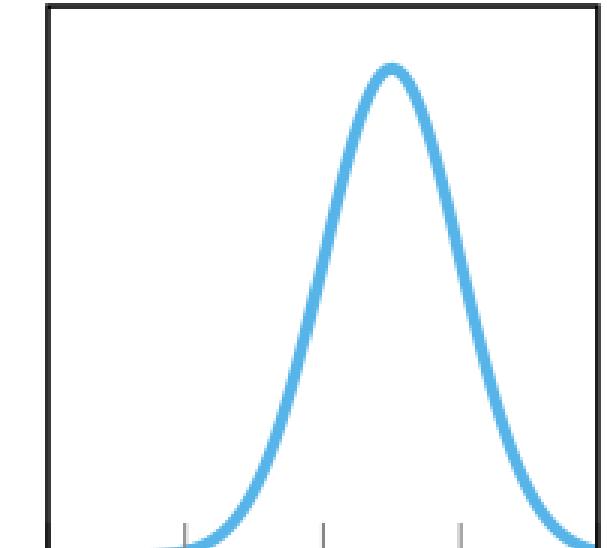
$$\sigma^2 = 1$$



$x$

$$\mu = 1$$

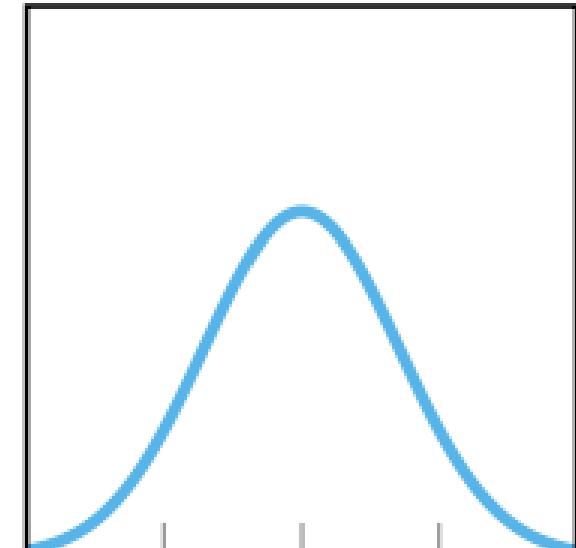
$$\sigma^2 = 1$$



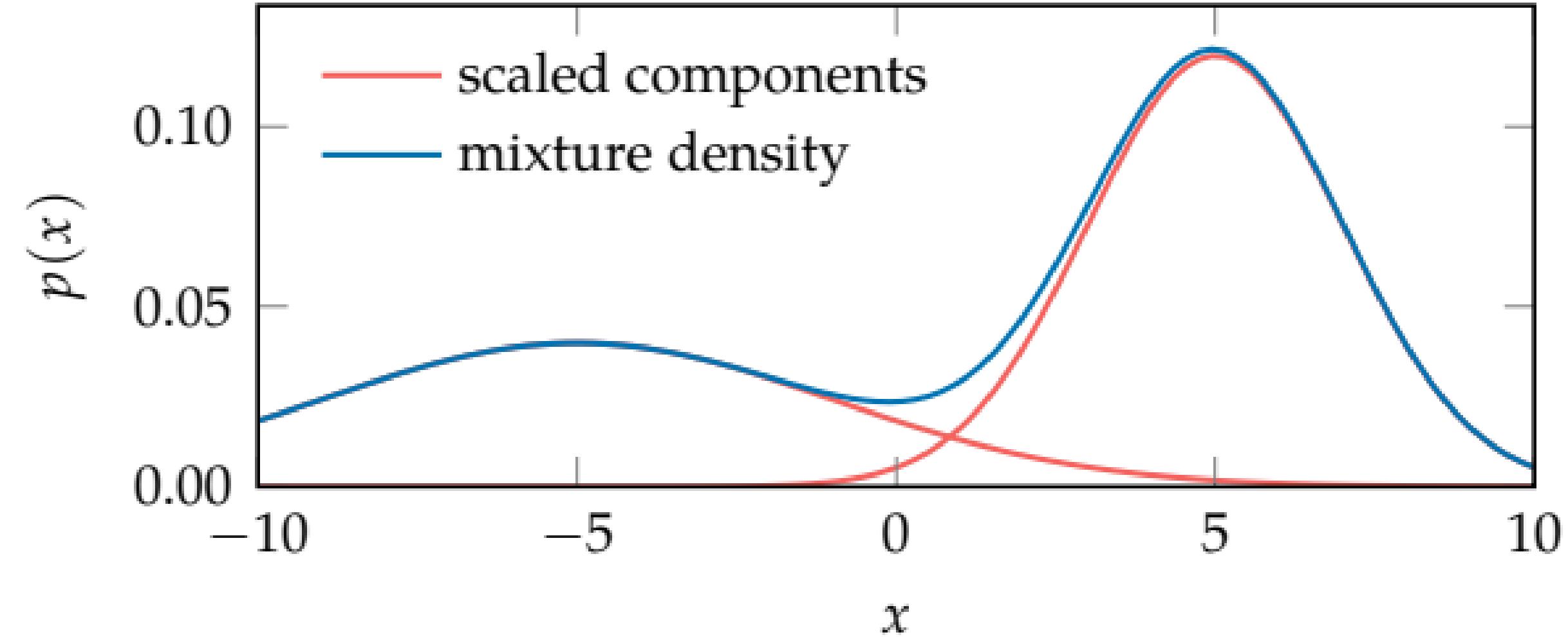
$x$

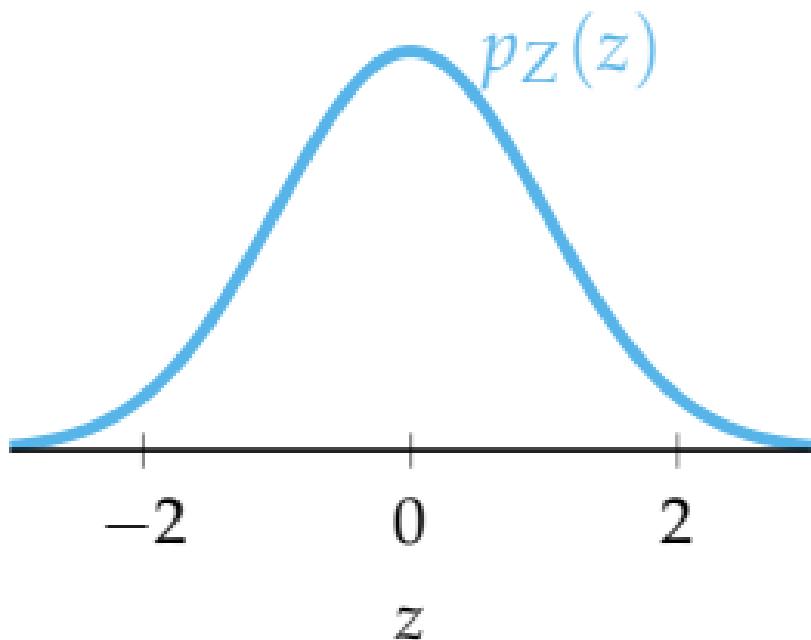
$$\mu = 0$$

$$\sigma^2 = 2$$

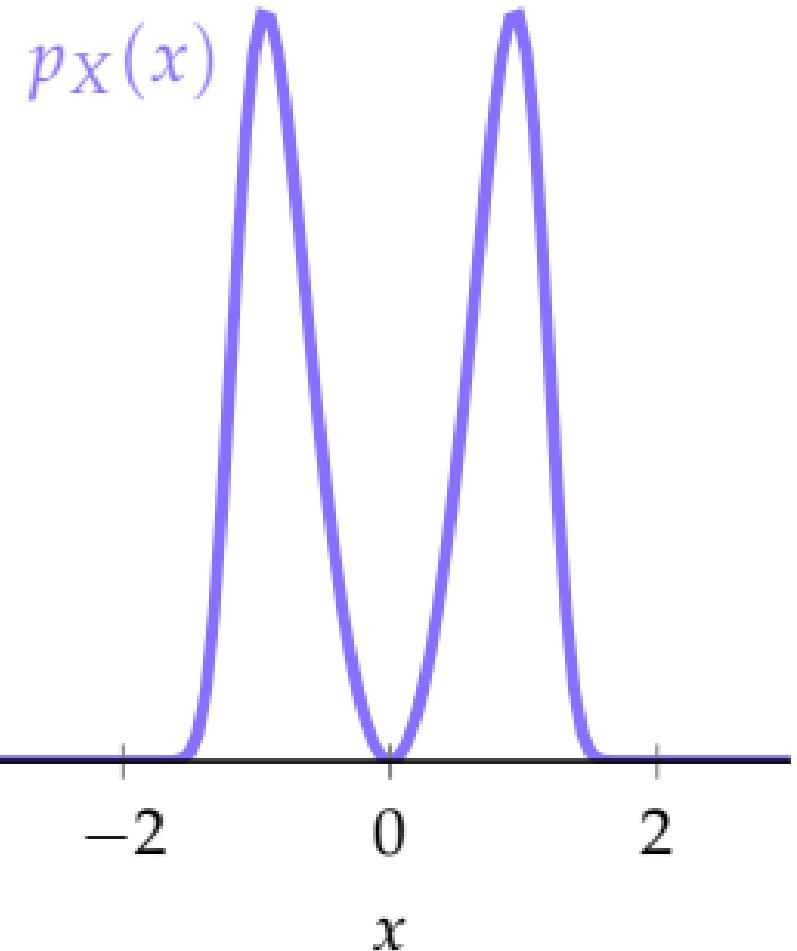


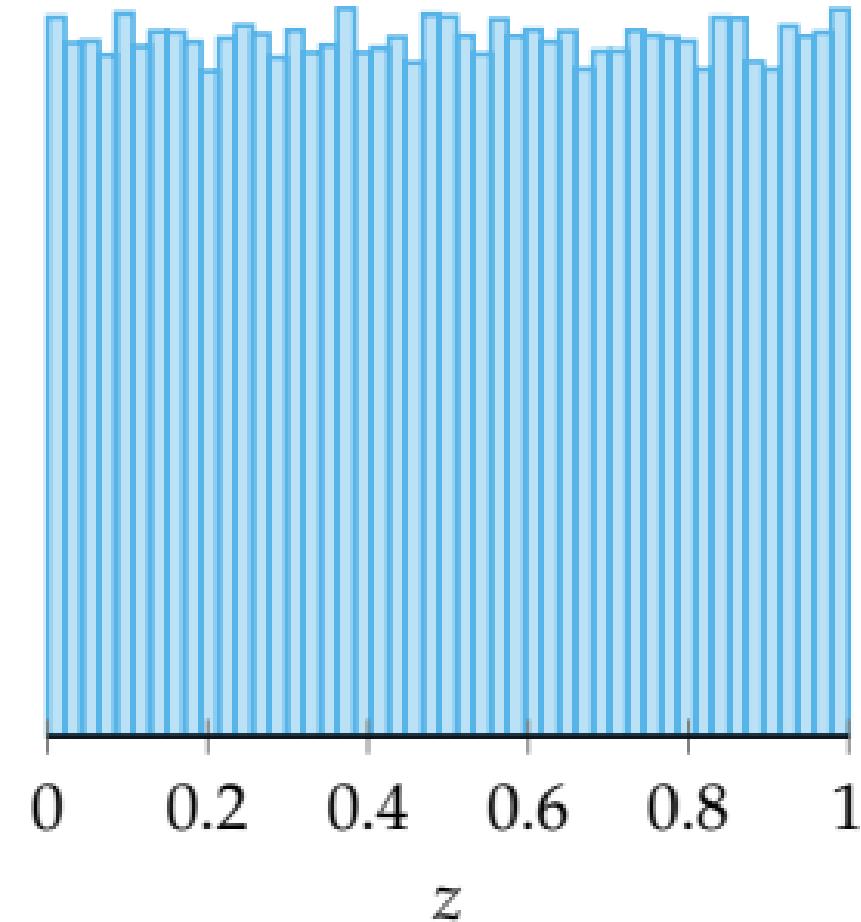
$x$





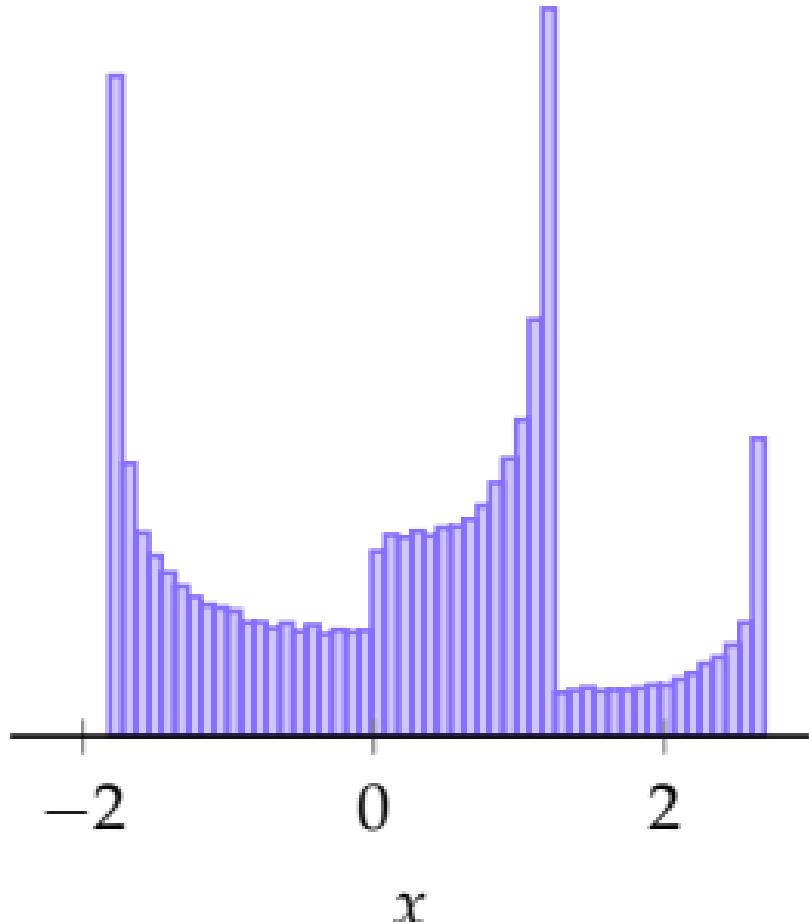
$$\xrightarrow{x = z^{1/3}}$$





$$x = \exp(z) \sin(8z)$$

---



$$\mu = [0, 0]$$

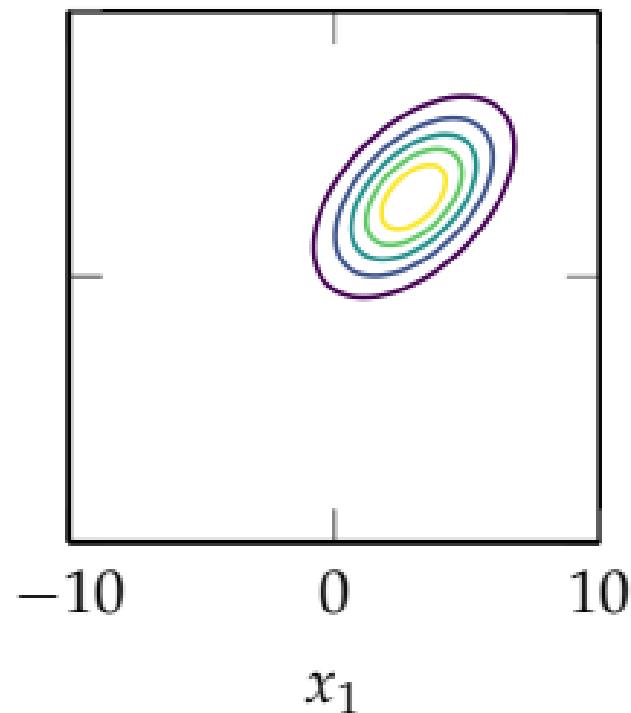
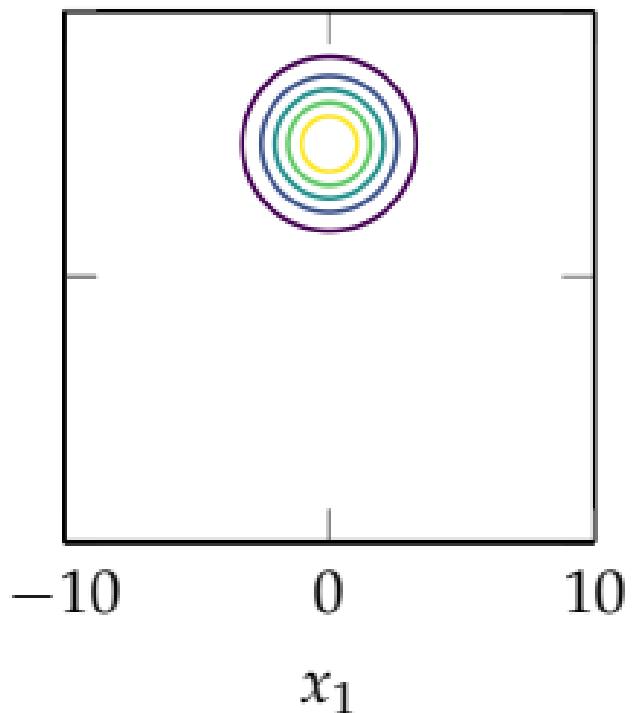
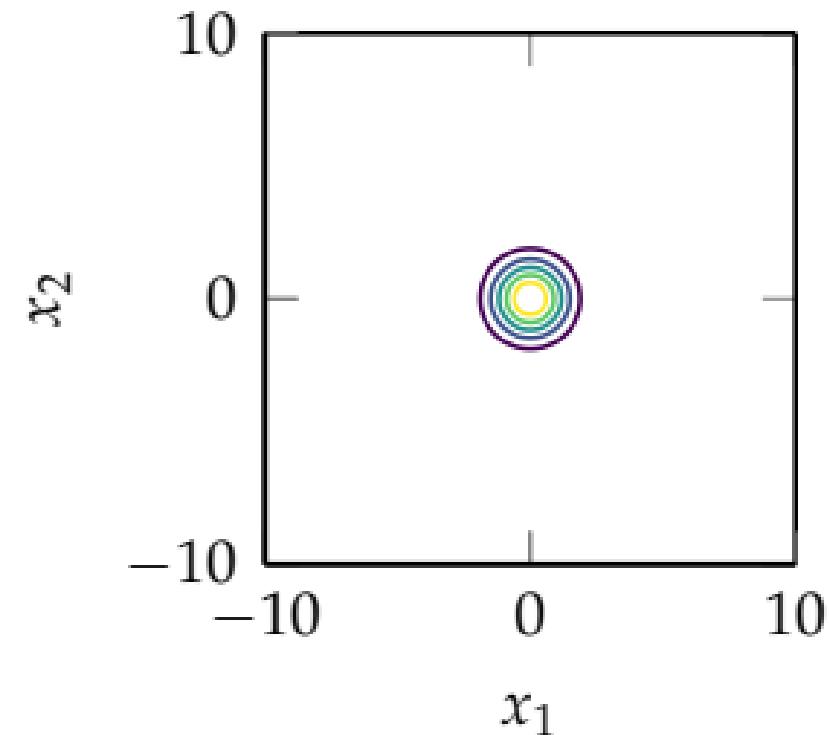
$$\Sigma = [1 \ 0; \ 0 \ 1]$$

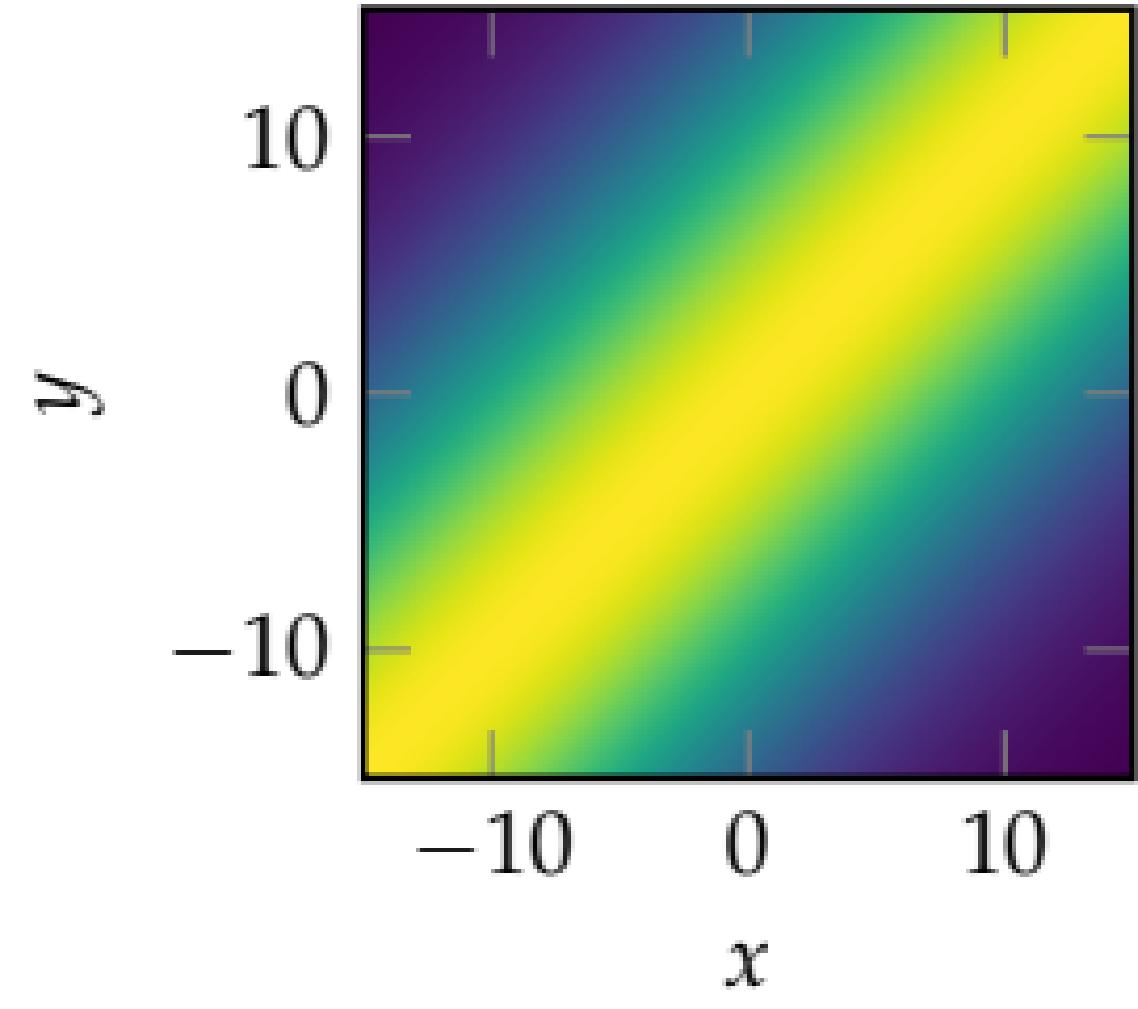
$$\mu = [0, 5]$$

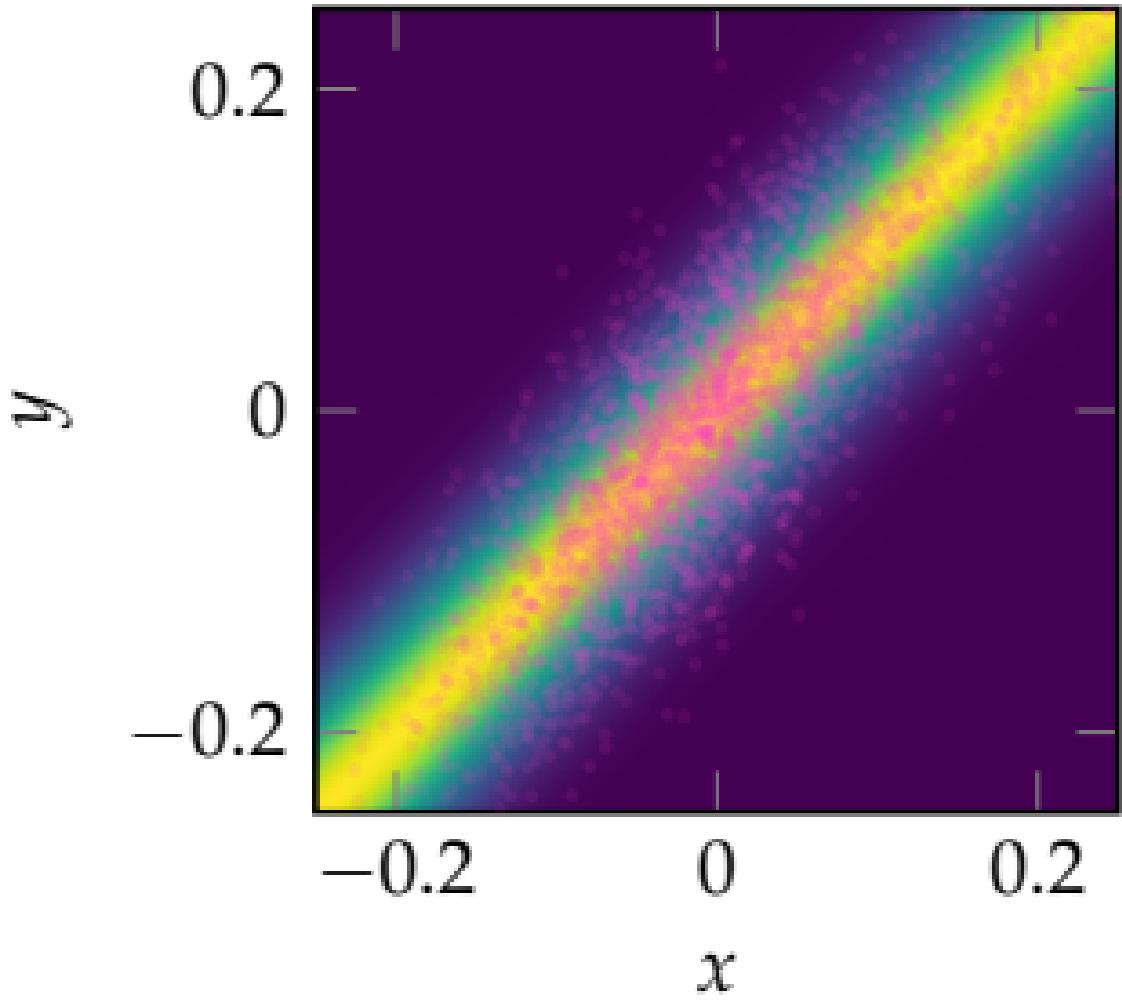
$$\Sigma = [3 \ 0; \ 0 \ 3]$$

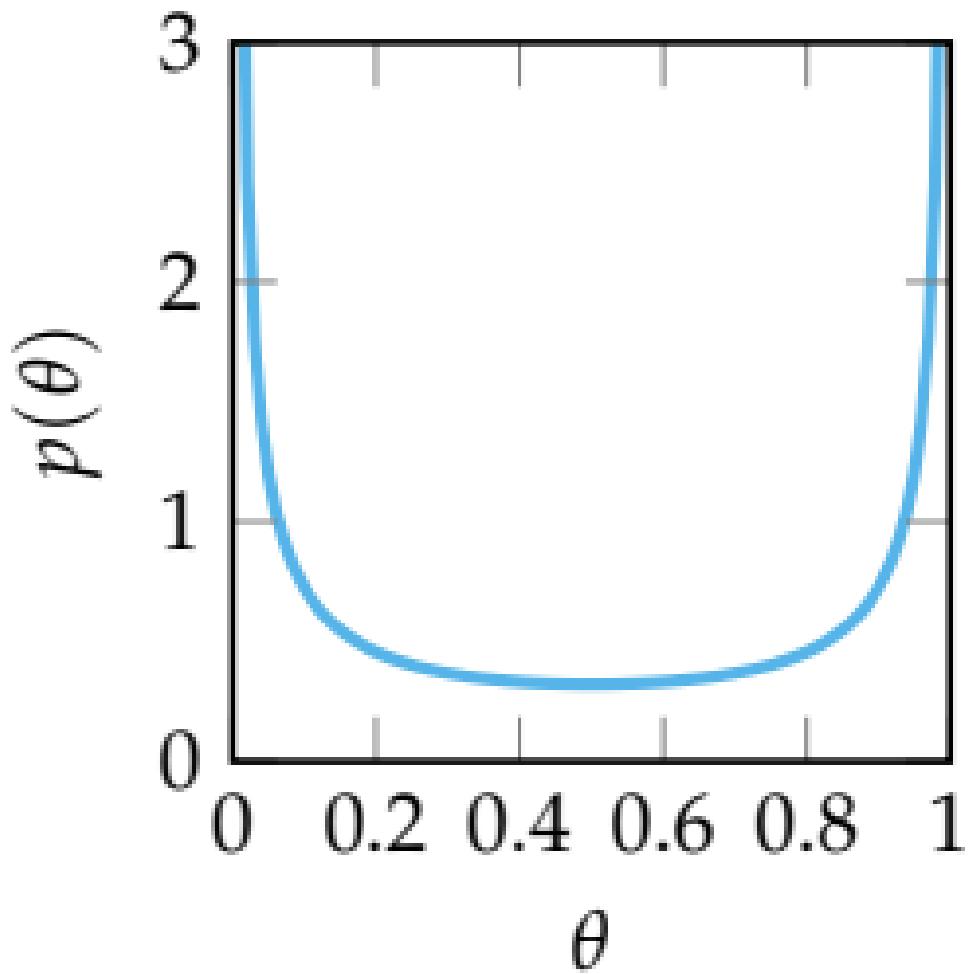
$$\mu = [3, 3]$$

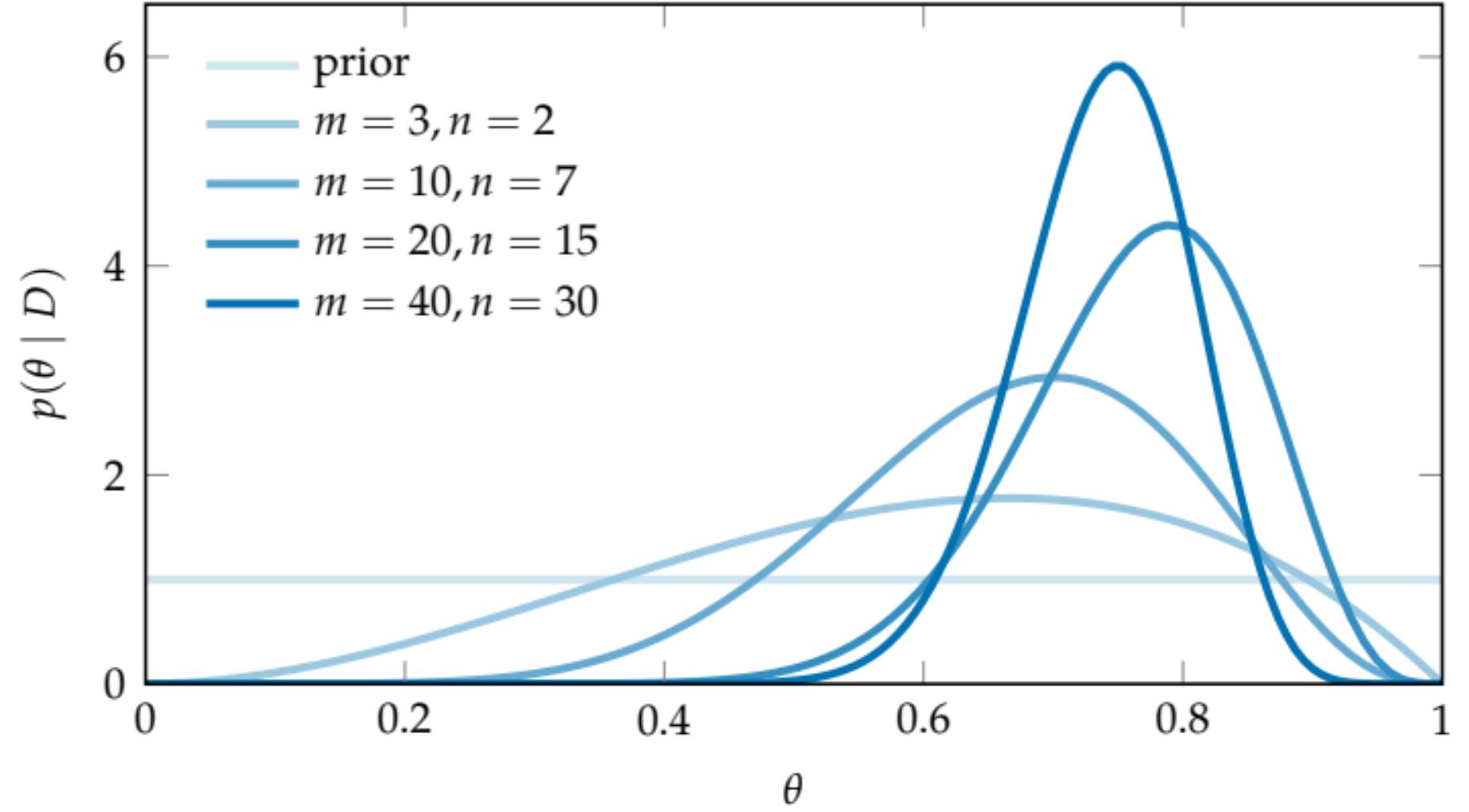
$$\Sigma = [4 \ 2; \ 2 \ 4]$$







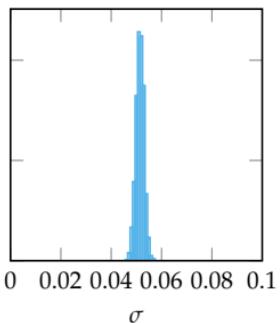
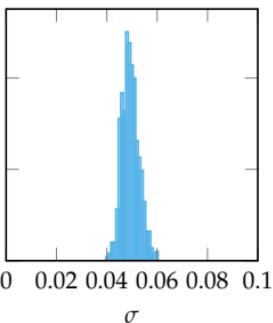
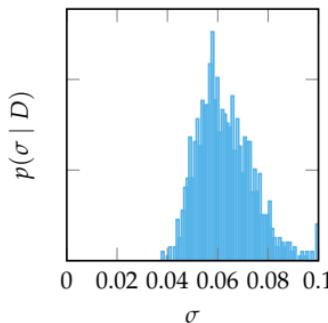
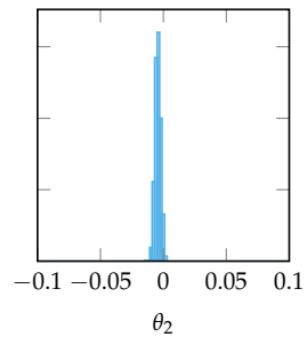
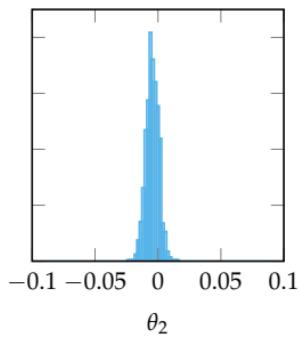
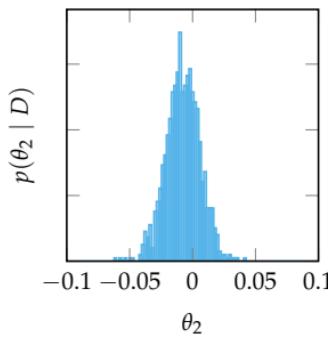
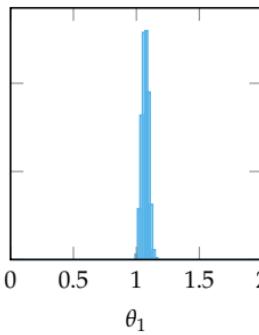
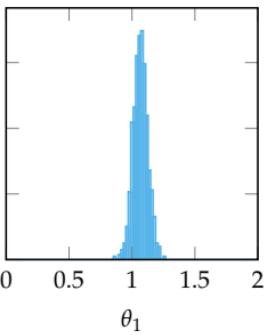
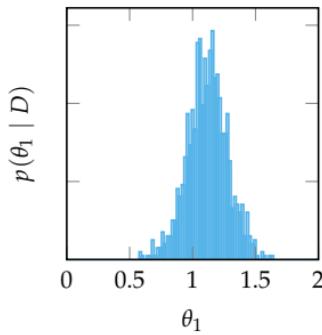
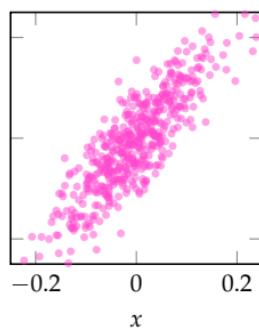
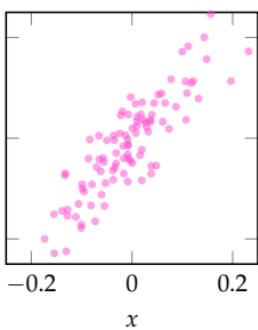
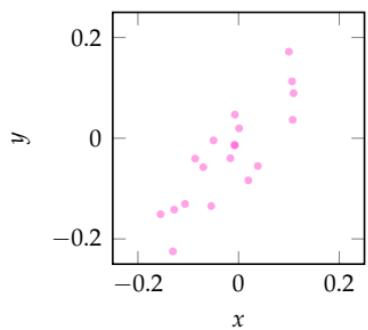


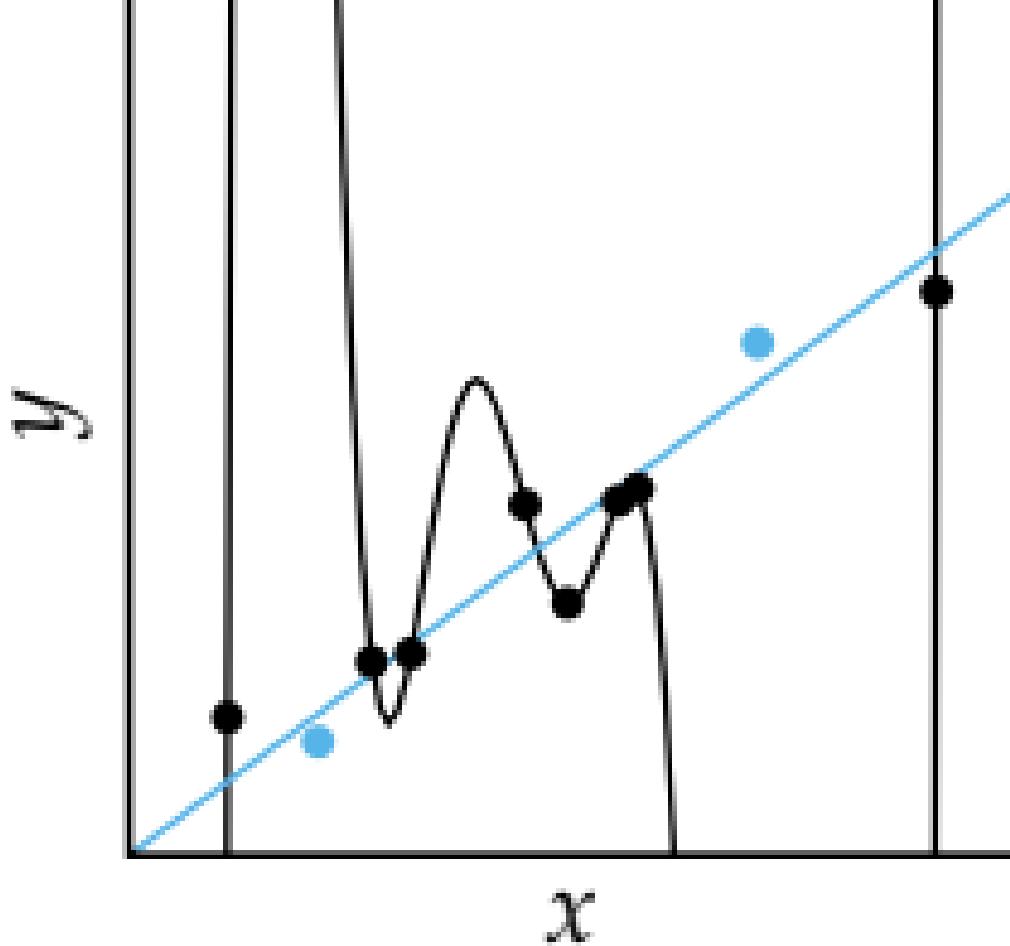


20 samples

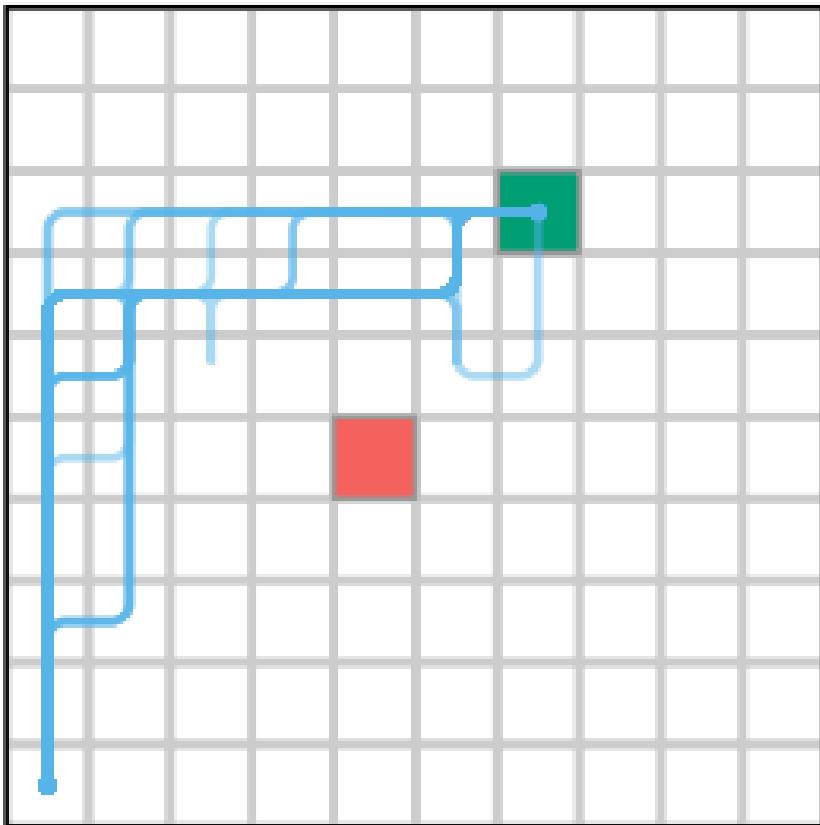
100 samples

500 samples

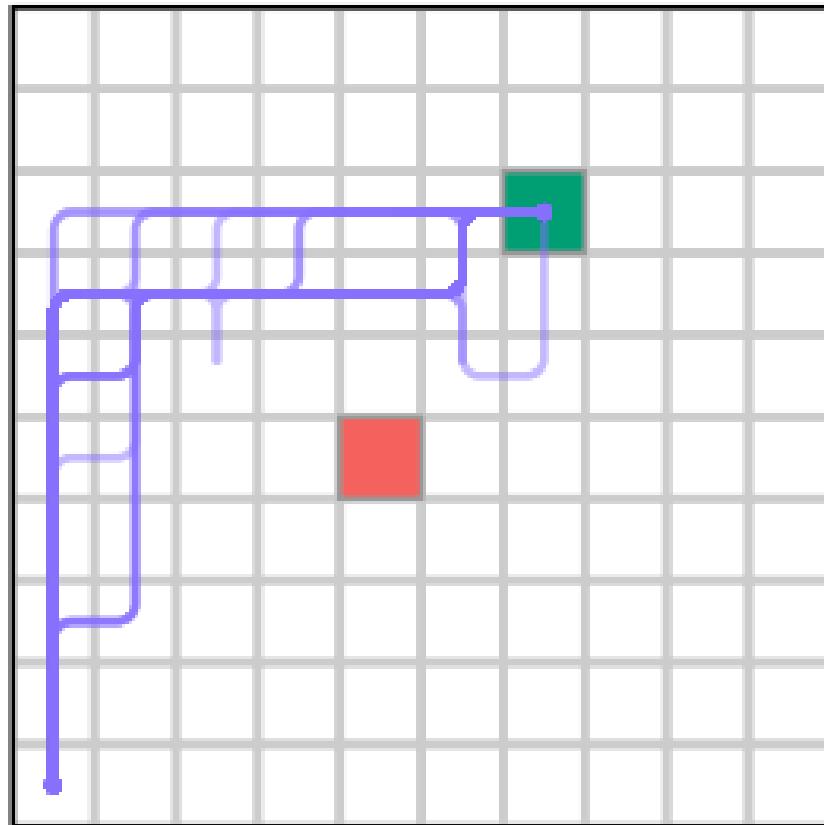


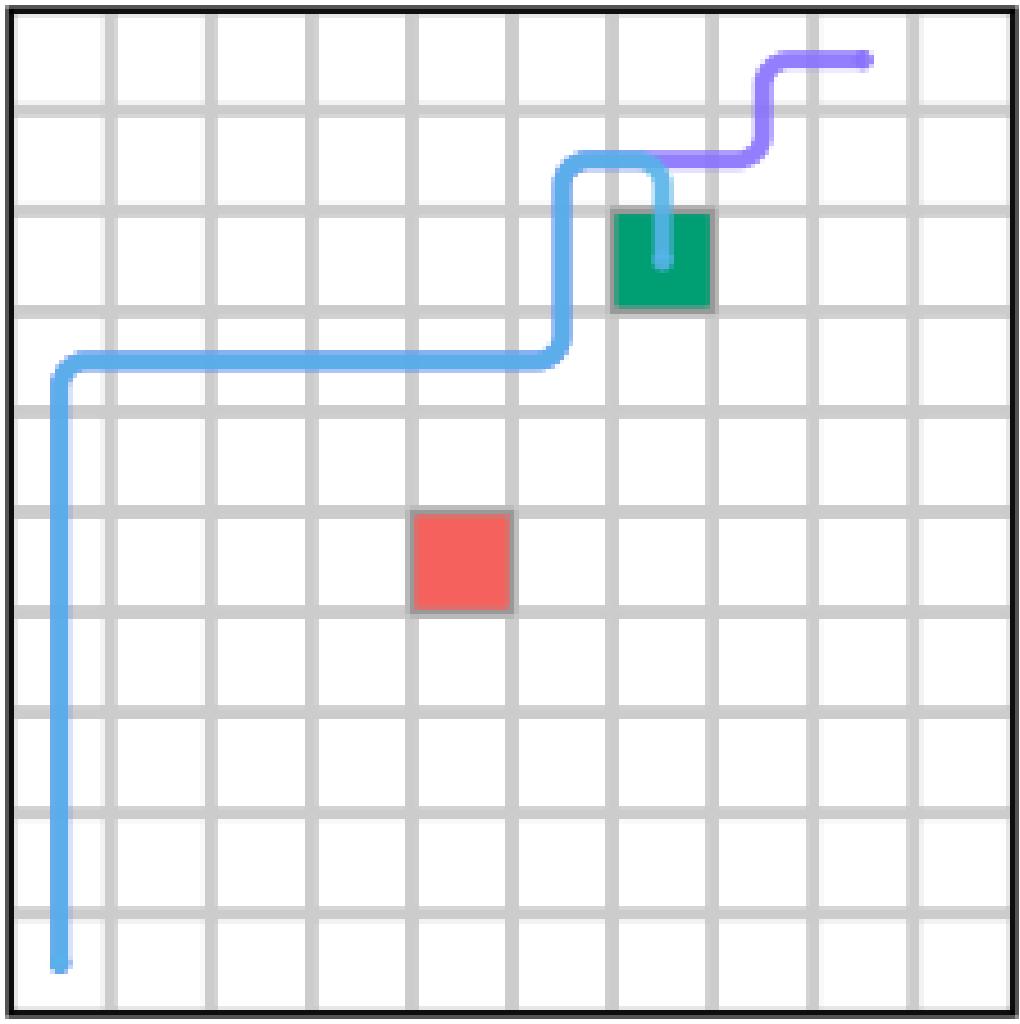


# Original Agent



# Cloned Agent



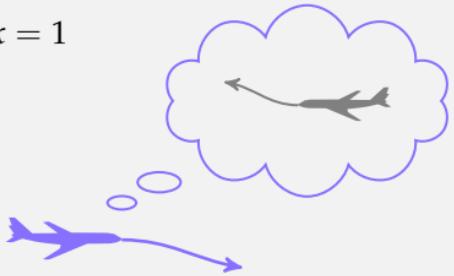




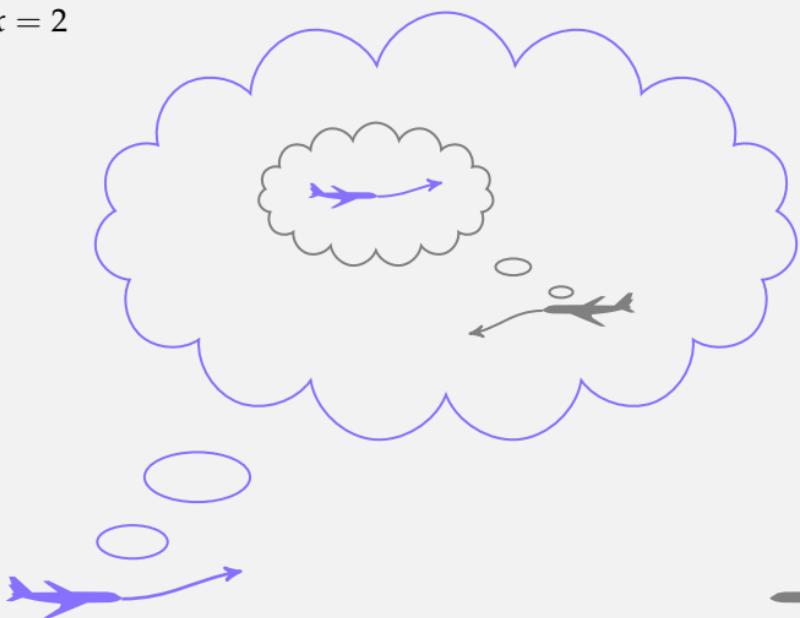
$k = 0$



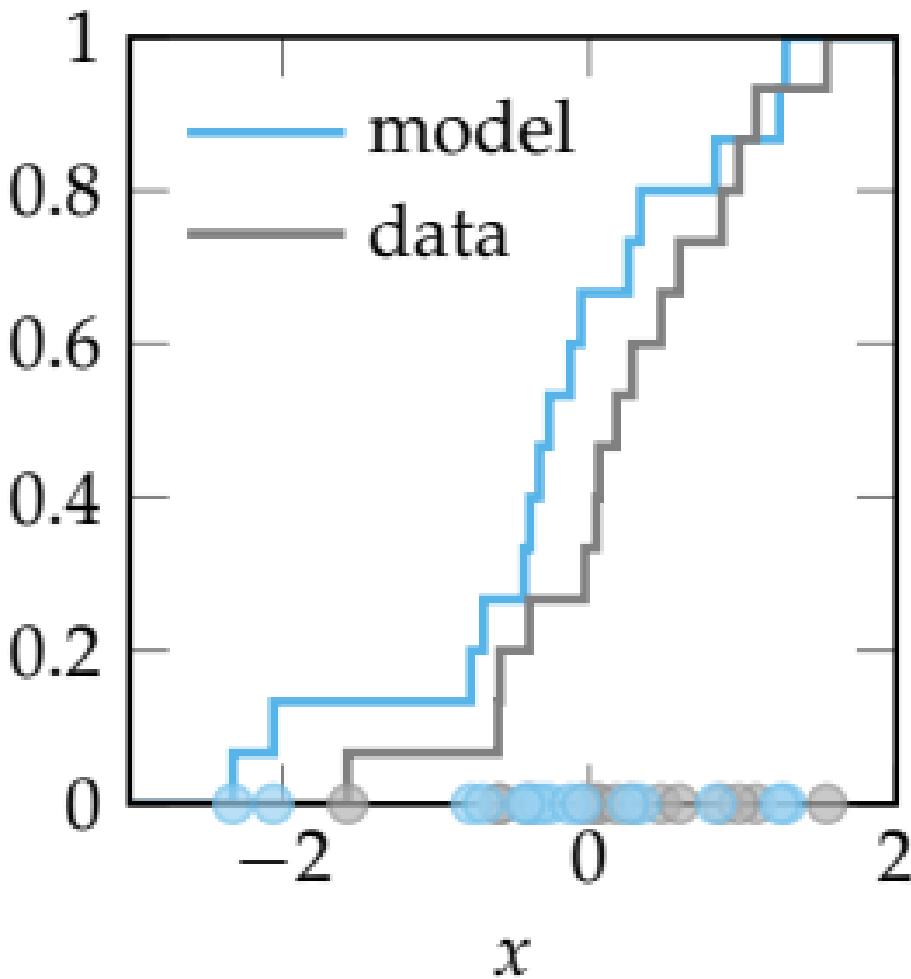
$k = 1$

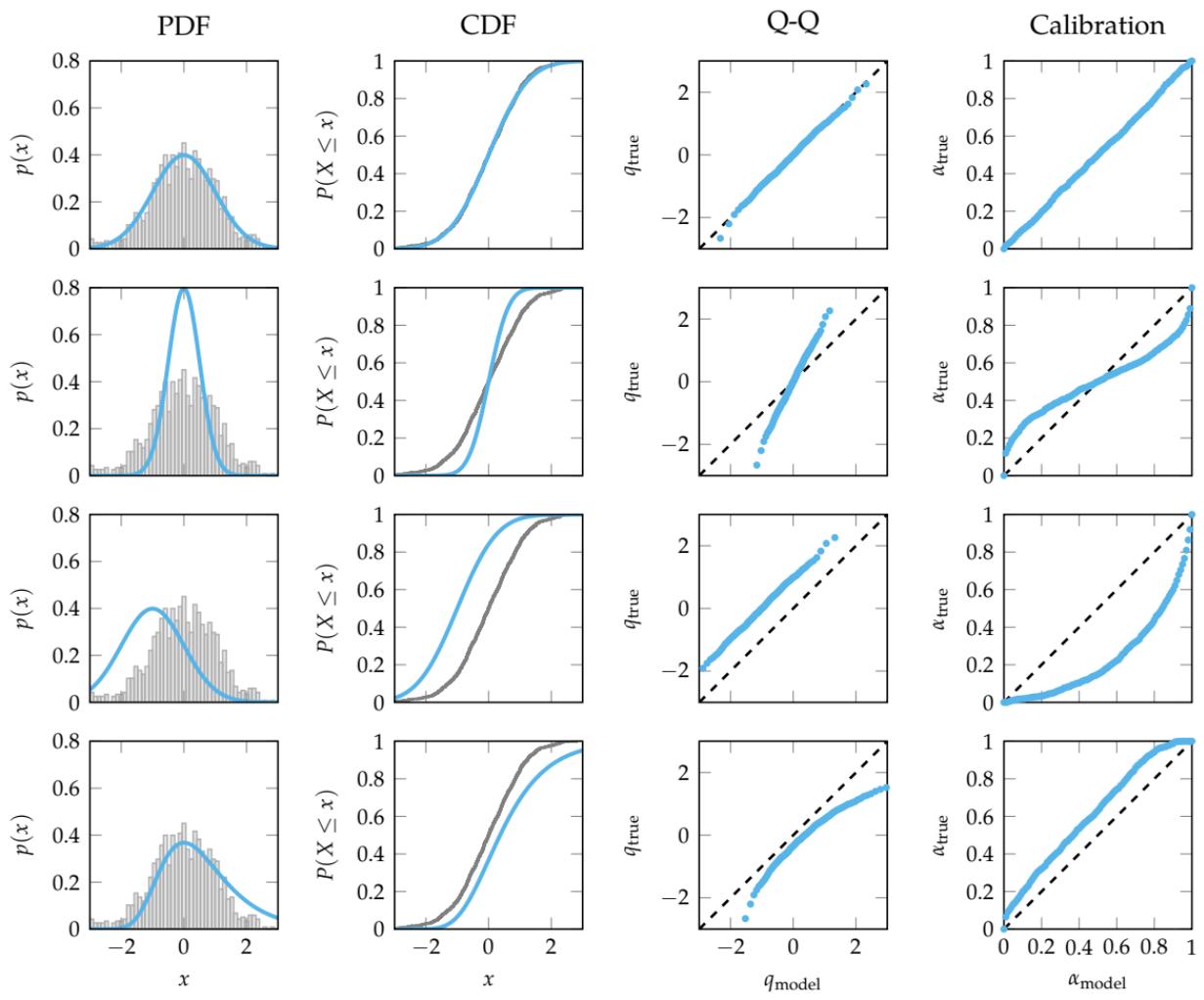


$k = 2$



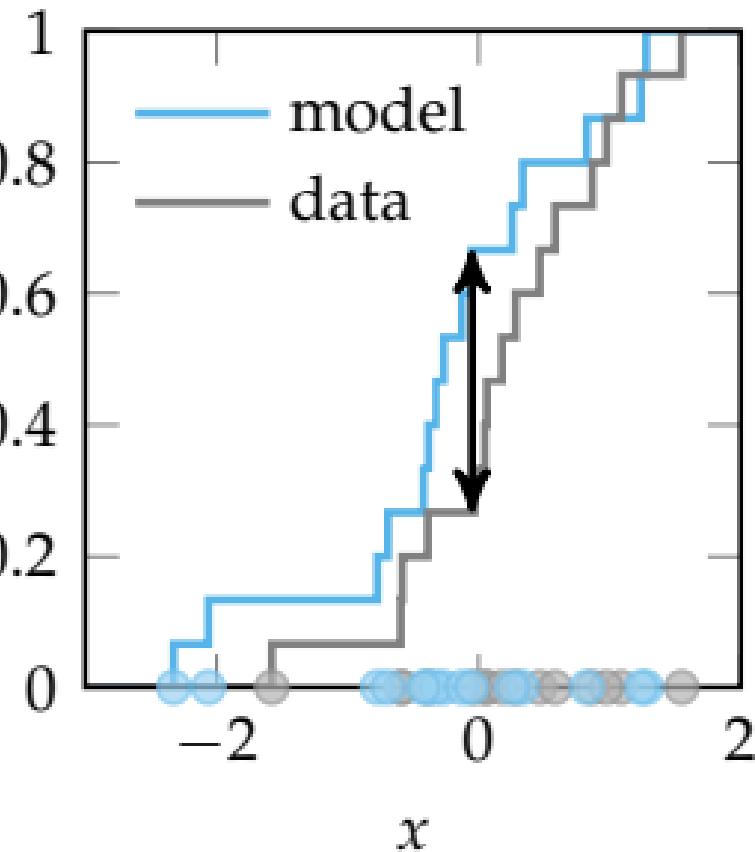
Empirical CDF



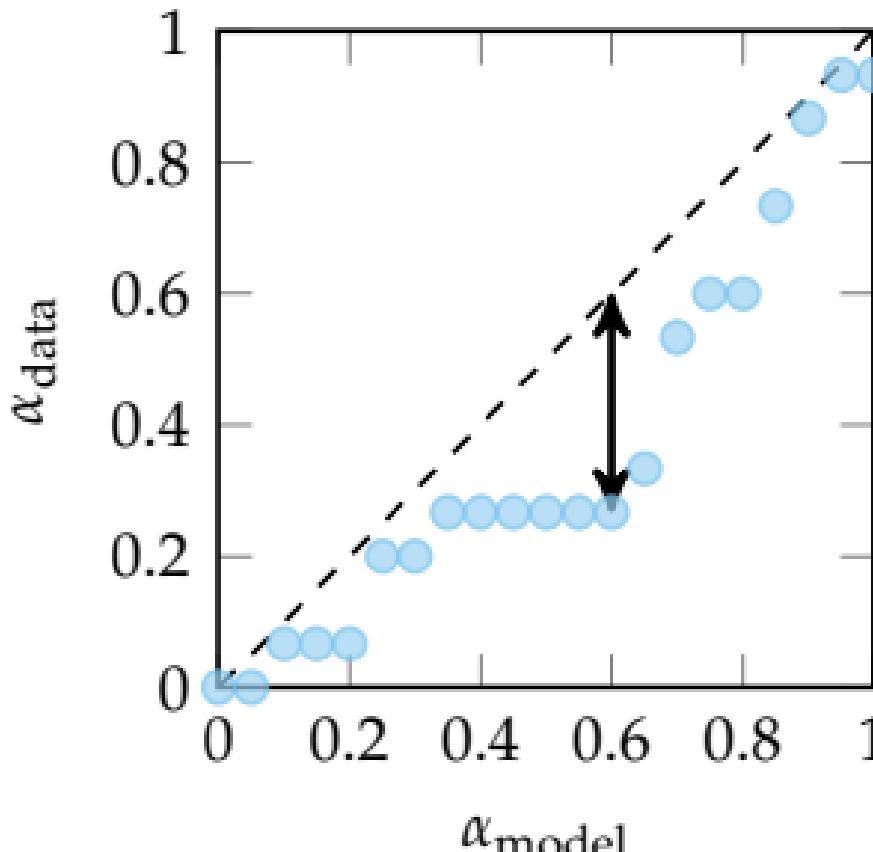


# K-S Statistic

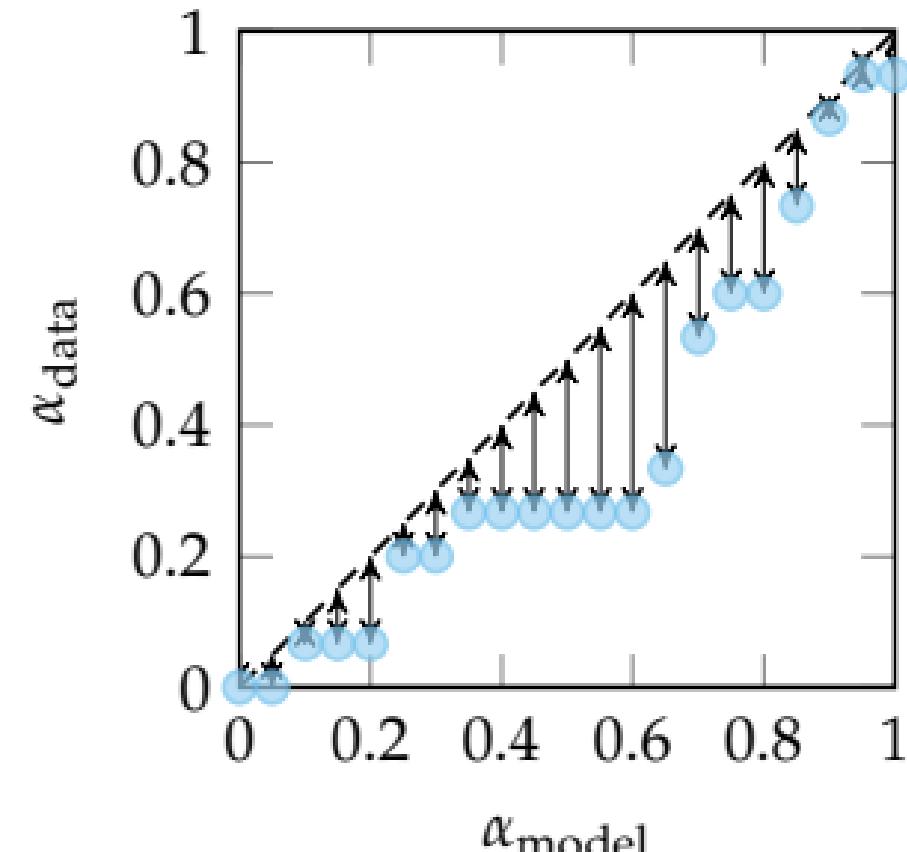
Empirical CDF

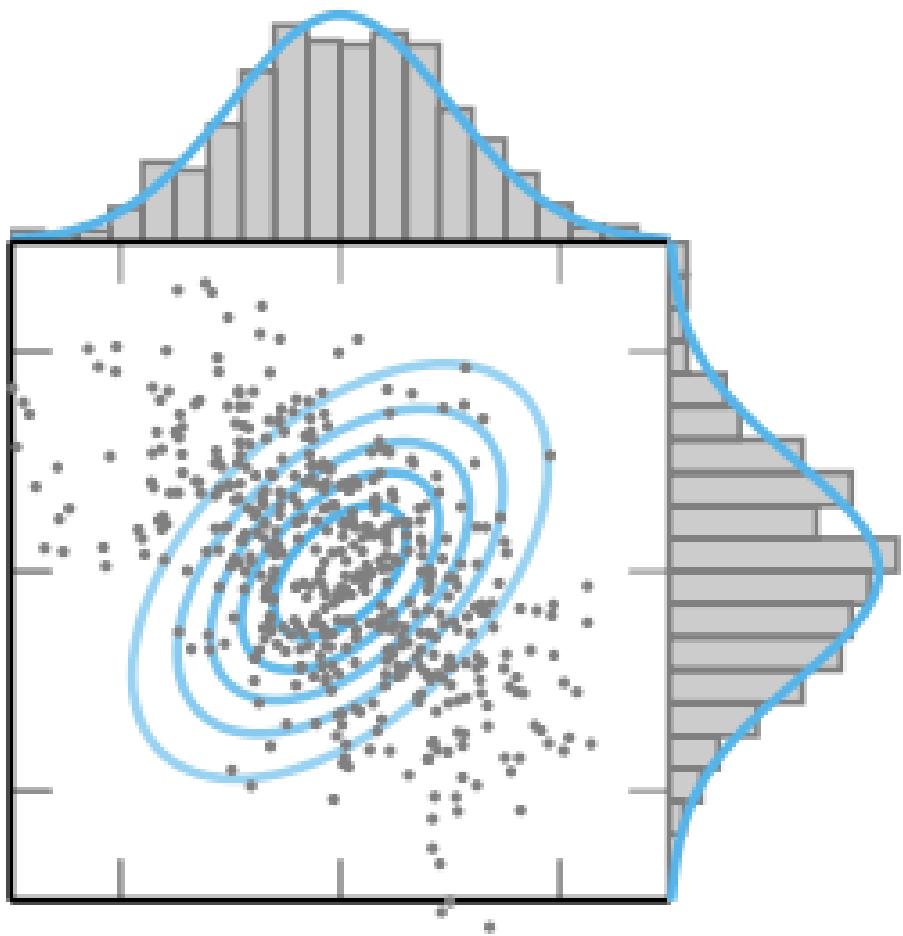


# MCE

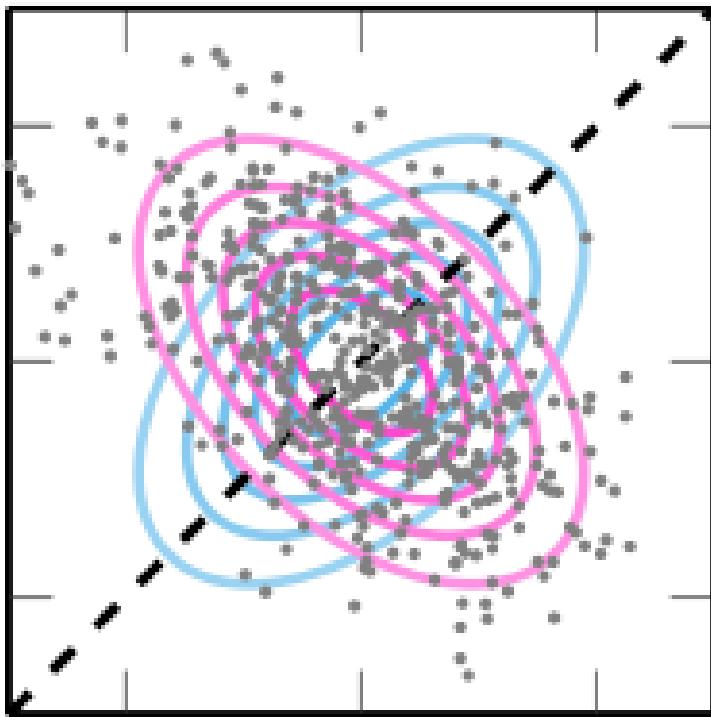


# ECE

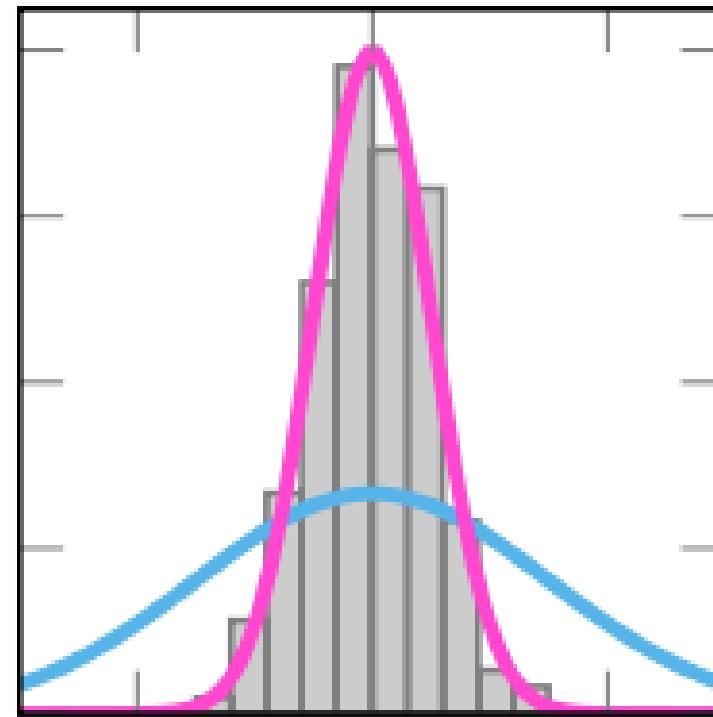




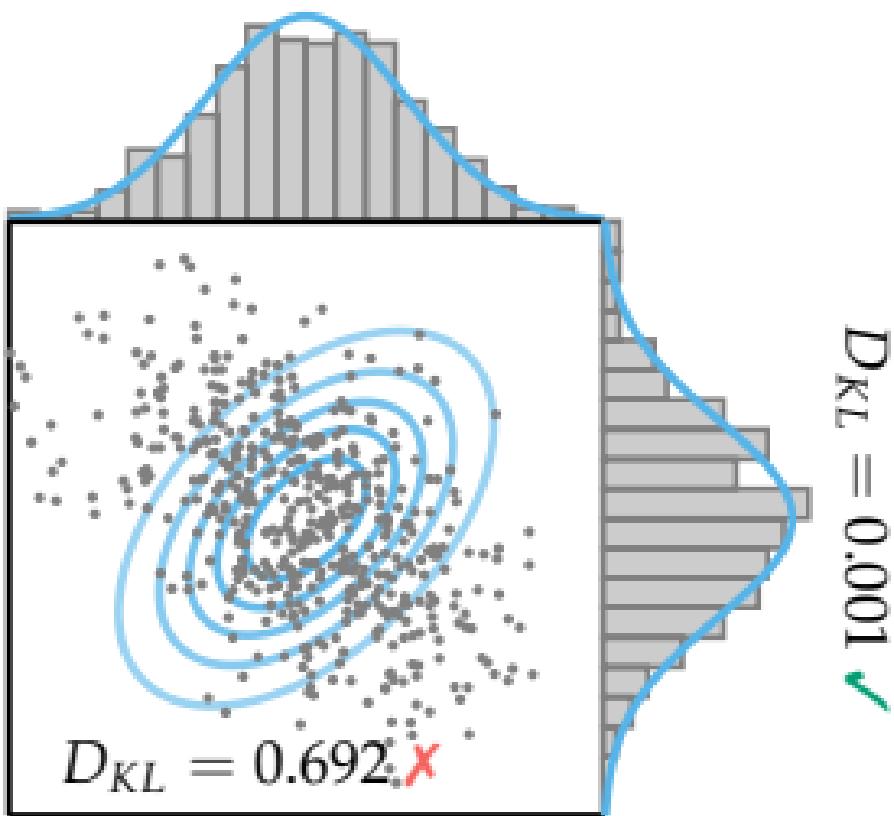
# Original Data



# Projected Distribution



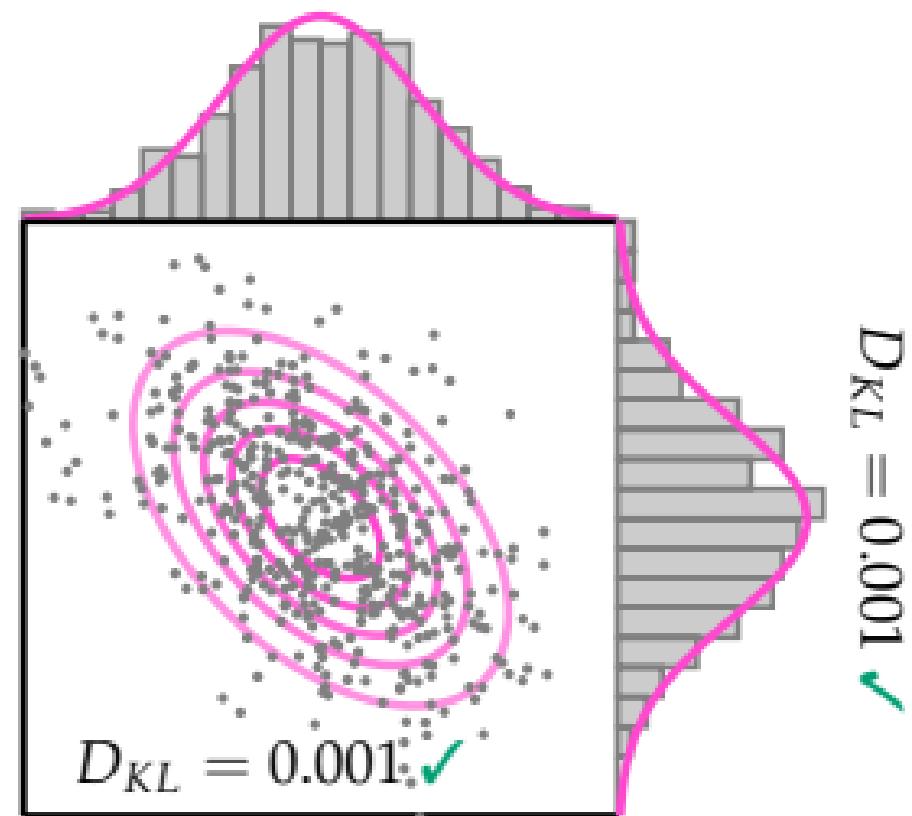
$D_{KL} = 0.001$  ✓



$D_{KL} = 0.692$  ✗

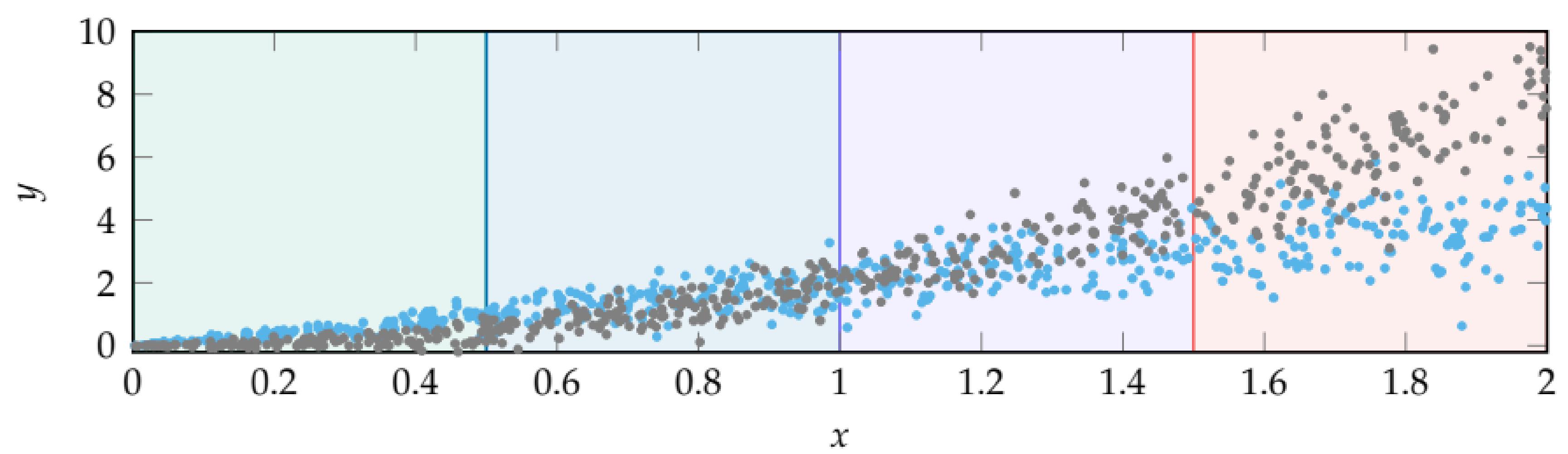
$D_{KL} = 0.001$  ✓

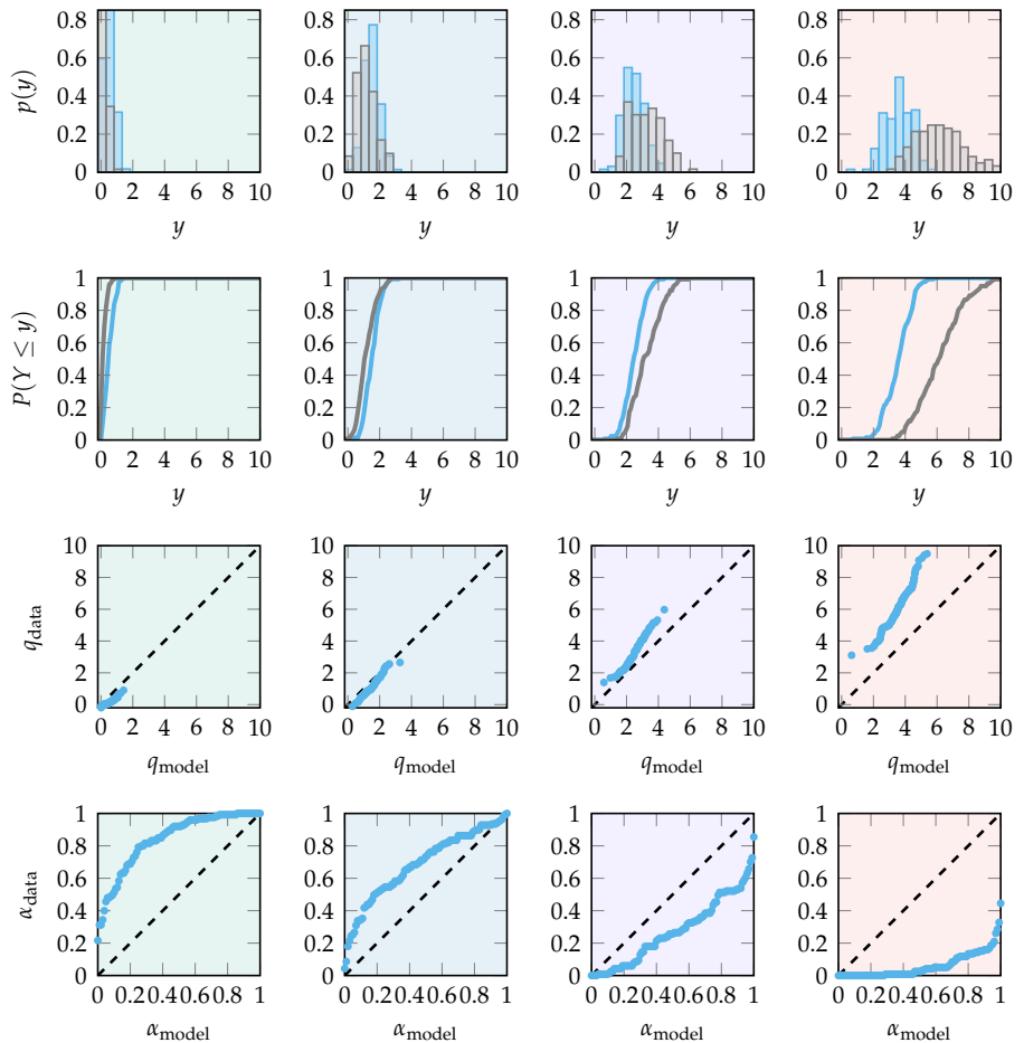
$D_{KL} = 0.001$  ✓



$D_{KL} = 0.001$  ✓

$D_{KL} = 0.001$  ✓





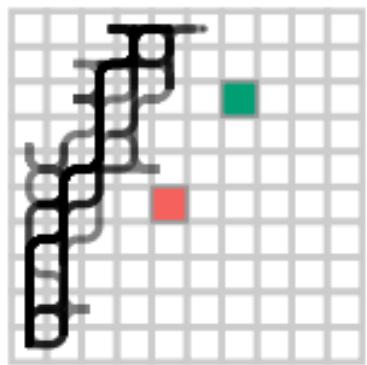
# Poor Model Representation ✗



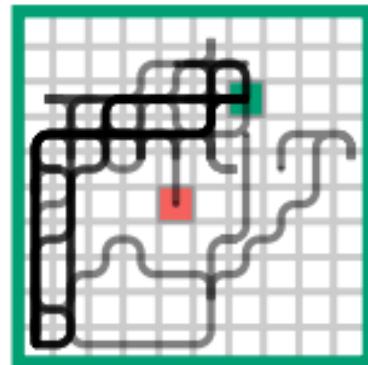
Pair 1



Pair 2



Pair 3

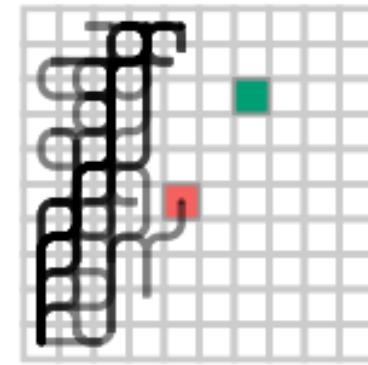
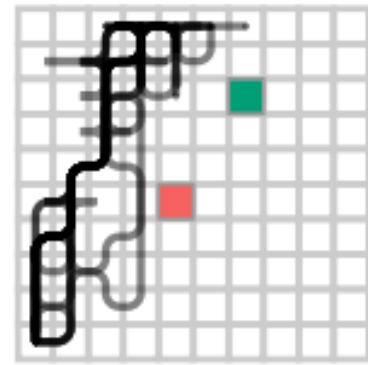
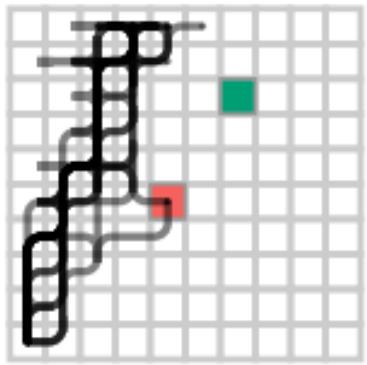


Pair 4



Accuracy: 100%

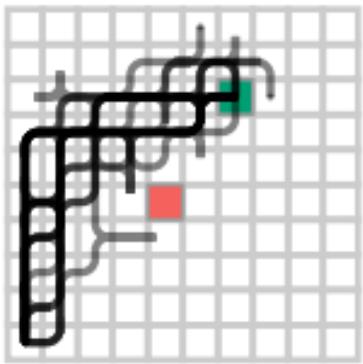
Test Failed ✗



# Good Model Representation ✓



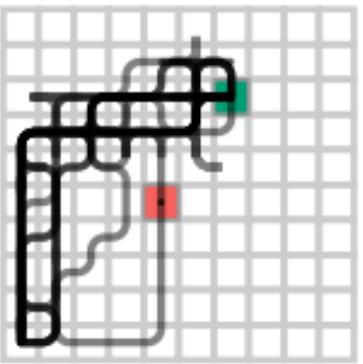
Pair 1



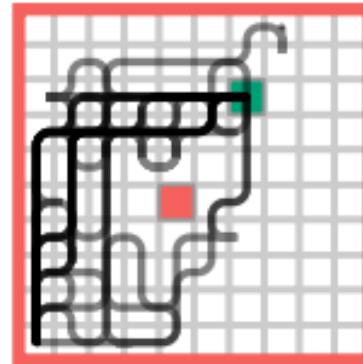
Pair 2



Pair 3

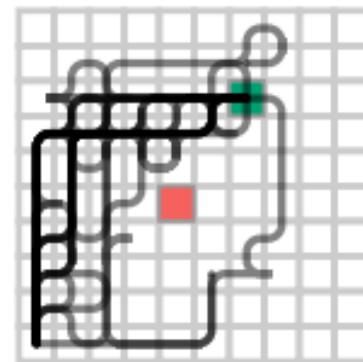
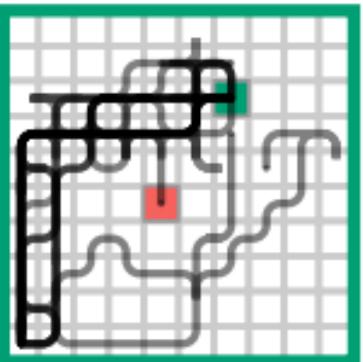
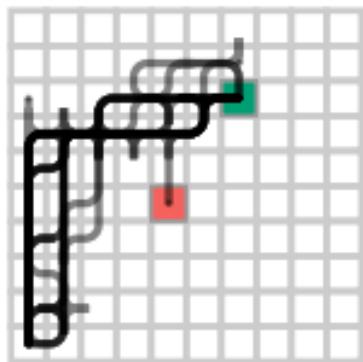


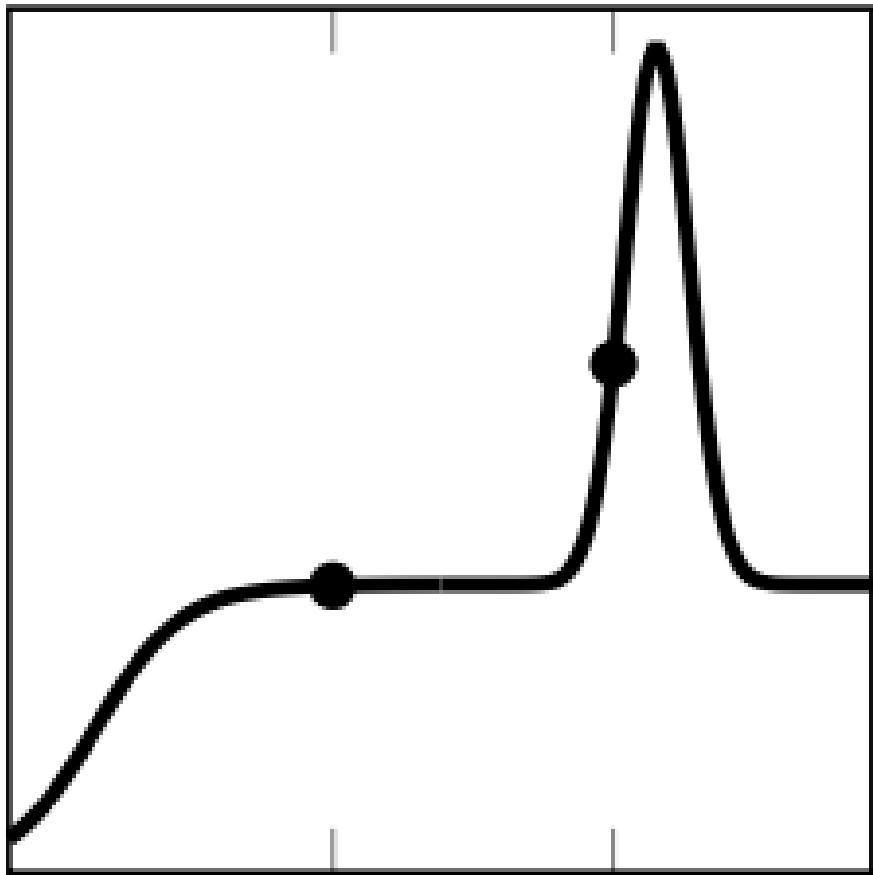
Pair 4

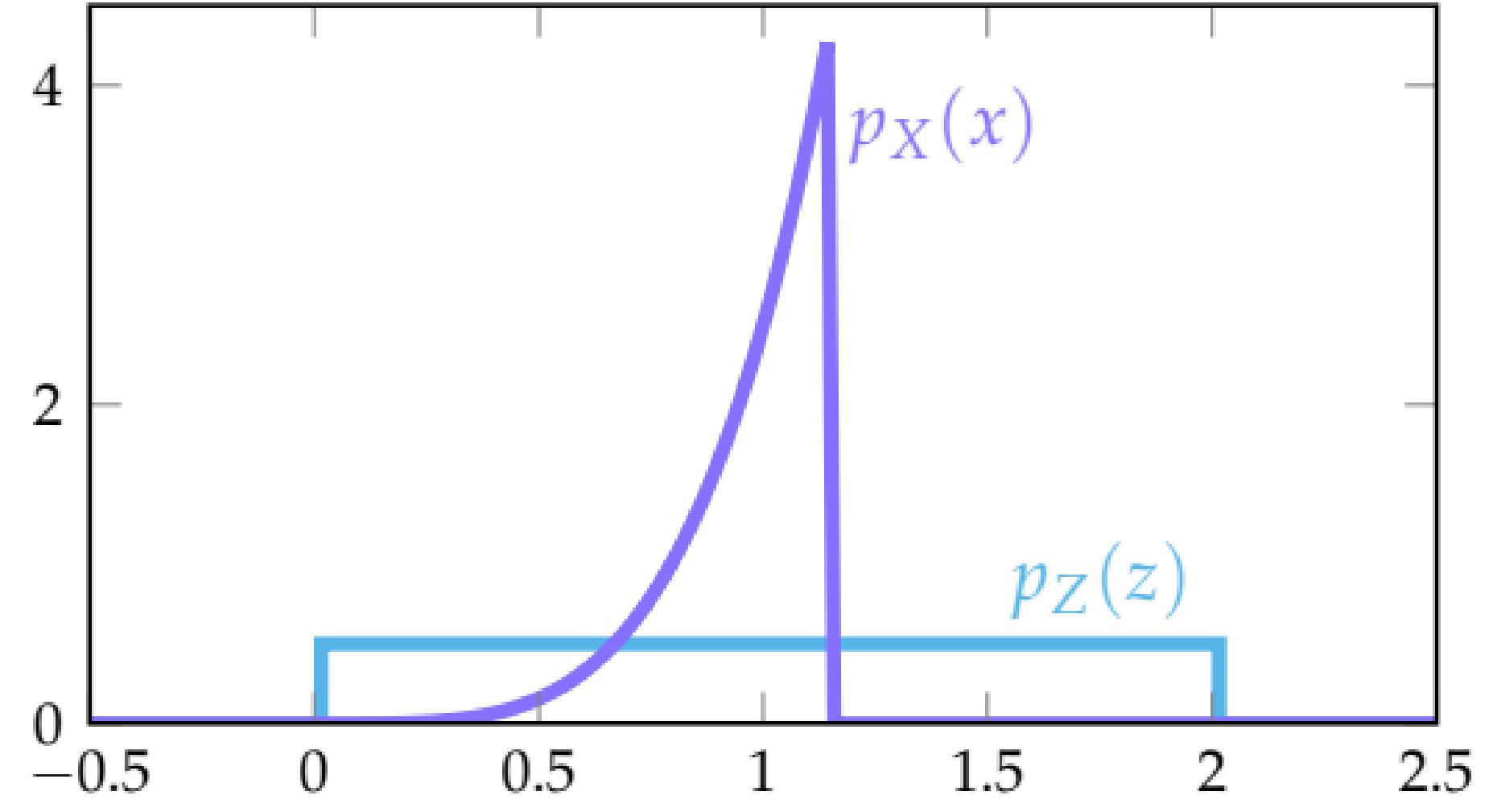


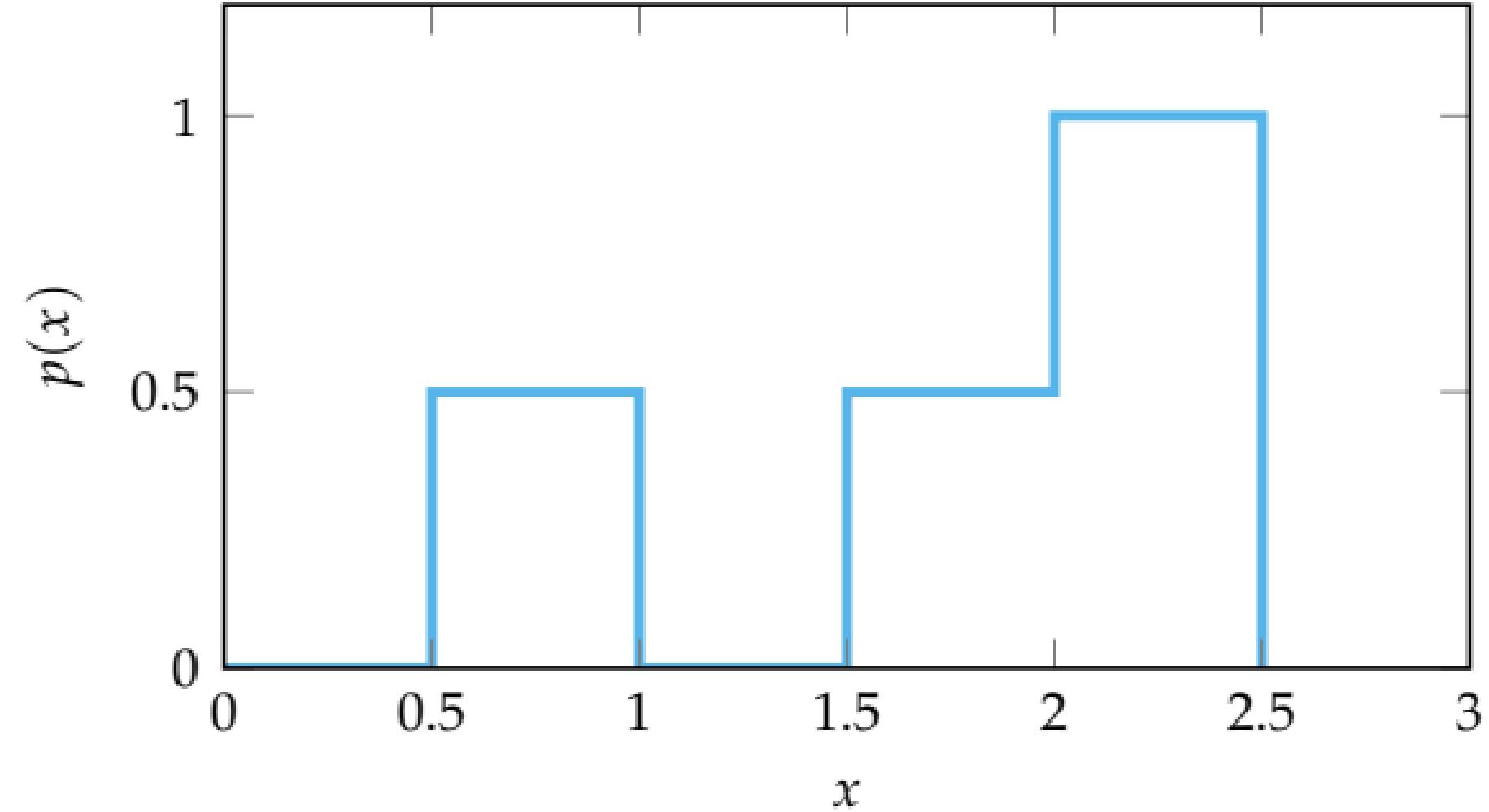
Accuracy: 50%

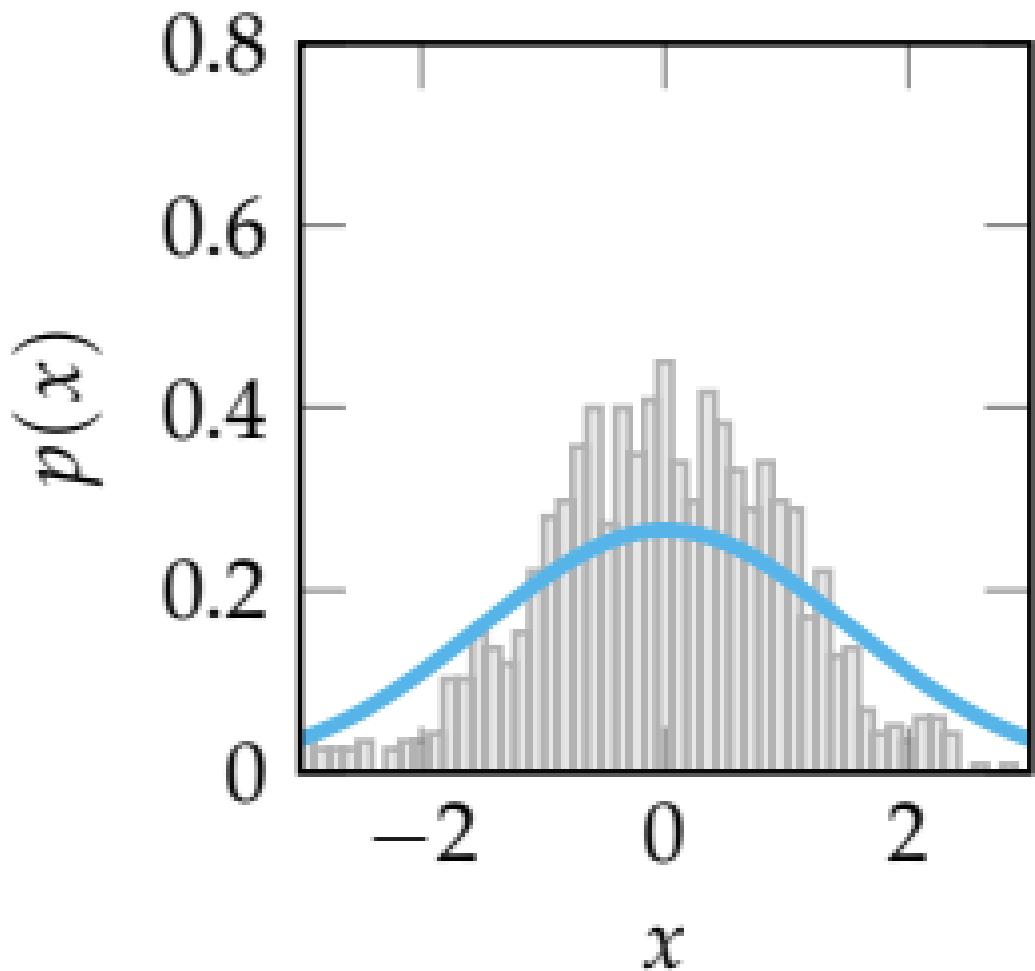
Test Passed ✓

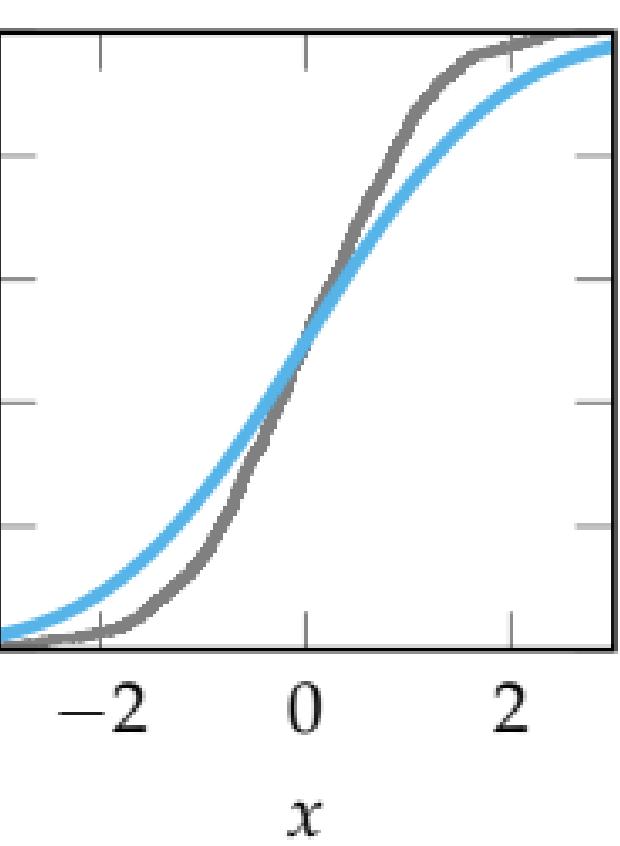
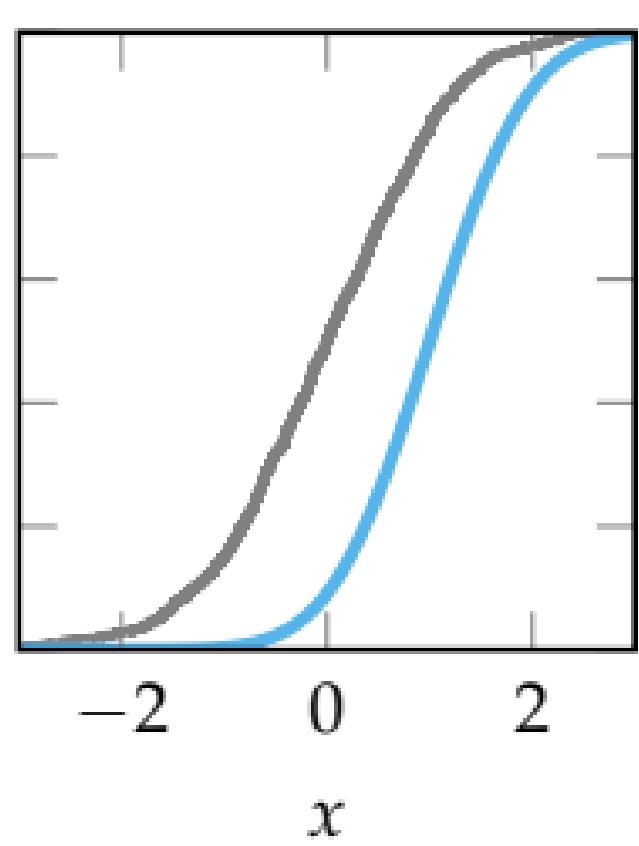
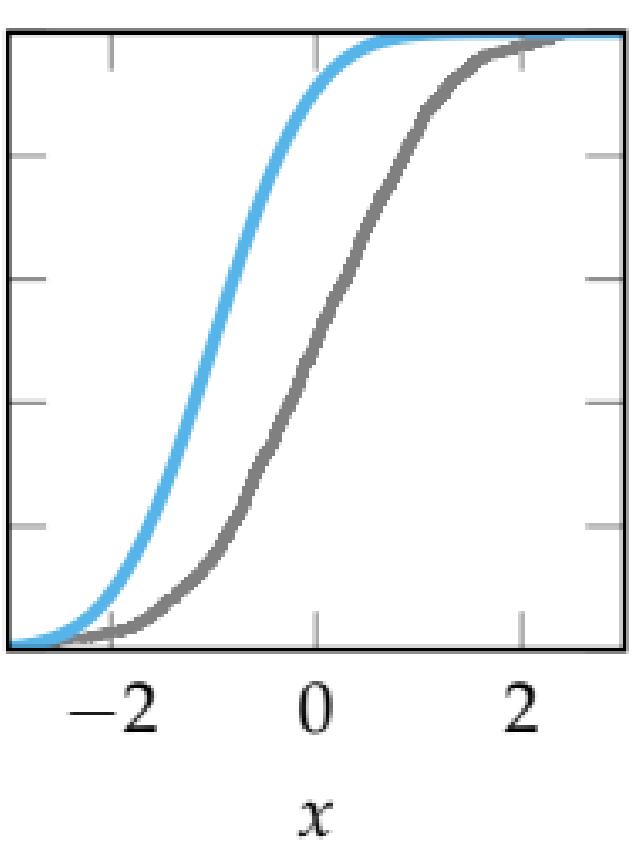
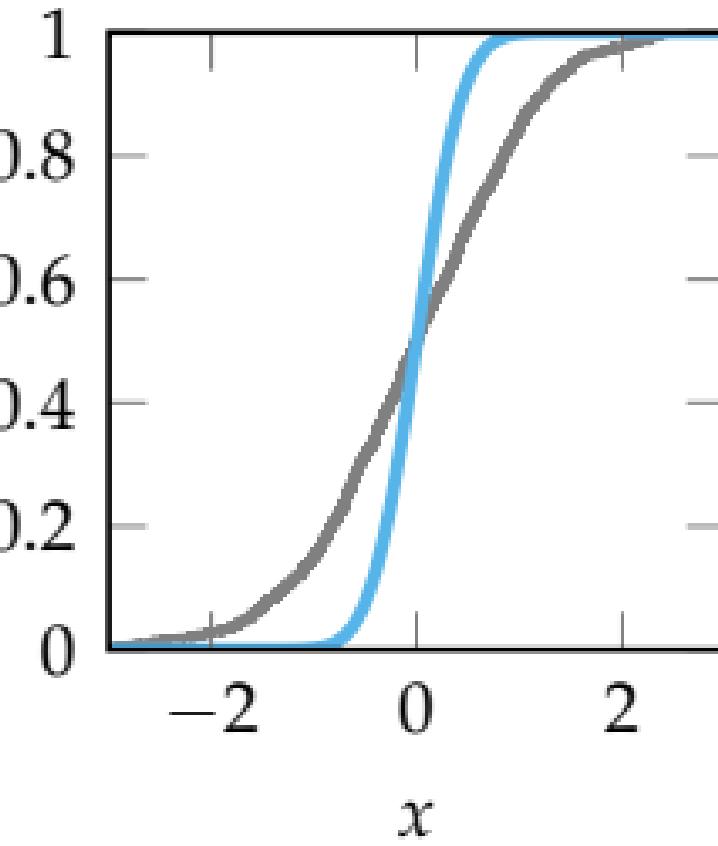


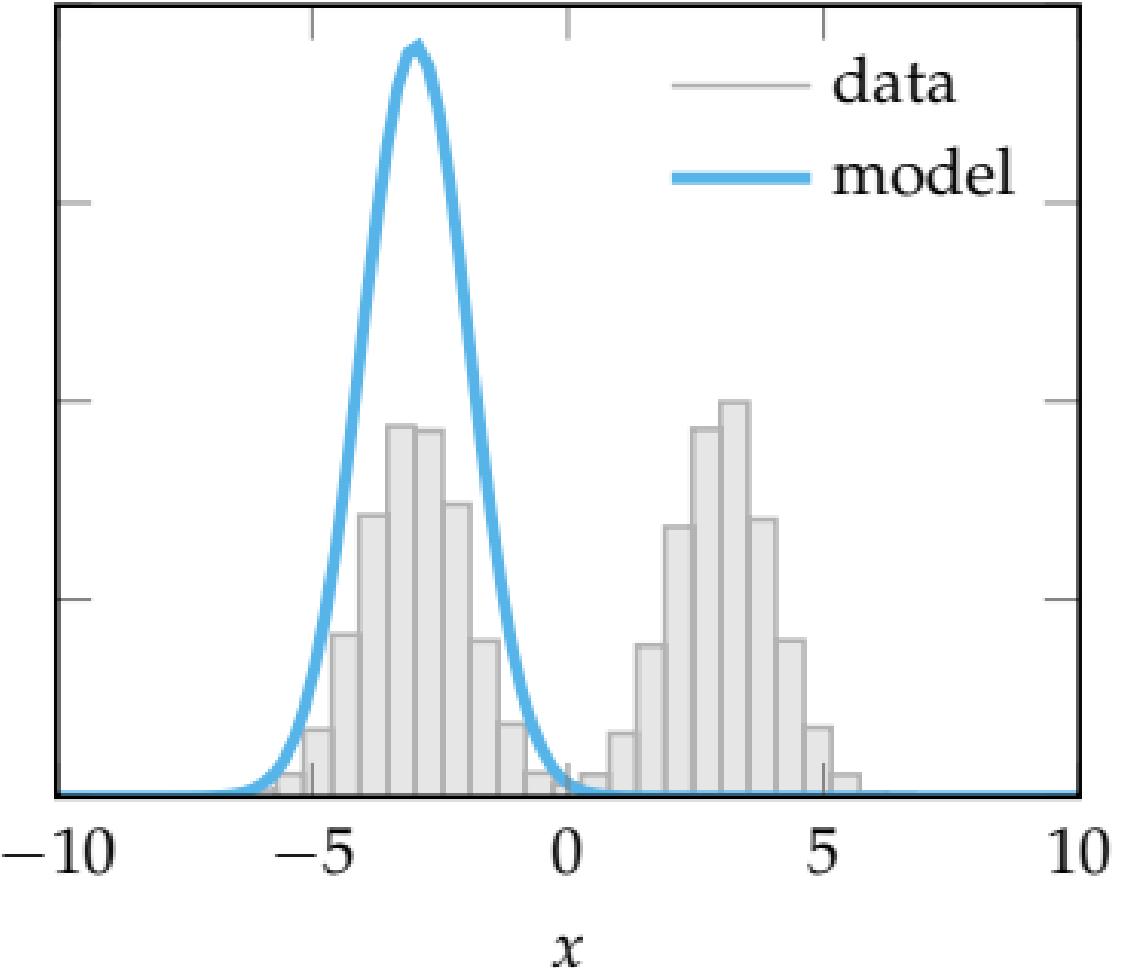
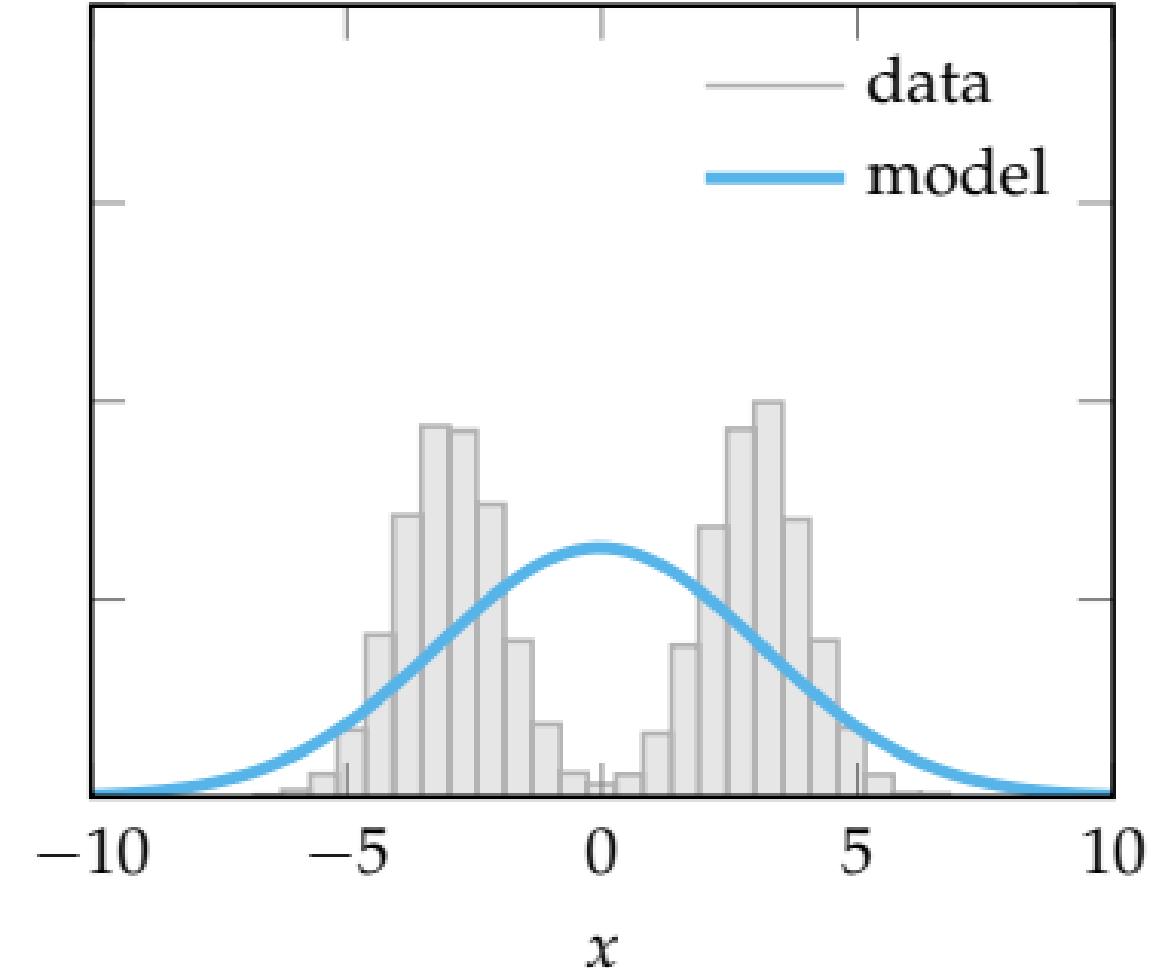
$f(\theta)$  $\theta_1$  $\theta_2$  $\theta$

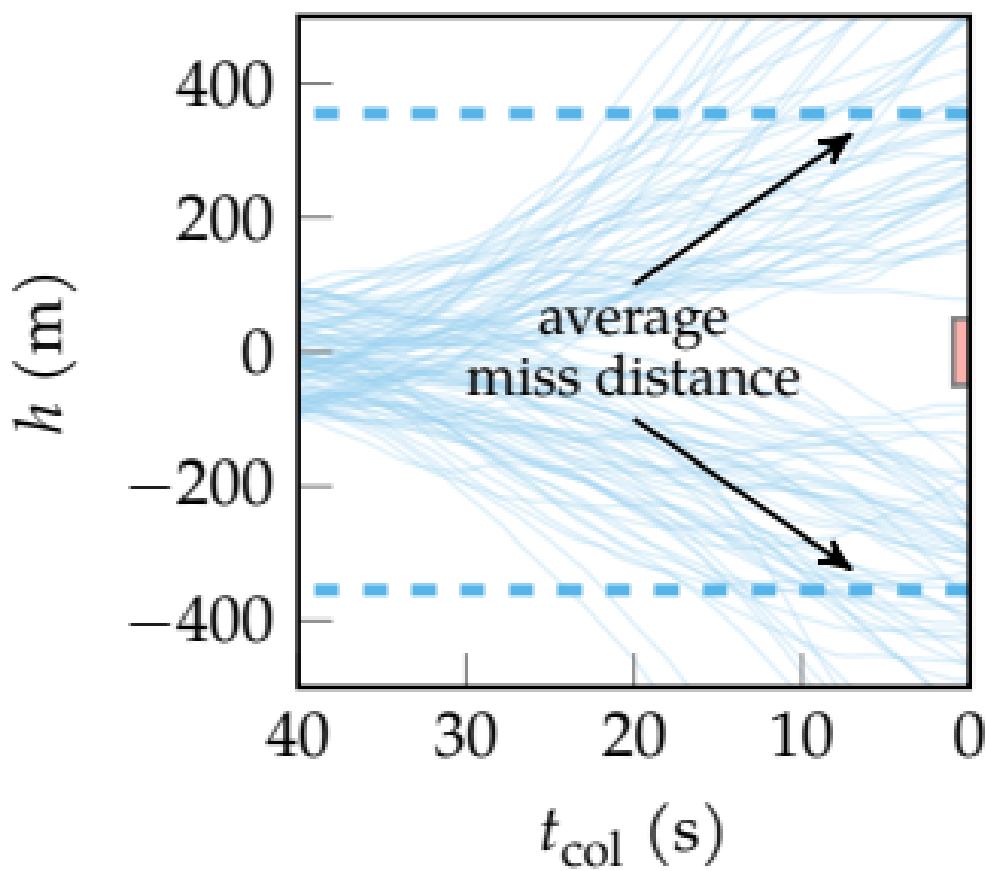
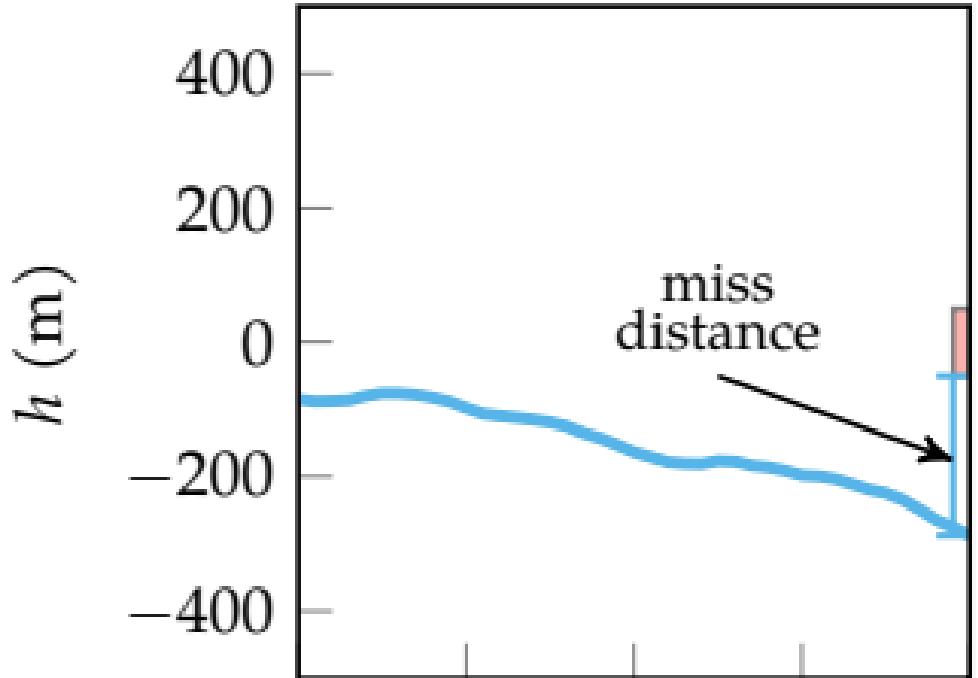


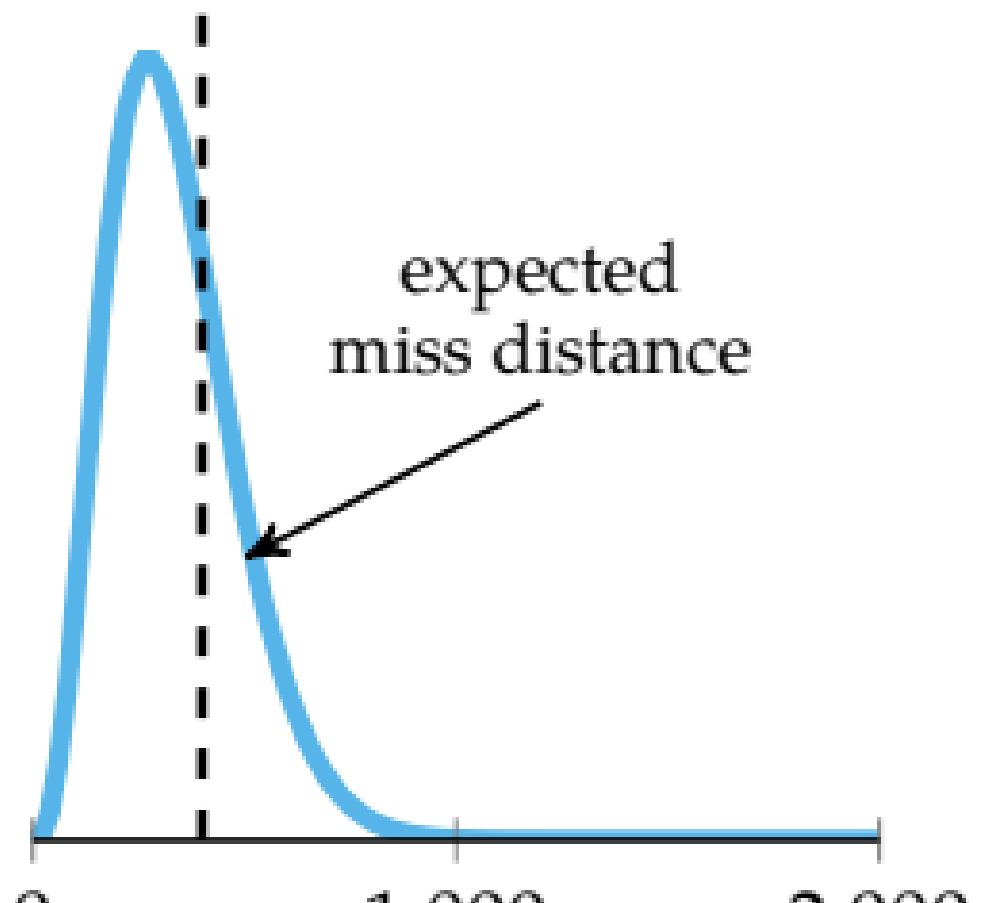




$P(X \leq x)$ 





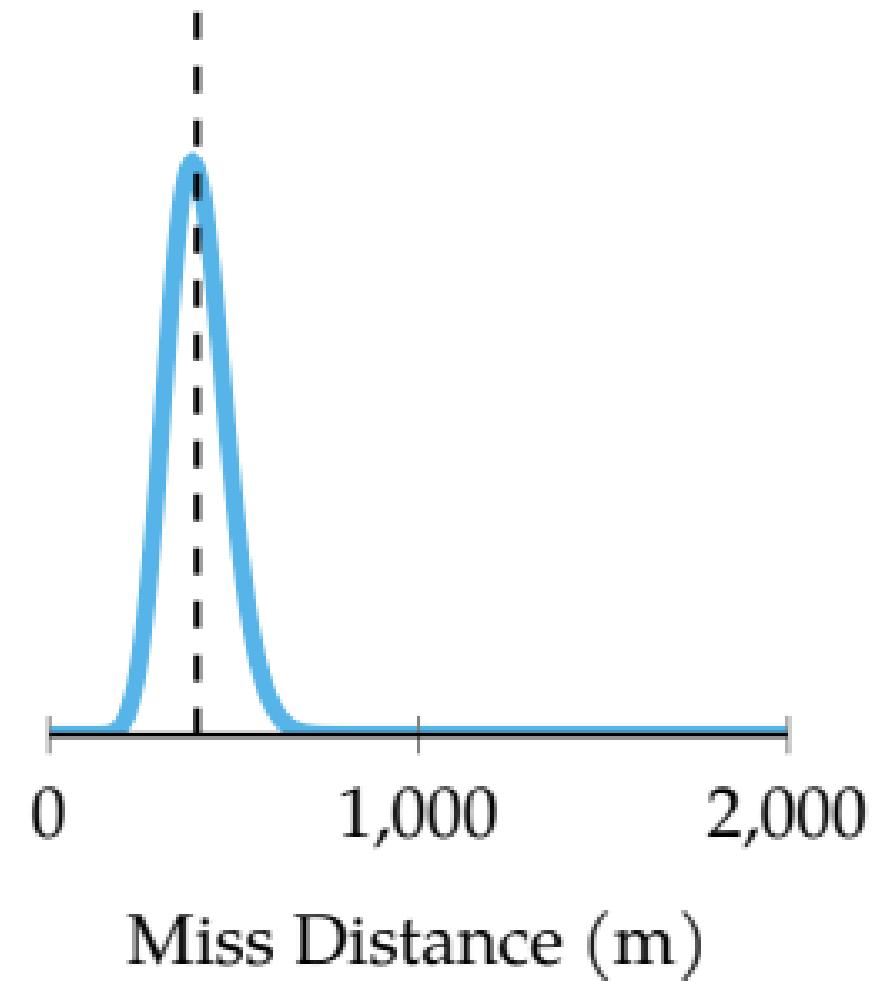
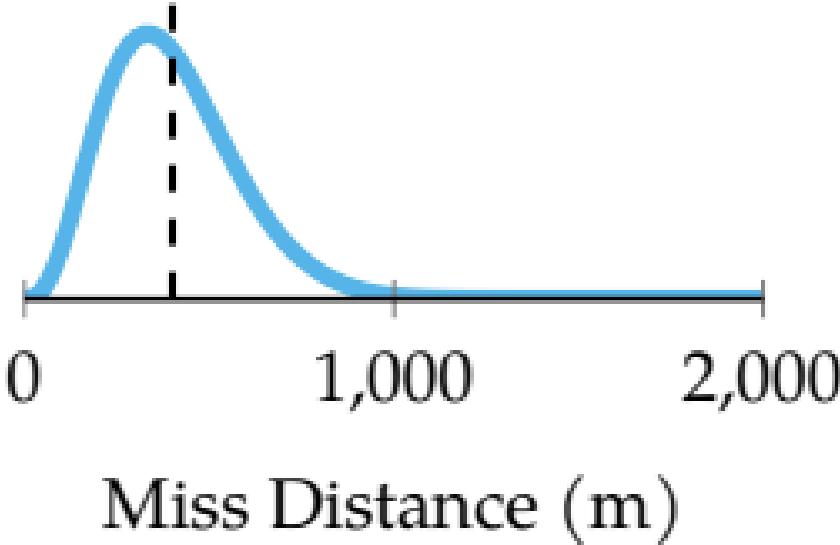
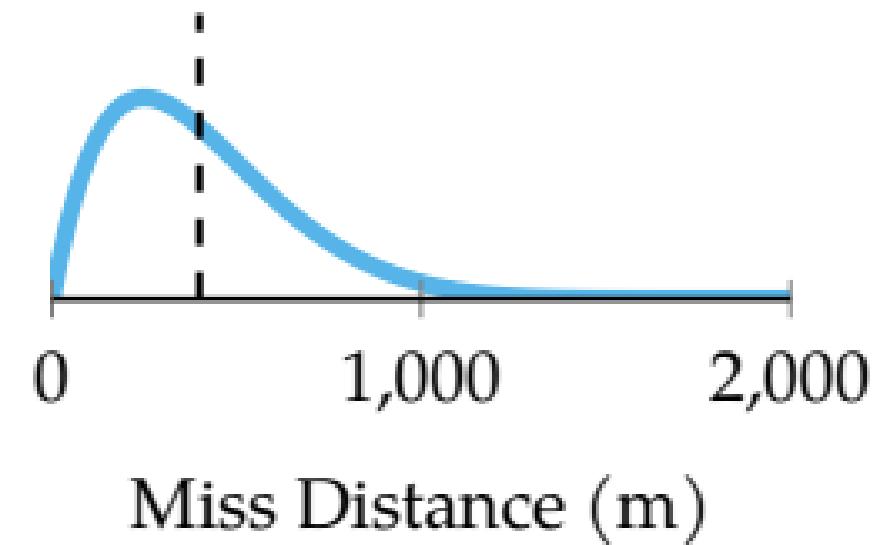


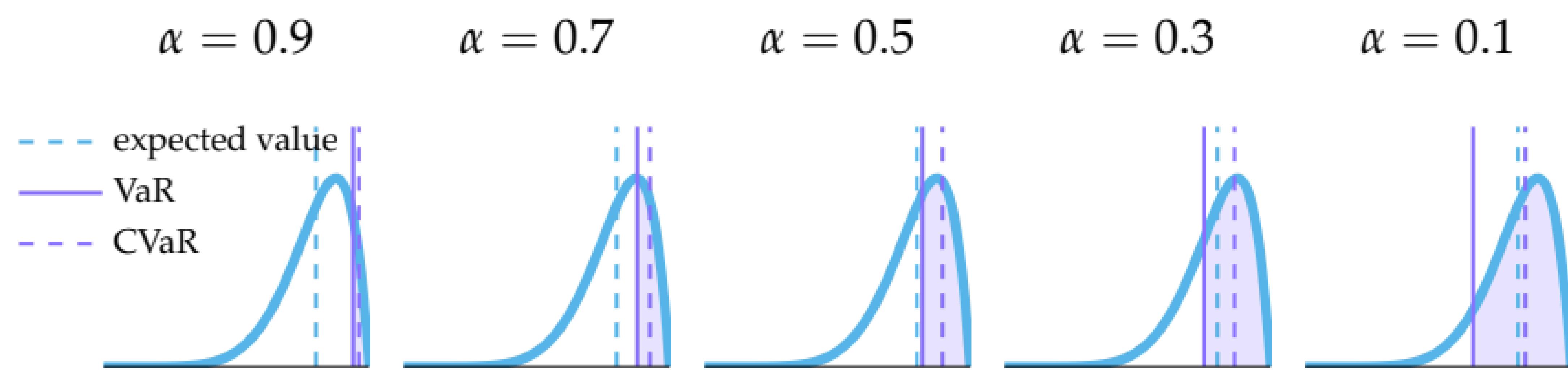
expected  
miss distance

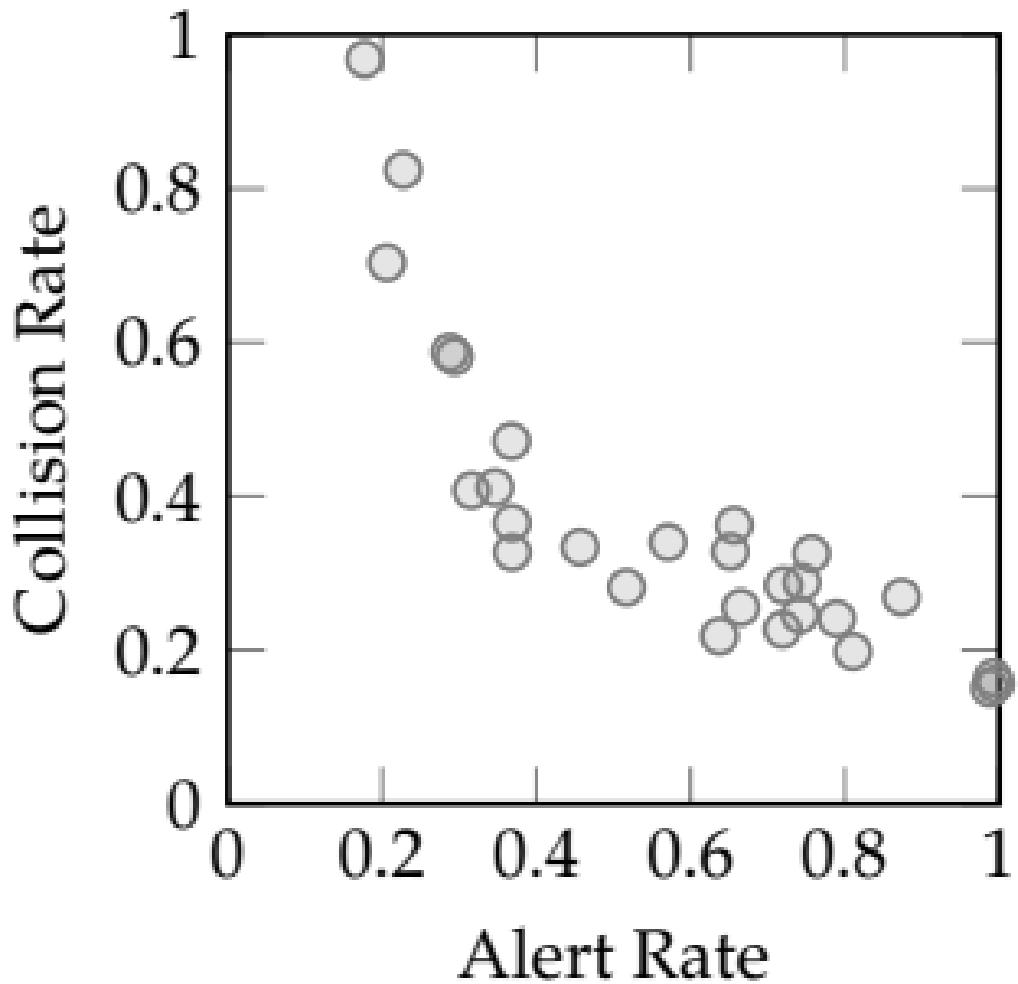
A blue bell-shaped curve is plotted against a horizontal axis labeled "Miss Distance (m)". The curve starts at zero, rises to a peak, and then decays back towards zero. A vertical dashed line is drawn from the peak of the curve down to the horizontal axis, marking the expected miss distance. An arrow points from the text "expected miss distance" to this dashed line.

0 1,000 2,000

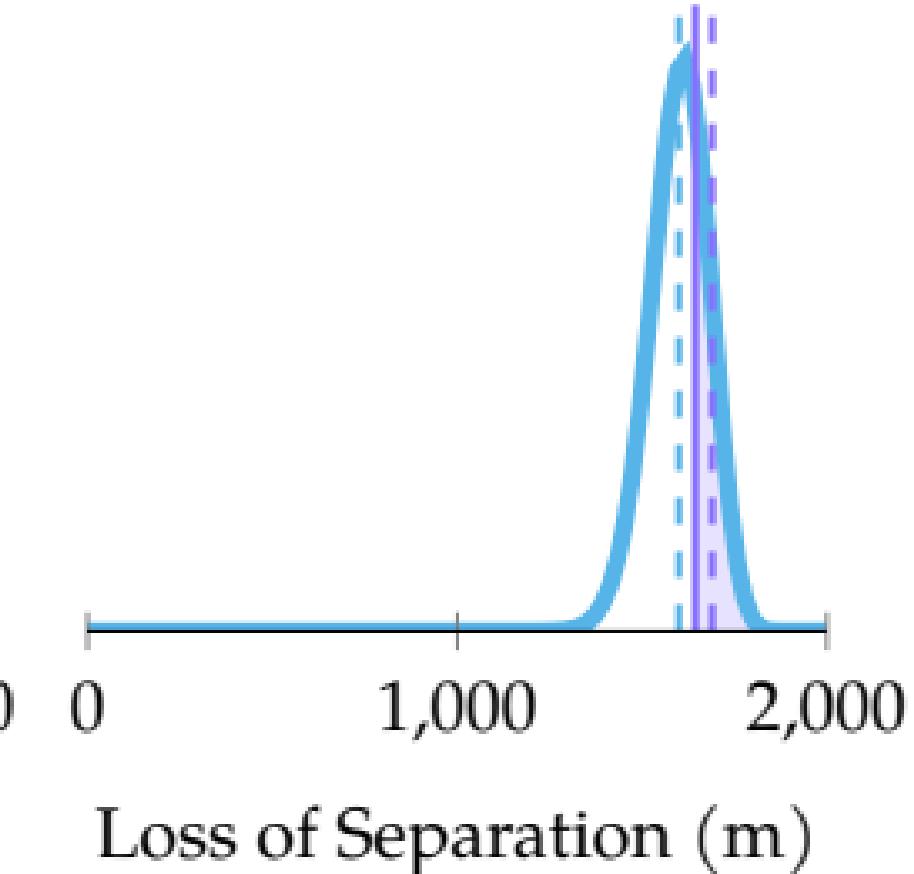
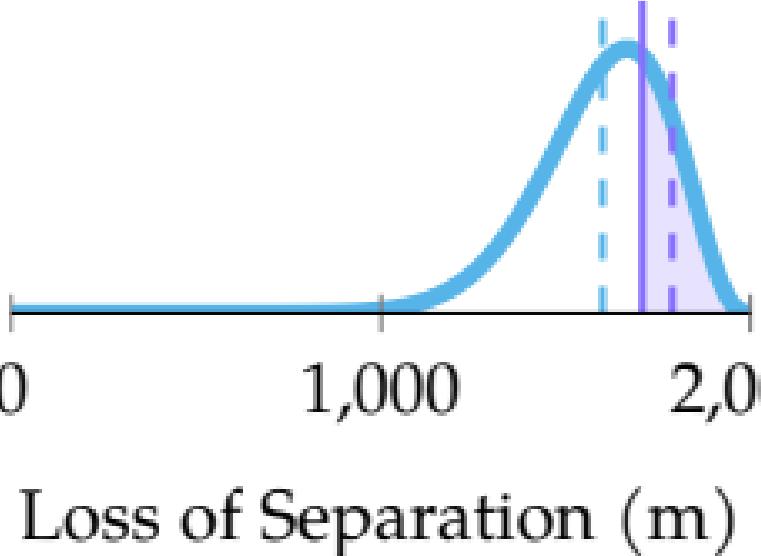
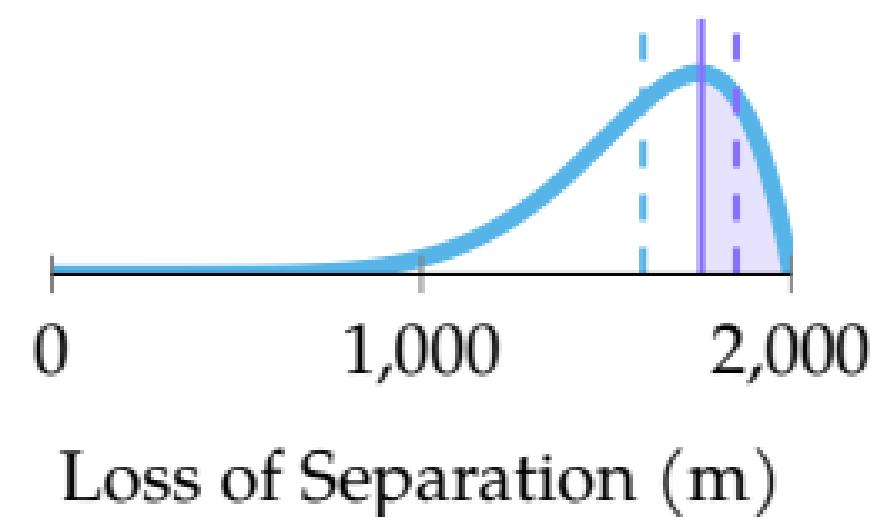
Miss Distance (m)

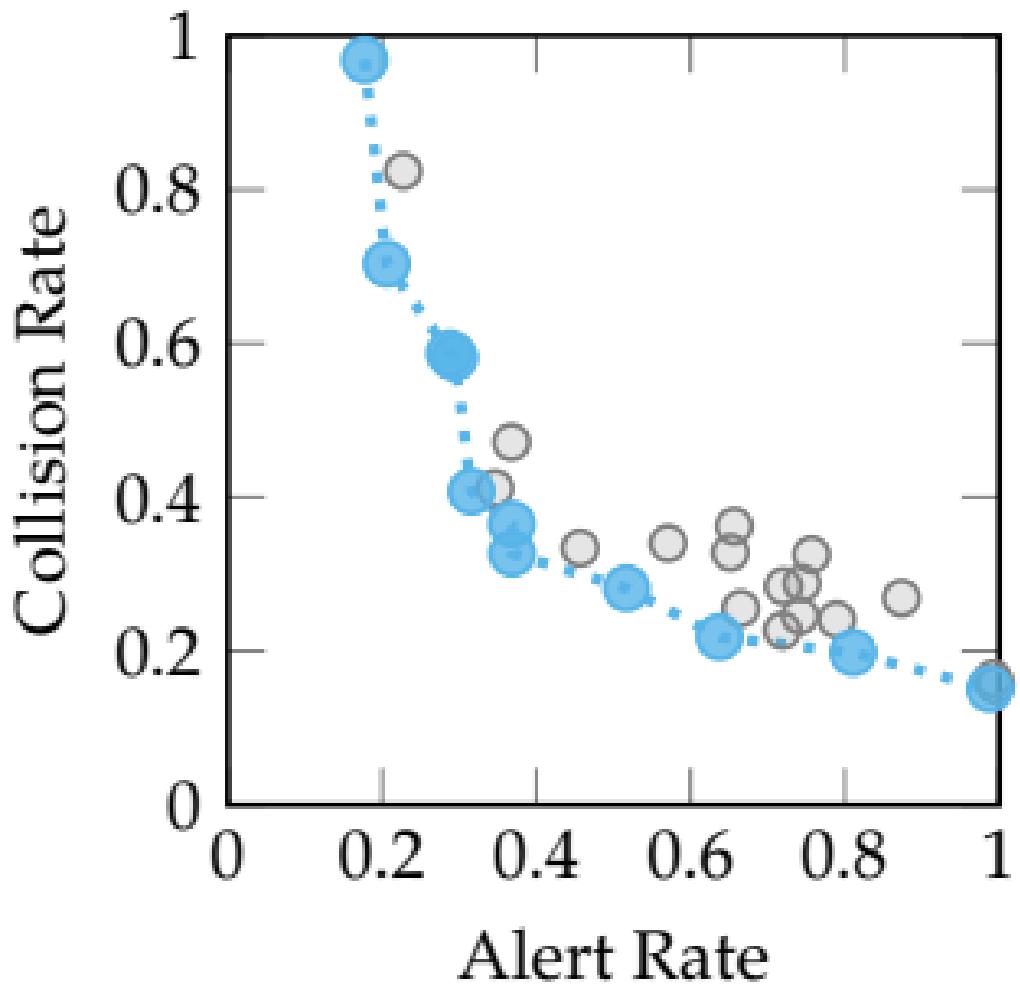




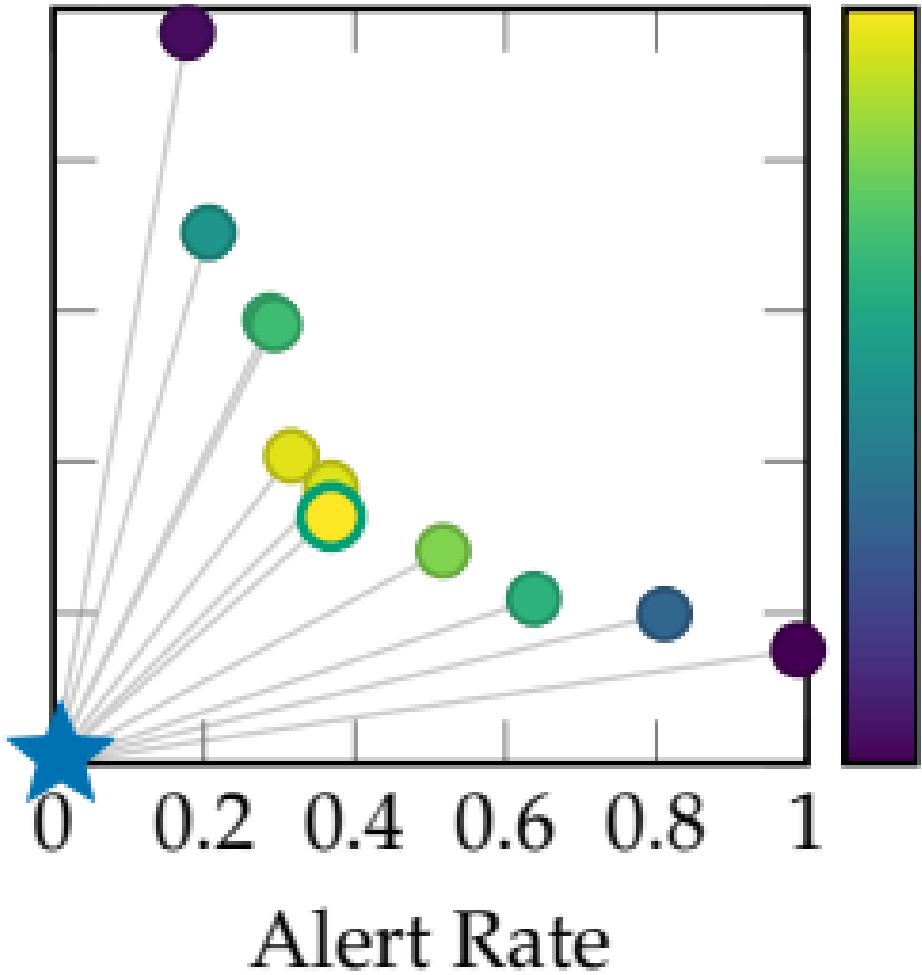


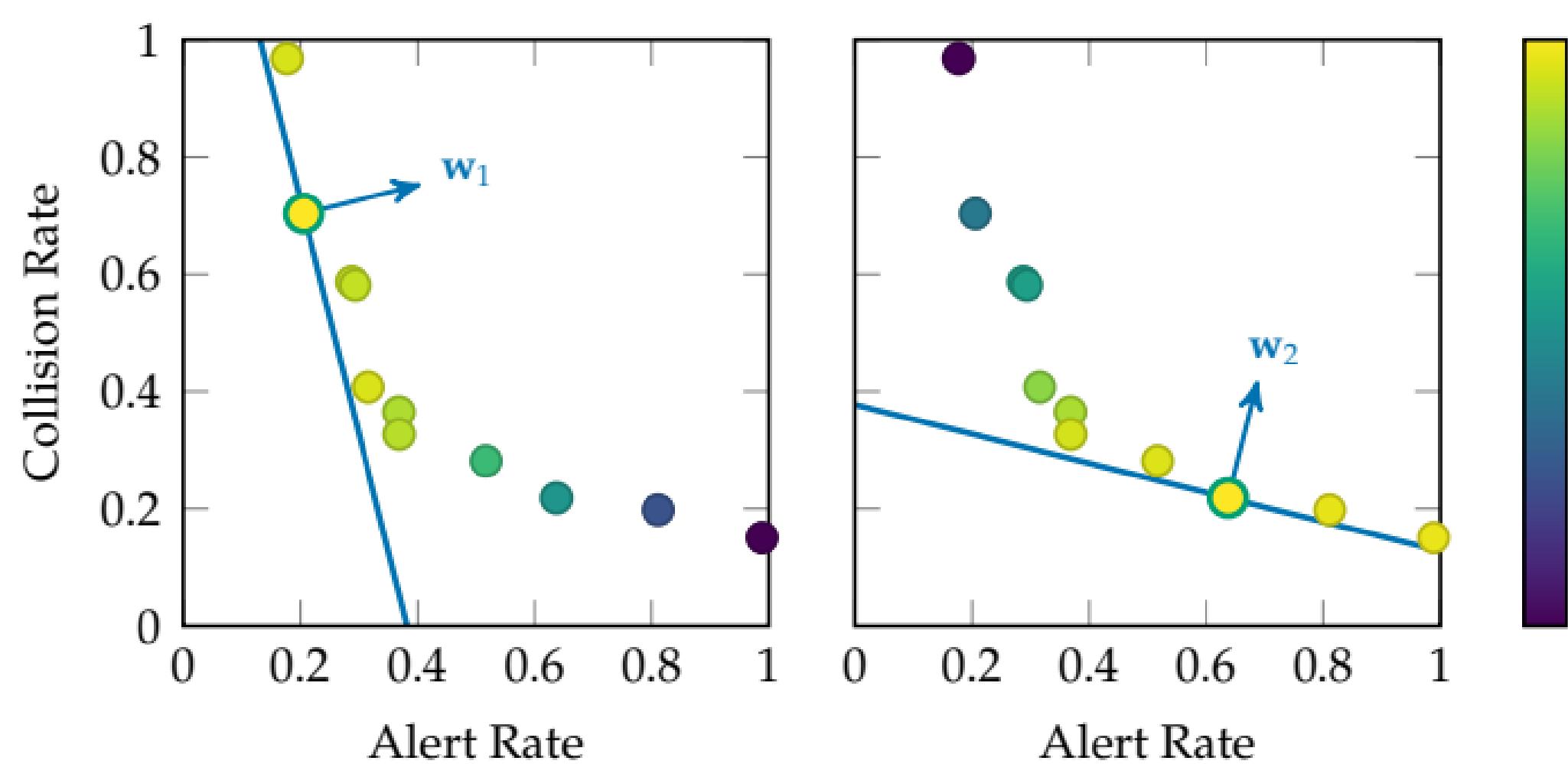
-- expected value  
— VaR  
- - - CVaR

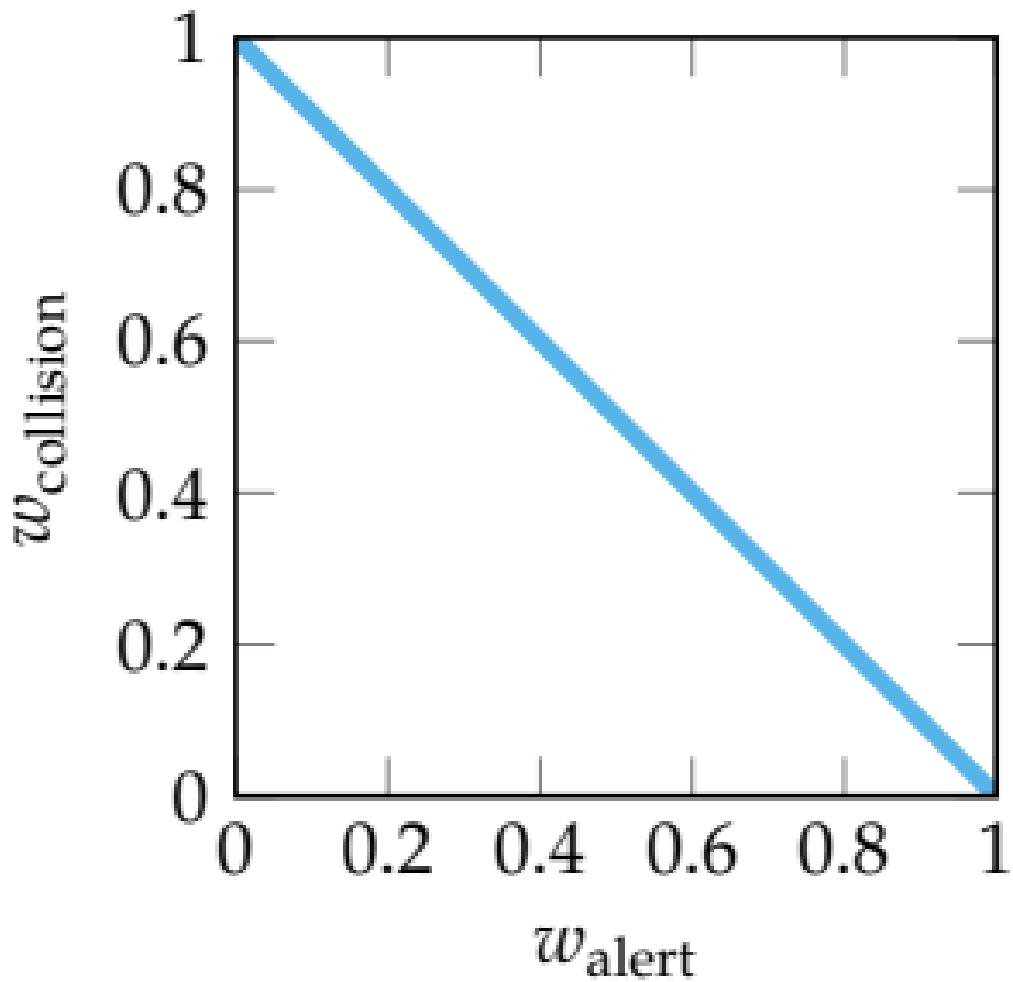


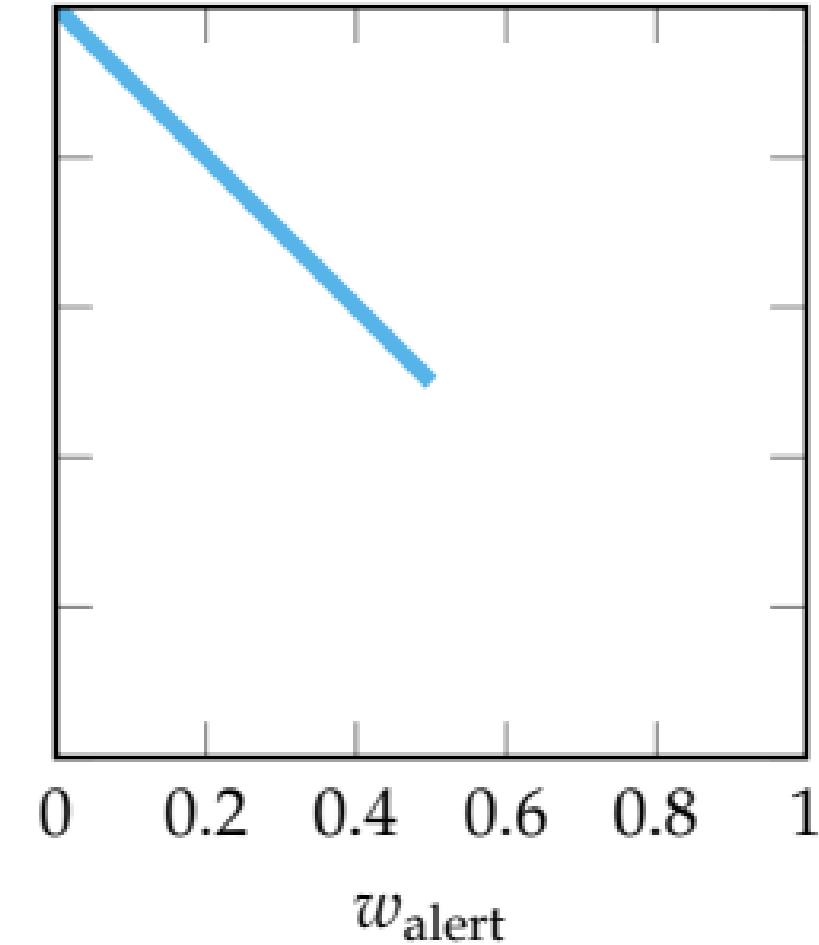
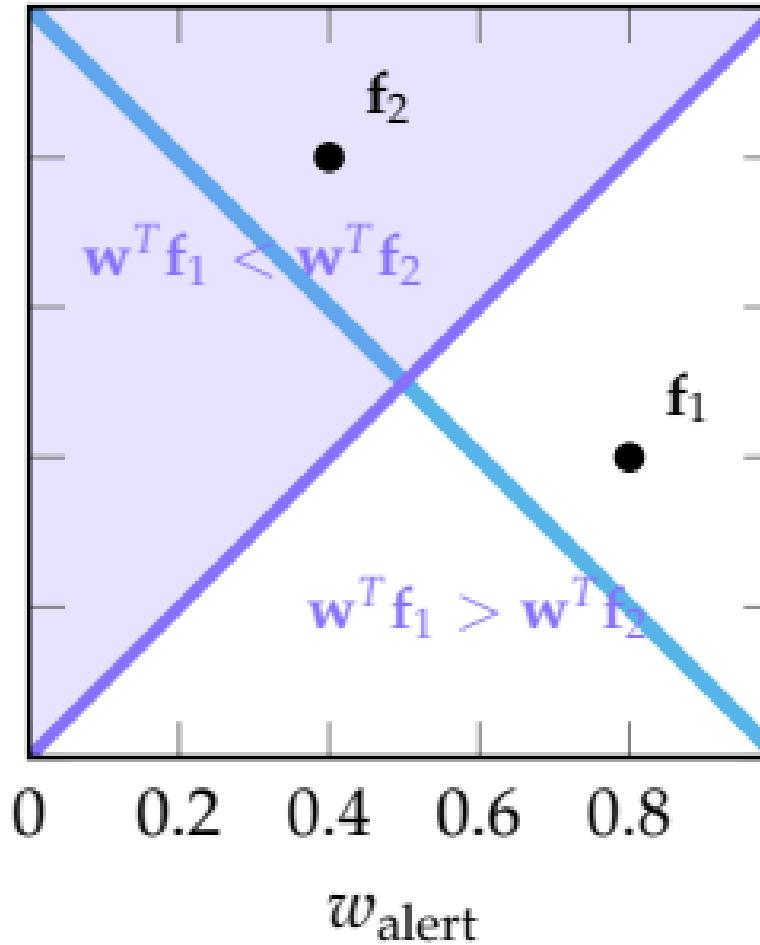
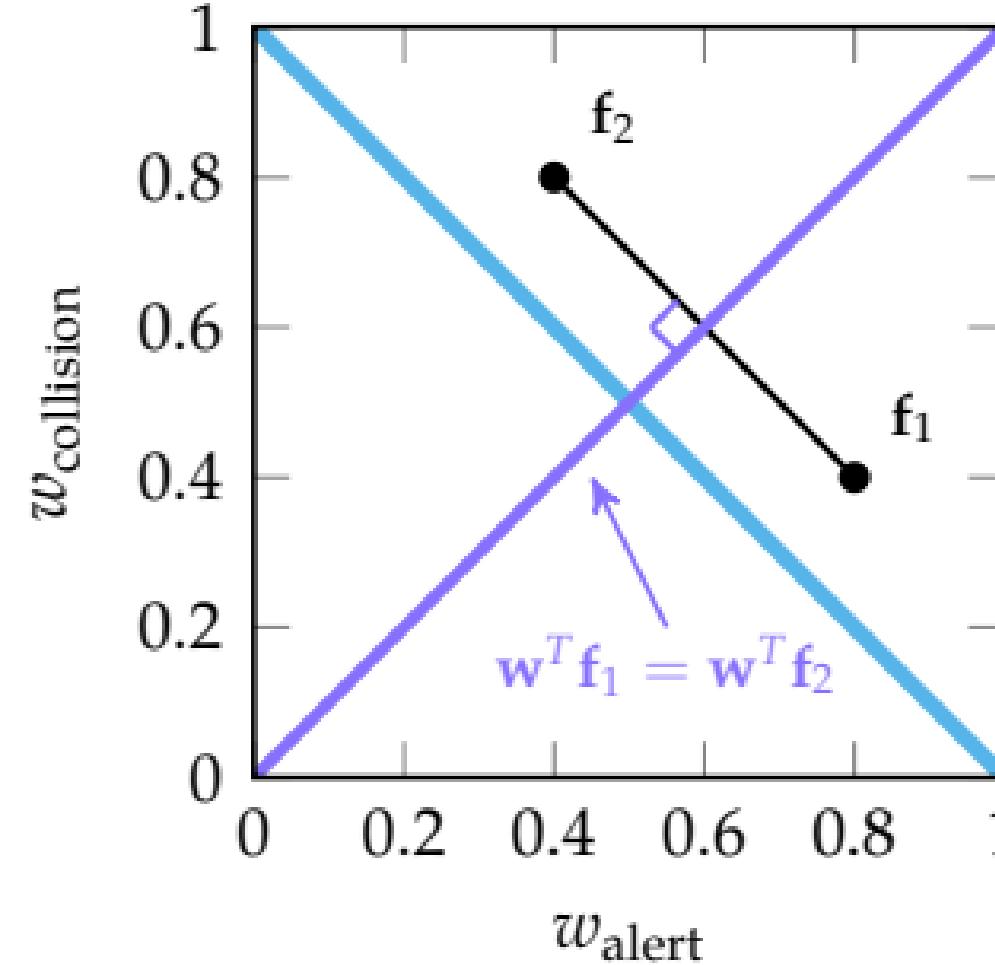


Collision Rate

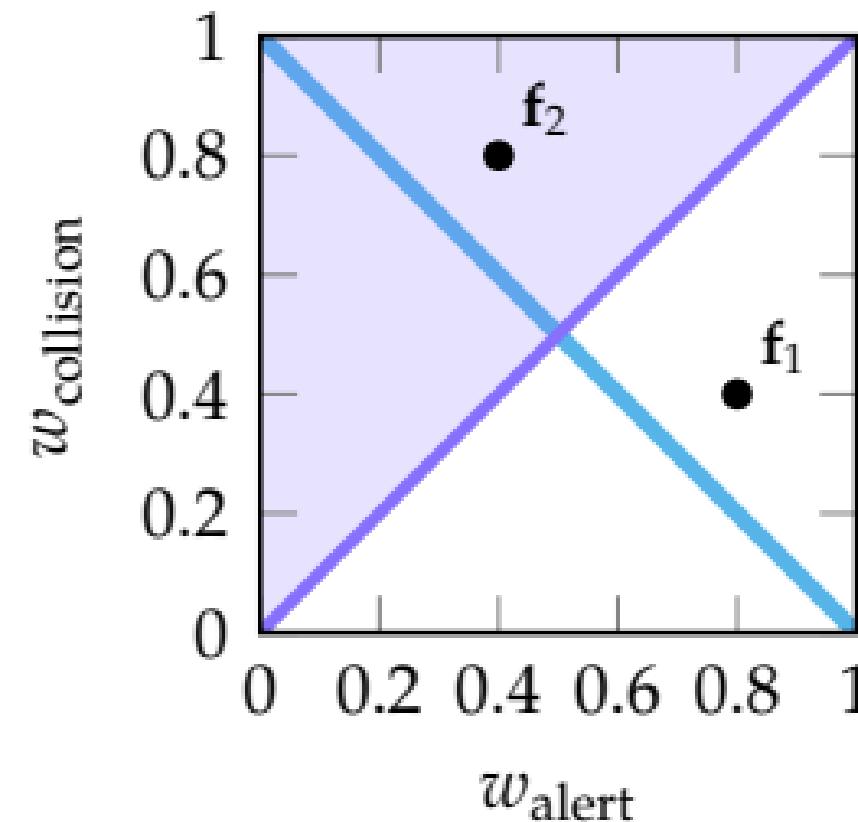




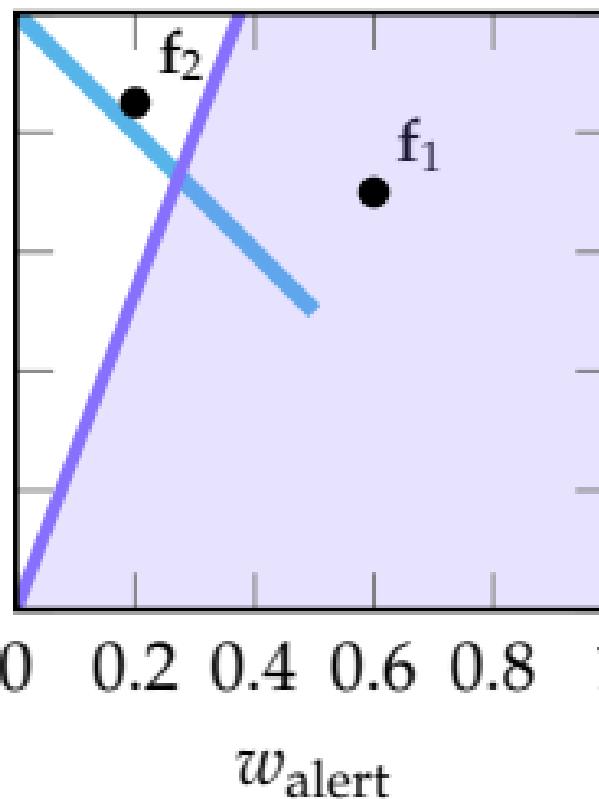




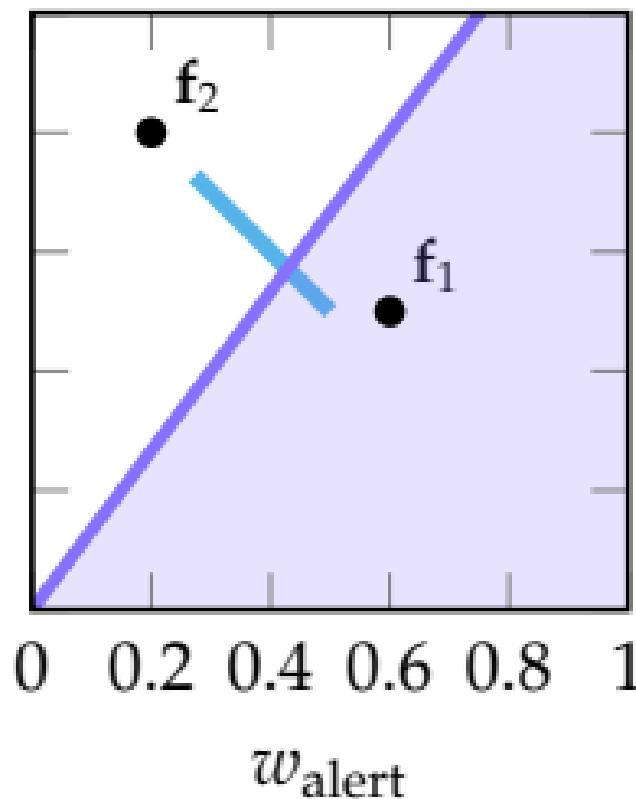
Query 1



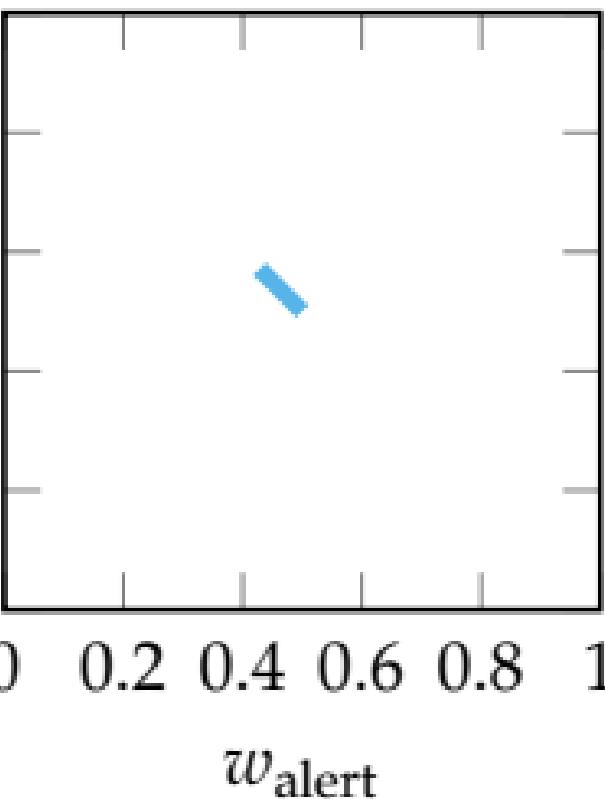
Query 2



Query 3

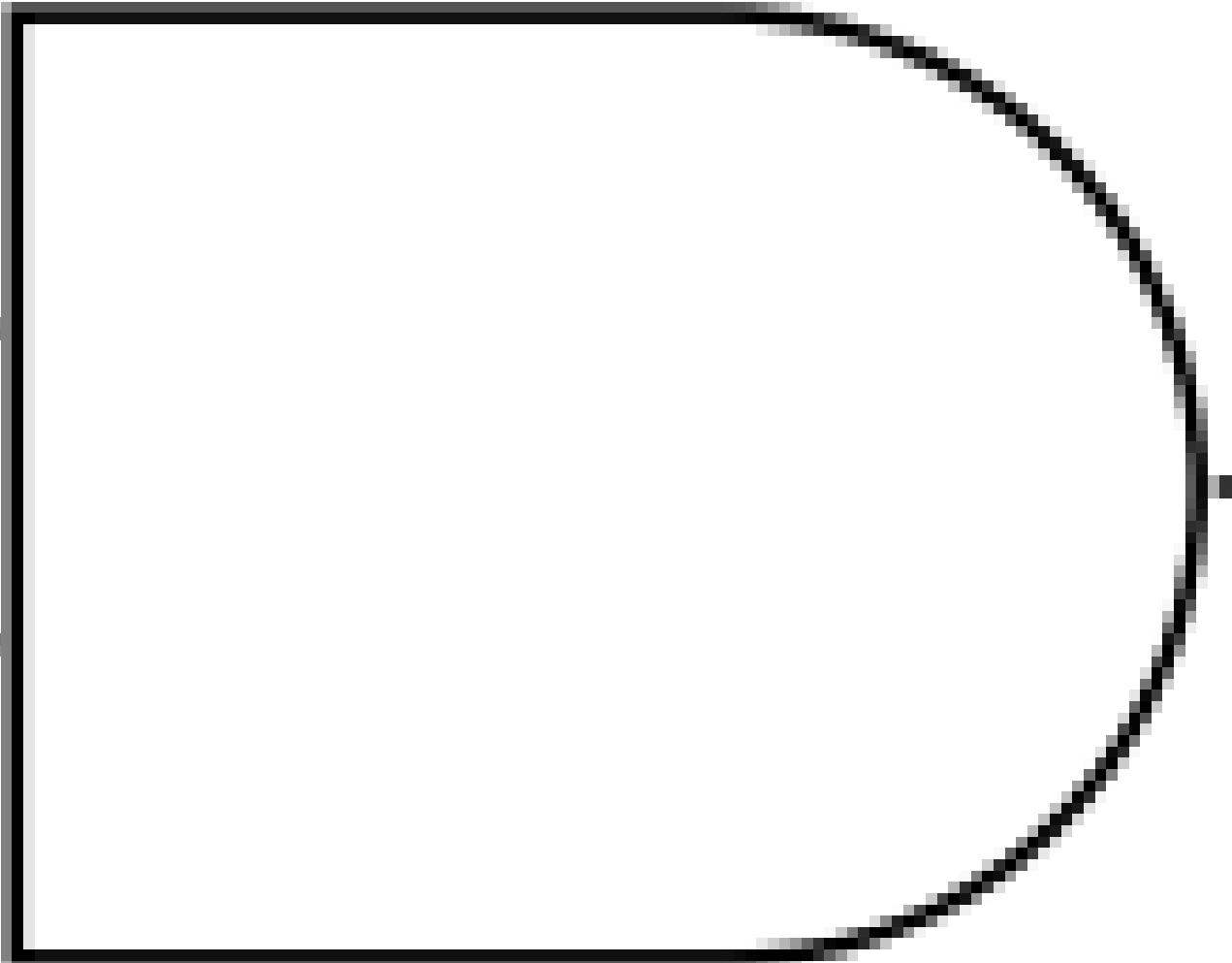
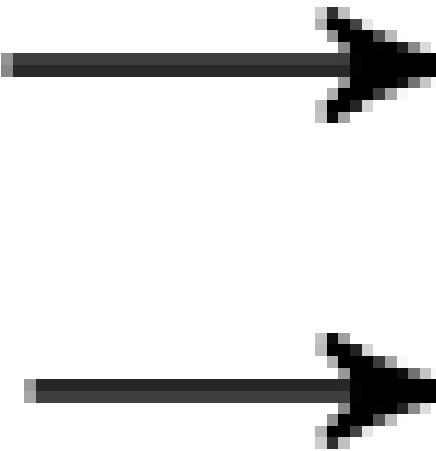


Final Weight Space



$P$

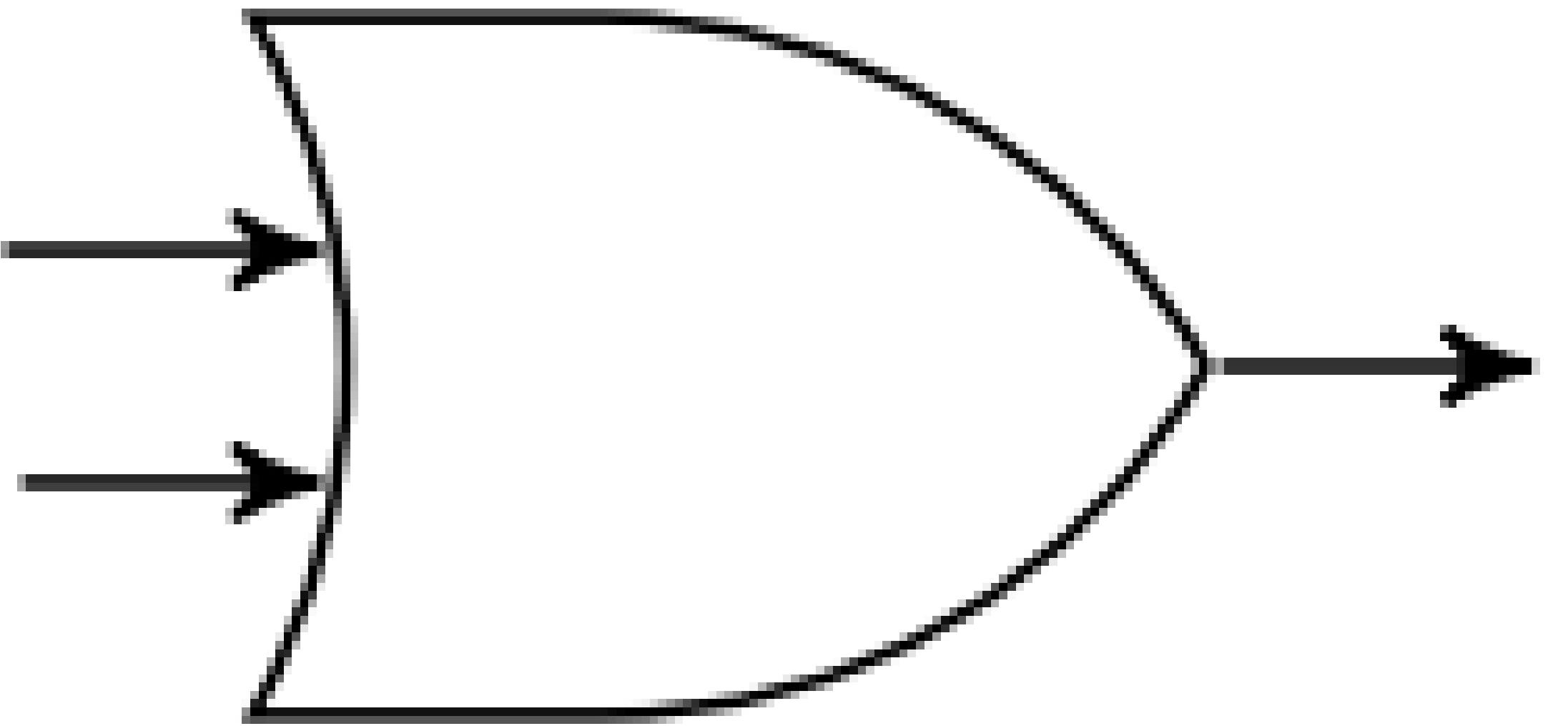
$Q$



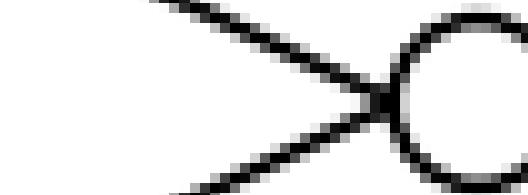
$P \wedge Q$

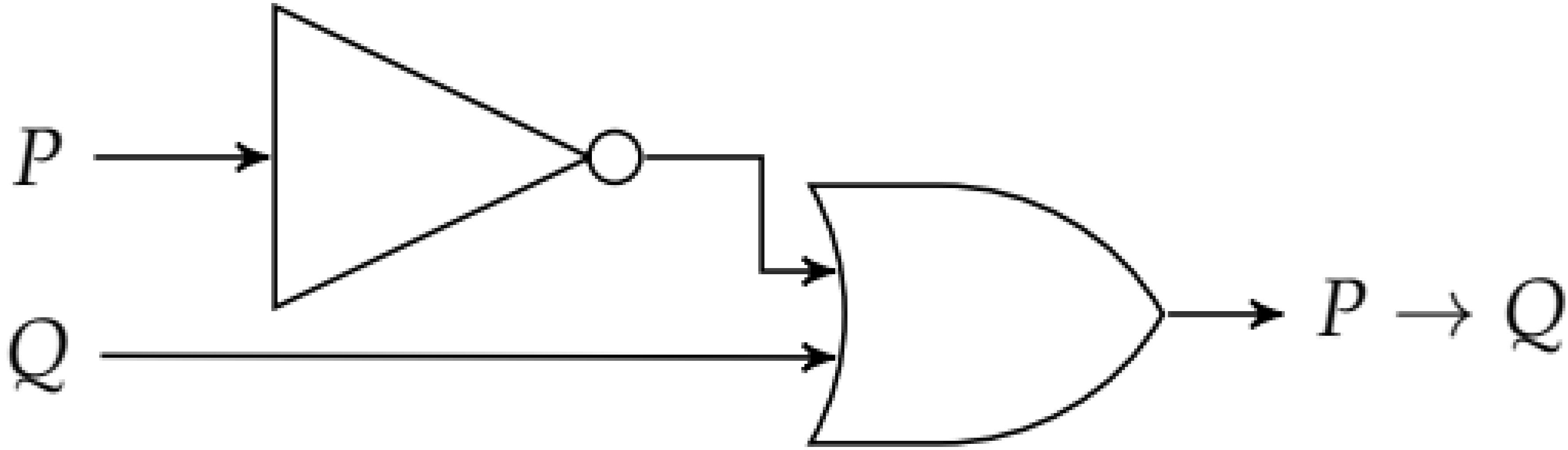
$\varnothing$

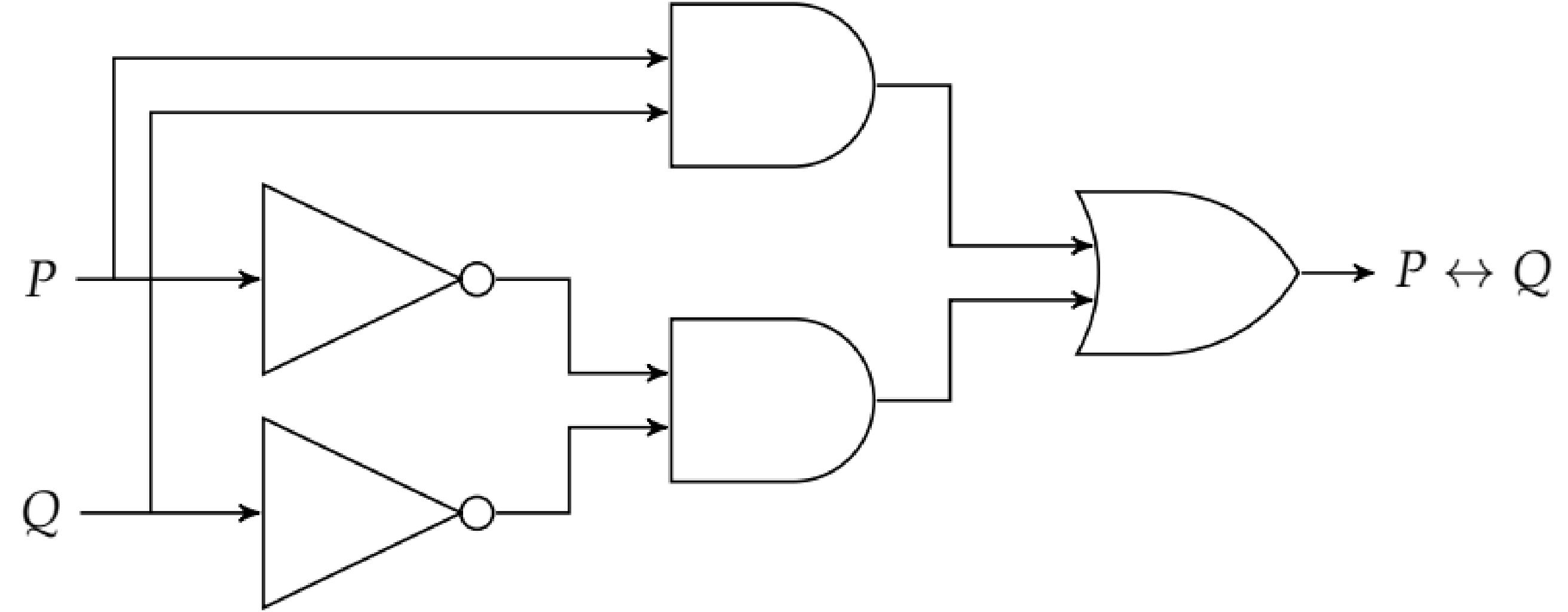
$P$   
 $Q$

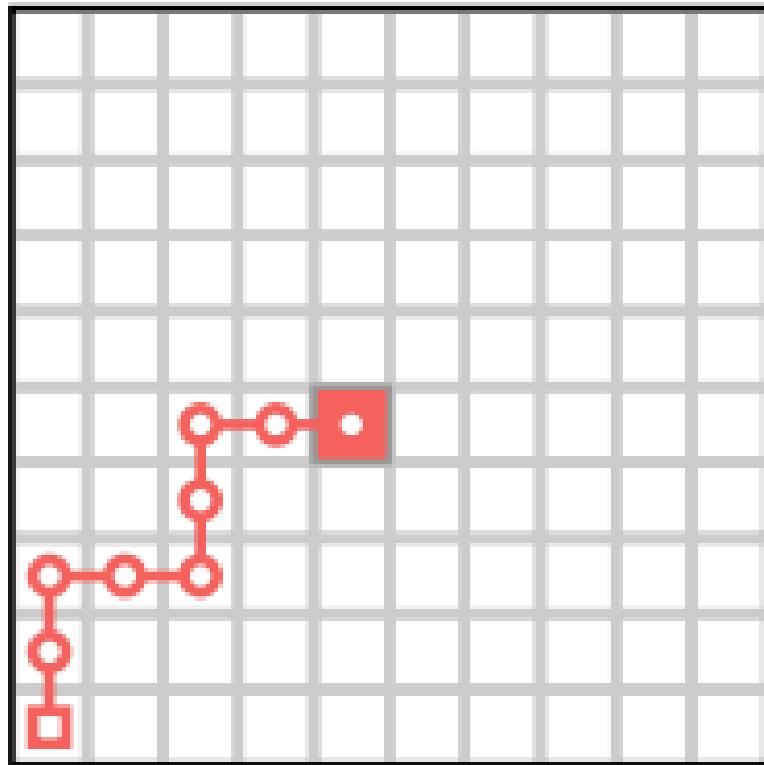
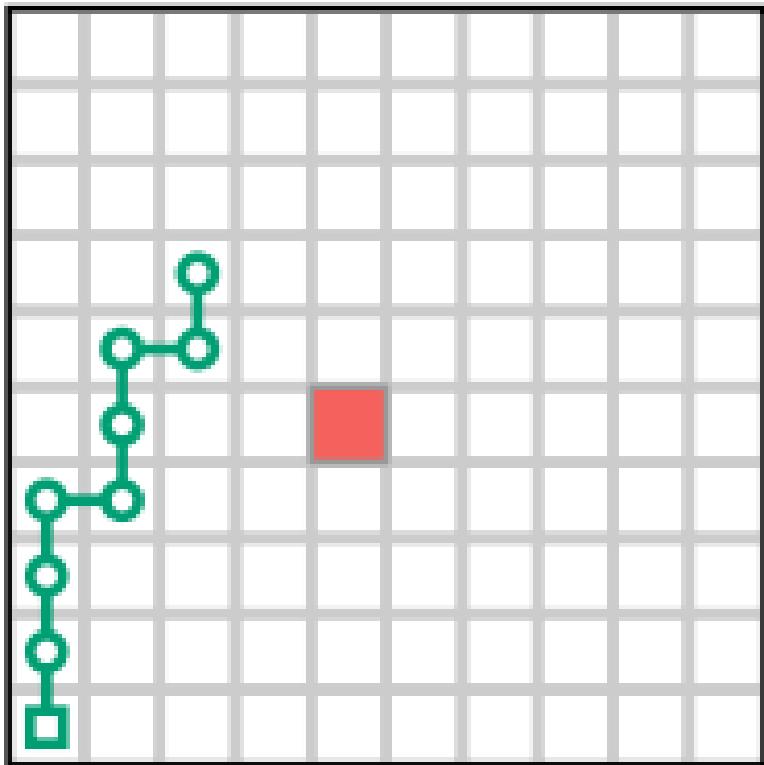


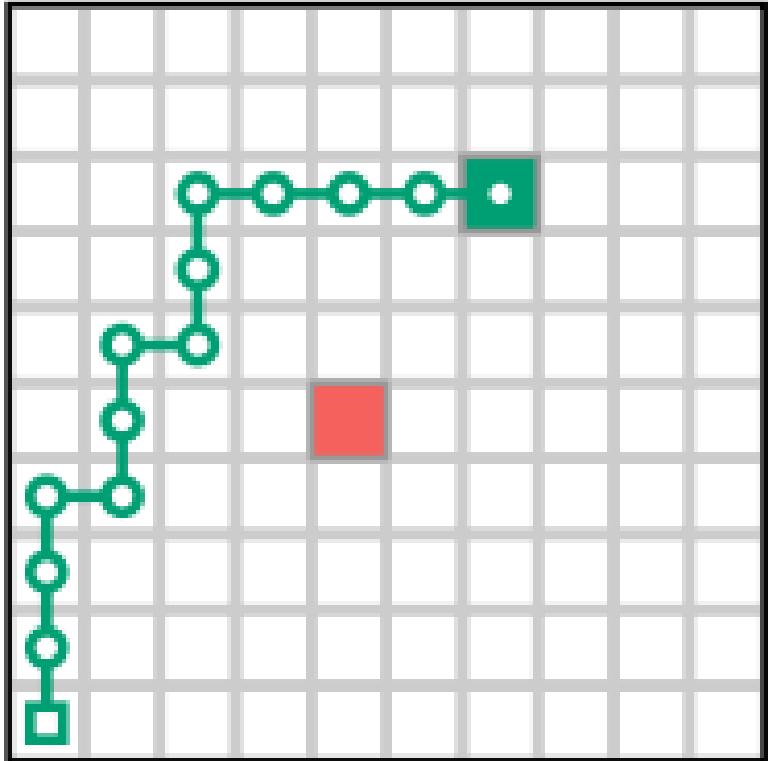
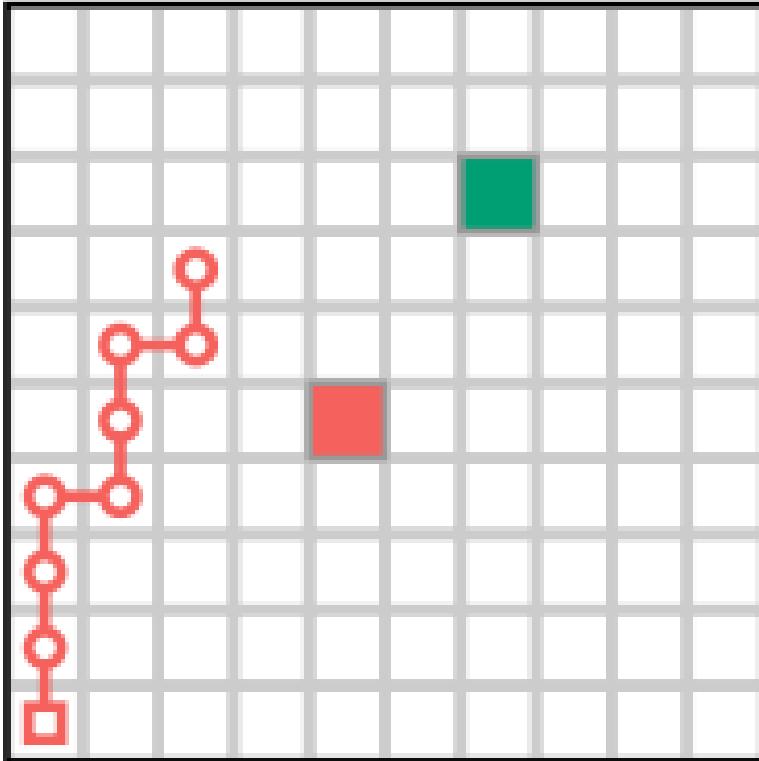
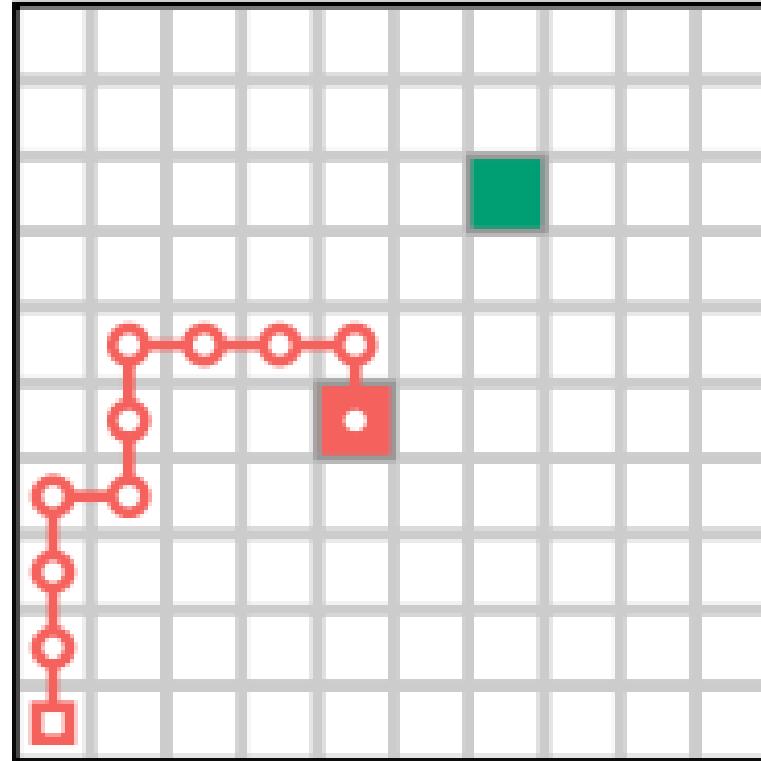
$P \vee Q$

$P$  $\neg P$

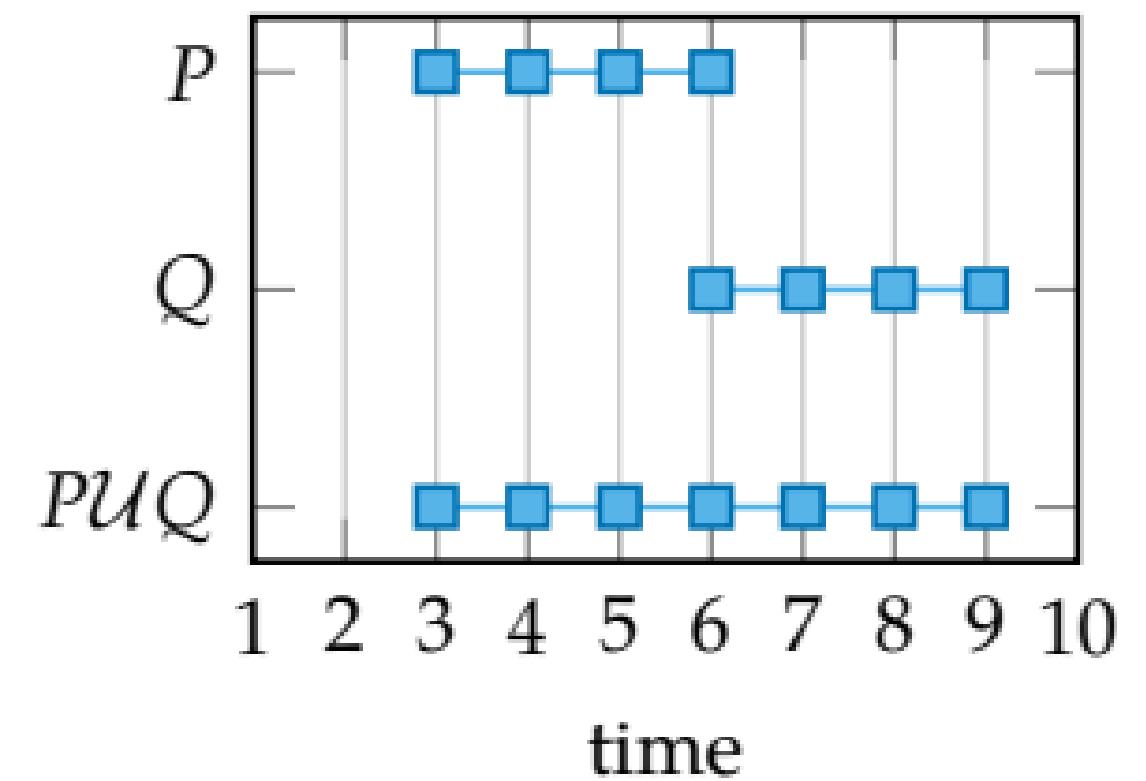




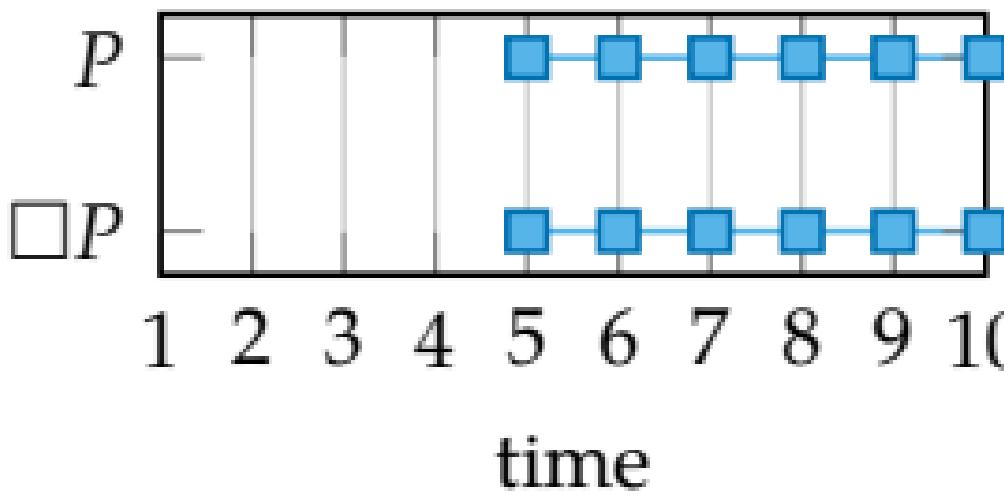
$\psi_1 = \text{true}$  $\psi_1 = \text{false}$ 

$\psi_2 = \text{true}$  $\psi_2 = \text{false}$  $\psi_3 = \text{false}$ 

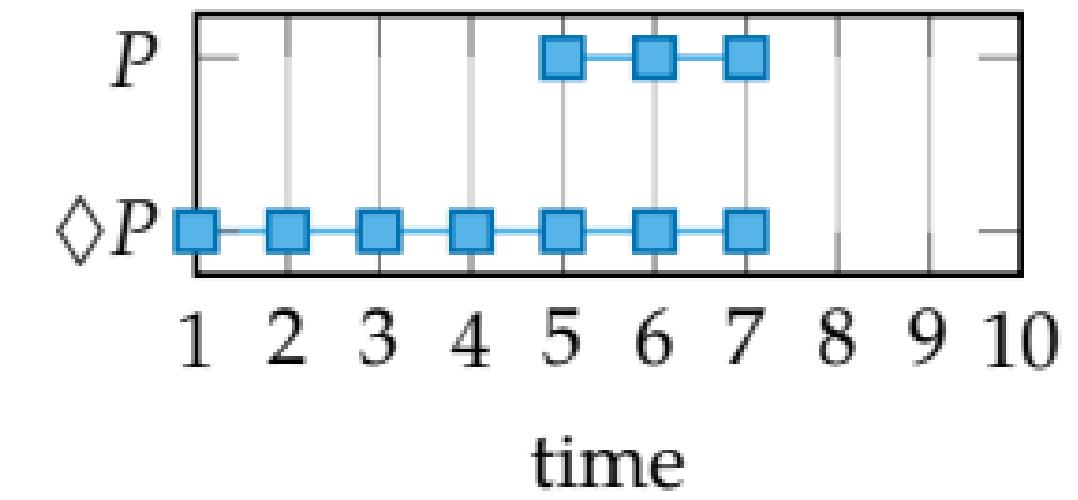
Until:  $P \mathcal{U} Q$

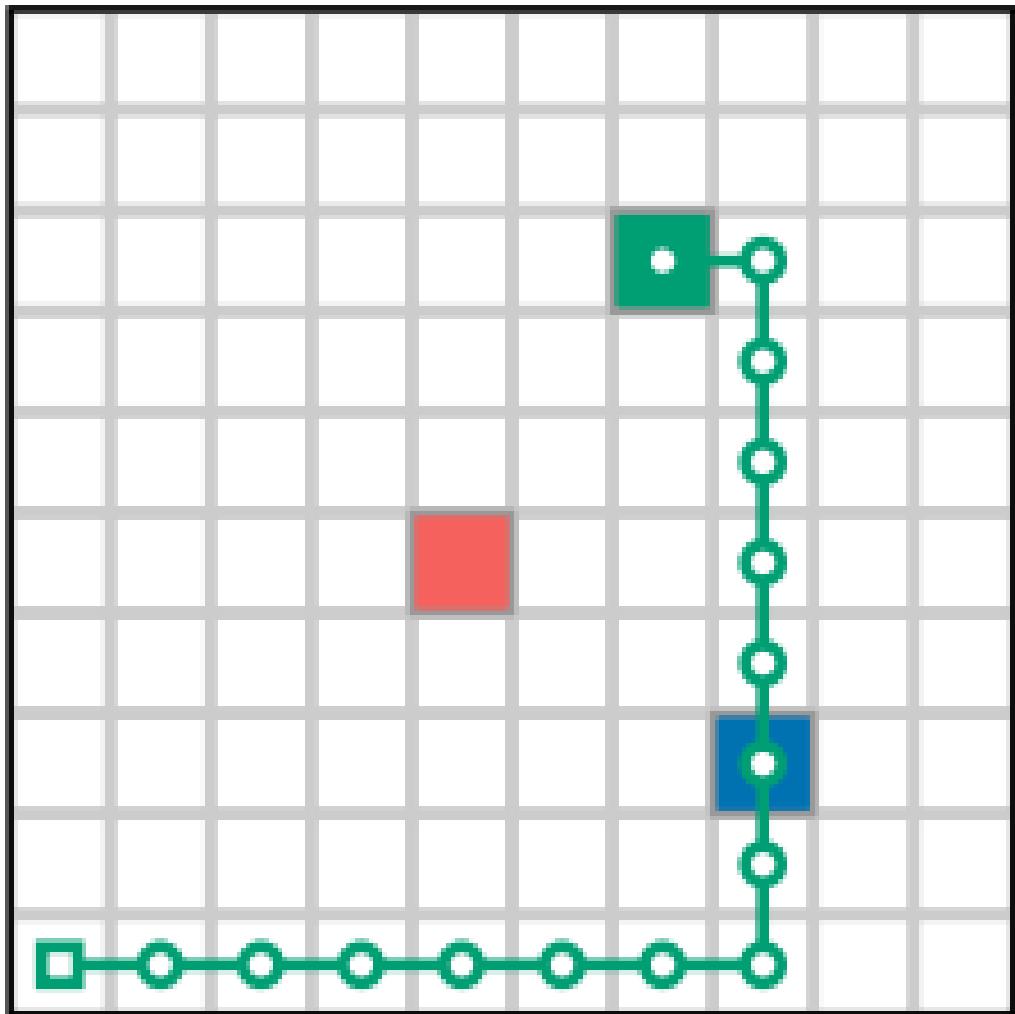


Always:  $\Box P$

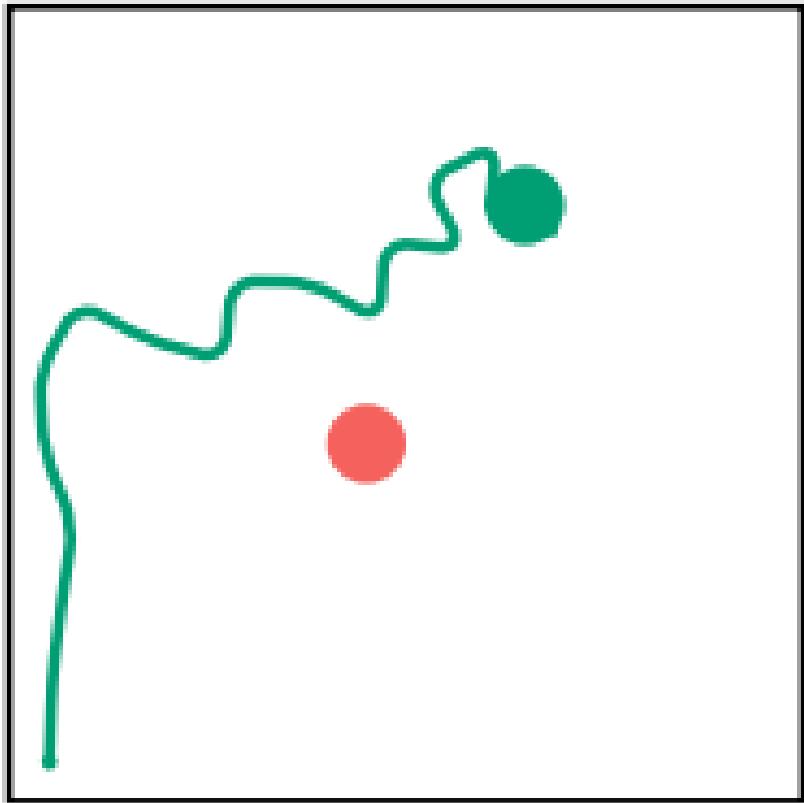


Eventually:  $\Diamond P$



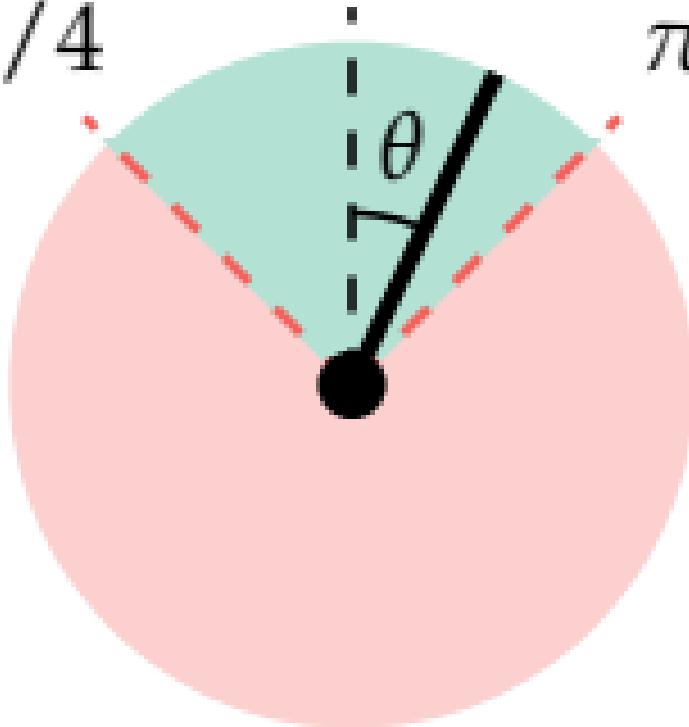
$\psi = \text{true}$ 

# Continuum World

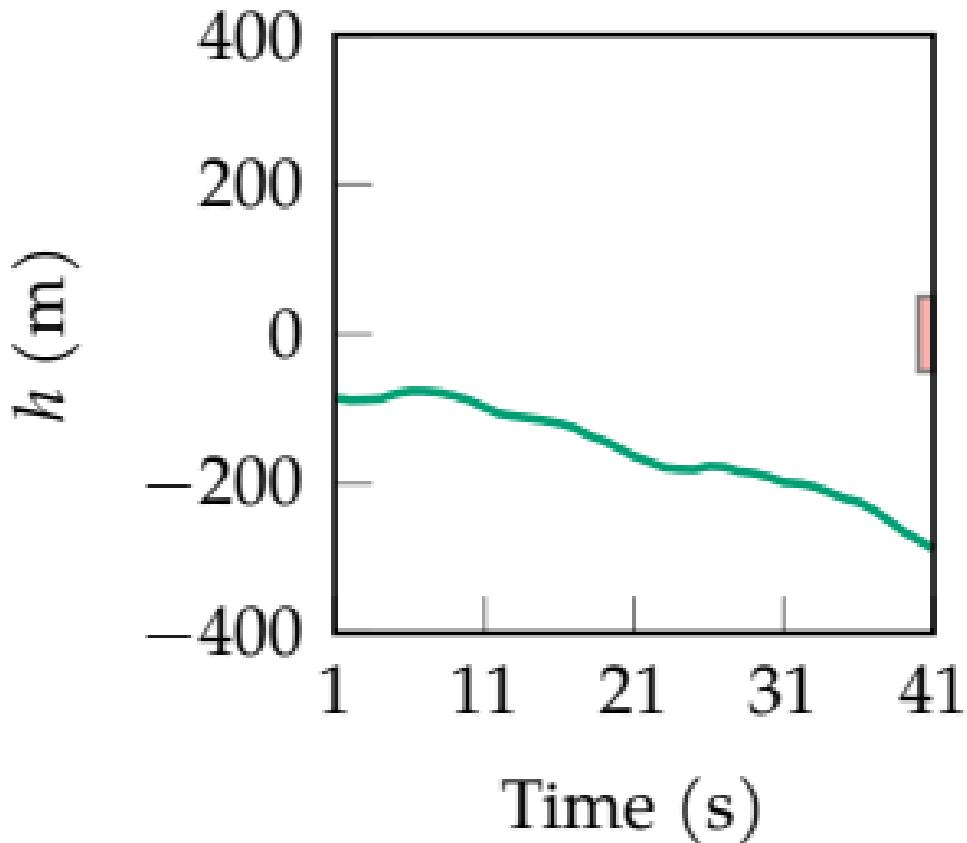


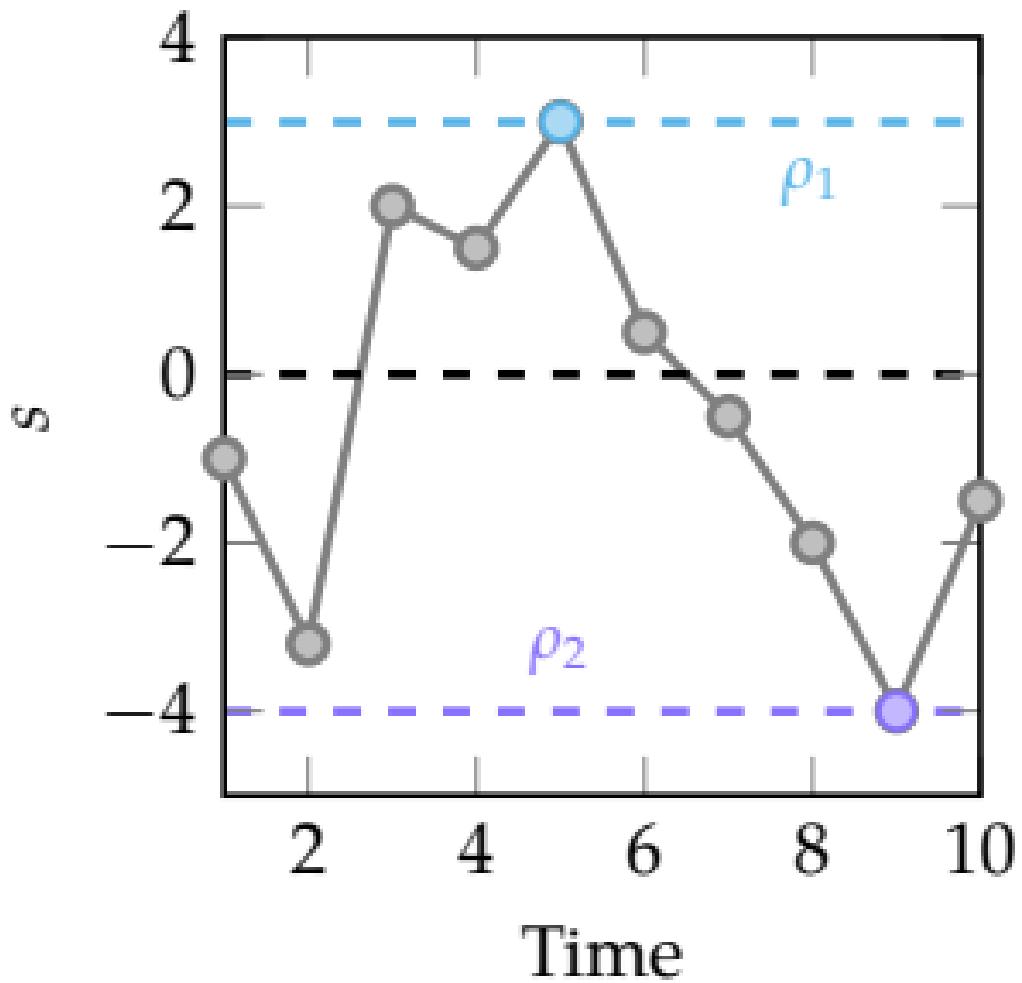
# Inverted Pendulum

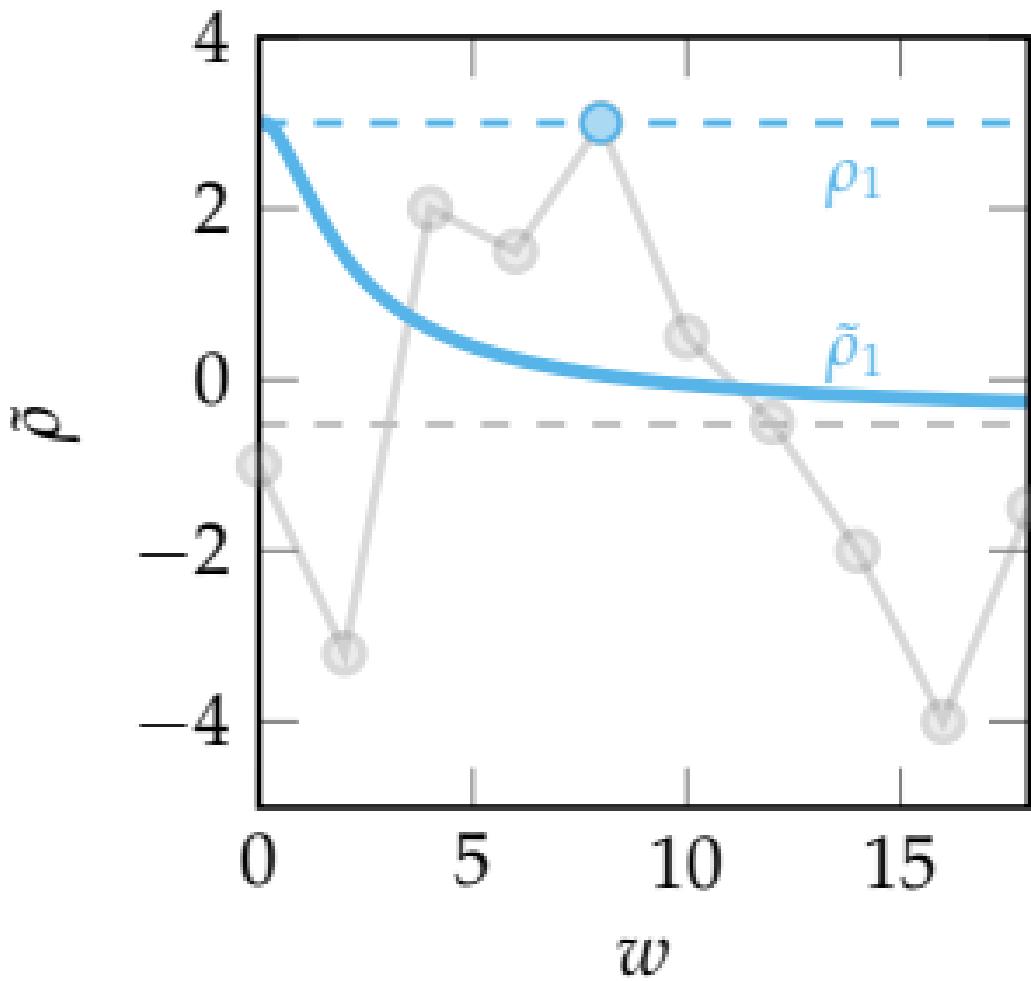
$-\pi/4$        $\pi/4$

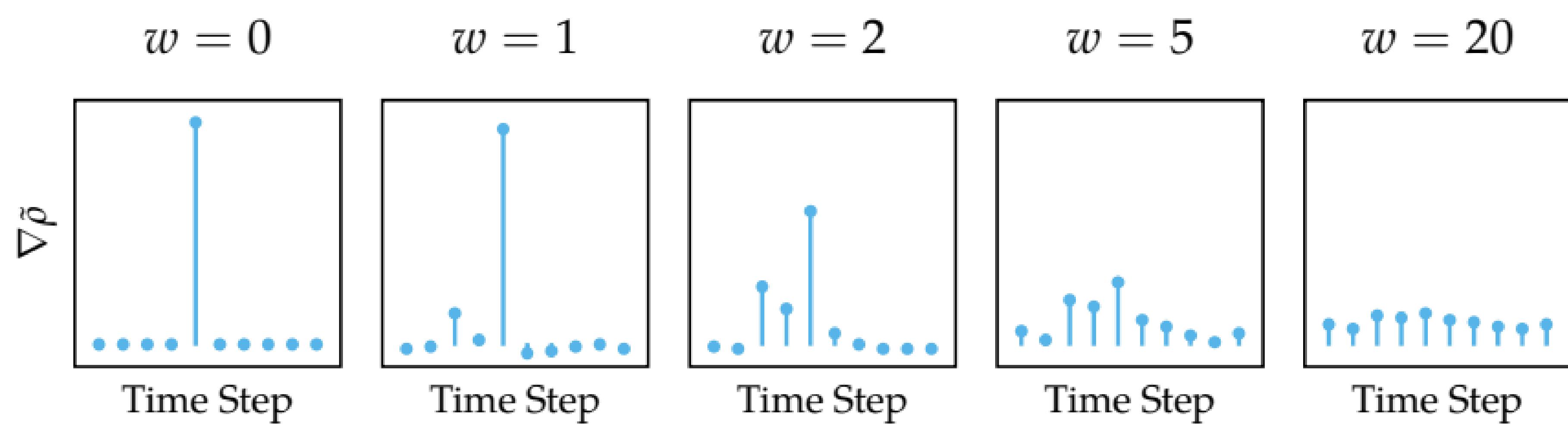


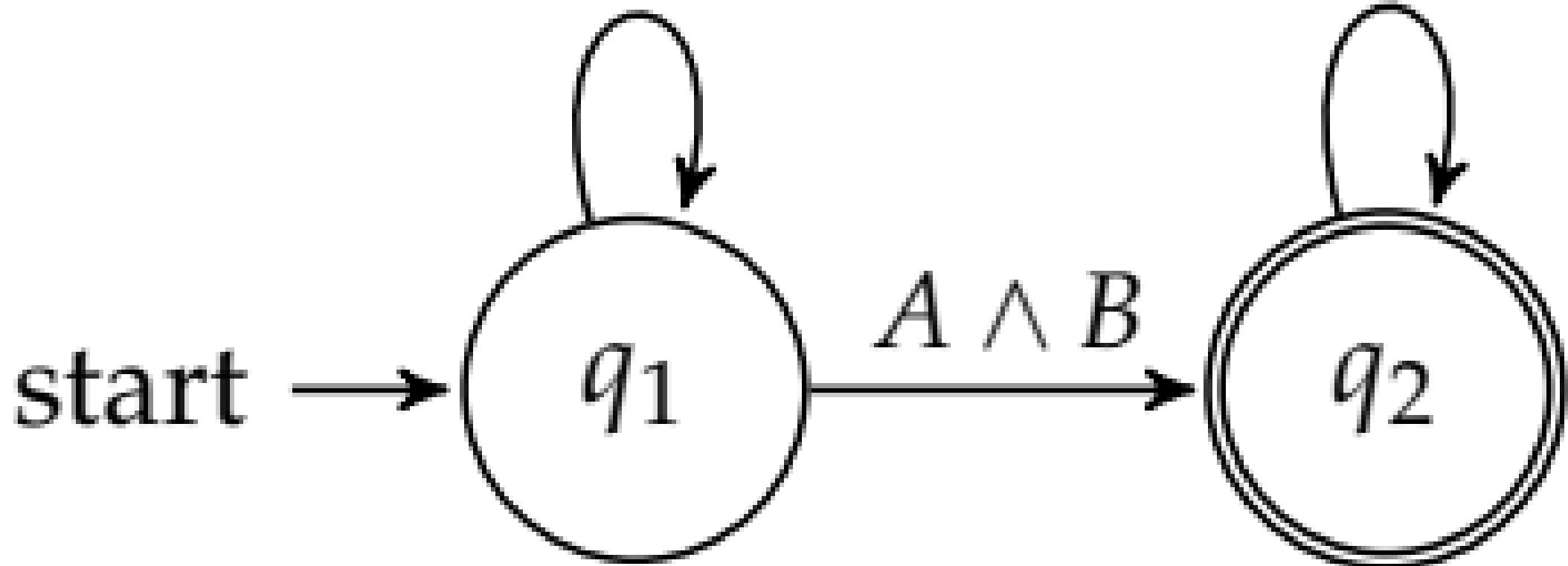
# Aircraft Collision Avoidance

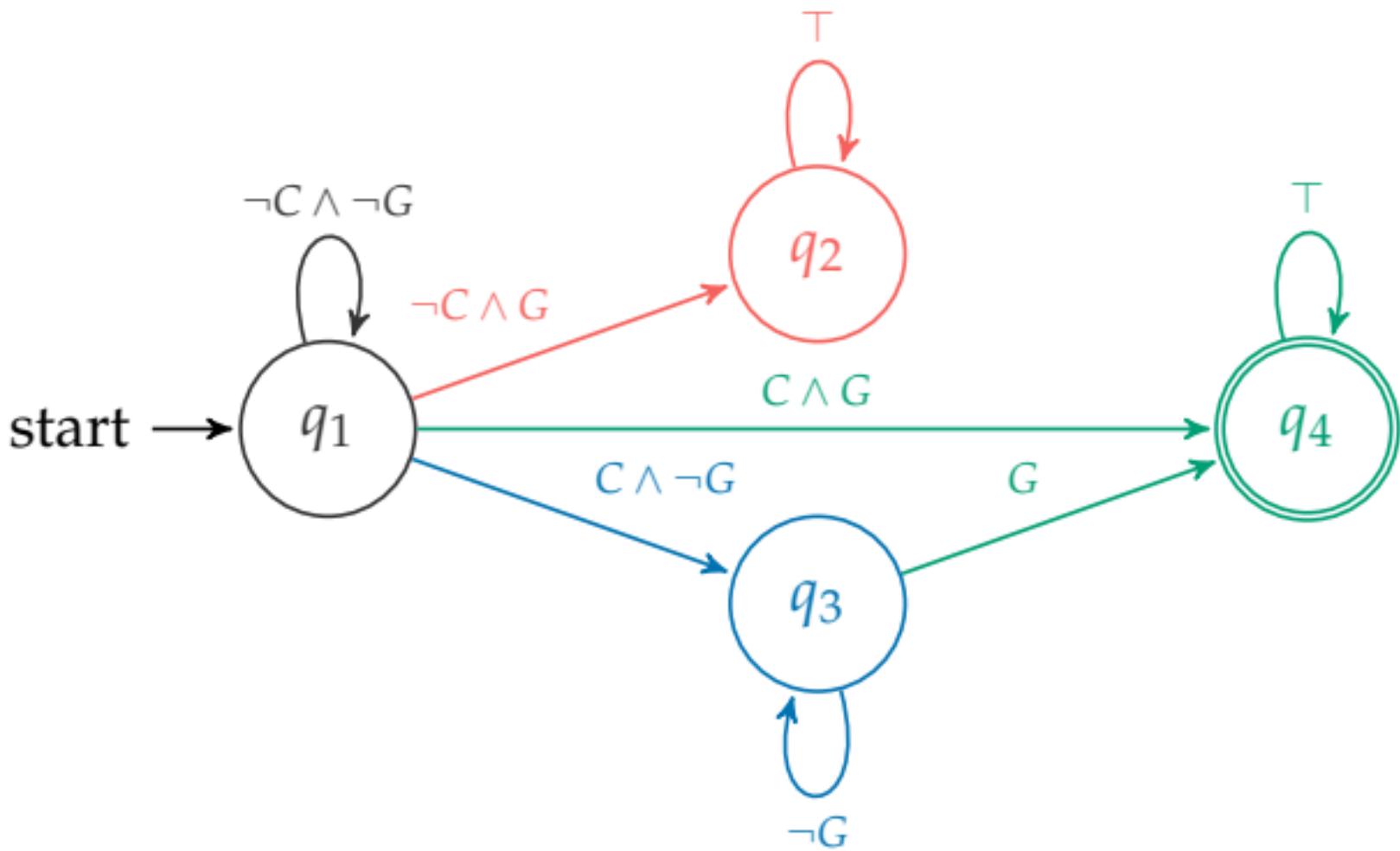


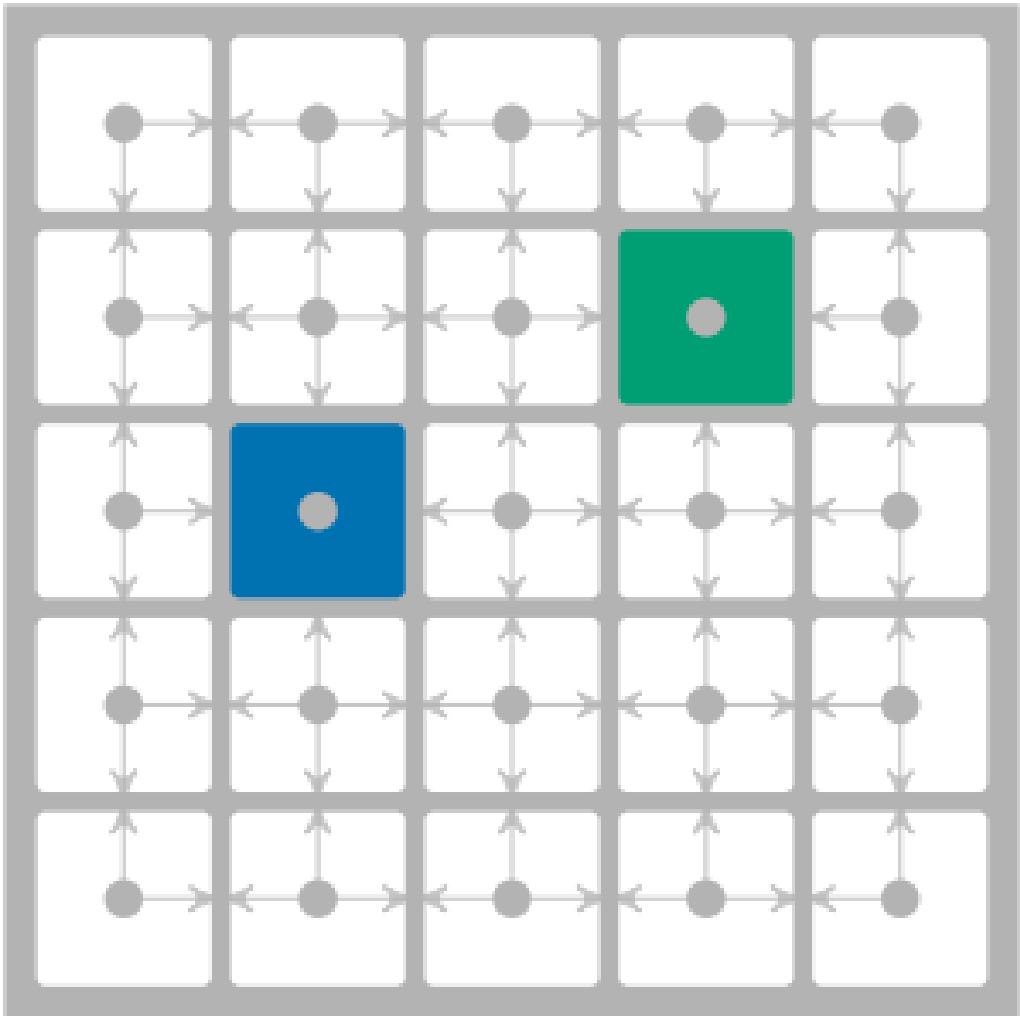




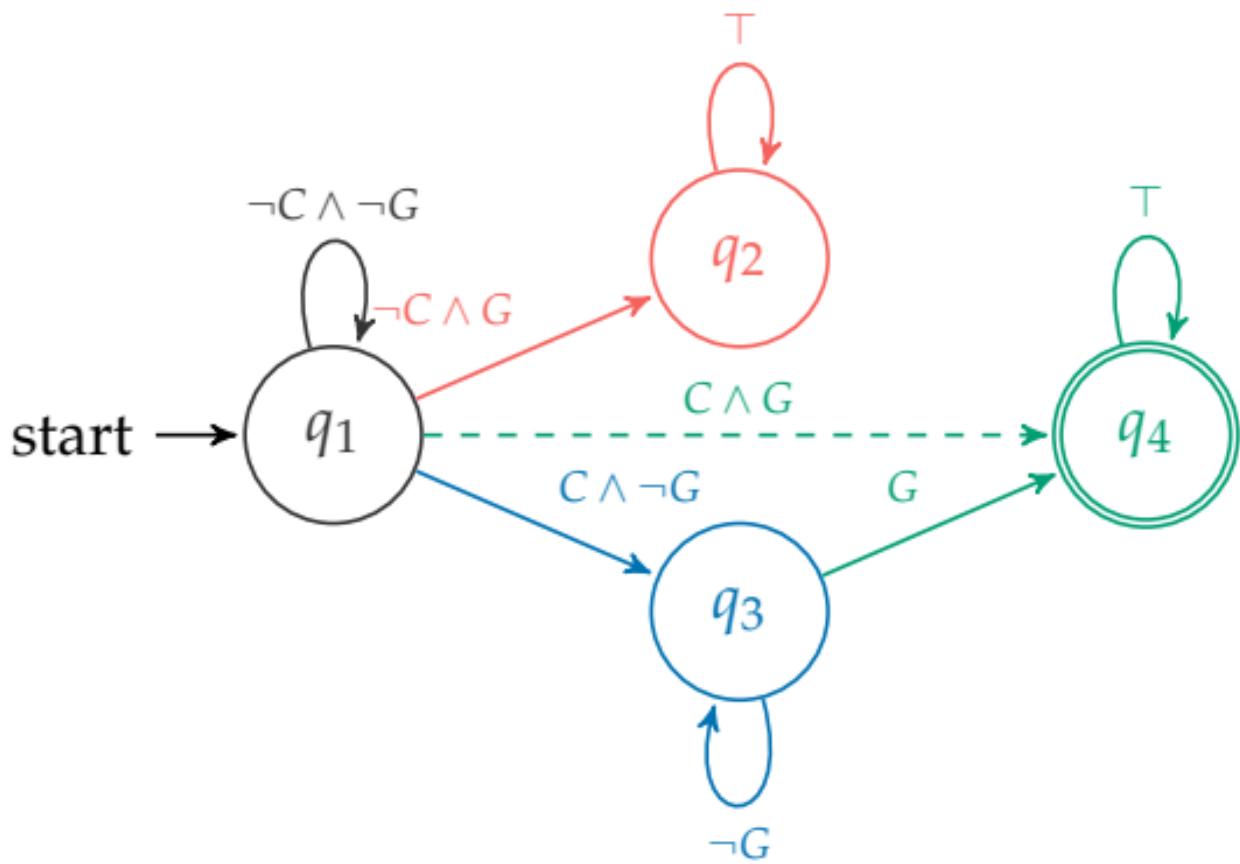


$\neg(A \wedge B)$  $\top$ 

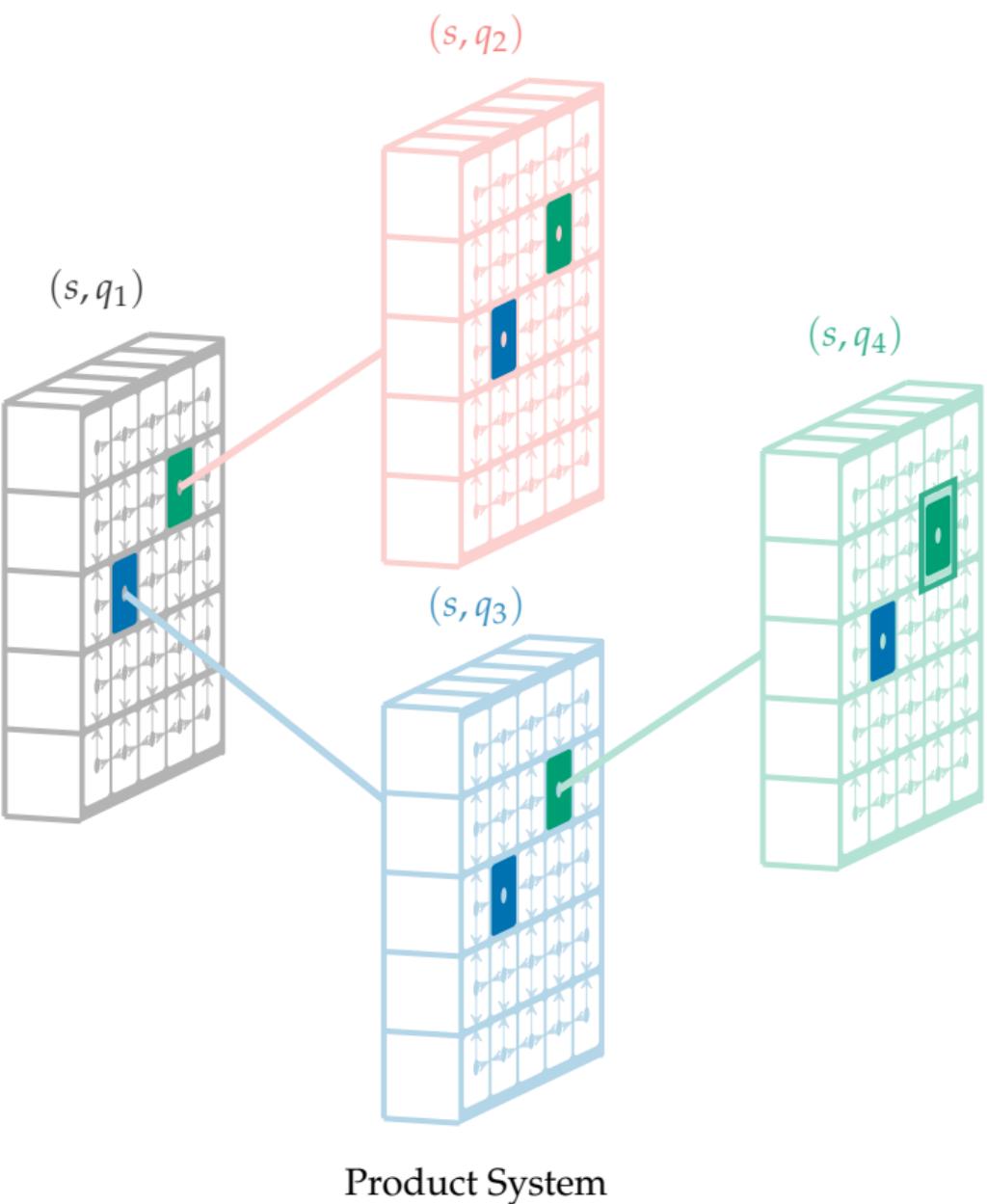


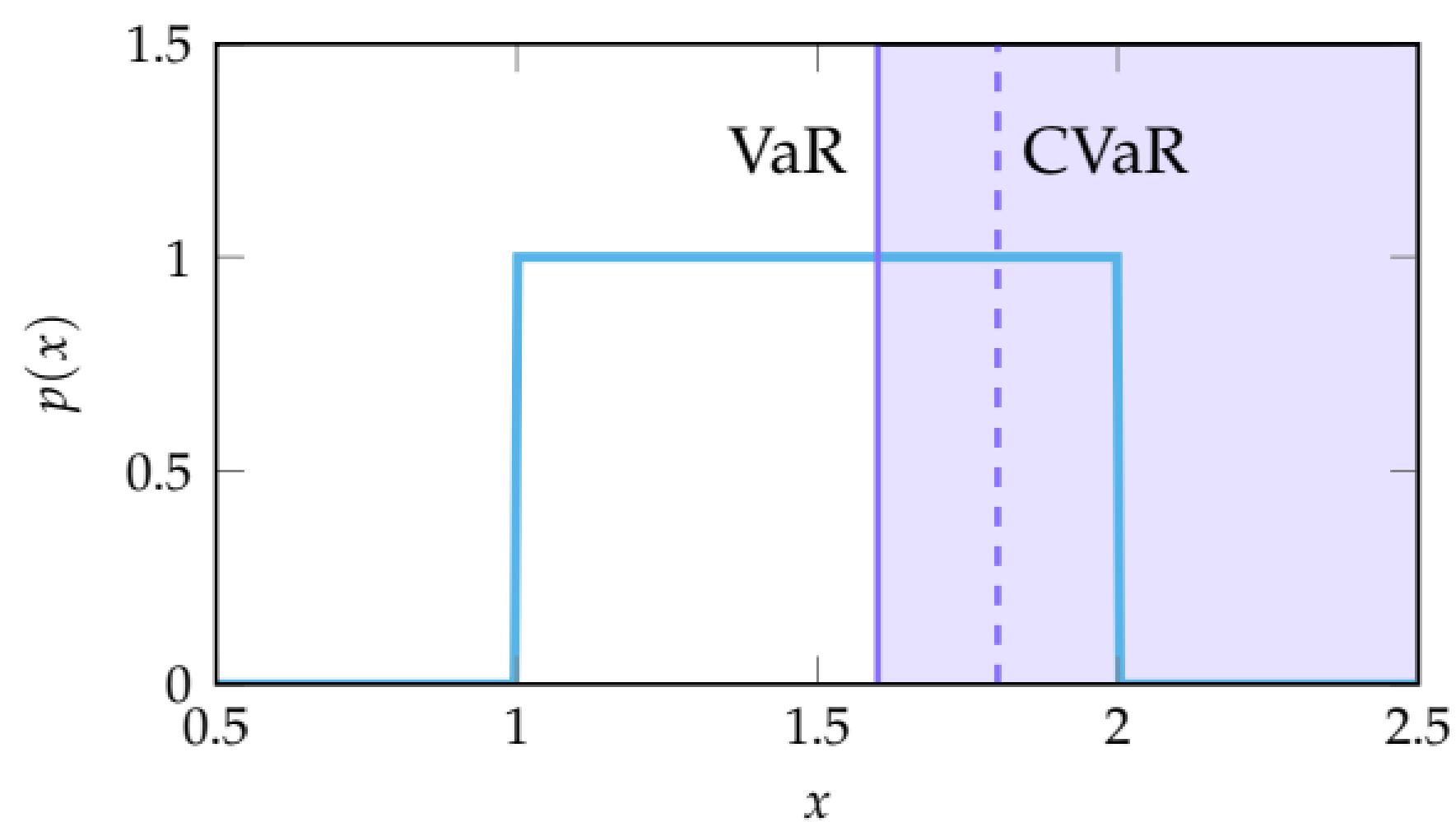


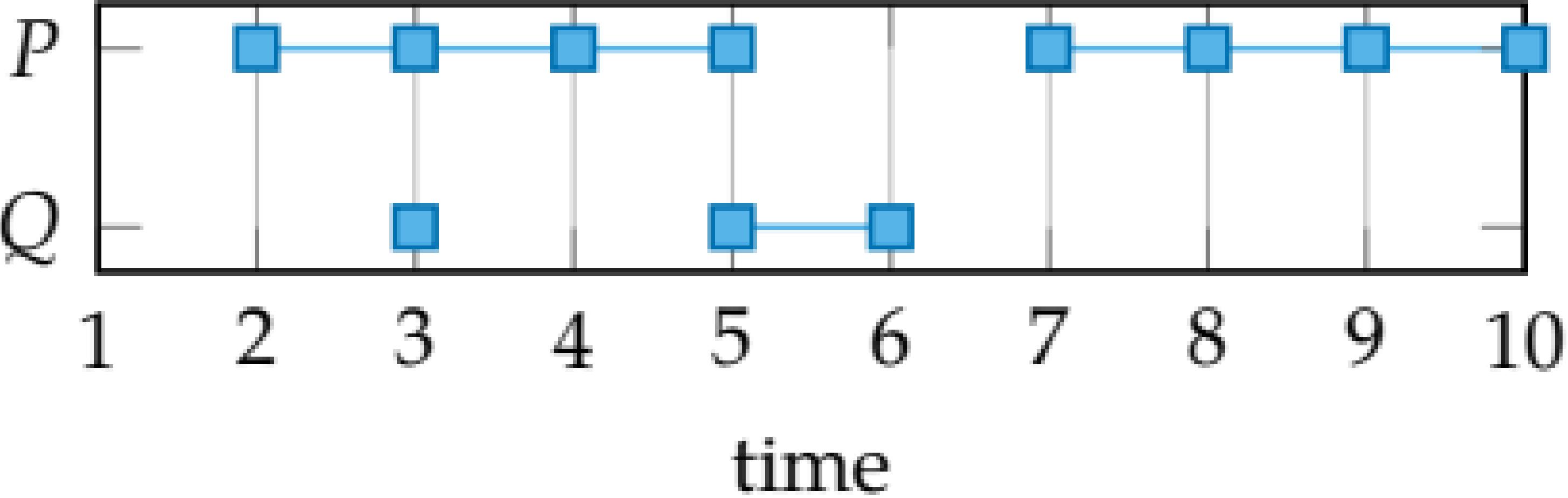
Original System

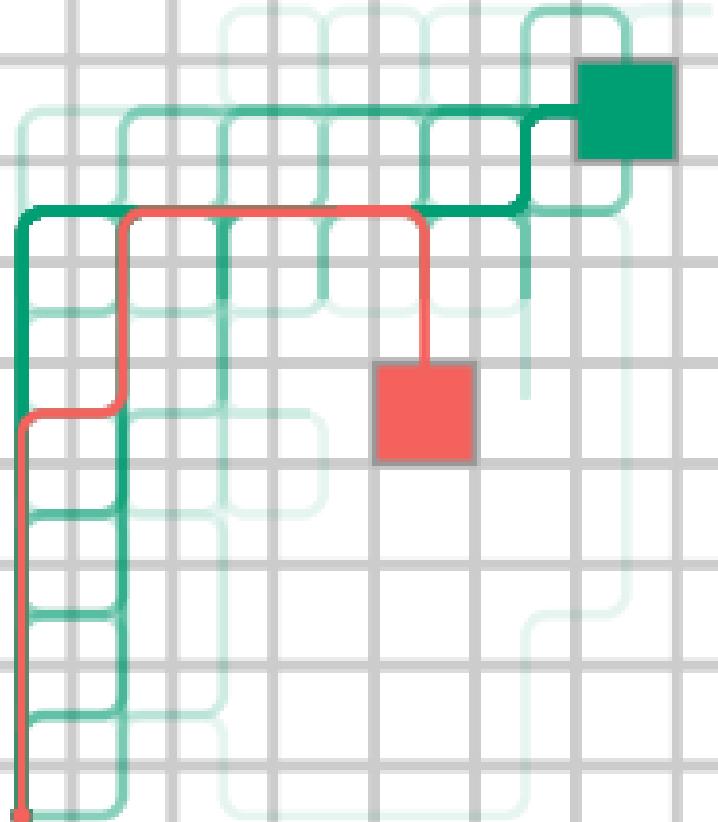


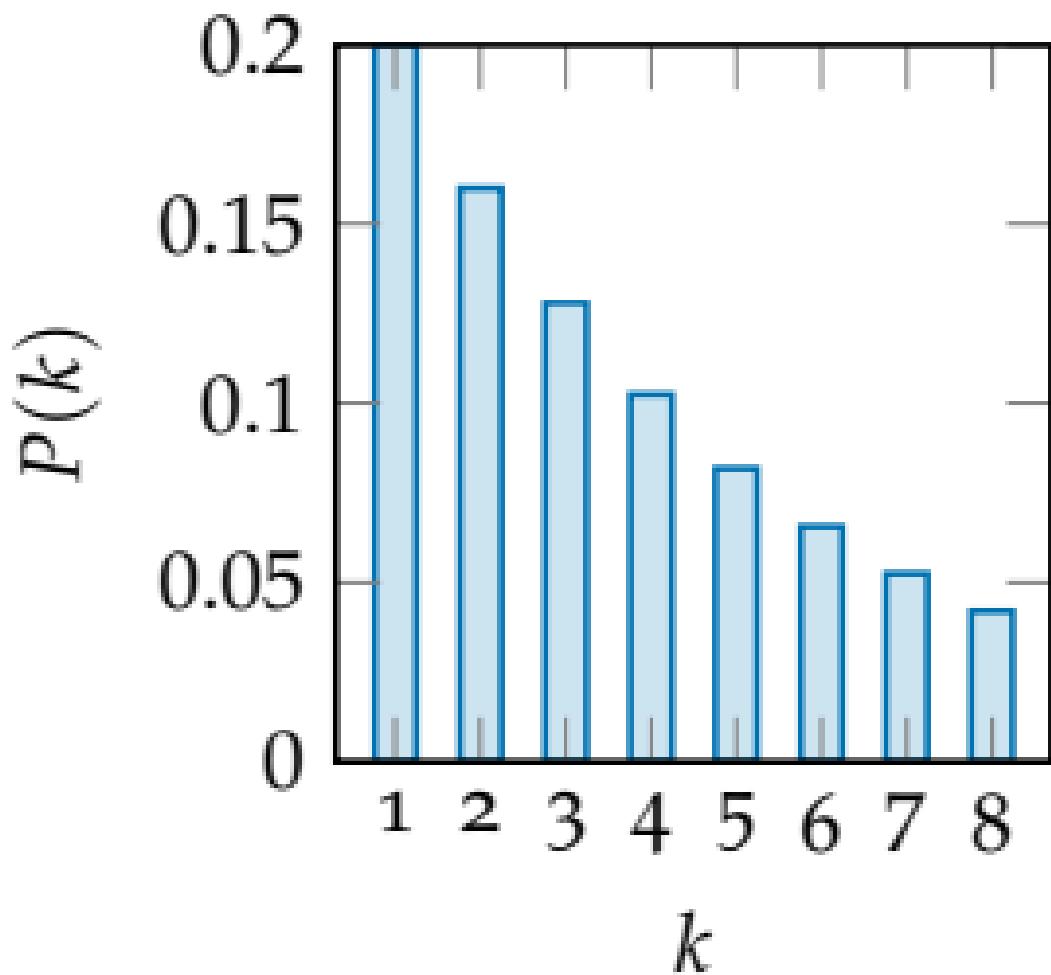
Büchi Automaton

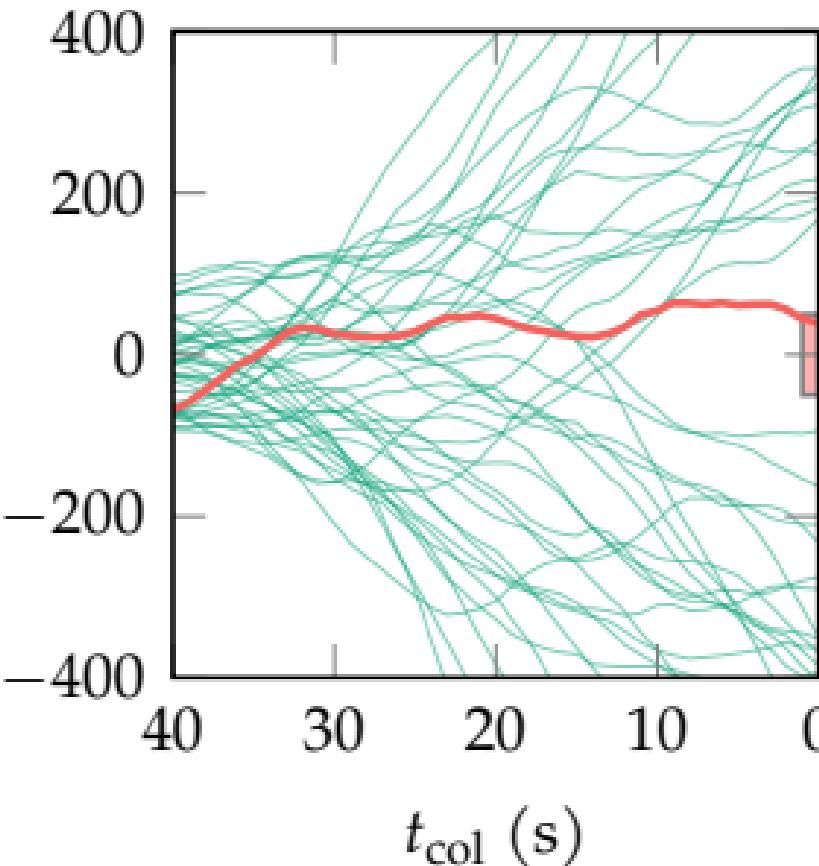
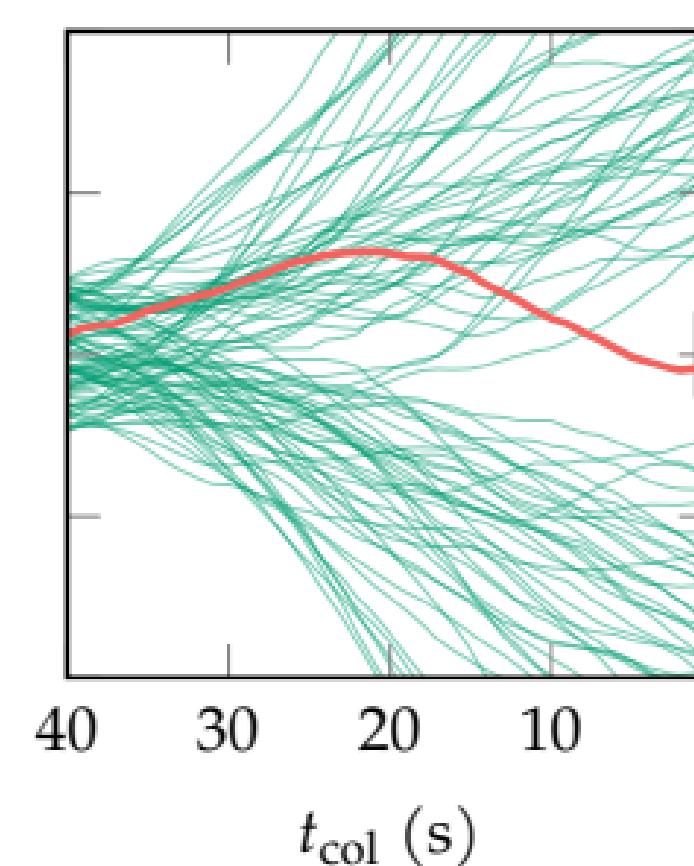
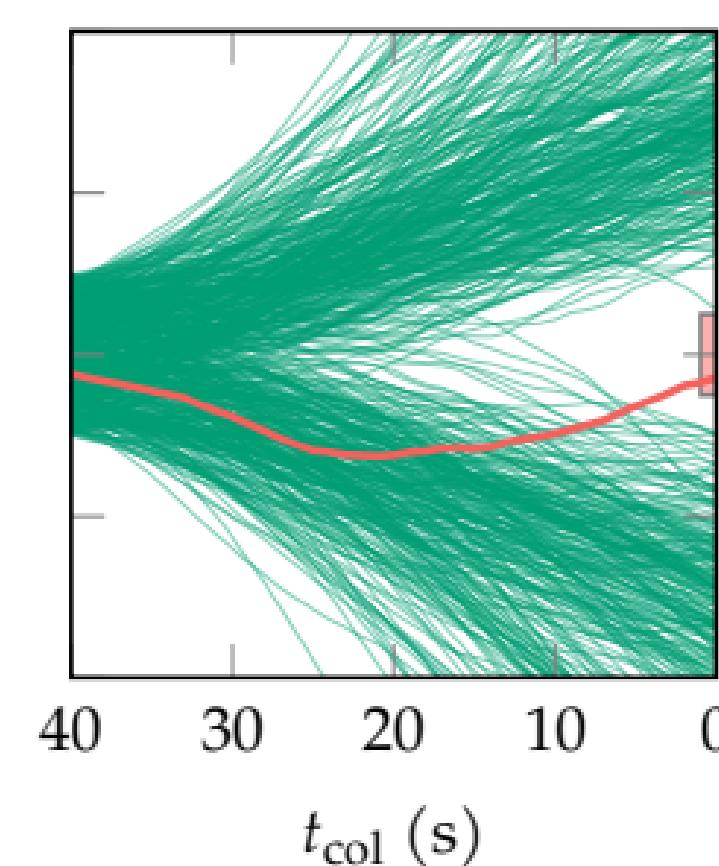








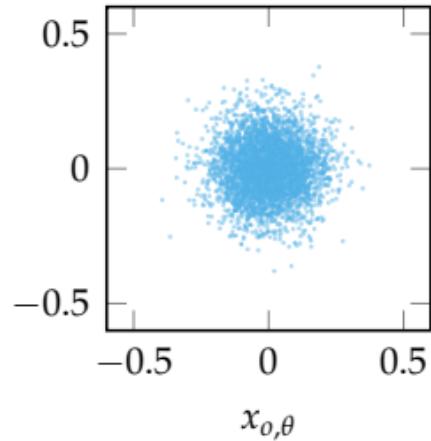


$\sigma = 5 \text{ m}$  $h \text{ (m)}$  $\sigma = 3 \text{ m}$  $\sigma = 2 \text{ m}$ 

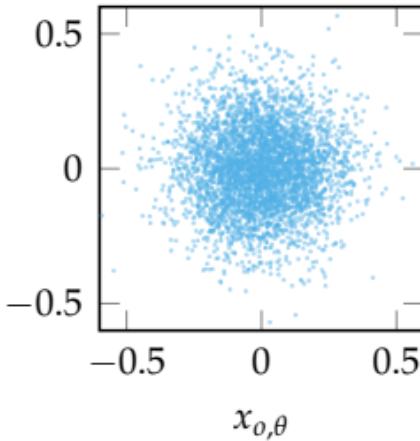
Nominal

Fuzzing

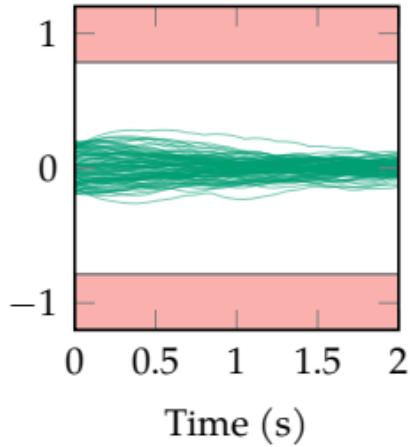
$x_{o,\omega}$



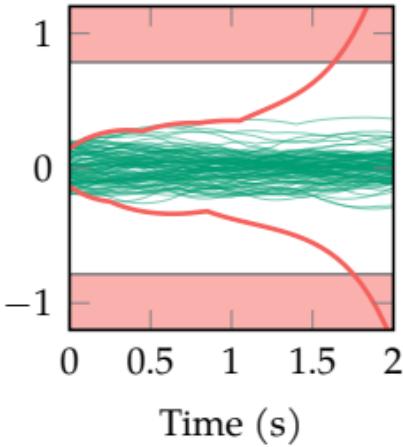
$x_{o,\omega}$



$\theta$  (rad)



$\theta$  (rad)



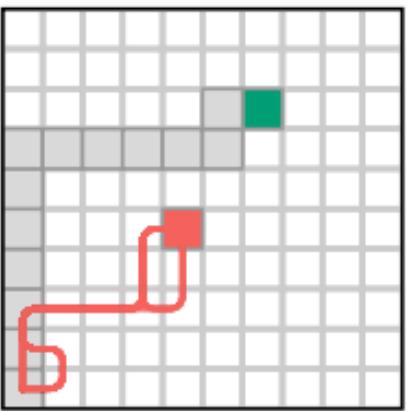
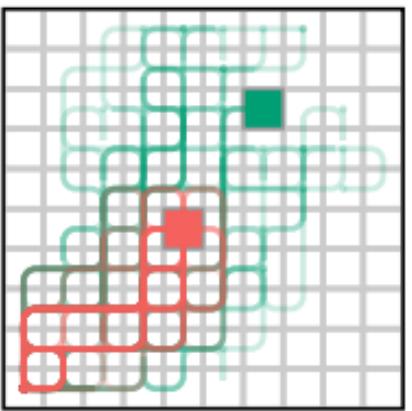
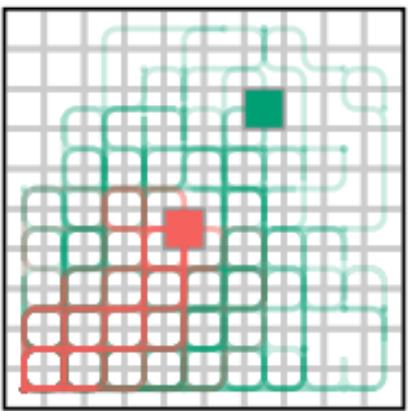
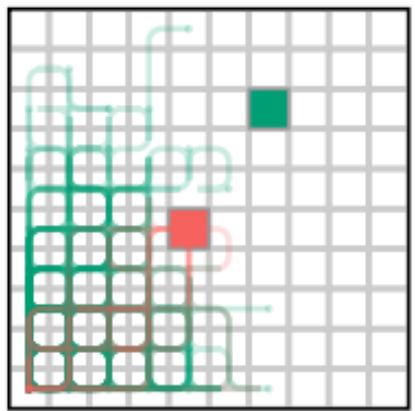
Iteration 1

Iteration 5

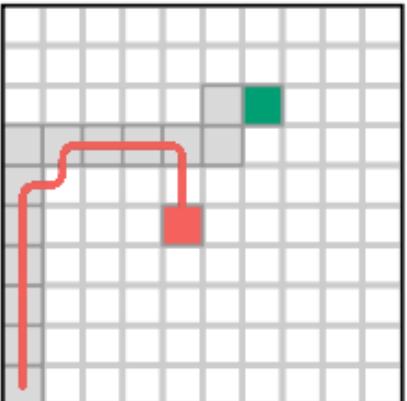
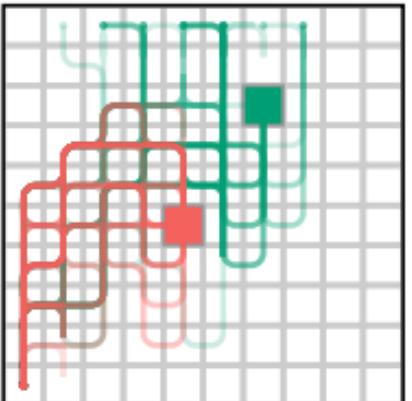
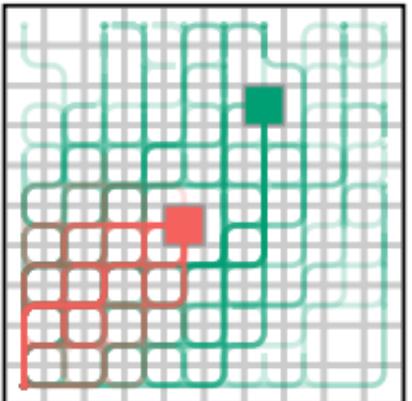
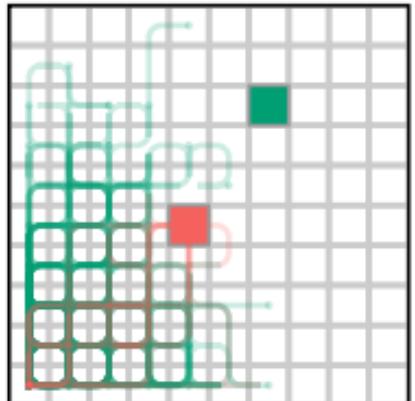
Iteration 8

Converged

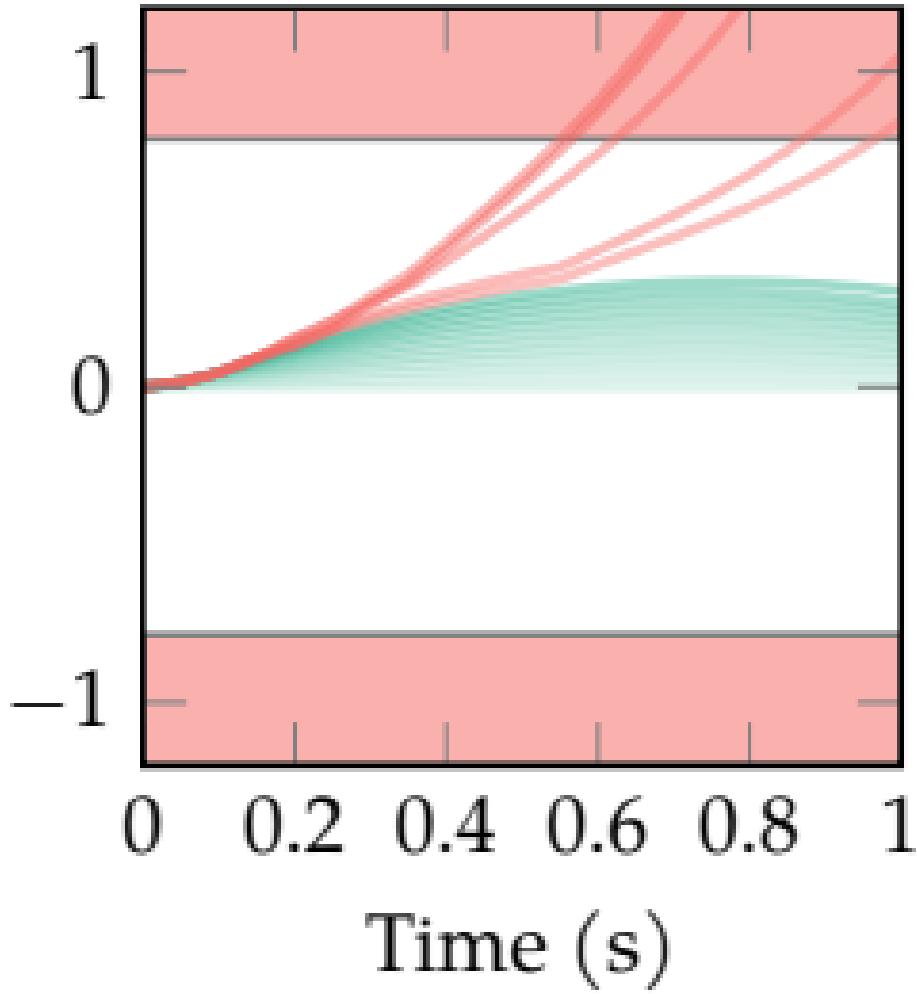
Robustness

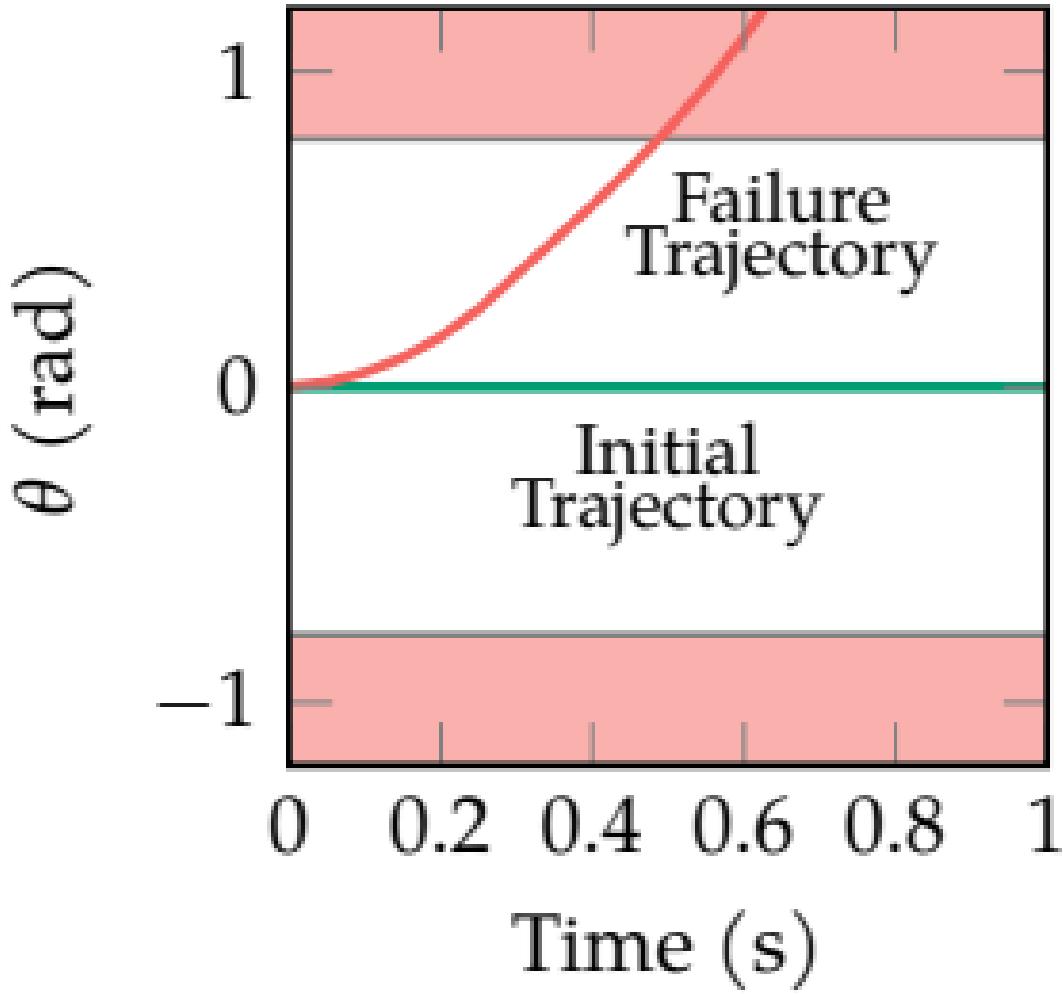


Likelihood

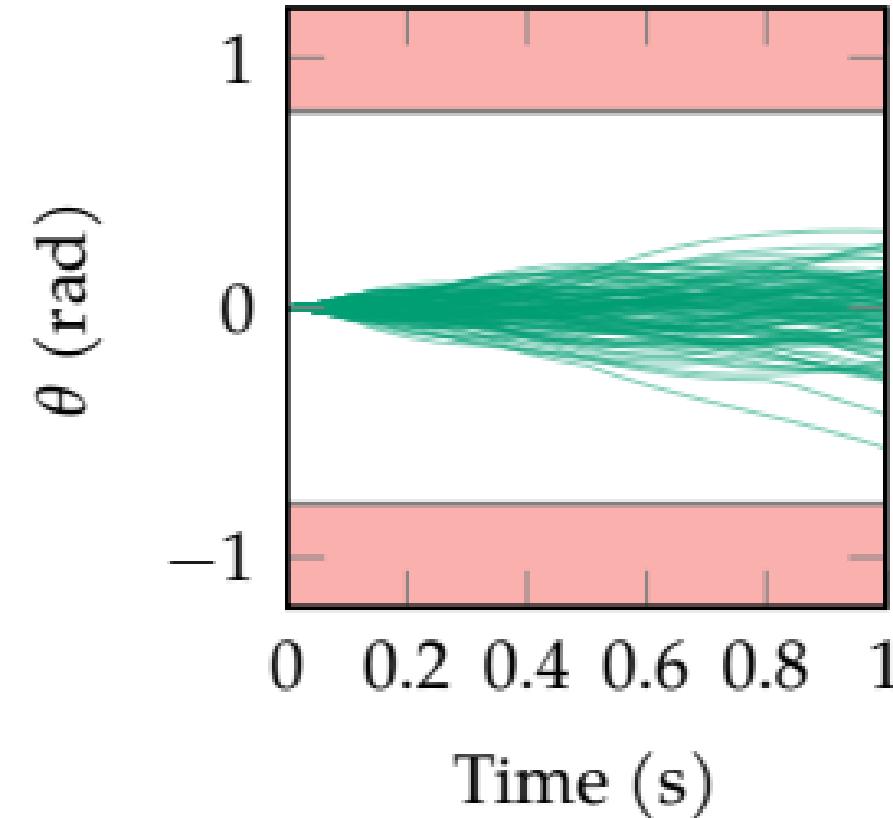


$\theta$  (rad)

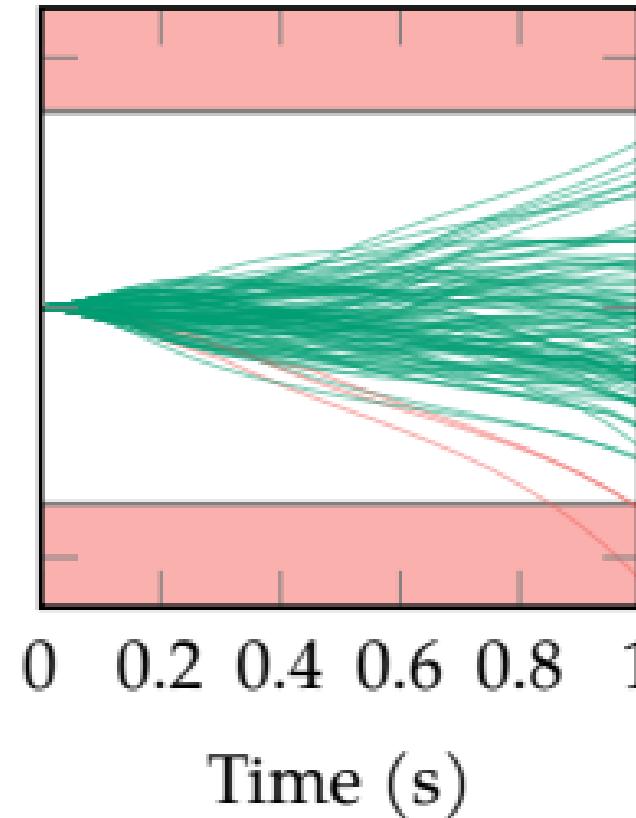




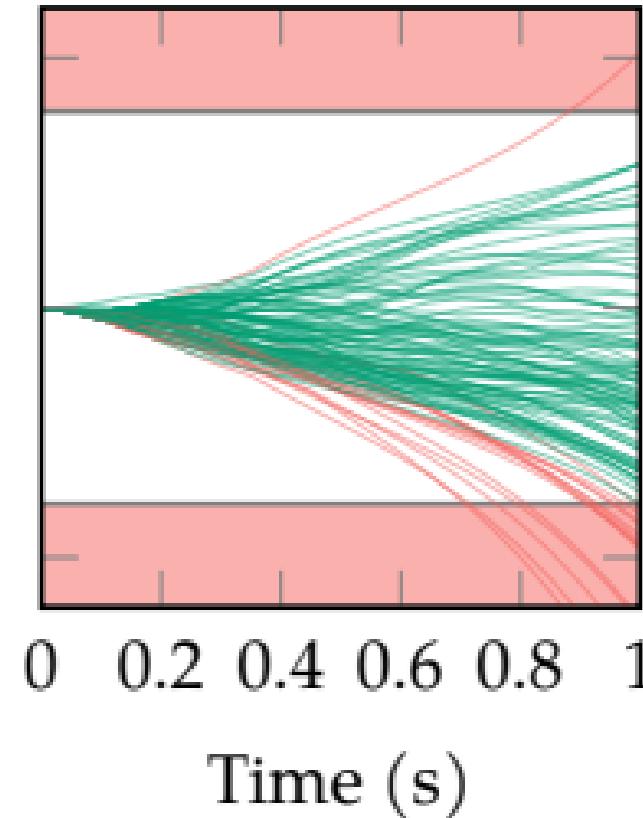
# Iteration 1



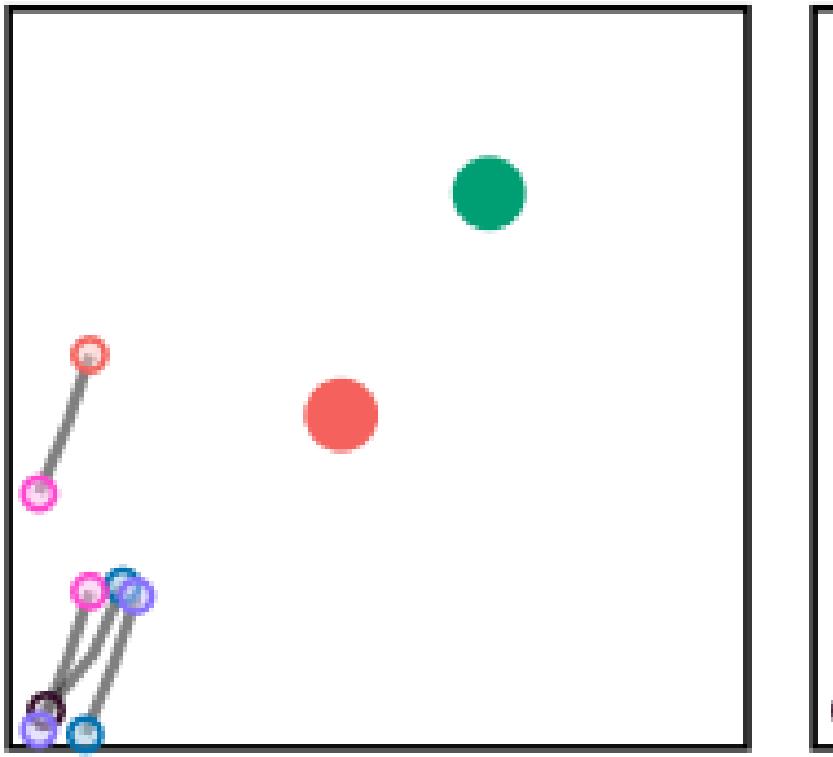
# Iteration 5



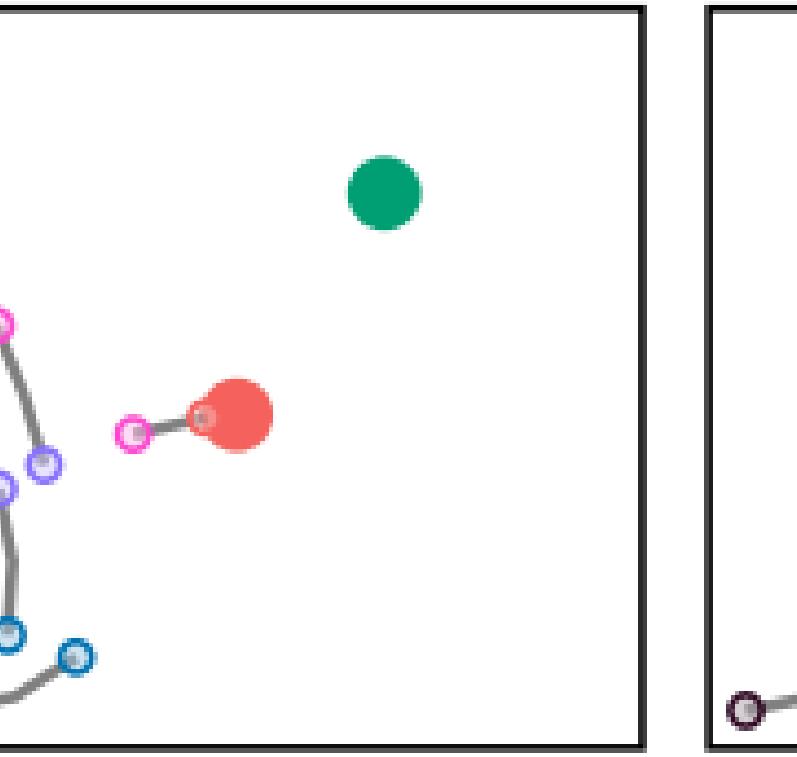
# Iteration 10



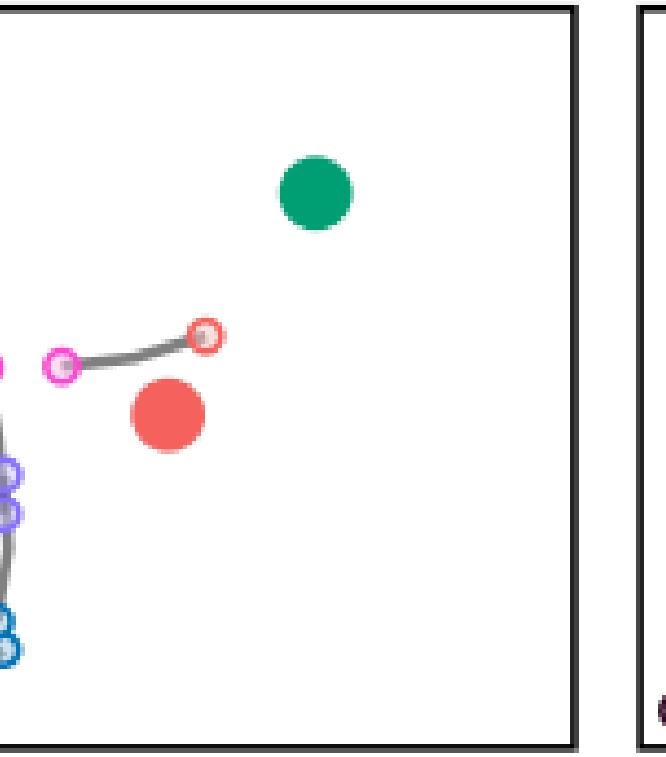
Iteration 2



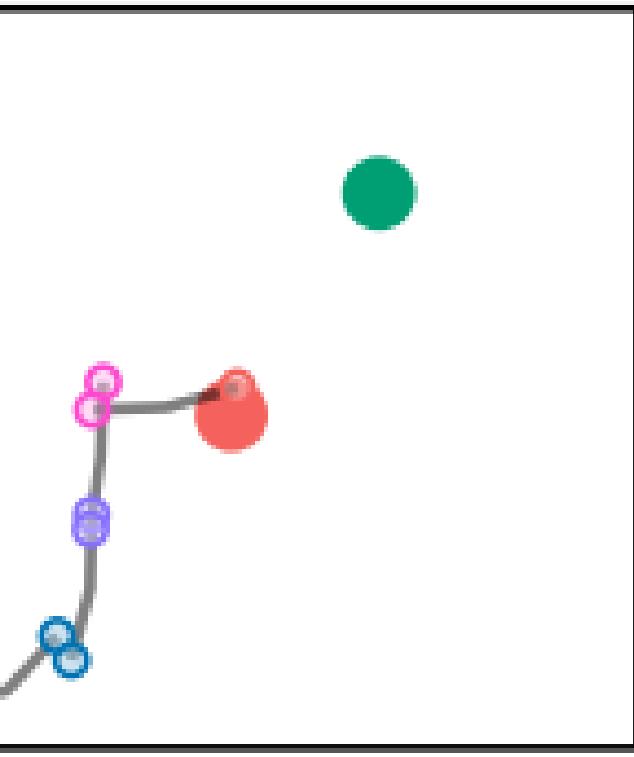
Iteration 4



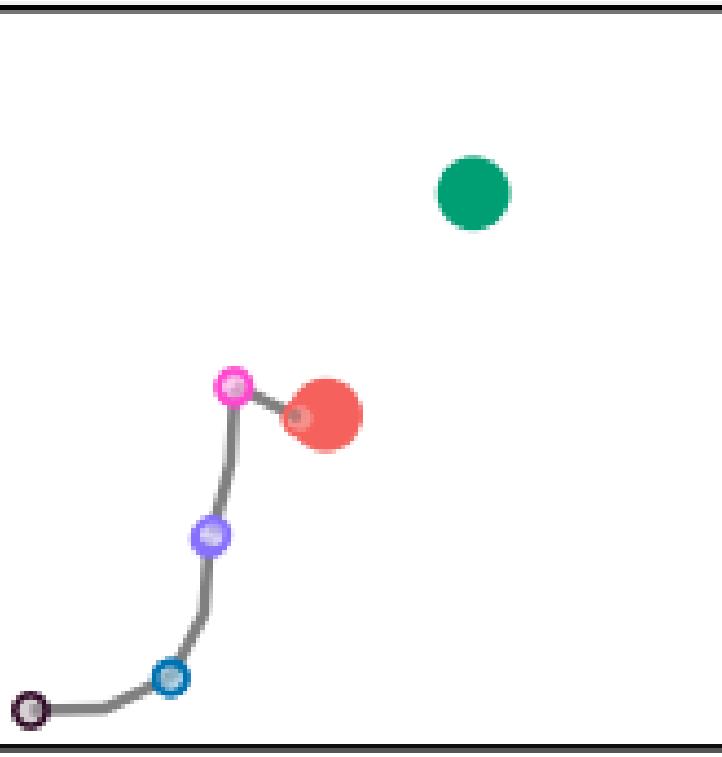
Iteration 6



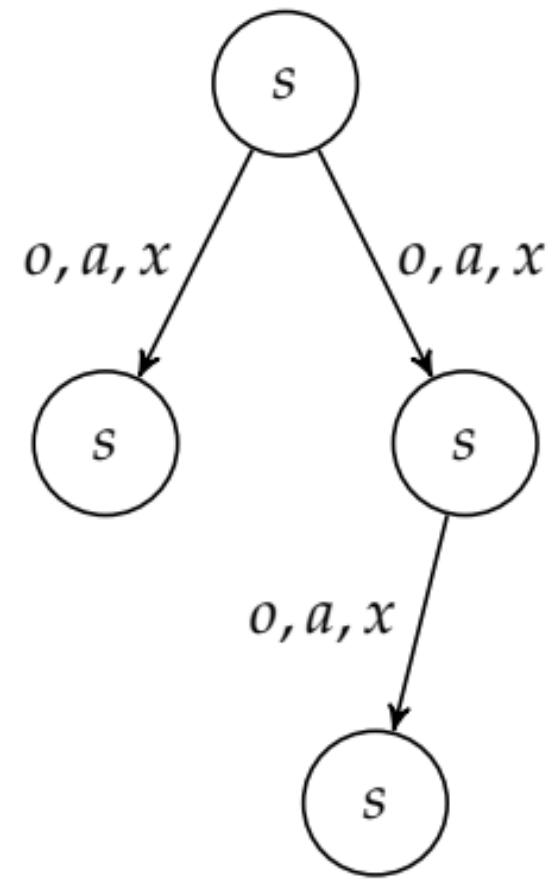
Iteration 8



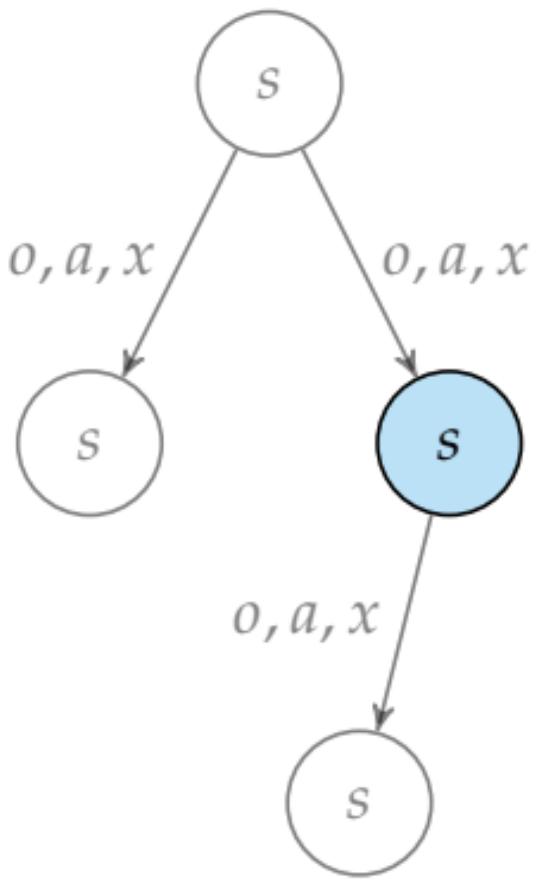
Converged



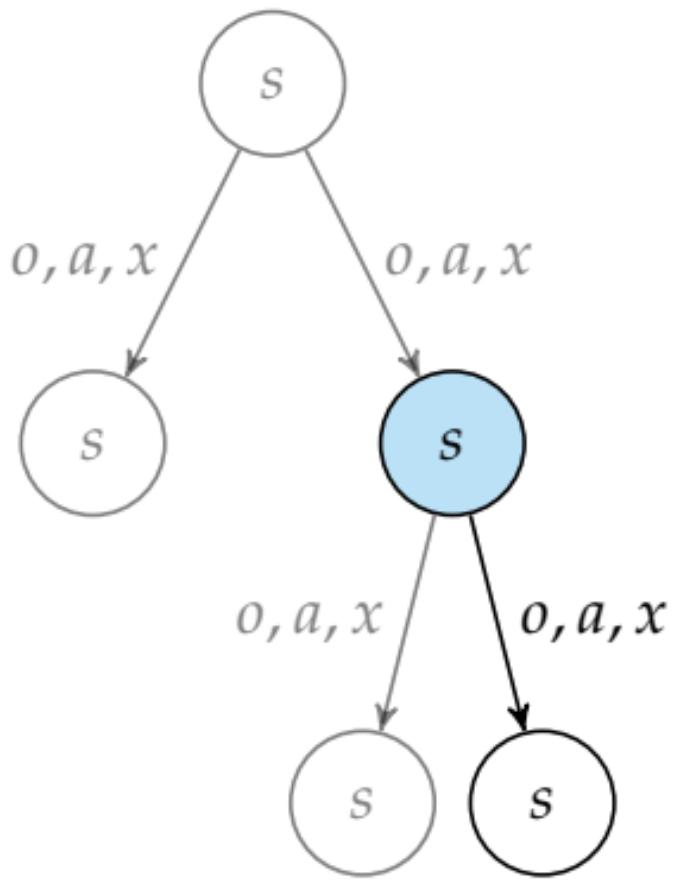
Current Tree



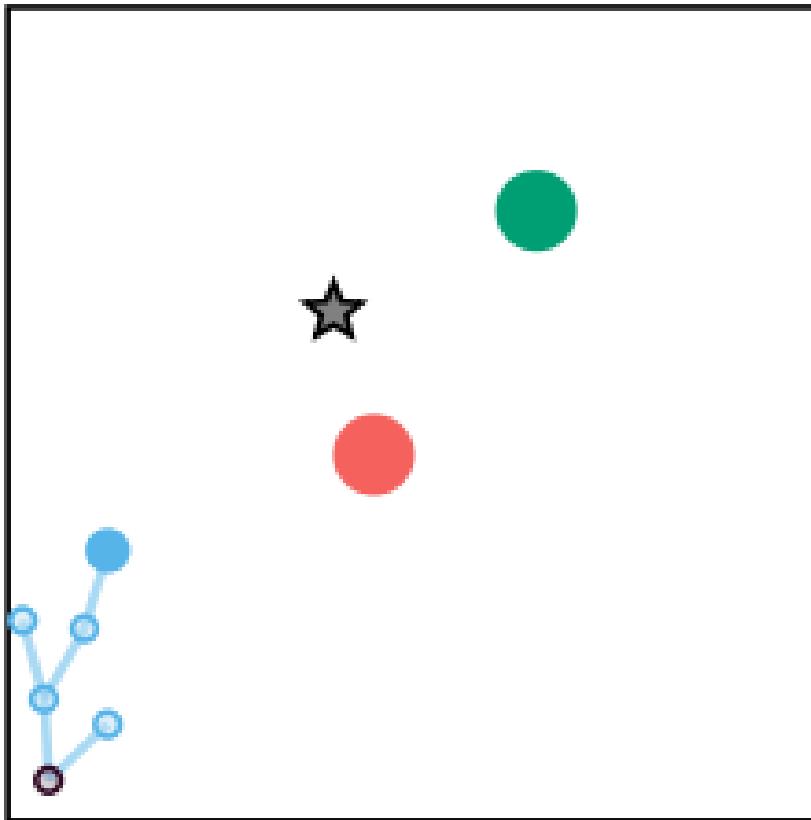
Select



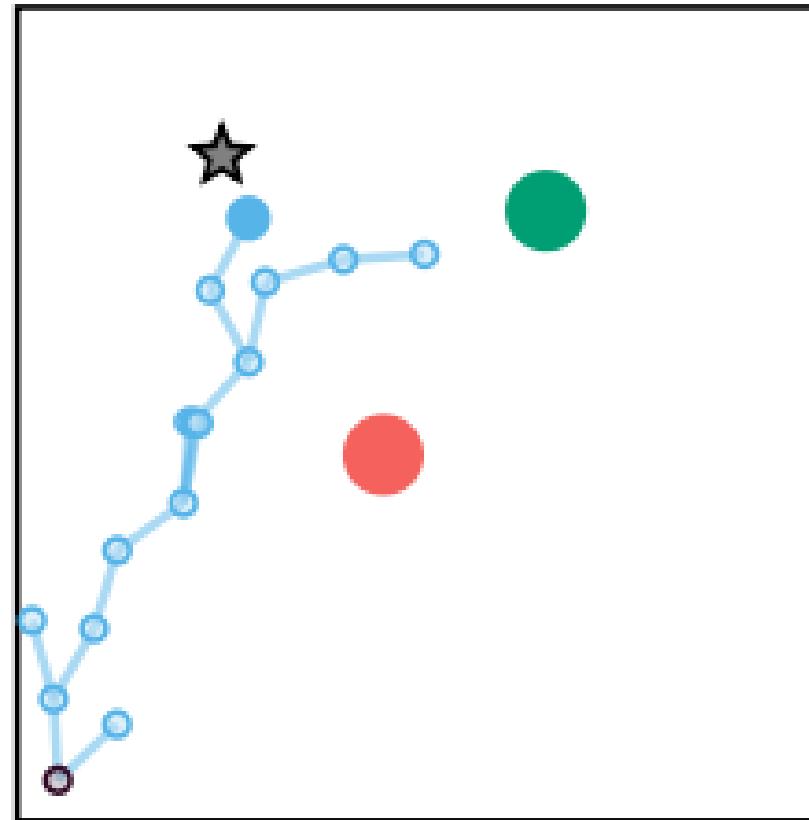
Extend



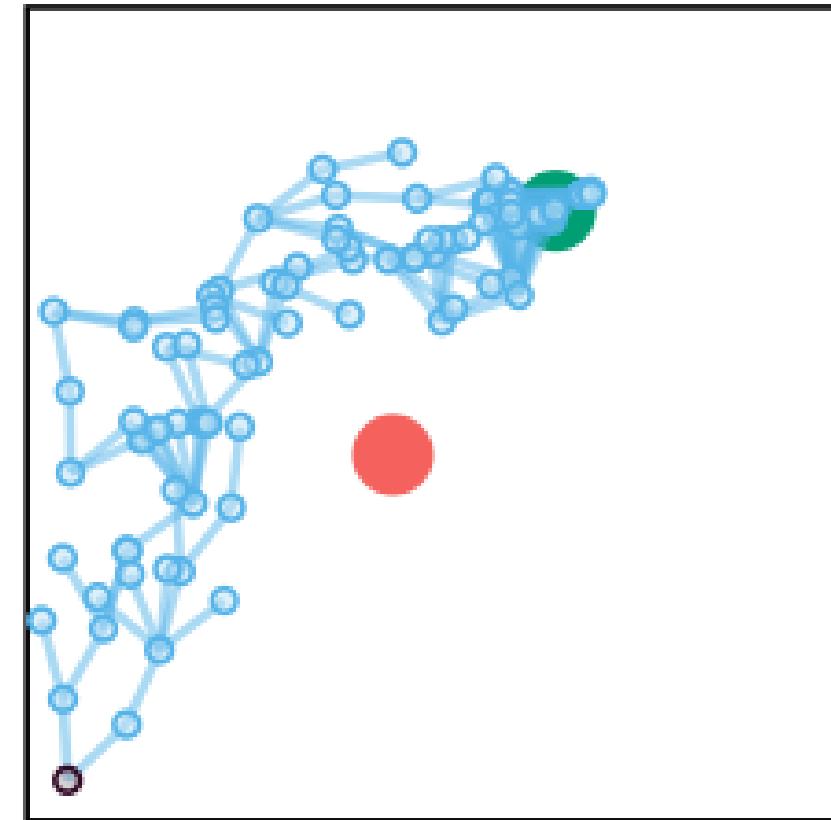
# Iteration 5



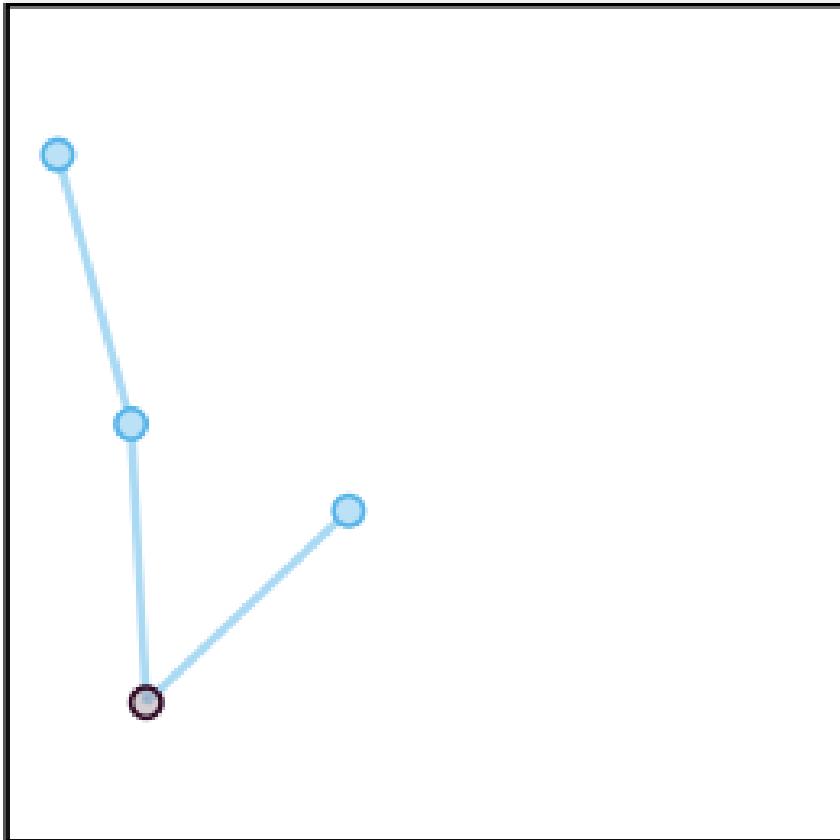
# Iteration 15



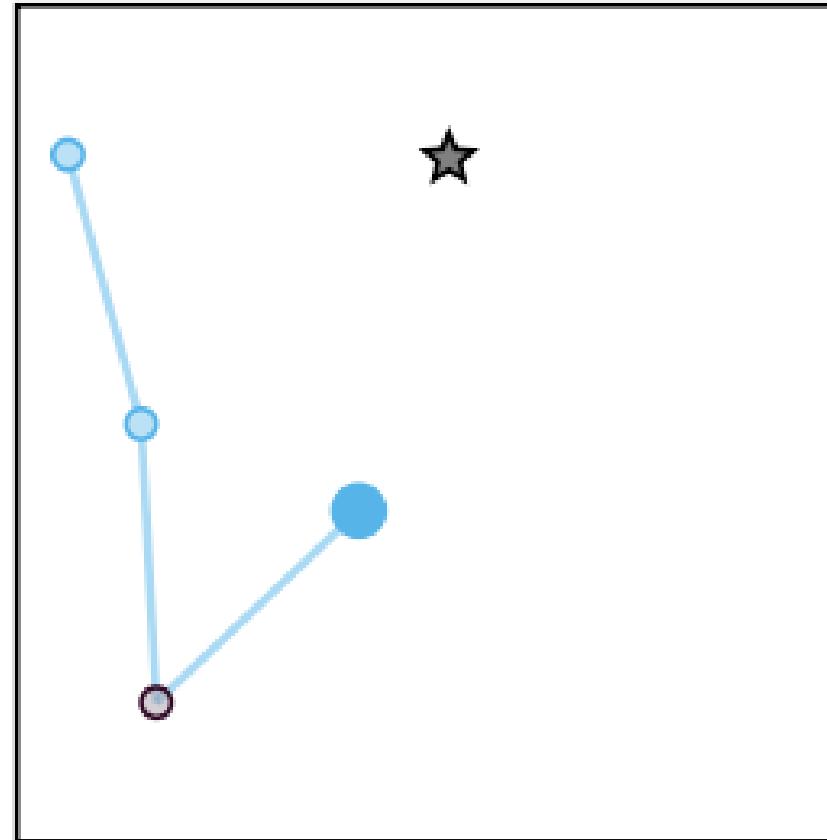
# Iteration 100



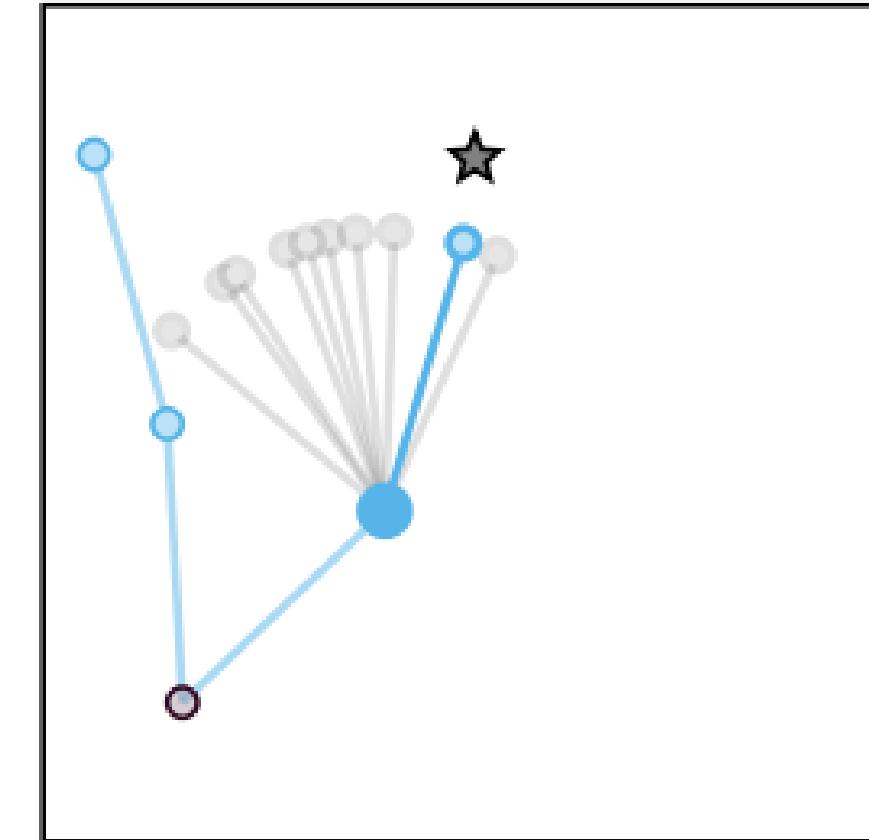
# Current Tree

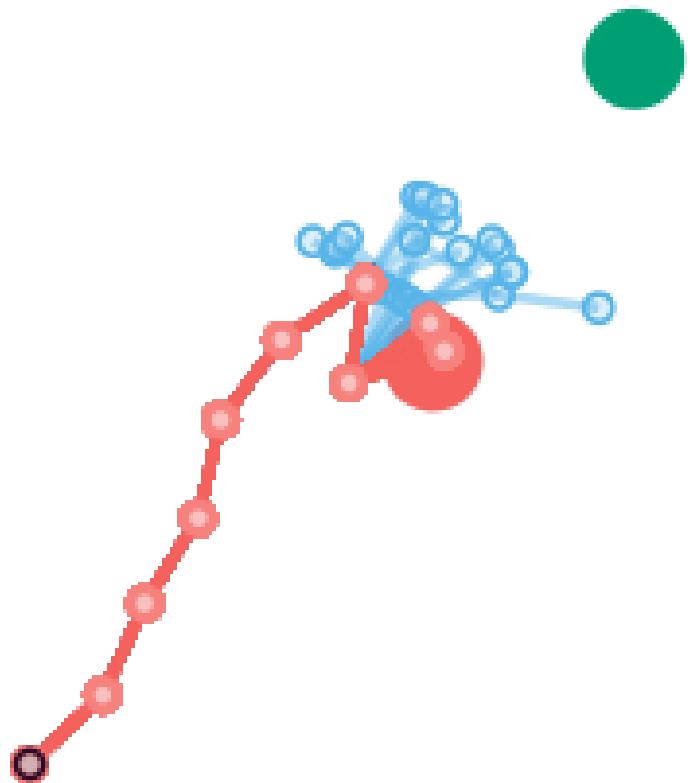


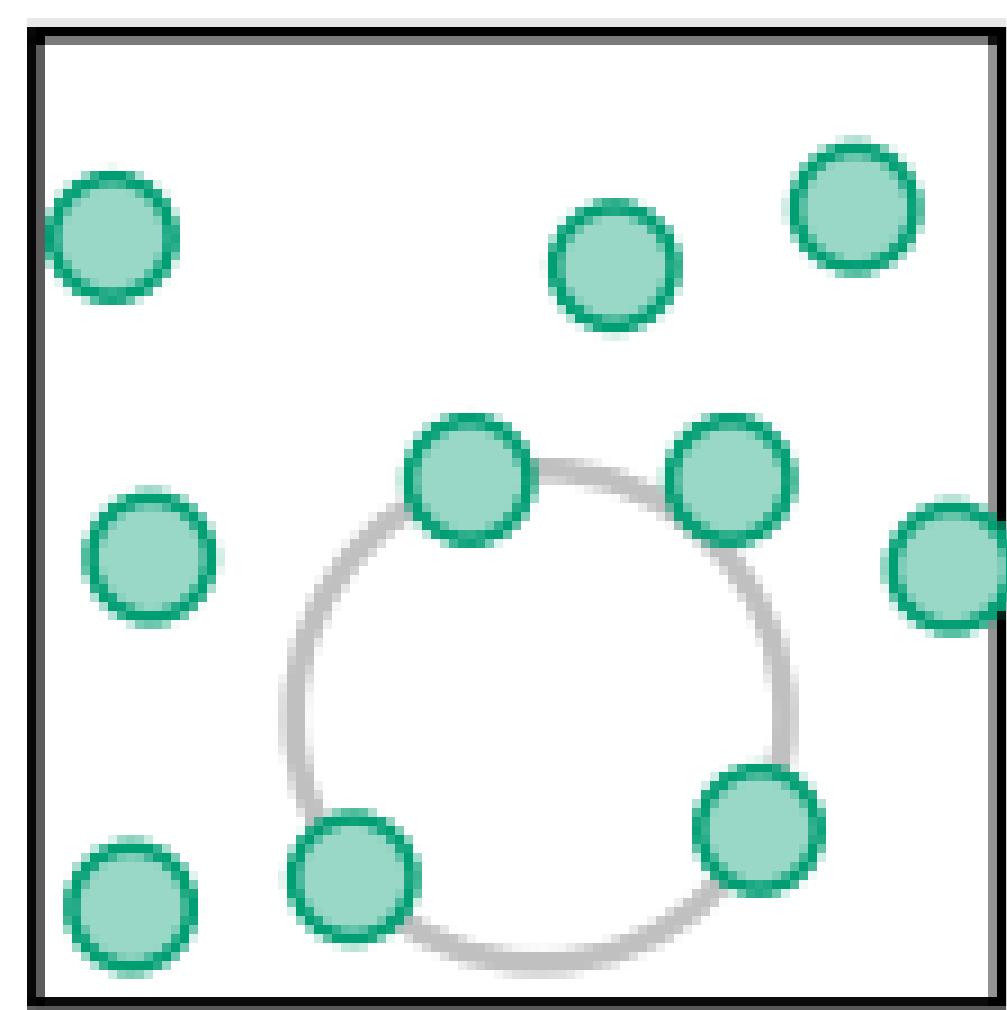
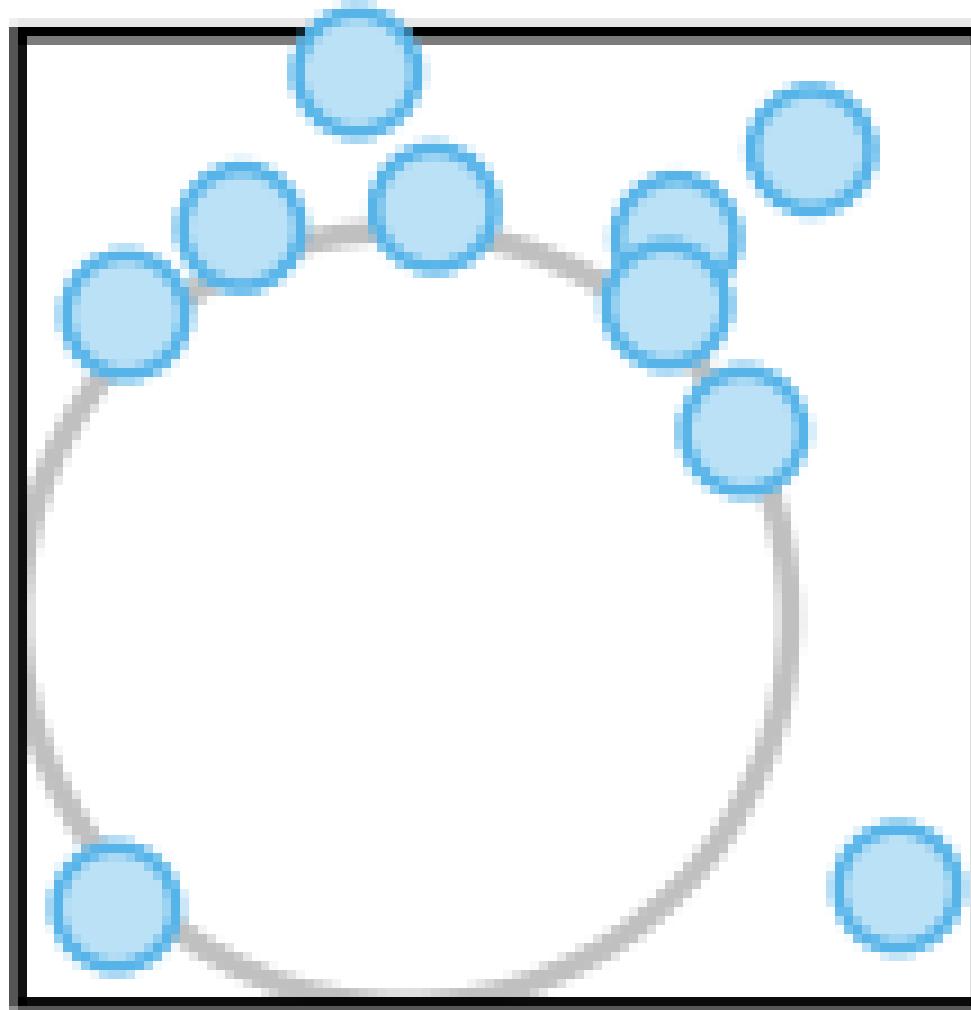
# Select



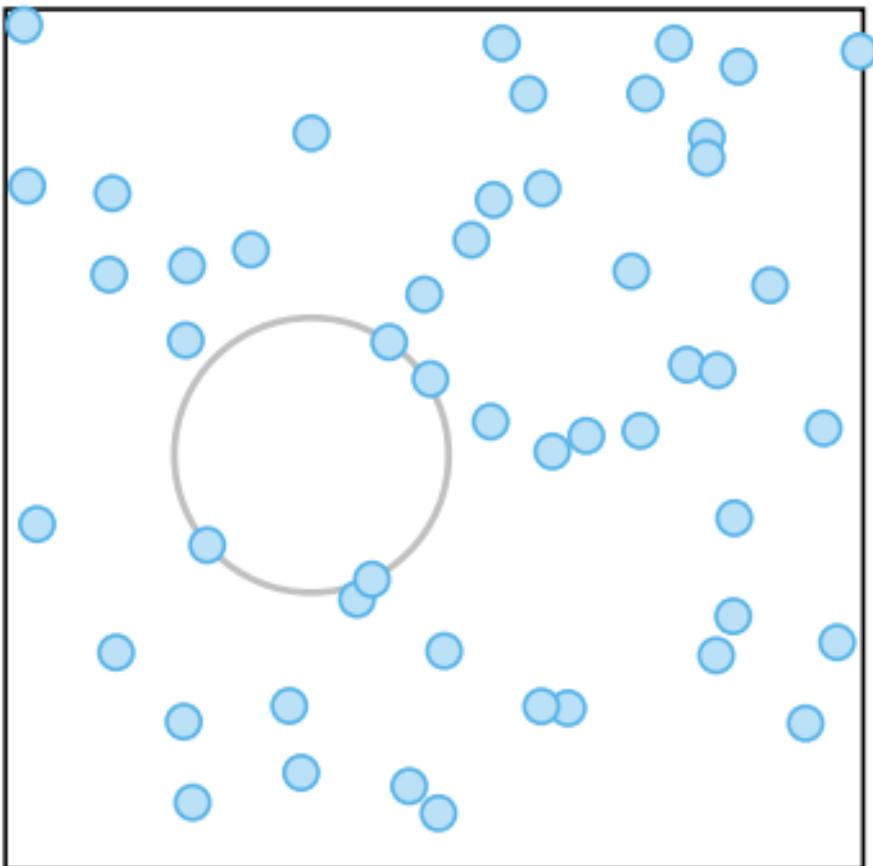
# Extend



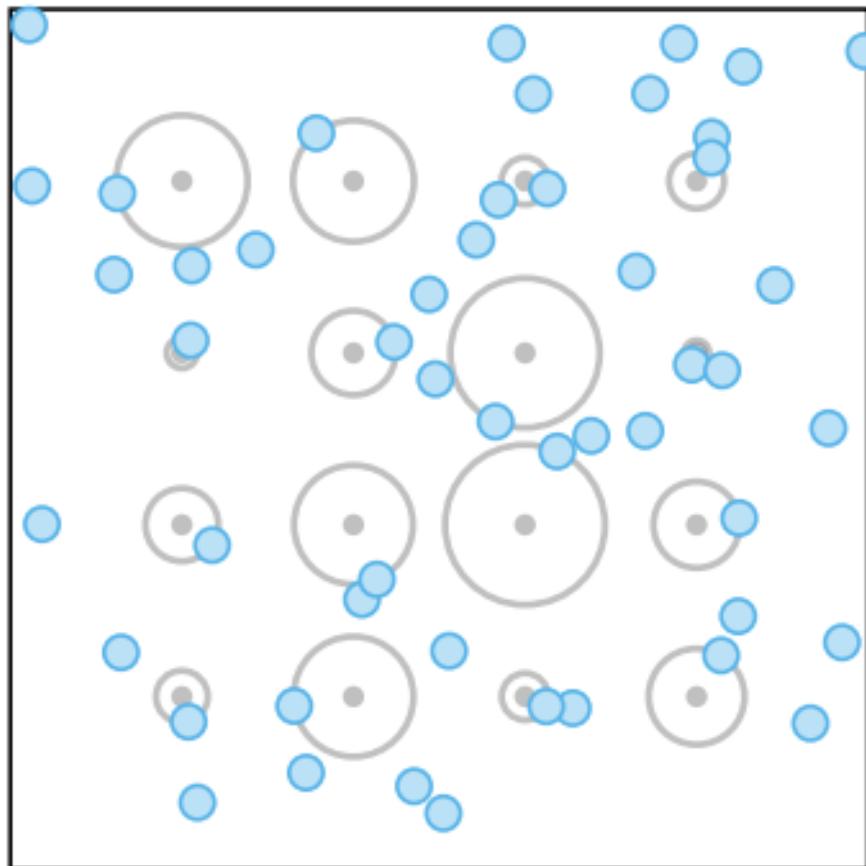


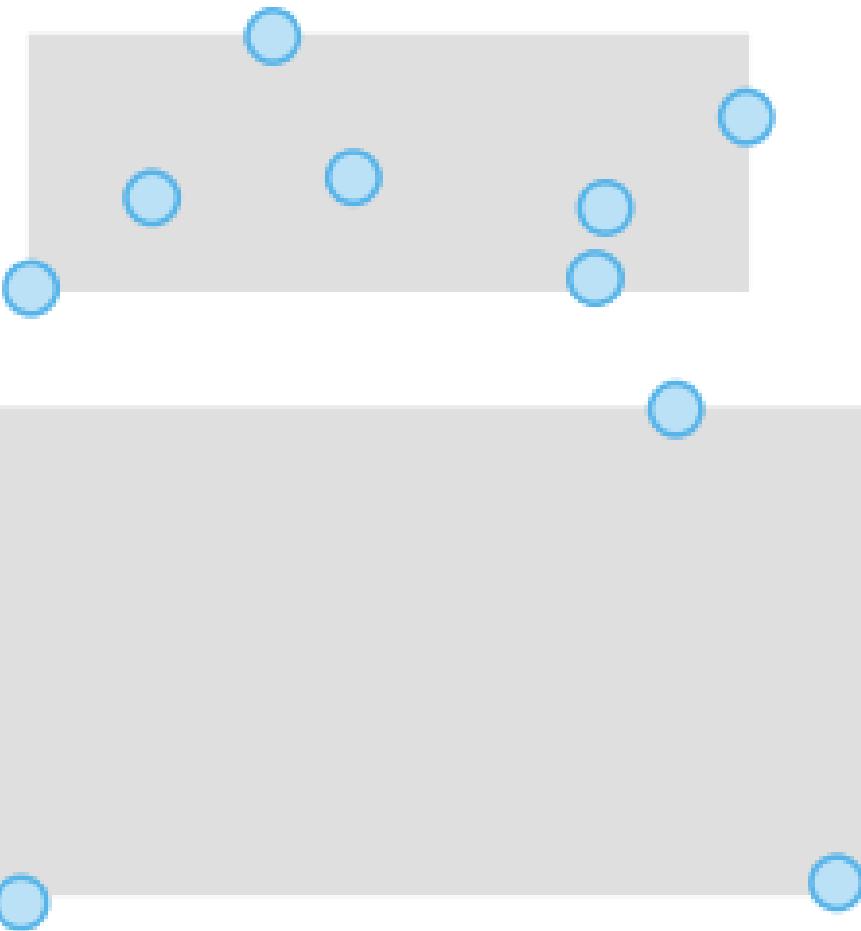


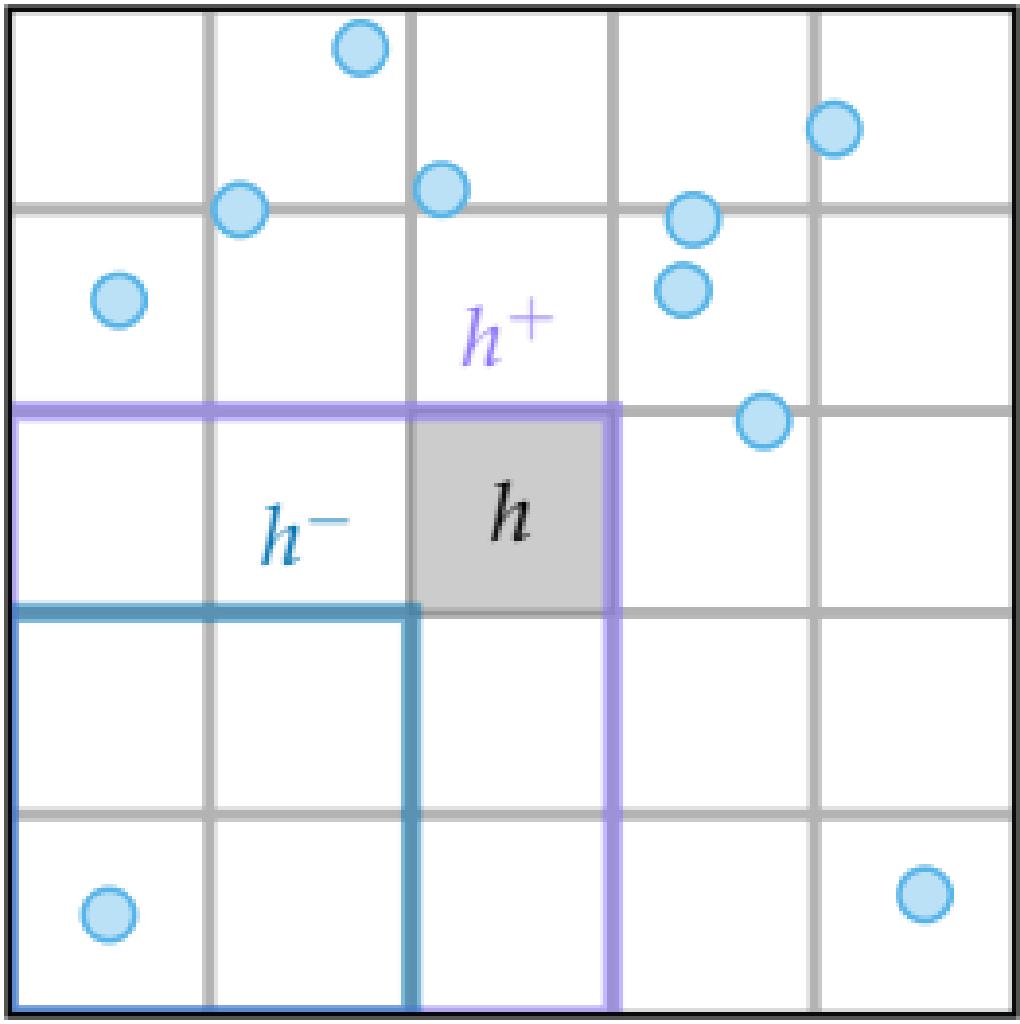
# Dispersion



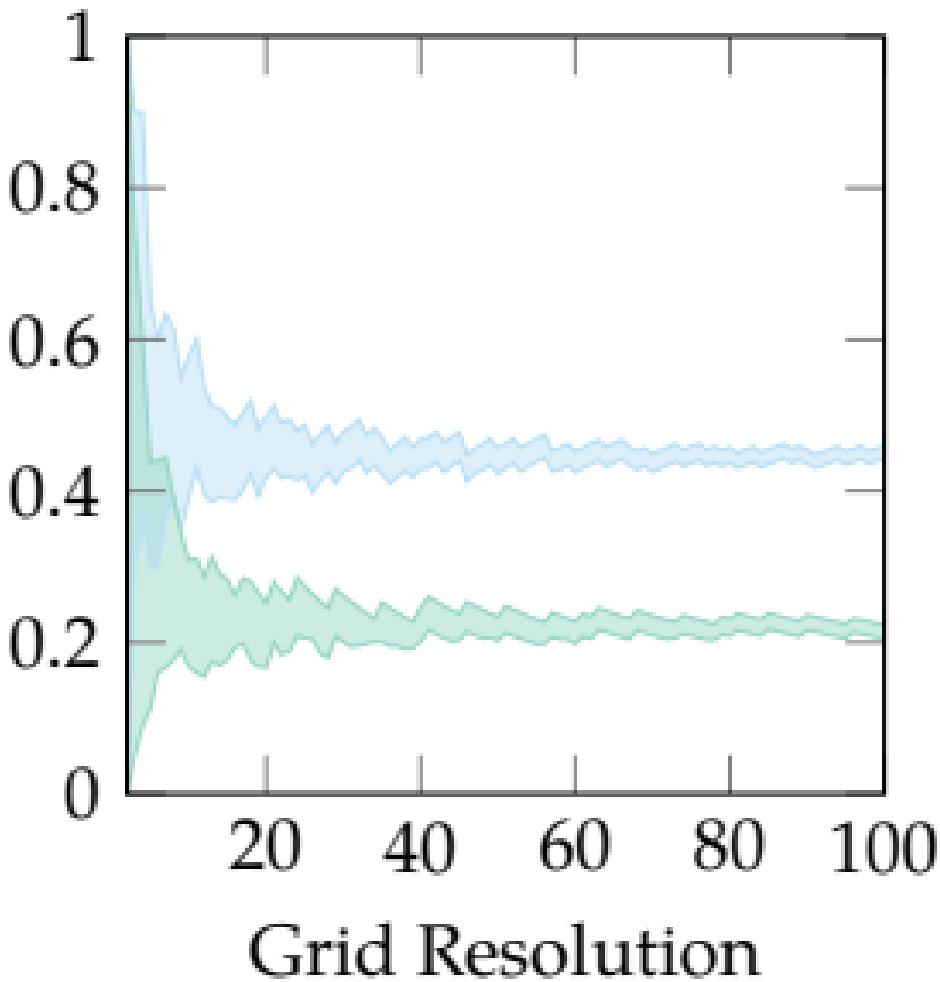
# Average Dispersion



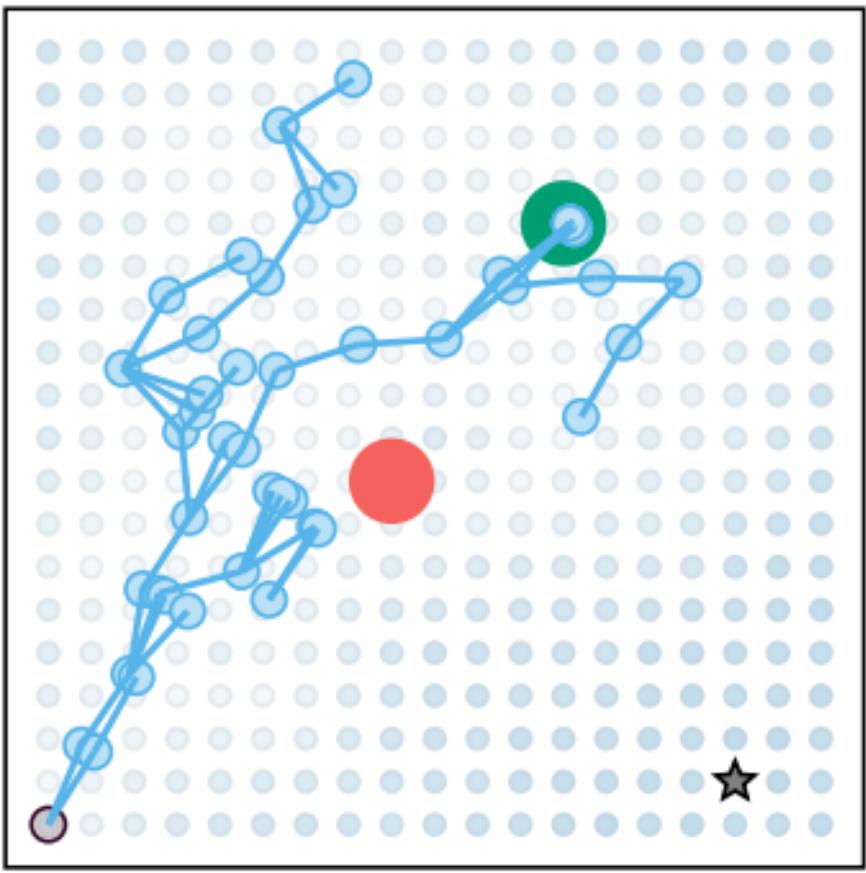




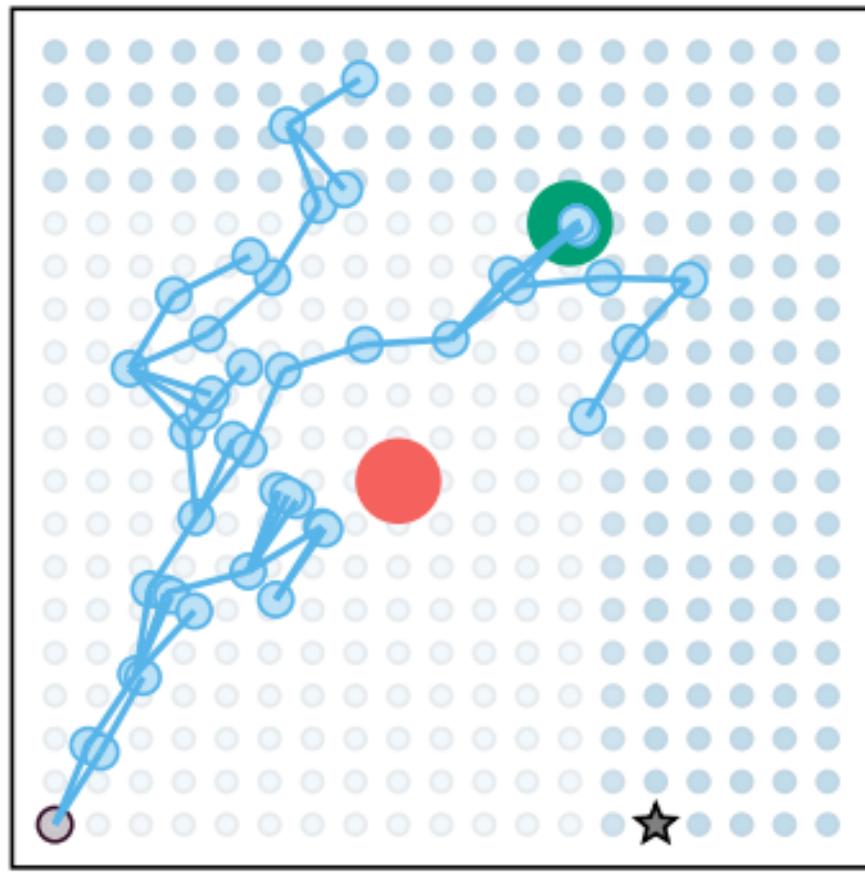
Star Discrepancy



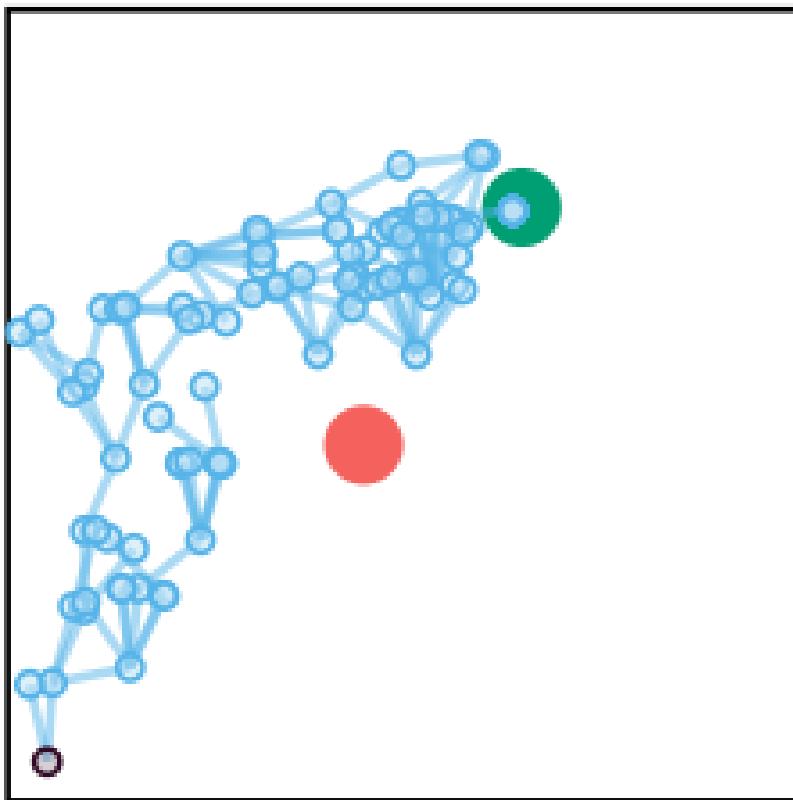
# Average Dispersion



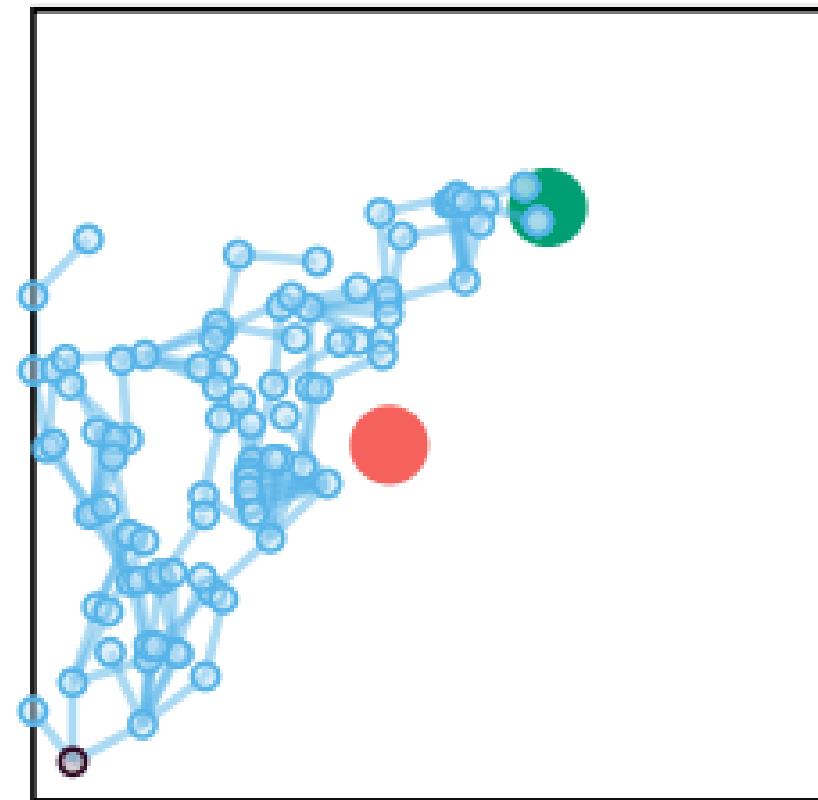
# Star Discrepancy

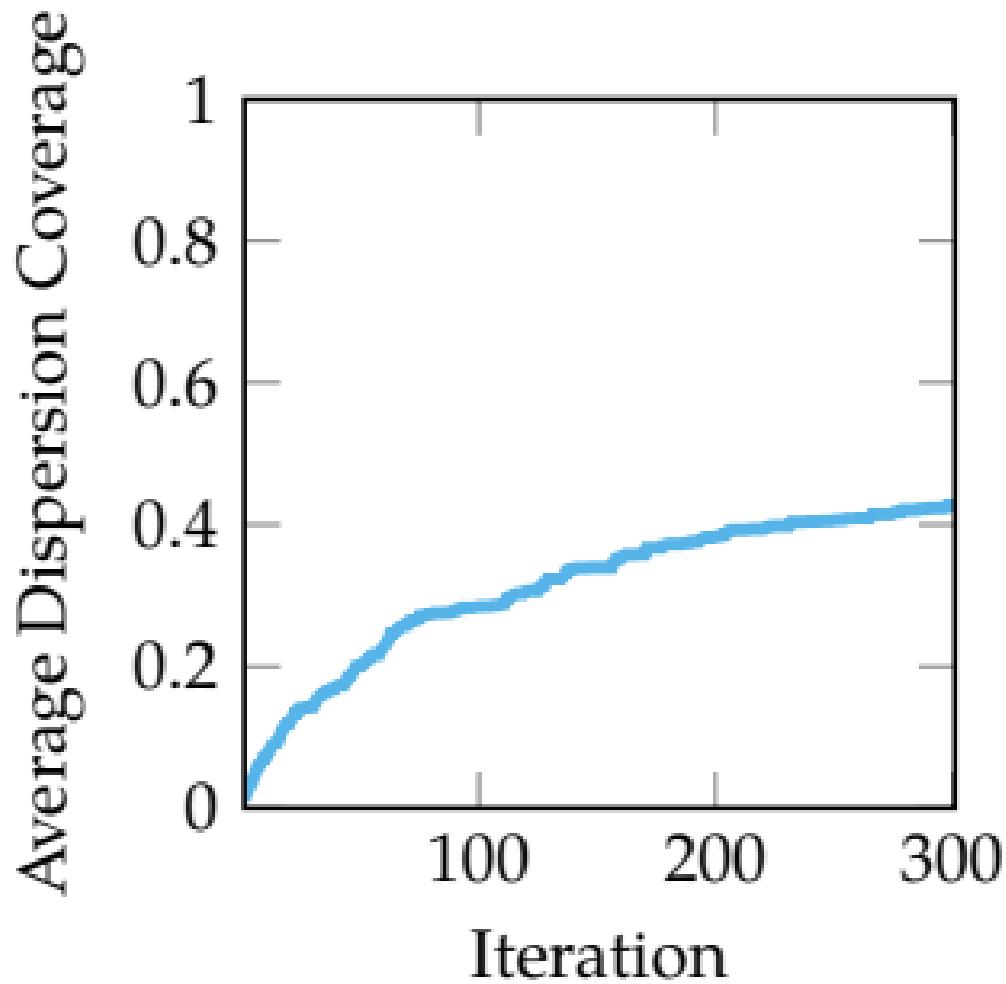


# Random Goal



# Coverage Goal





Nominal Path



Most Likely  
Failure

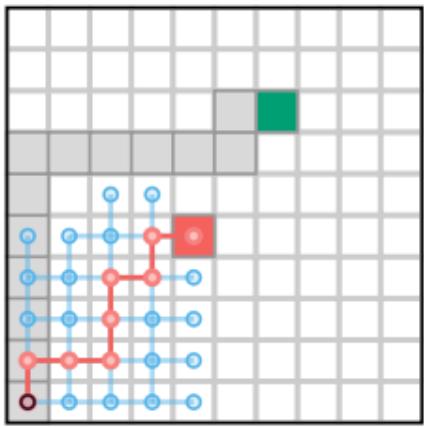
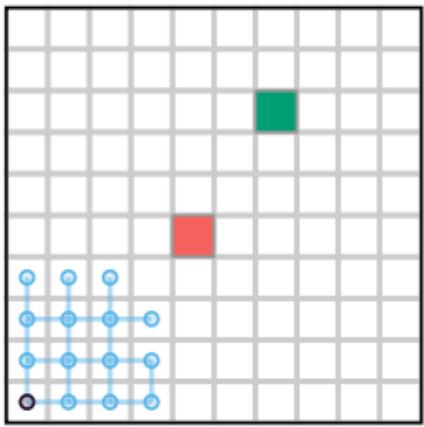
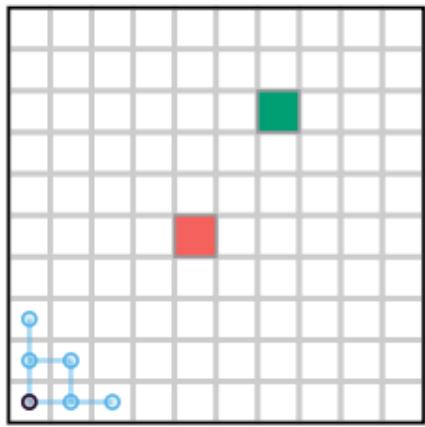
Shortest  
Failure

Iteration 10

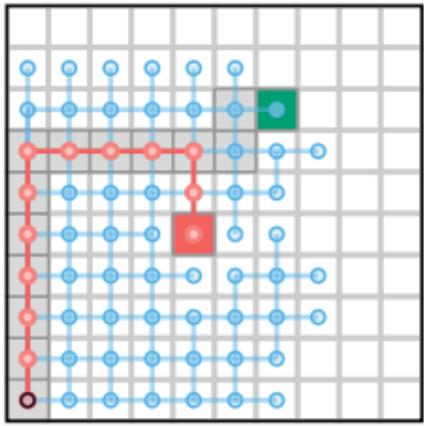
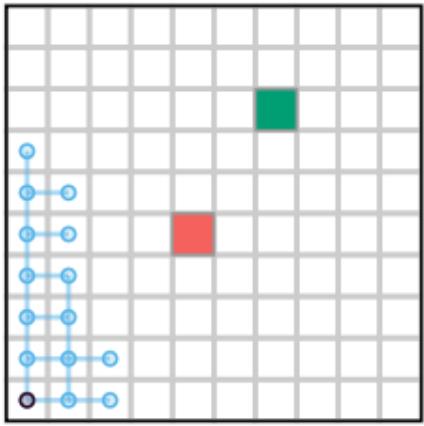
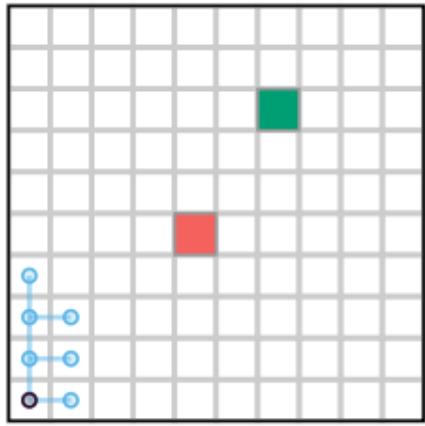
Iteration 25

Converged

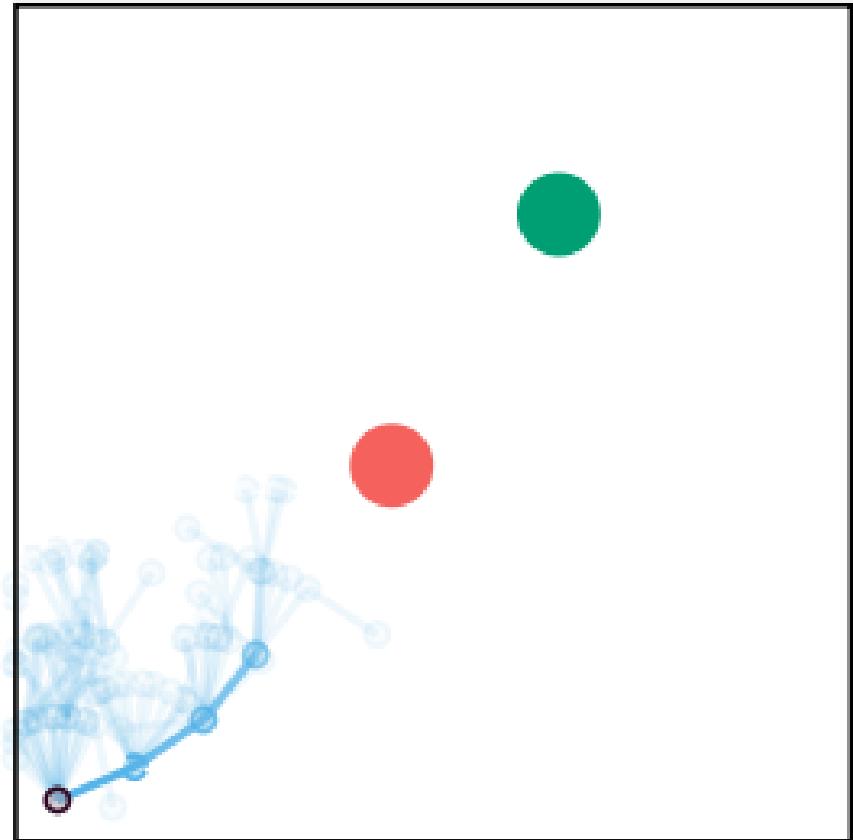
Shortest Path



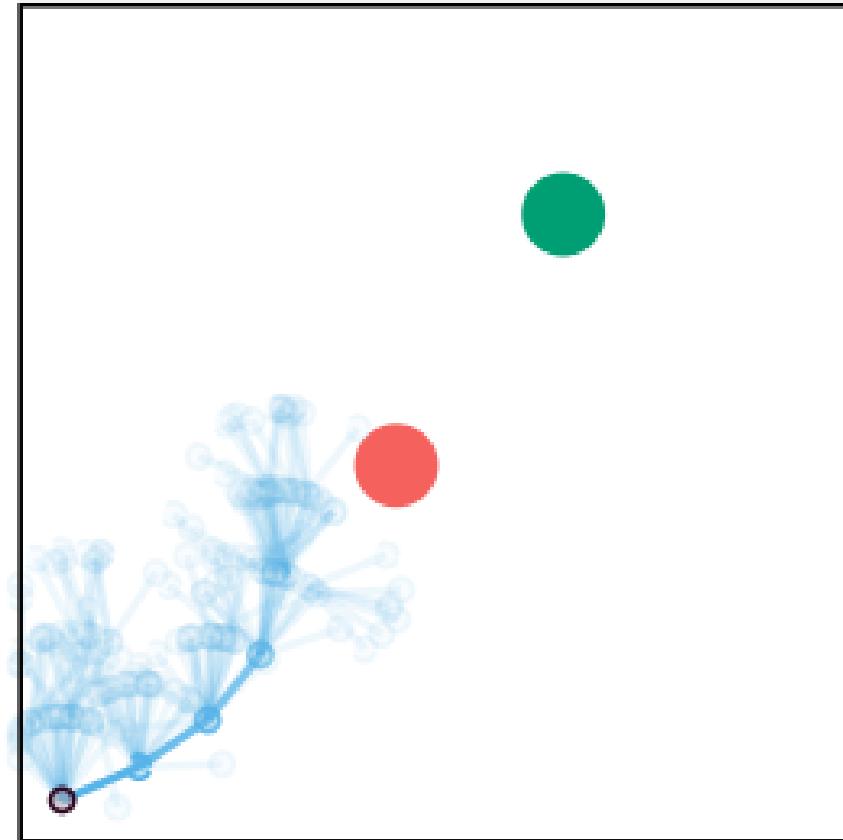
Most Likely Path



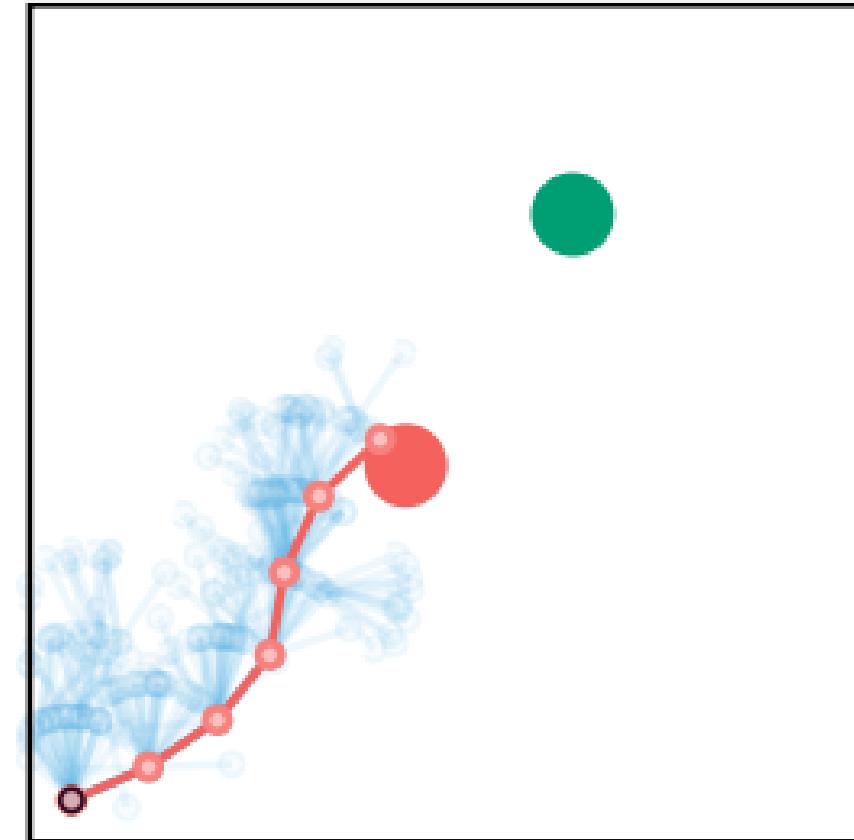
Iteration 100



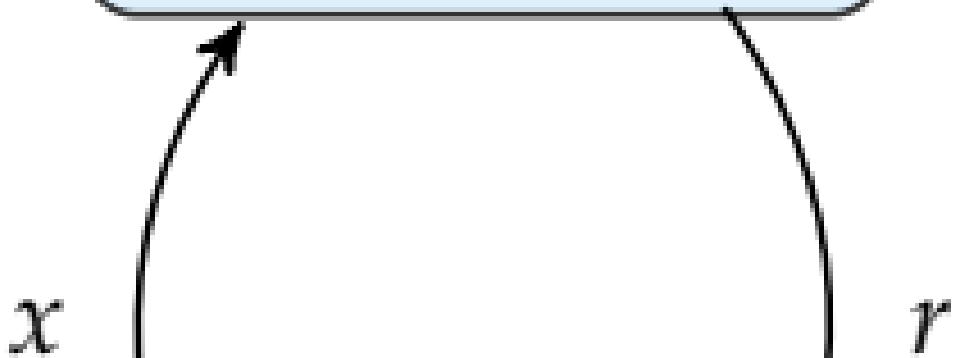
Iteration 200



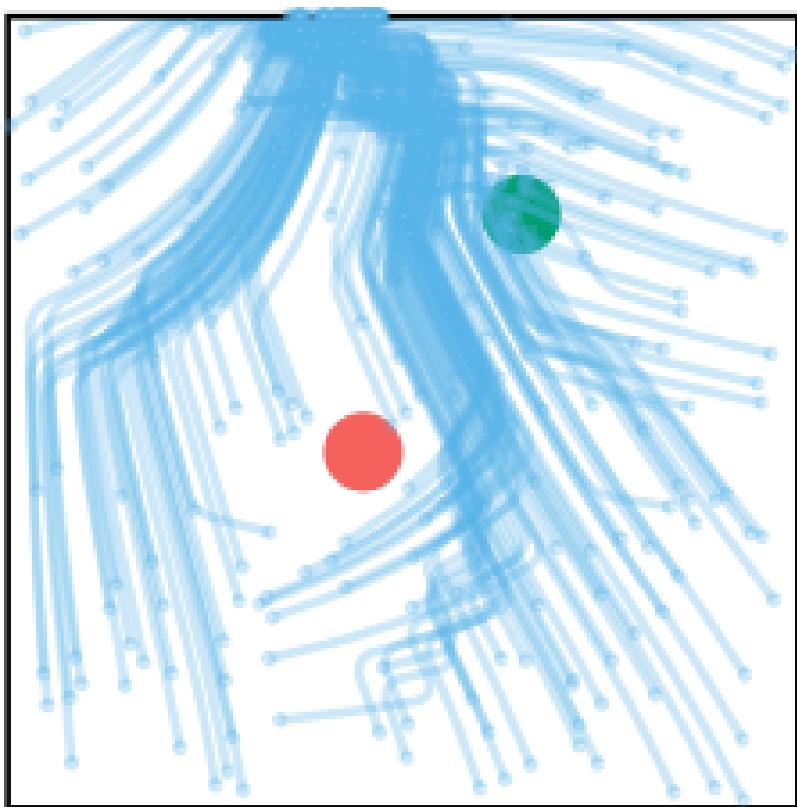
Failure Found



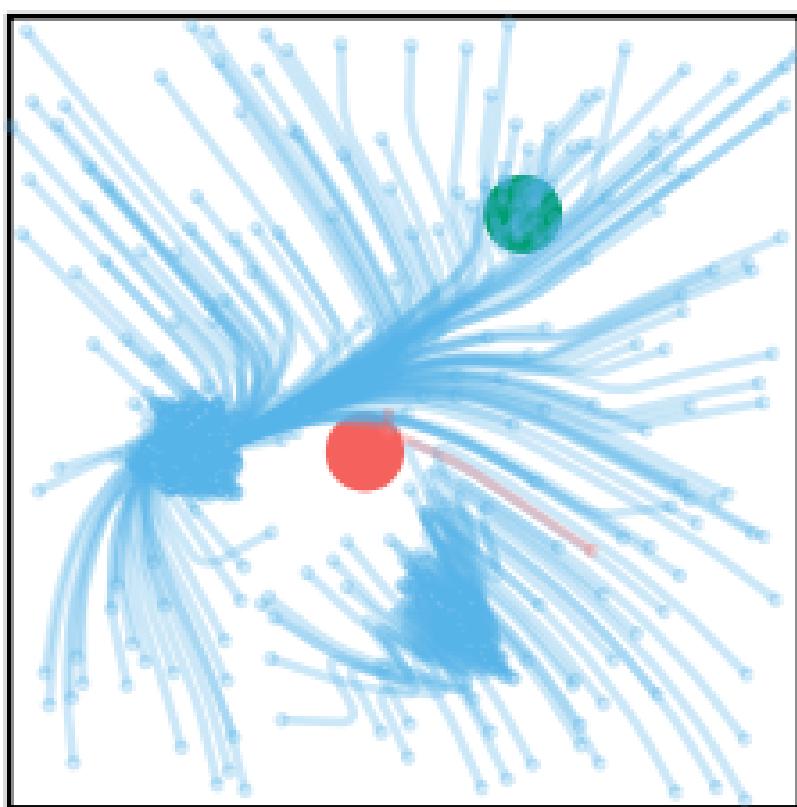
System



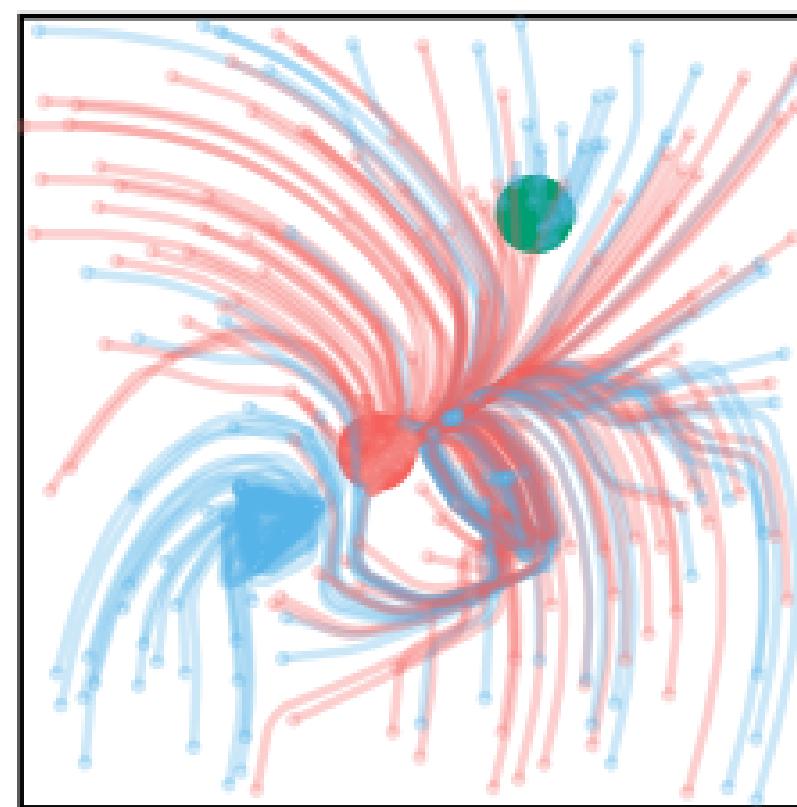
5,000 Episodes



30,000 Episodes



50,000 Episodes



## Algorithm Categories

## Simulator Requirements

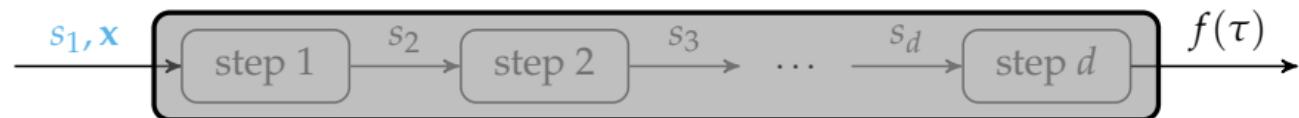
direct sampling

fuzzing

direct methods

population methods

### Black-Box Rollout



### White-Box Rollout



### Episode

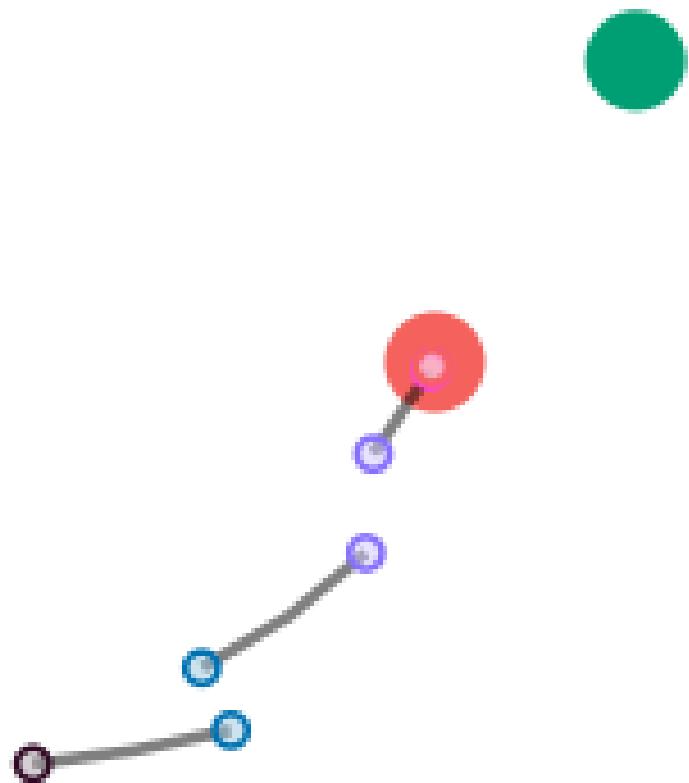
reinforcement learning

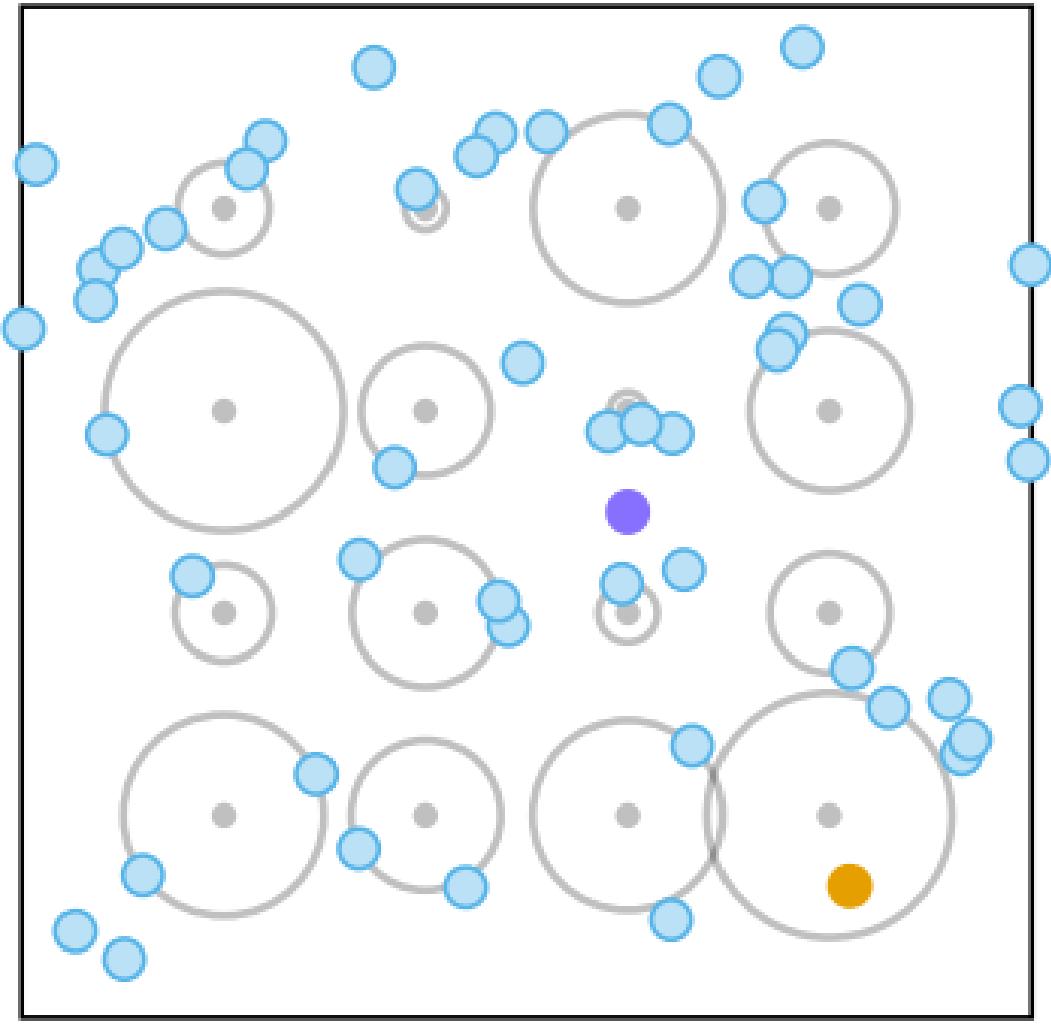


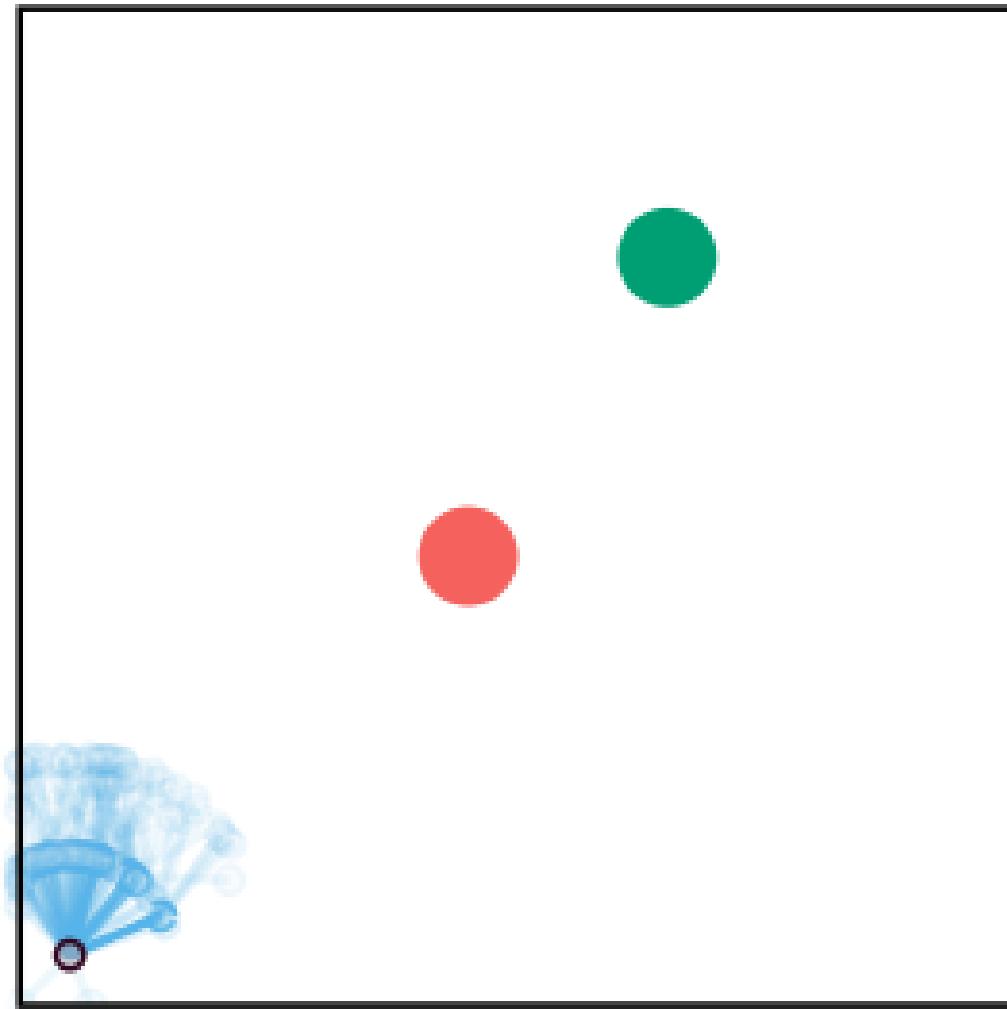
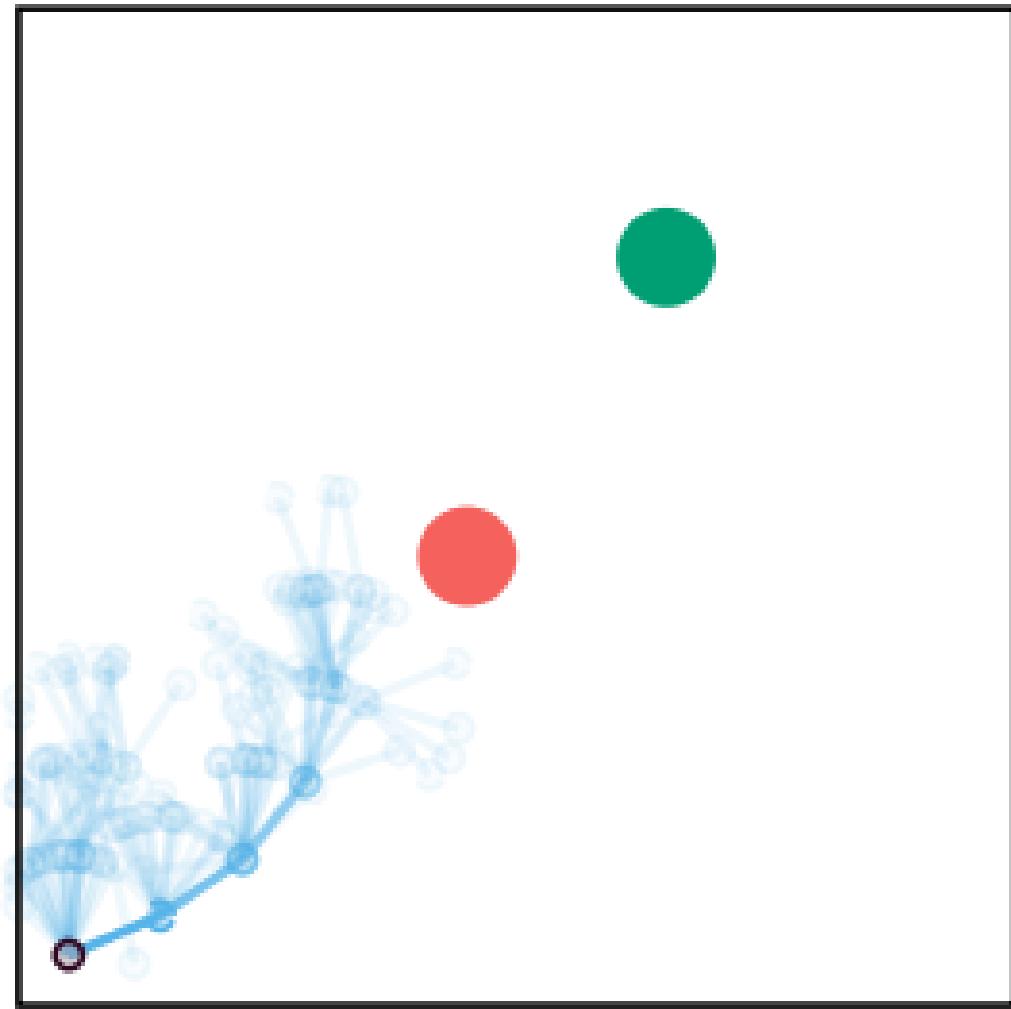
### Extend

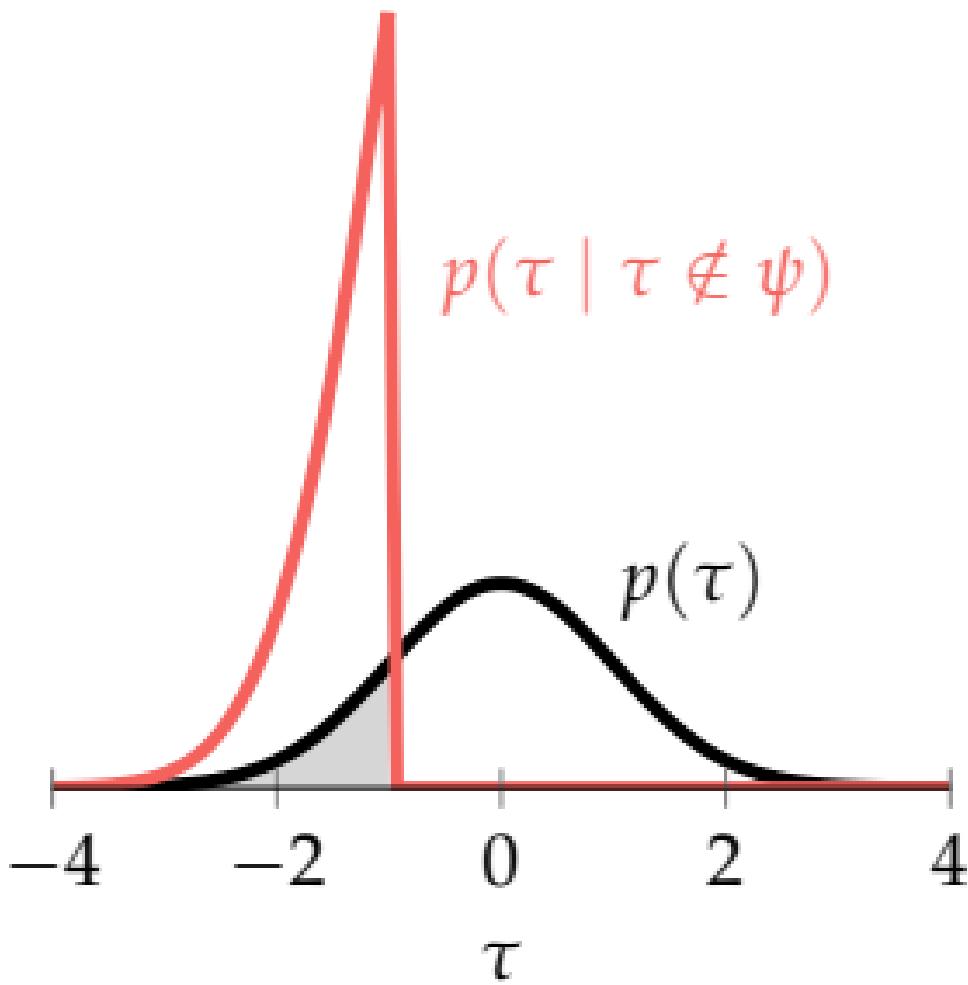
tree search  
multiple shooting

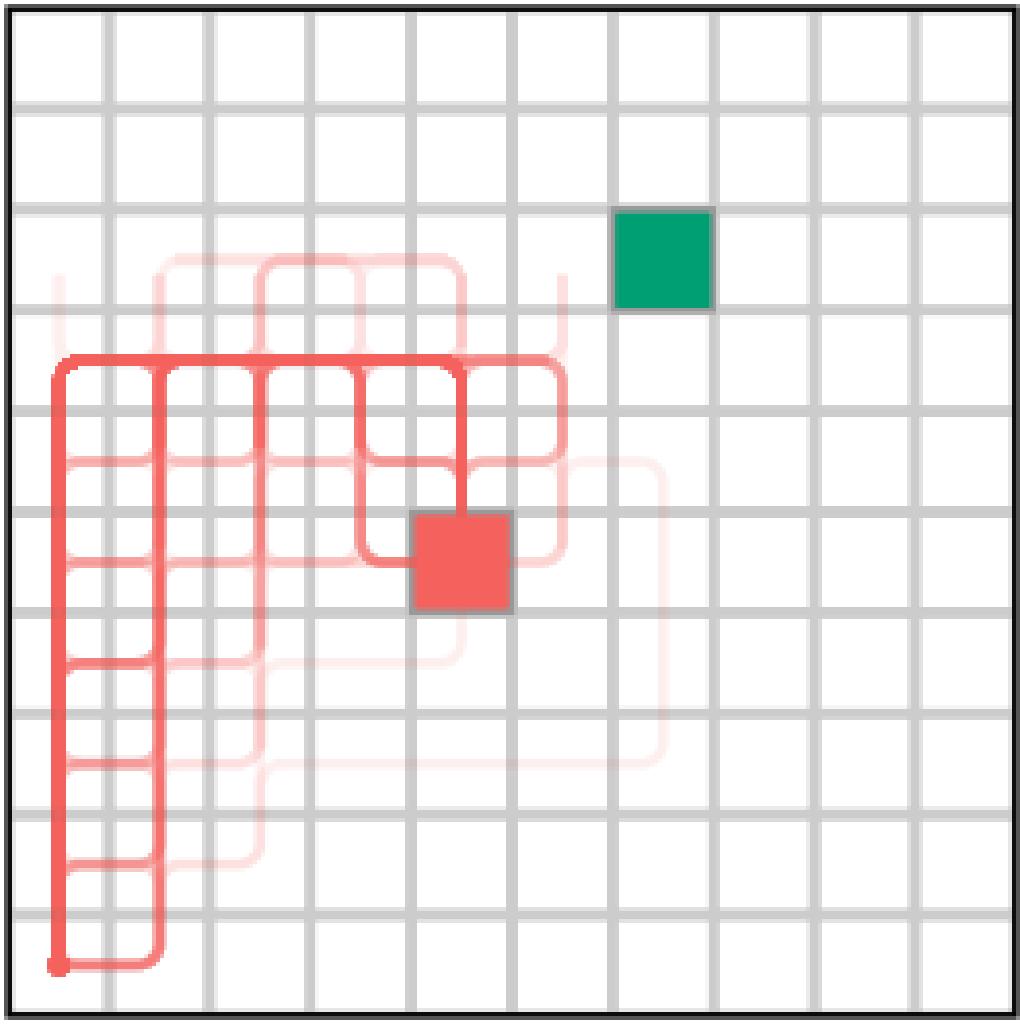


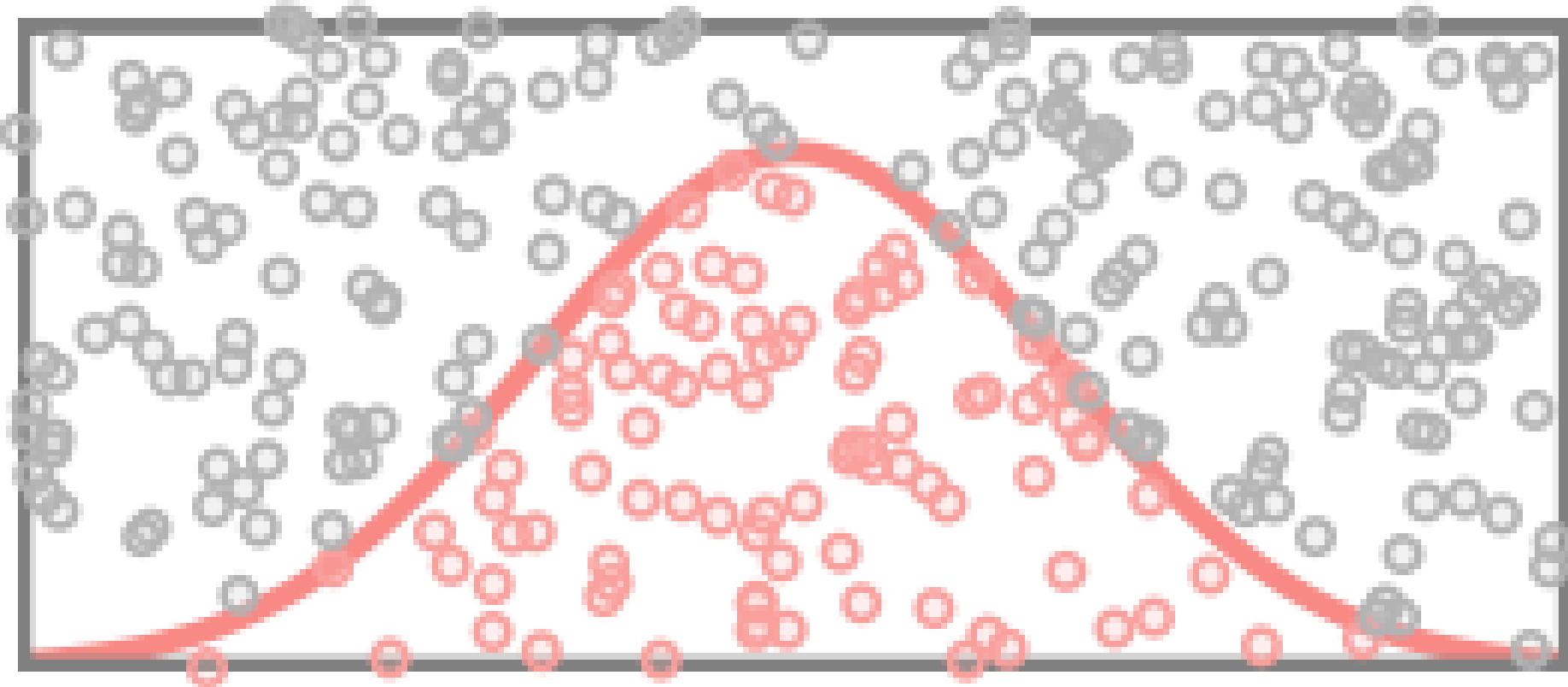


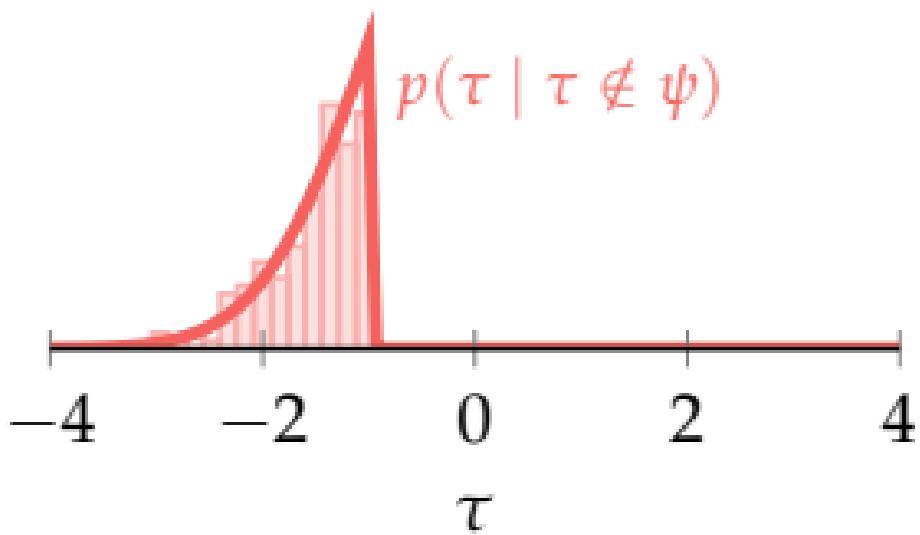
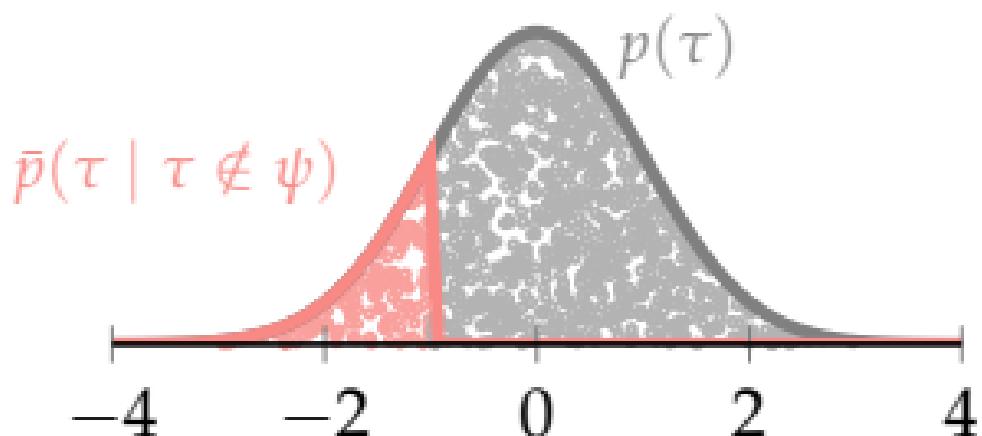




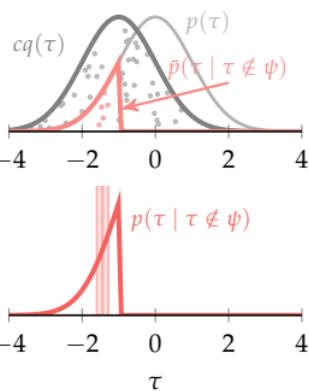




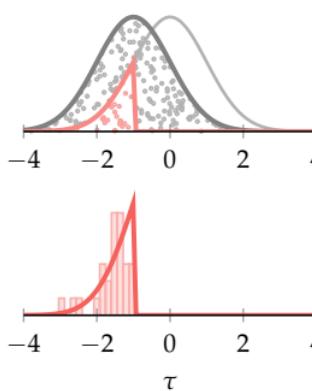




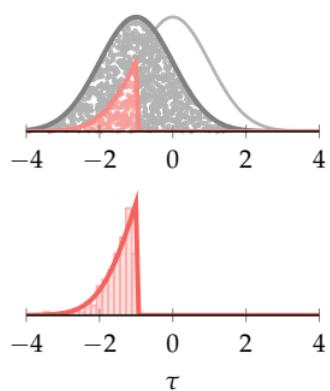
50 samples



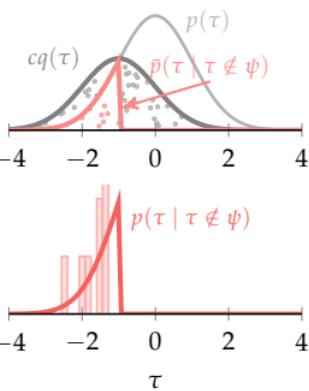
200 samples



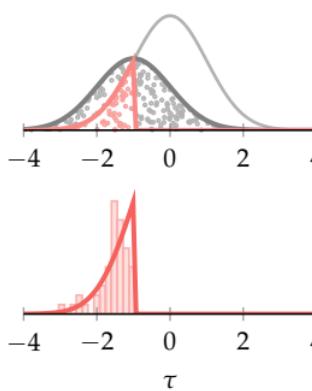
1,000 samples



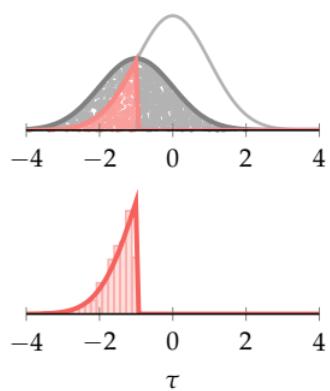
50 samples

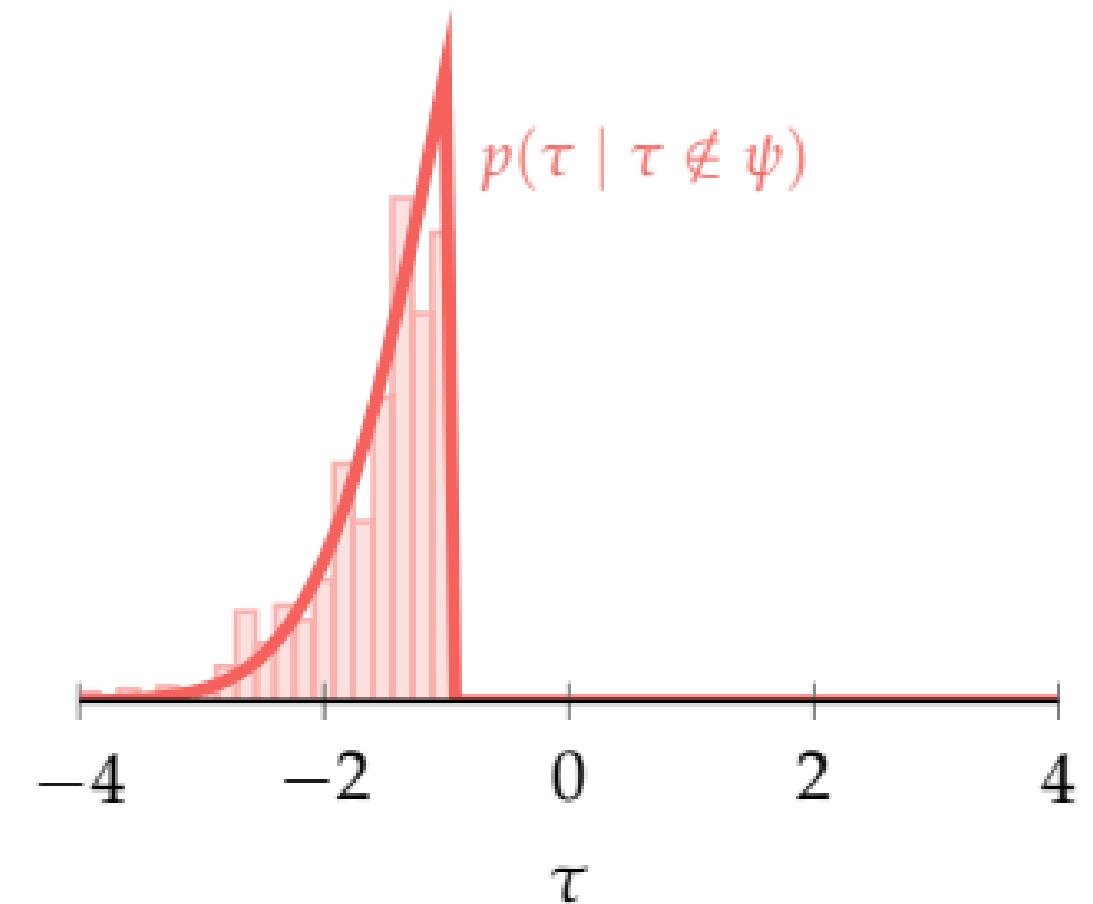
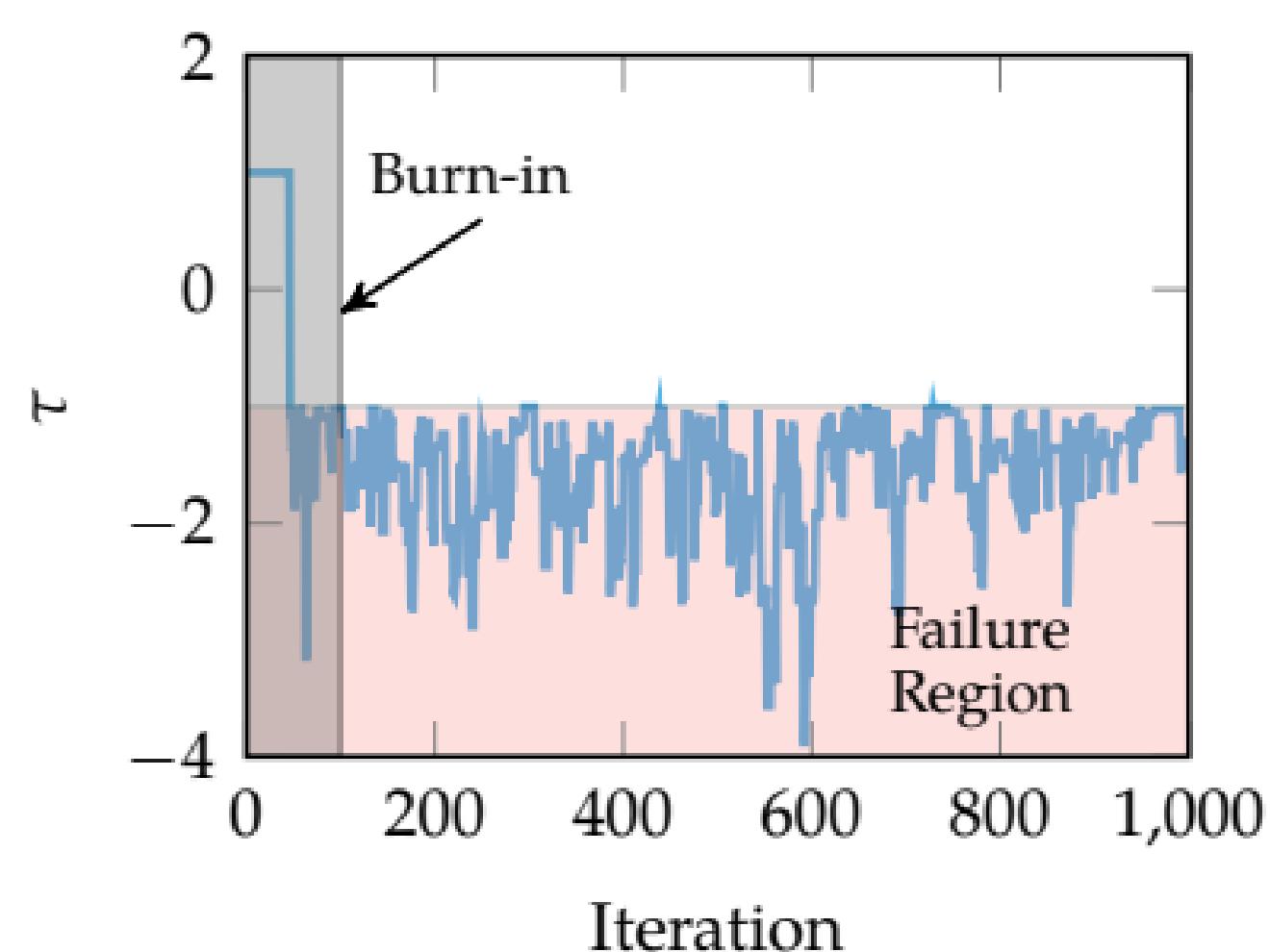


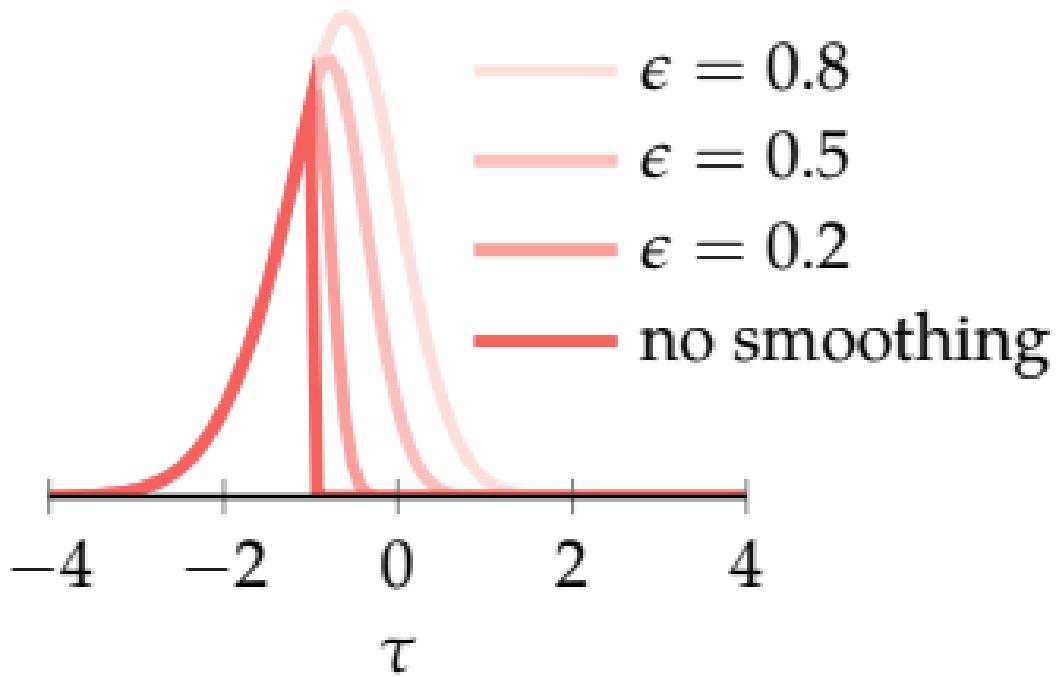
200 samples

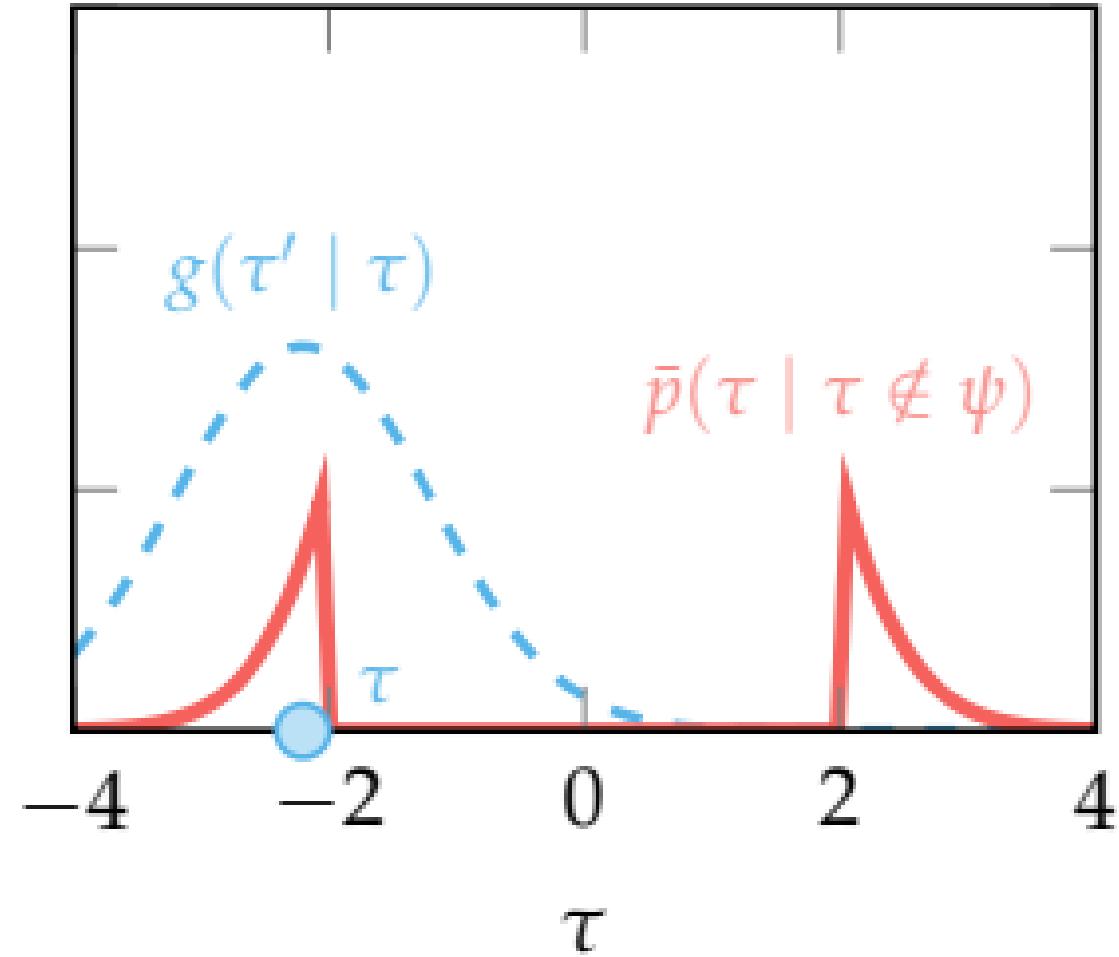
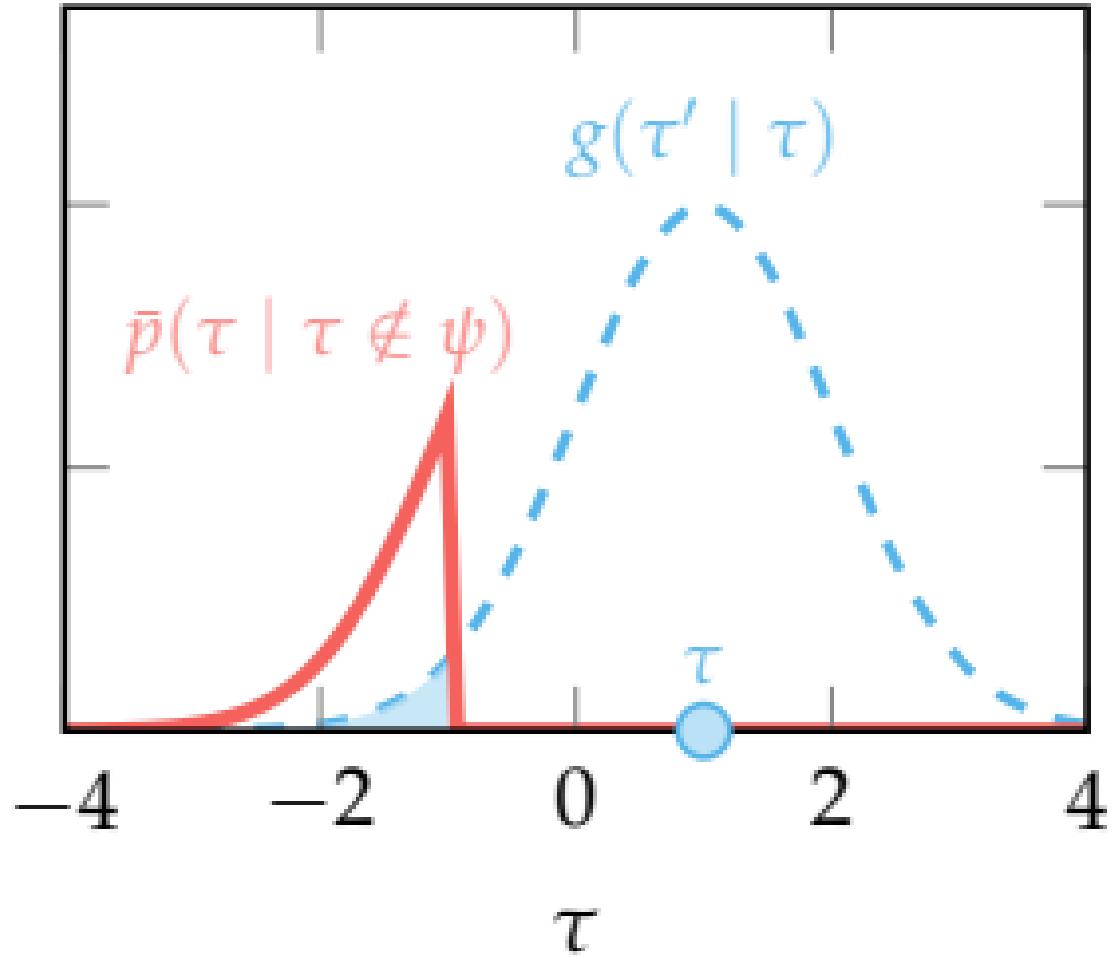


1,000 samples

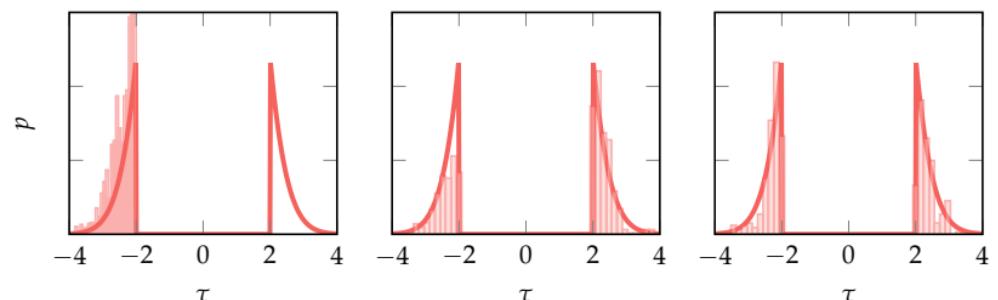
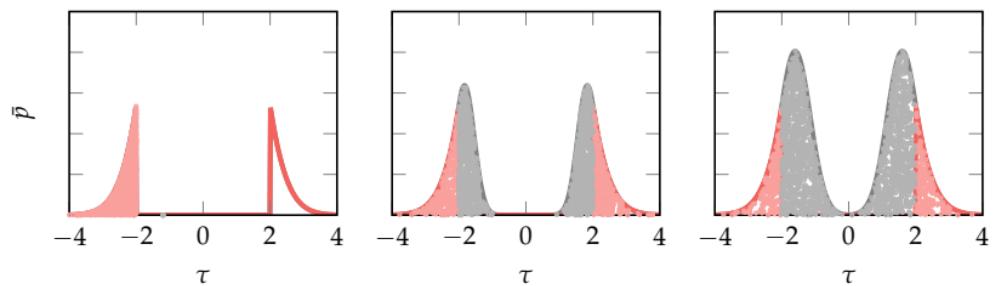
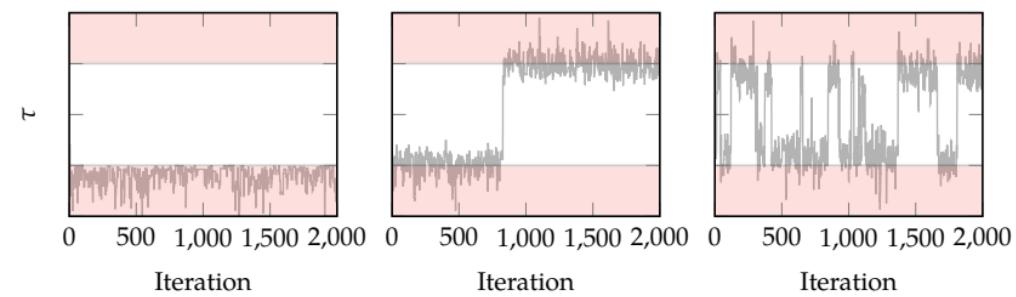
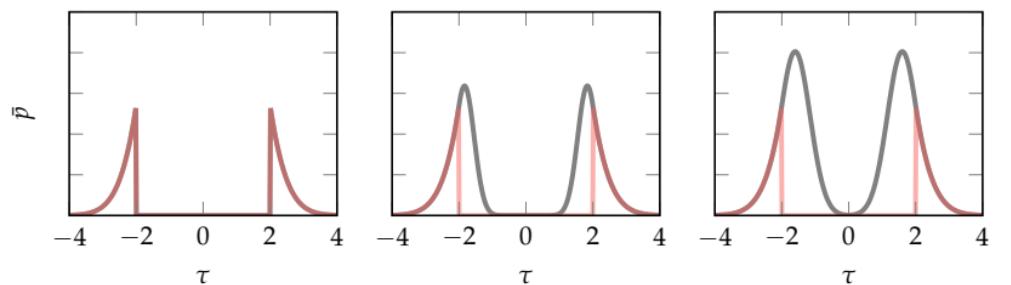




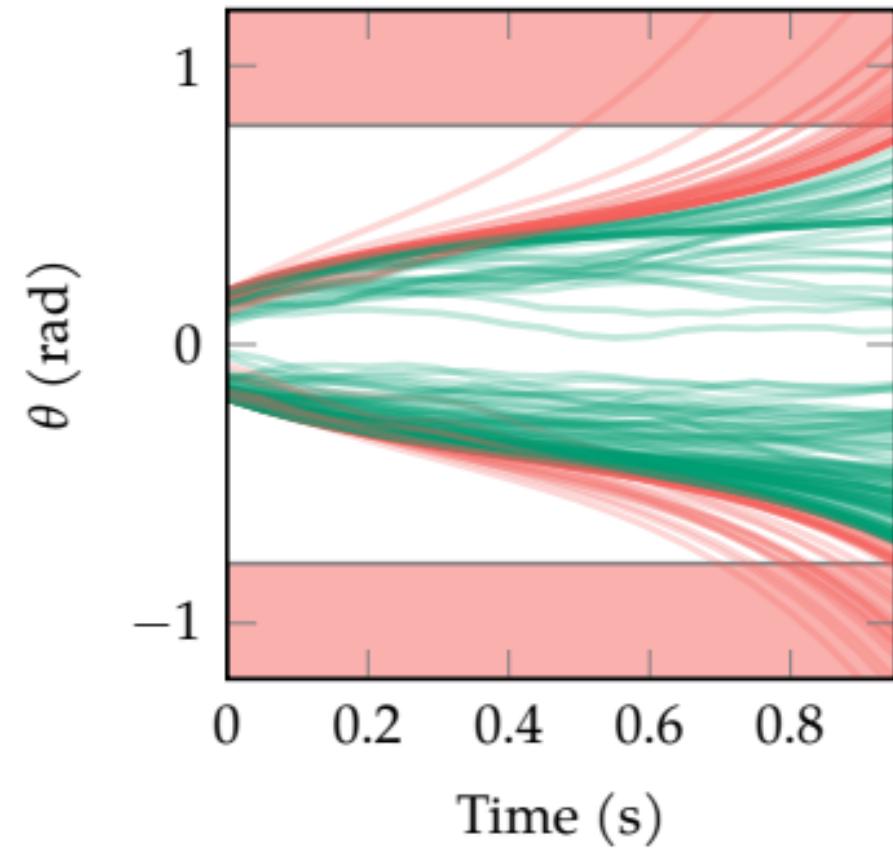




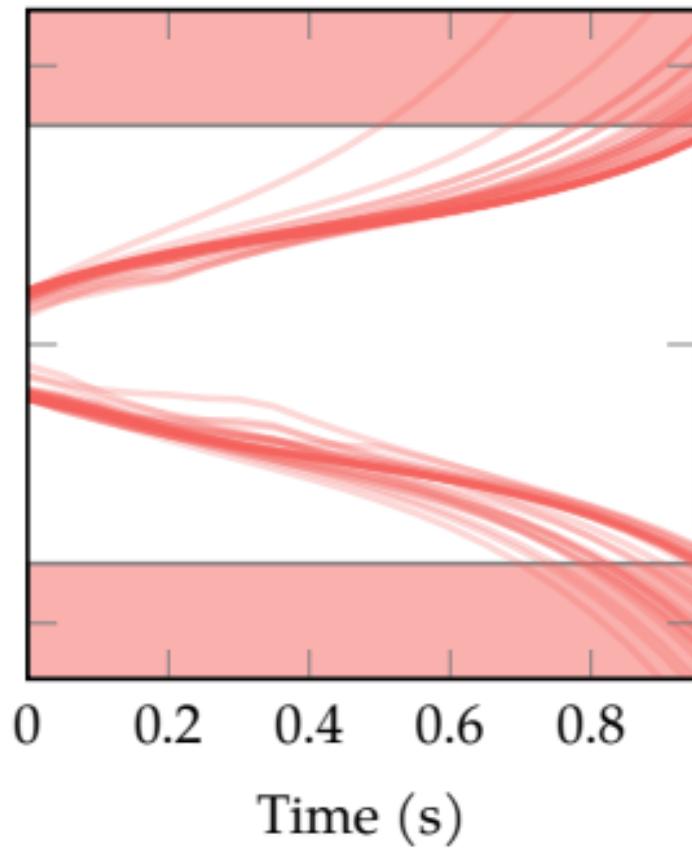
No Smoothing

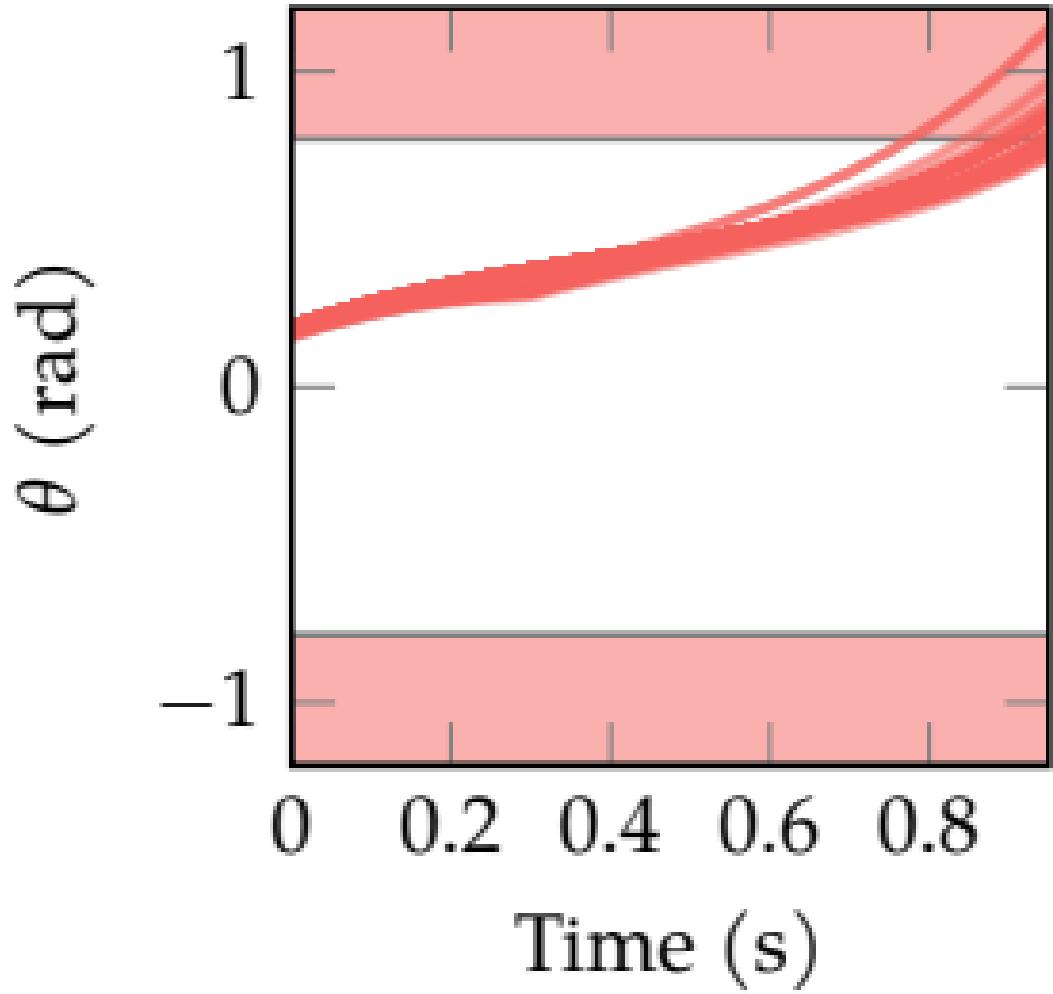
 $\epsilon = 0.3$  $\epsilon = 0.5$ 

# MCMC Output

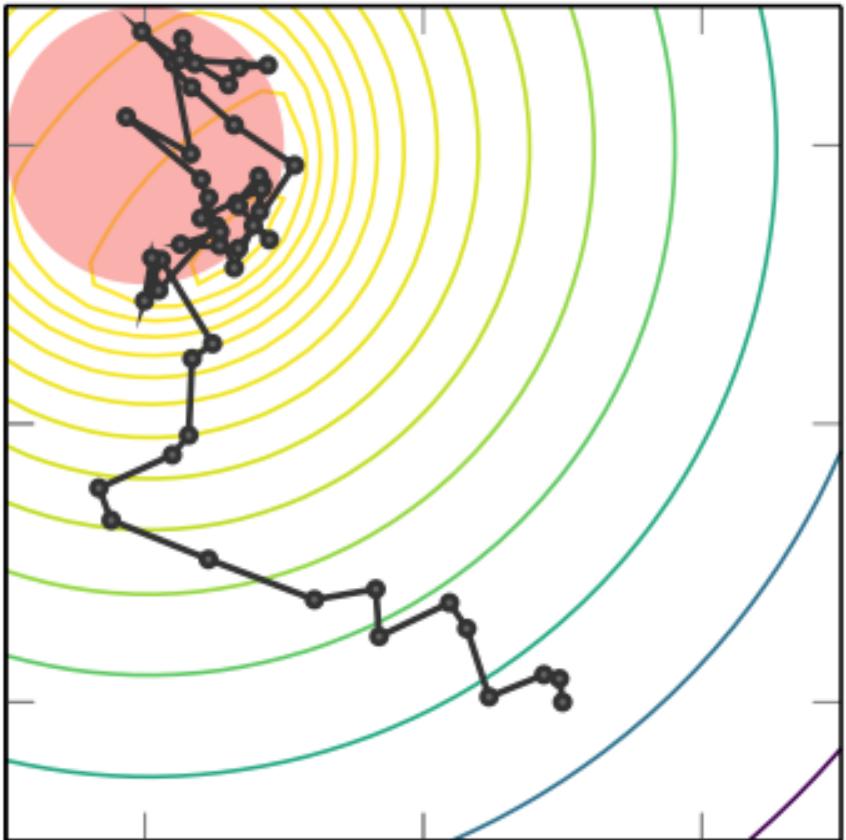


# Failure Distribution

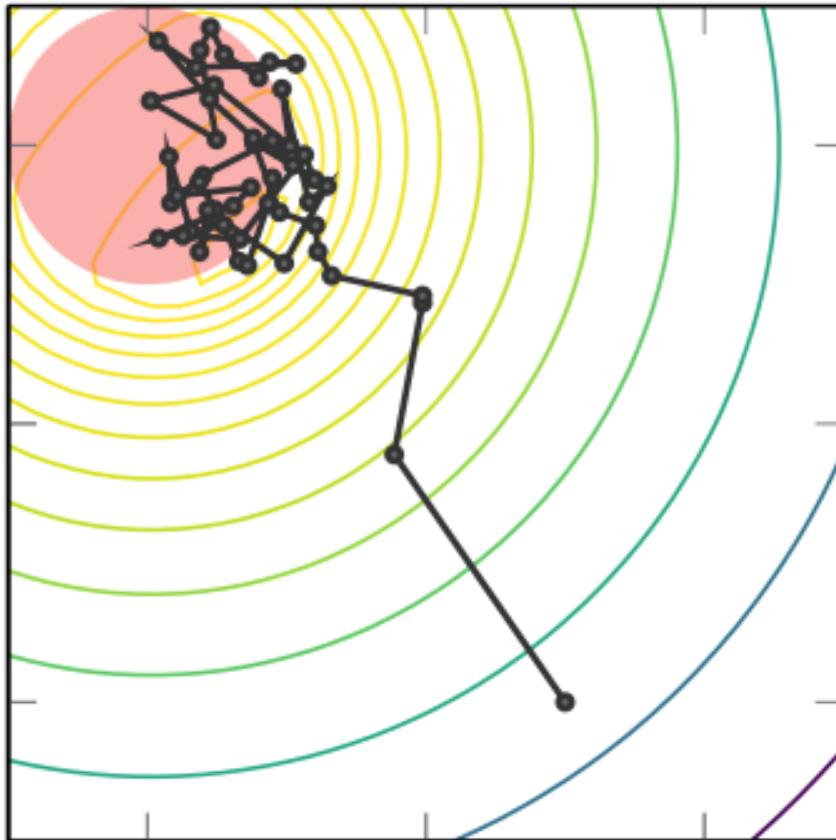


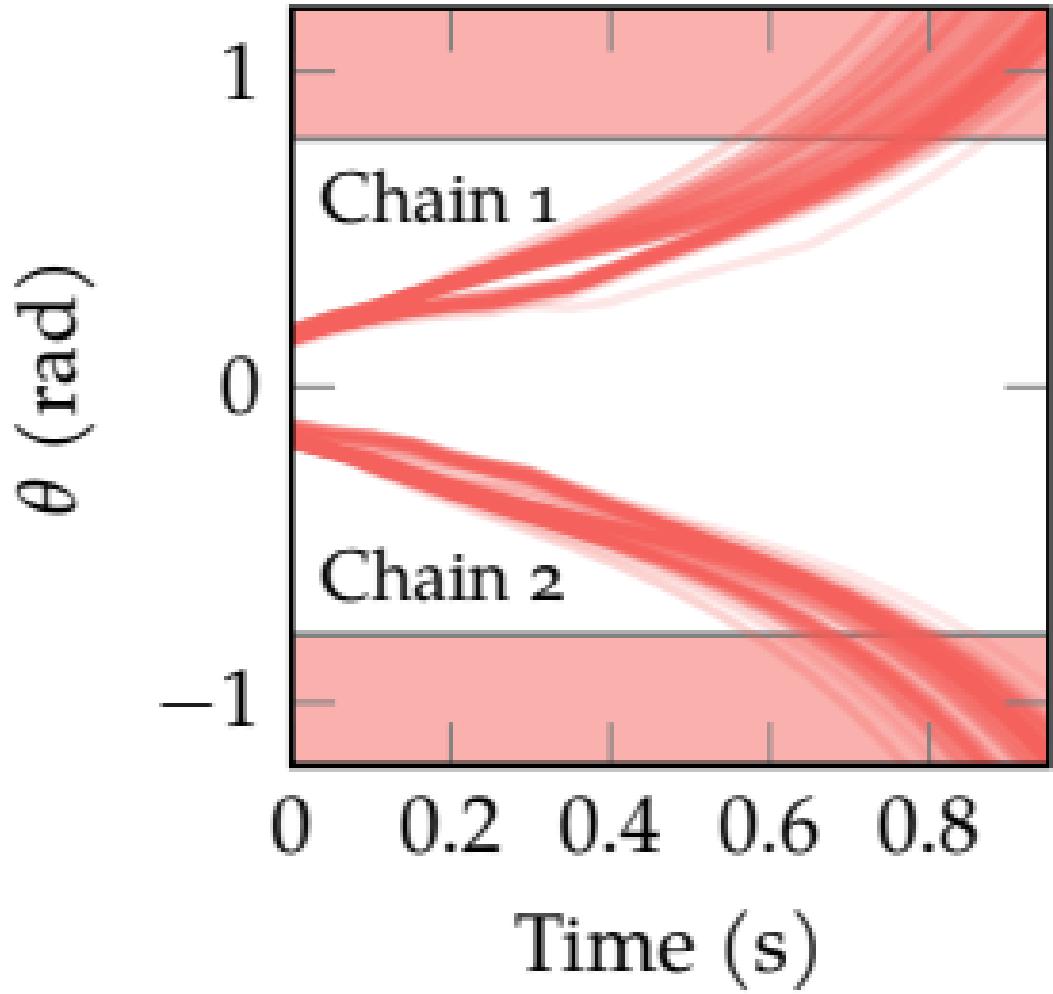


# Gaussian Kernel



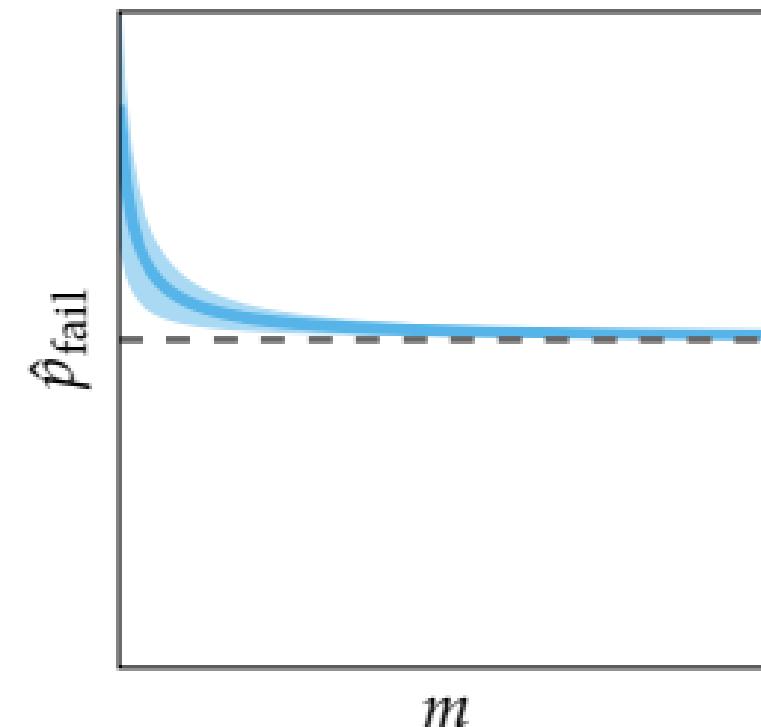
# MALA Kernel





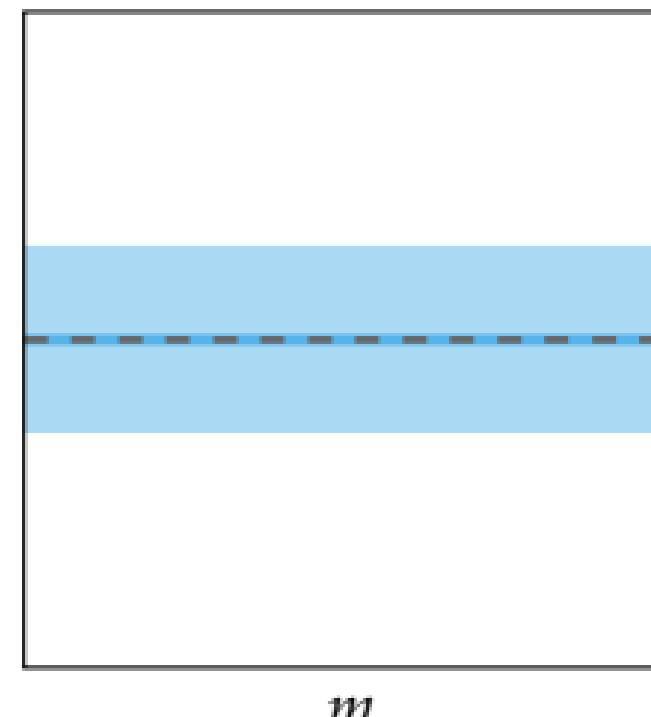
Unbiased: ✗

Consistent: ✓



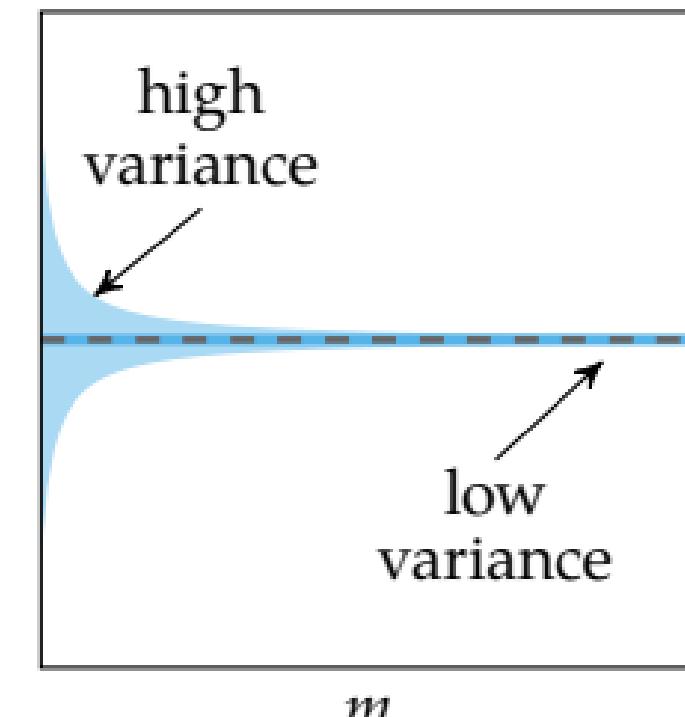
Unbiased: ✓

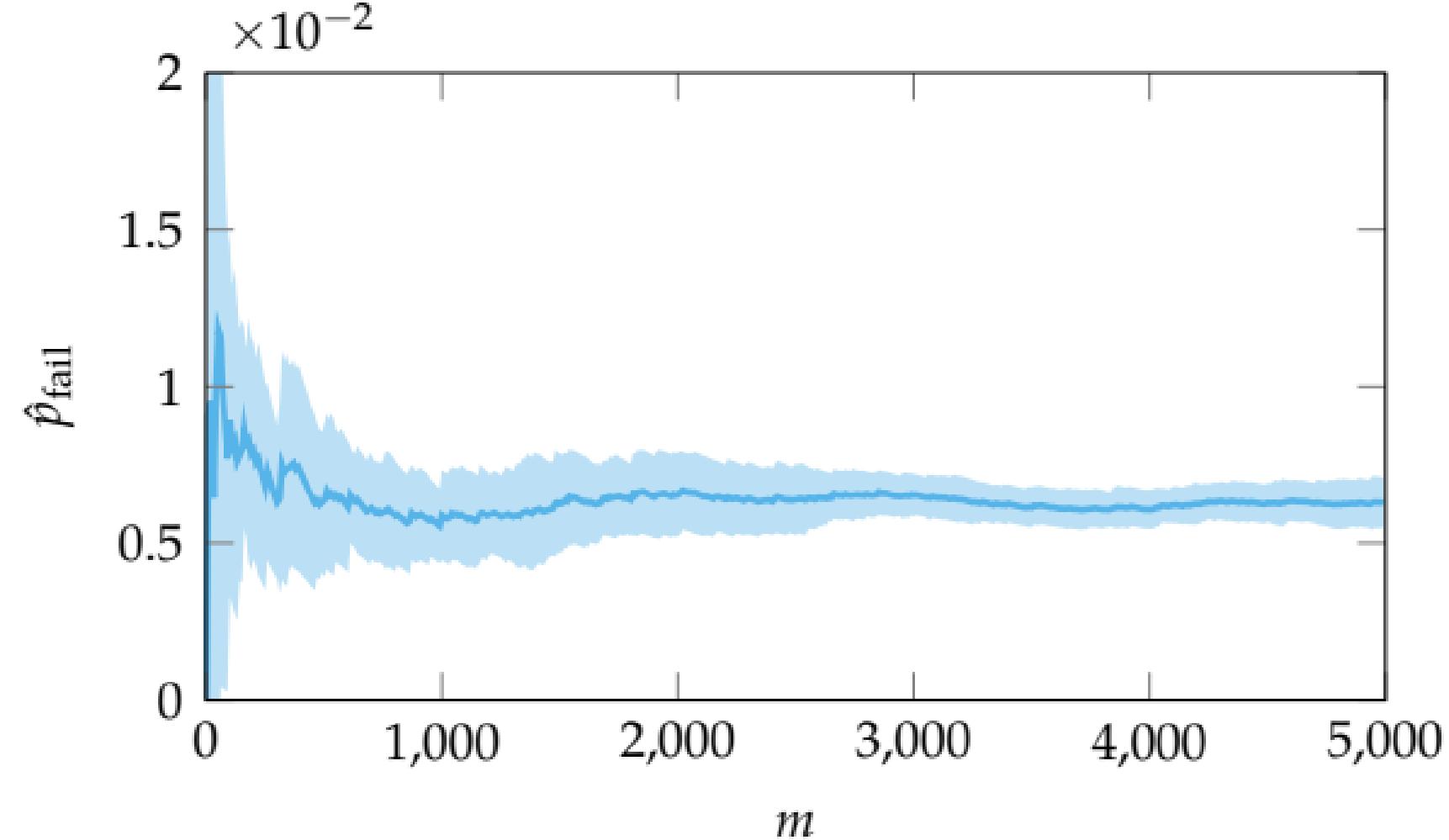
Consistent: ✗

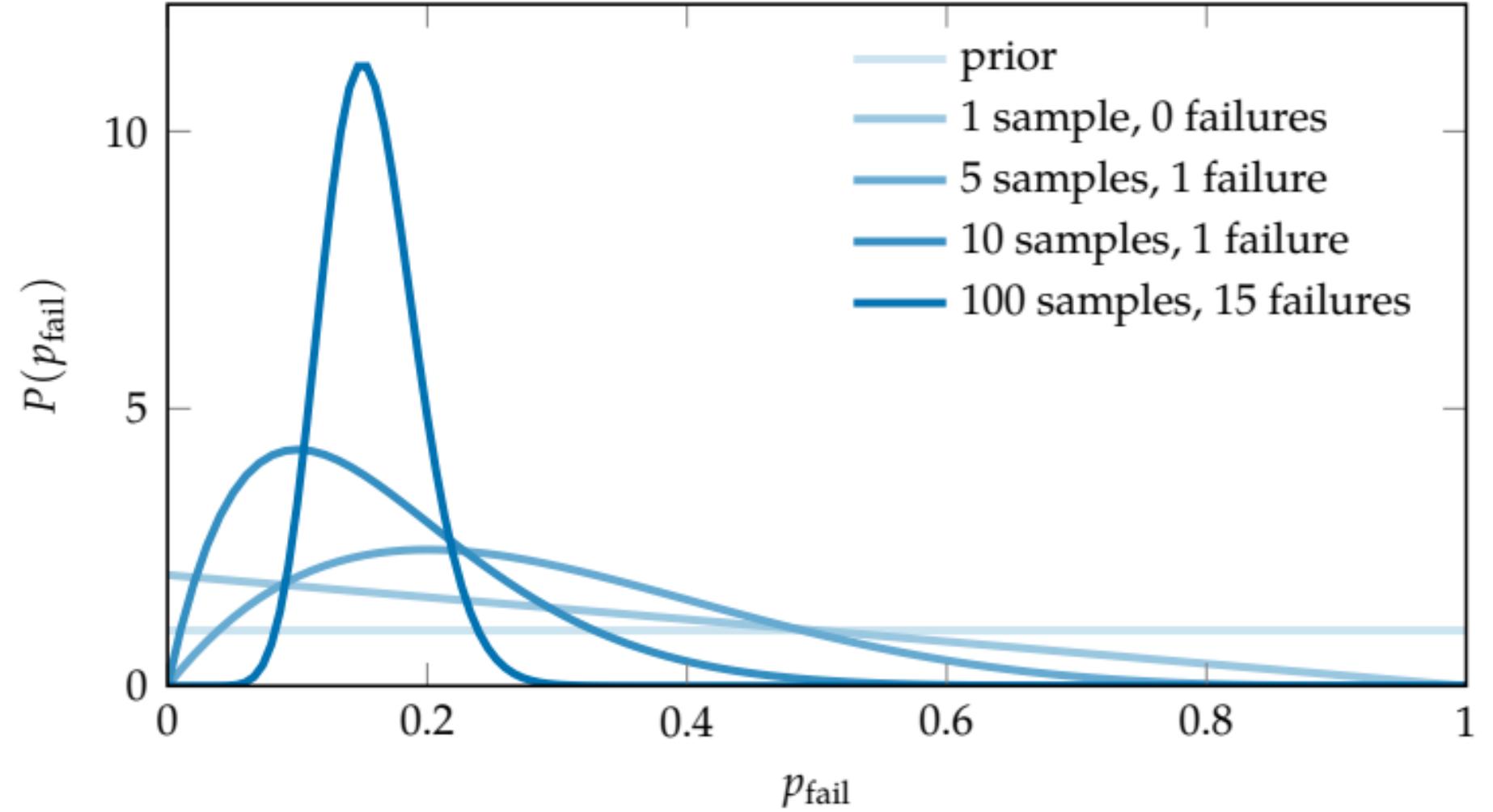


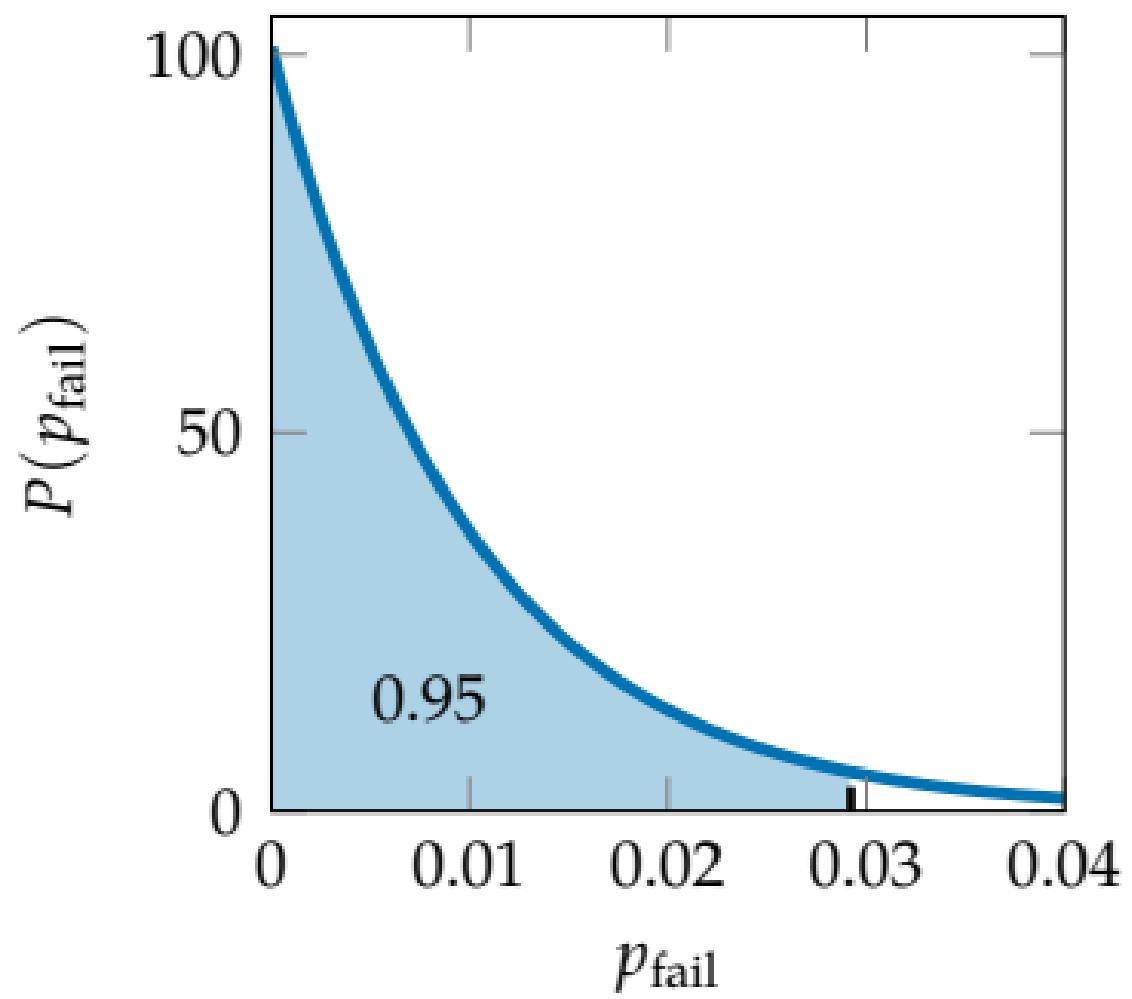
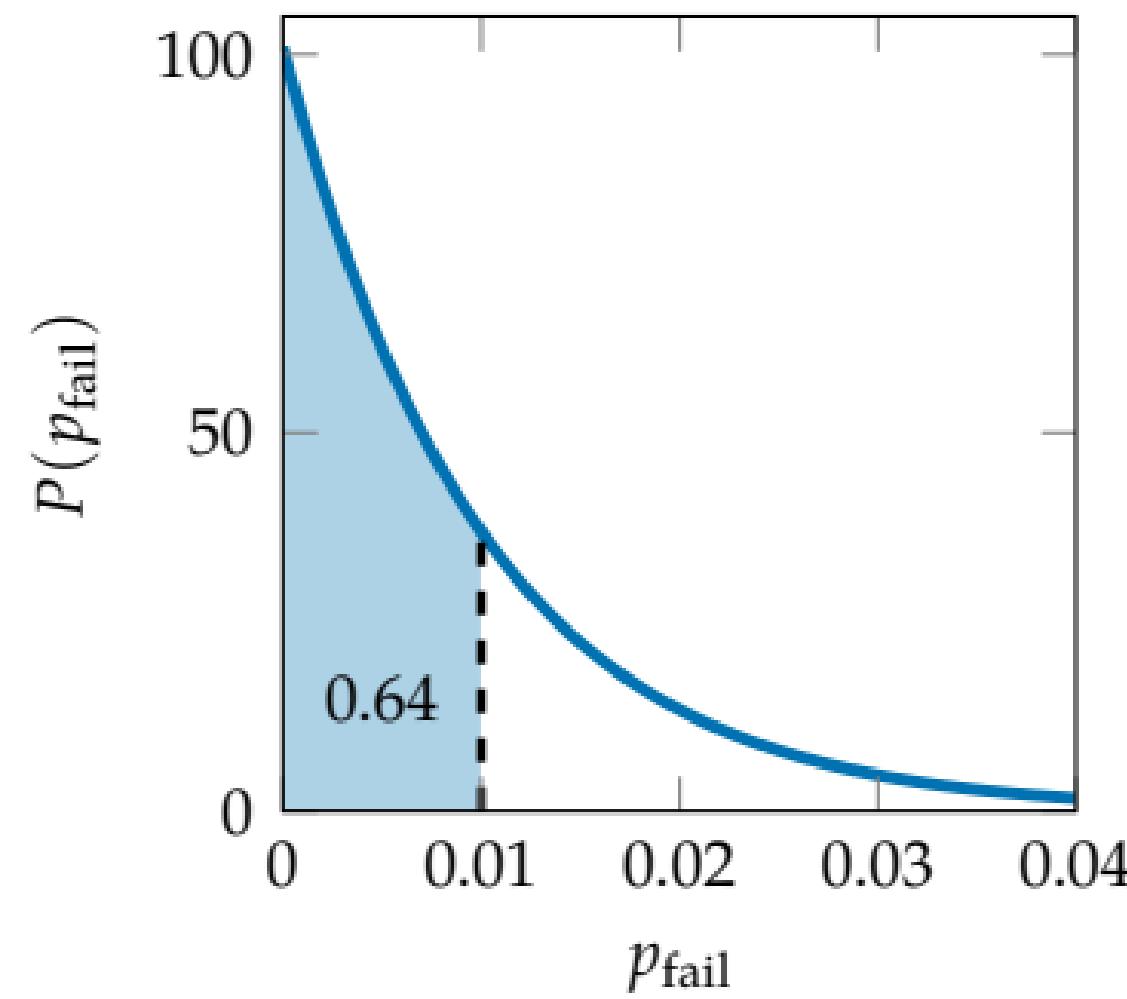
Unbiased: ✓

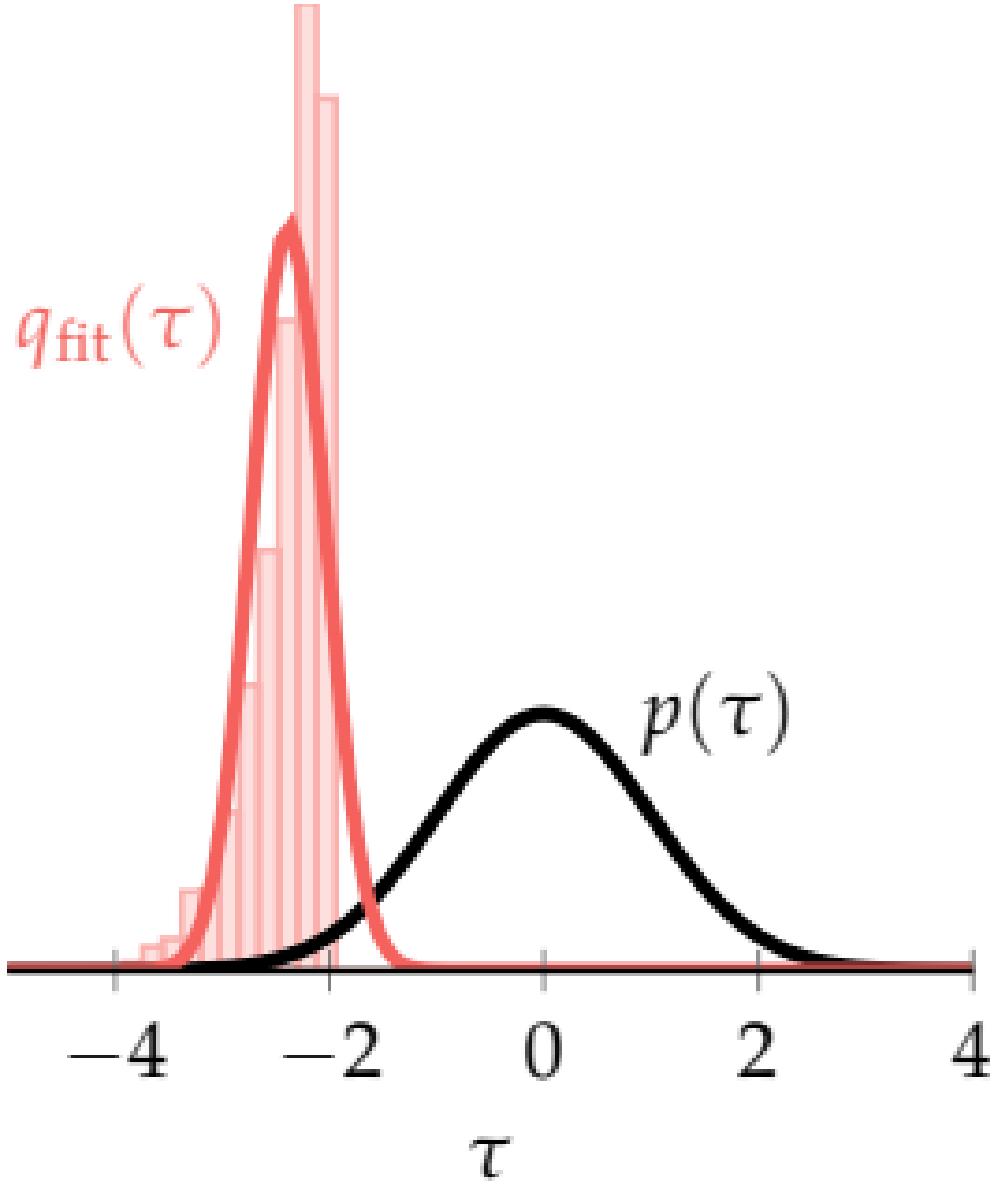
Consistent: ✓

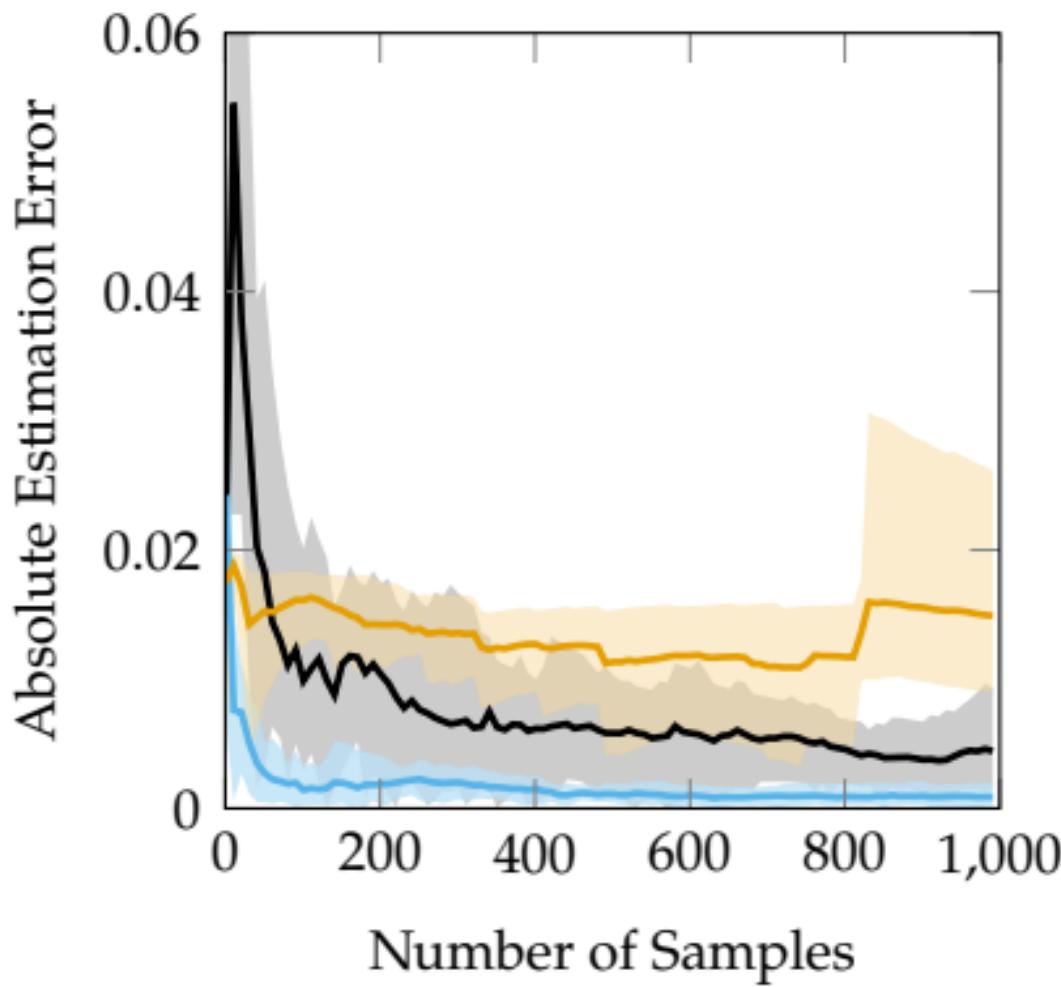
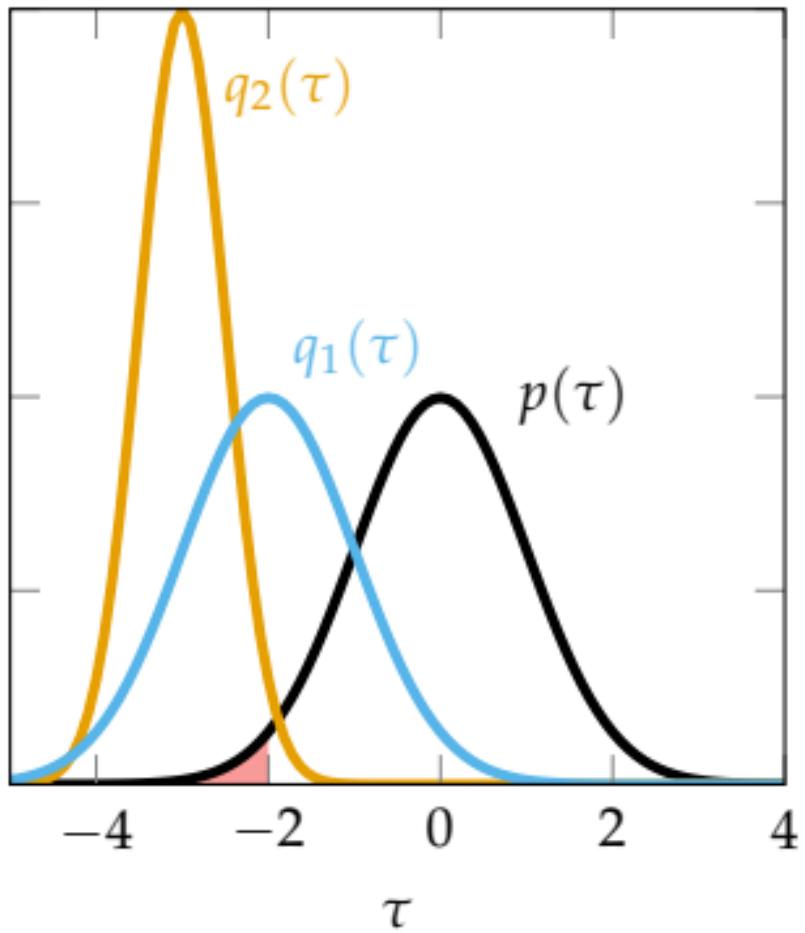




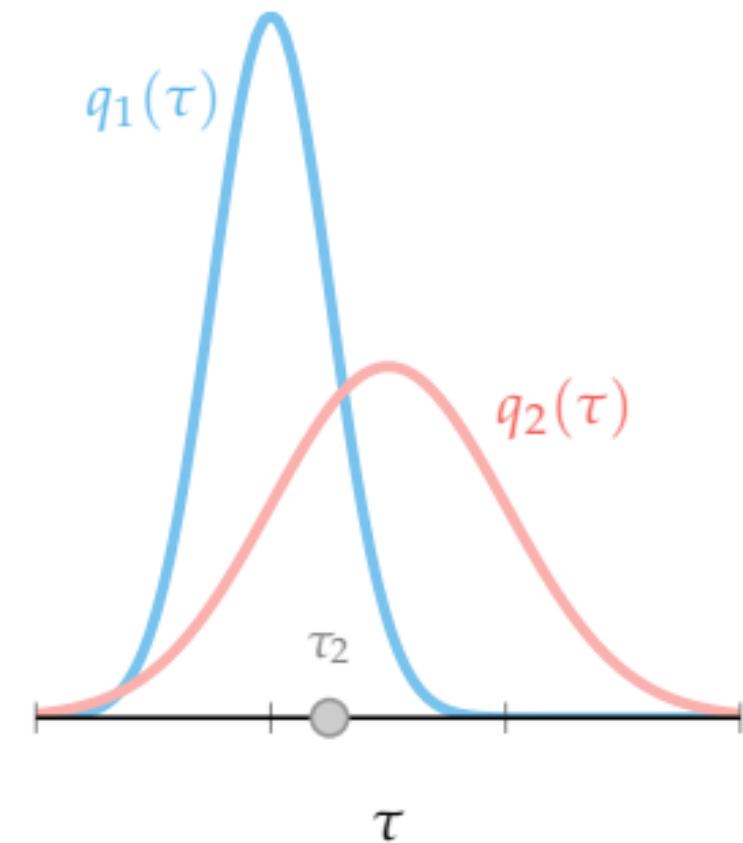




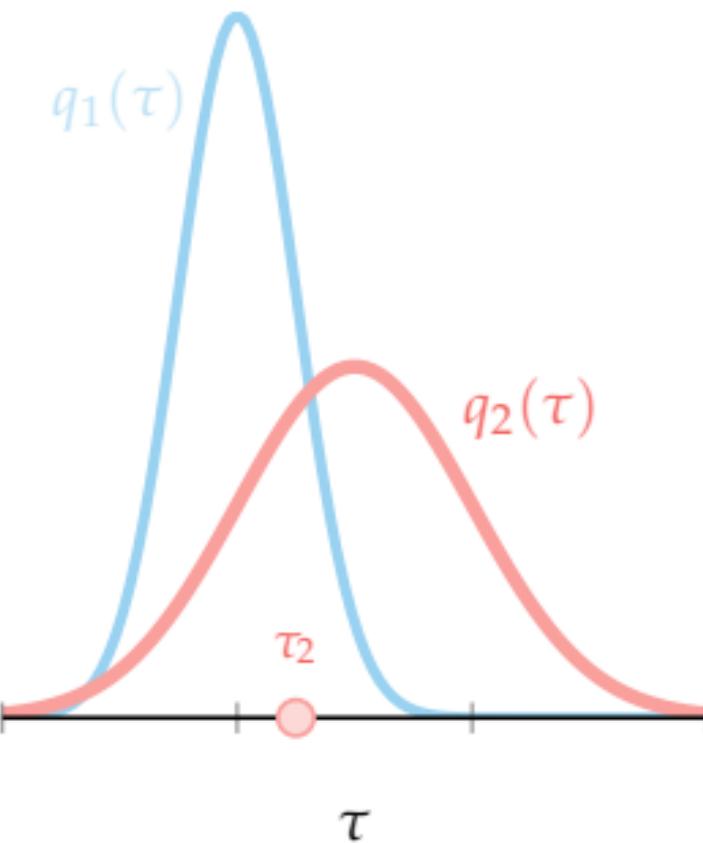




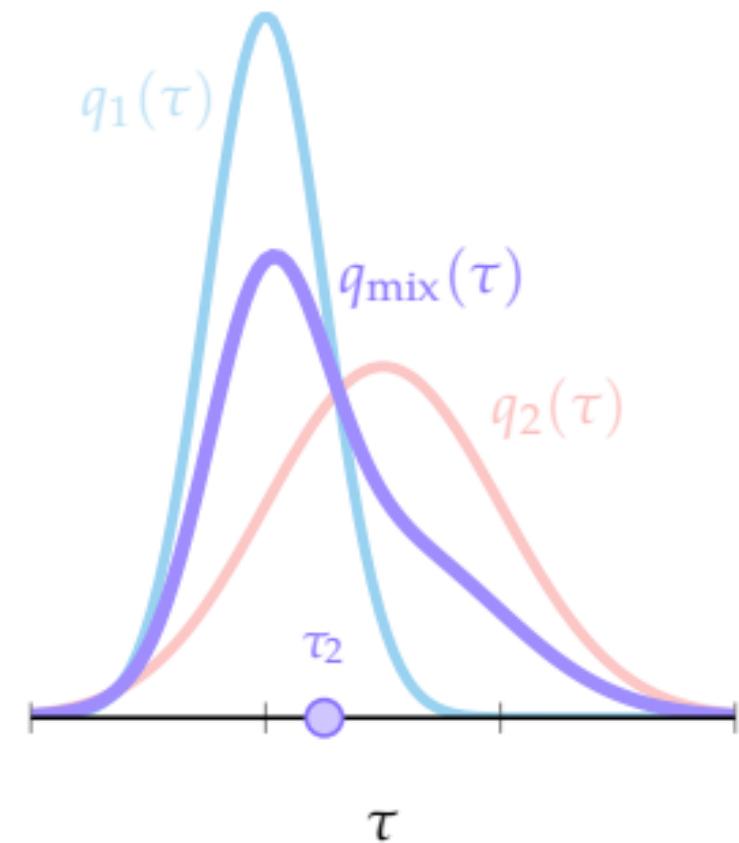
# Proposal Distributions



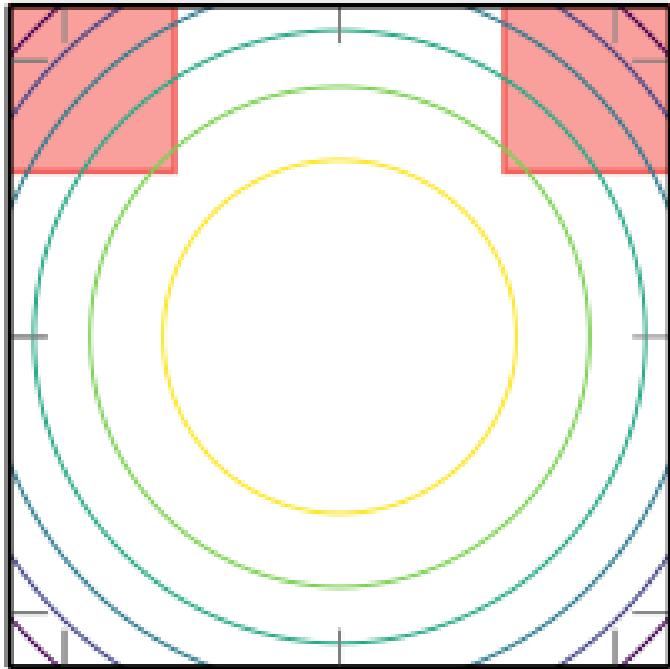
# s-MIS



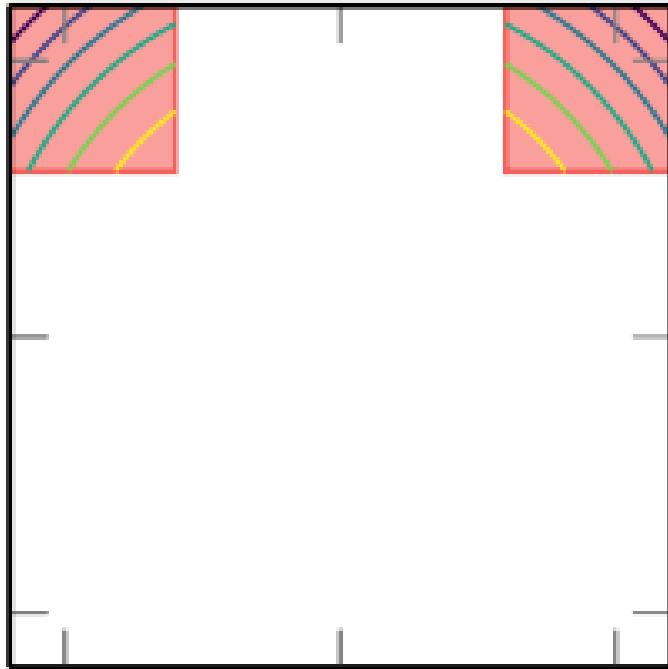
# DM-MIS



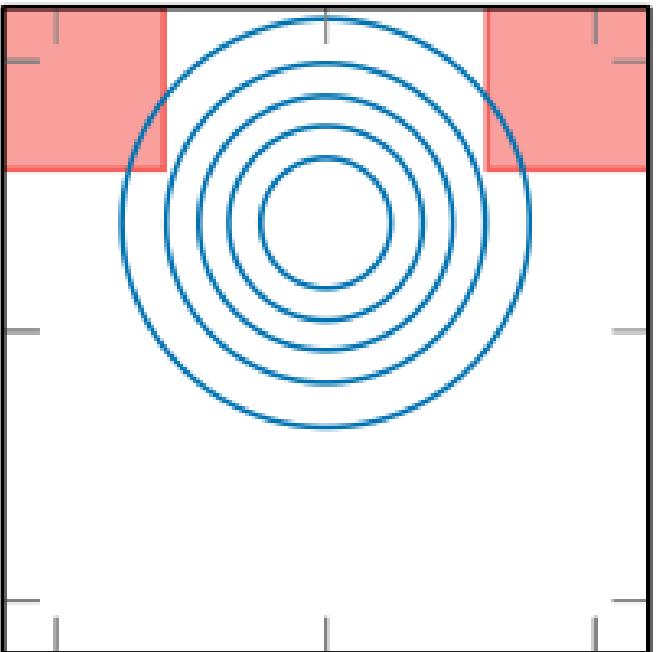
# Nominal Distribution

 $\tau_2$  $\tau_1$ 

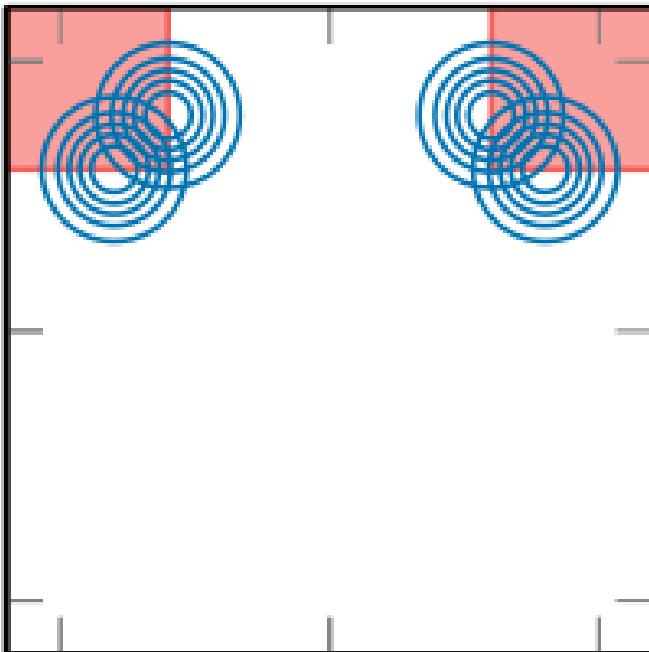
# Failure Distribution

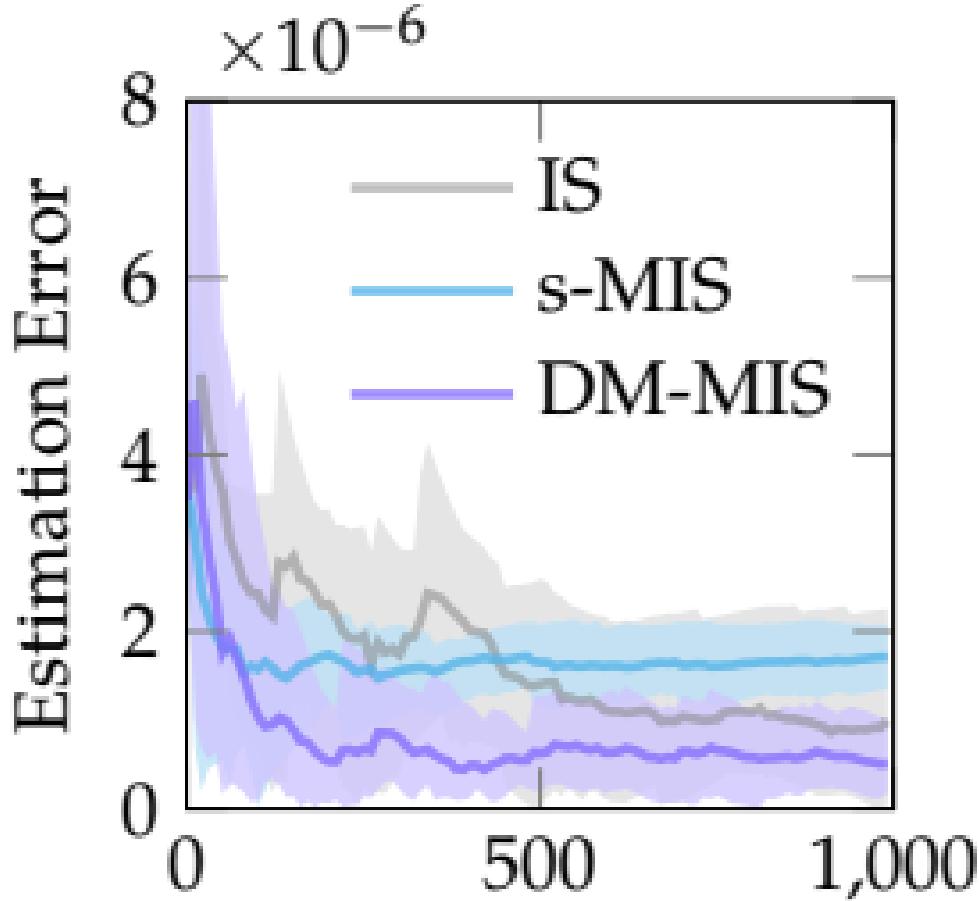
 $\tau_2$  $\tau_1$

# IS Proposal

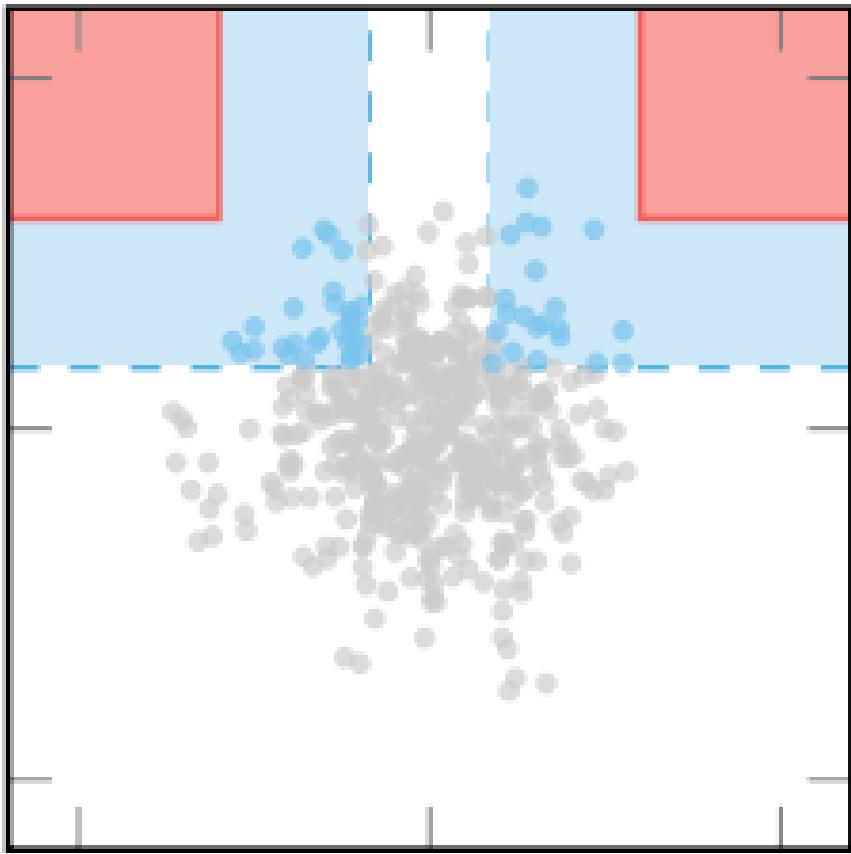
 $\tau_2$  $\tau_1$ 

# MIS Proposals

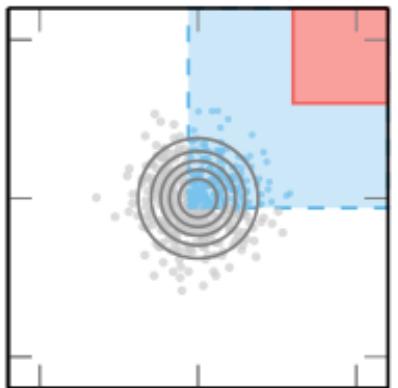
 $\tau_2$  $\tau_1$



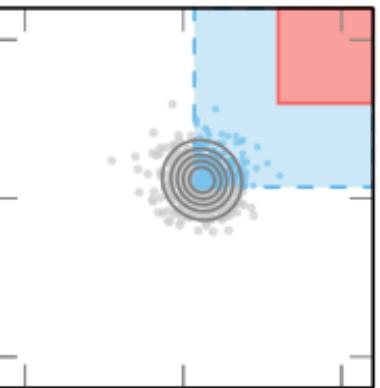
Number of Samples

$\tau_2$  $\tau_1$ 

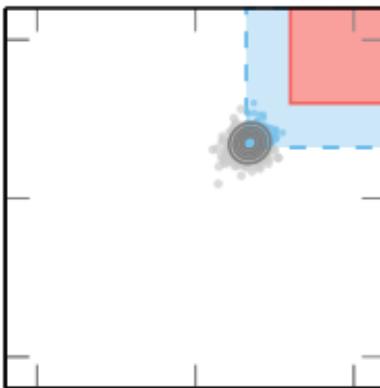
Iteration 1



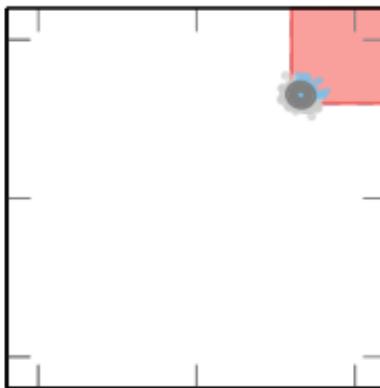
Iteration 2

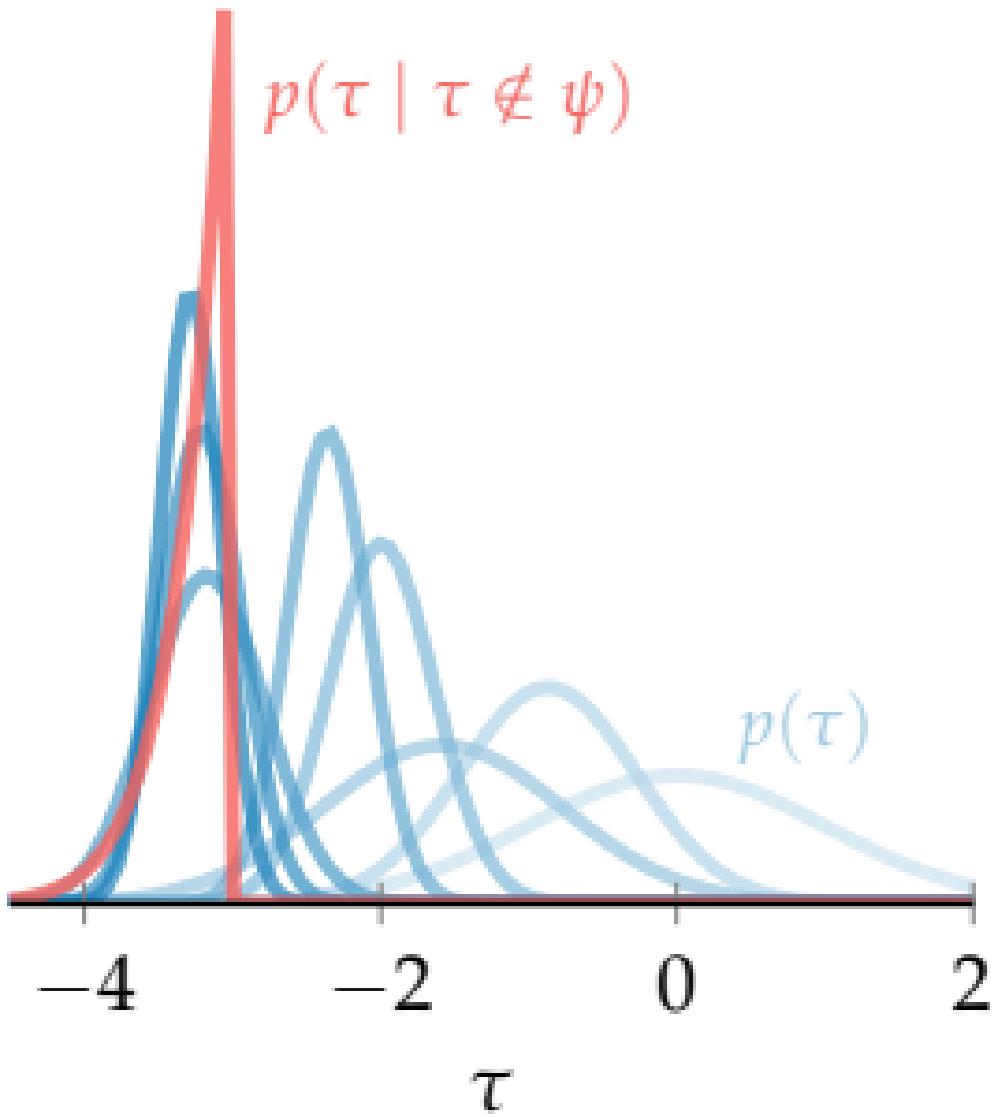


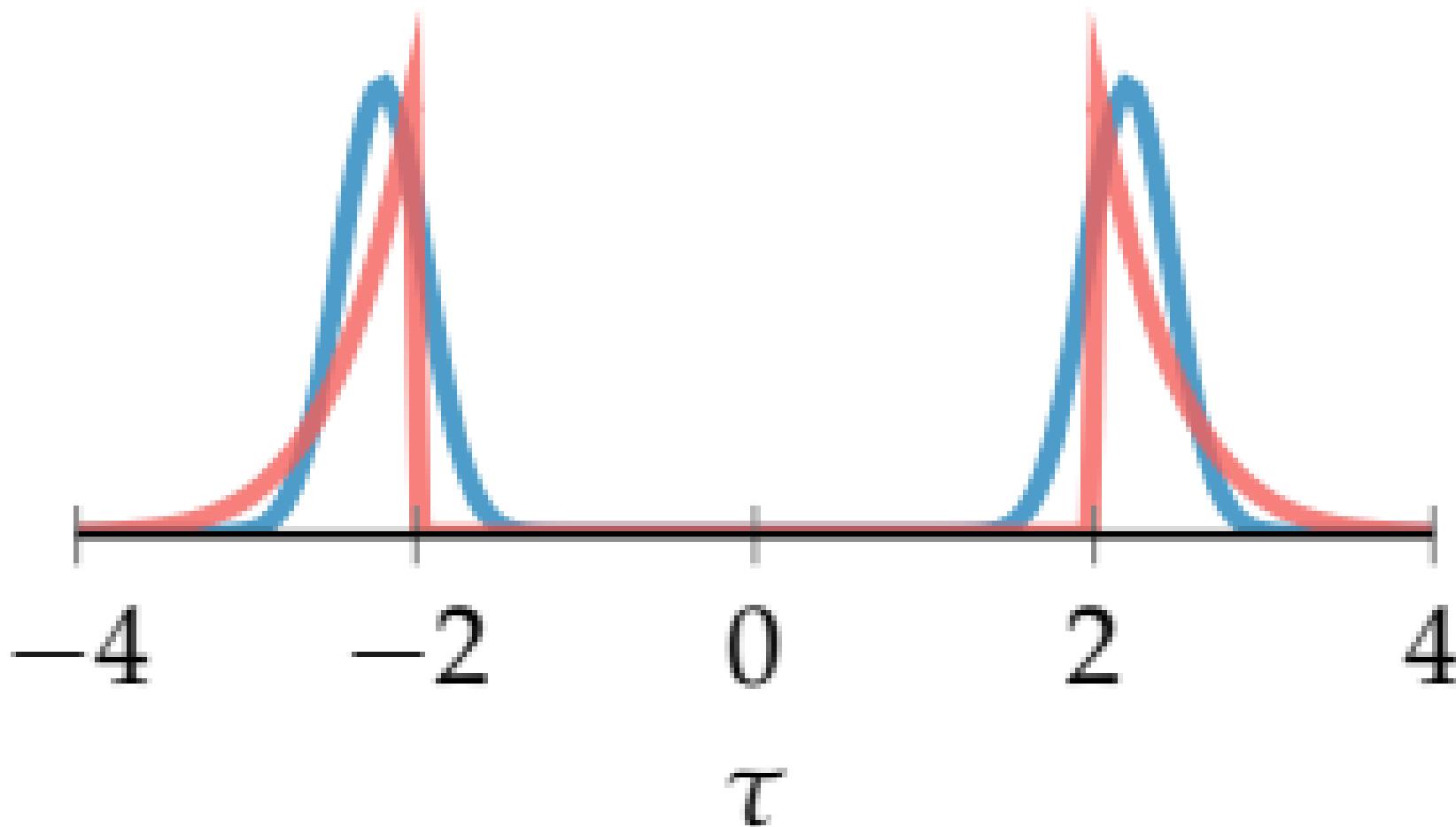
Iteration 5



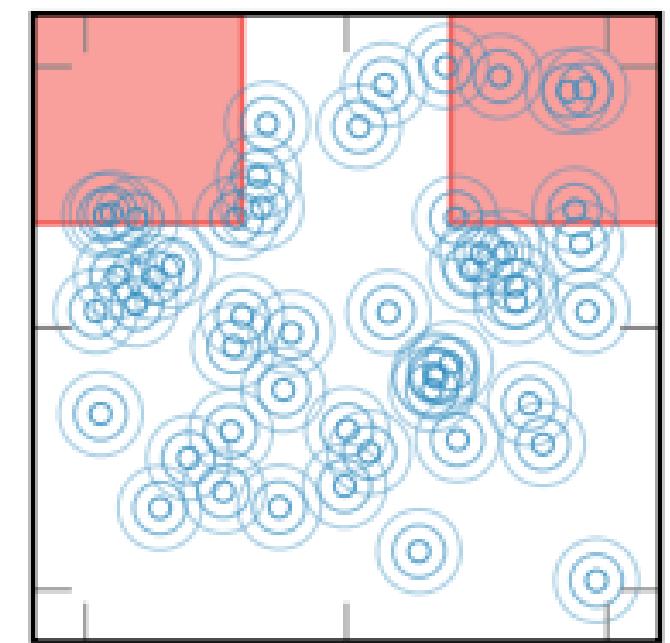
Iteration 20

 $\tau_2$  $\tau_2$  $\tau_1$  $\tau_1$  $\tau_1$  $\tau_1$

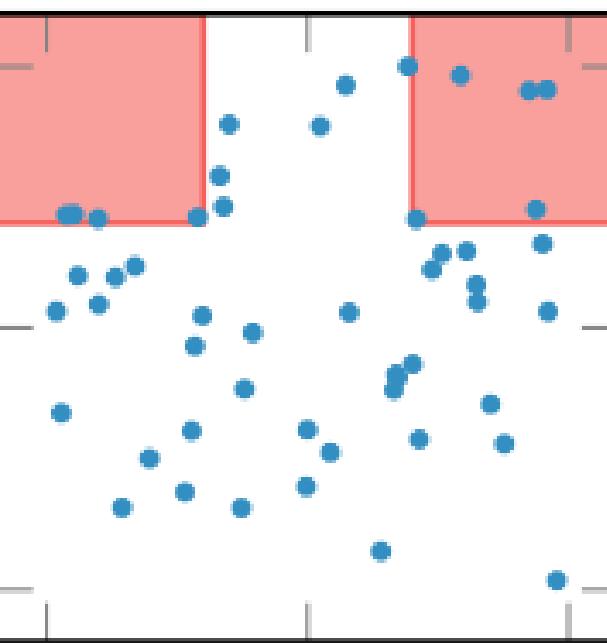




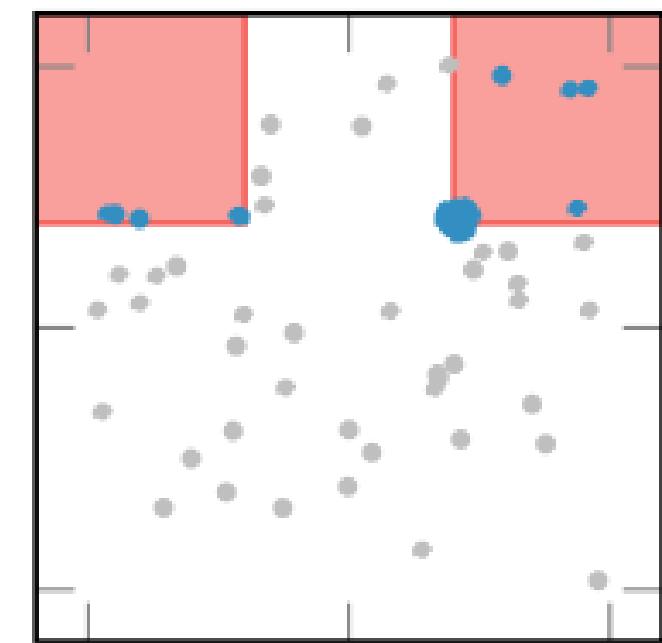
# Initial Proposals



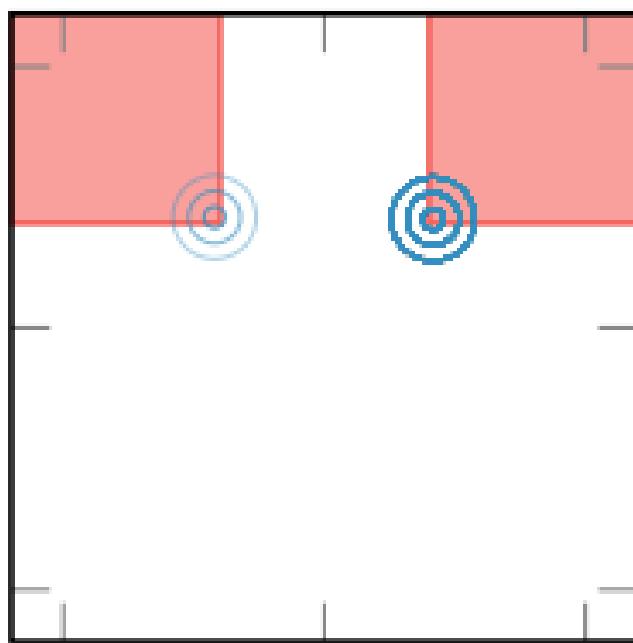
# Sampling



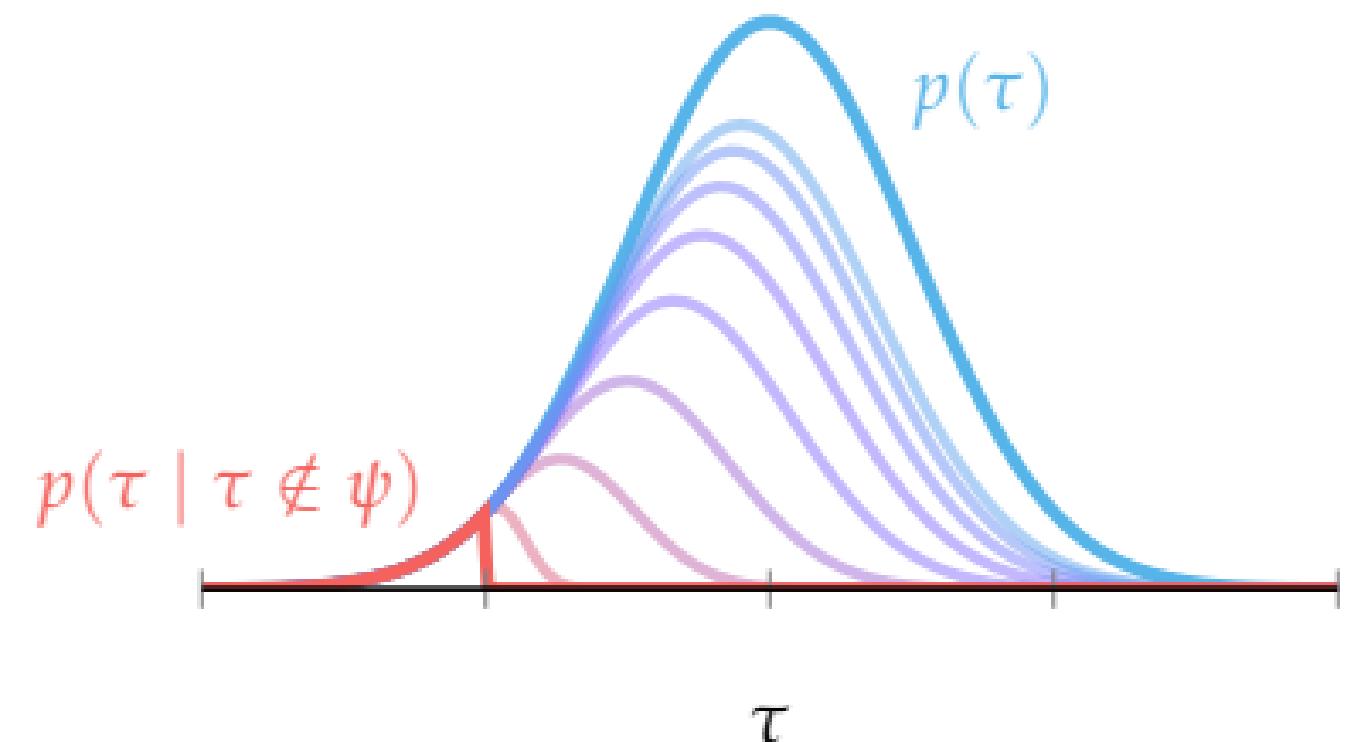
# Weighting



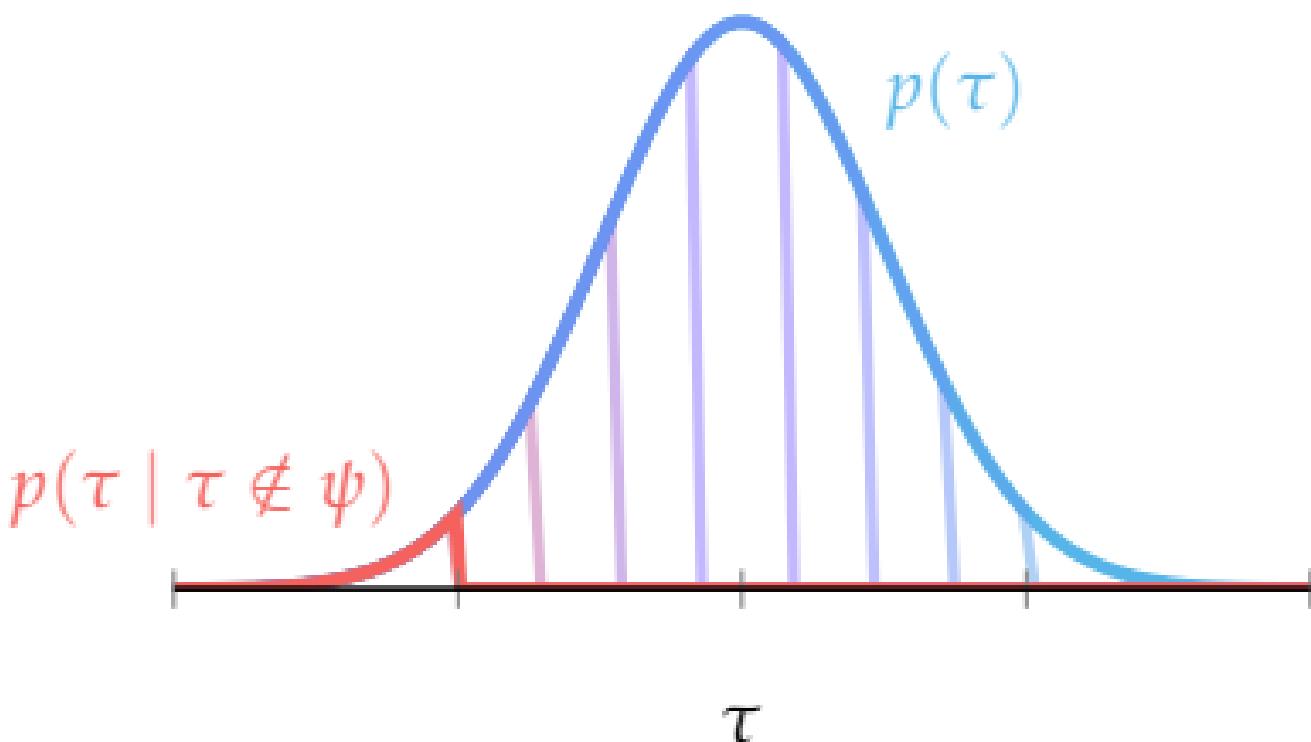
# Resampling

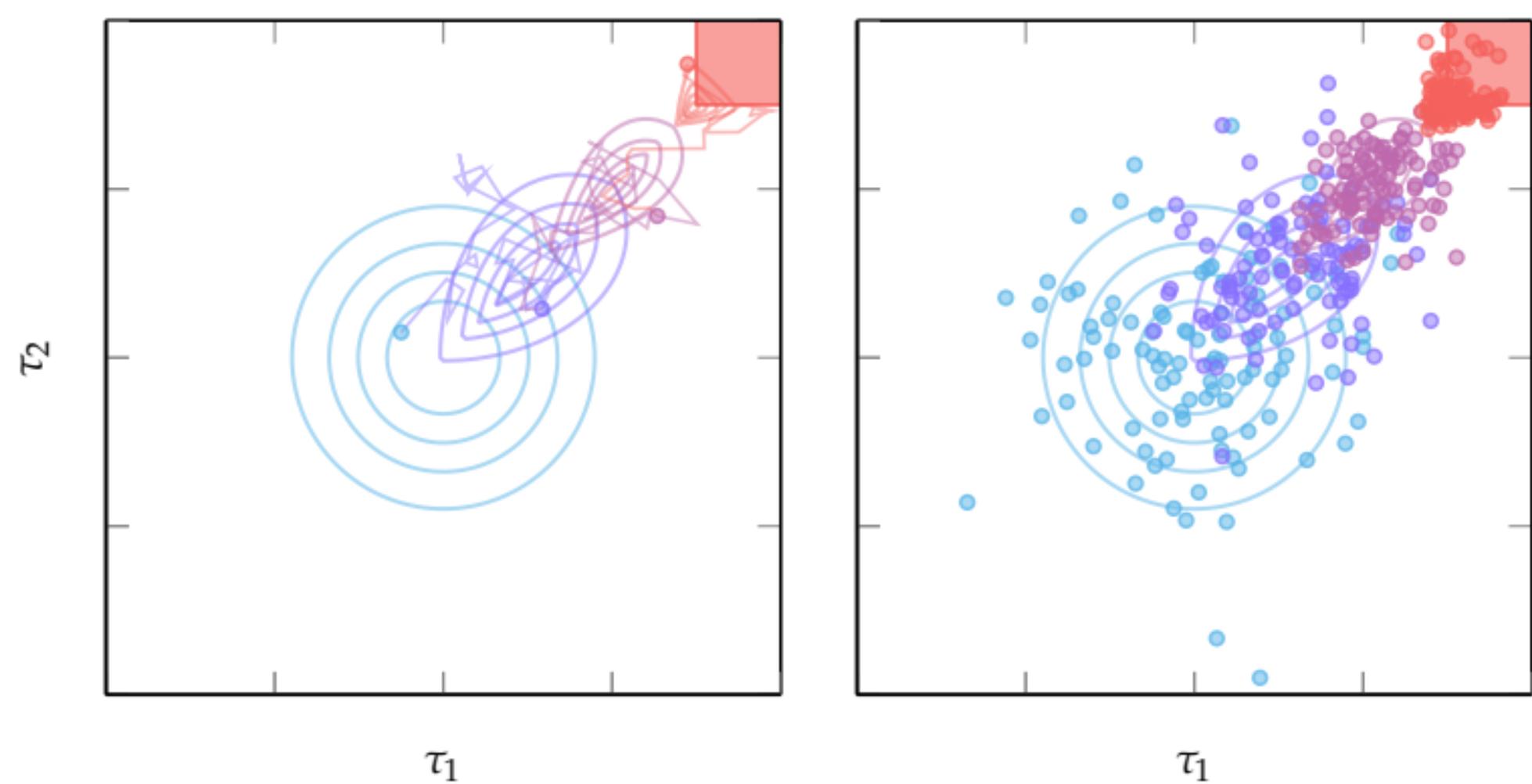


# Smoothing



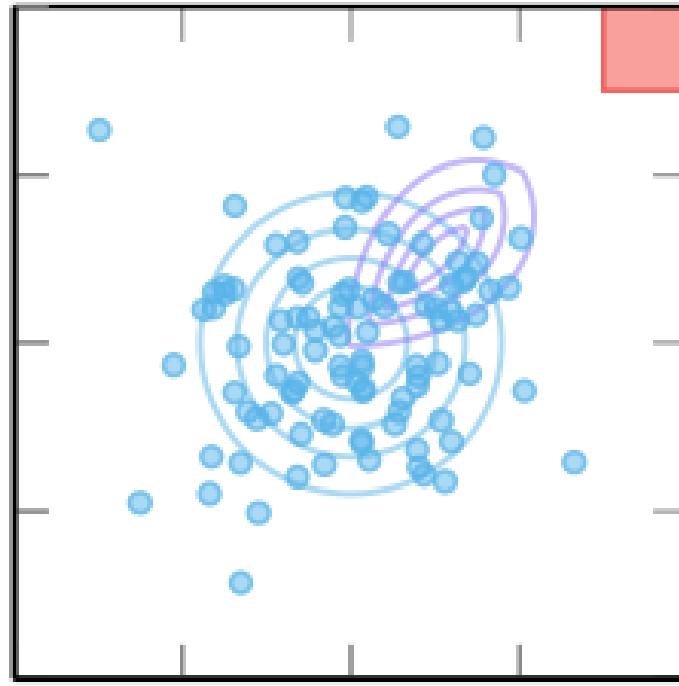
# Thresholding



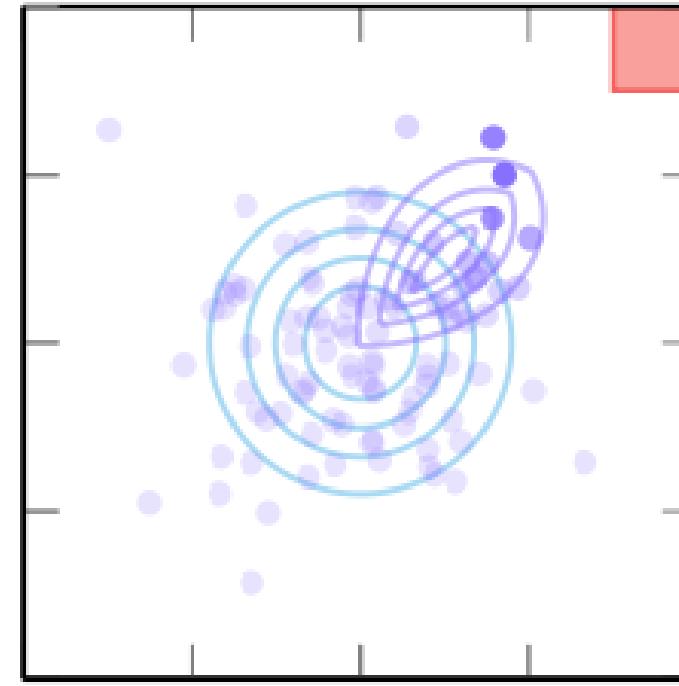


Sample

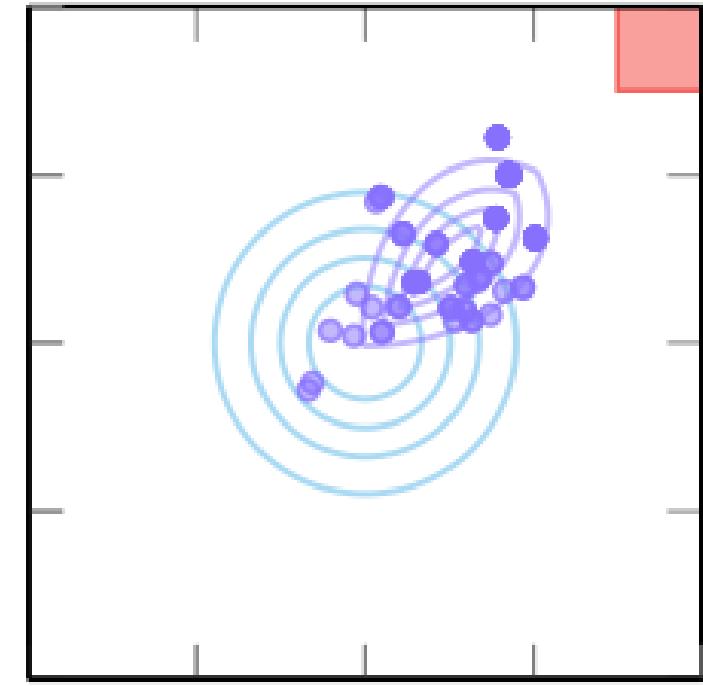
$\tau_2$



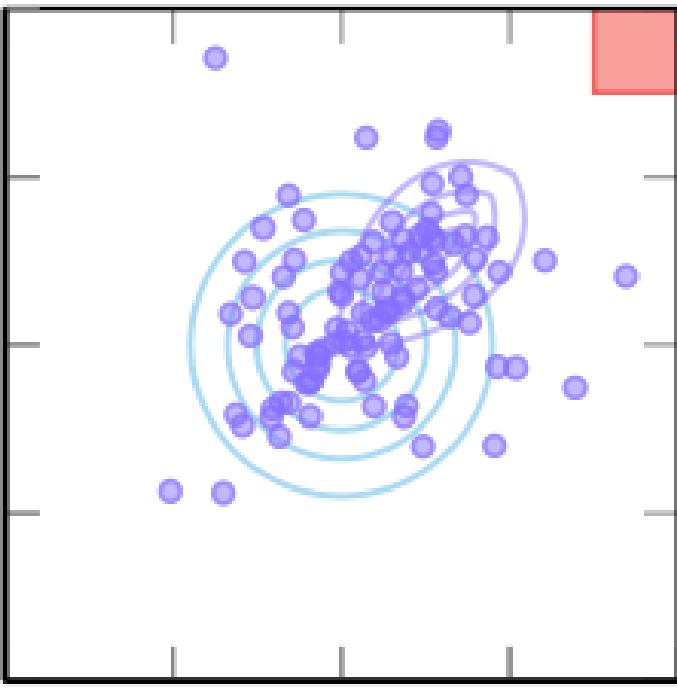
Weight



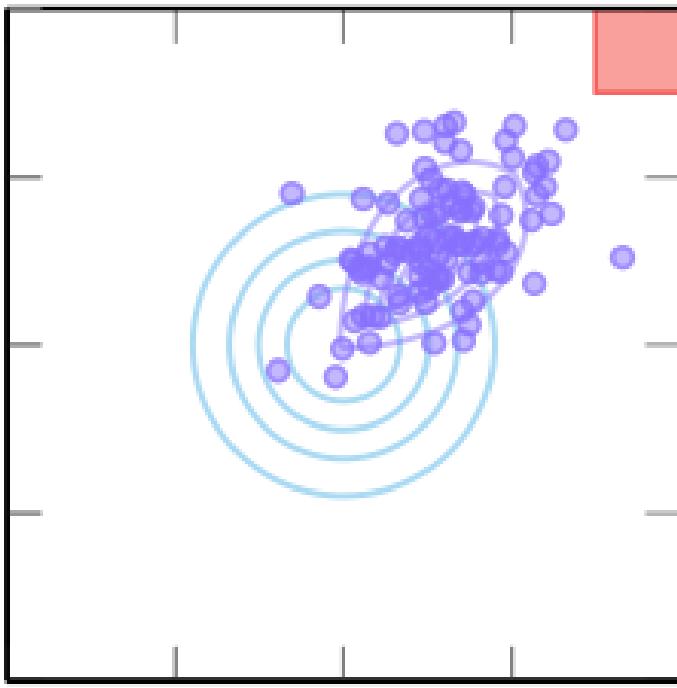
Resample

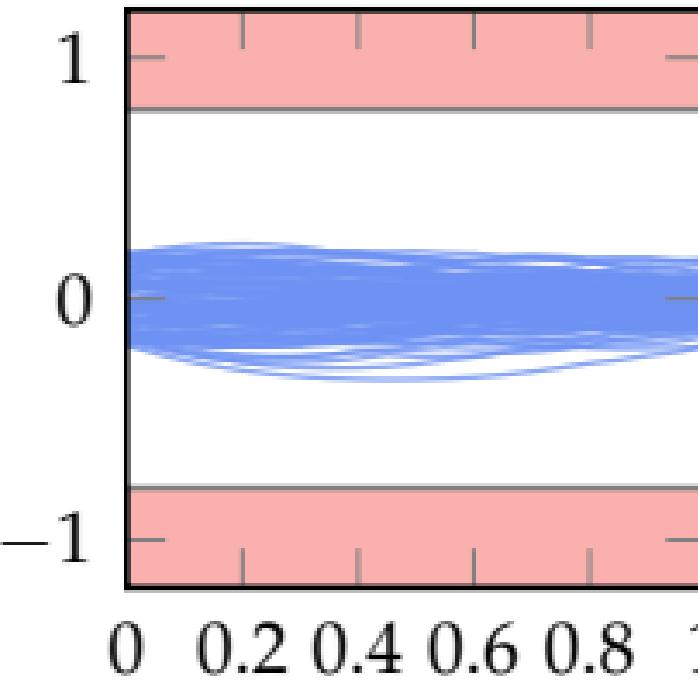
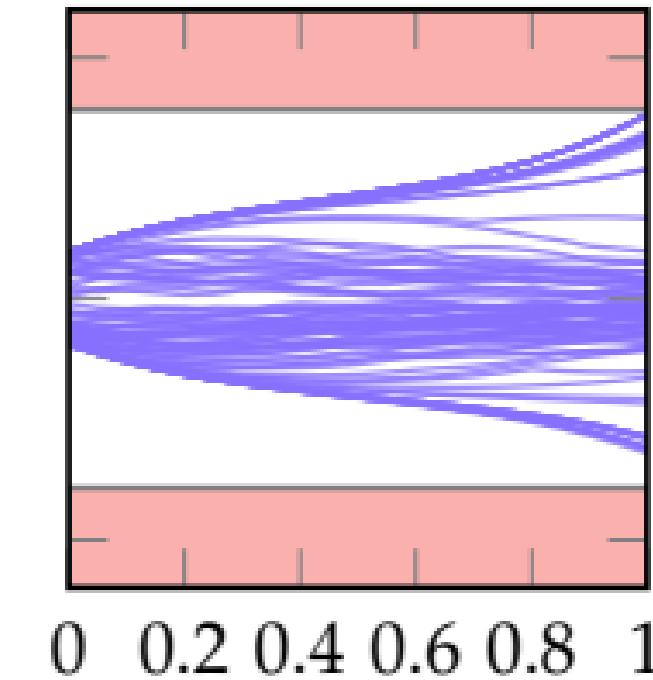
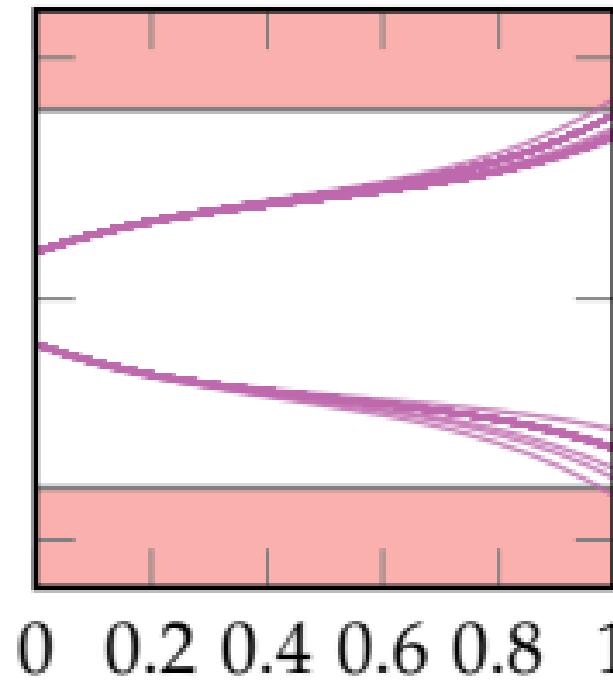
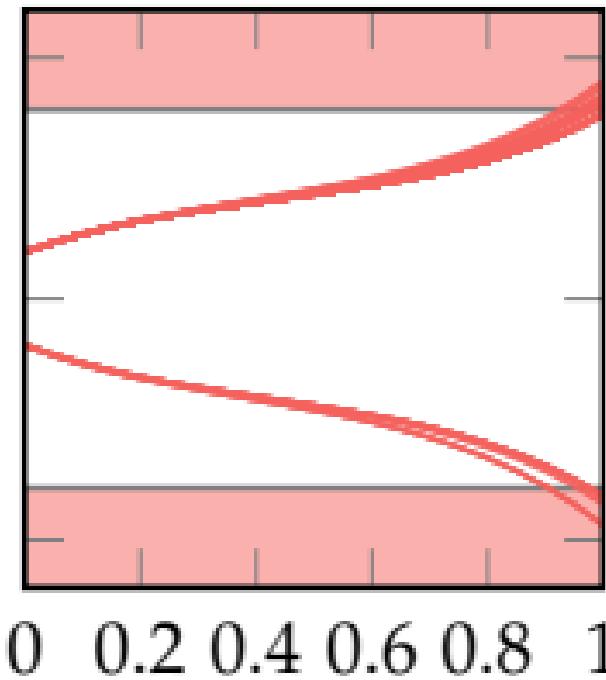


# Without Resampling

 $\tau_2$ 

# With Resampling

 $\tau_1$ 

$\epsilon = 0.5$  $\theta$  (rad) $\epsilon = 0.2$  $\epsilon = 0.1$  $\epsilon = 0.01$ 

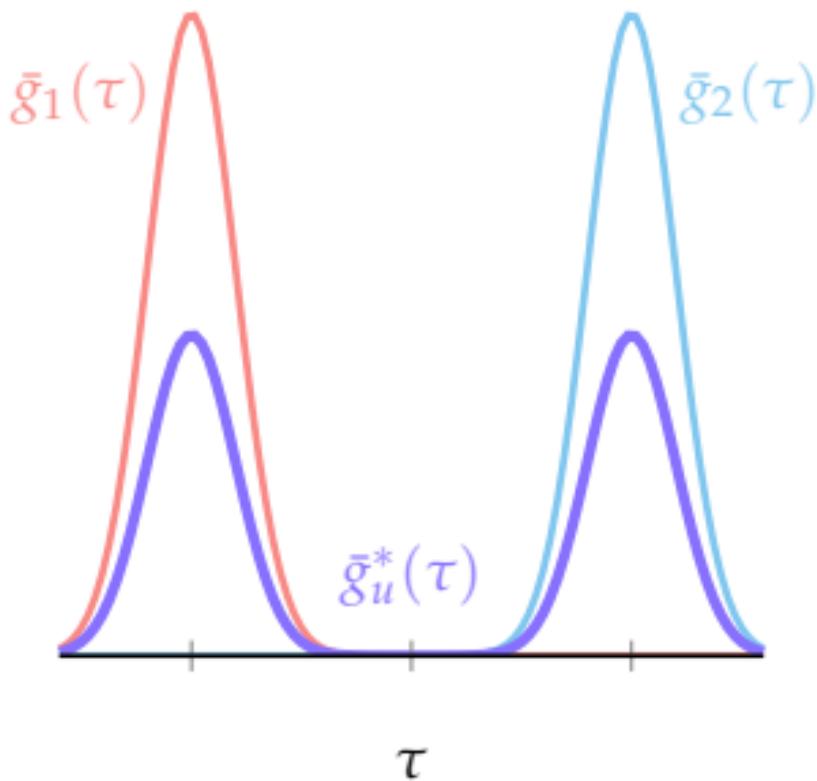
Time (s)

Time (s)

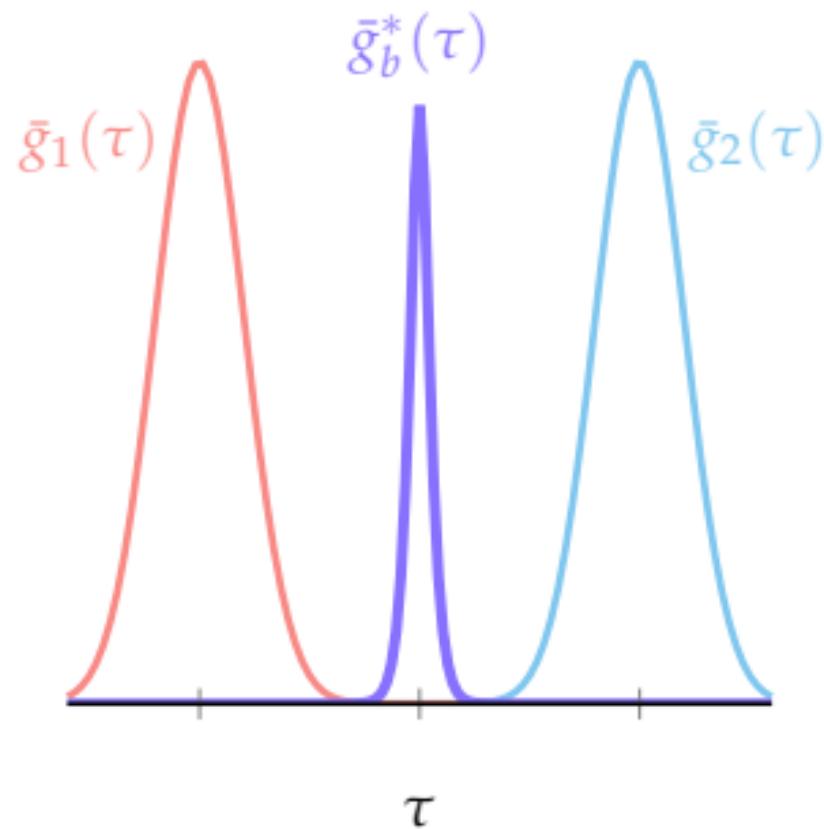
Time (s)

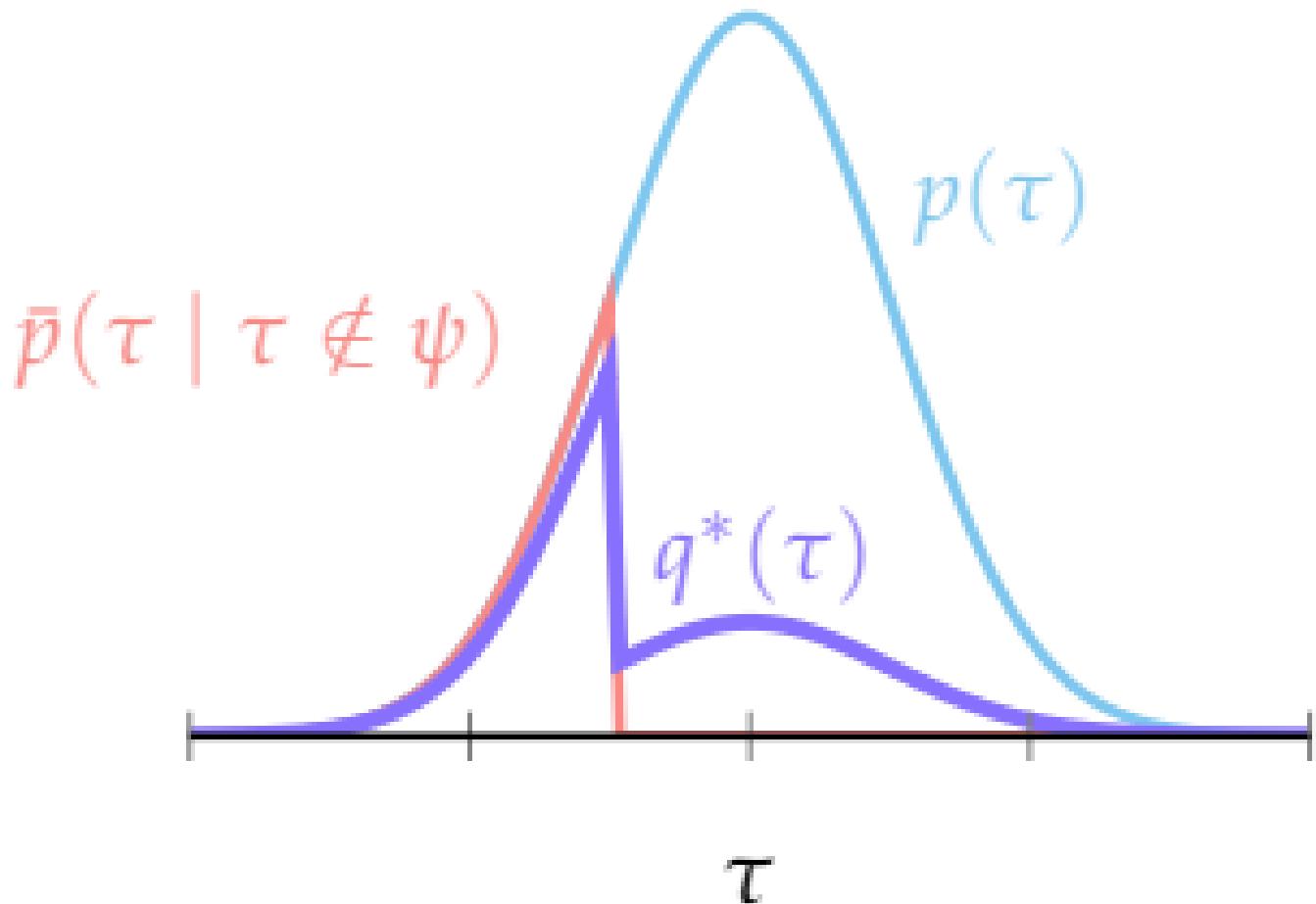
Time (s)

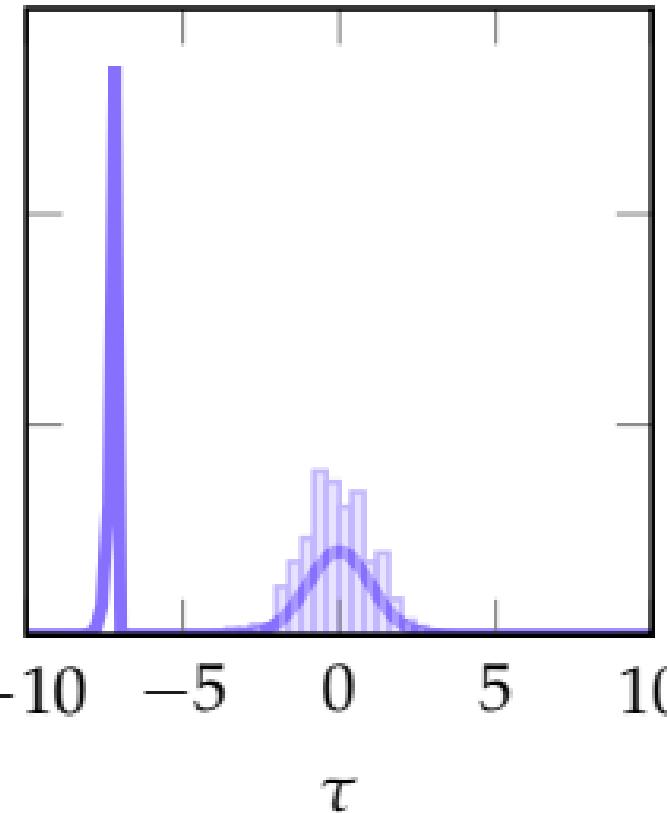
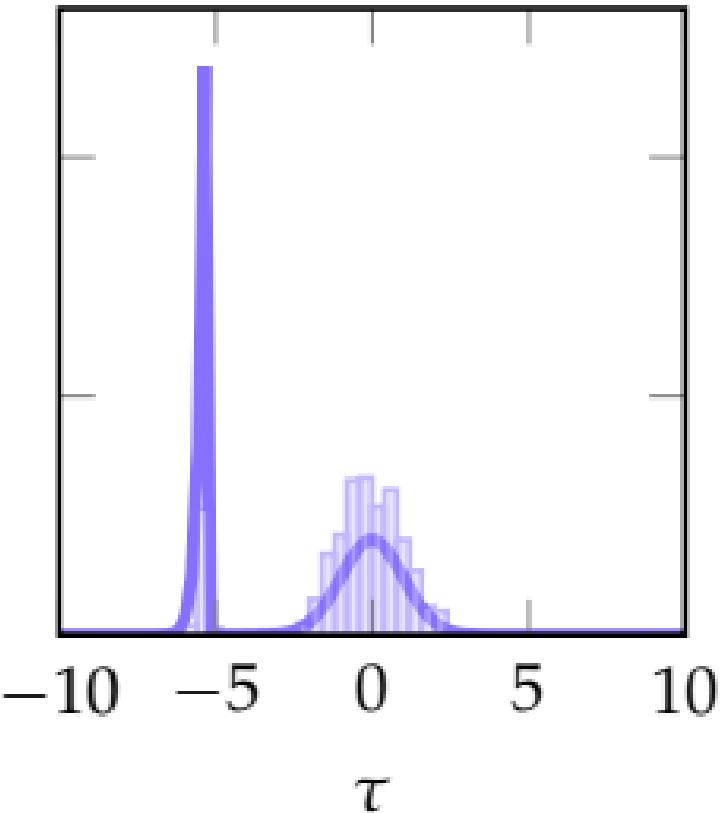
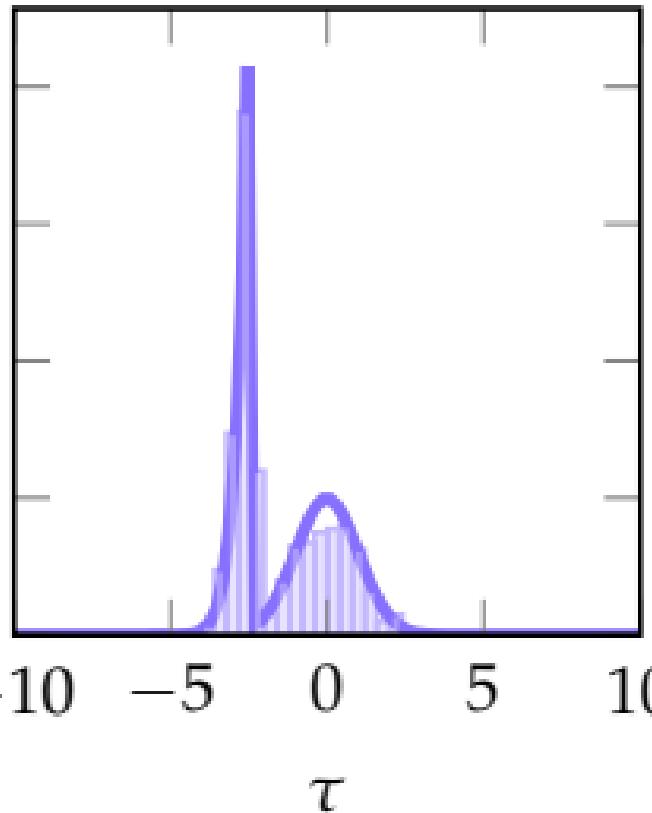
# Optimal Umbrella

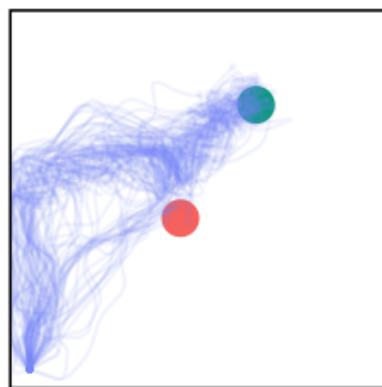
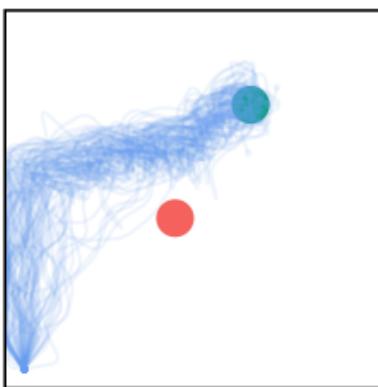
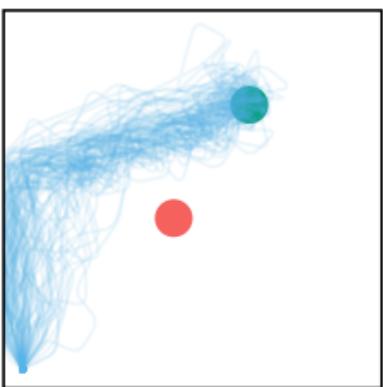
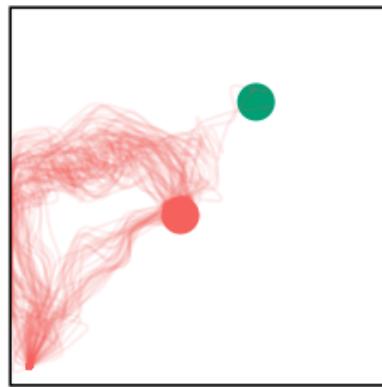
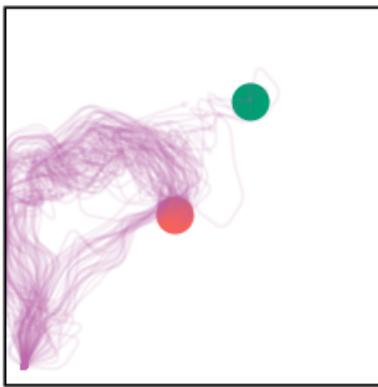
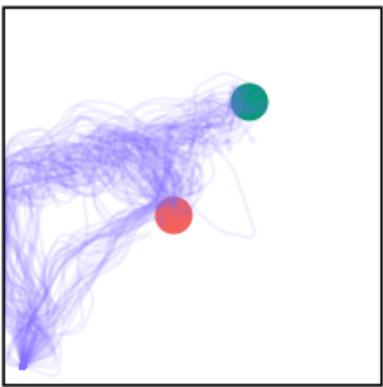


# Optimal Bridge

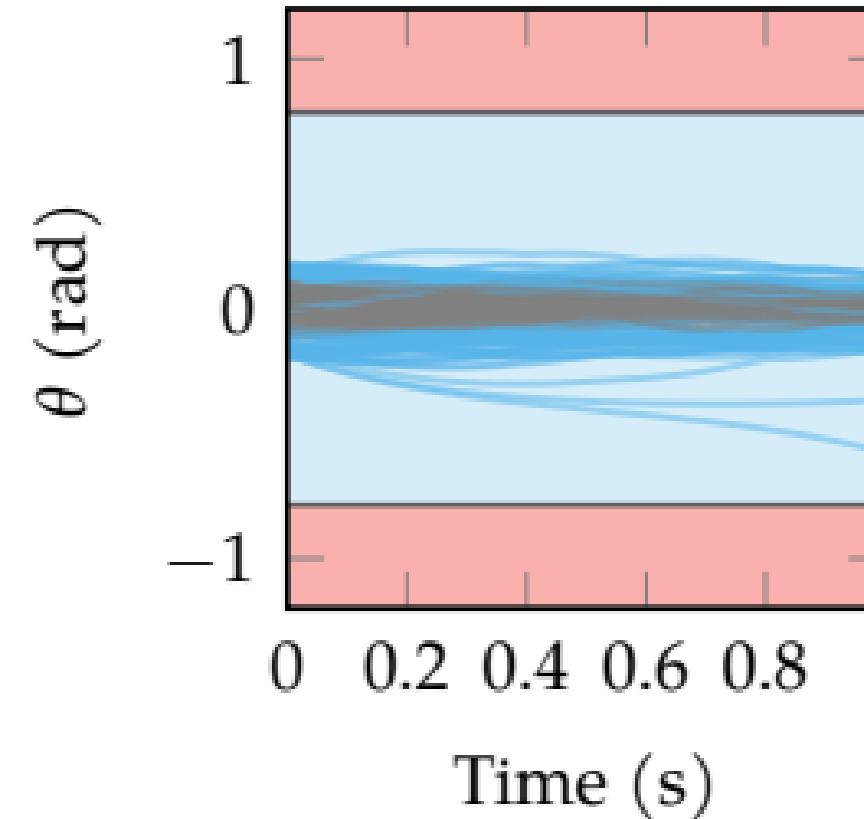




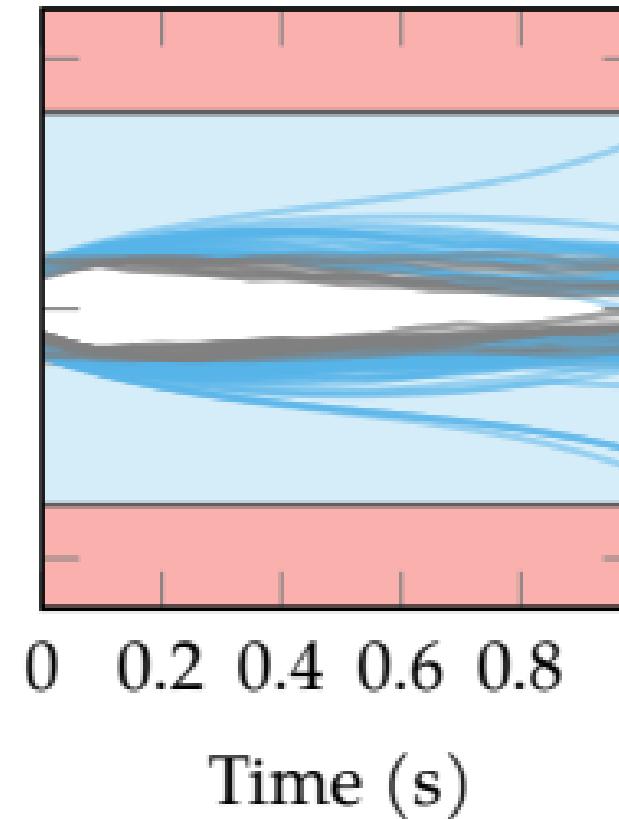
$\alpha = 10^{-2}$  $\alpha = 10^{-7}$  $\alpha = 10^{-12}$ 

$p(\tau)$  $\epsilon = 0.5$  $\epsilon = 0.2$  $\epsilon = 0.1$  $\epsilon = 0.01$  $p(\tau \mid \tau \notin \psi)$ 

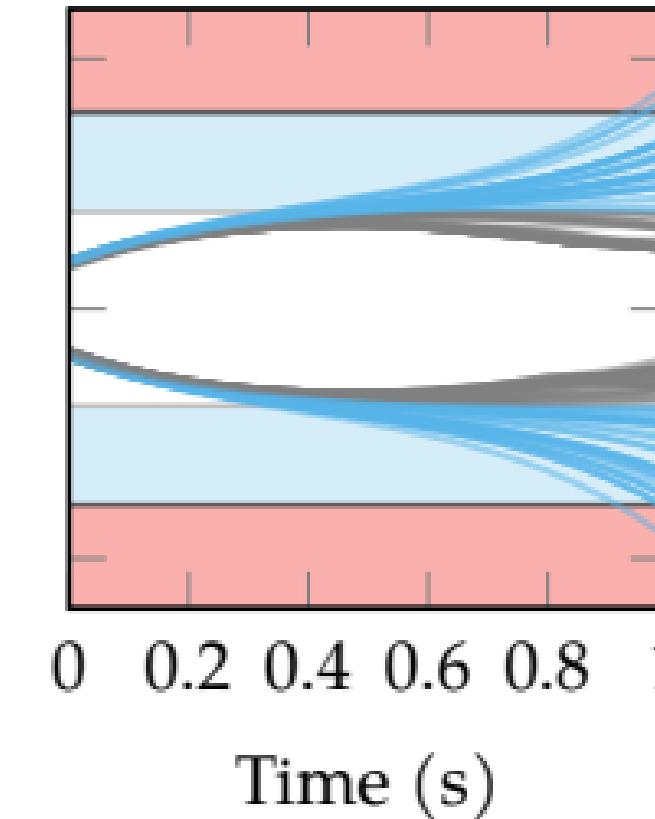
Iteration 1



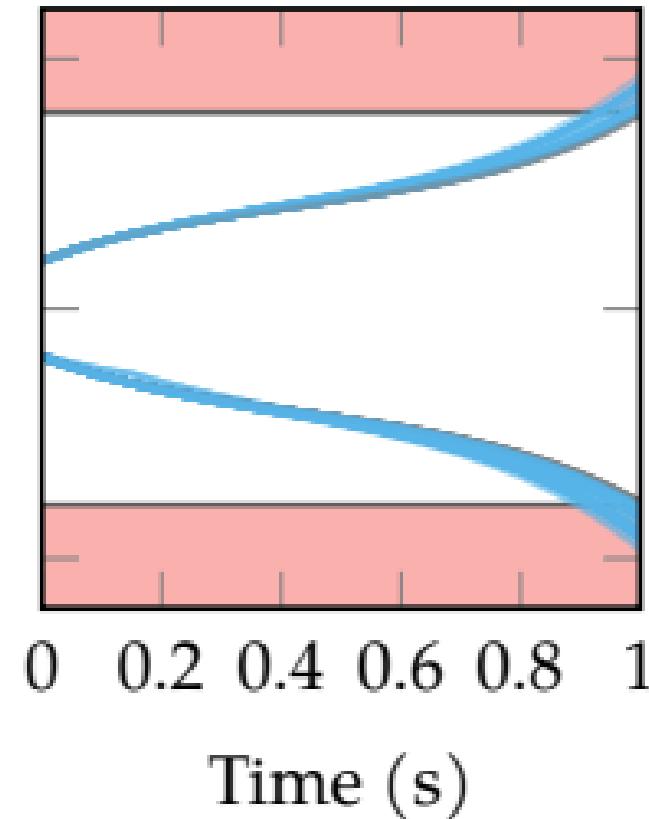
Iteration 4



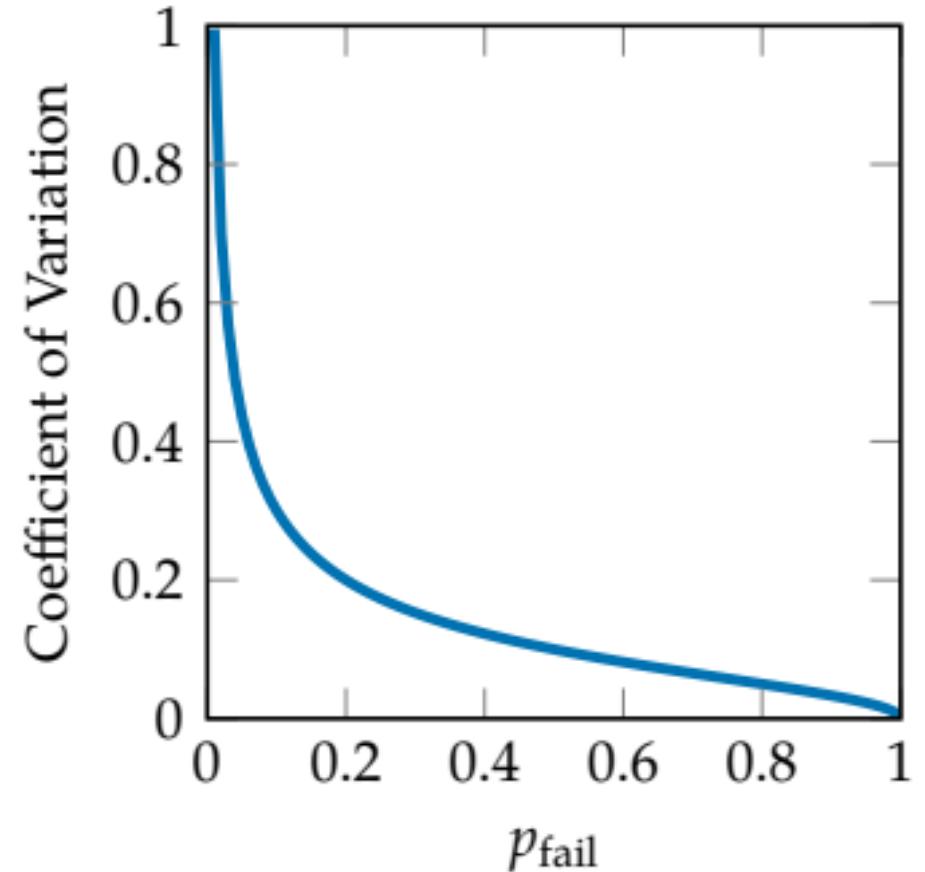
Iteration 8



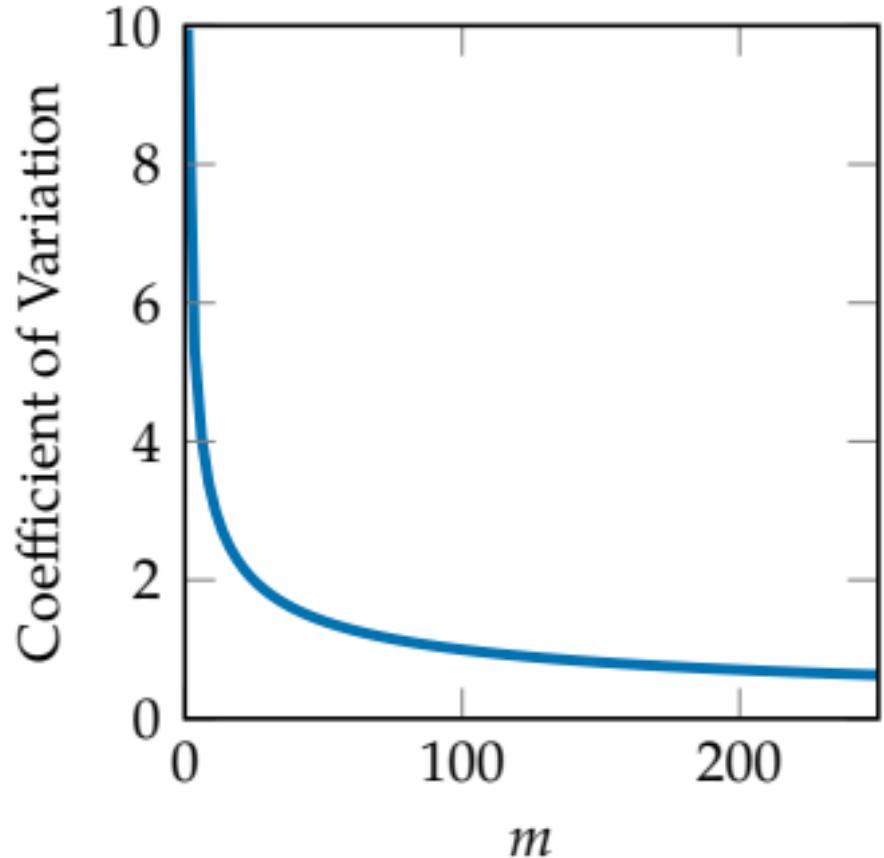
Iteration 12

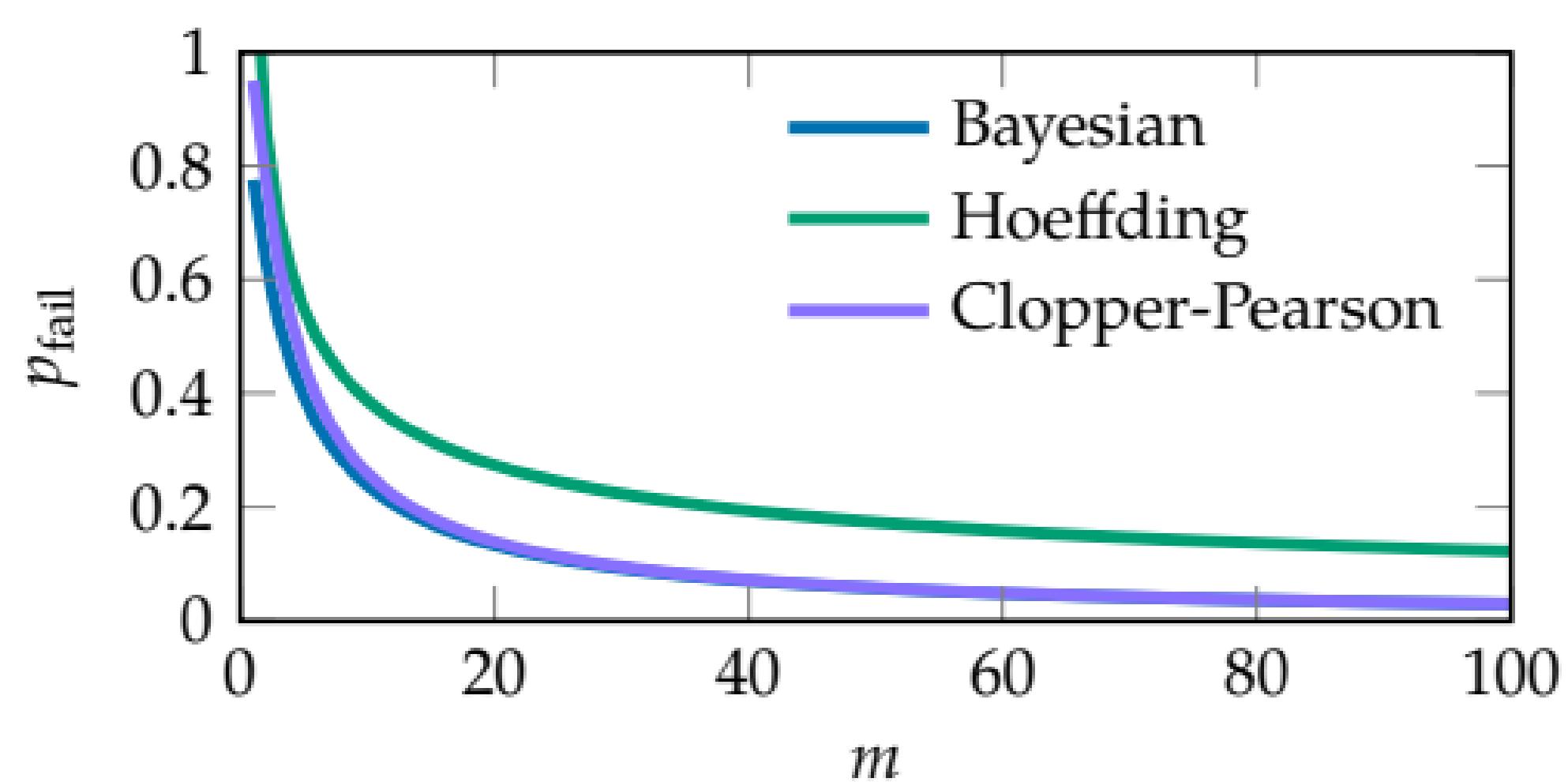


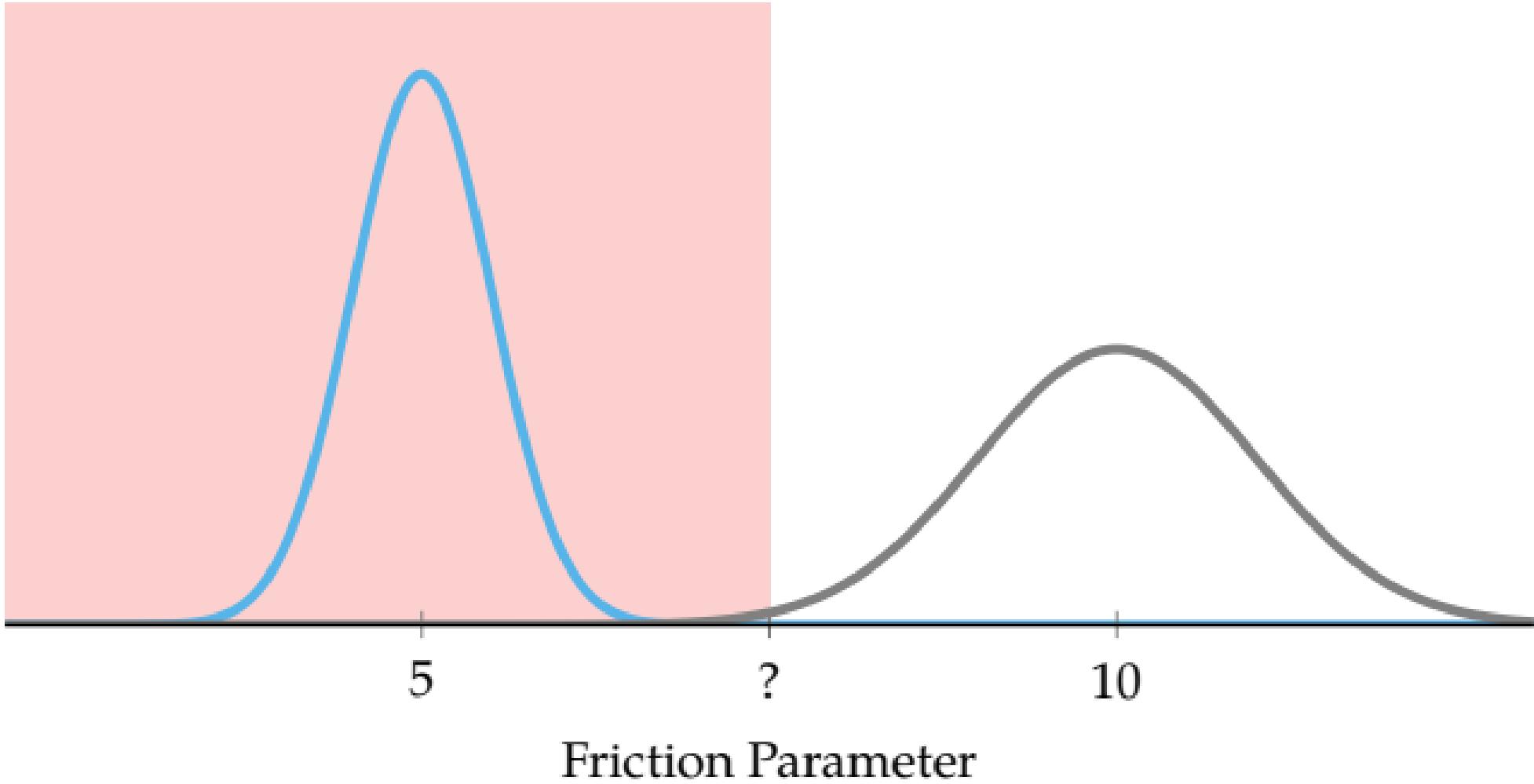
$m = 100$



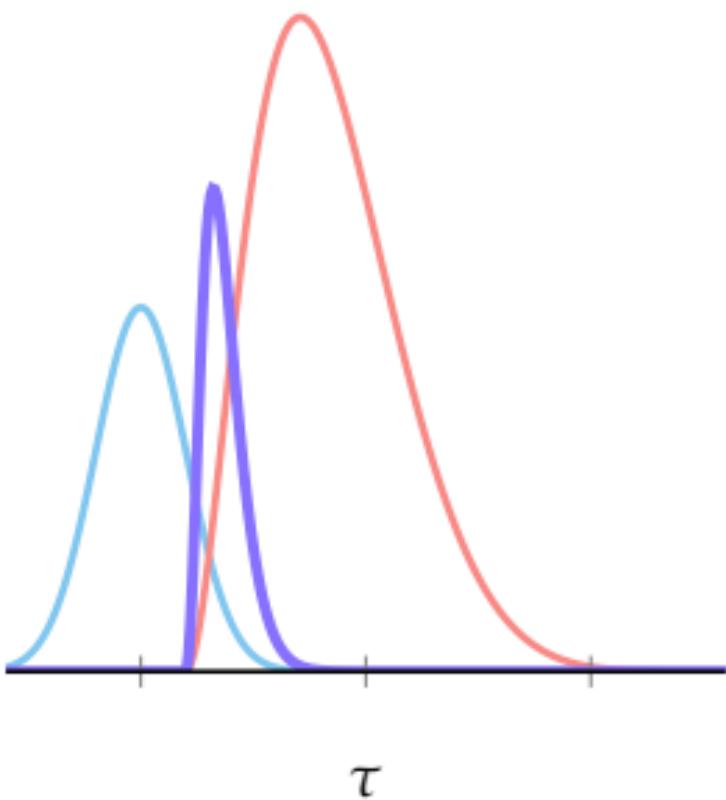
$p_{\text{fail}} = 0.01$



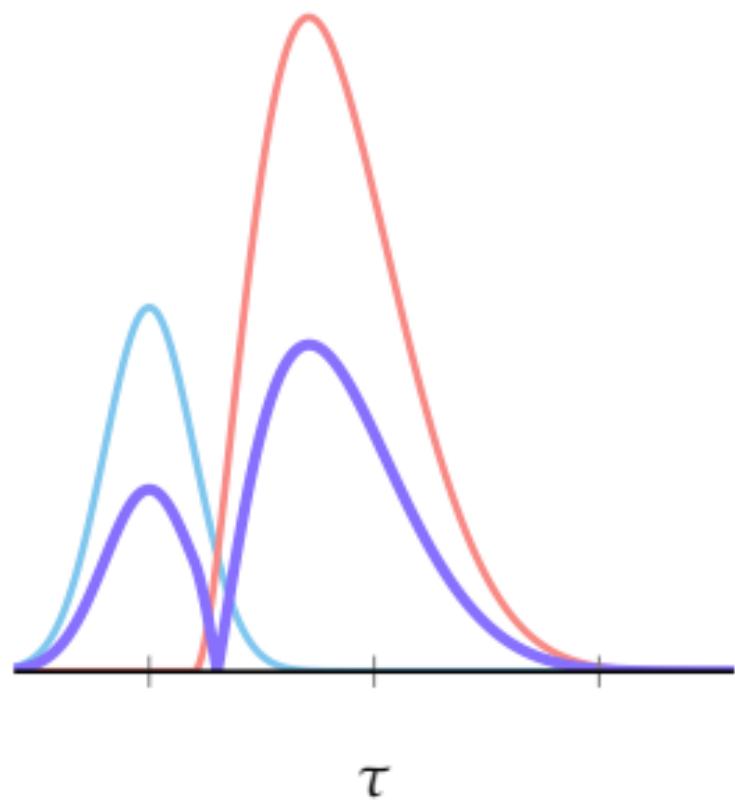


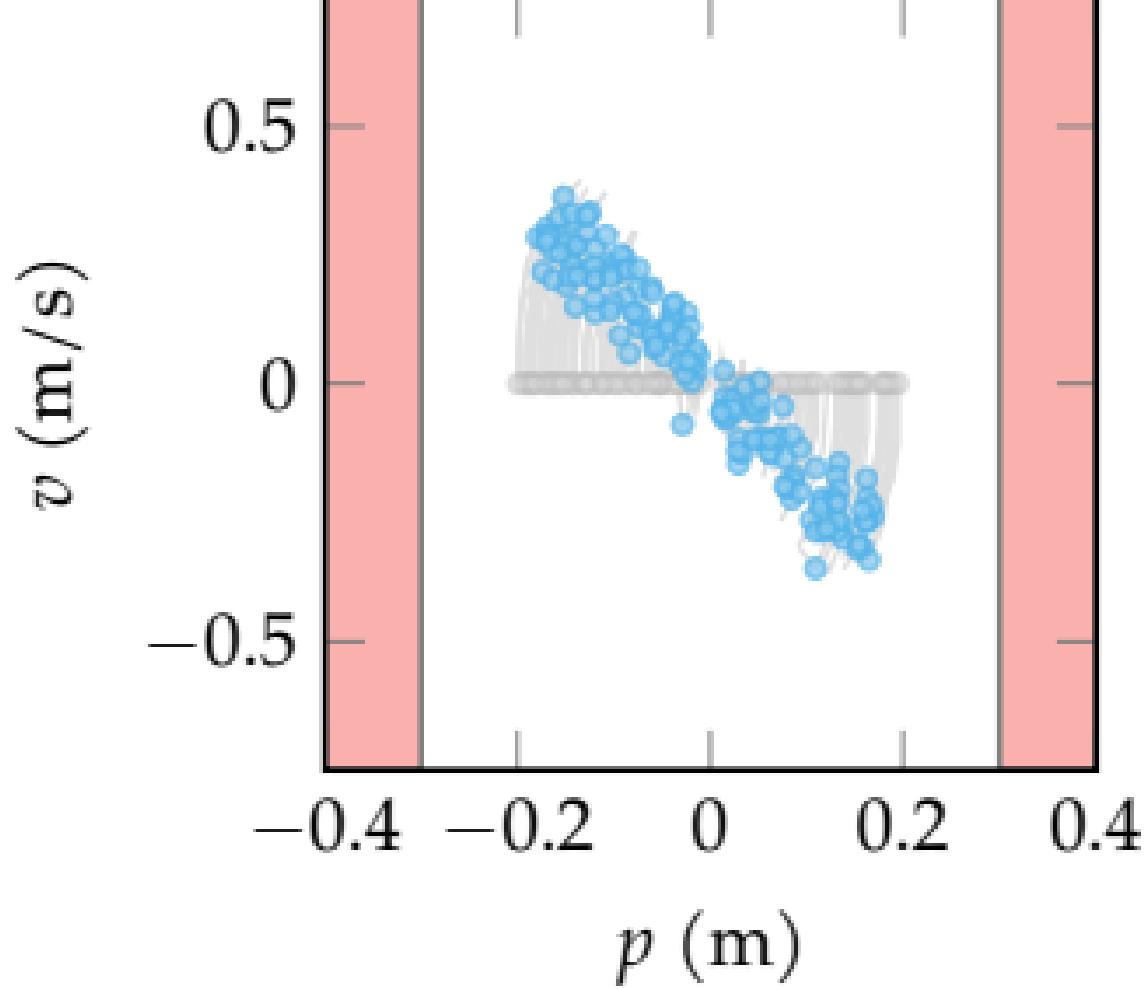


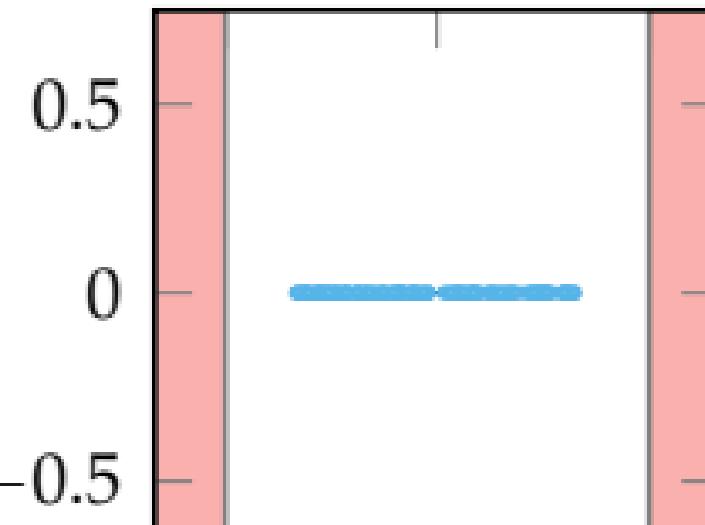
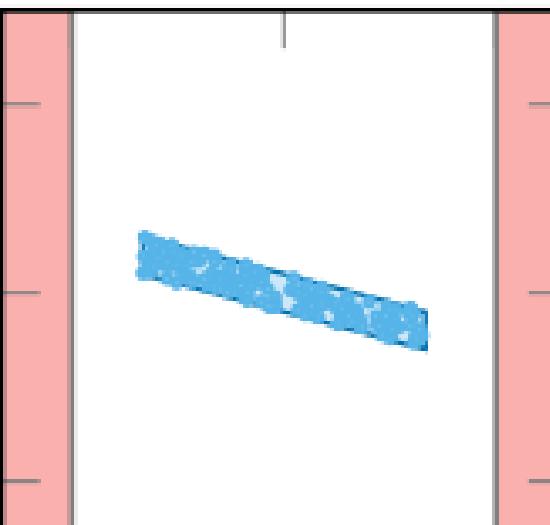
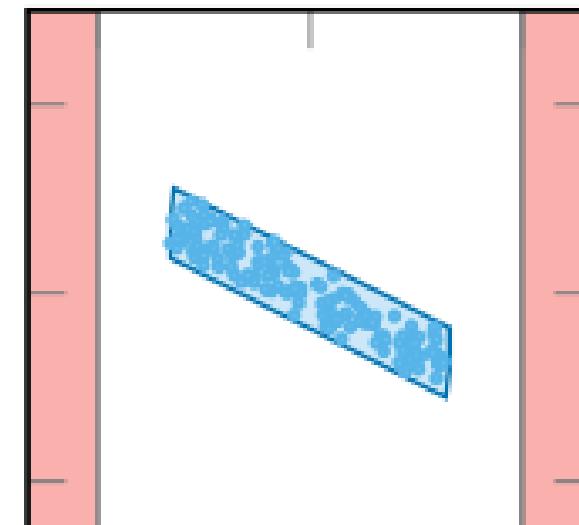
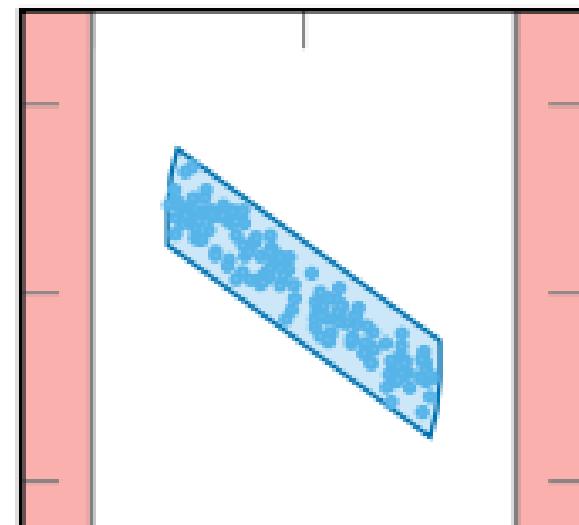
A



B



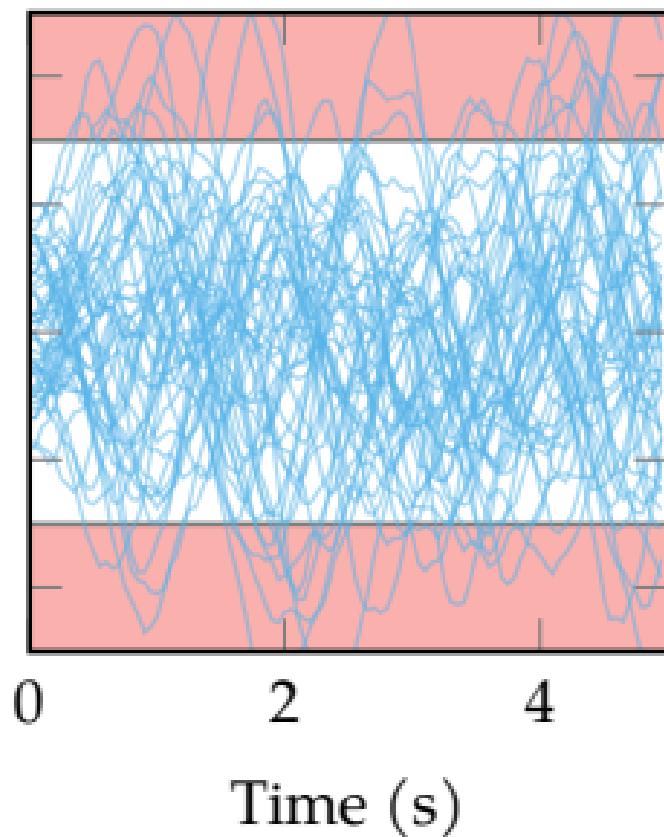
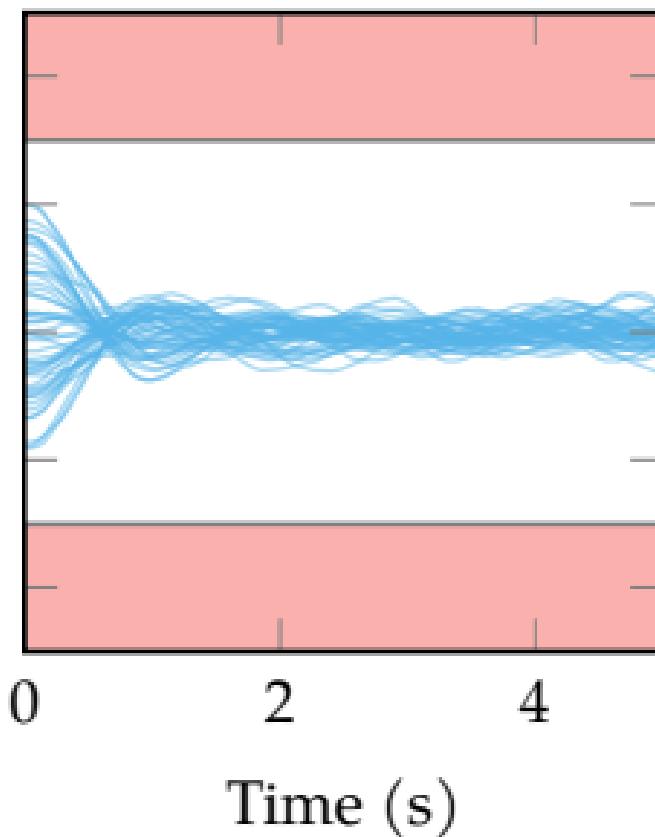
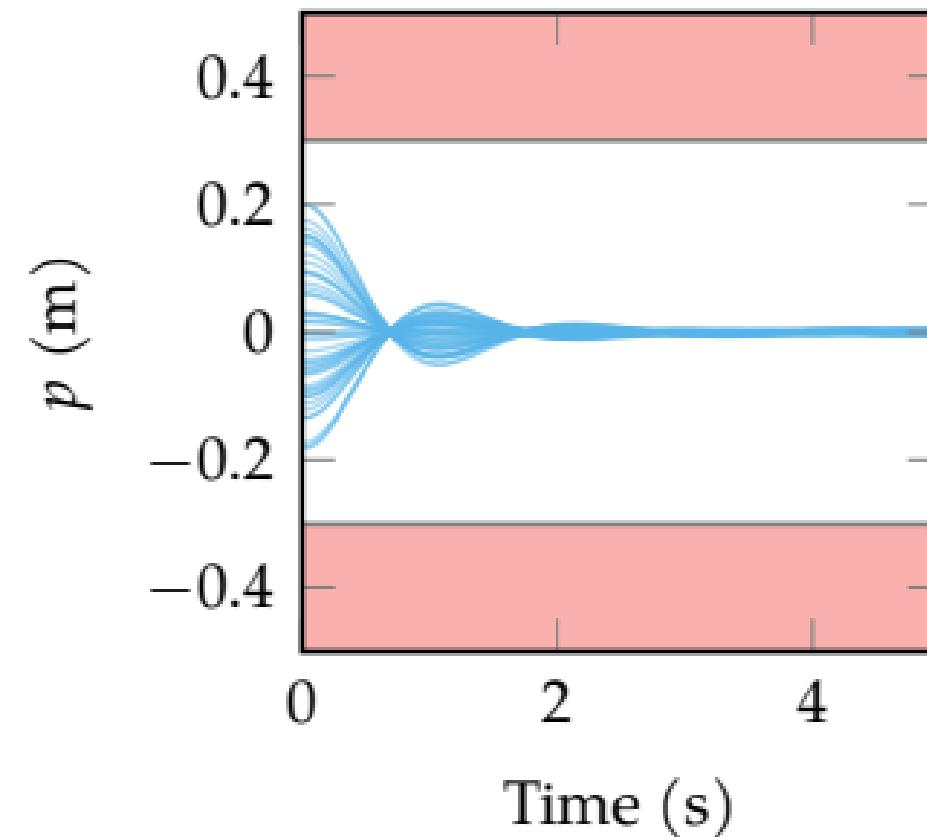


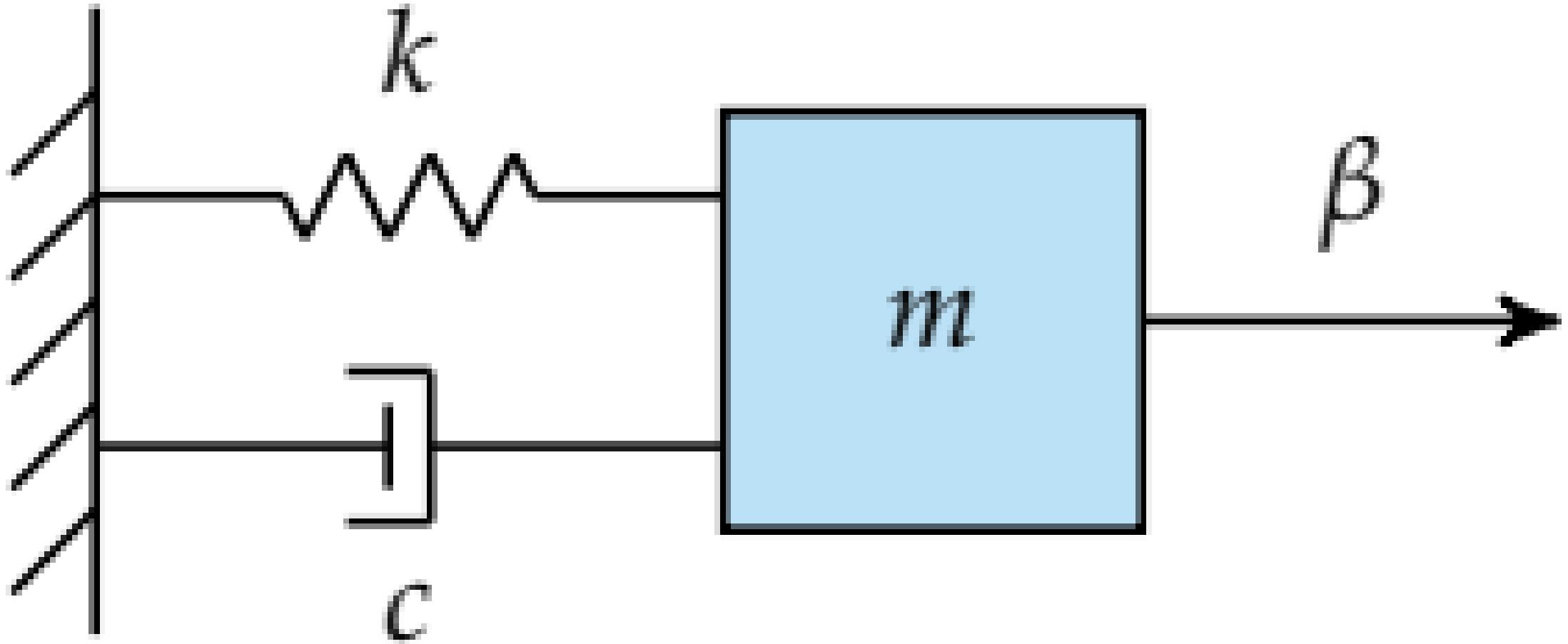
$\mathcal{R}_1 = \mathcal{S}$  $v \text{ (m/s)}$  $\mathcal{R}_2$  $p \text{ (m)}$  $\mathcal{R}_3$  $p \text{ (m)}$  $\mathcal{R}_4$  $p \text{ (m)}$ 

$$-0.1 \leq \mathbf{x}_0 \leq 0.1$$

$$-1 \leq \mathbf{x}_0 \leq 1$$

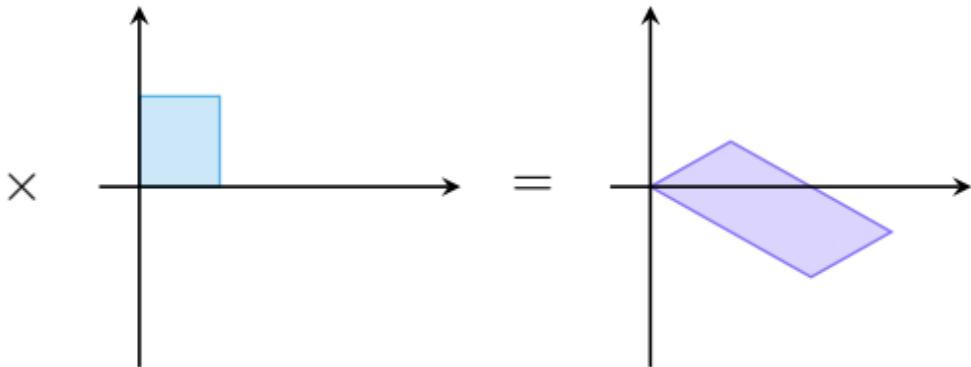
$$-10 \leq \mathbf{x}_0 \leq 10$$



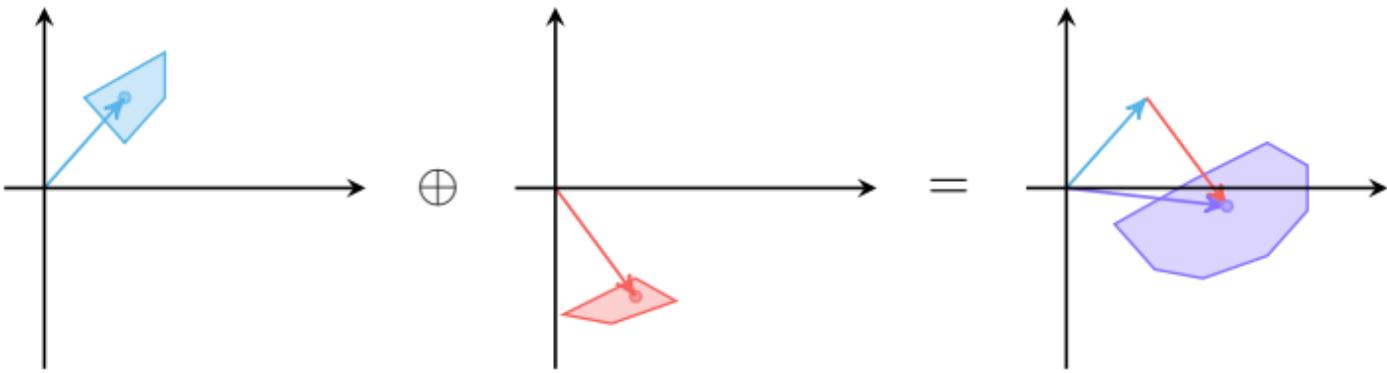


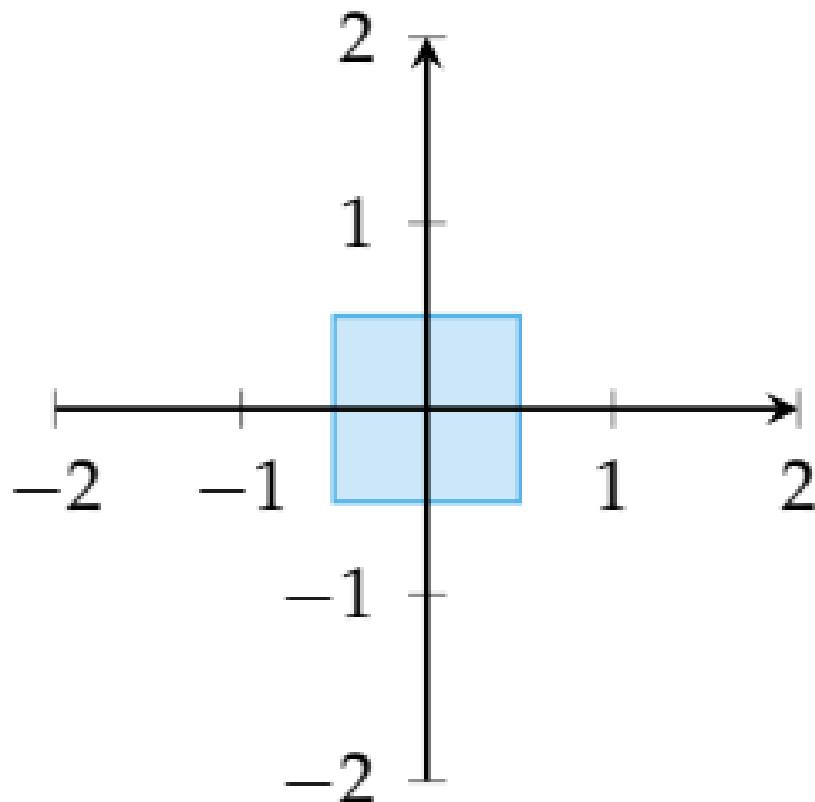
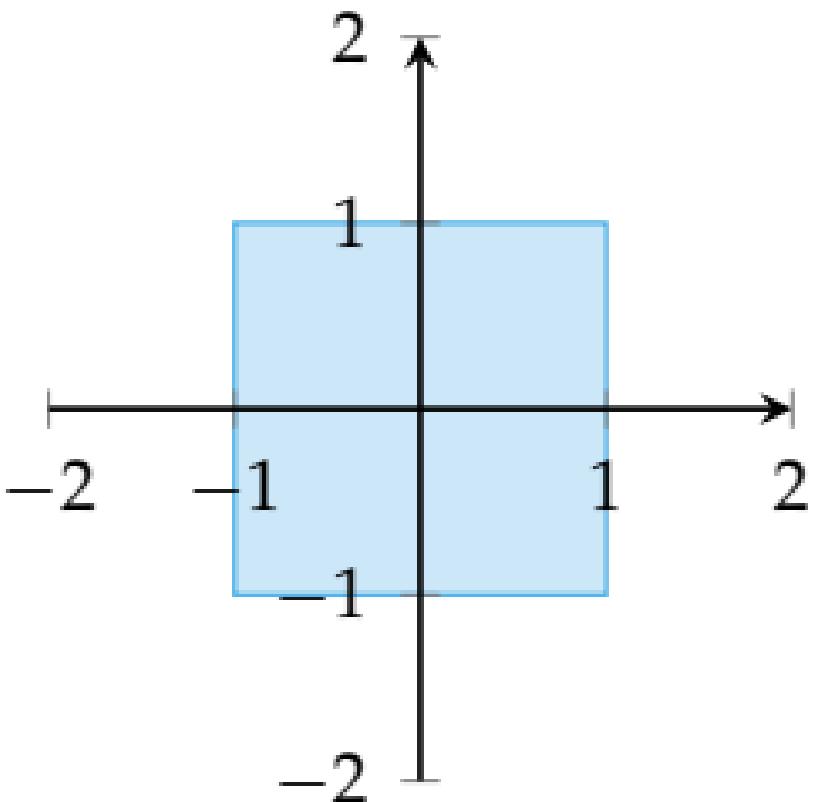
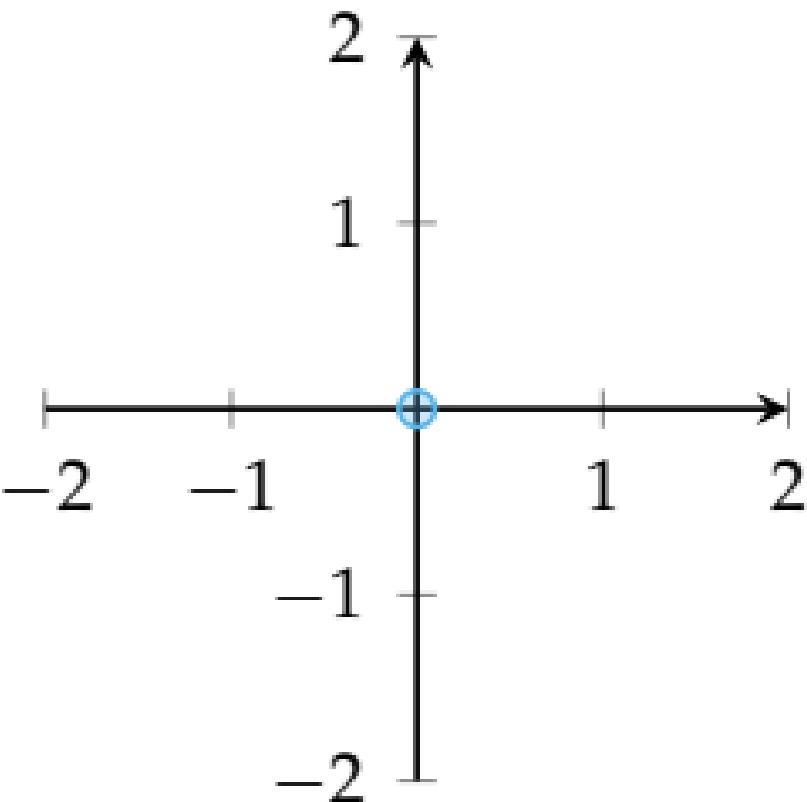
# Linear Transformation

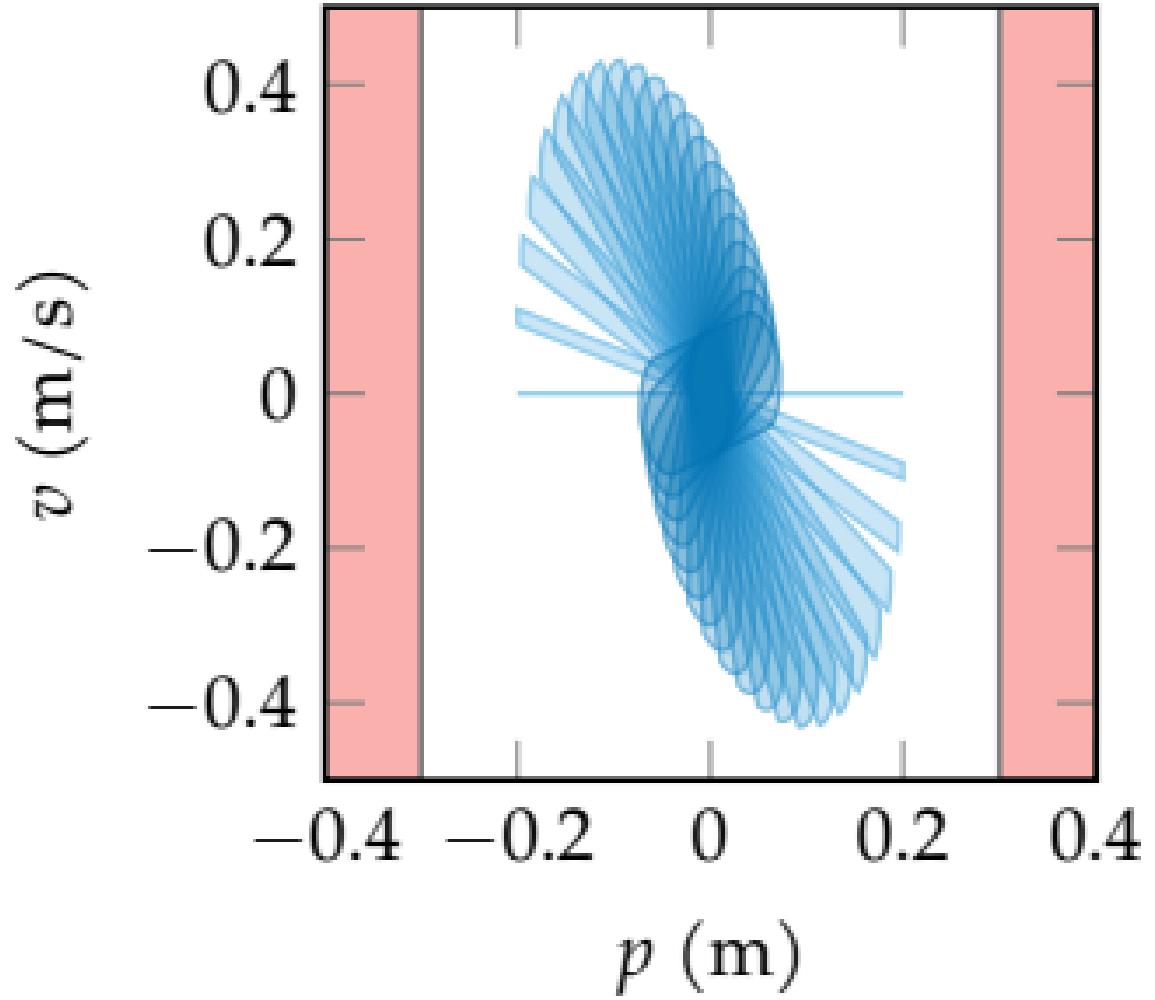
$$\begin{bmatrix} 1.0 & 2.0 \\ 0.5 & -1.0 \end{bmatrix}$$



# Minkowski Sum



$\mathcal{S}$  $\mathcal{S}'$  (no simplification)True  $\mathcal{S}'$ 

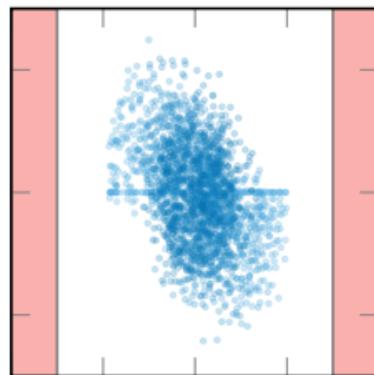
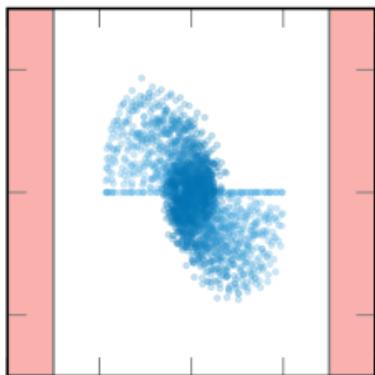
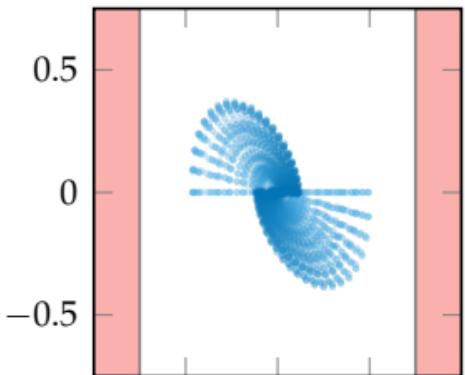


$$-0.1 \leq x_0 \leq 0.1$$

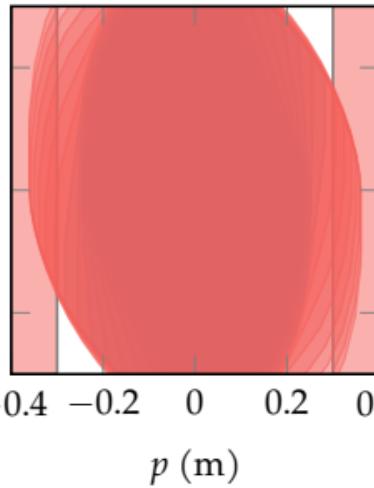
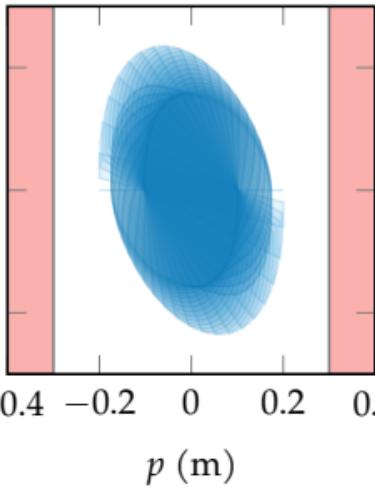
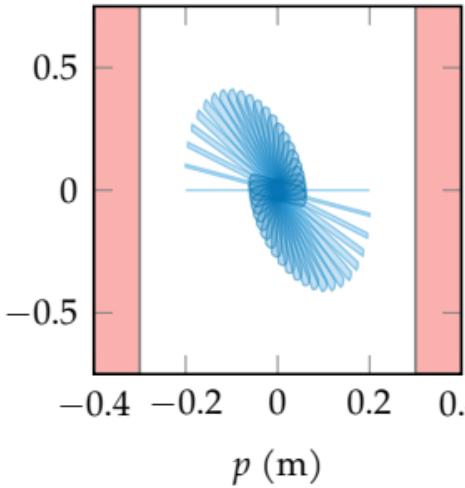
$$-1.0 \leq x_0 \leq 1.0$$

$$-2.5 \leq x_0 \leq 2.5$$

$a$  (m/s)



$a$  (m/s)



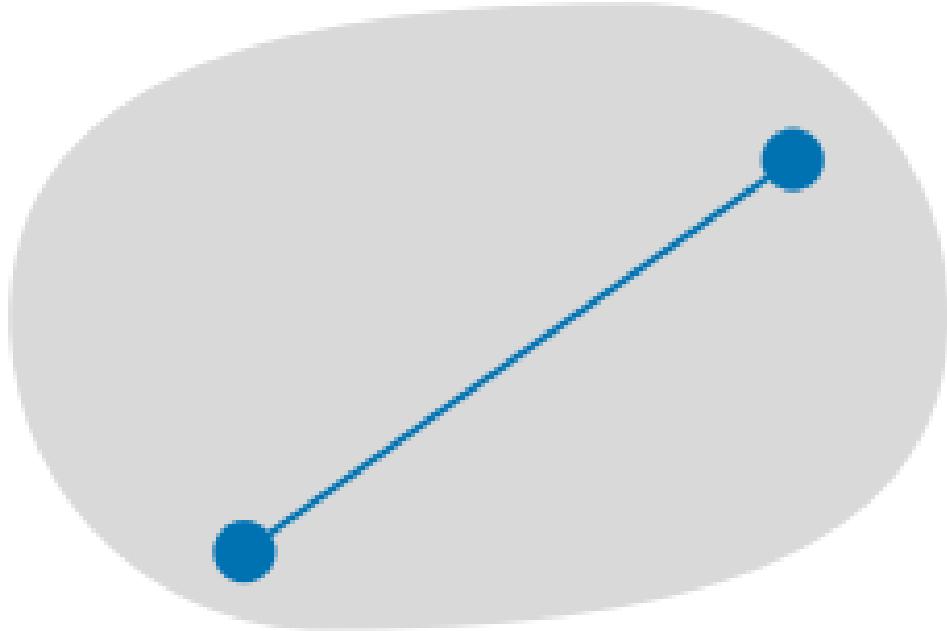
$p$  (m)

$p$  (m)

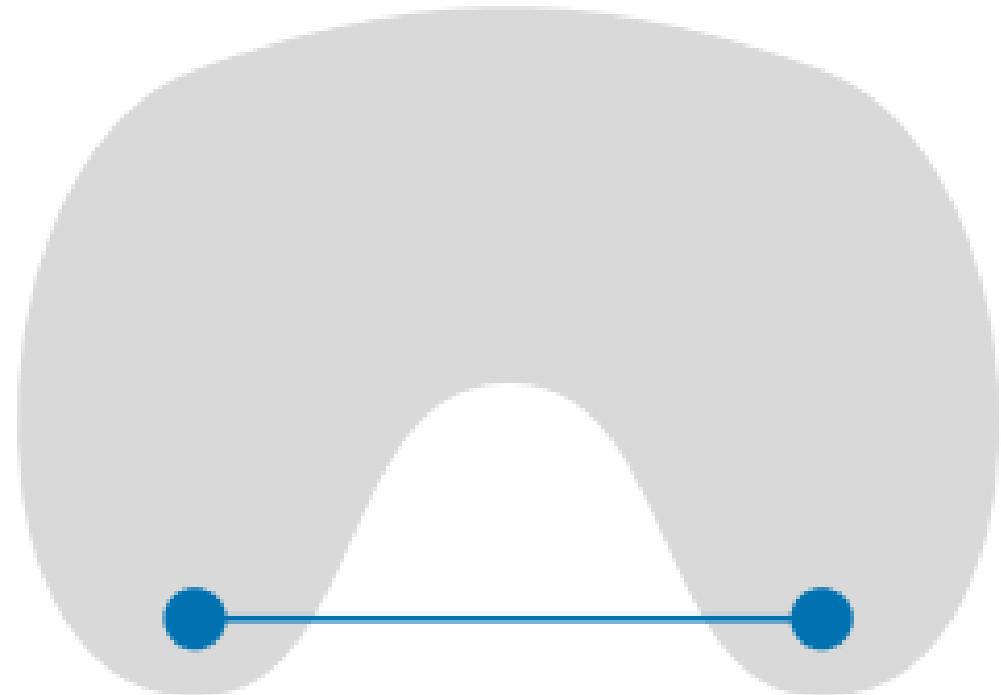
$p$  (m)

$\mathcal{R}_{d-1}$

$\mathcal{R}_d$

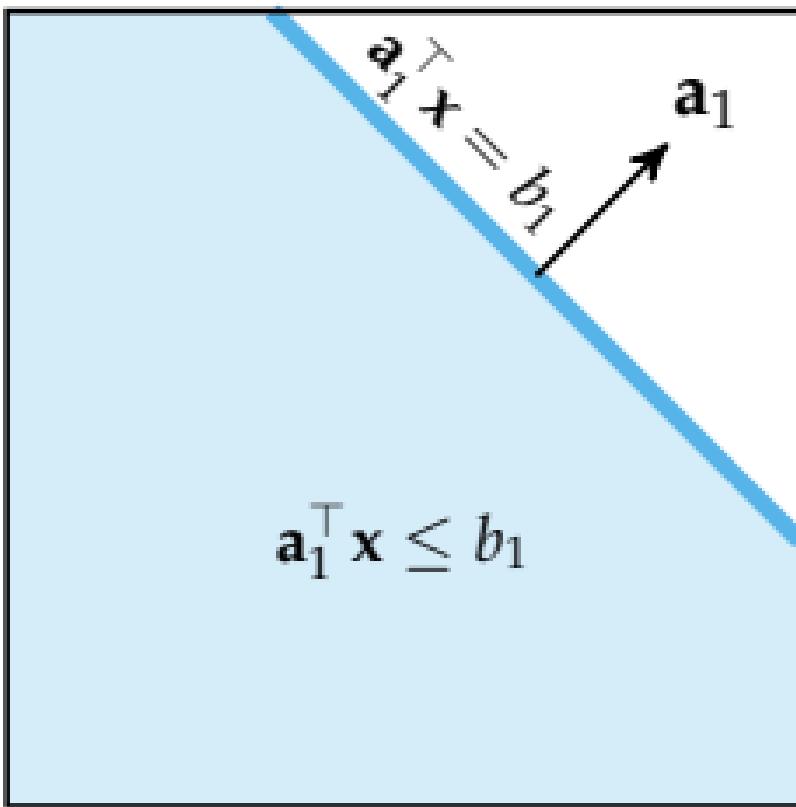


a convex set

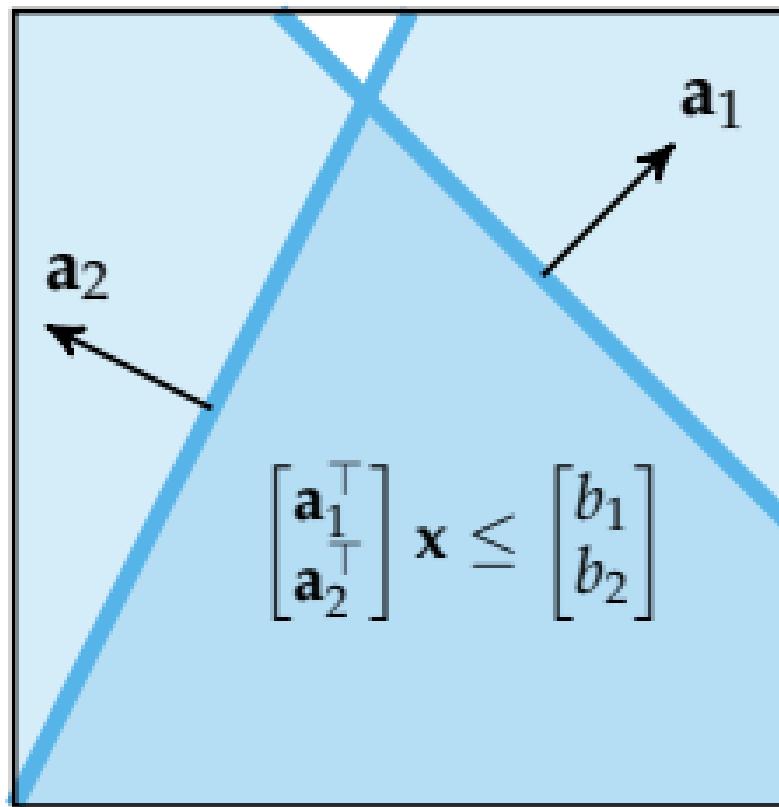


a nonconvex set

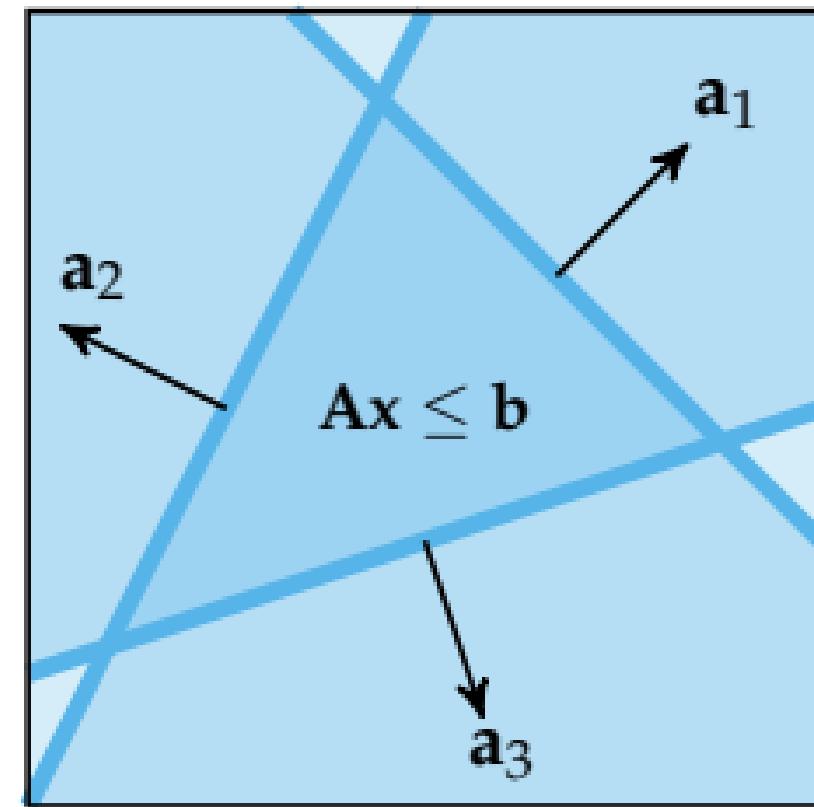
# Half Space

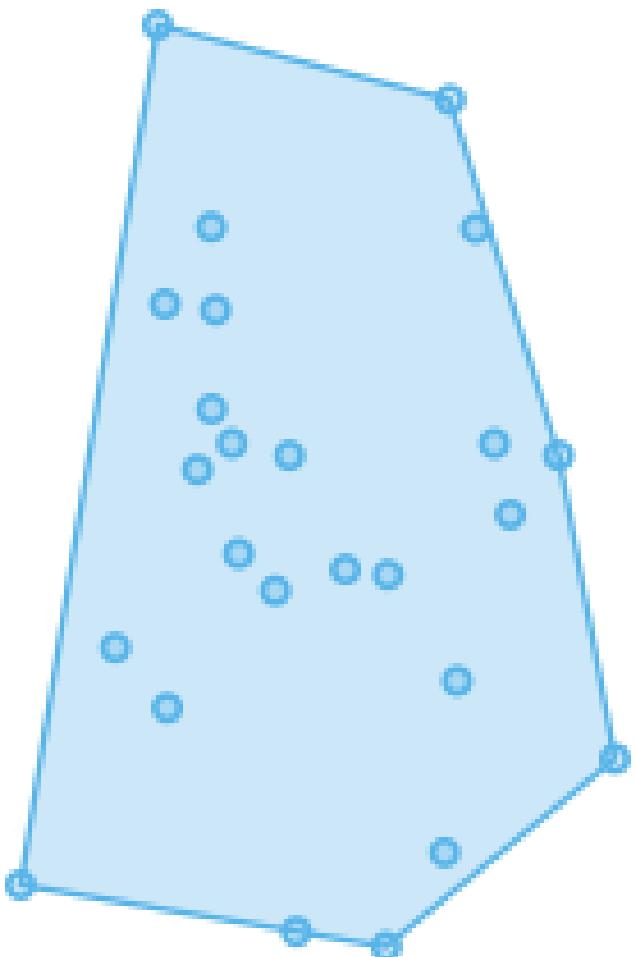


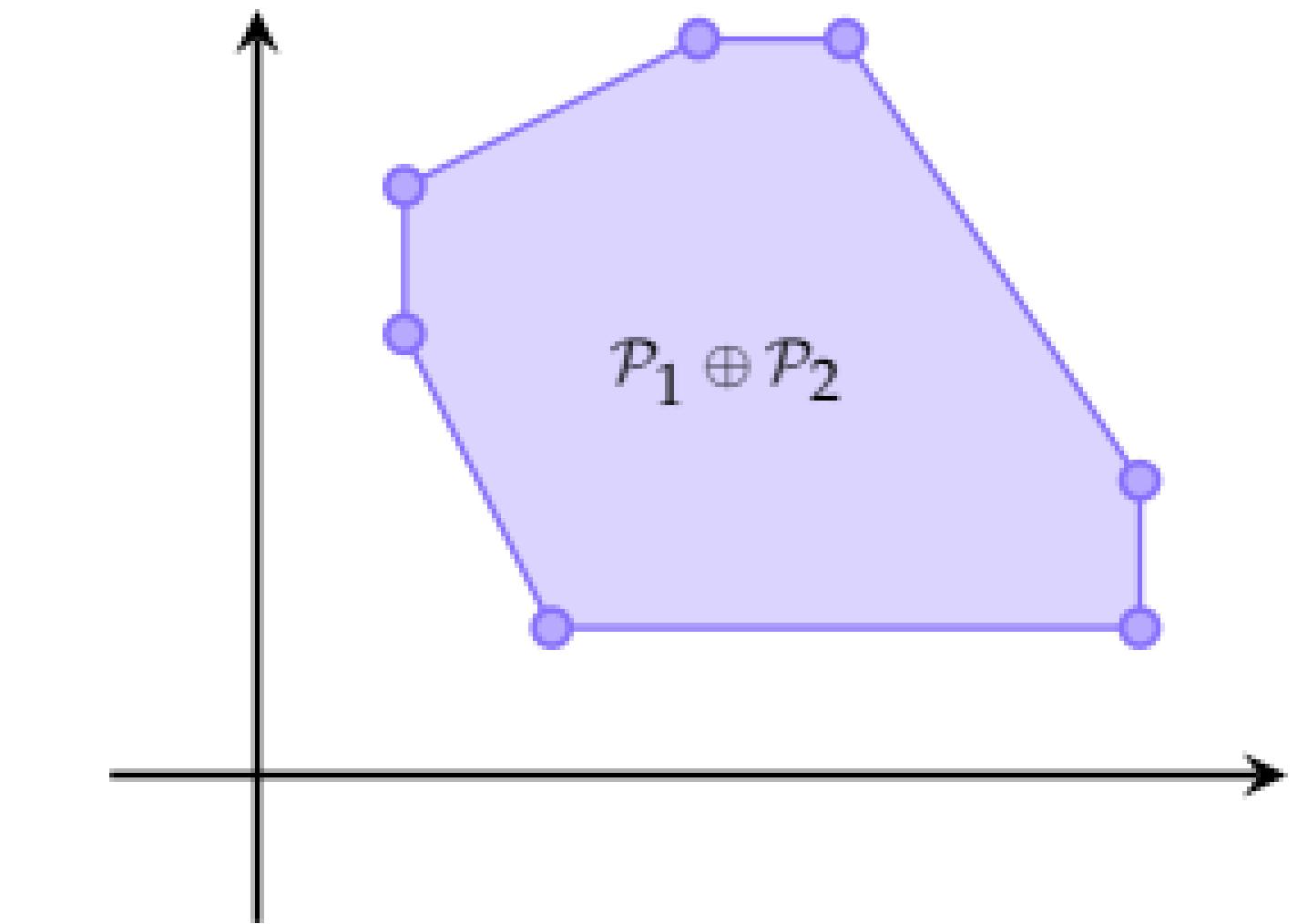
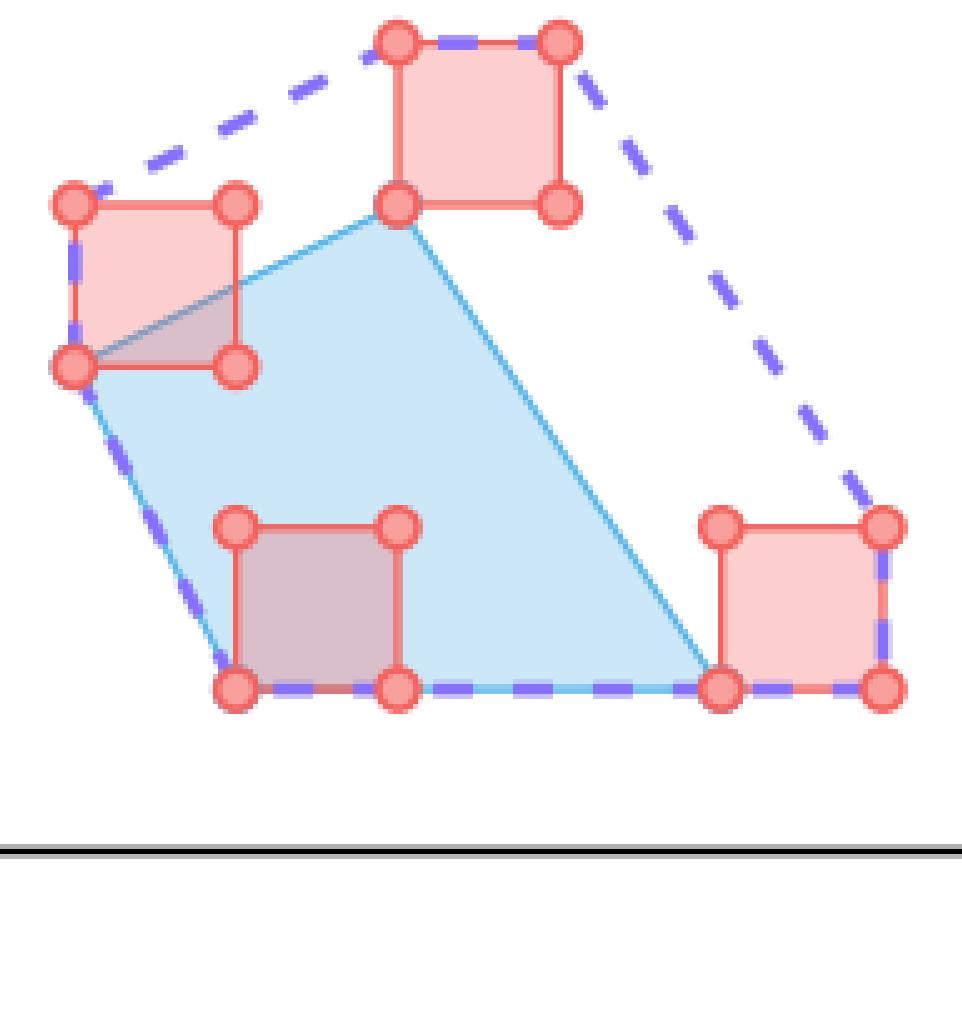
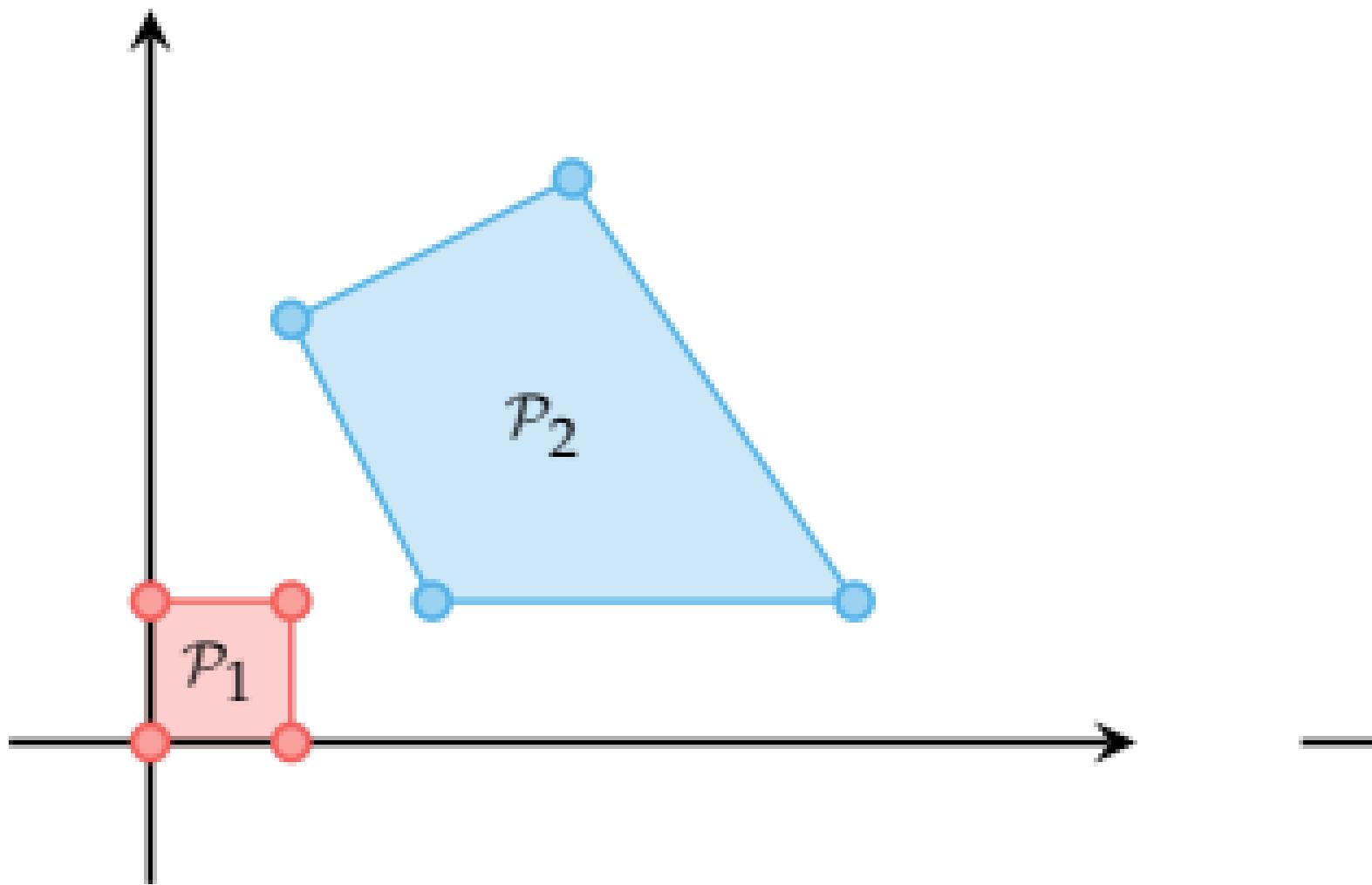
# Polyhedron

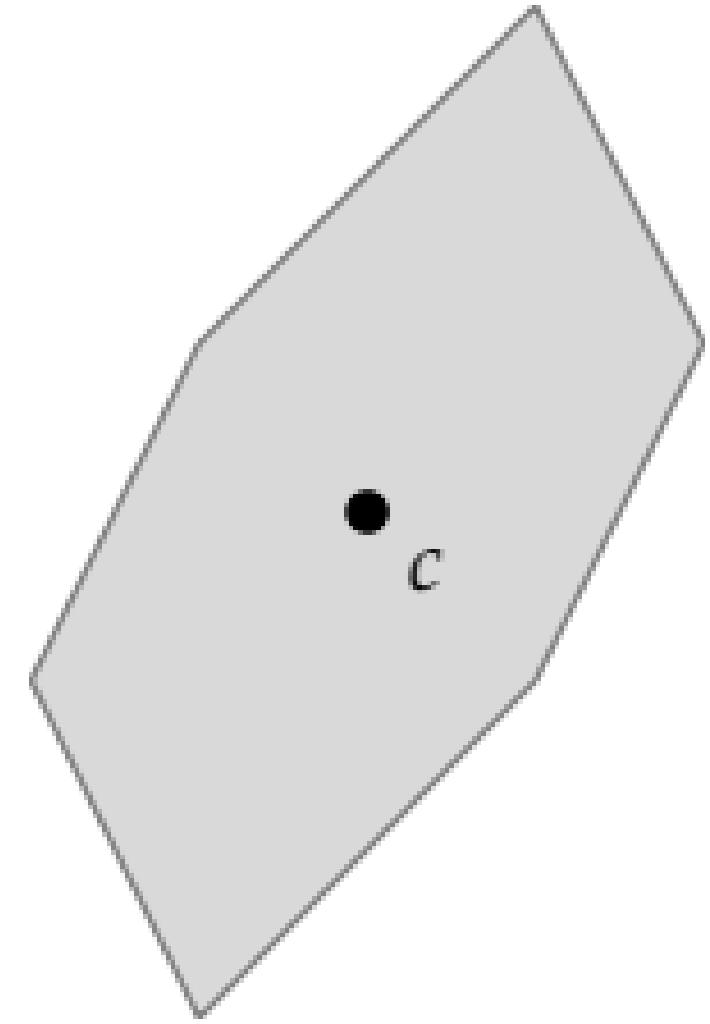
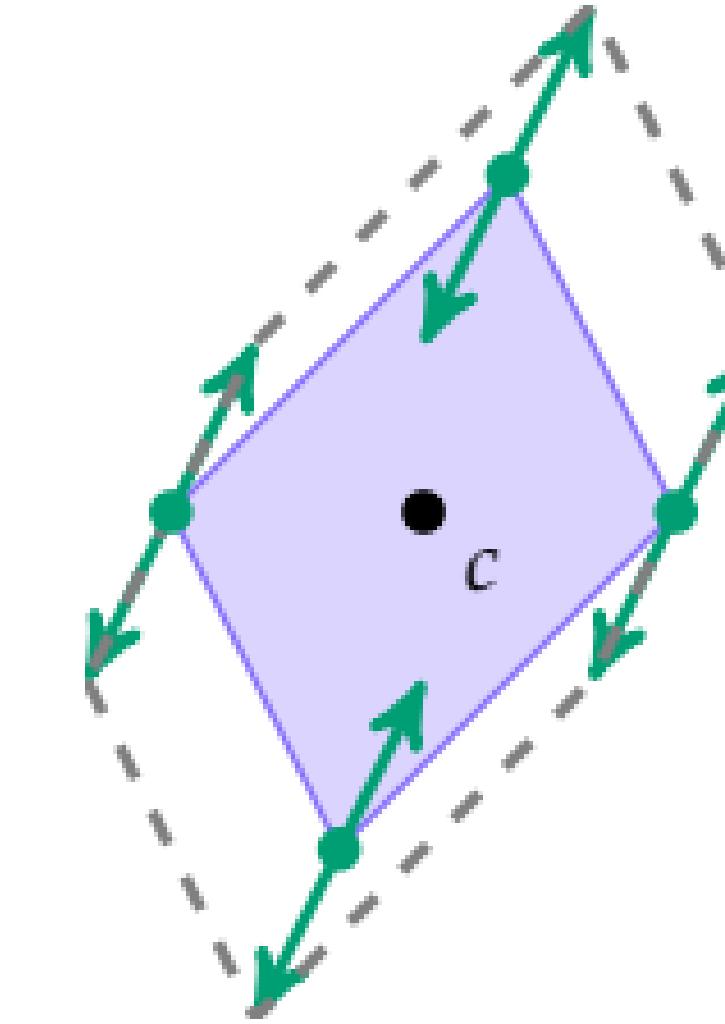
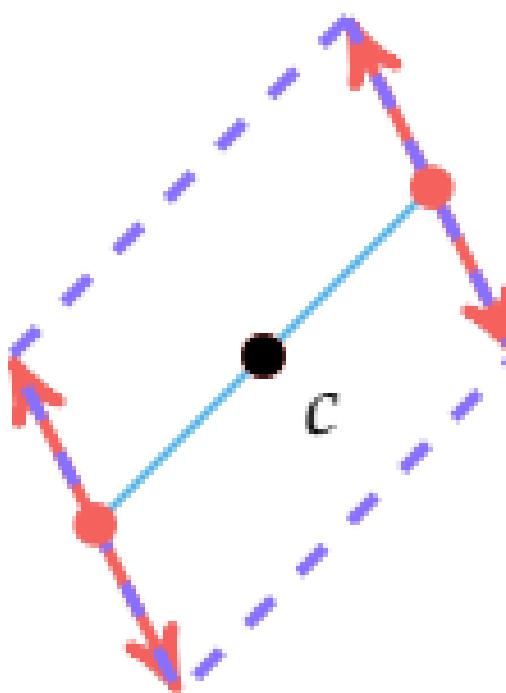
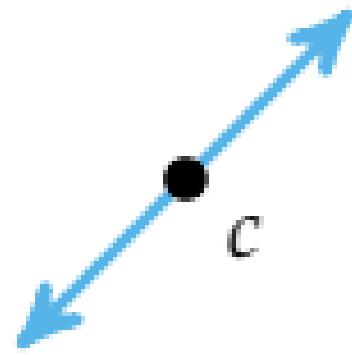


# Polytope







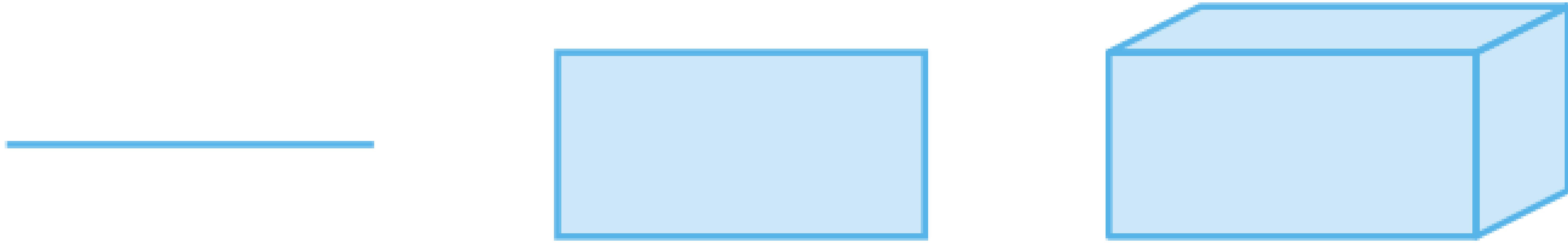


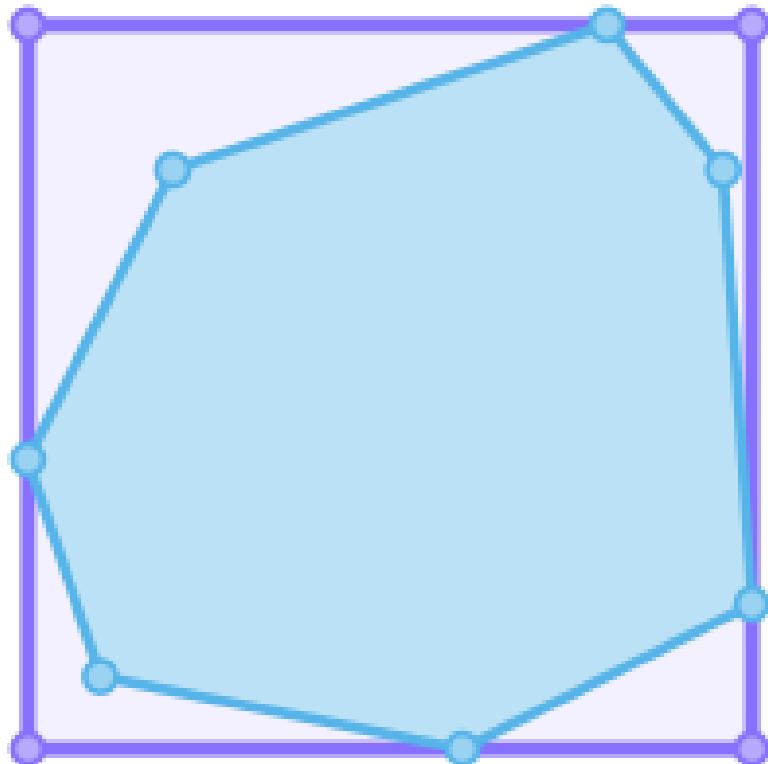


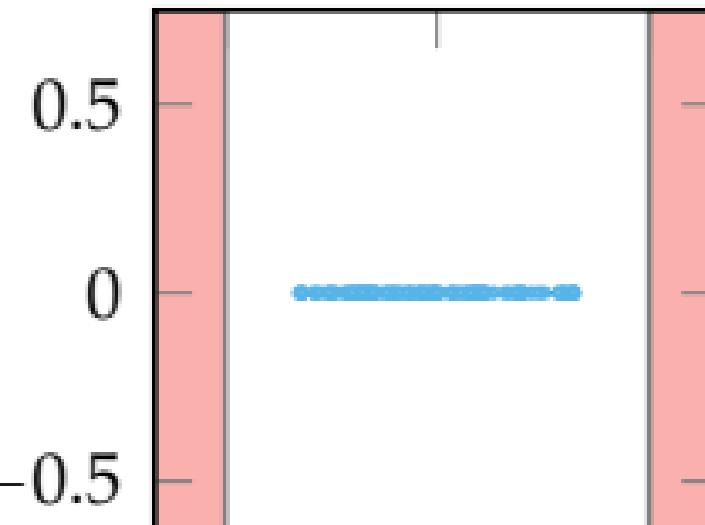
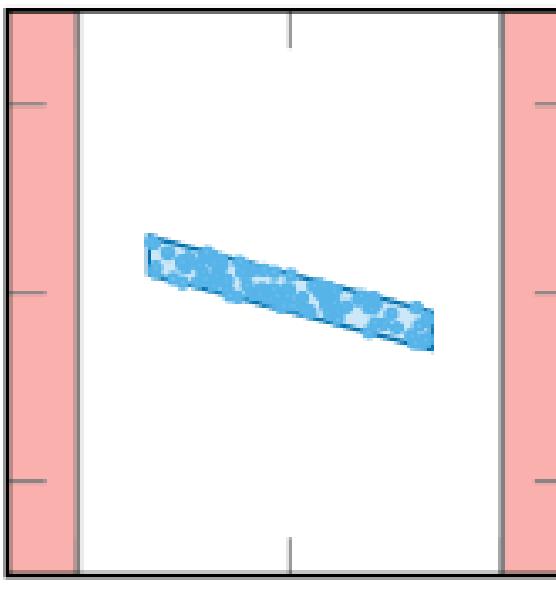
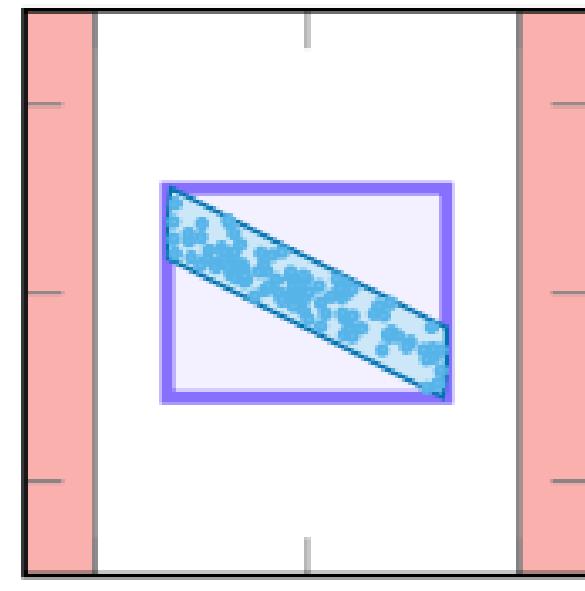
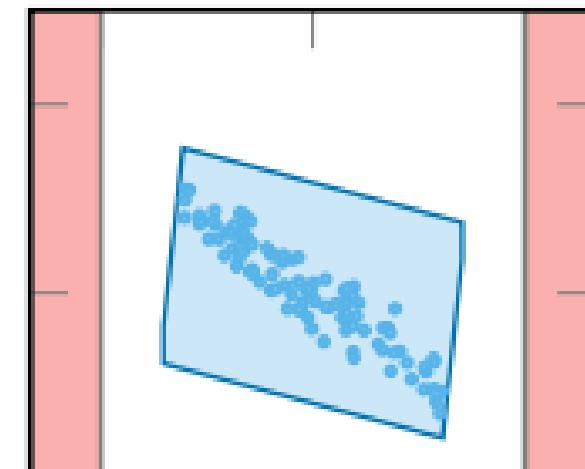
# Polytopes

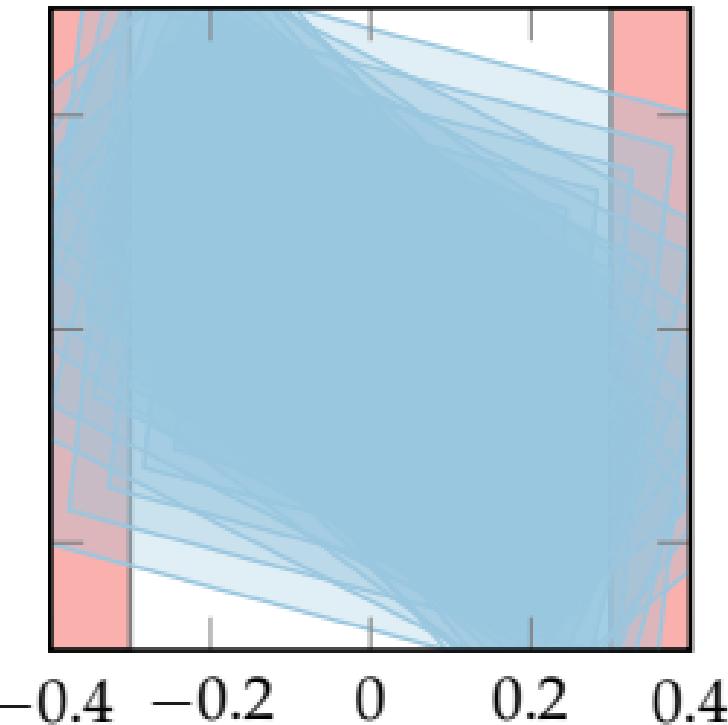
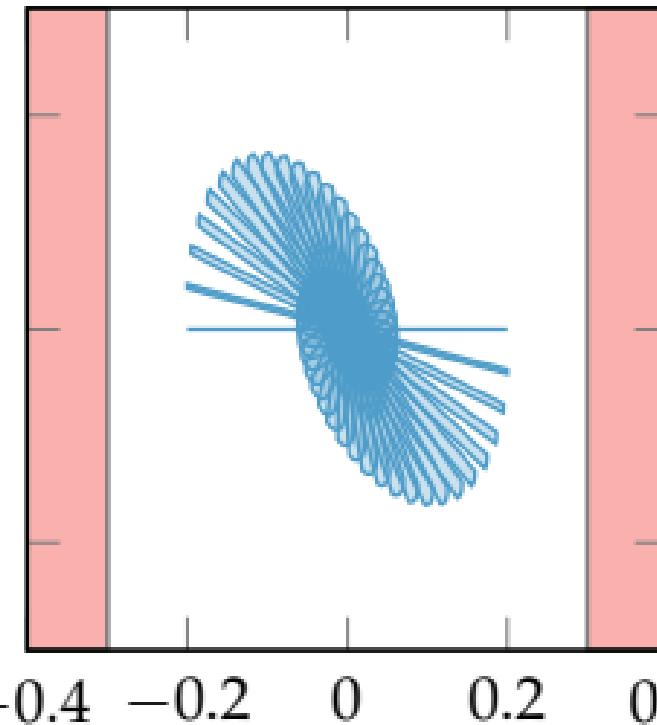
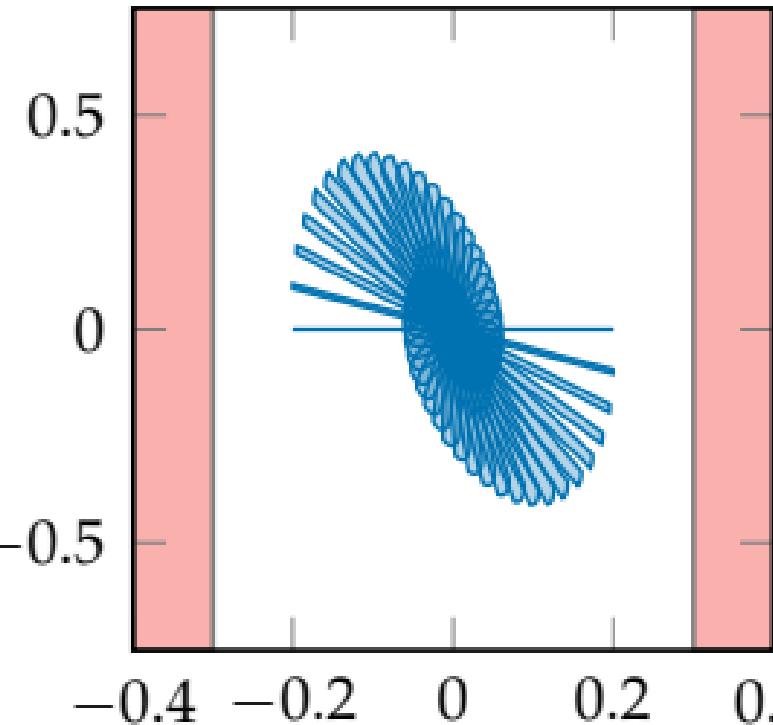
## Zonotopes

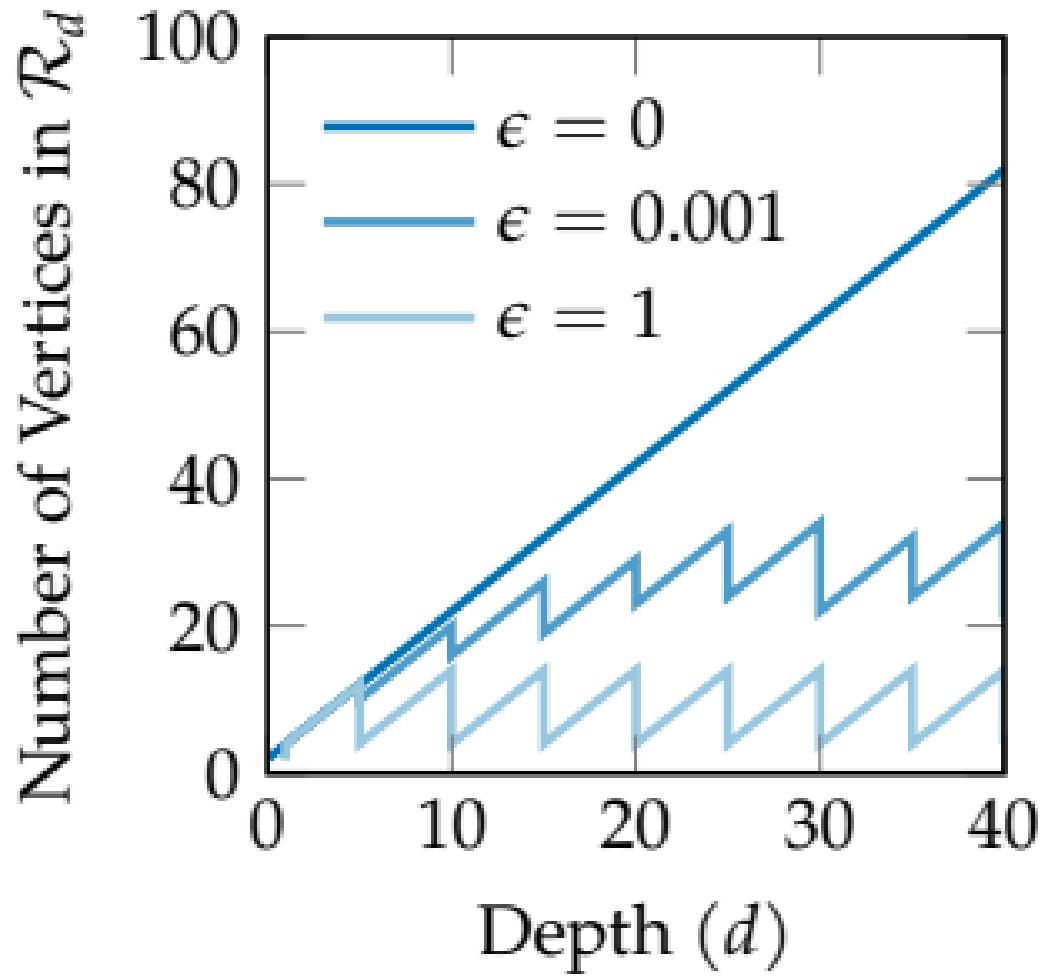
## Hyperrectangles





$\mathcal{R}_1 = \mathcal{S}$  $v \text{ (m/s)}$  $p \text{ (m)}$  $\mathcal{R}_2$  $p \text{ (m)}$  $\mathcal{R}_3$  $p \text{ (m)}$  $\mathcal{R}_4$  $p \text{ (m)}$ 

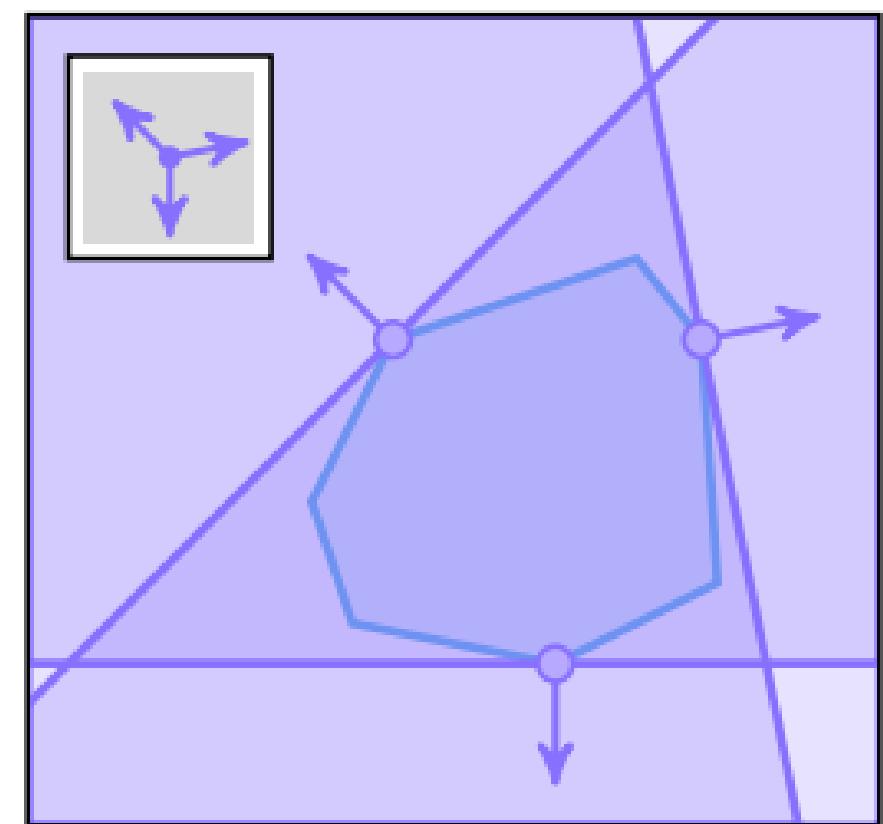
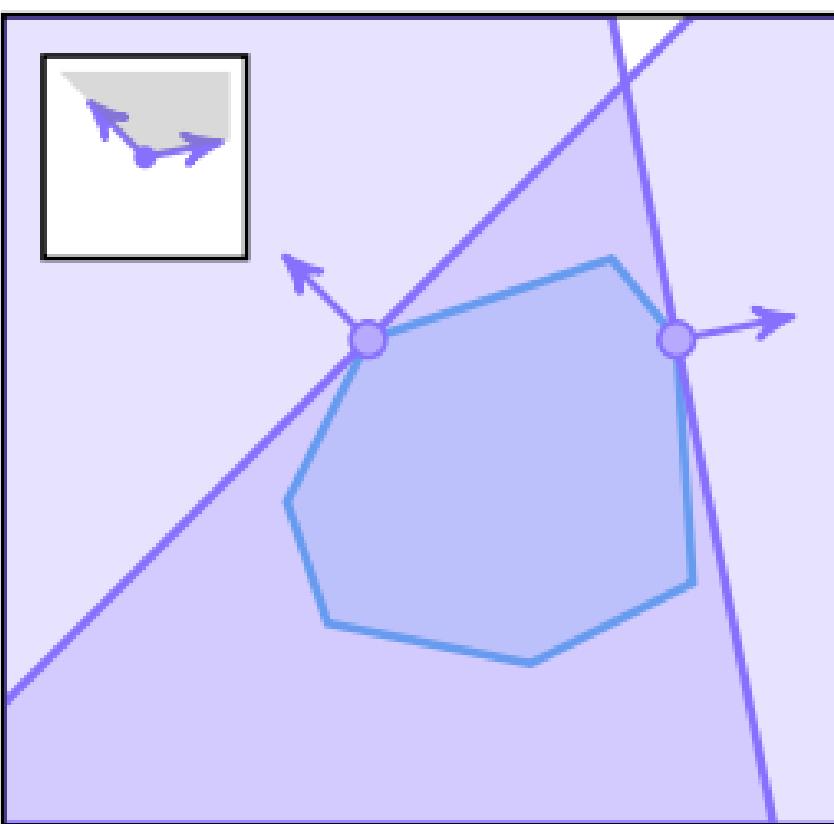
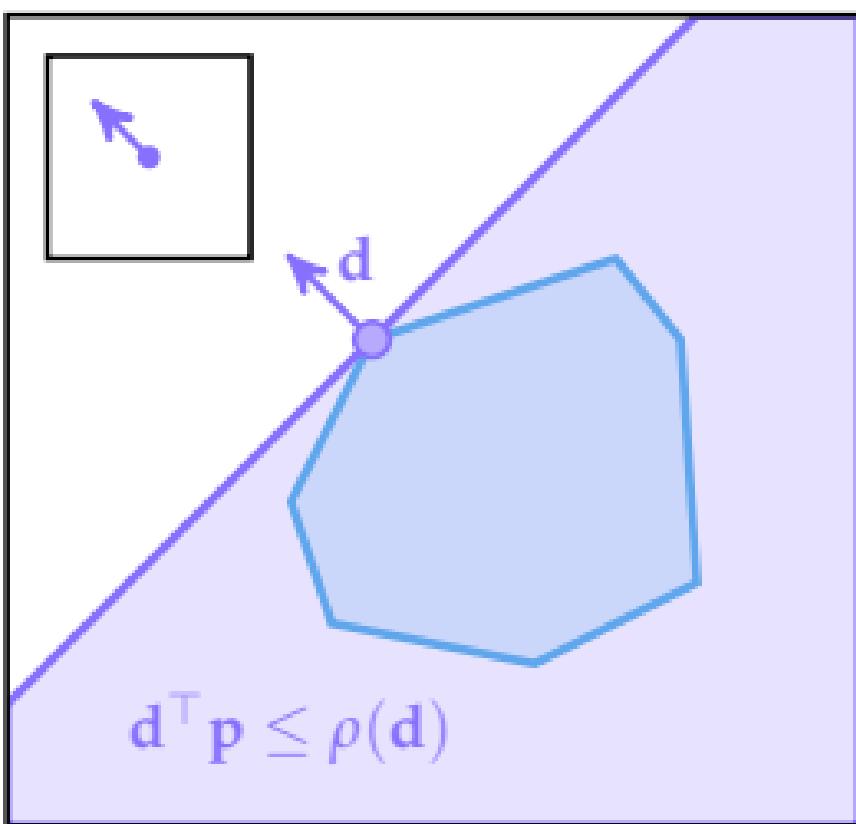
$\epsilon = 0$  $\epsilon = 0.001$  $\epsilon = 1$  $v \text{ (m/s)}$  $p \text{ (m)}$  $p \text{ (m)}$  $p \text{ (m)}$

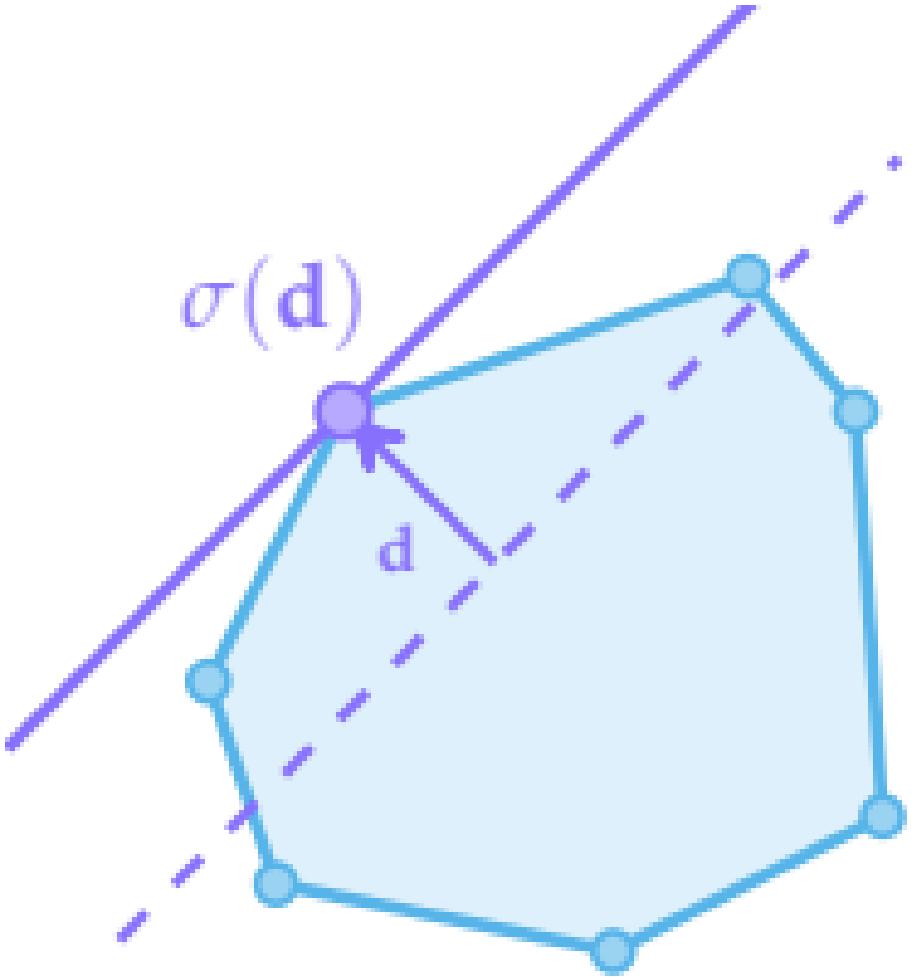


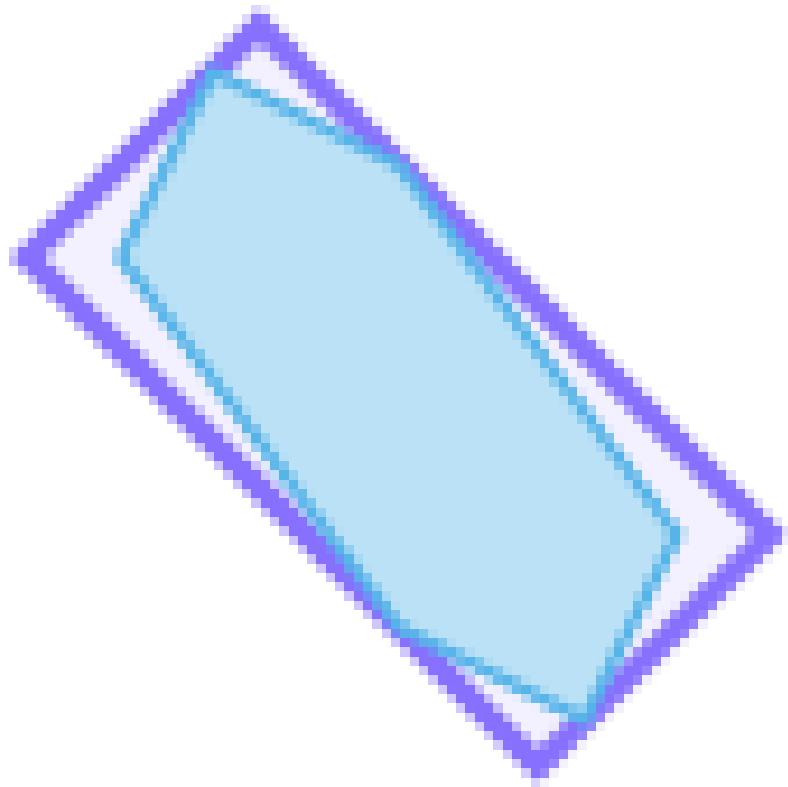
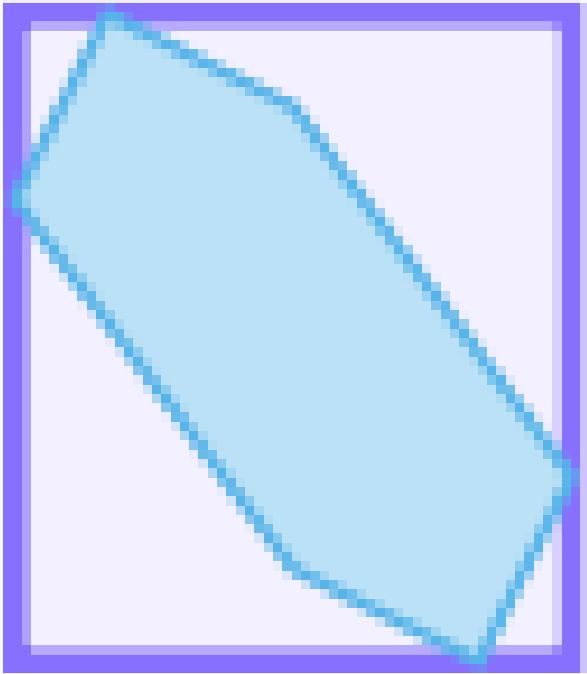
# Bounding Halfspace

# Bounding Polyhedron

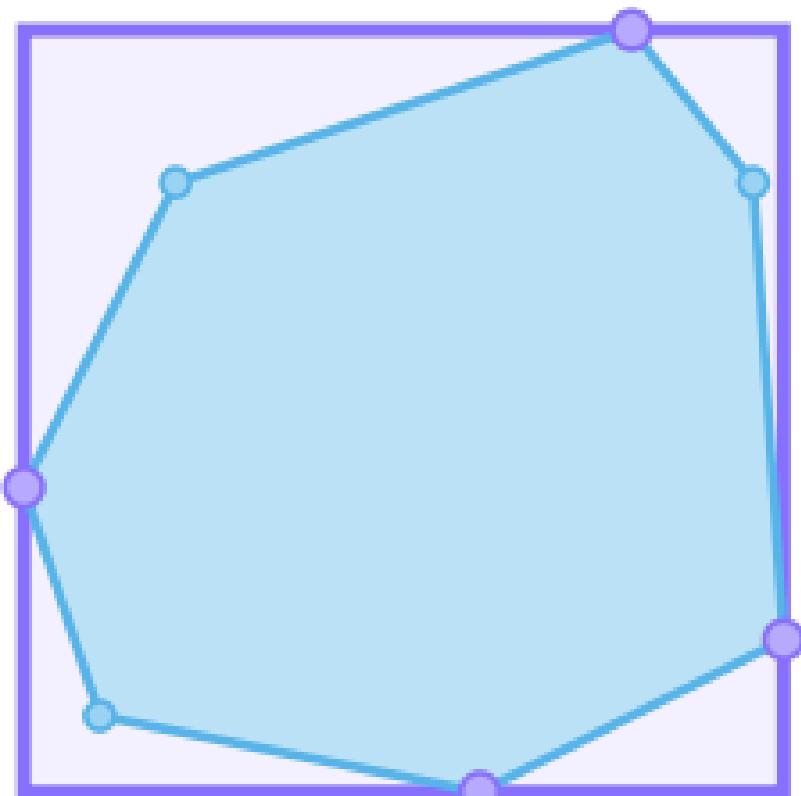
# Bounding Polytope



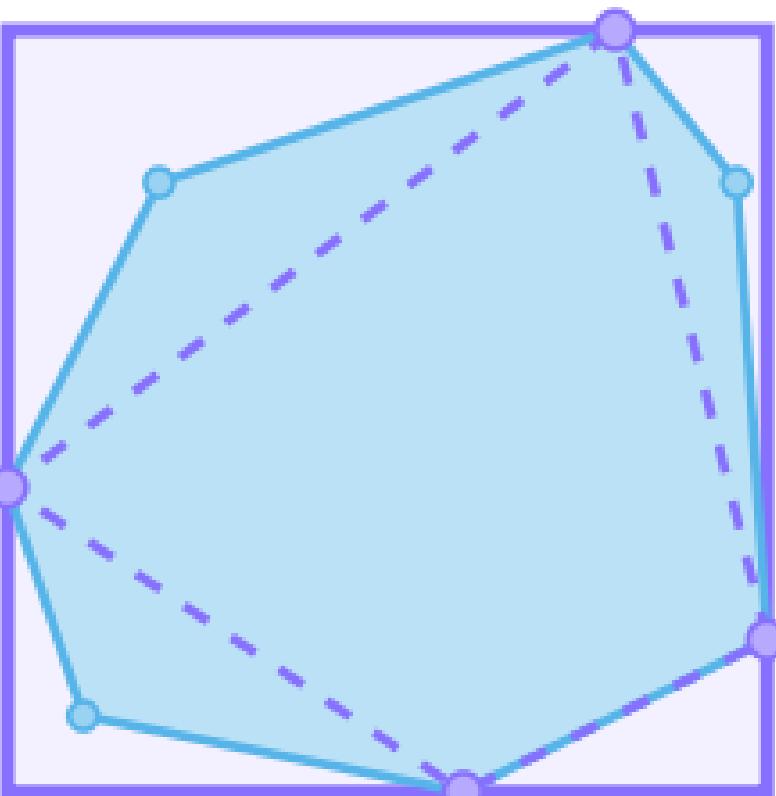




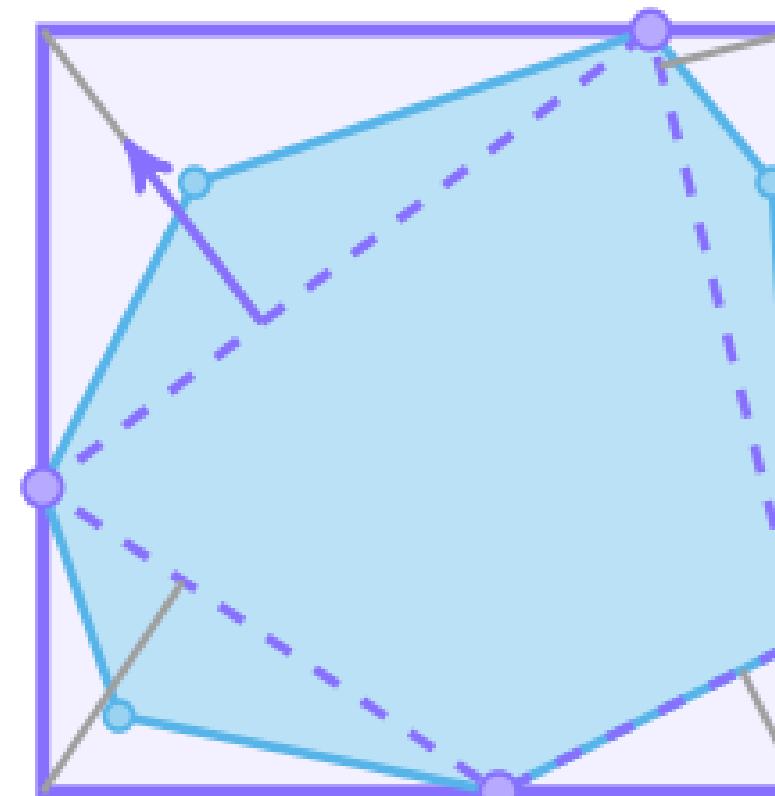
Step 1



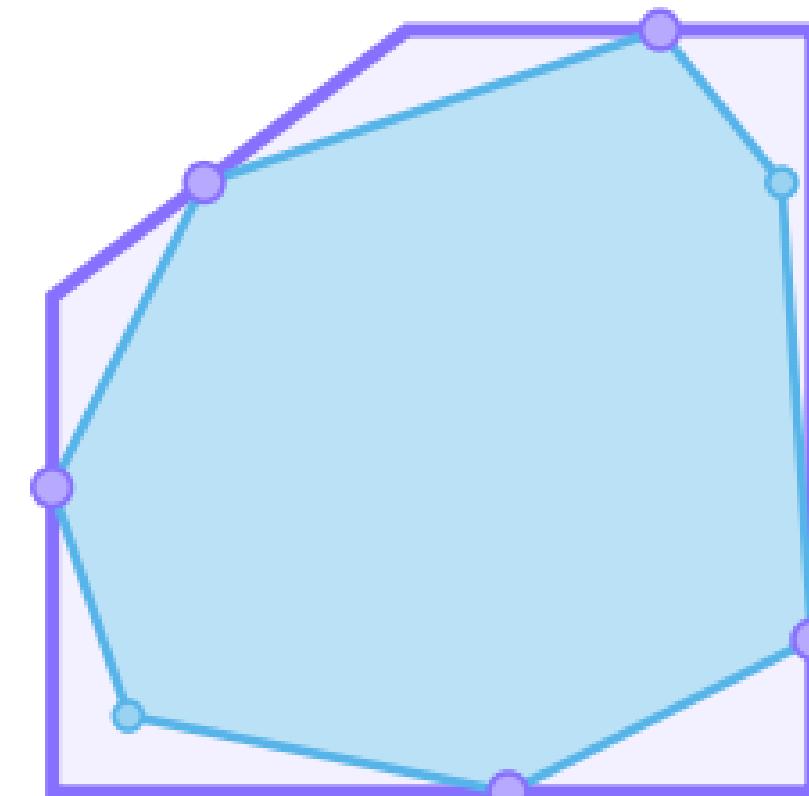
Step 2



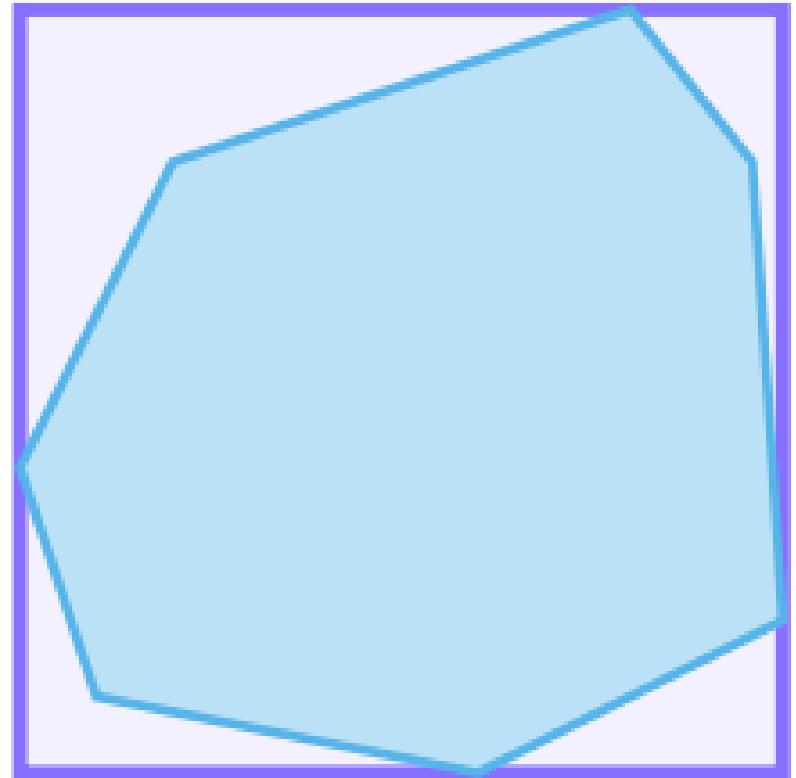
Step 3



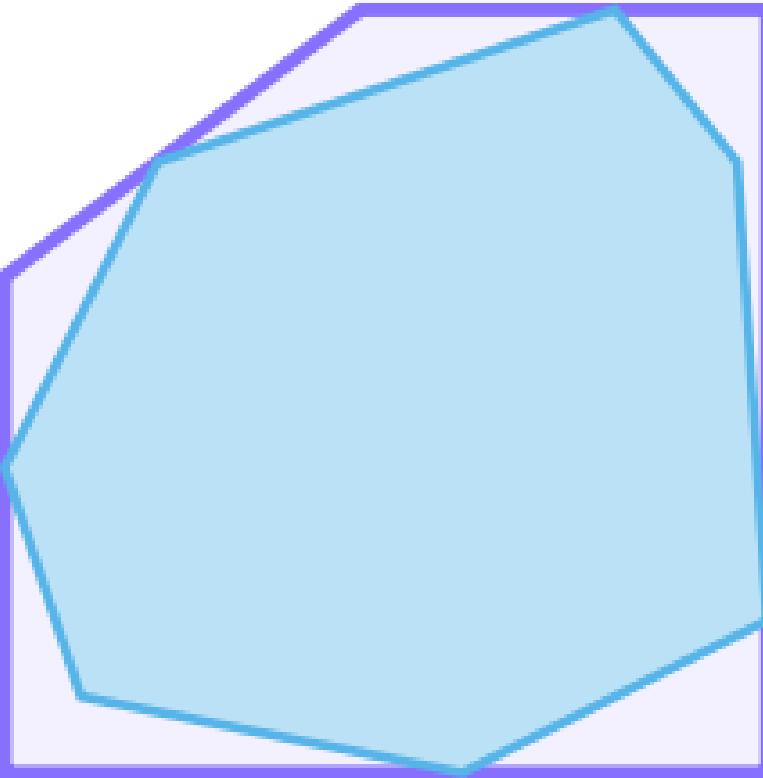
Step 4



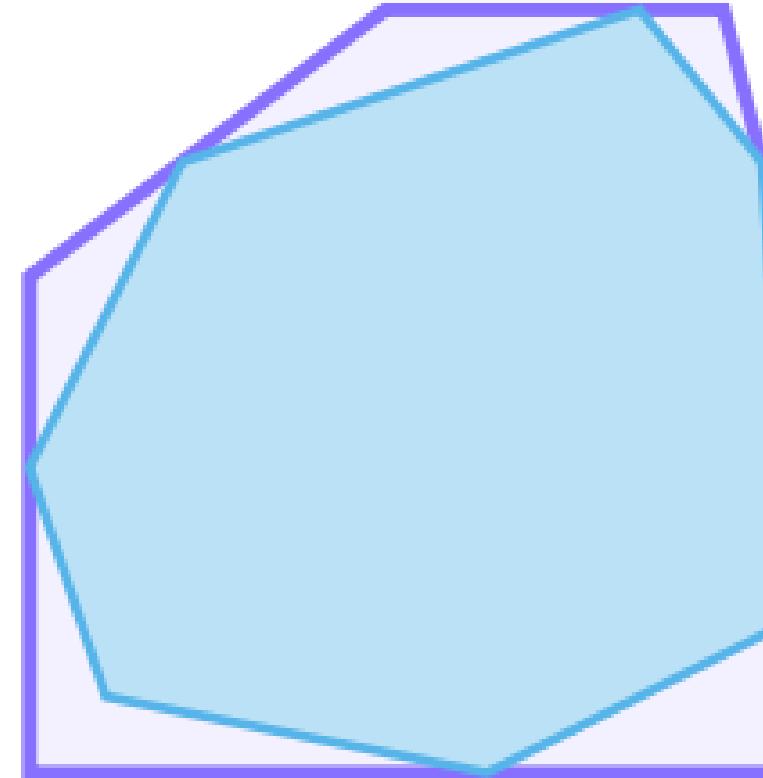
Iteration 1



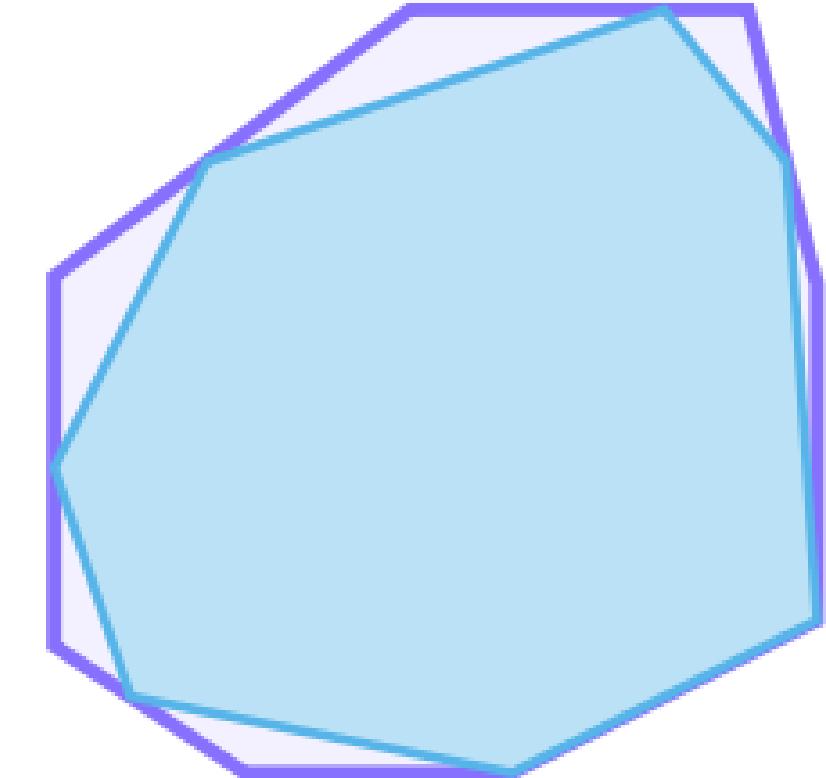
Iteration 2



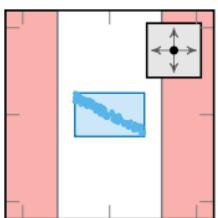
Iteration 3



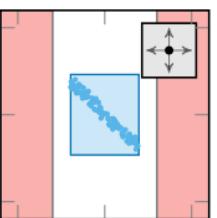
Converged



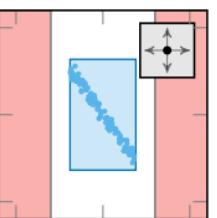
R<sub>2</sub>



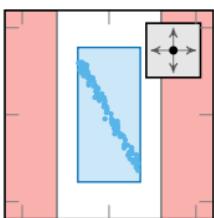
R<sub>3</sub>



R<sub>4</sub>



R<sub>5</sub>



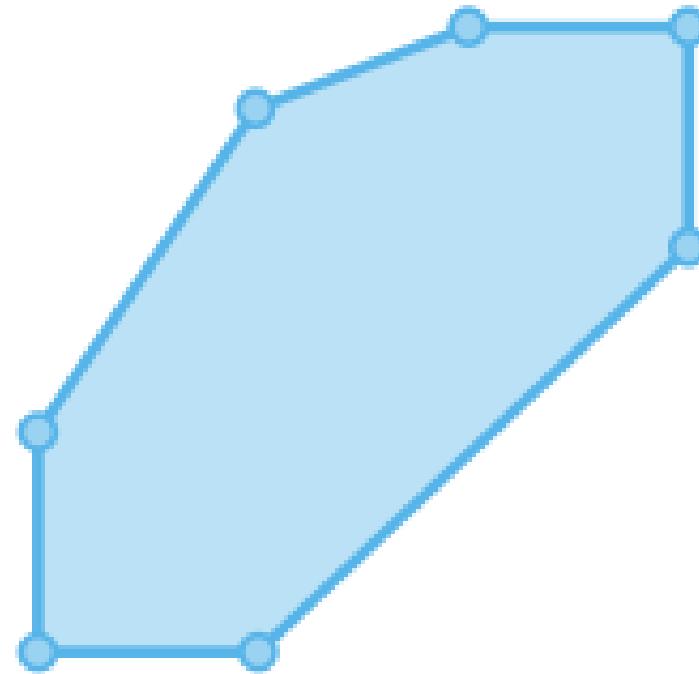
Axis Aligned

10 Random

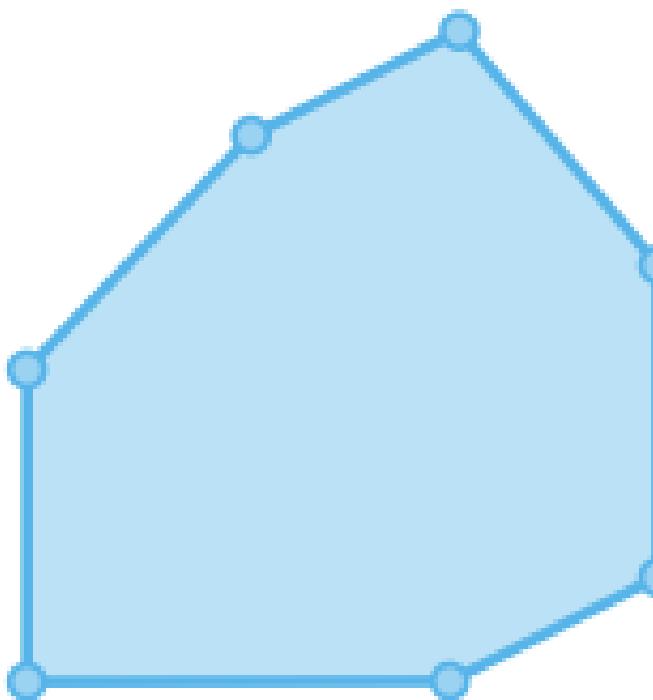
50 Random

PCA

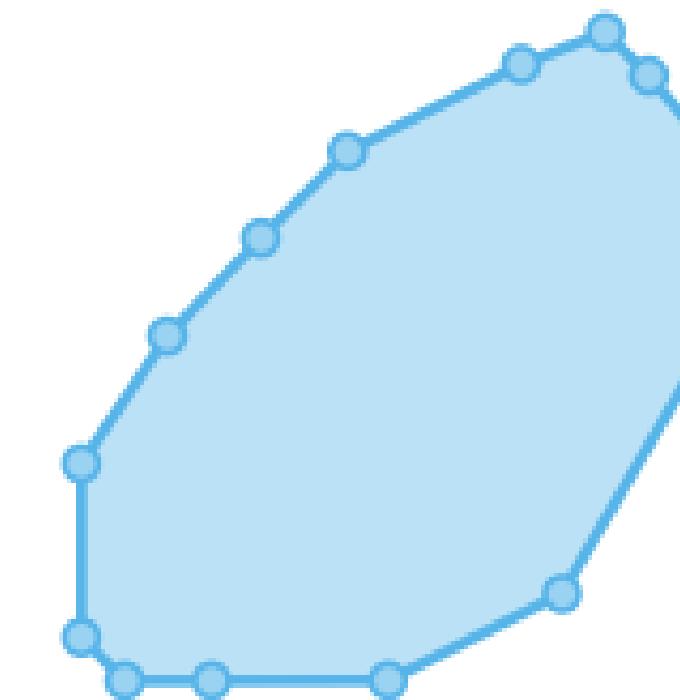
**A**



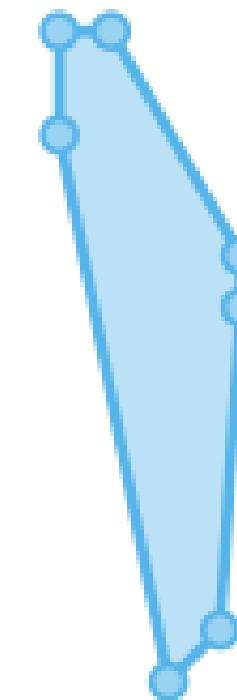
**B**

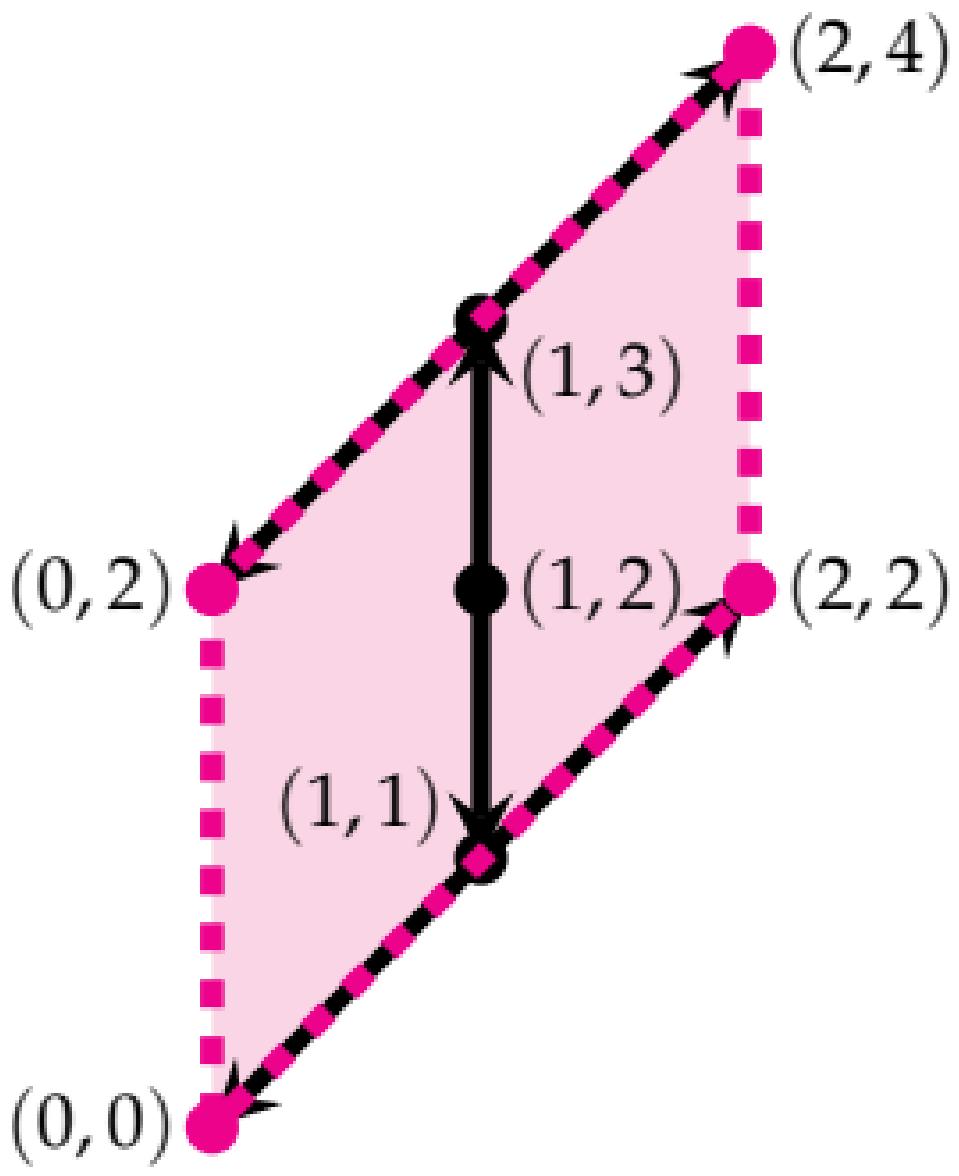


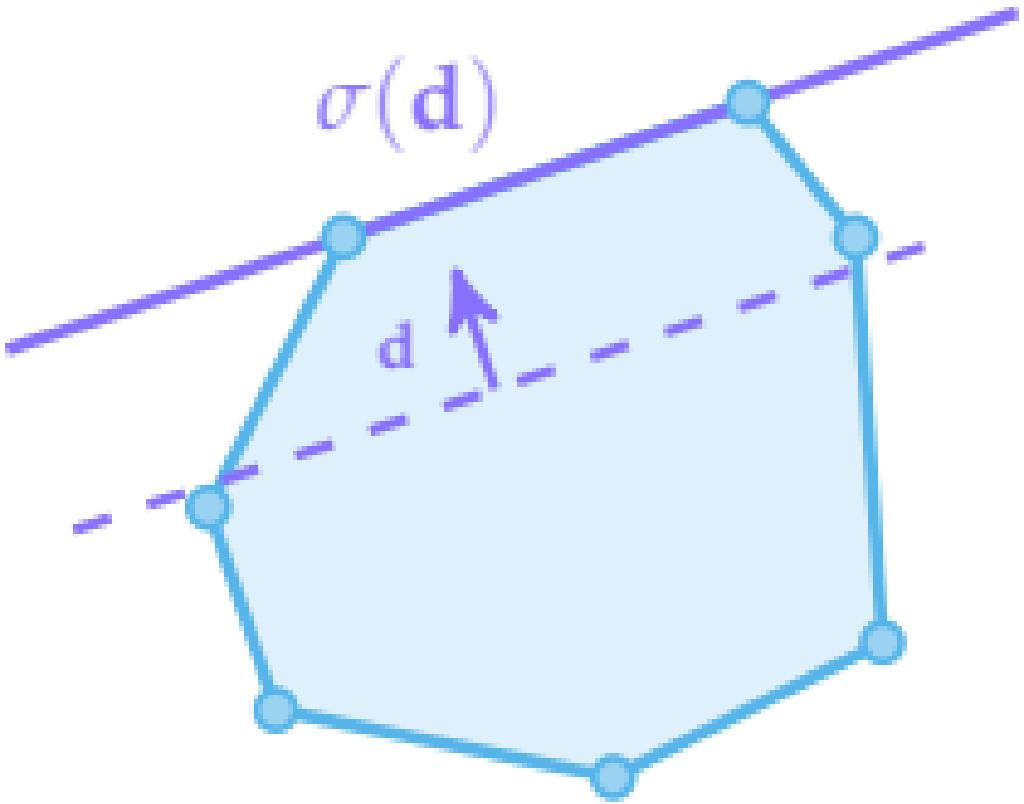
**C**



**D**





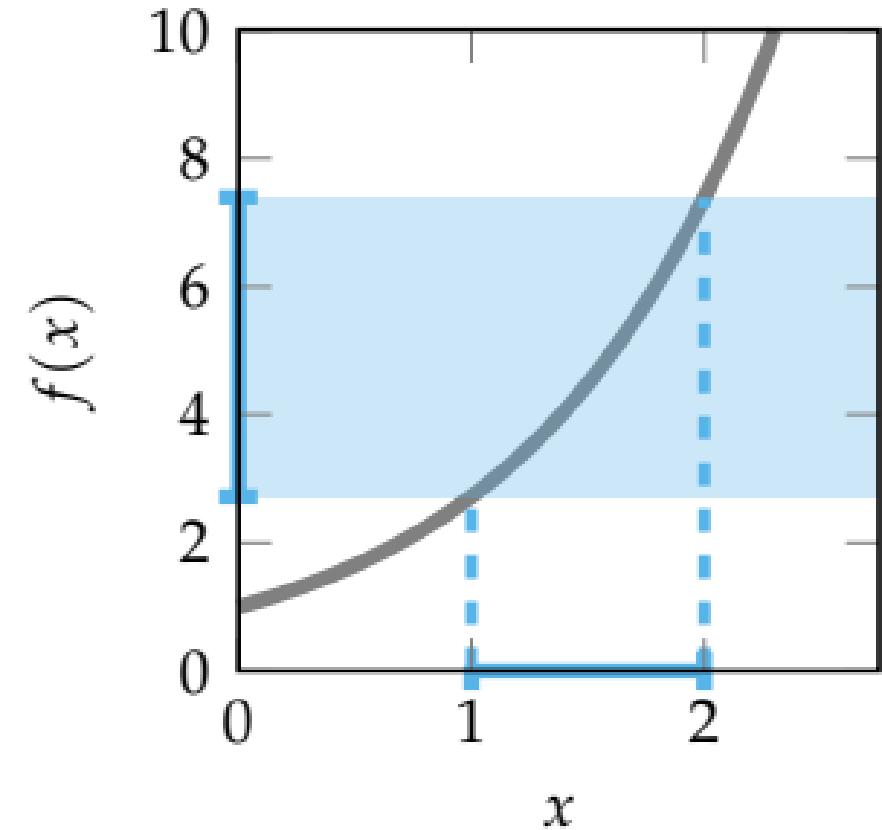
$\sigma(\mathbf{d})$ 

$[x_2]$

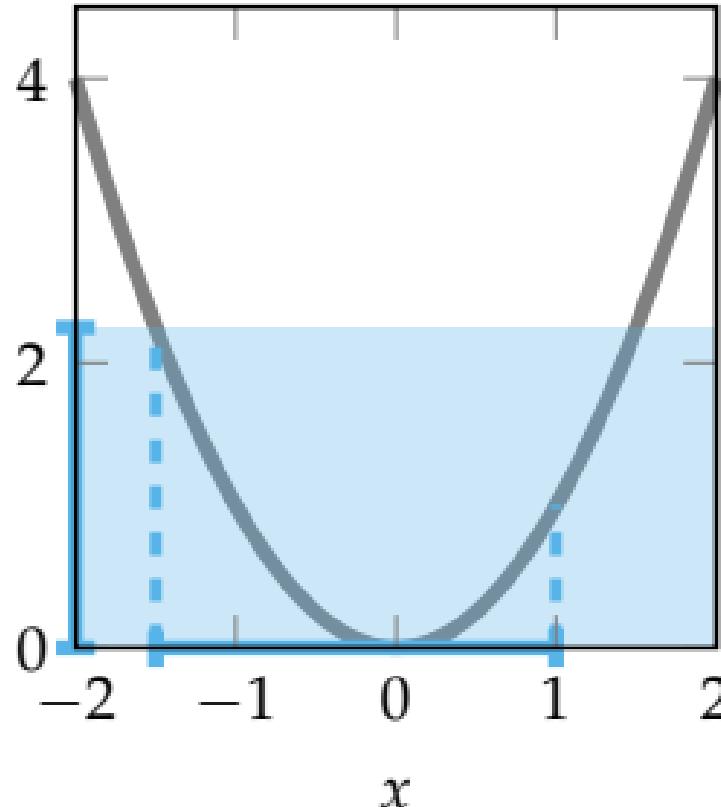
$$[\mathbf{x}] = [x_1] \times [x_2]$$

$[x_1]$

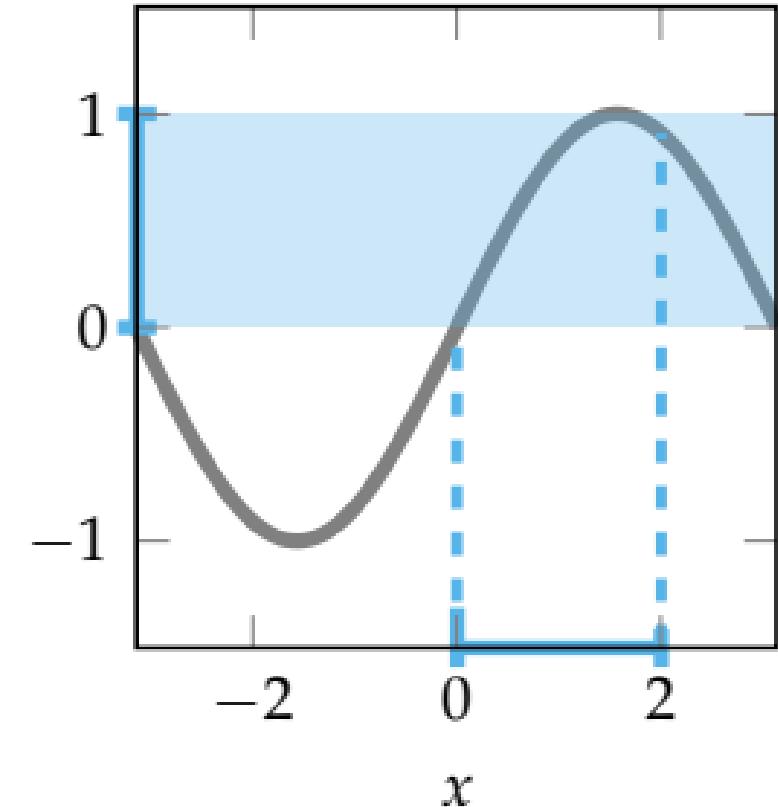
$$f(x) = \exp(x)$$



$$f(x) = x^2$$



$$f(x) = \sin(x)$$



$f(x)$

2

0

-2

-2

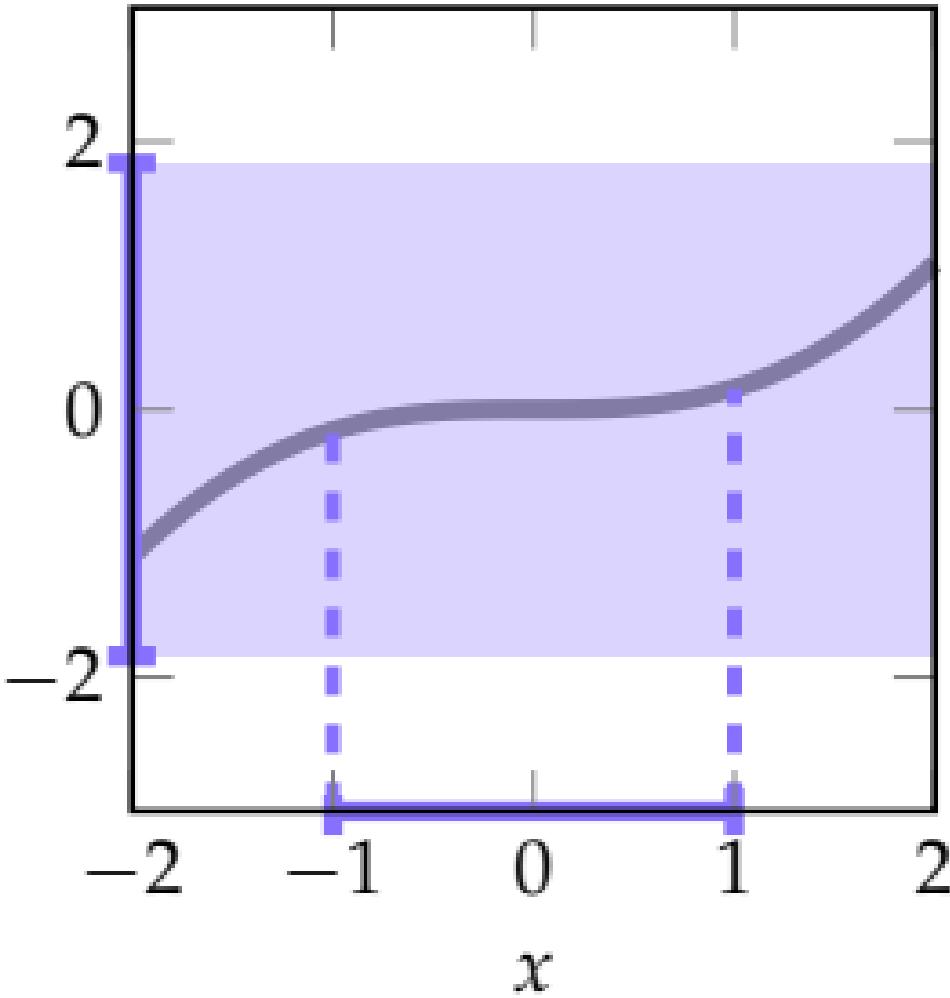
-1

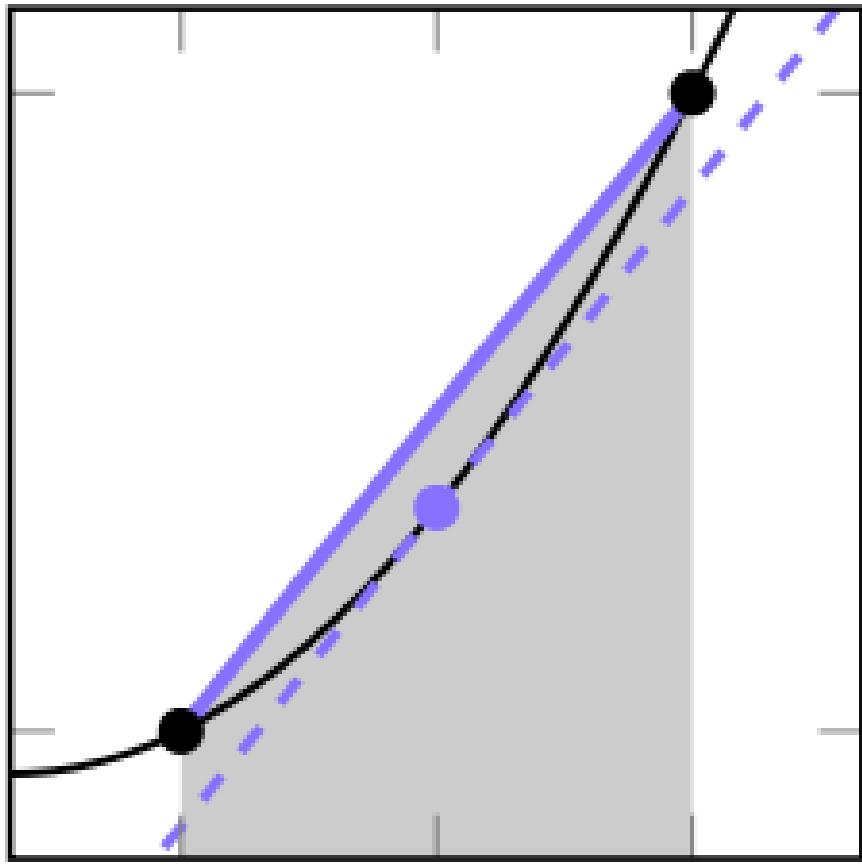
0

1

2

$x$



$f(x)$  $f(\underline{x})$  $\underline{x}$  $x'$  $\bar{x}$ 

$\omega$  (rad/s)

2

0

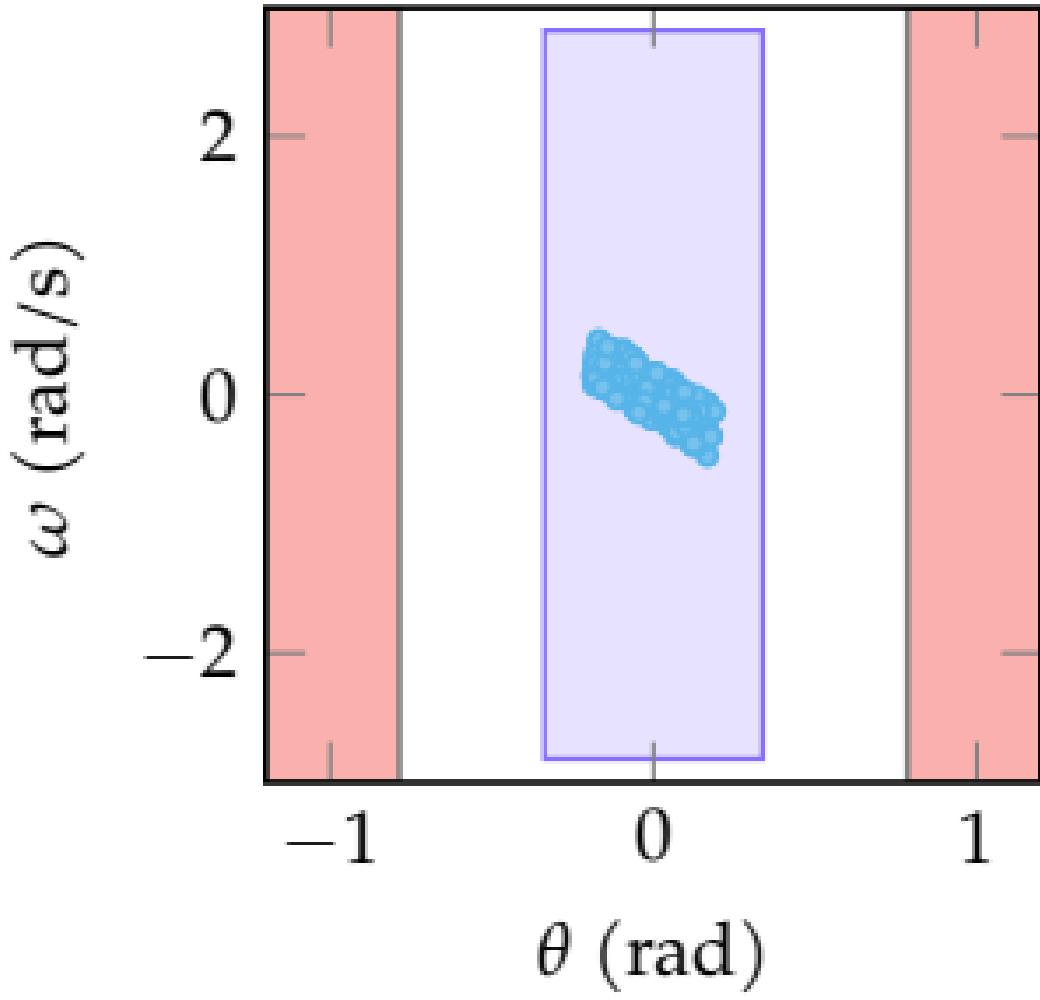
-2

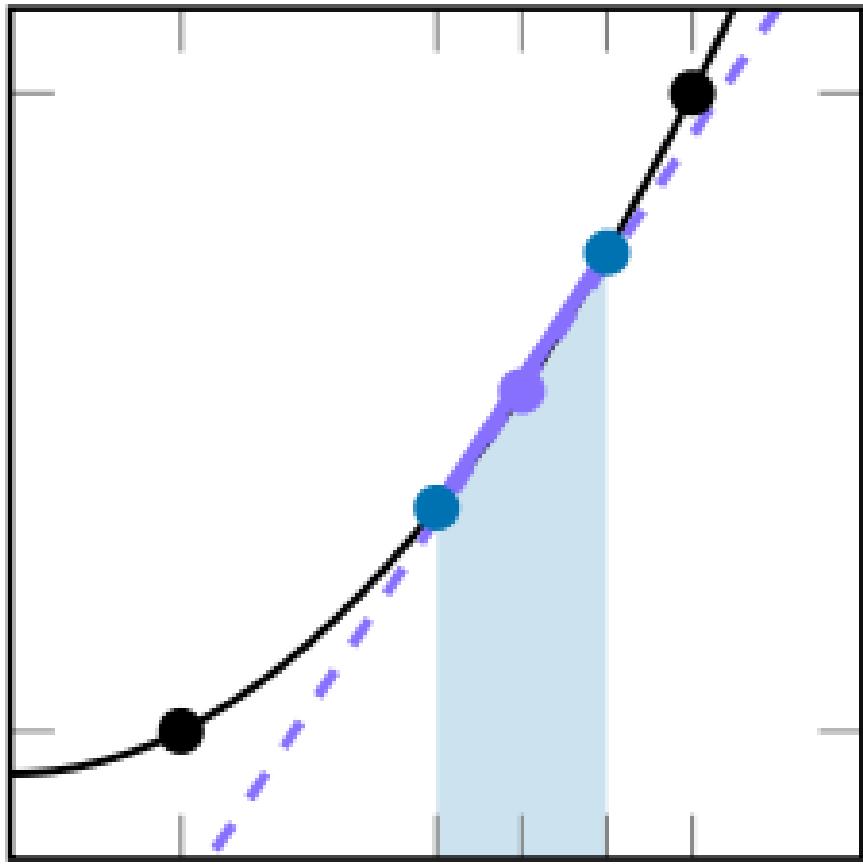
-1

0

1

$\theta$  (rad)



$f(x)$  $f(\underline{x})$  $\underline{x}$  $c \ x' \ x \ \bar{x}$ 

$f(x)$

2

0

-2

-2

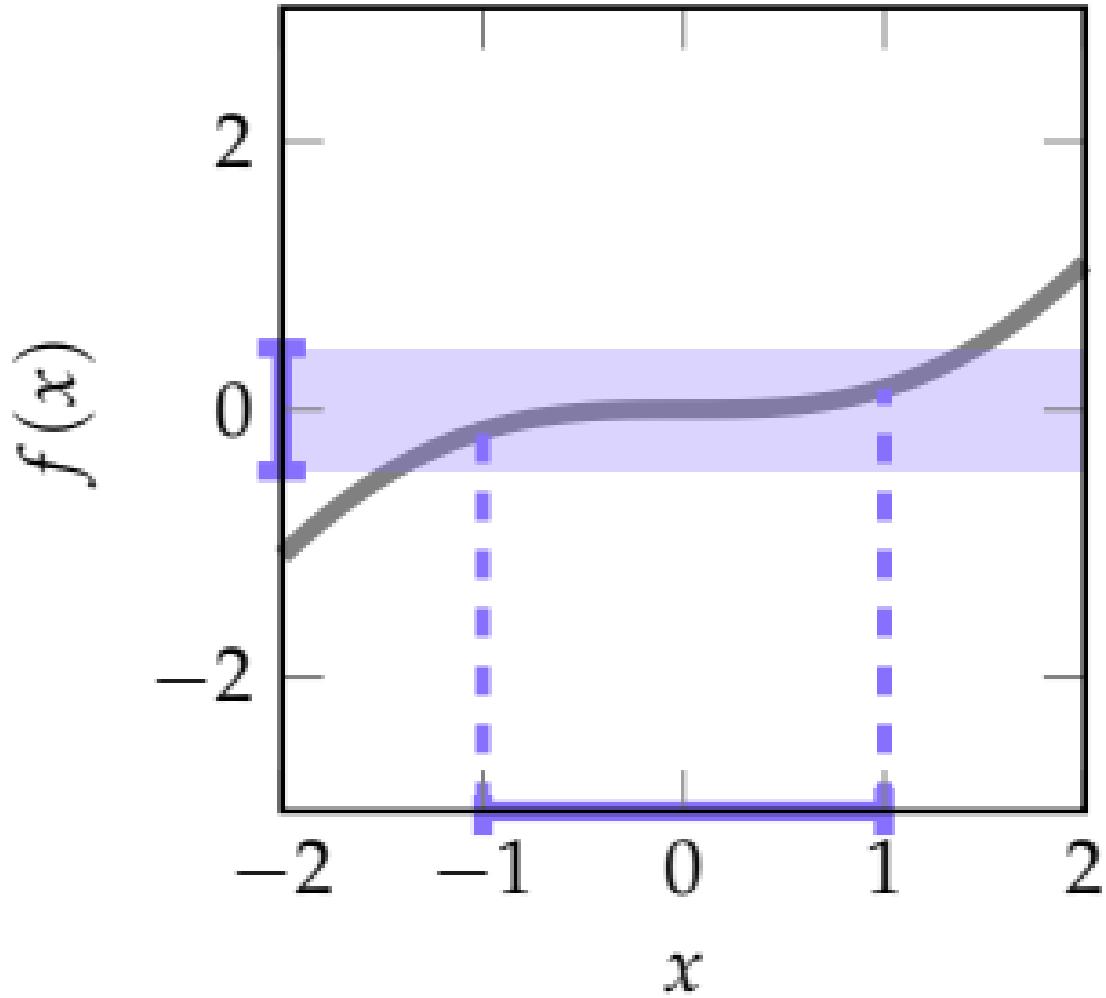
-1

0

1

2

$x$



$f(x)$

2

0

-2

-2

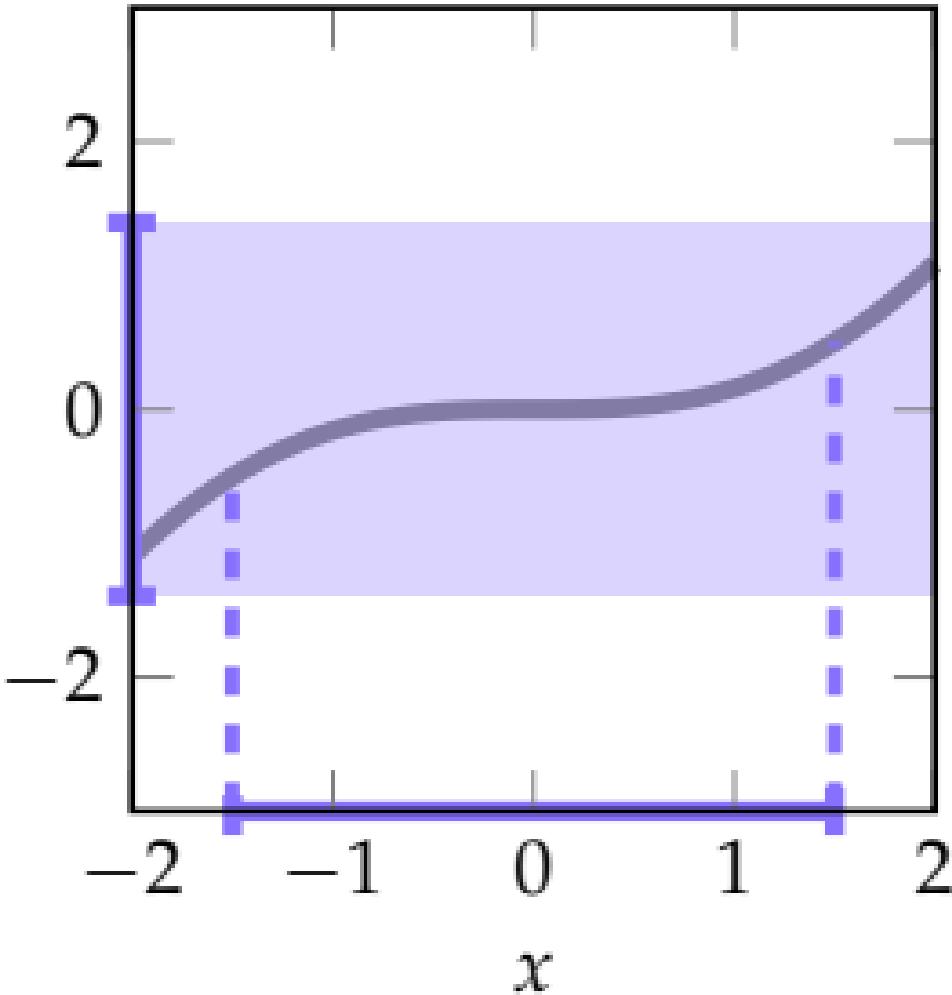
-1

0

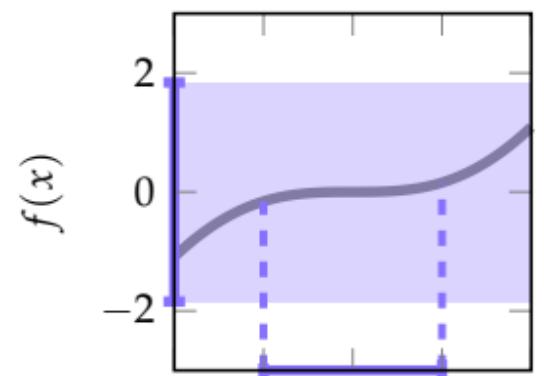
1

2

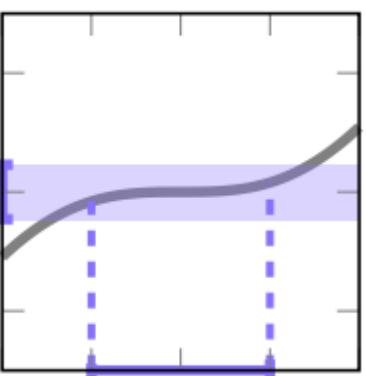
$x$



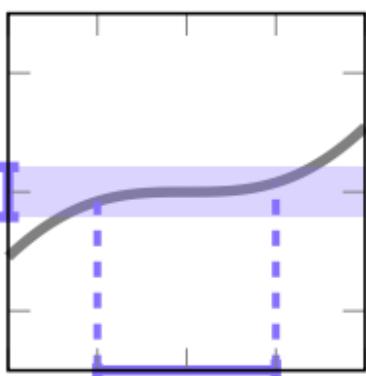
Natural Inclusion



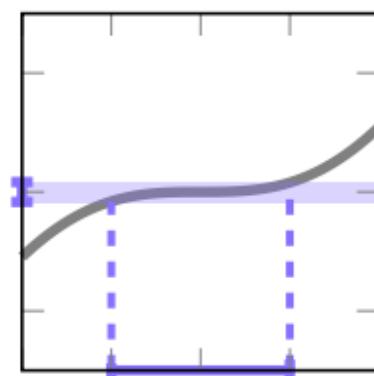
First Order



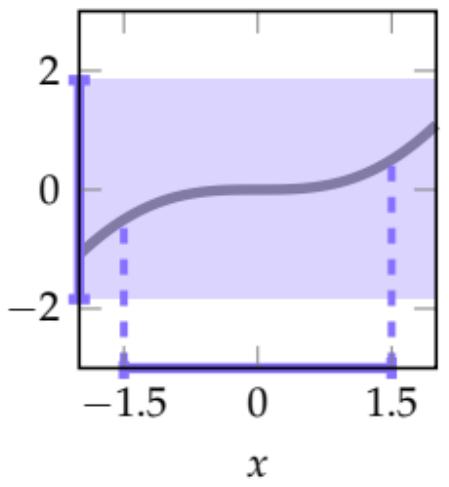
## Second Order



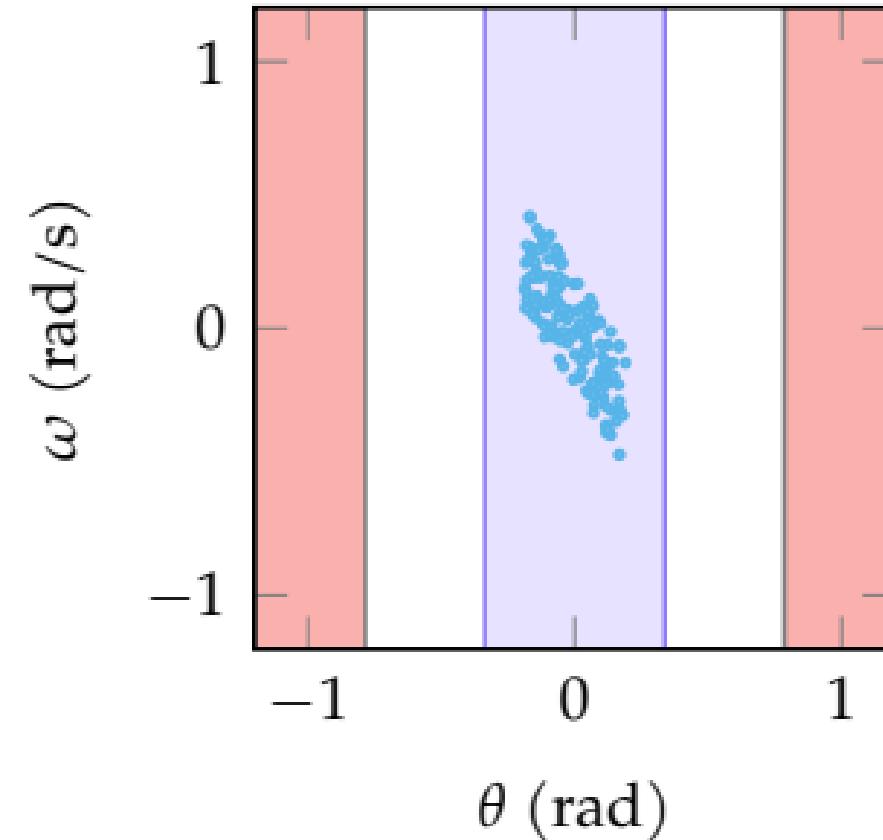
Third Order



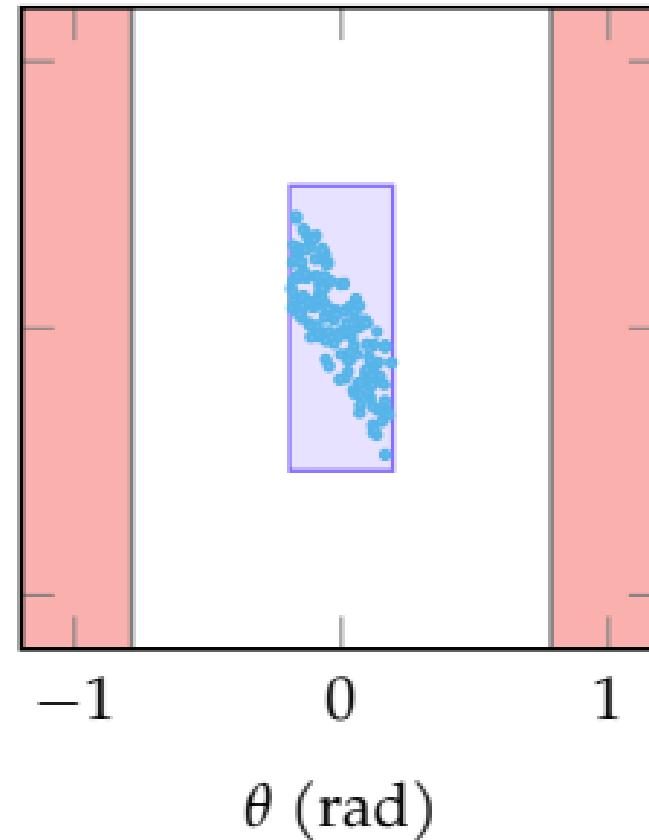
f(x)



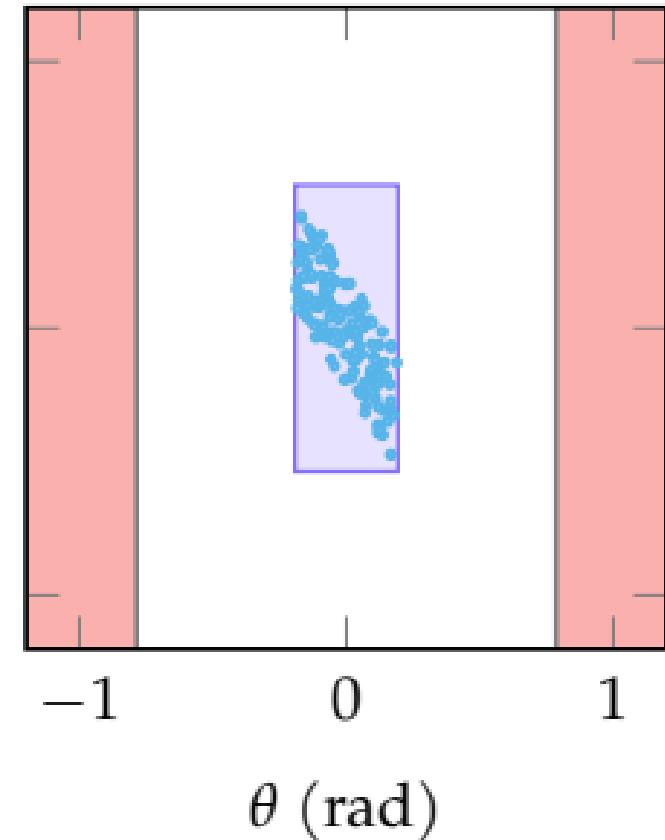
# Natural Inclusion

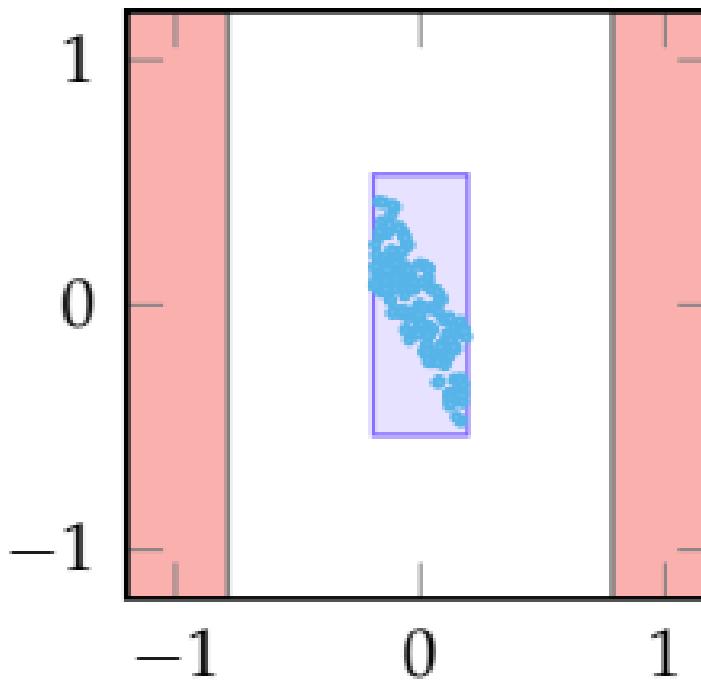
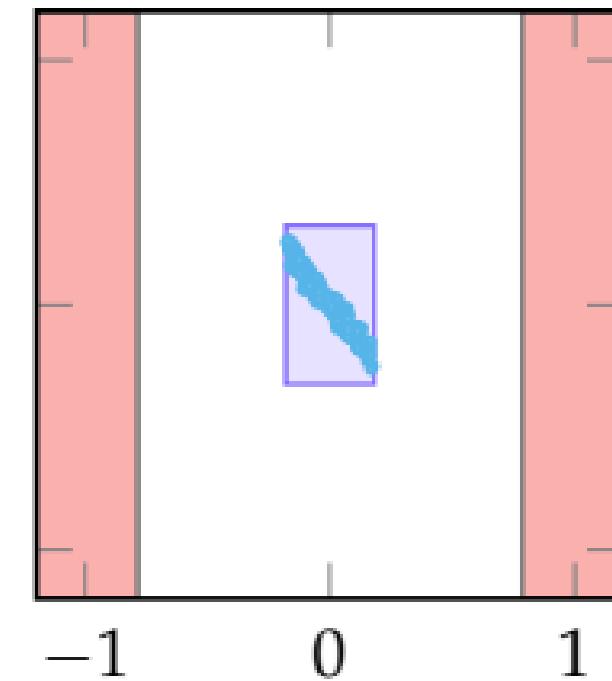
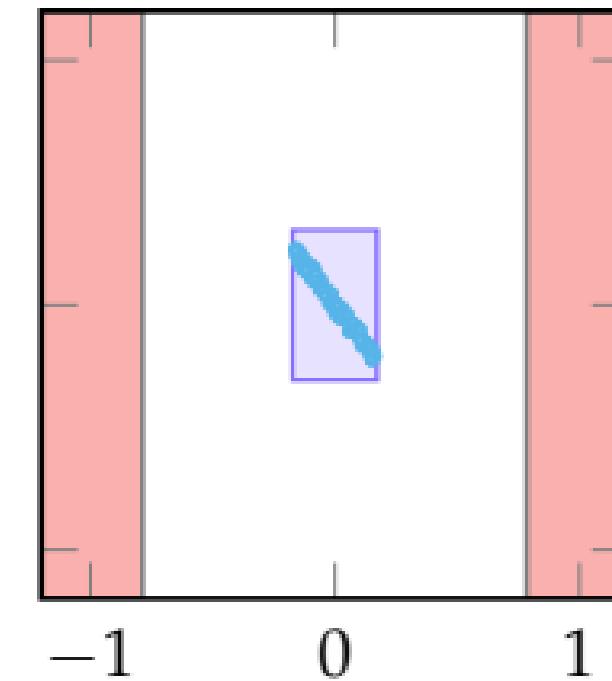
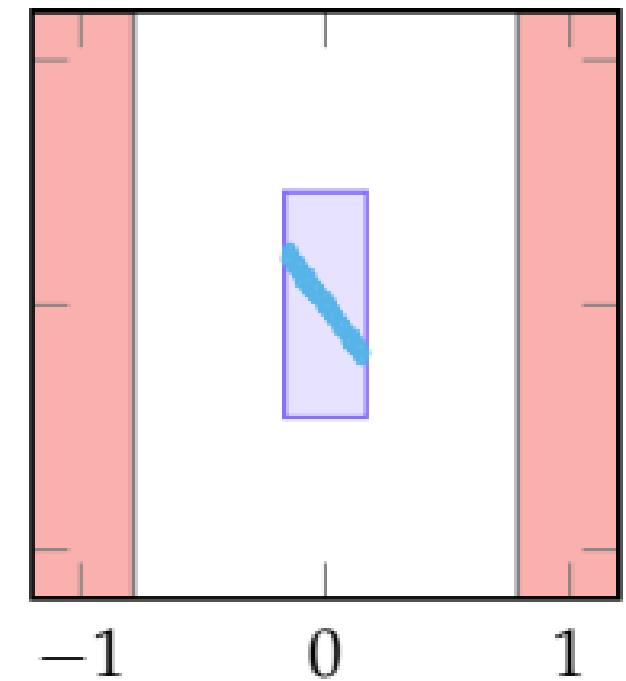


# First Order

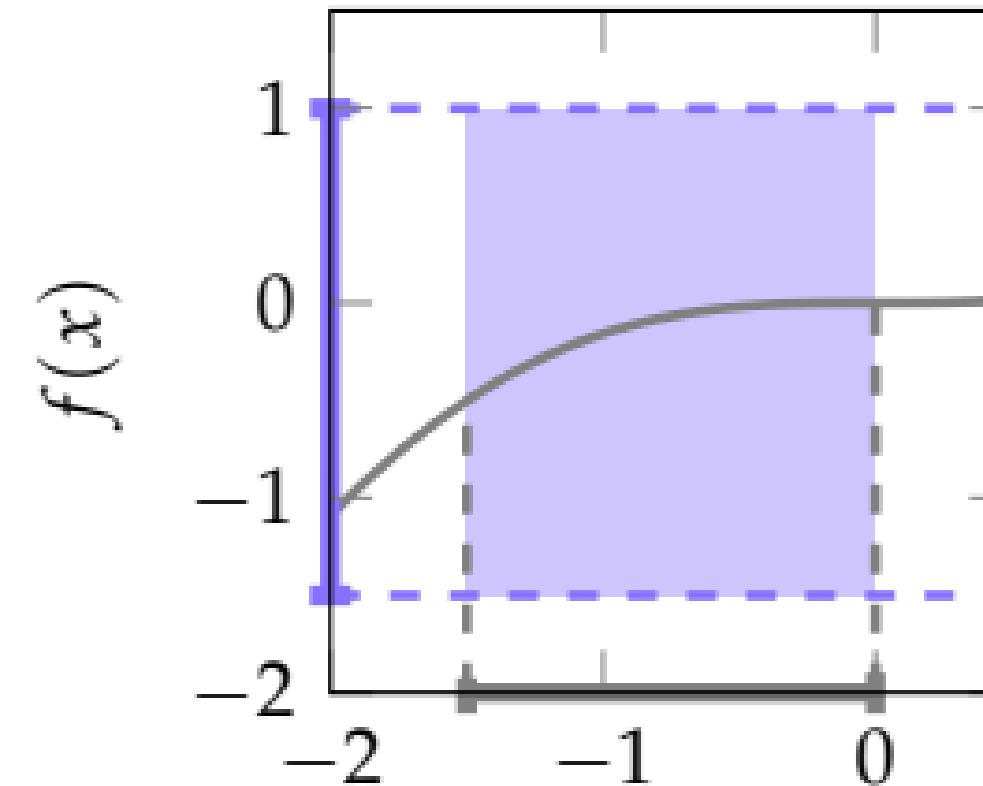


# Second Order

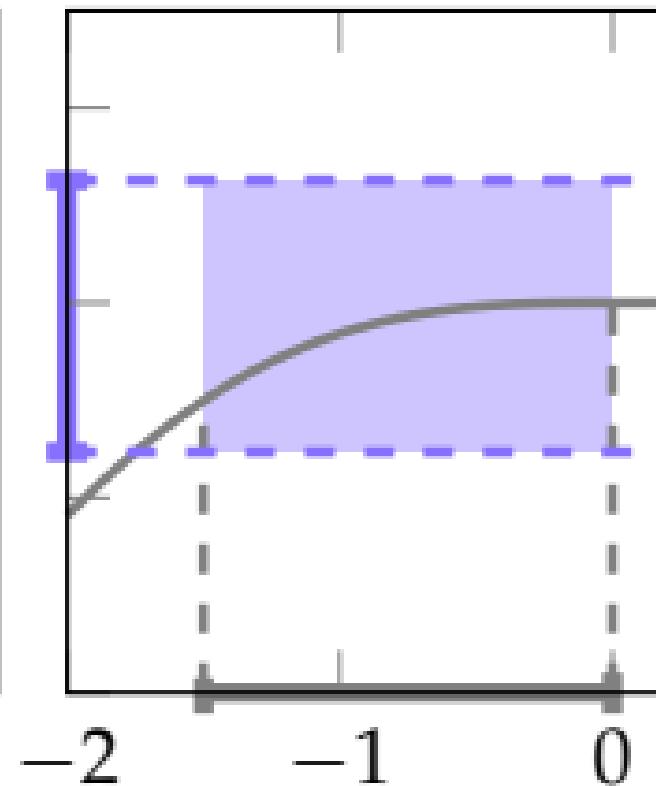


$\omega$  (rad/s) $\mathcal{R}_2$  $\mathcal{R}_3$  $\mathcal{R}_4$  $\mathcal{R}_5$  $\theta$  (rad) $\theta$  (rad) $\theta$  (rad) $\theta$  (rad)

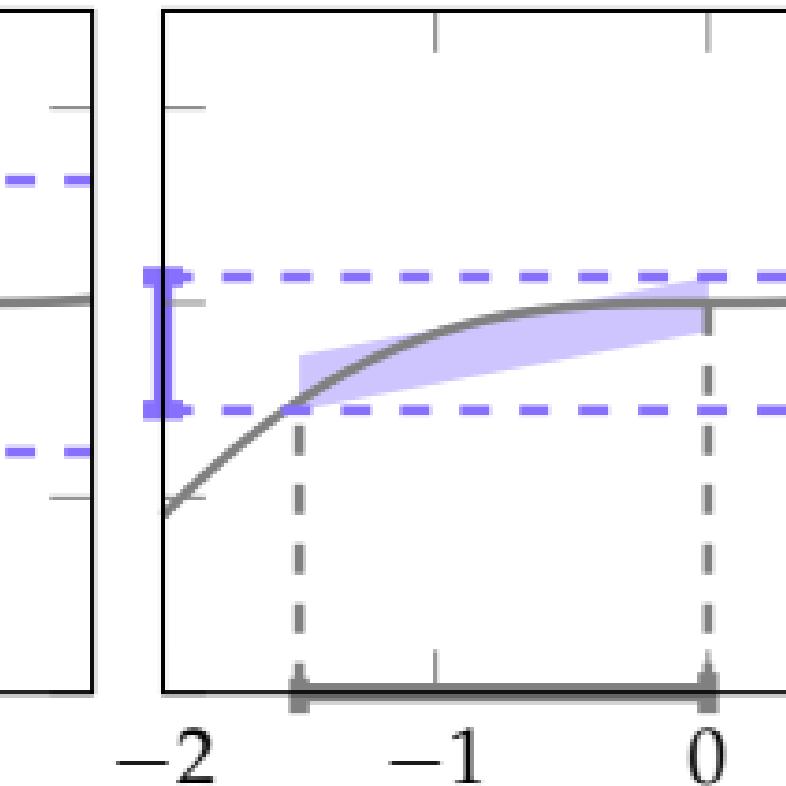
Zero Order



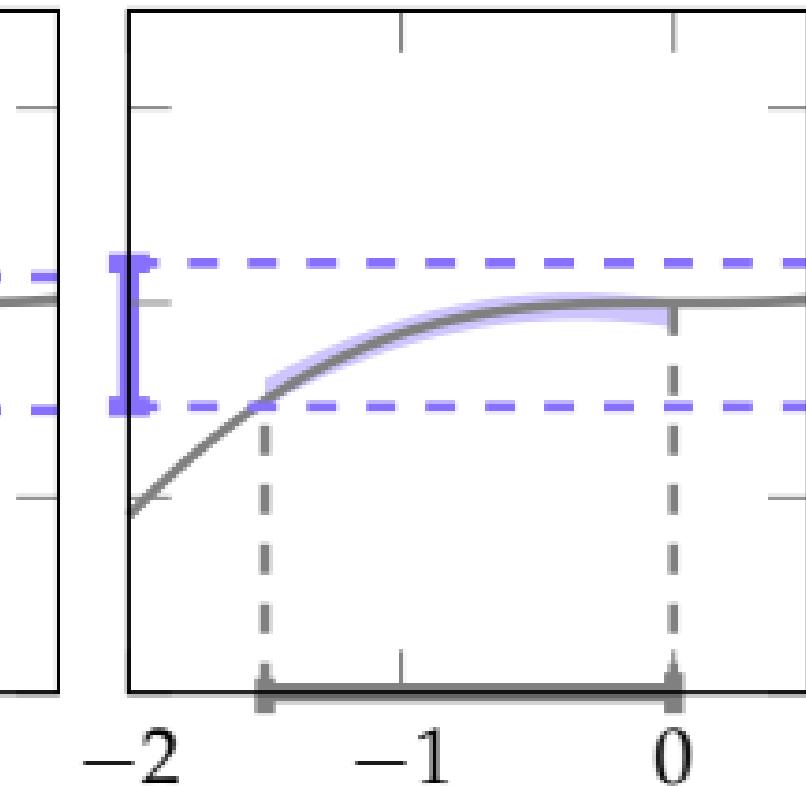
First Order



Second Order

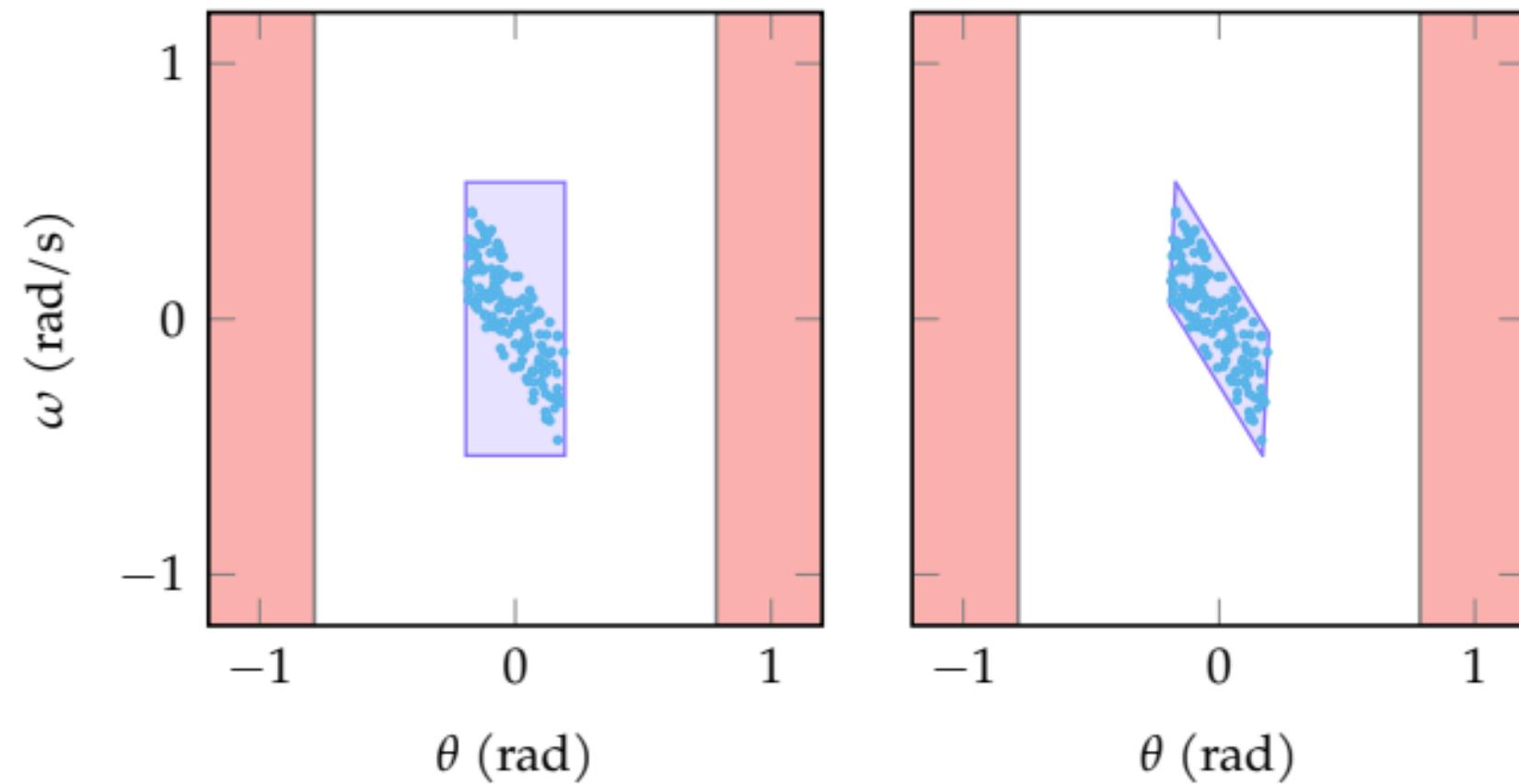


Third Order

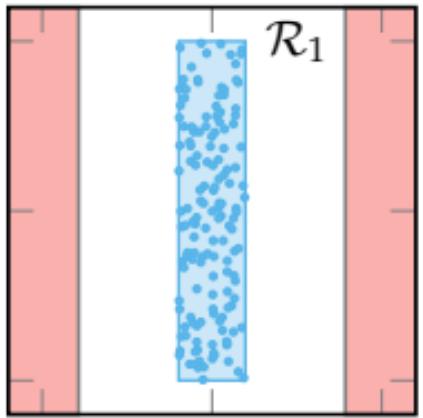


# Taylor Inclusion

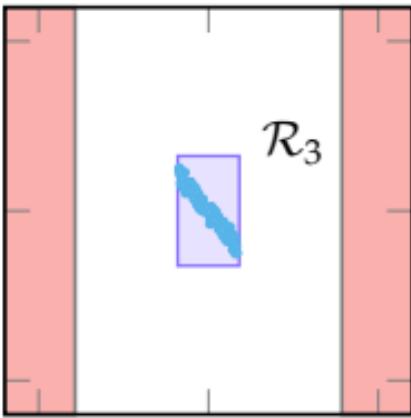
# Conservative Linearization



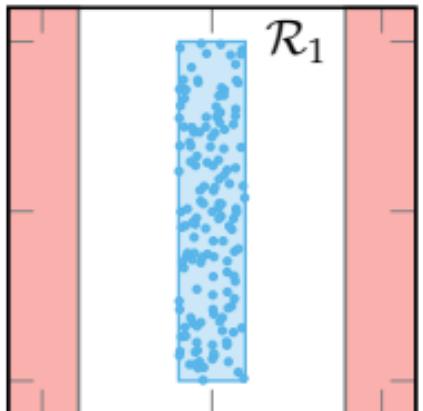
Symbolic



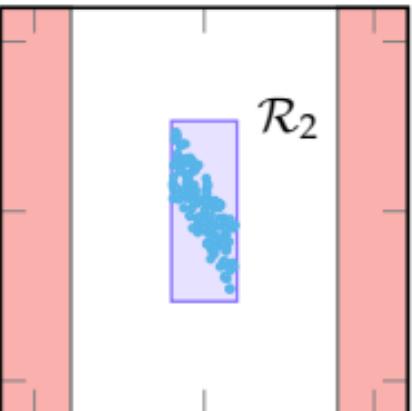
$$r(\mathbf{s}_1, \mathbf{x}_{1:3})$$



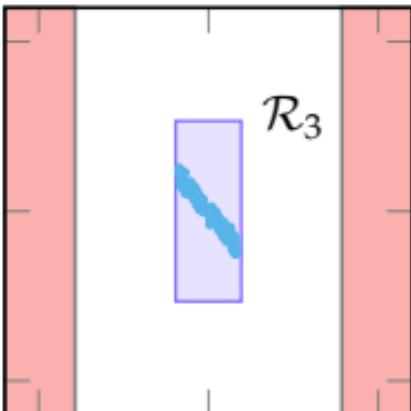
Concrete



$$r(\mathbf{s}_1, \mathbf{x}_{1:2})$$

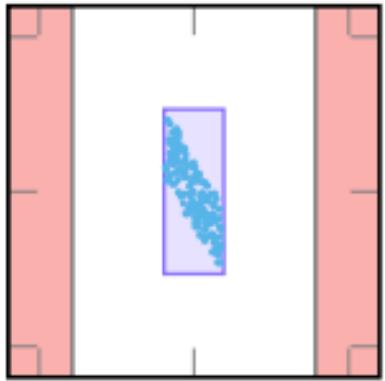


$$r(\mathbf{s}_2, \mathbf{x}_{2:3})$$

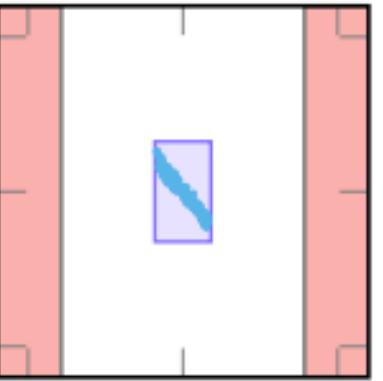


Symbolic

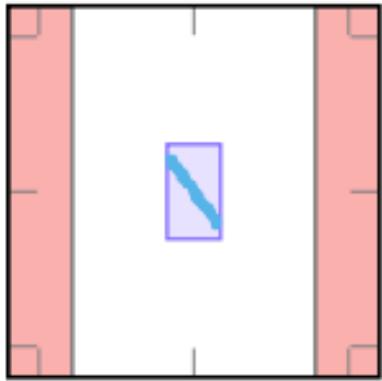
$\mathcal{R}_2$



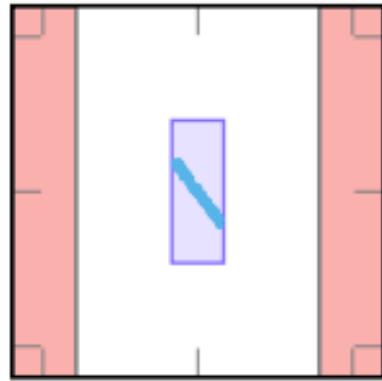
$\mathcal{R}_3$



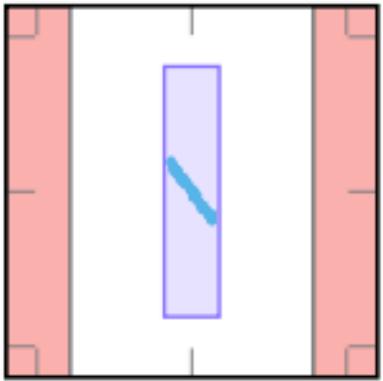
$\mathcal{R}_4$



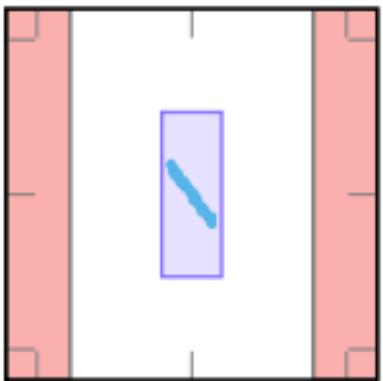
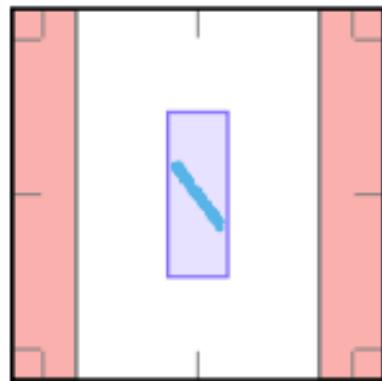
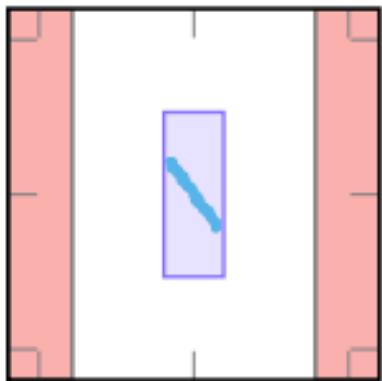
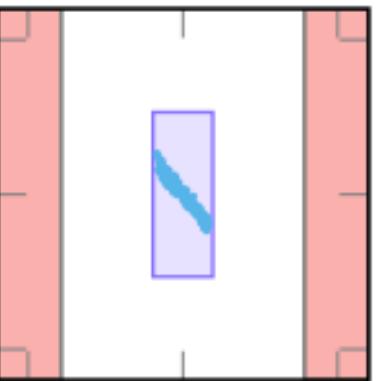
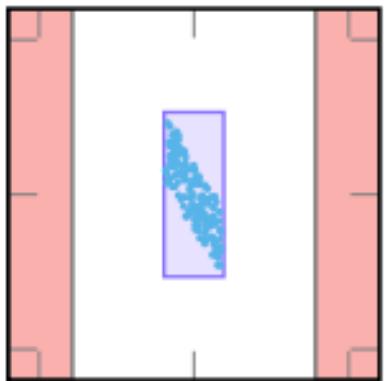
$\mathcal{R}_5$



$\mathcal{R}_6$

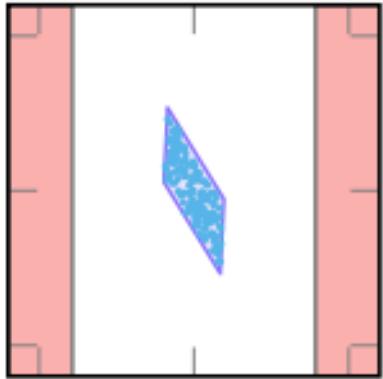


Concrete

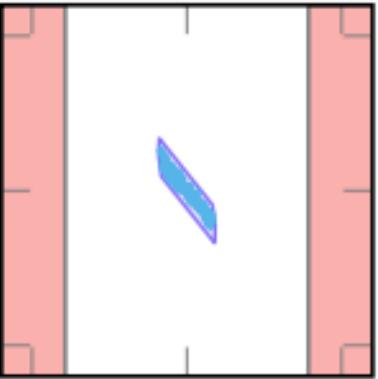


Symbolic

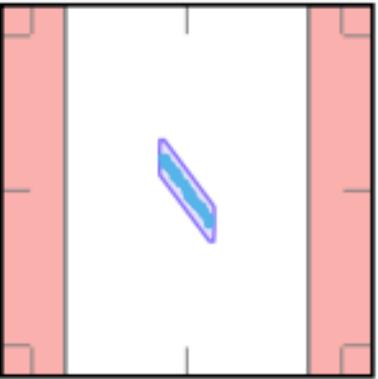
$\mathcal{R}_2$



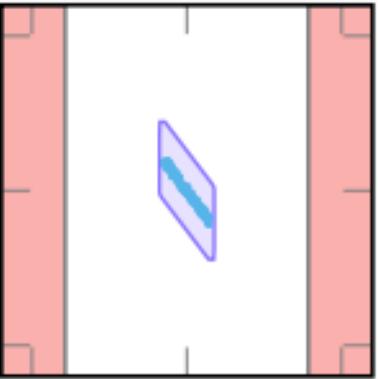
$\mathcal{R}_3$



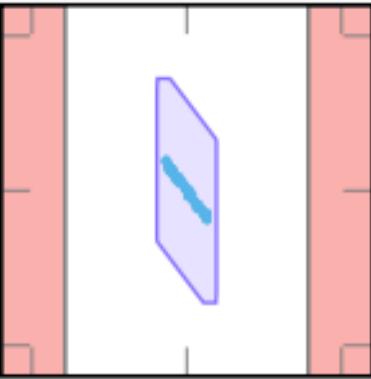
$\mathcal{R}_4$



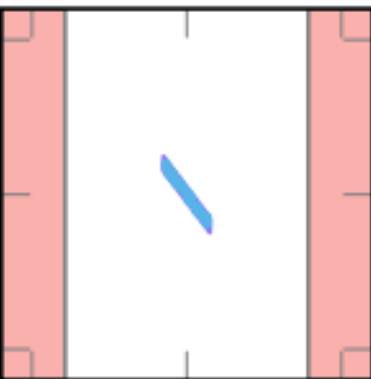
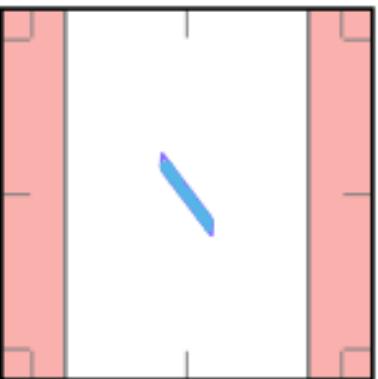
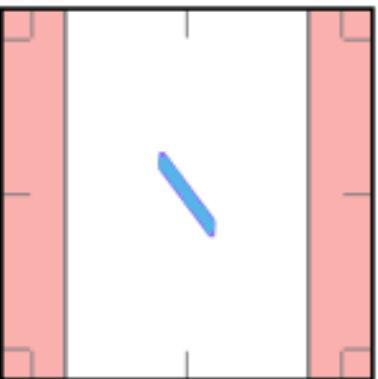
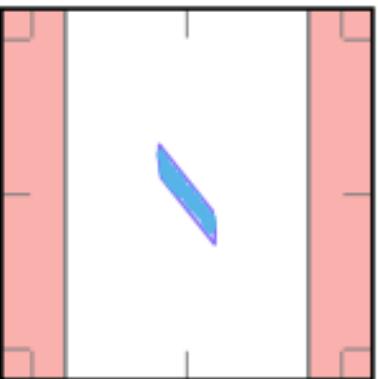
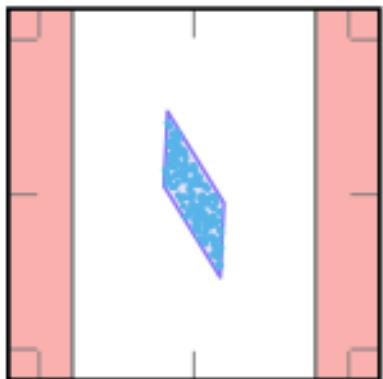
$\mathcal{R}_5$

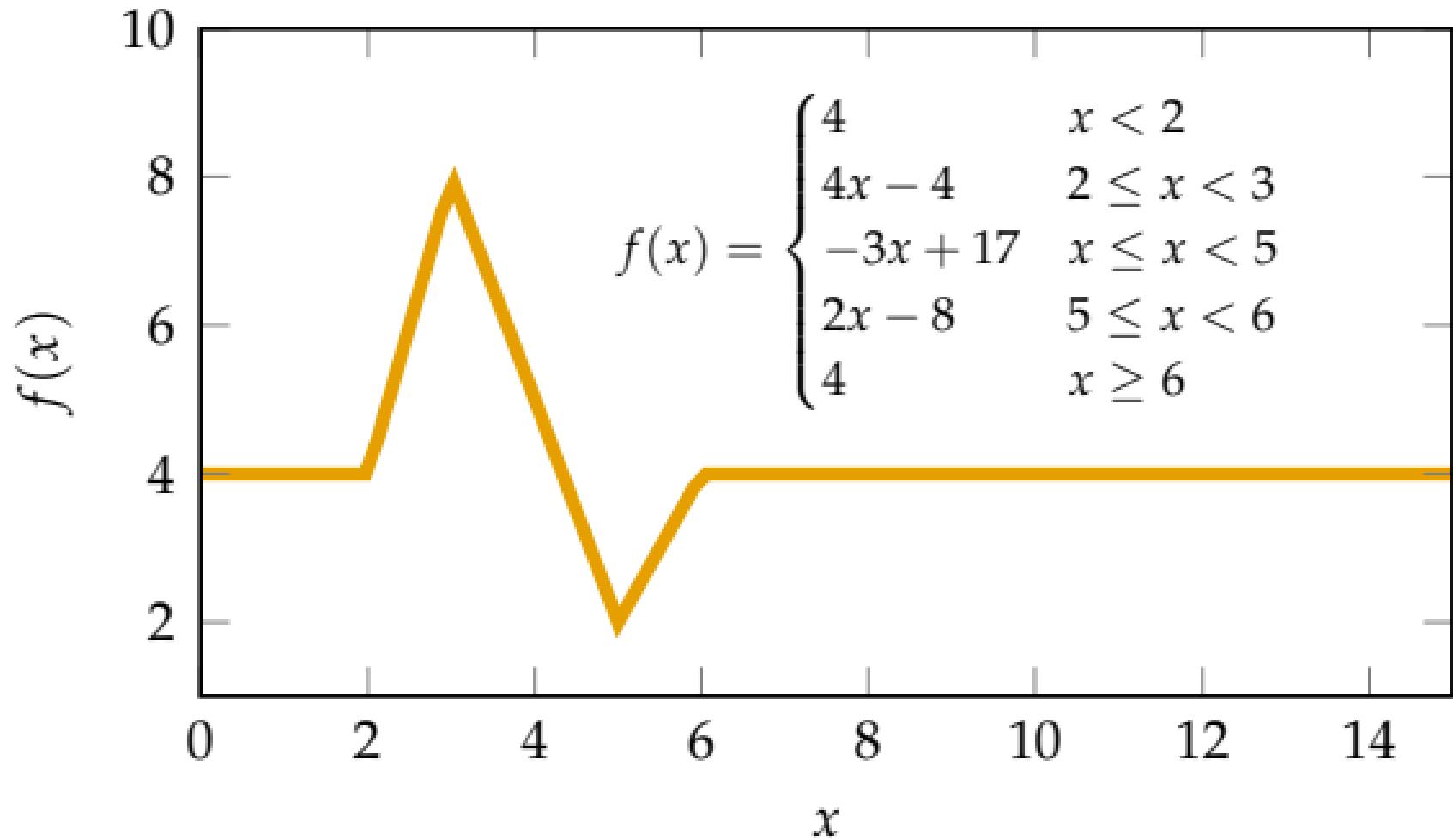


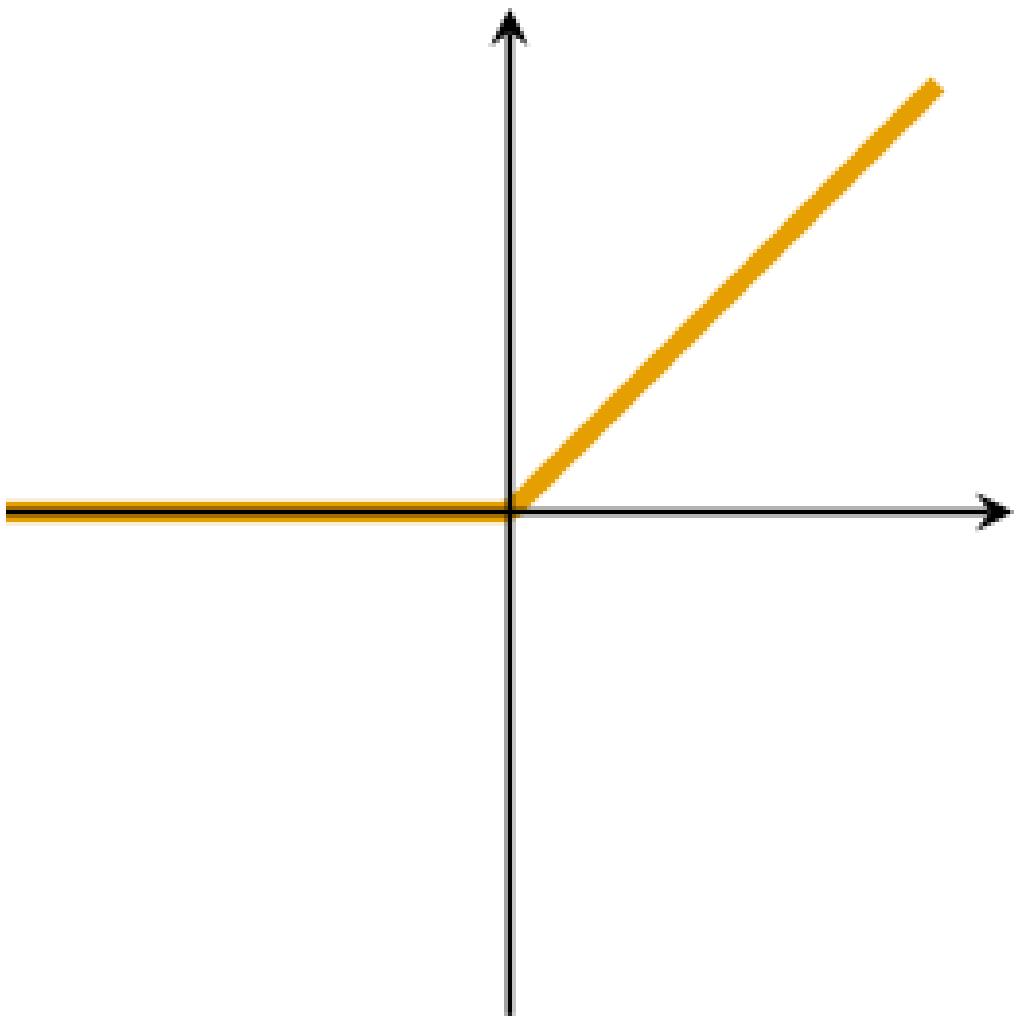
$\mathcal{R}_6$



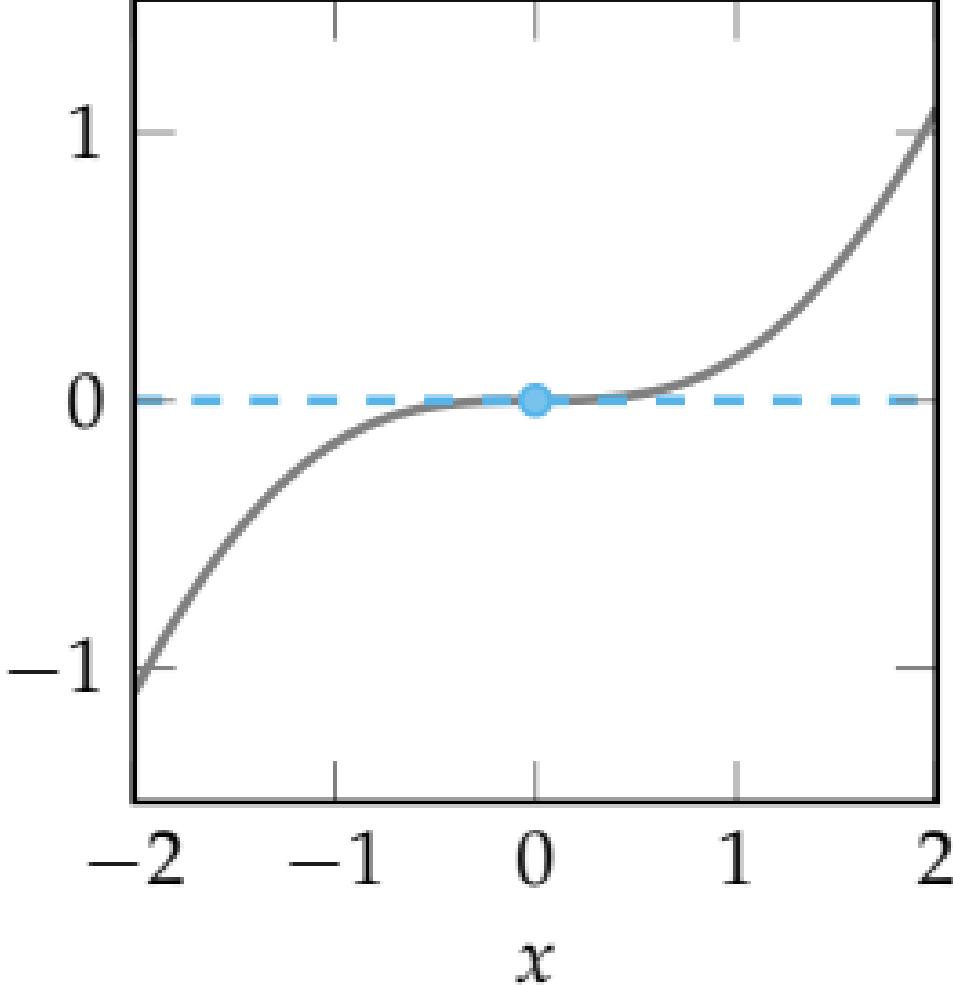
Concrete



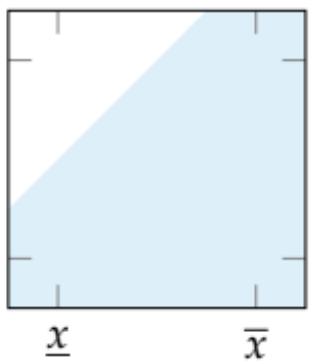




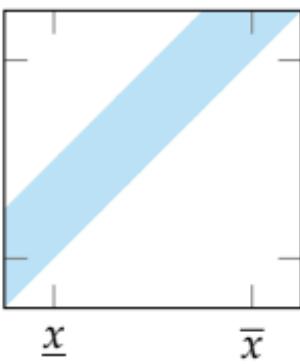
$f(x)$



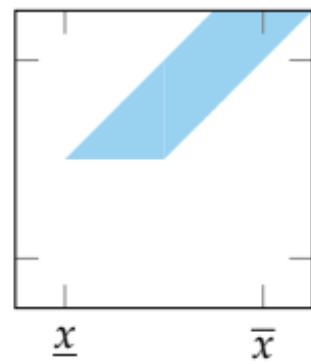
$$y \leq x - \underline{x}$$



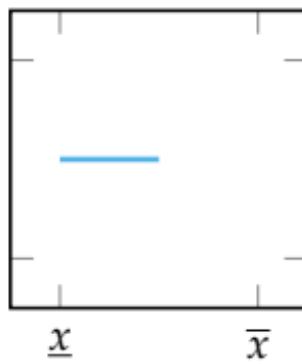
$$y \geq x$$



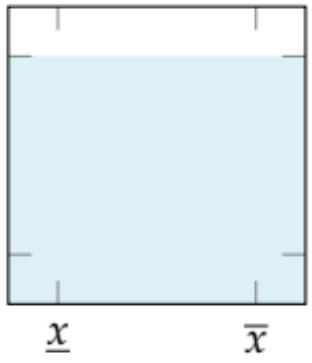
$$y \geq 0$$



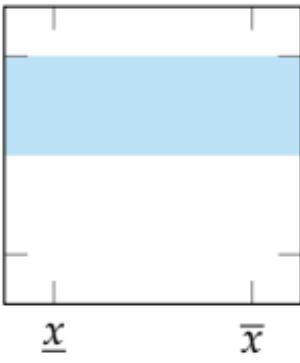
$$y \leq 0$$



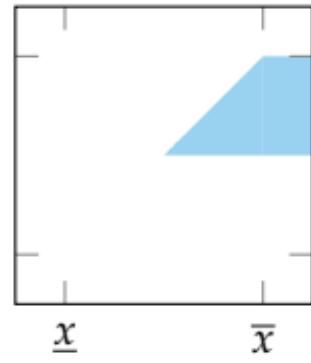
$$y \leq \bar{x}$$



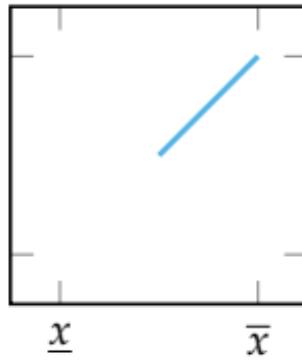
$$y \geq 0$$

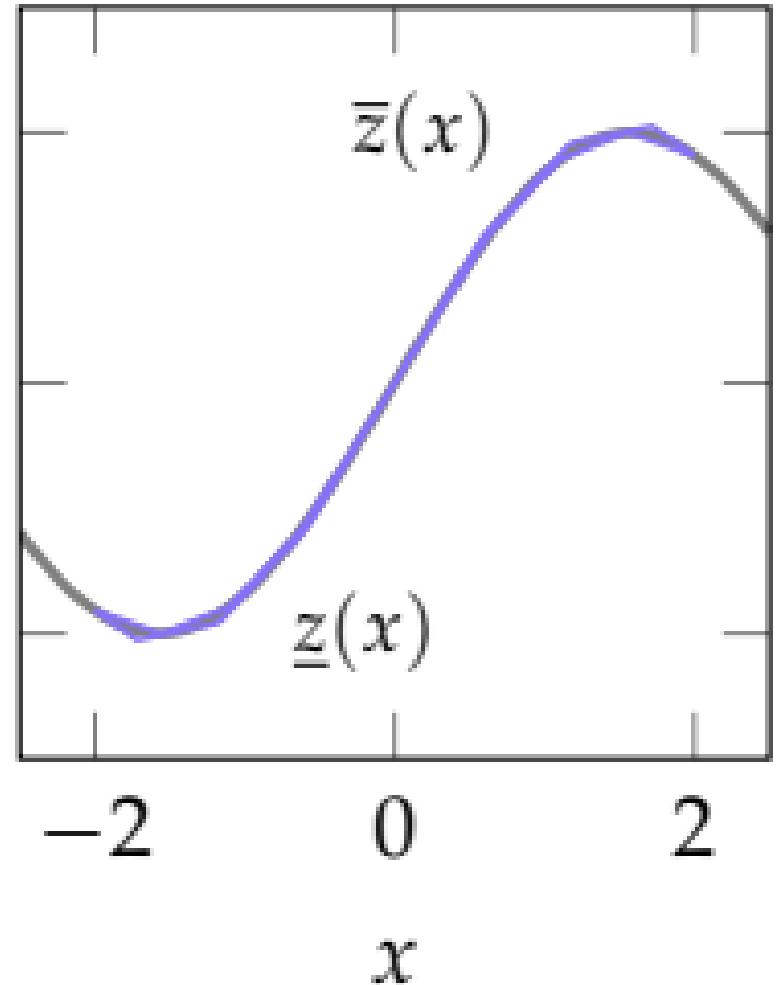
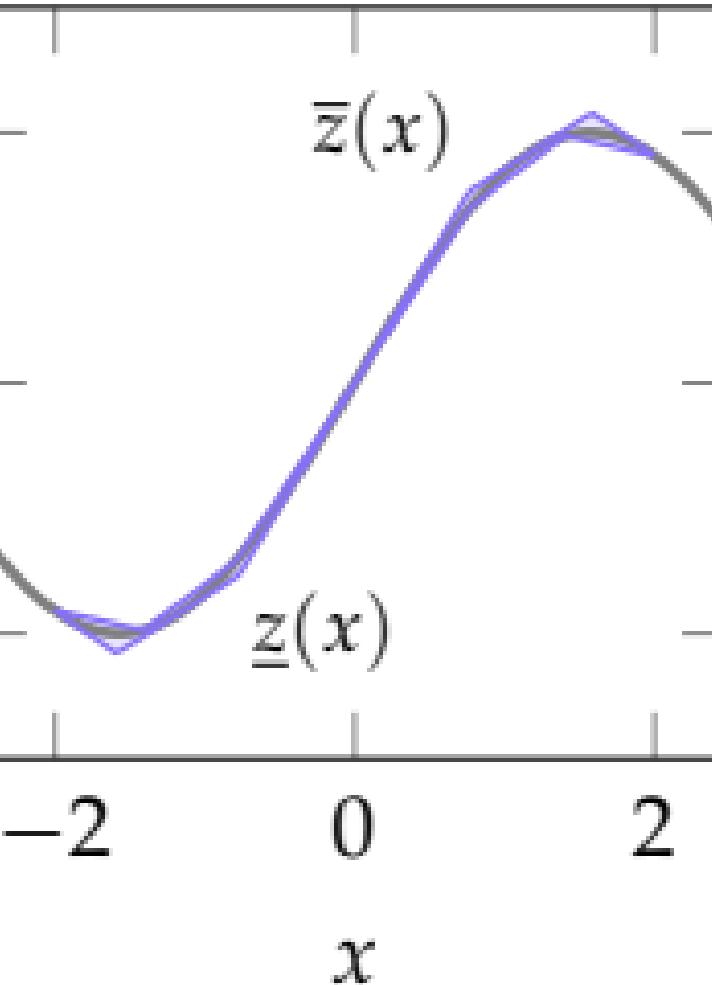
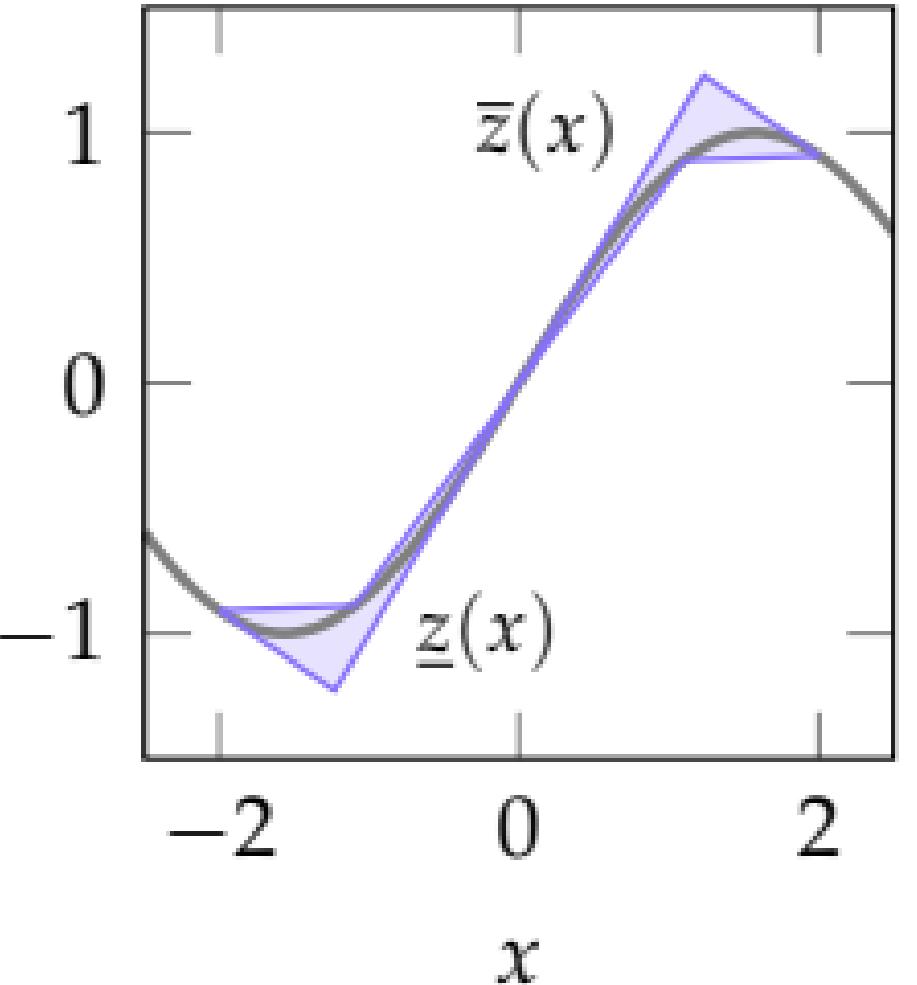


$$y \leq x$$

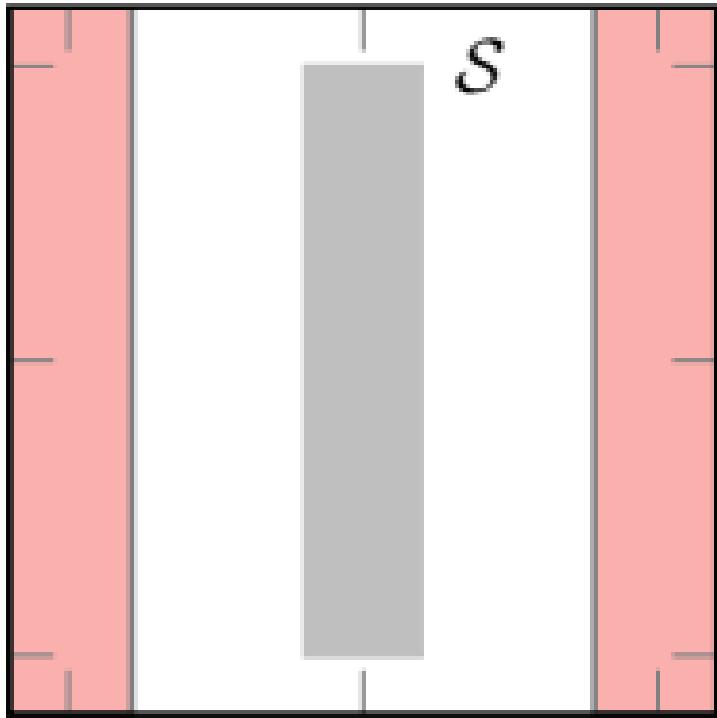


$$y \geq x$$

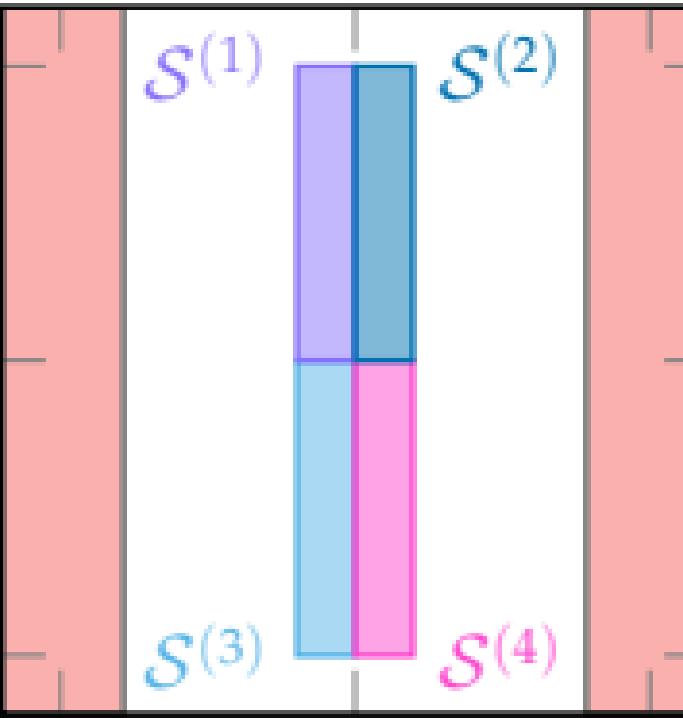


$f(x)$ 

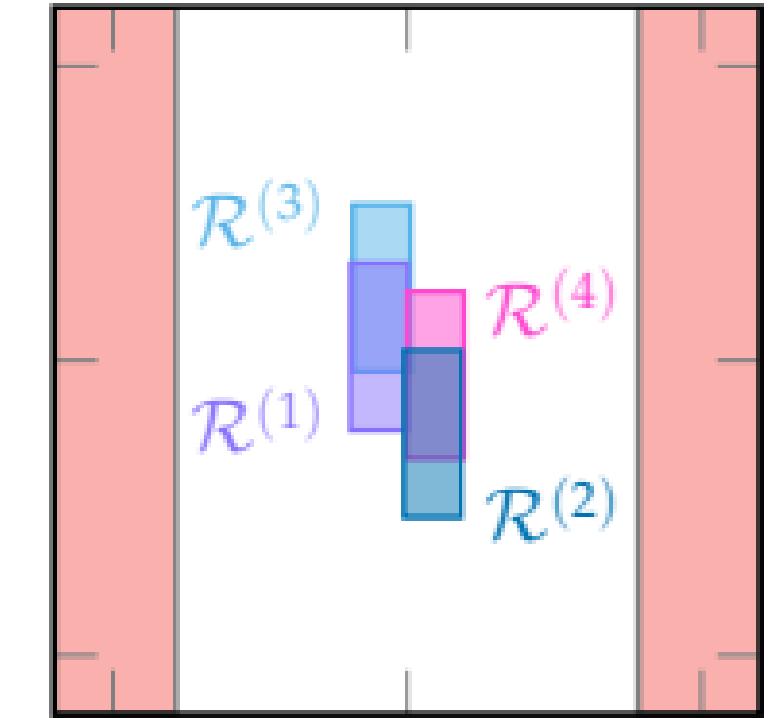
# Input Set



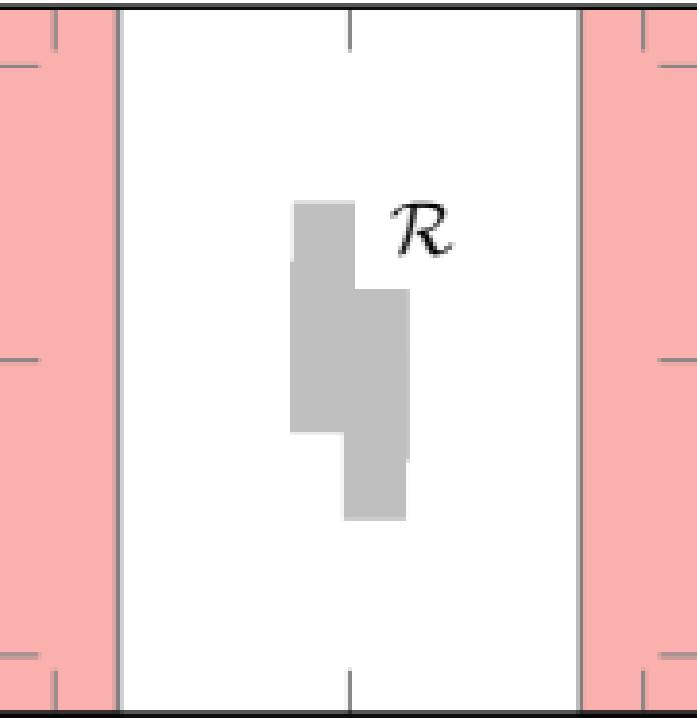
# Input Partition



# Output Partition

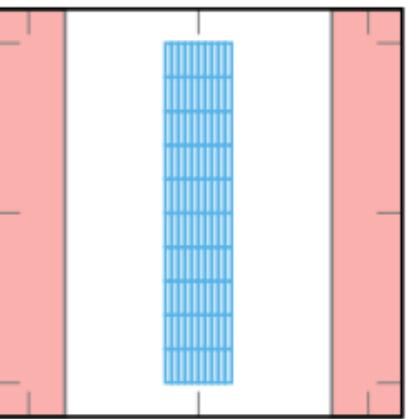
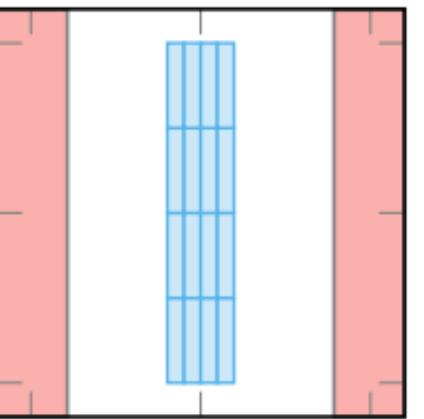
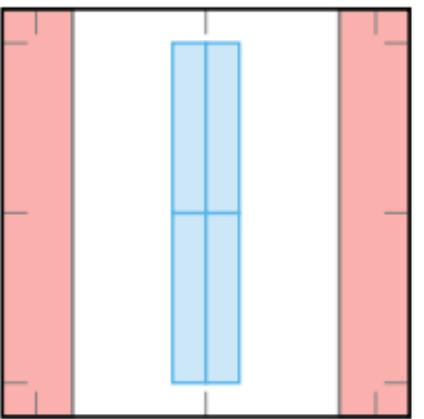
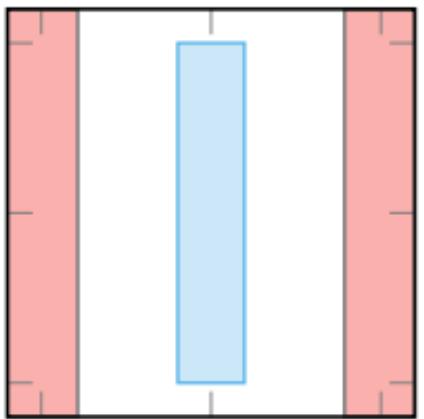


# Output Set

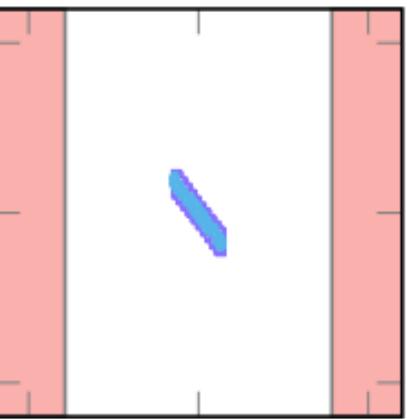
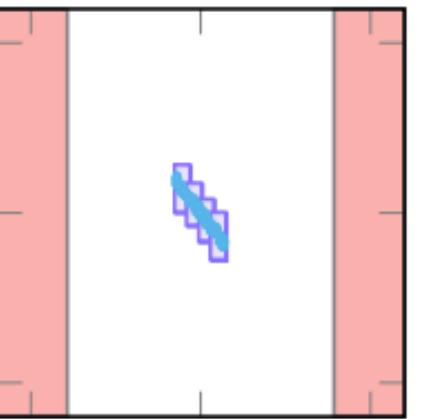
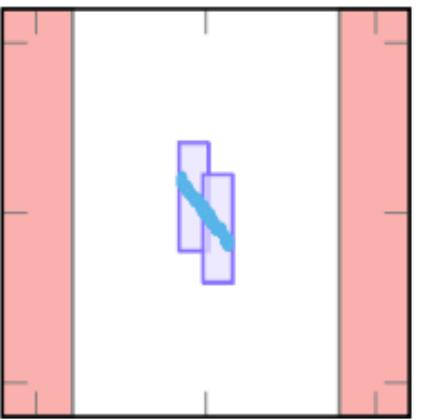
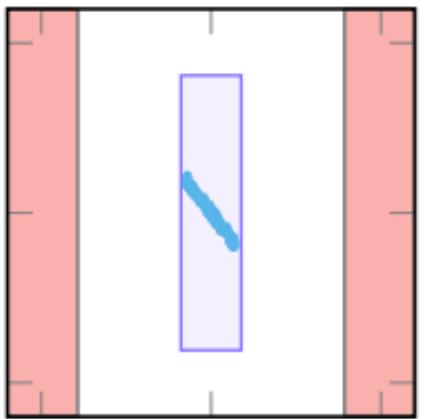


$m = 1$  $m = 4$  $m = 16$  $m = 100$ 

Input Partition



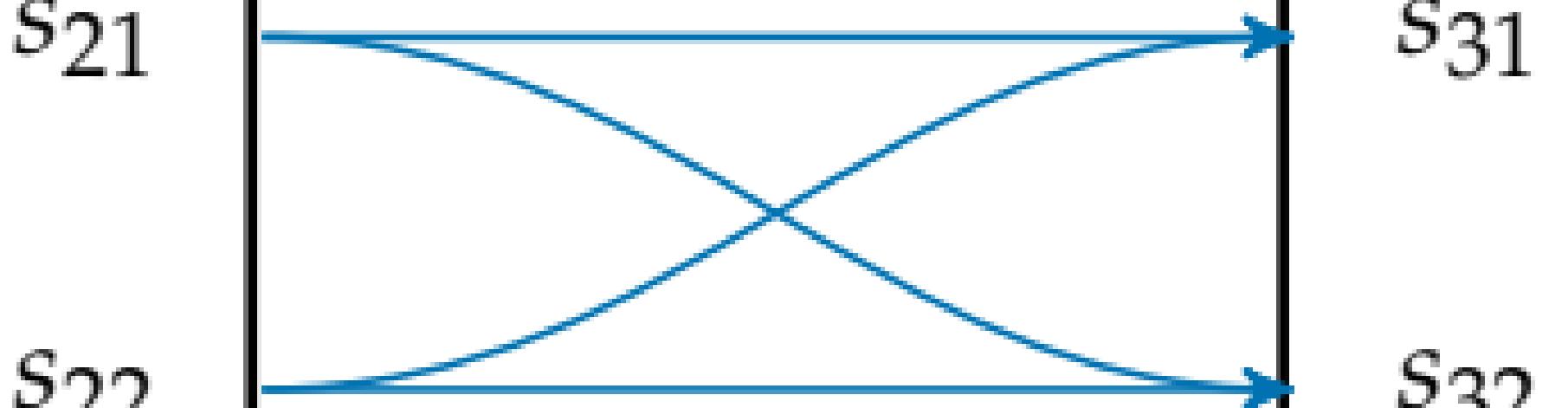
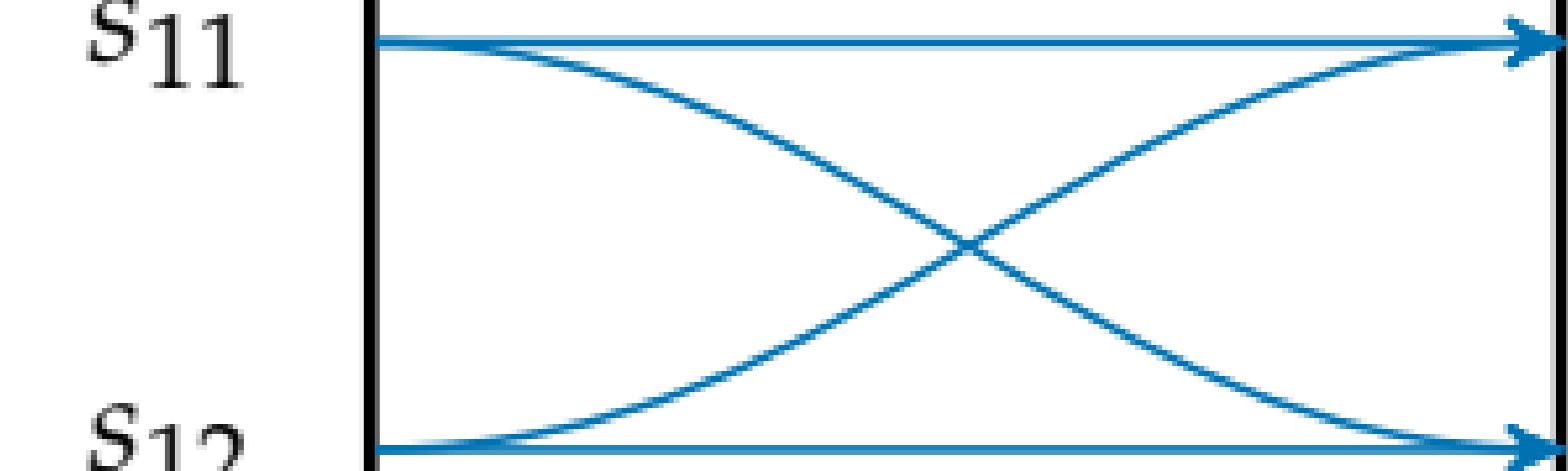
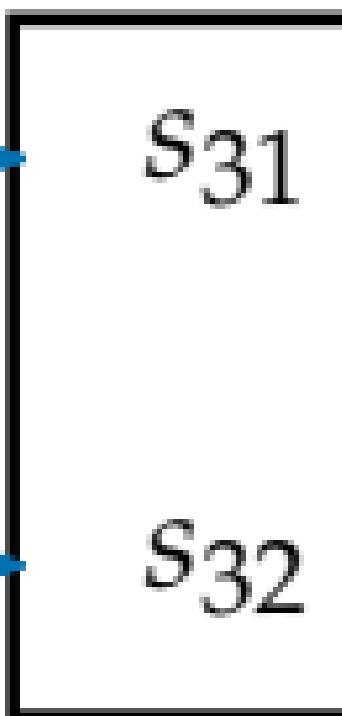
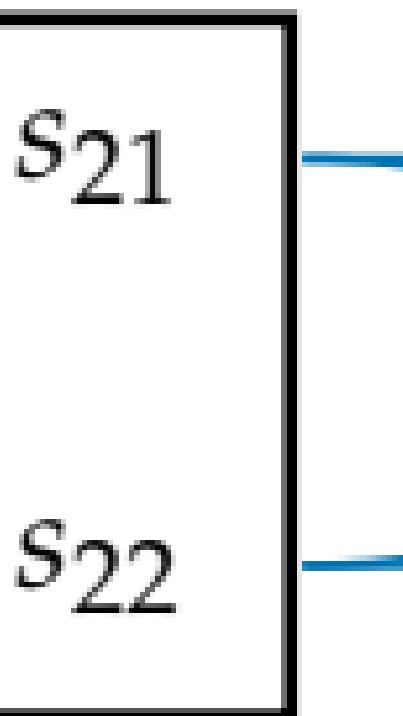
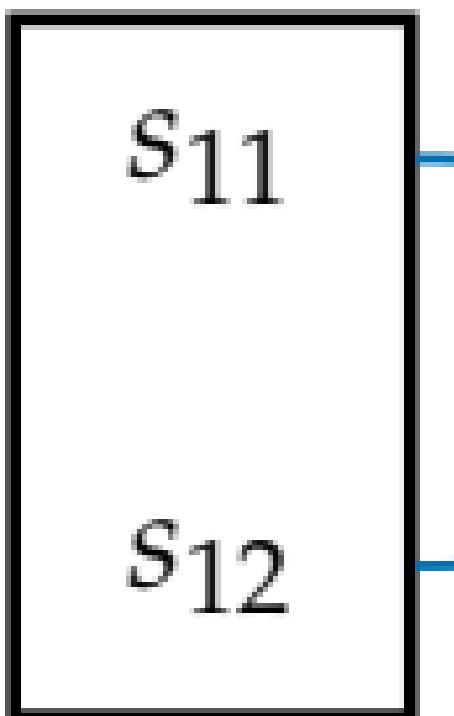
Output Partition



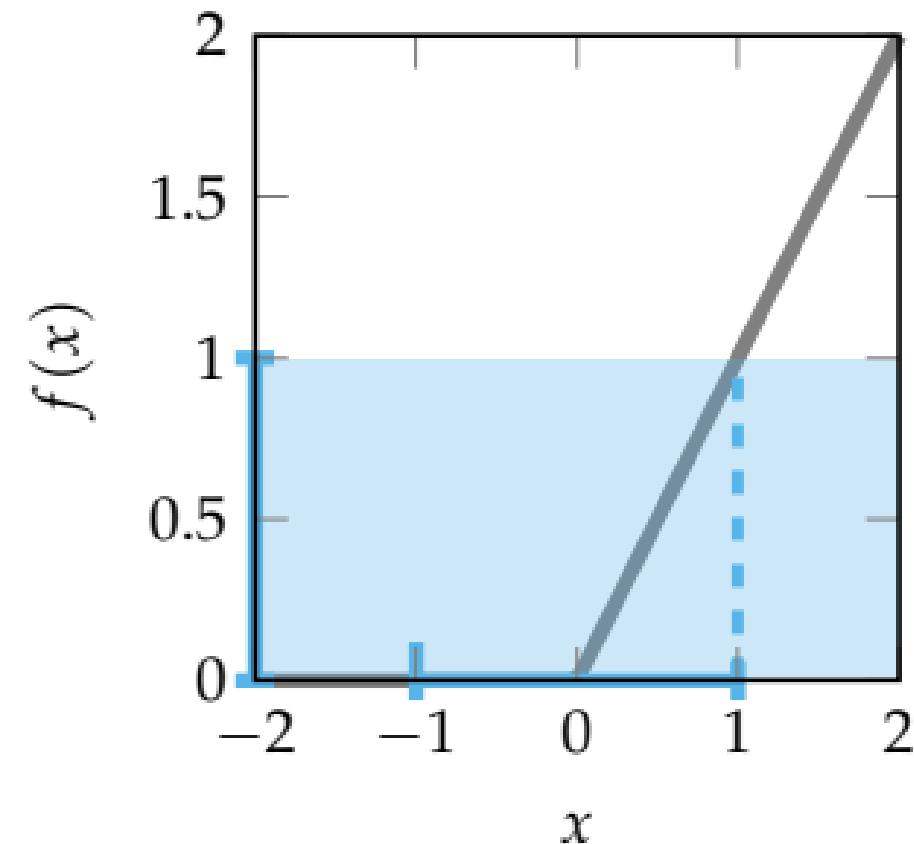
$s_1$ 

$$s_2 = \phi(W_1 s_1 + b_1)$$

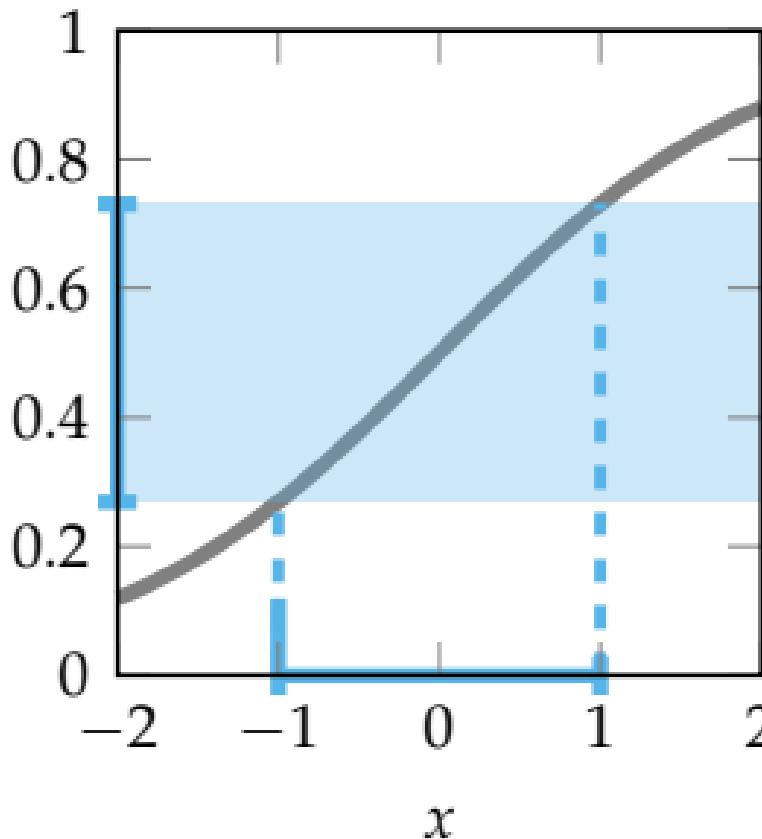
$$s_3 = \phi(W_2 s_2 + b_2)$$



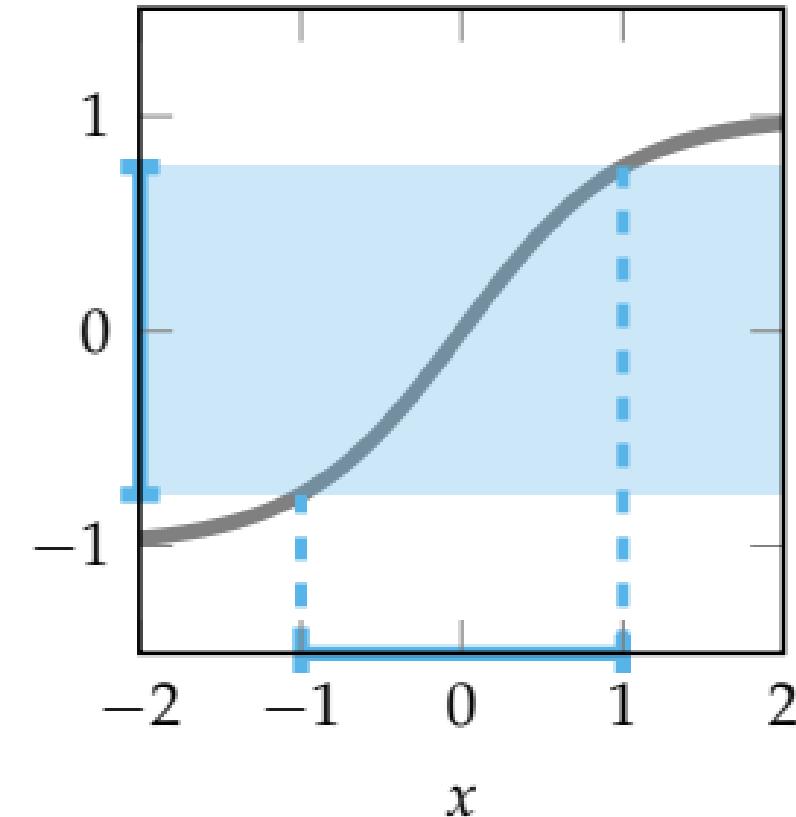
$$f(x) = \text{ReLU}(x)$$



$$f(x) = \text{sigmoid}(x)$$



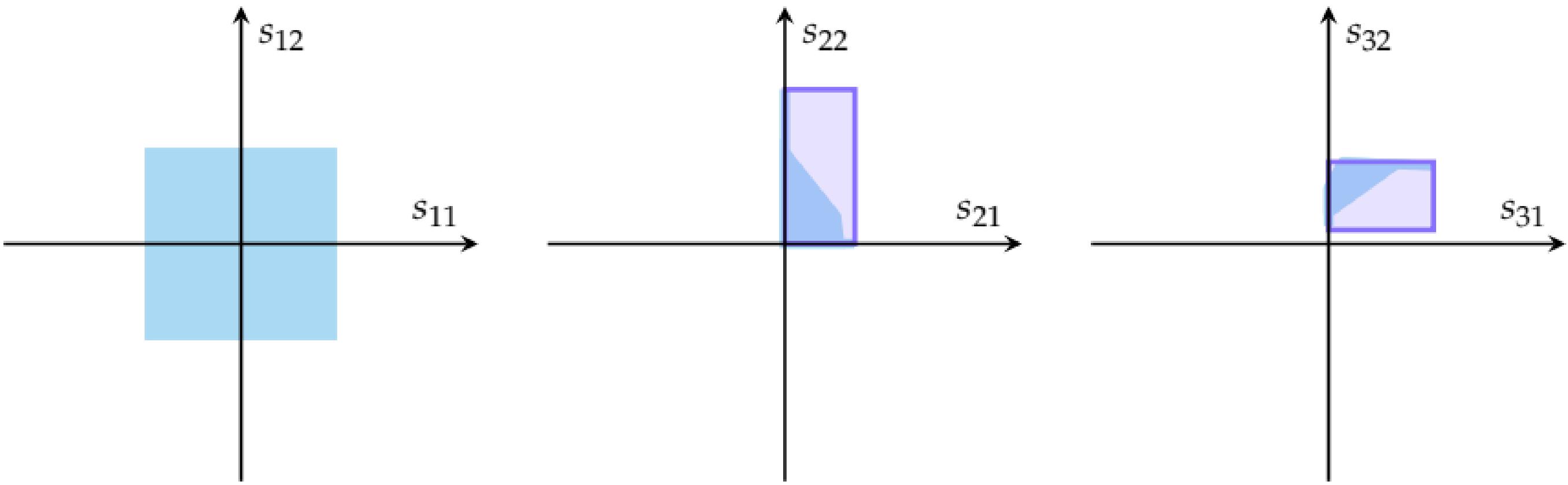
$$f(x) = \tanh(x)$$



Input Set

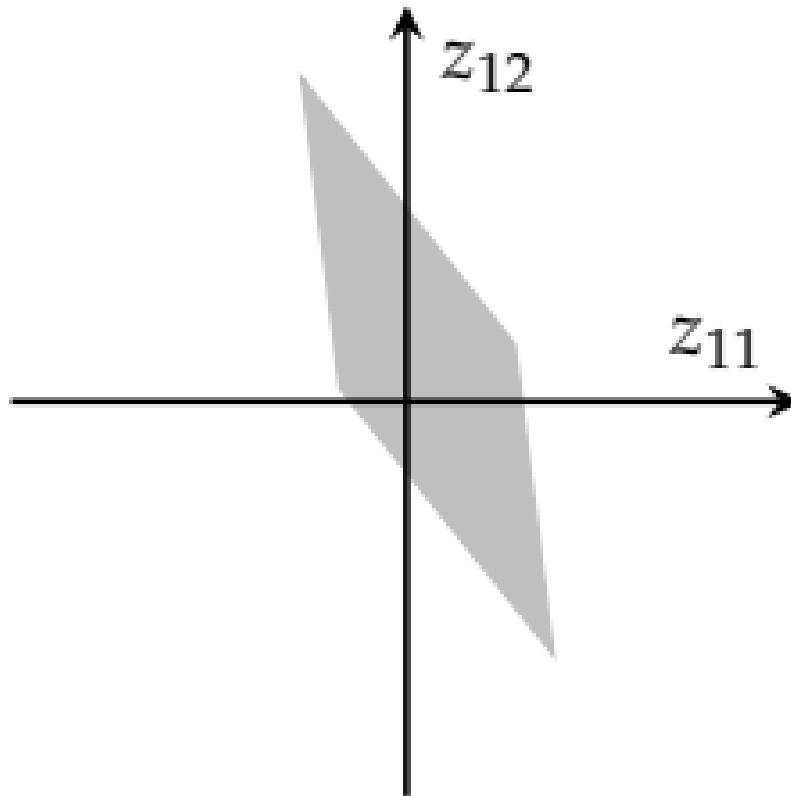
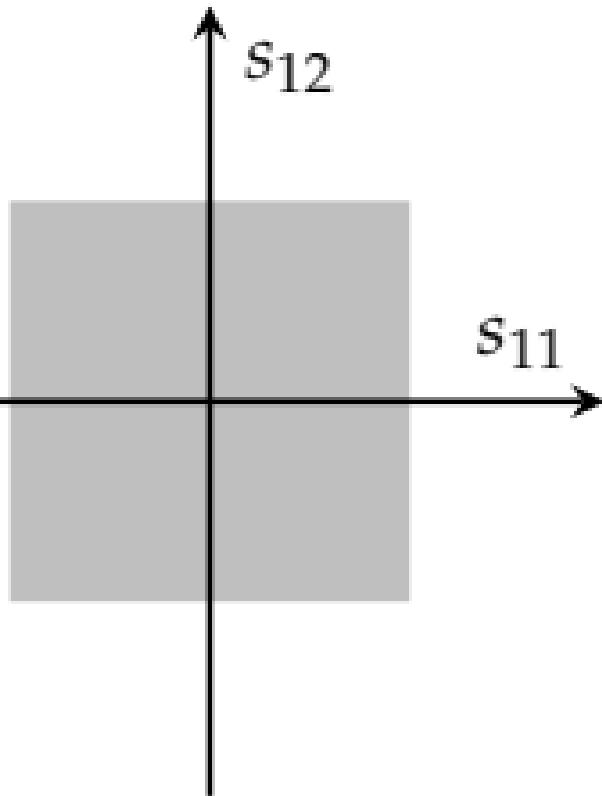
Layer 1 Output

Output Set

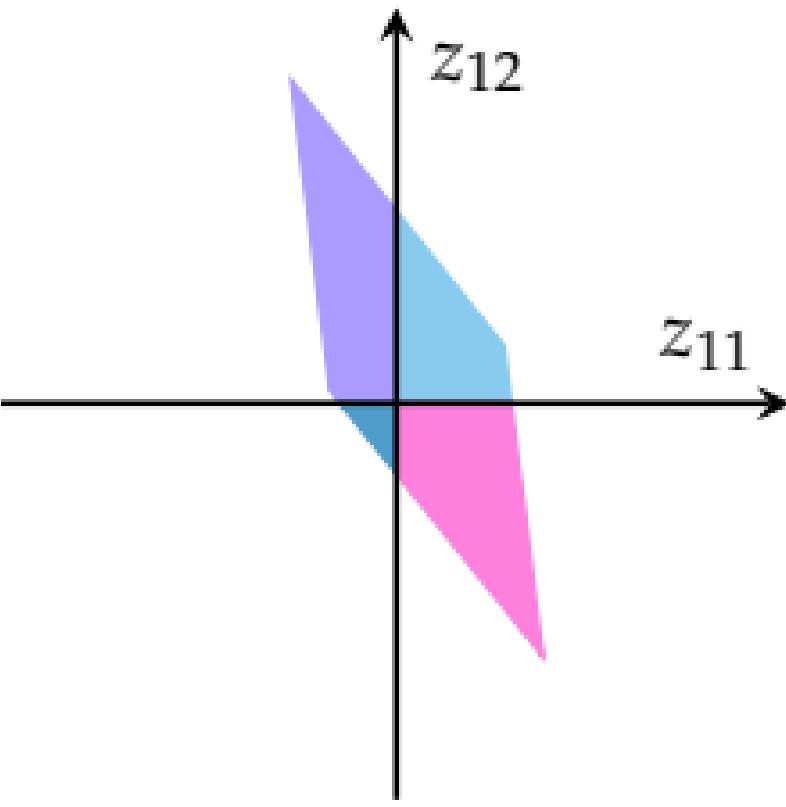


# Input Set

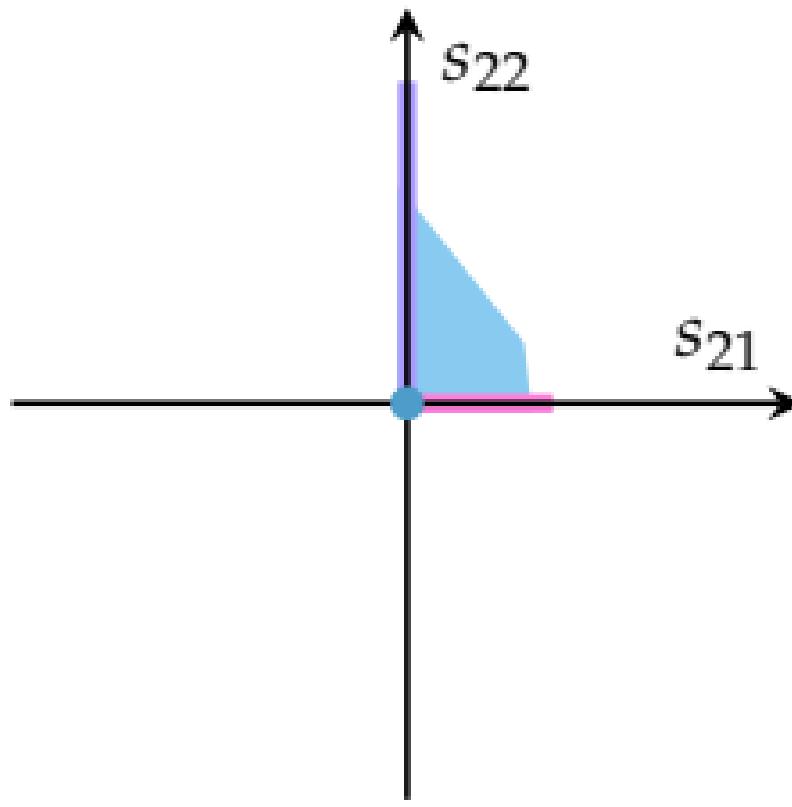
# Linear Transformation



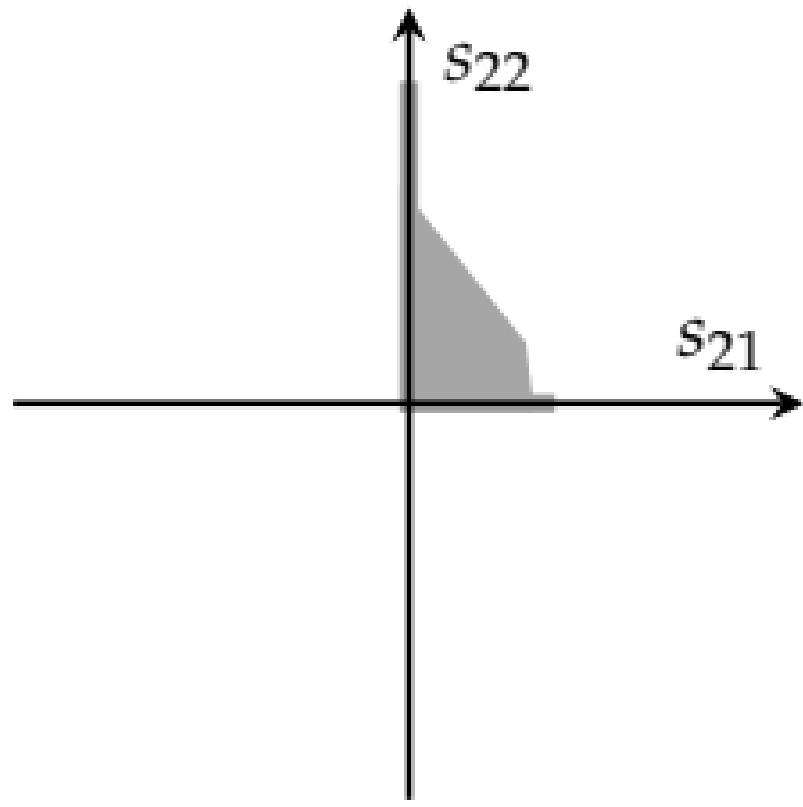
# Activation Patterns

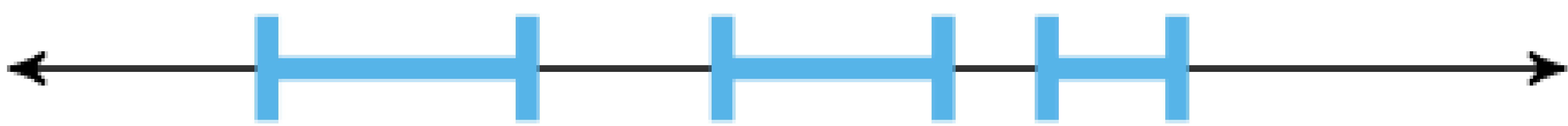


# Output Subsets



# Output Set





QUESTION

-2

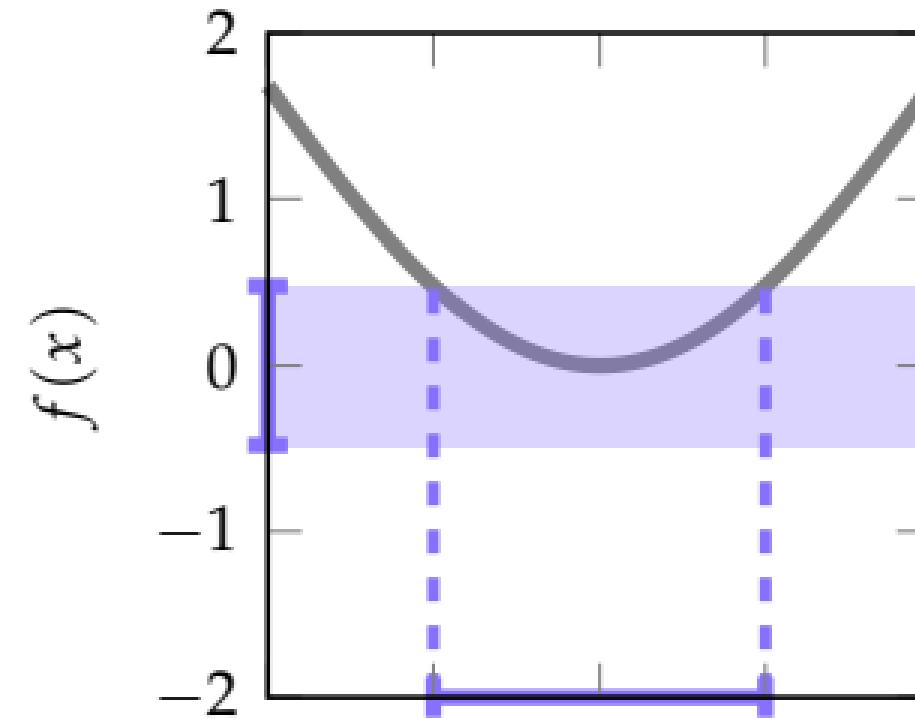
-1

-0.25

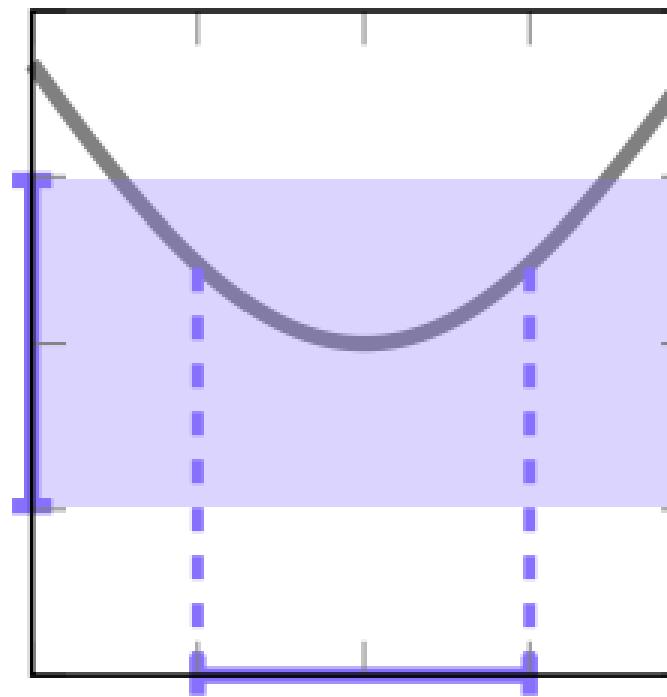
0.6 1

1.5

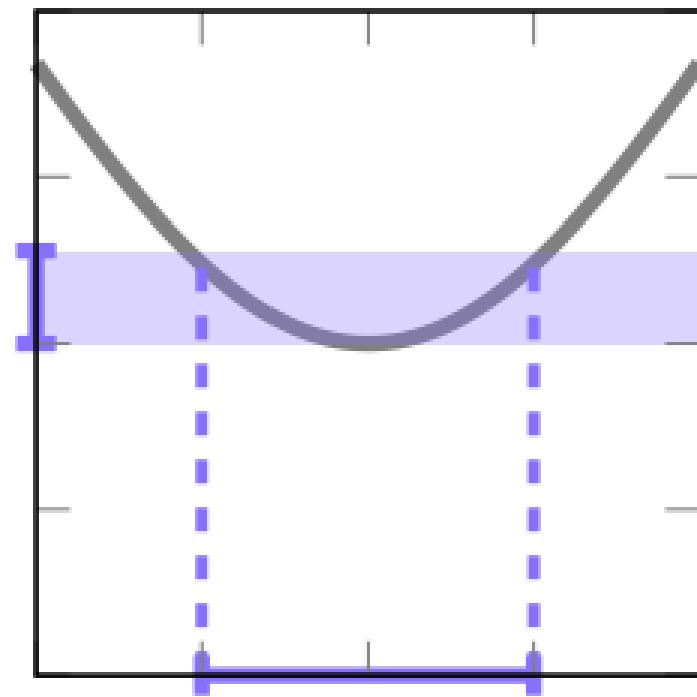
Natural  
Inclusion



Mean Value  
Inclusion

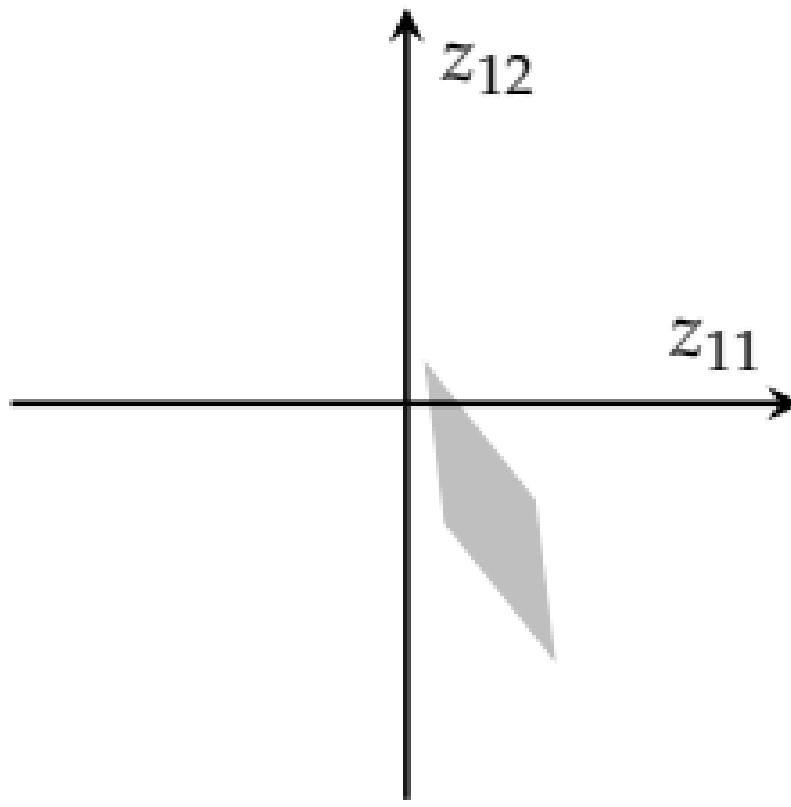
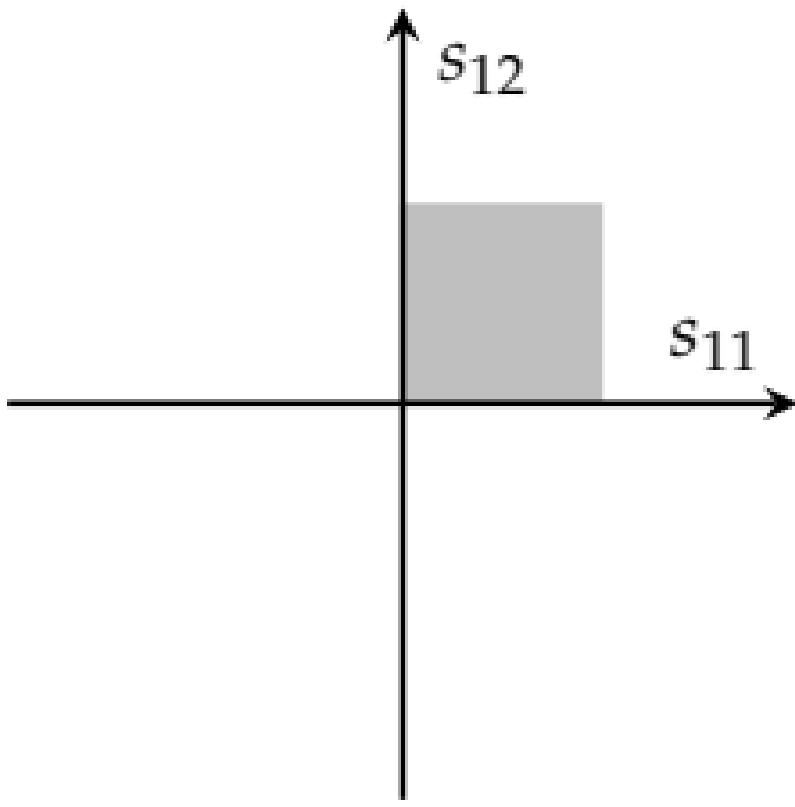


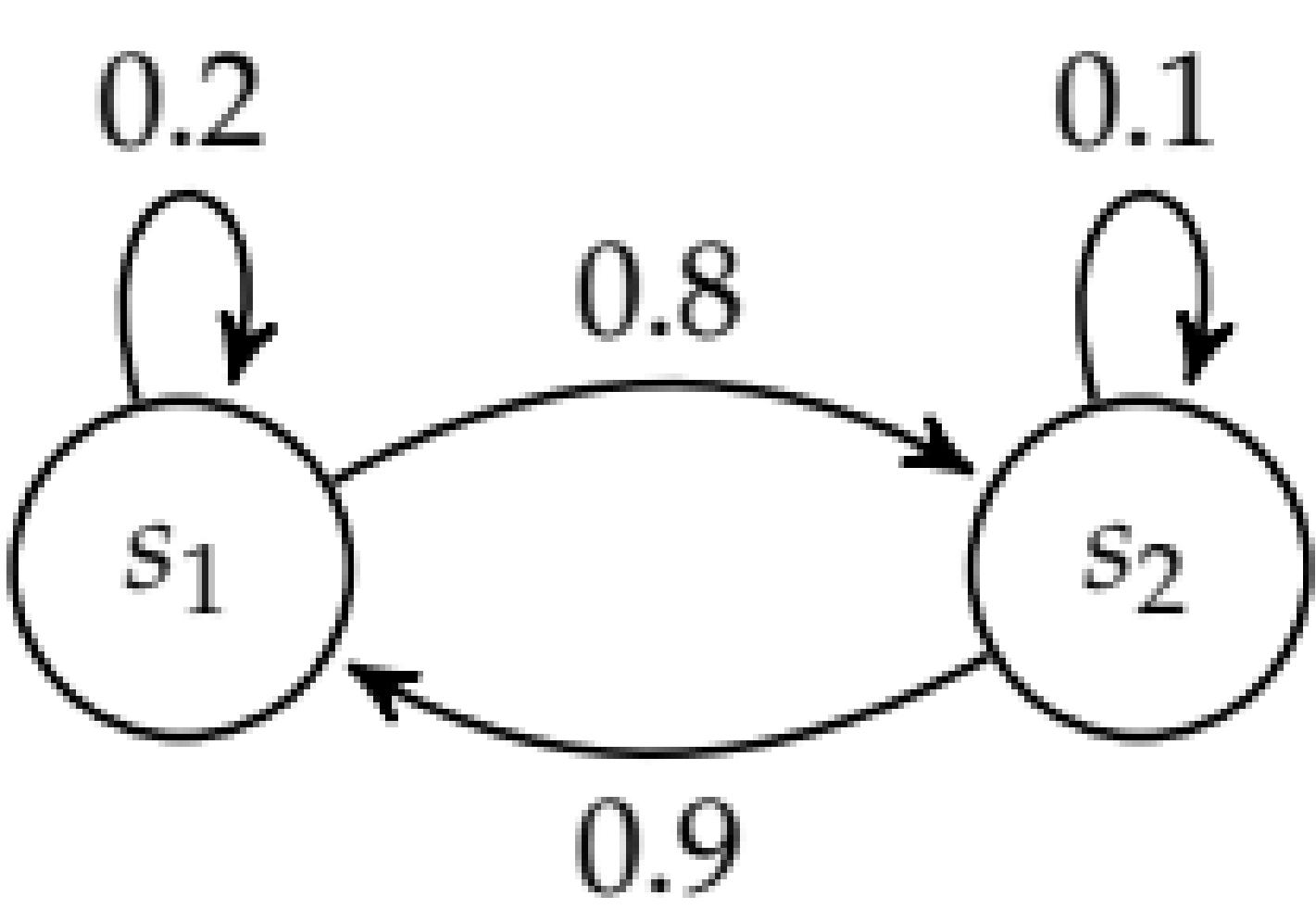
Second Order  
Taylor Inclusion

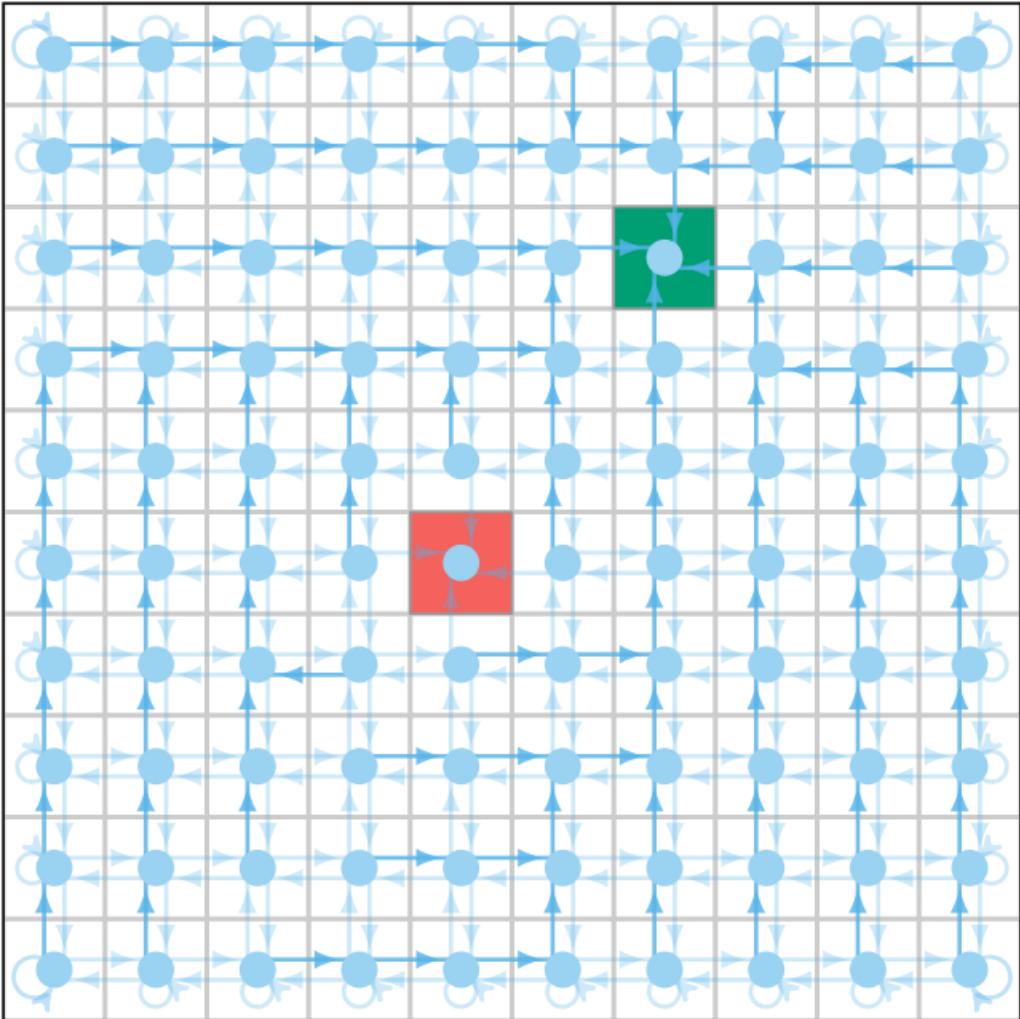


# Input Set

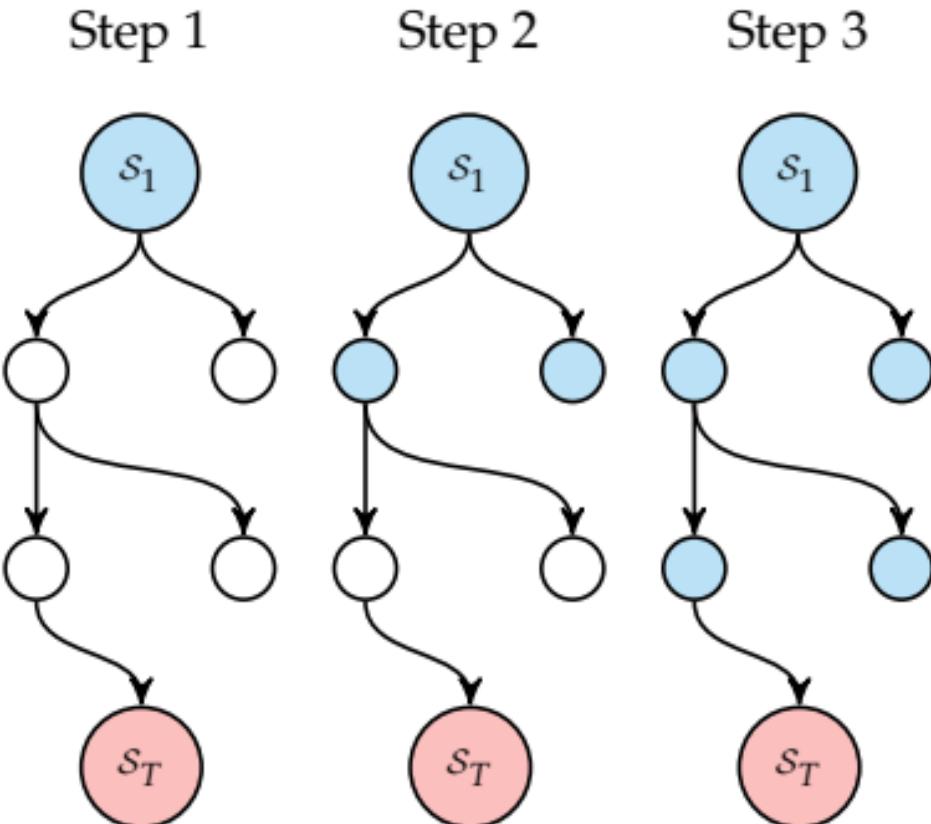
# Linear Transformation



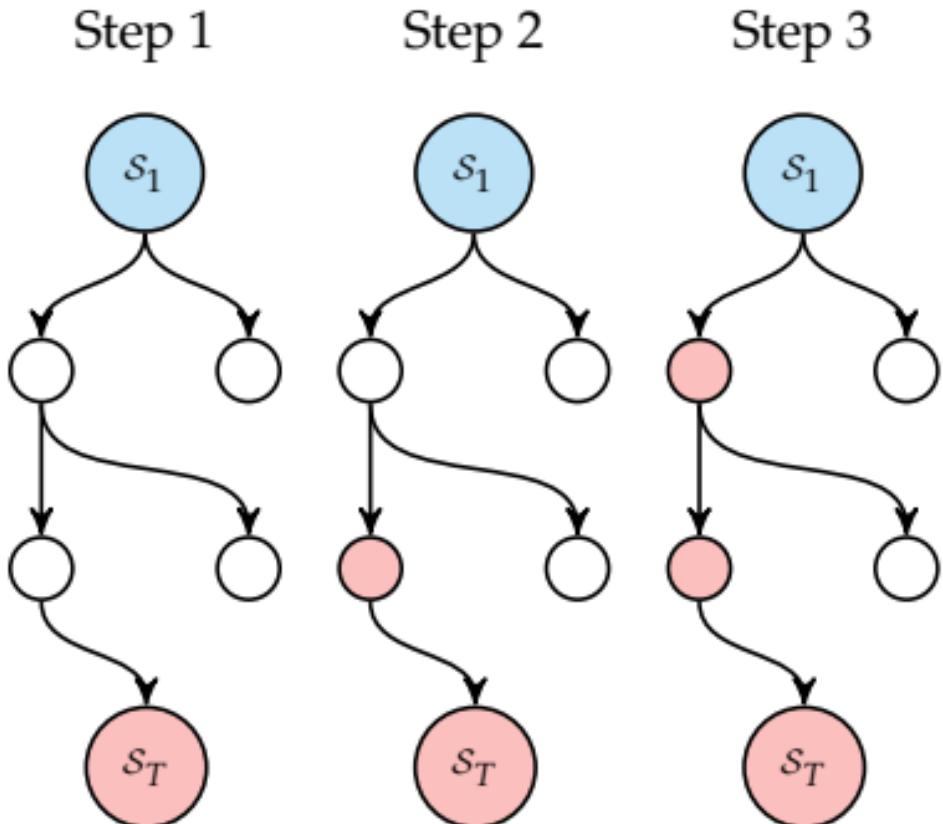


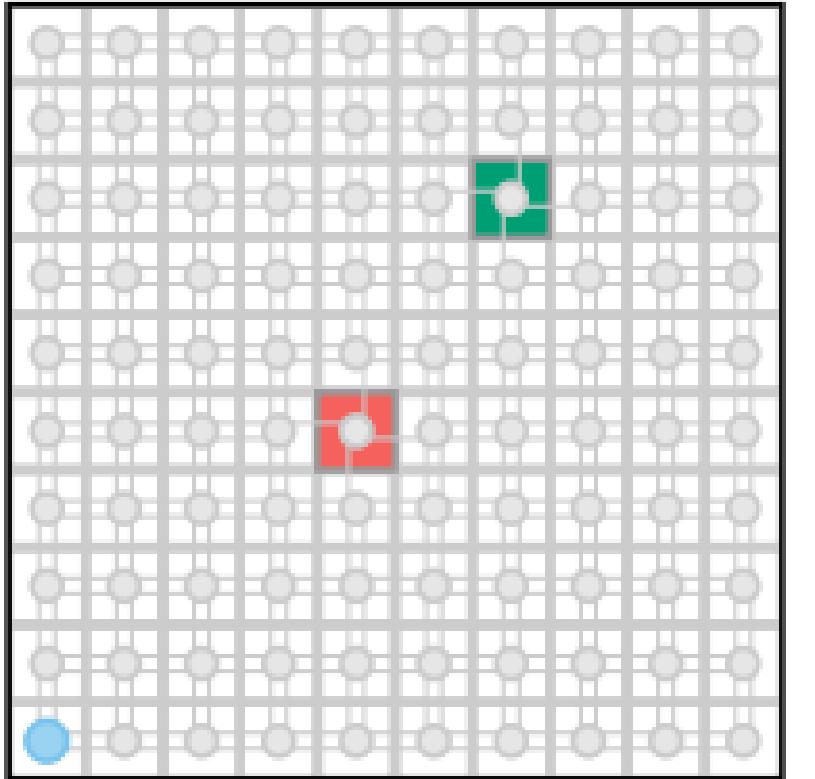
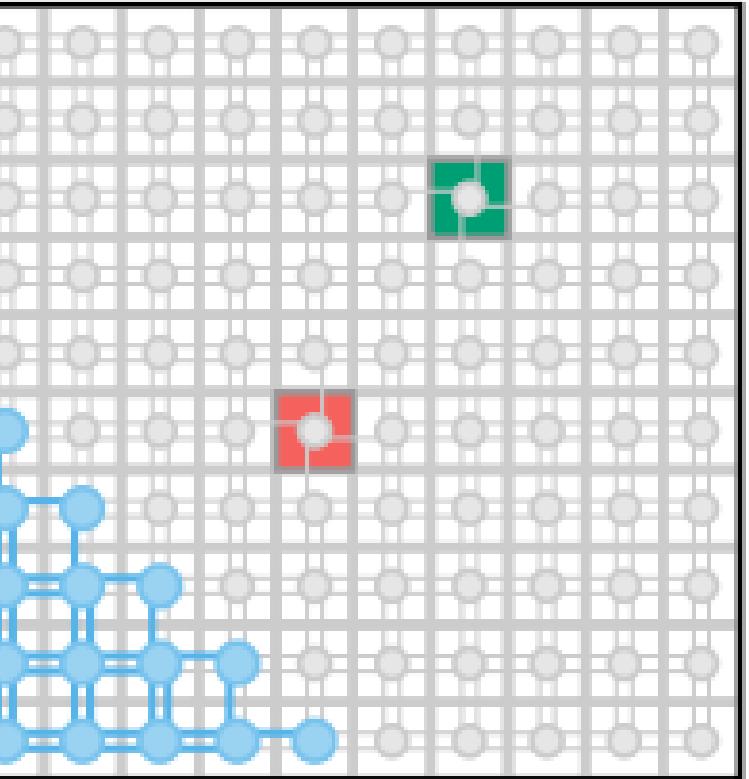
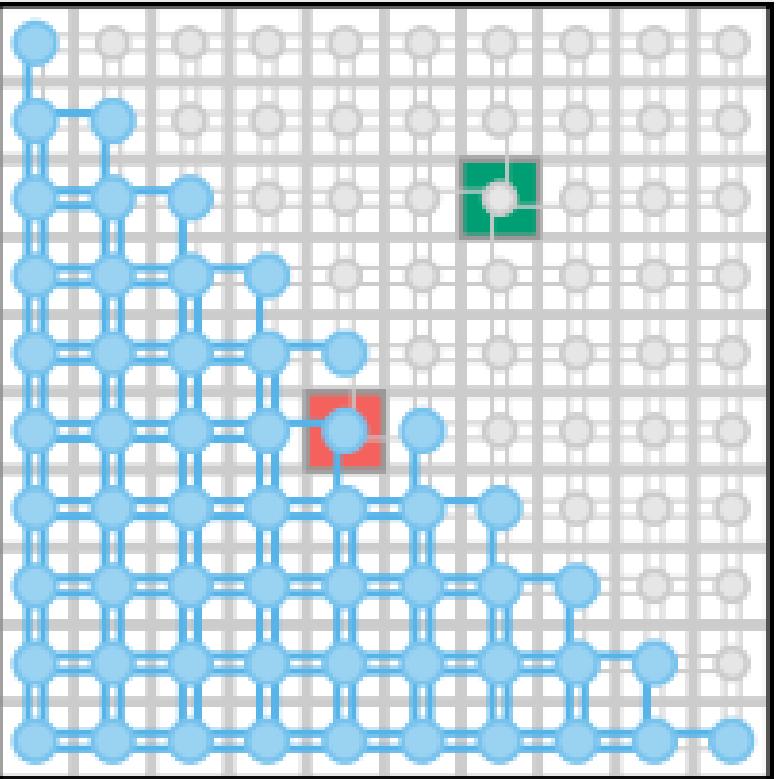


## Forward Reachability

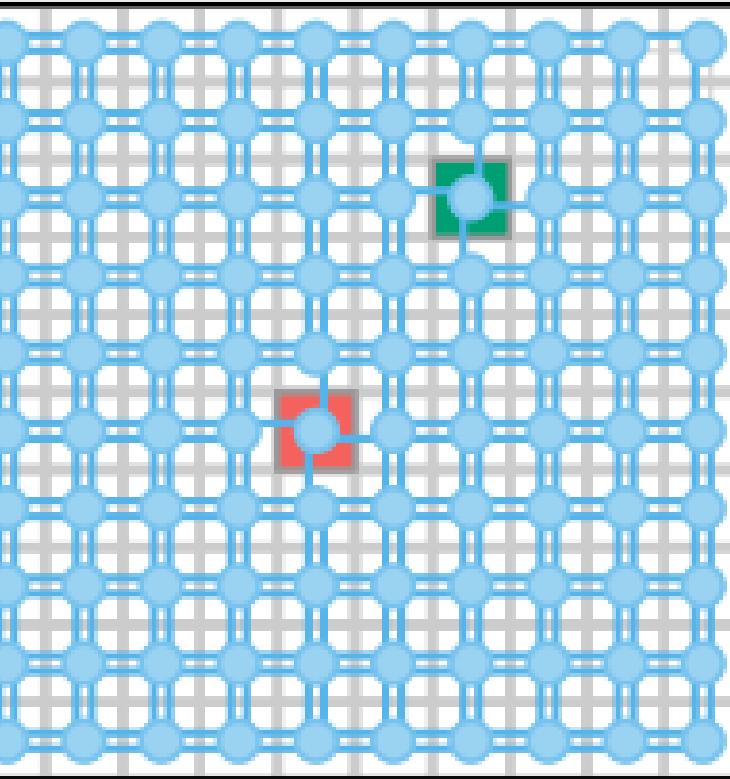


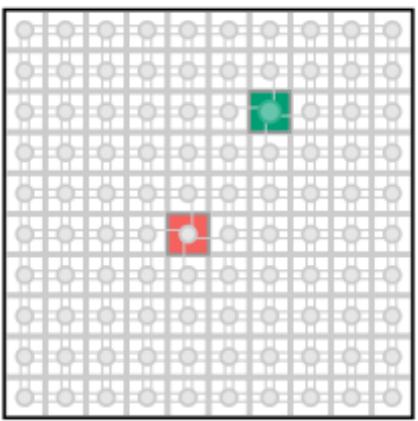
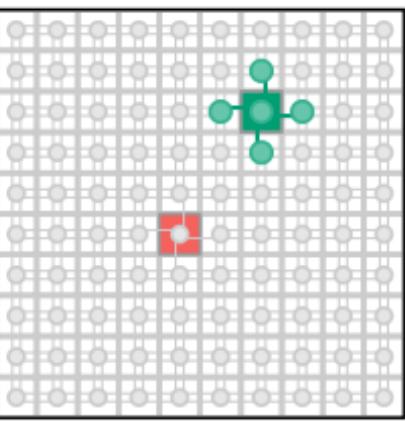
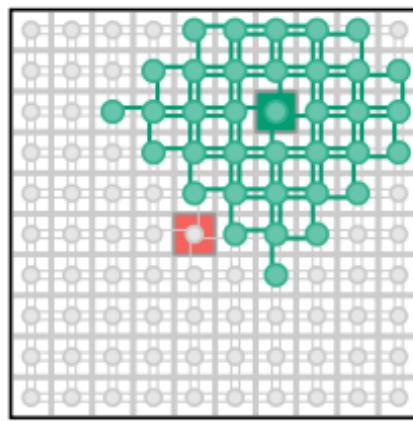
## Backward Reachability



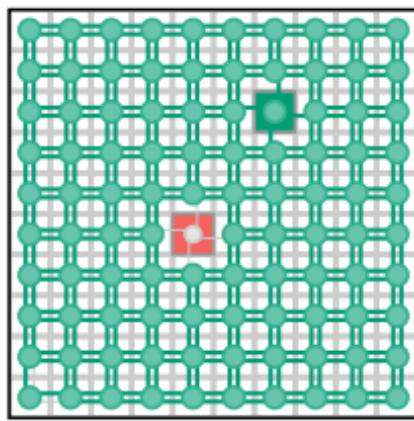
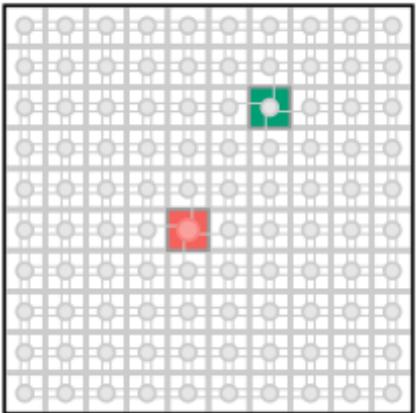
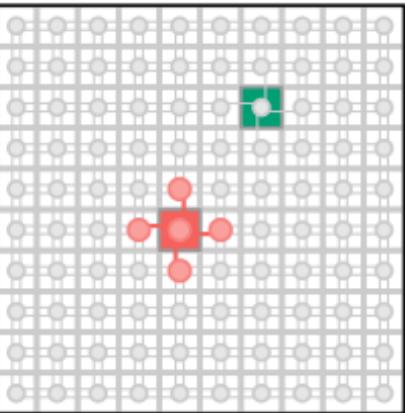
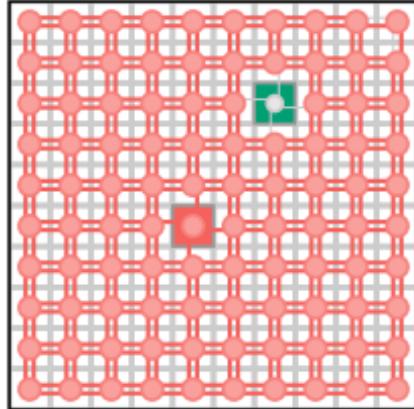
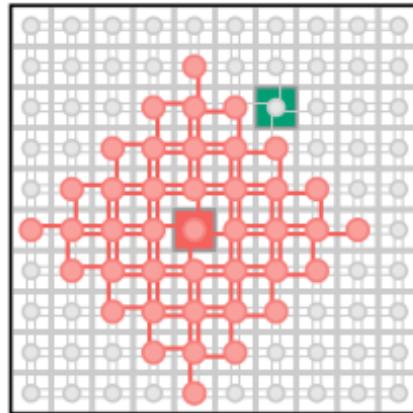
$\mathcal{R}_1$  $\mathcal{R}_5$  $\mathcal{R}_{10}$ 

Converged



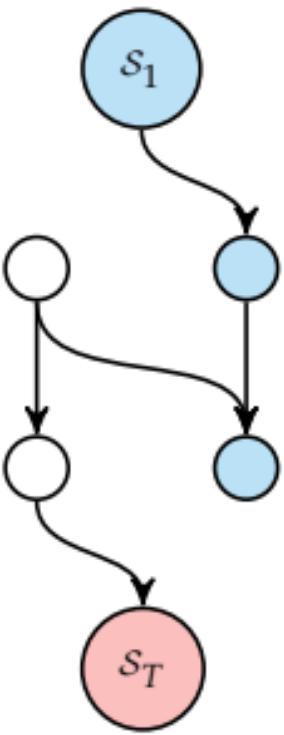
$\mathcal{B}_1$  $\mathcal{B}_2$  $\mathcal{B}_5$ 

Converged

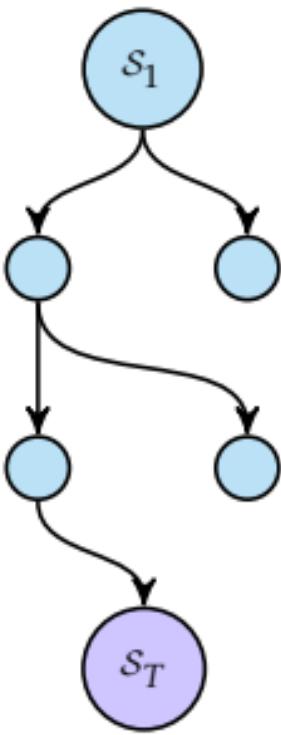
 $\mathcal{B}_1$  $\mathcal{B}_2$  $\mathcal{B}_5$ 

## Forward Reachability

Safe ✓

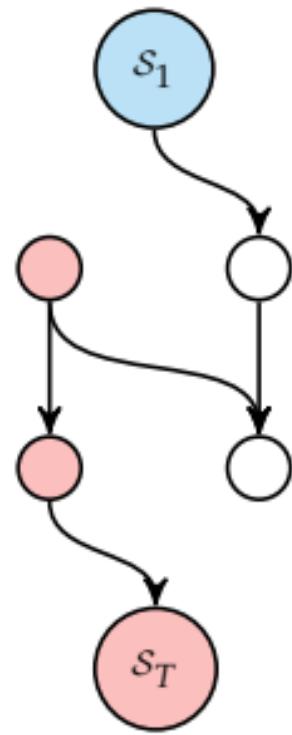


Unsafe ✗

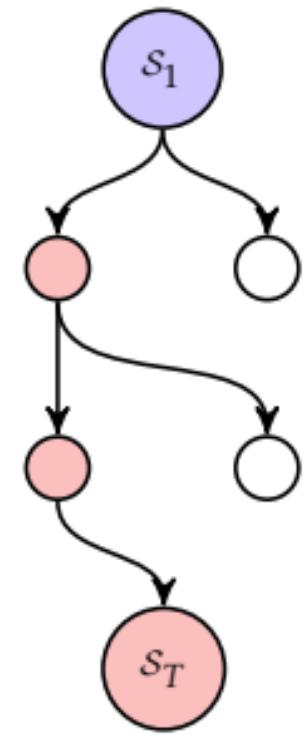


## Backward Reachability

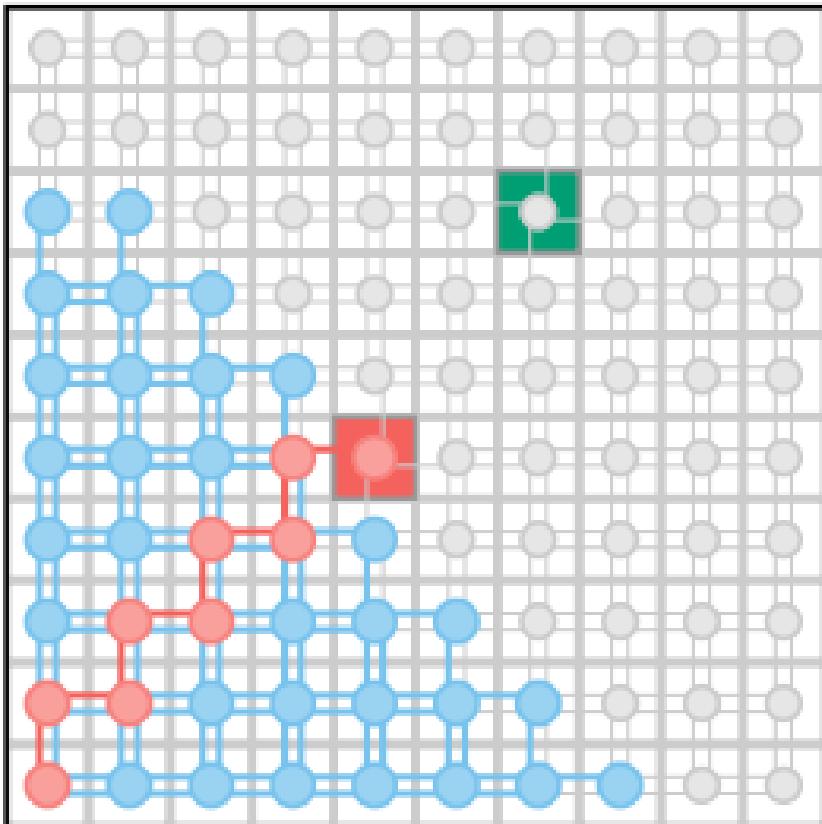
Safe ✓



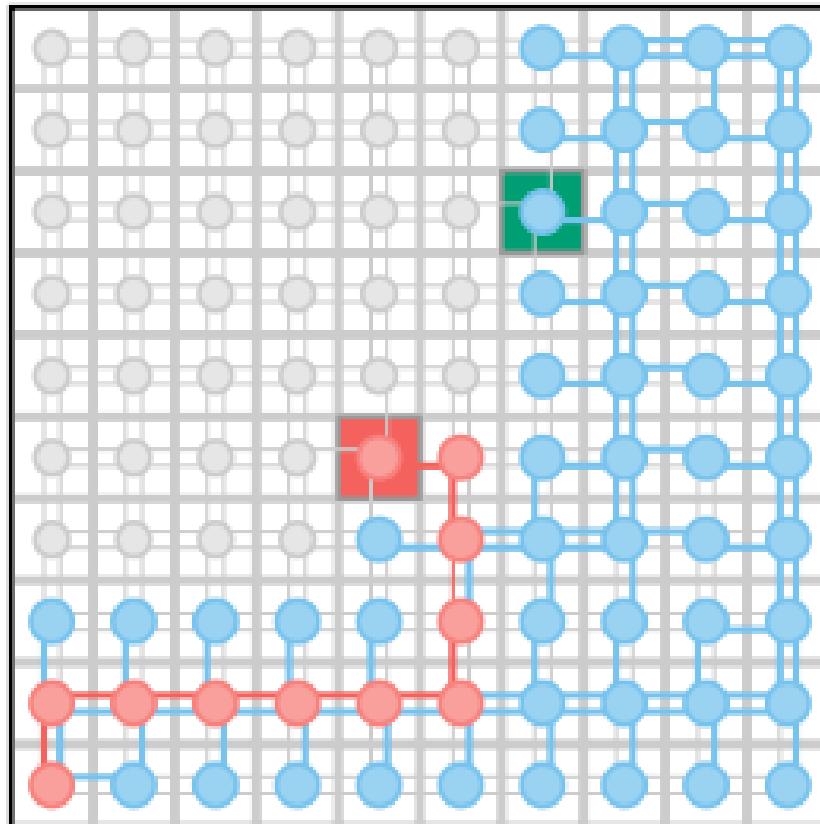
Unsafe ✗



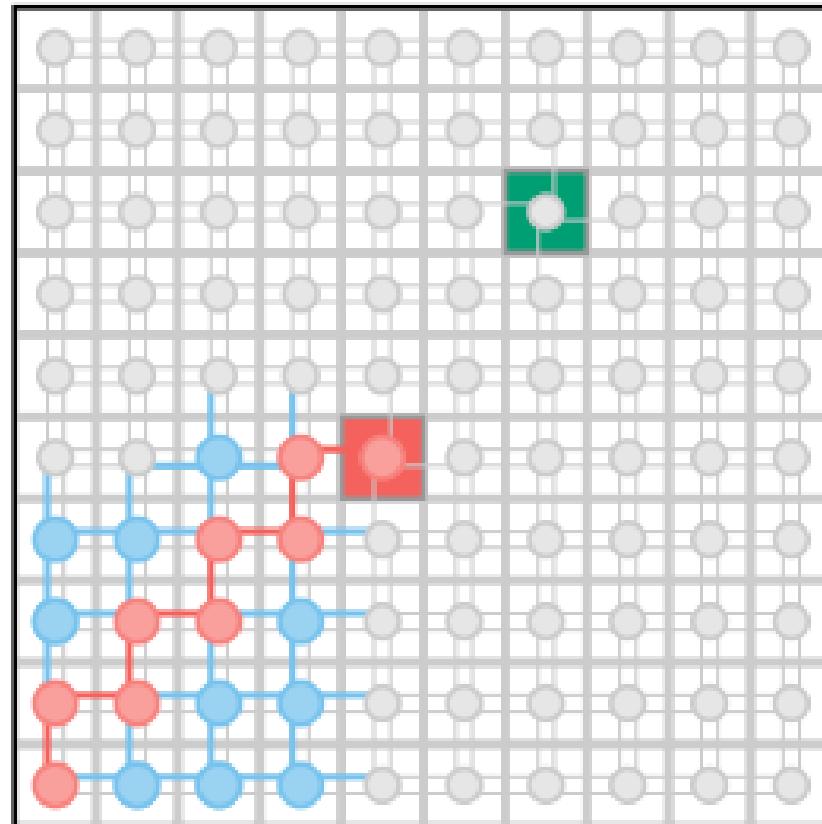
# Breadth First

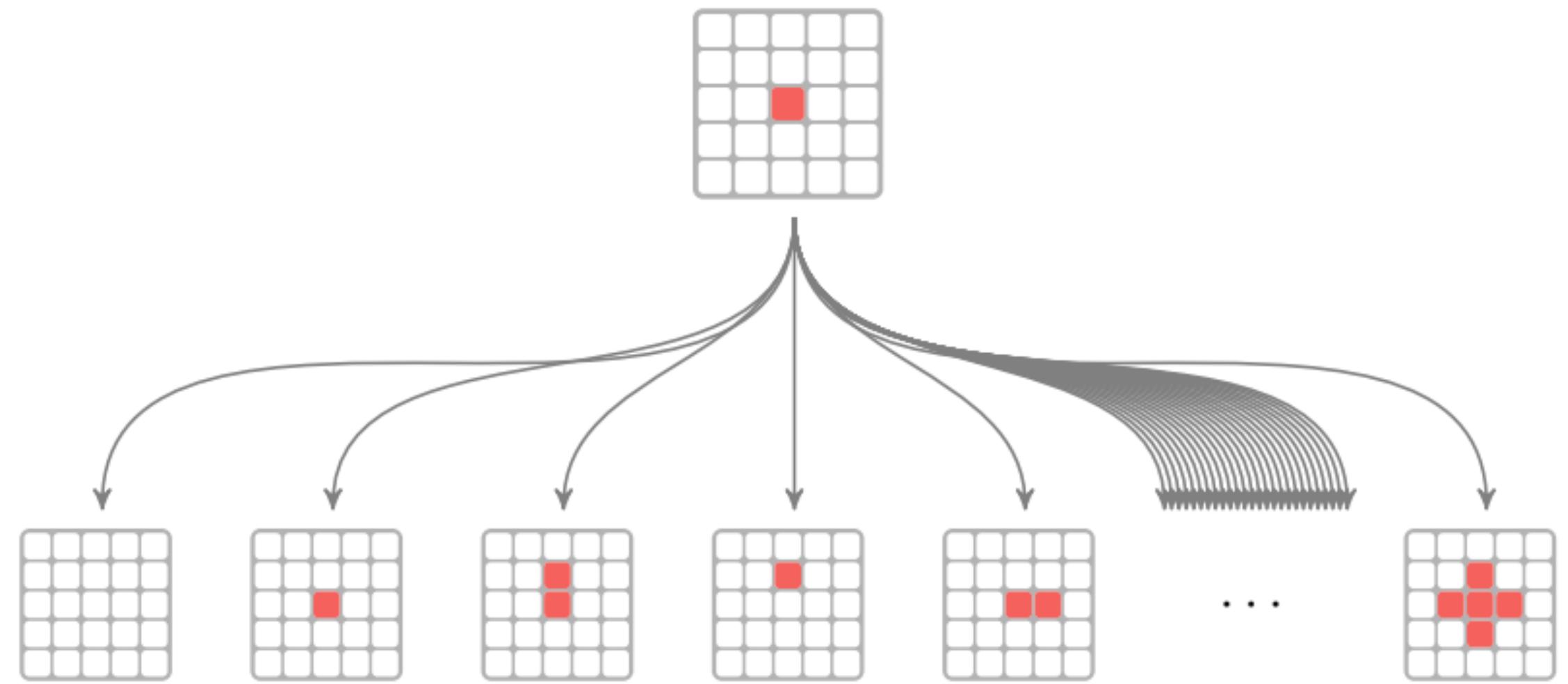


# Depth First



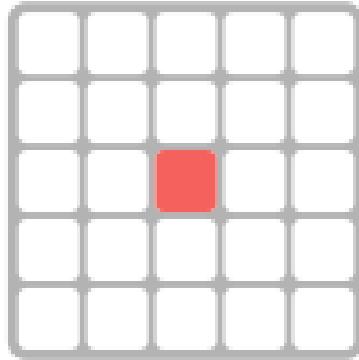
# Heuristic



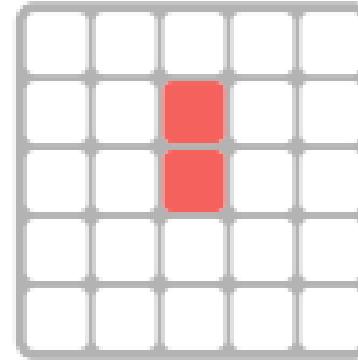


$T$

(



,

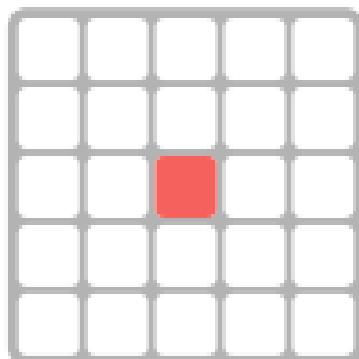


)

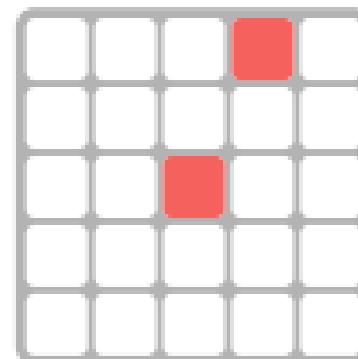
= true

$T$

(



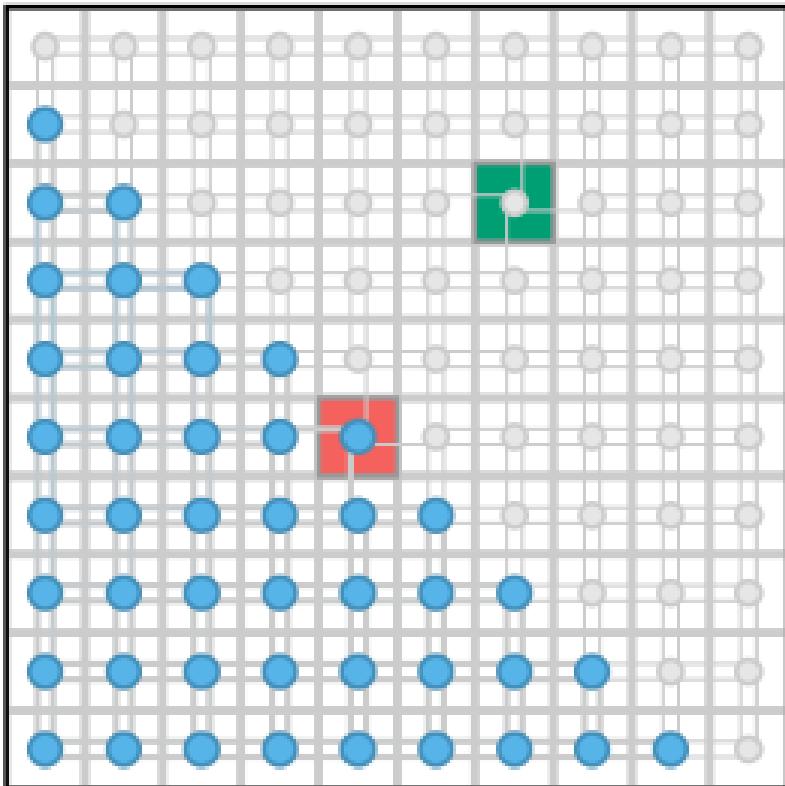
,



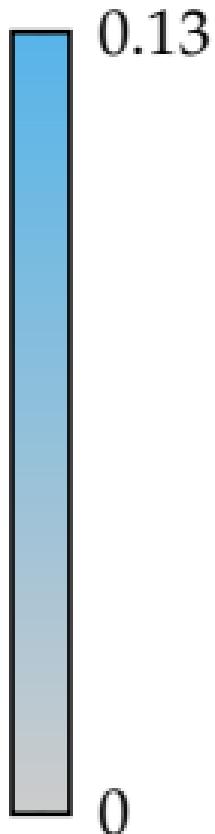
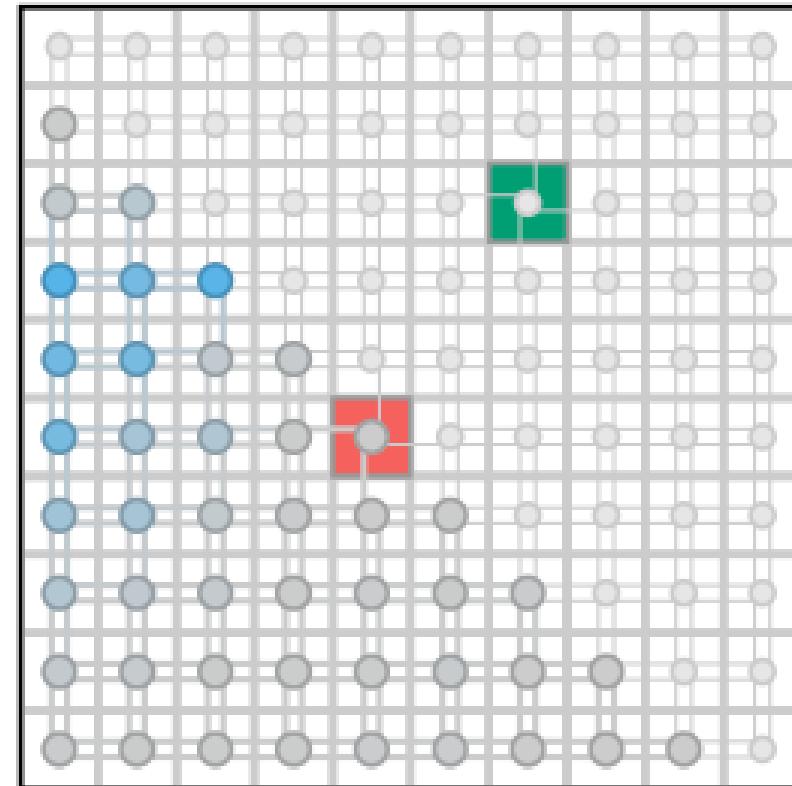
)

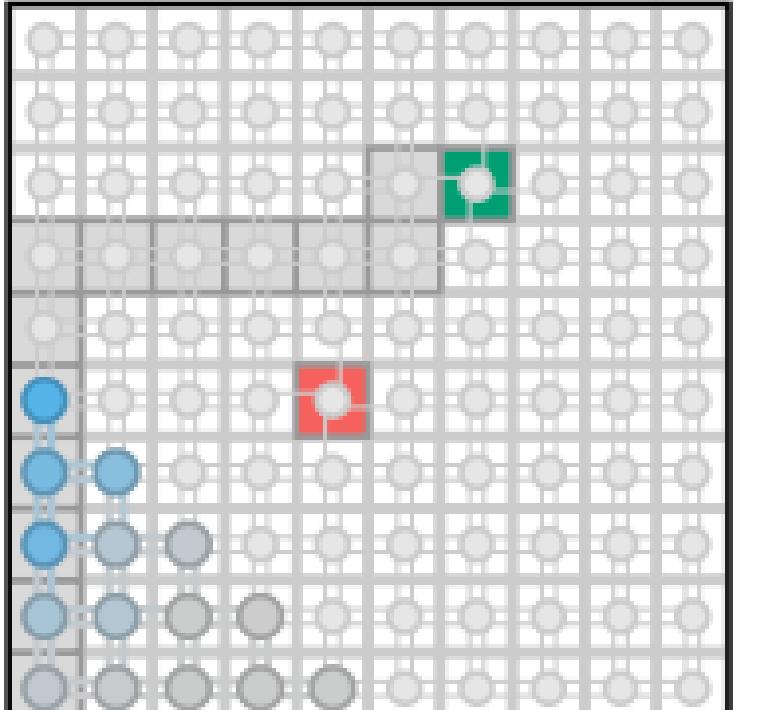
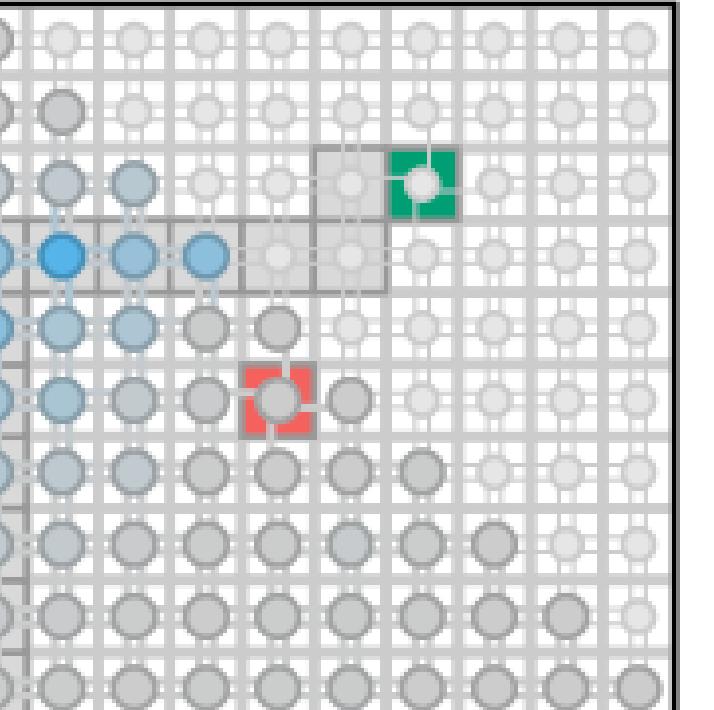
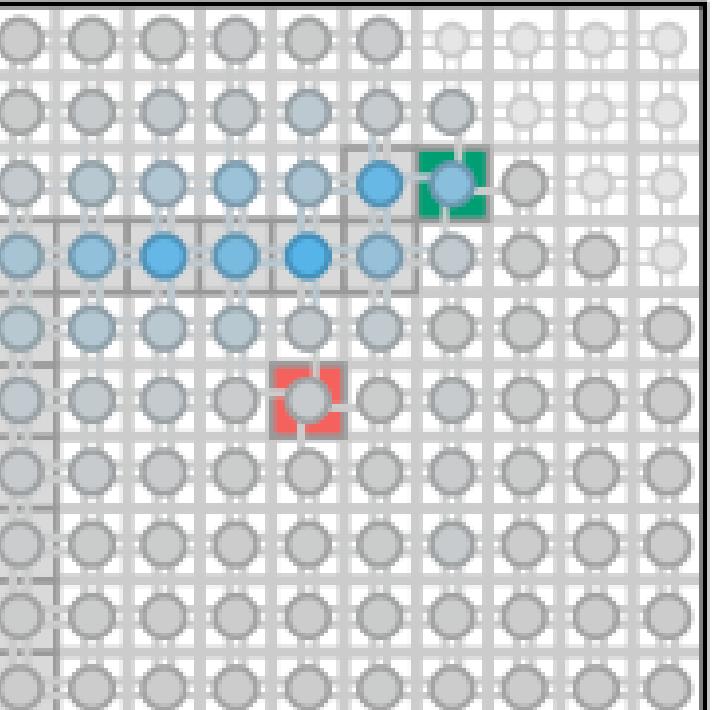
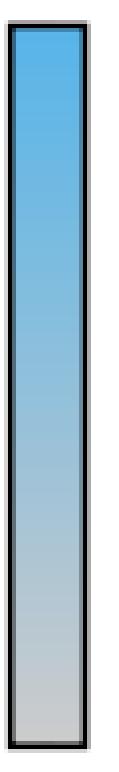
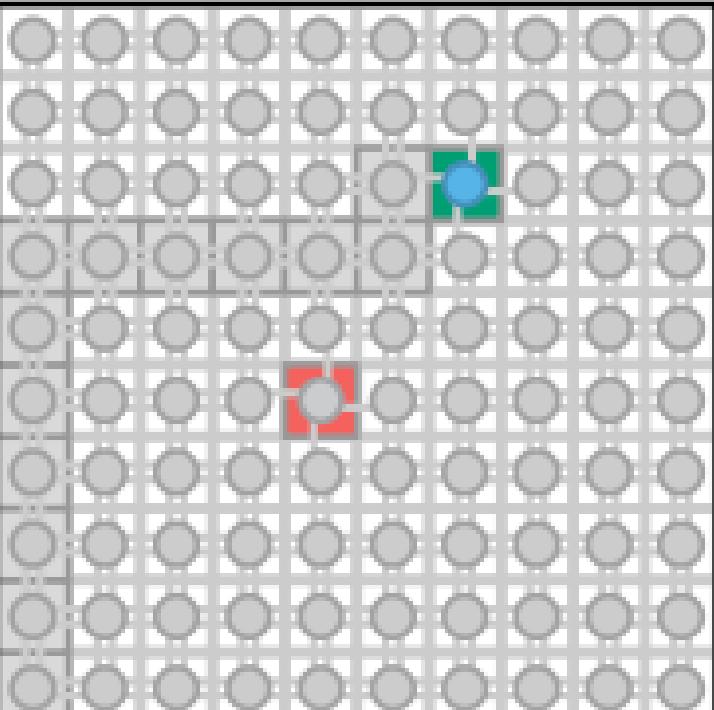
= false

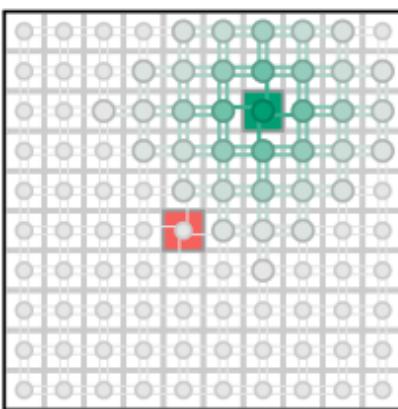
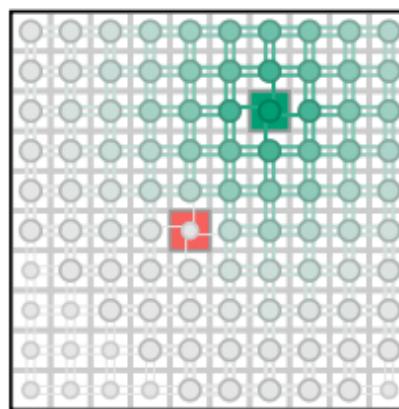
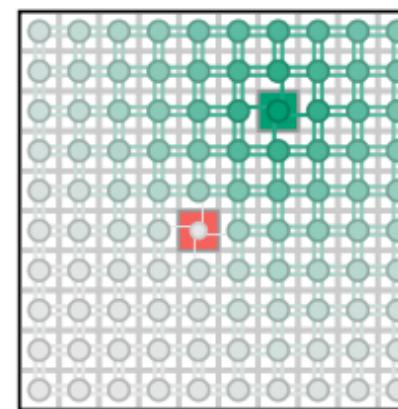
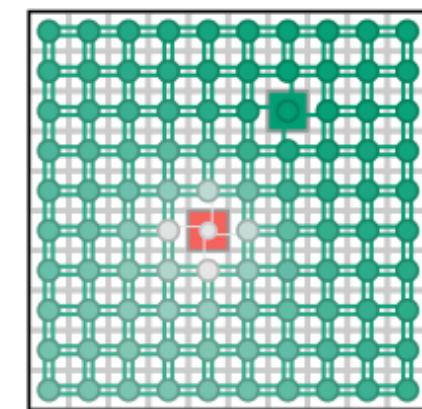
# Reachable Set



# Probabilistic Reachability



$P_5$  $P_{10}$  $P_{15}$  $P_{50}$ 

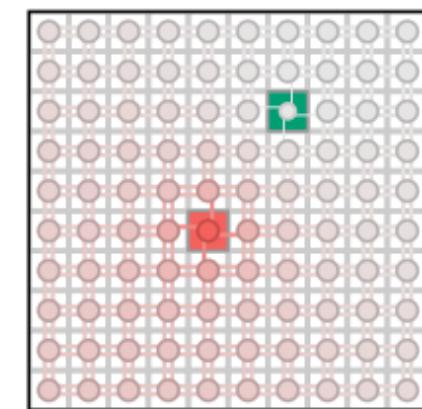
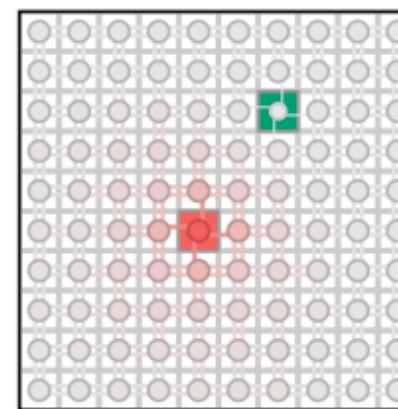
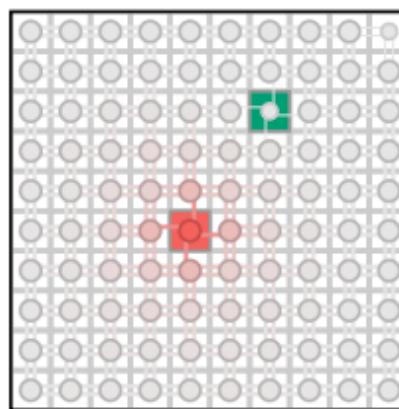
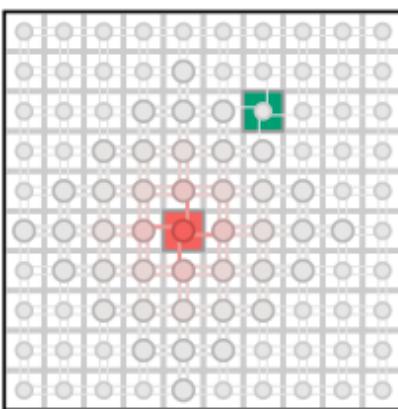
$R_5$  $R_{10}$  $R_{15}$  $R_{200}$ 

1

0

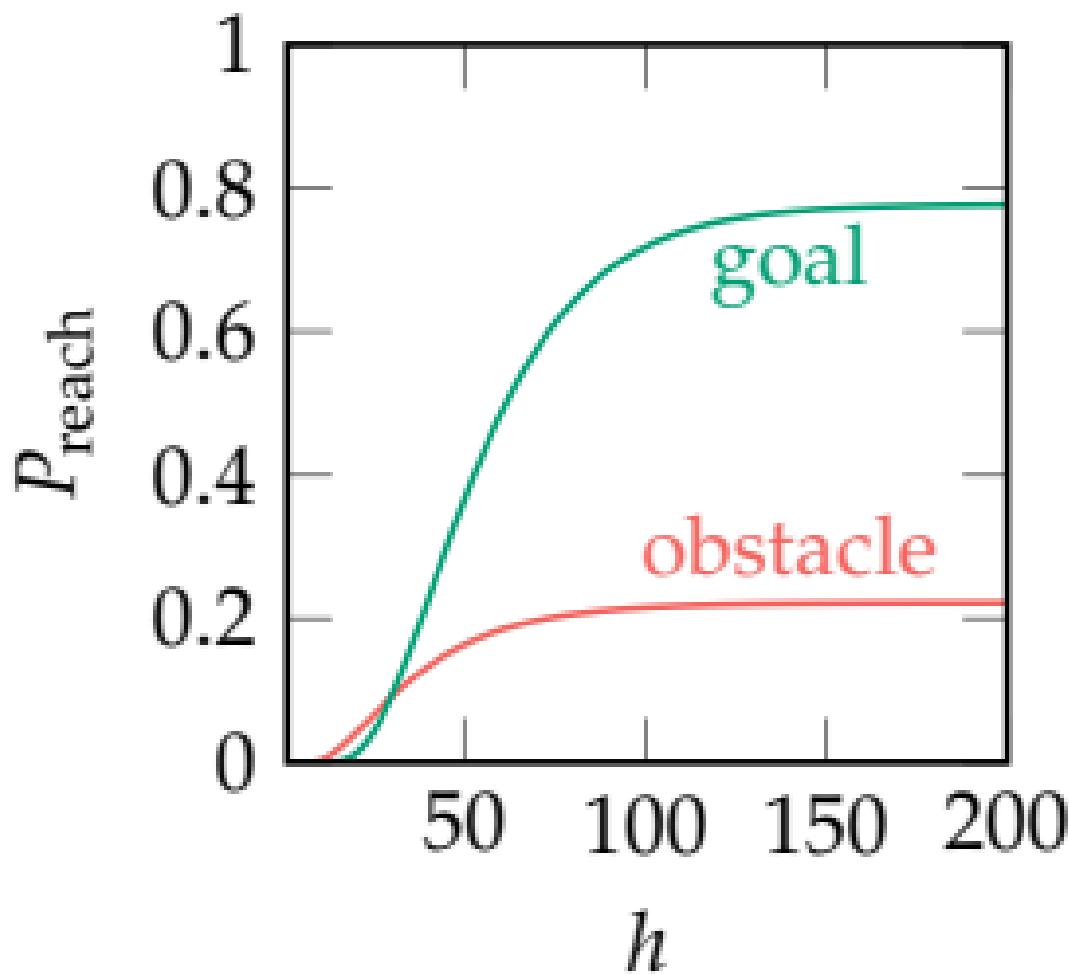
1

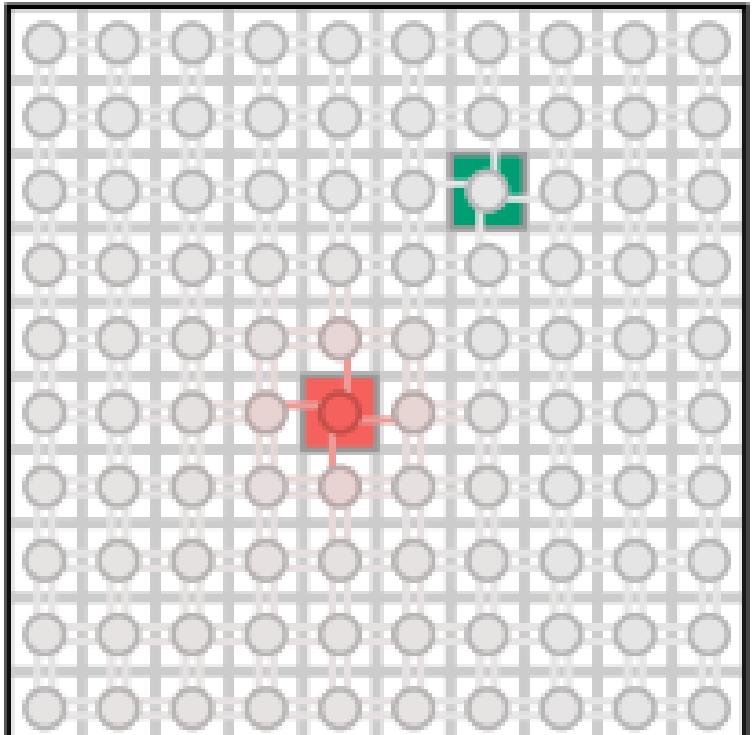
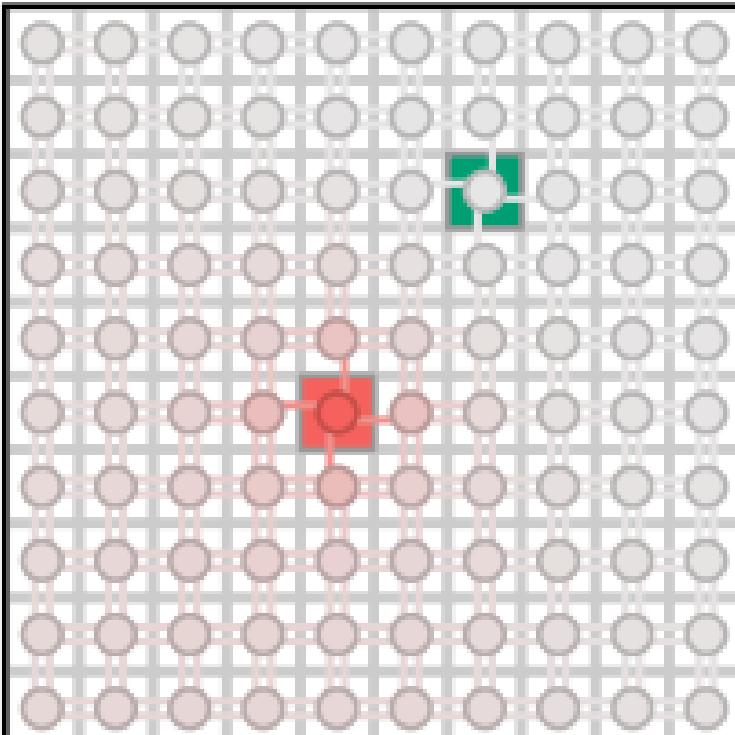
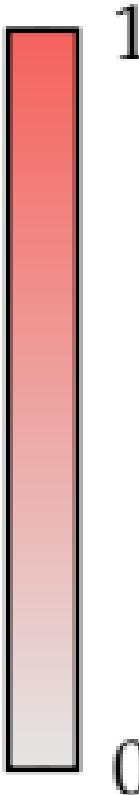
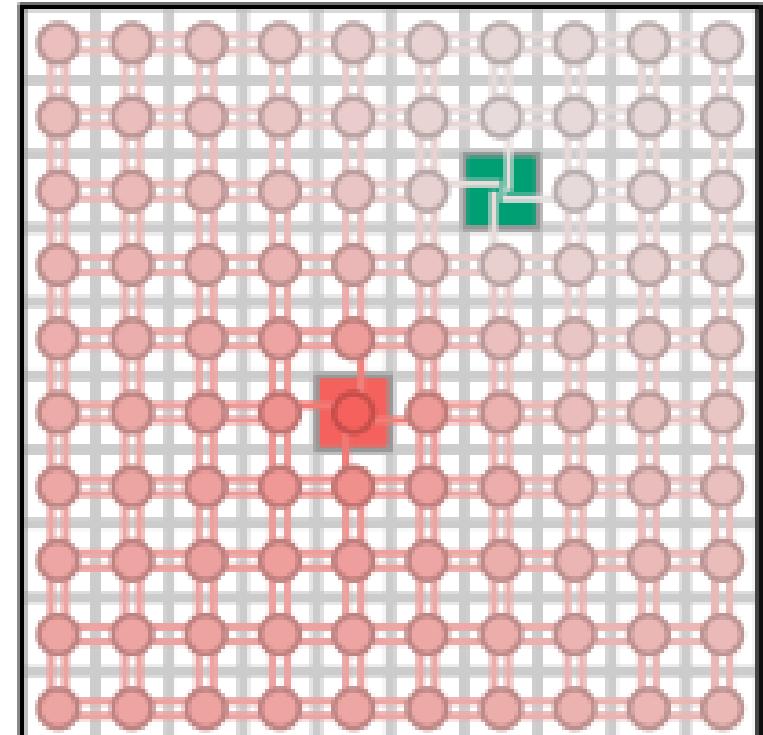
0



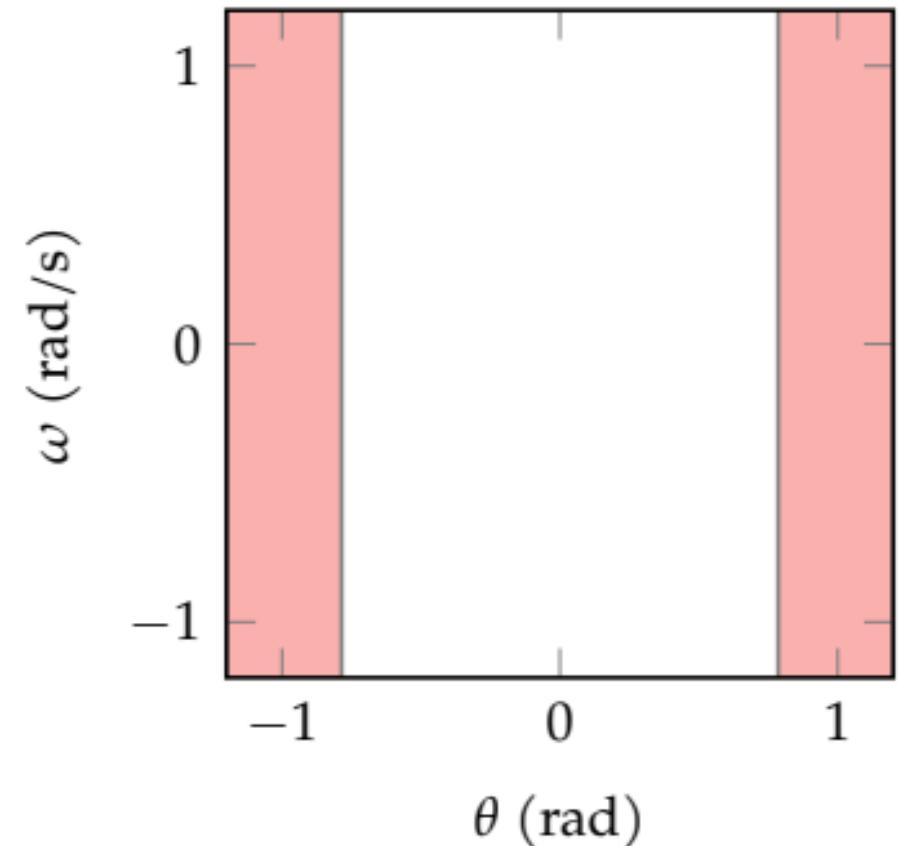
1

0

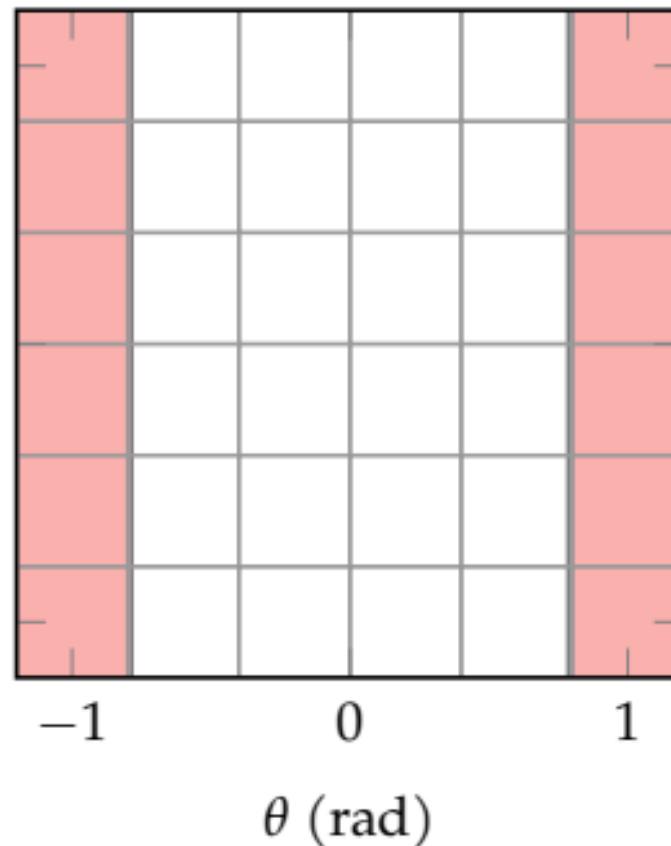


$P_{\text{fail}} = 0.018$  $P_{\text{fail}} = 0.102$  $P_{\text{fail}} = 0.49$ 

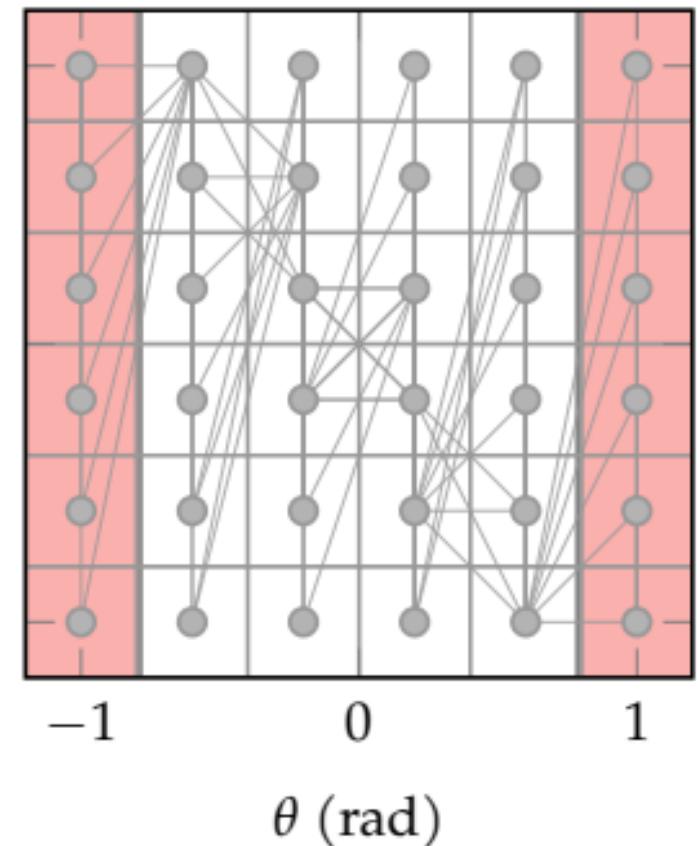
# Continuous State Space

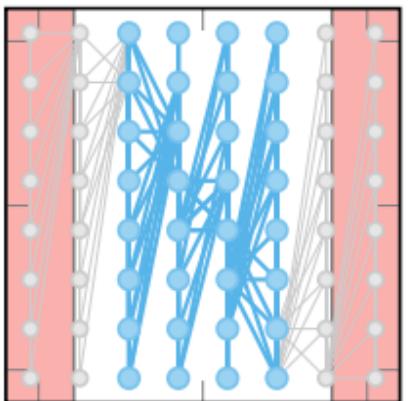
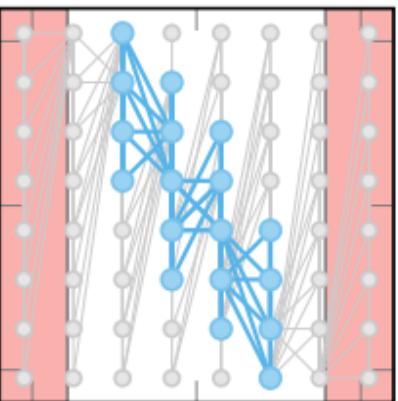
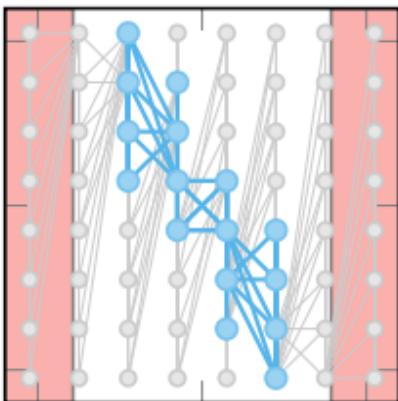


# Partition

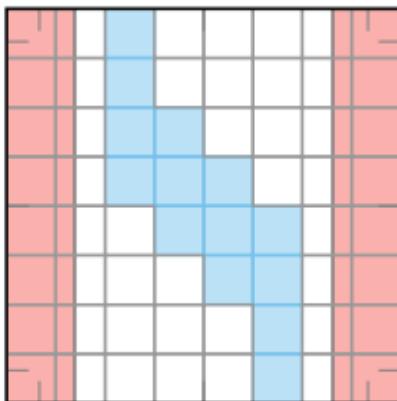
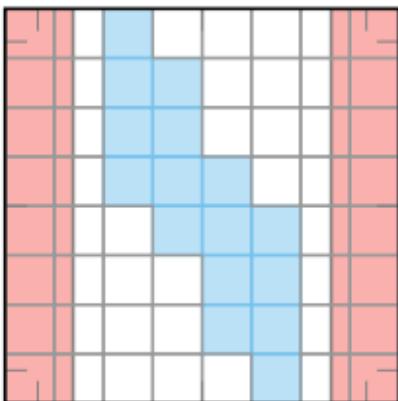
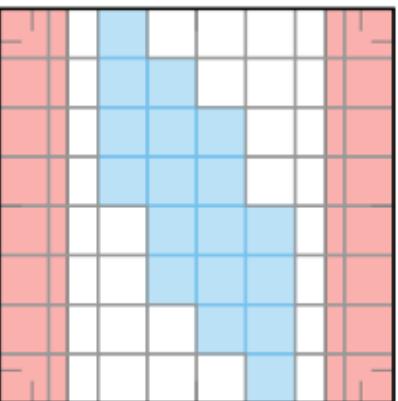
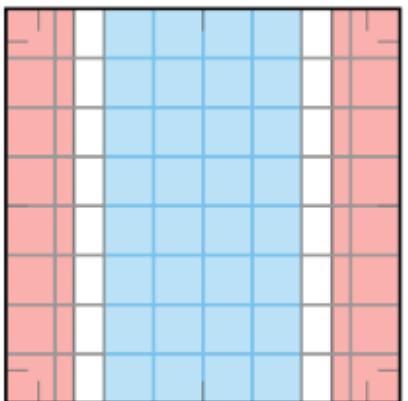
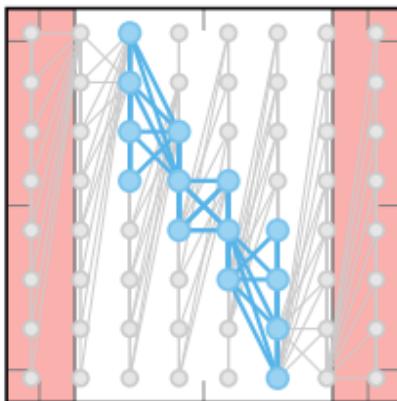


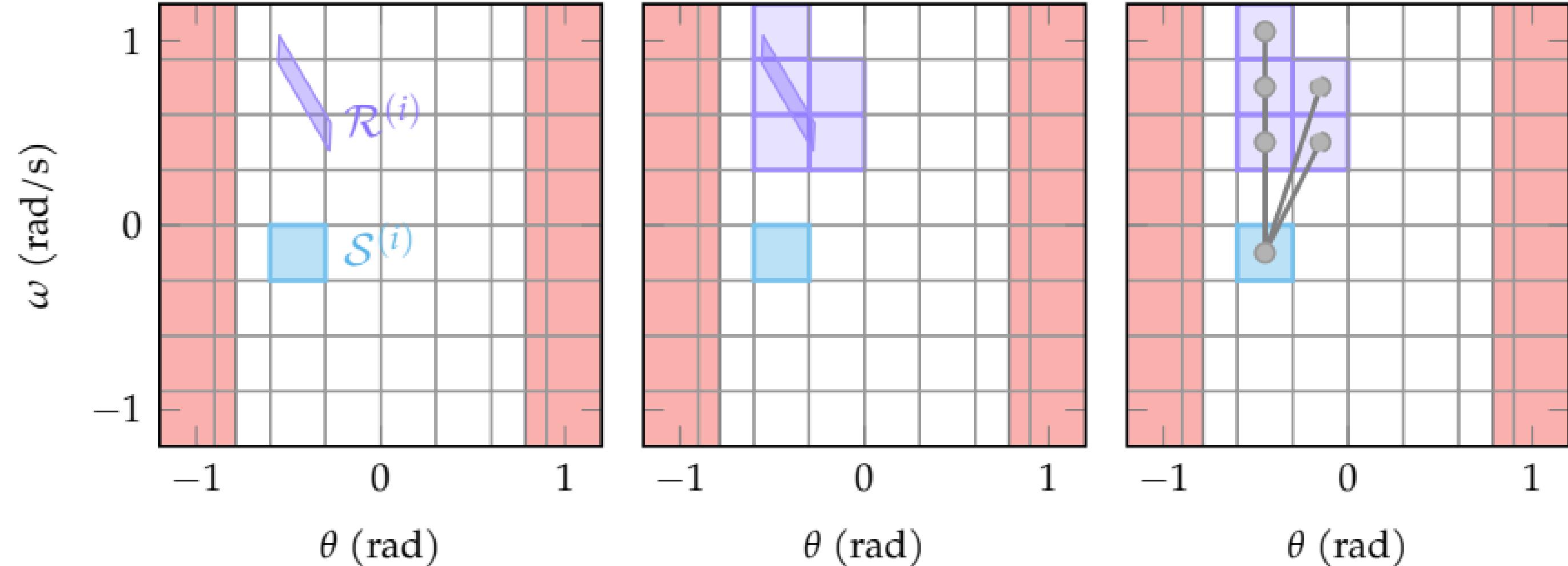
# DSA



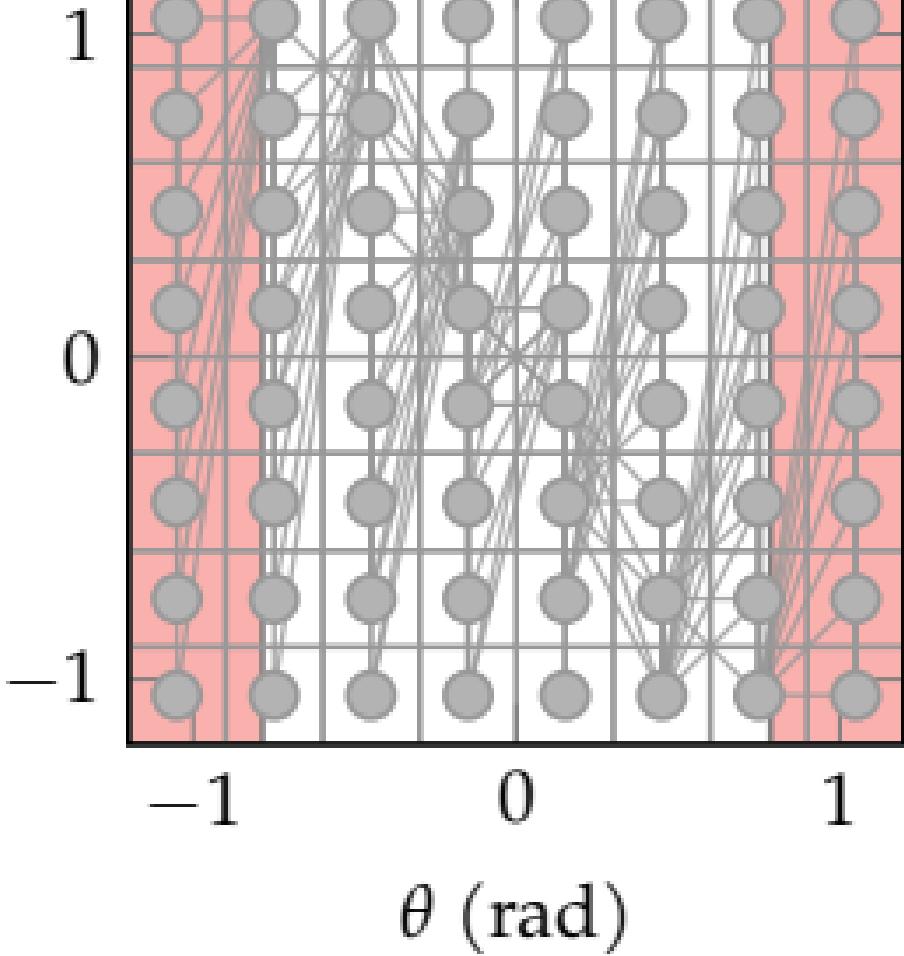
$\mathcal{R}_1$  $\mathcal{R}_2$  $\mathcal{R}_3$ 

Converged

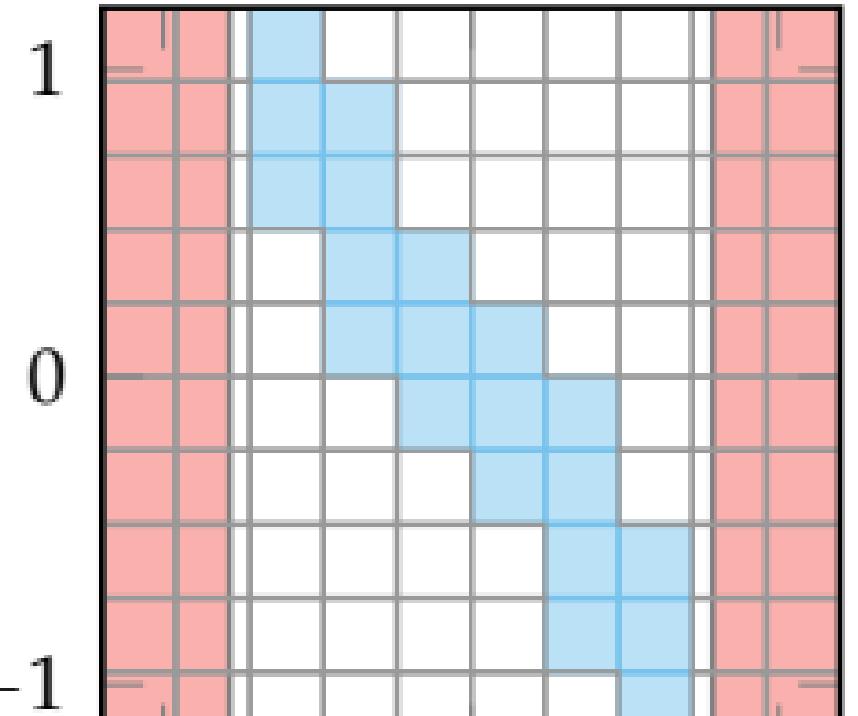




$\omega$  (rad/s)

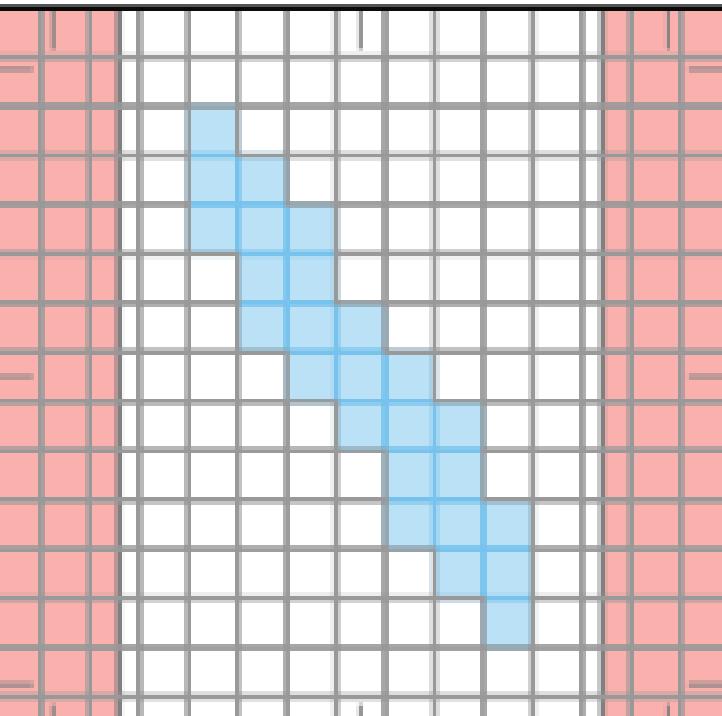


$\omega$  (rad/s)



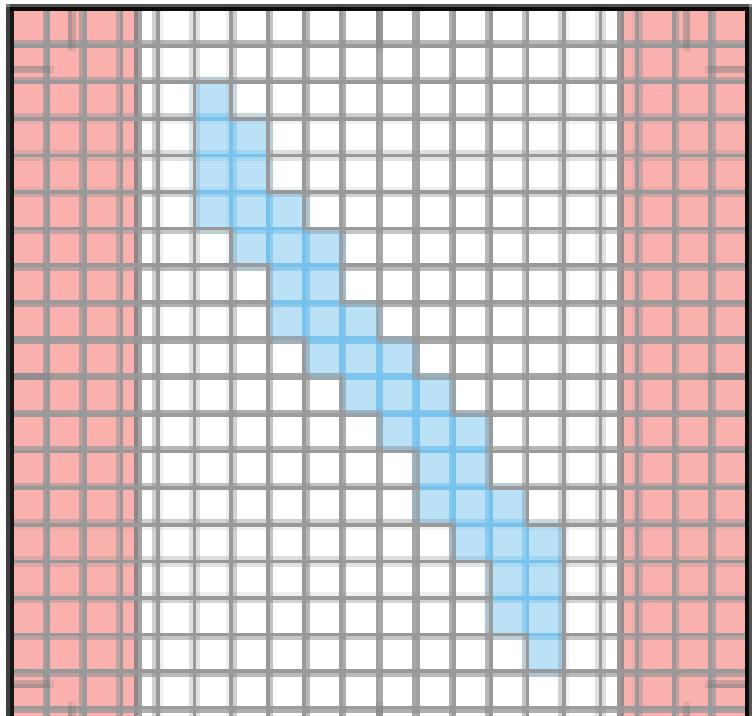
-1 0 1

$\theta$  (rad)



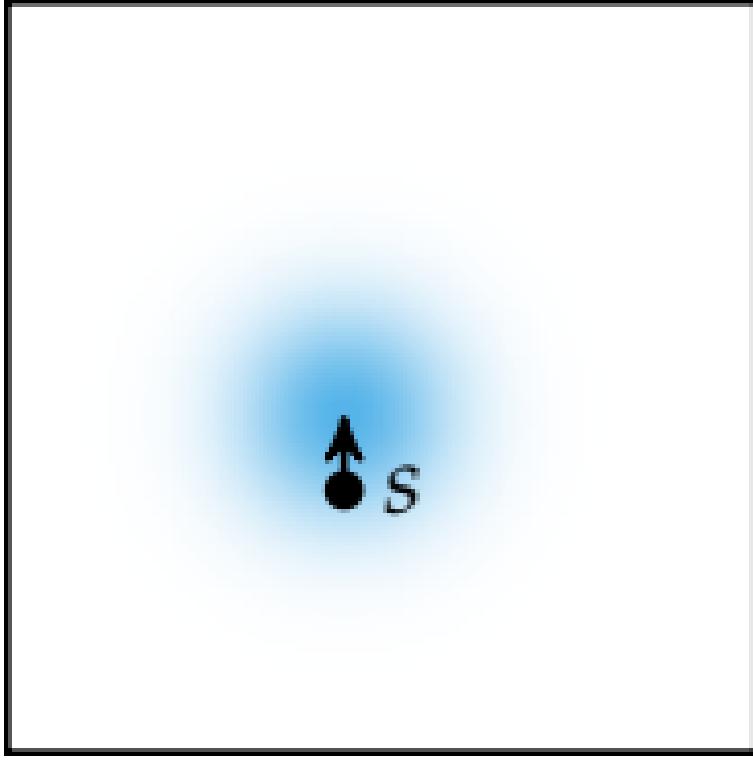
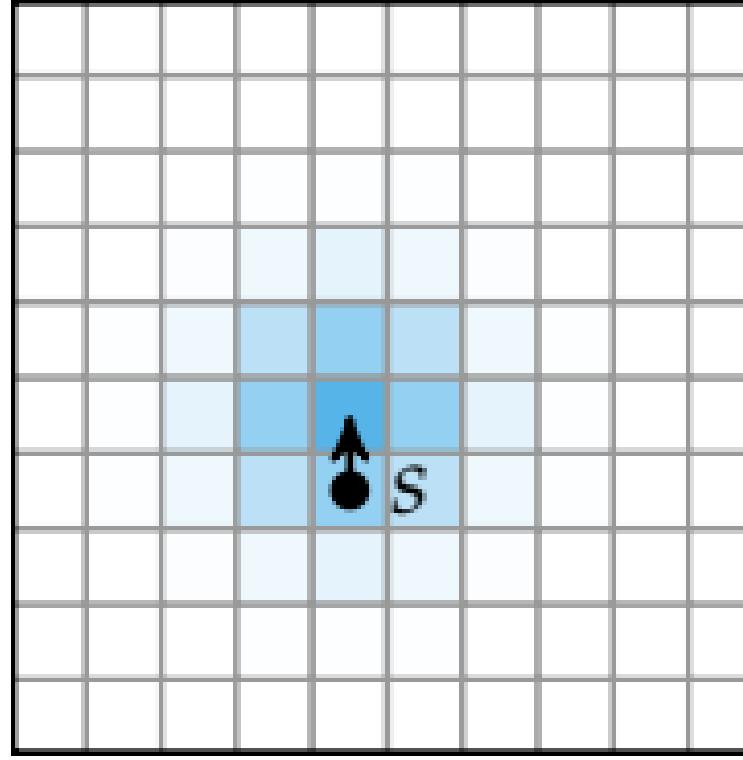
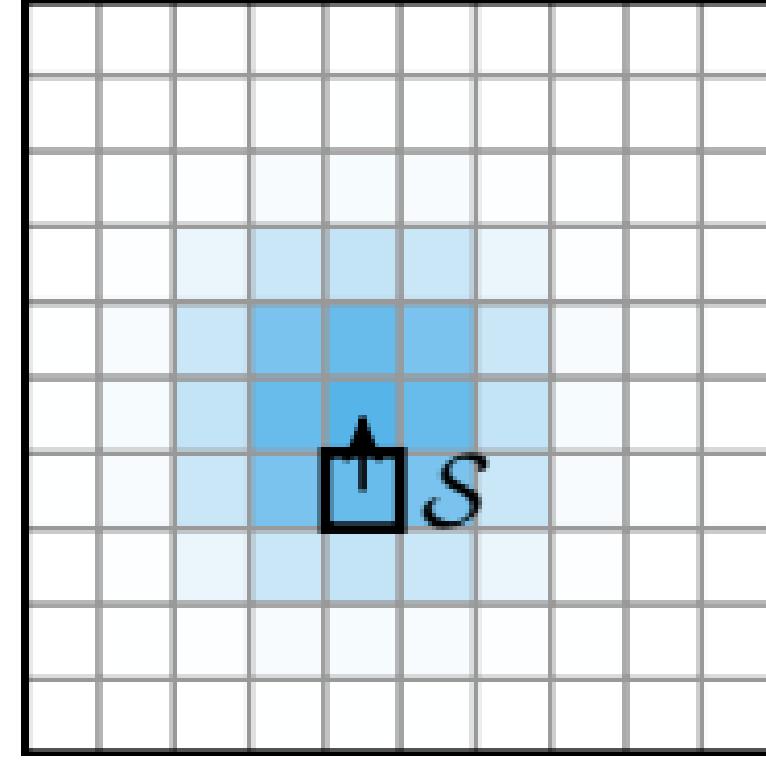
-1 0 1

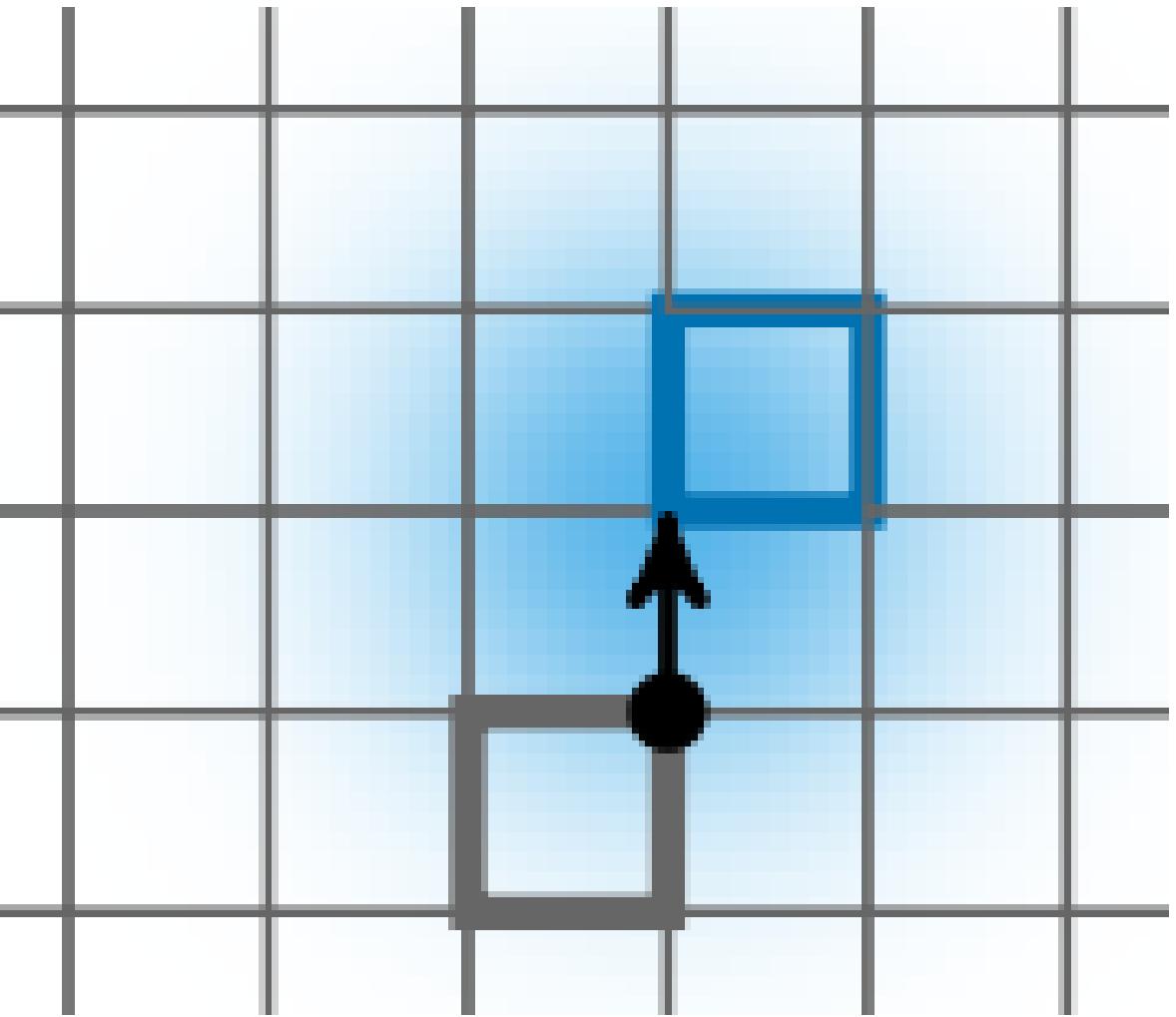
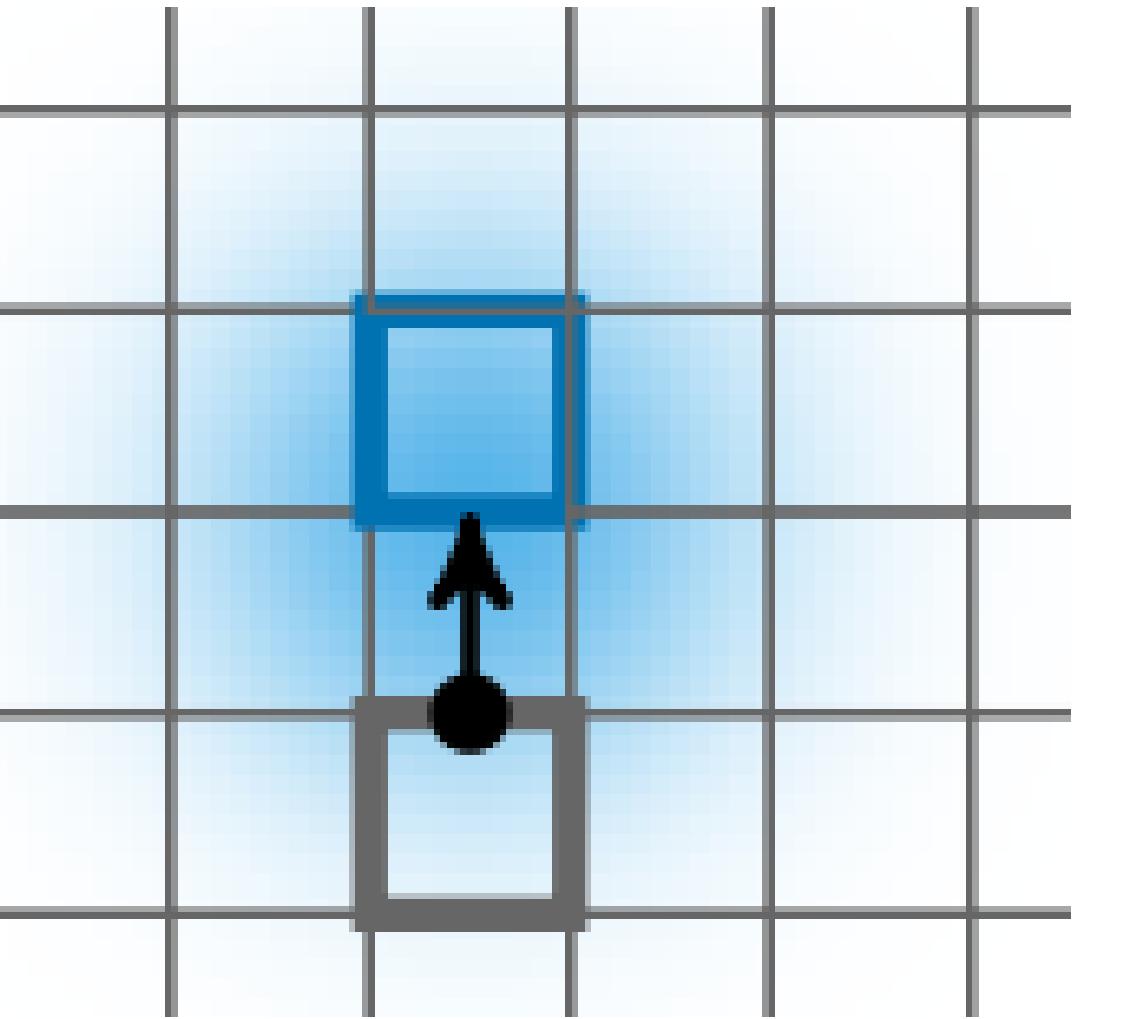
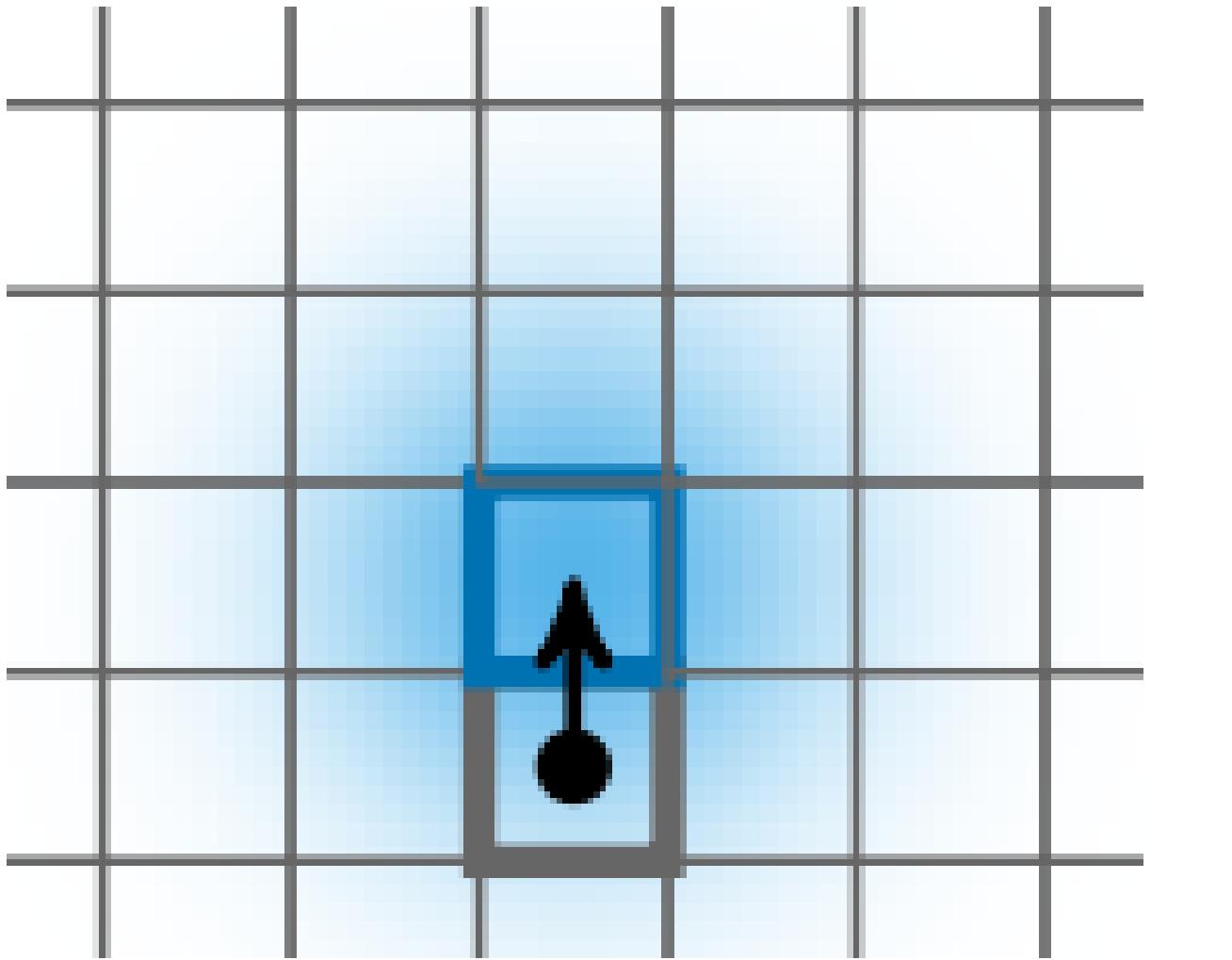
$\theta$  (rad)

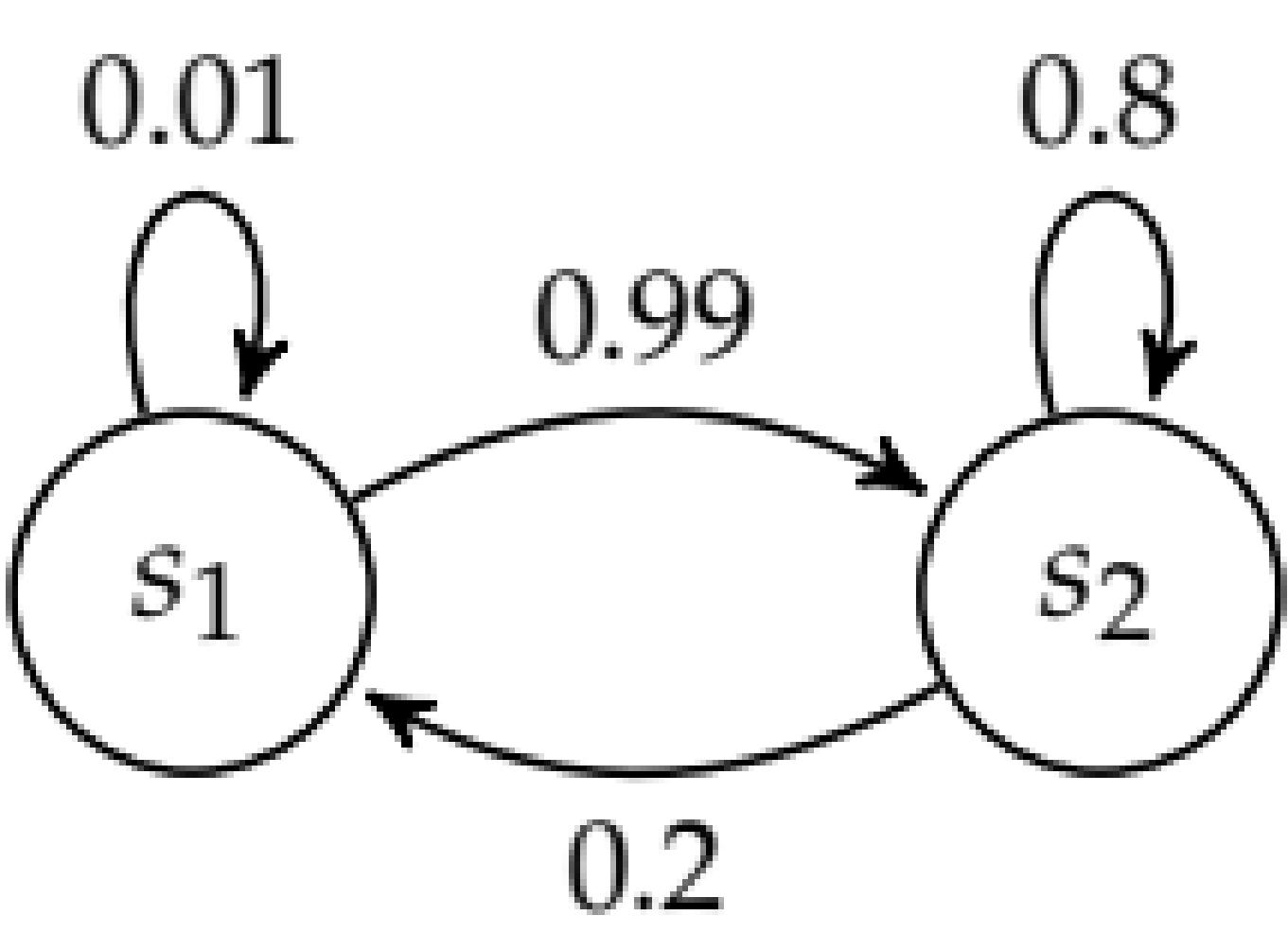


-1 0 1

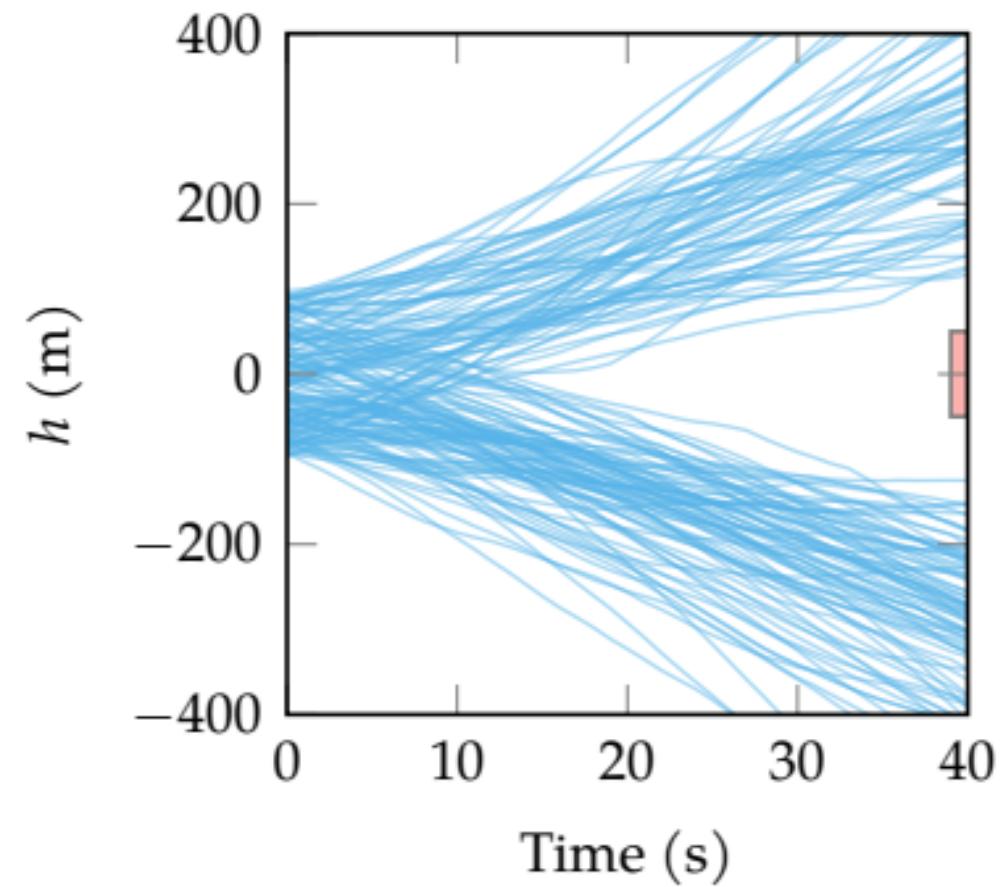
$\theta$  (rad)

$T(\mathbf{s}, \mathbf{s}')$  $T(\mathbf{s}, \mathcal{S}')$  $T(\mathcal{S}, \mathcal{S}')$ 

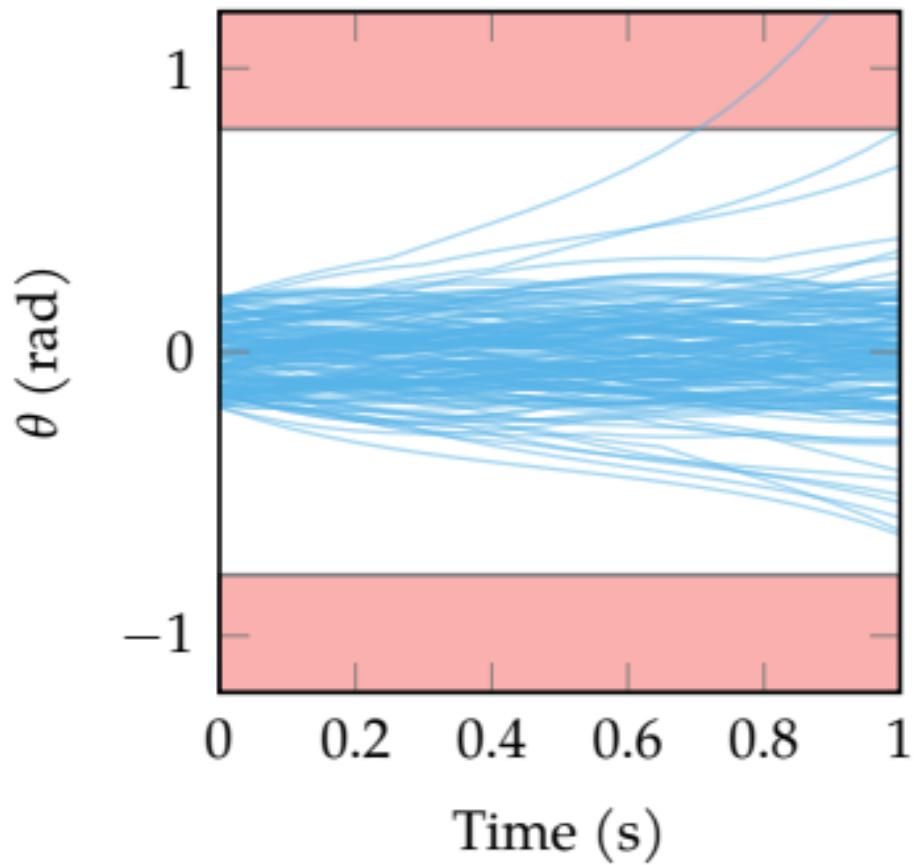


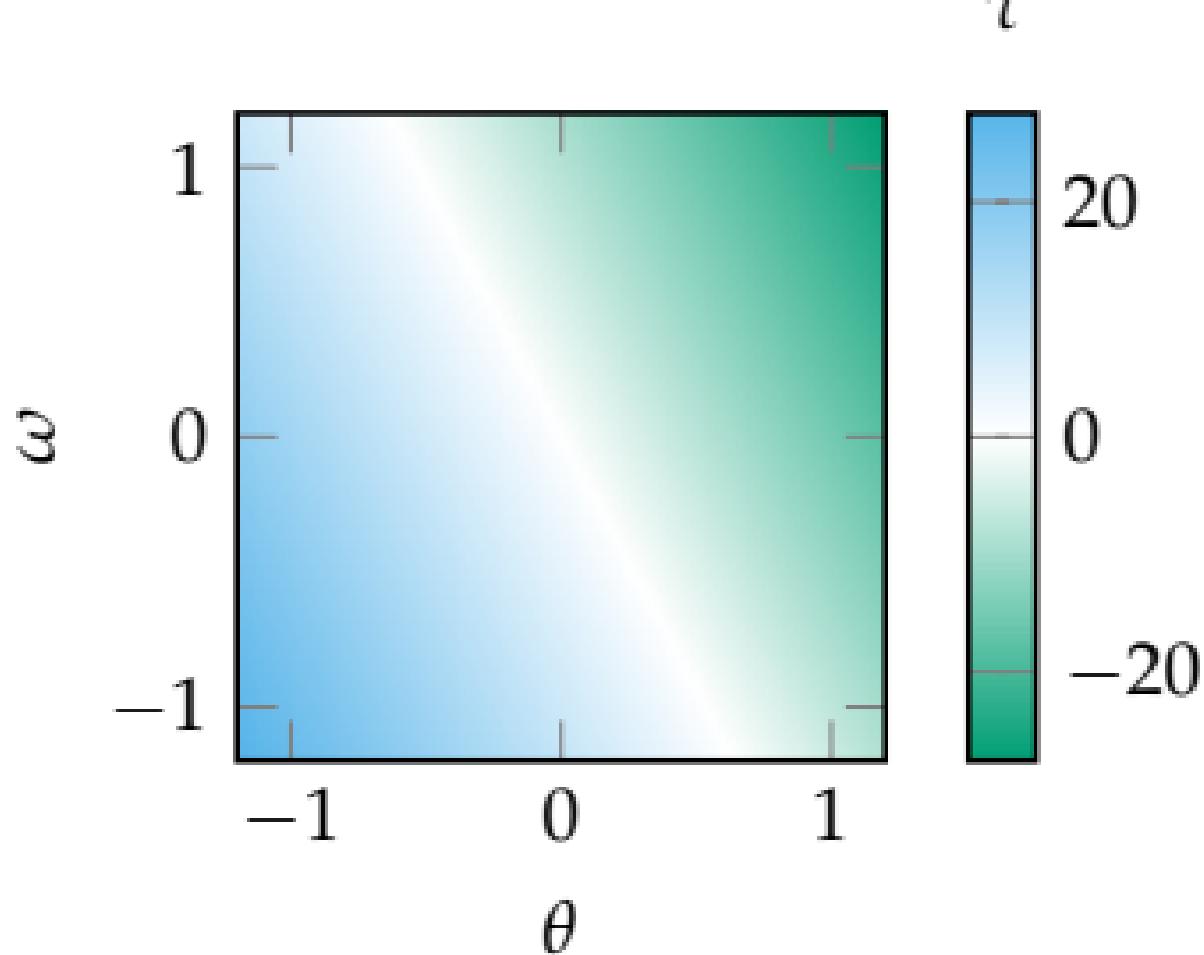


# Collision Avoidance

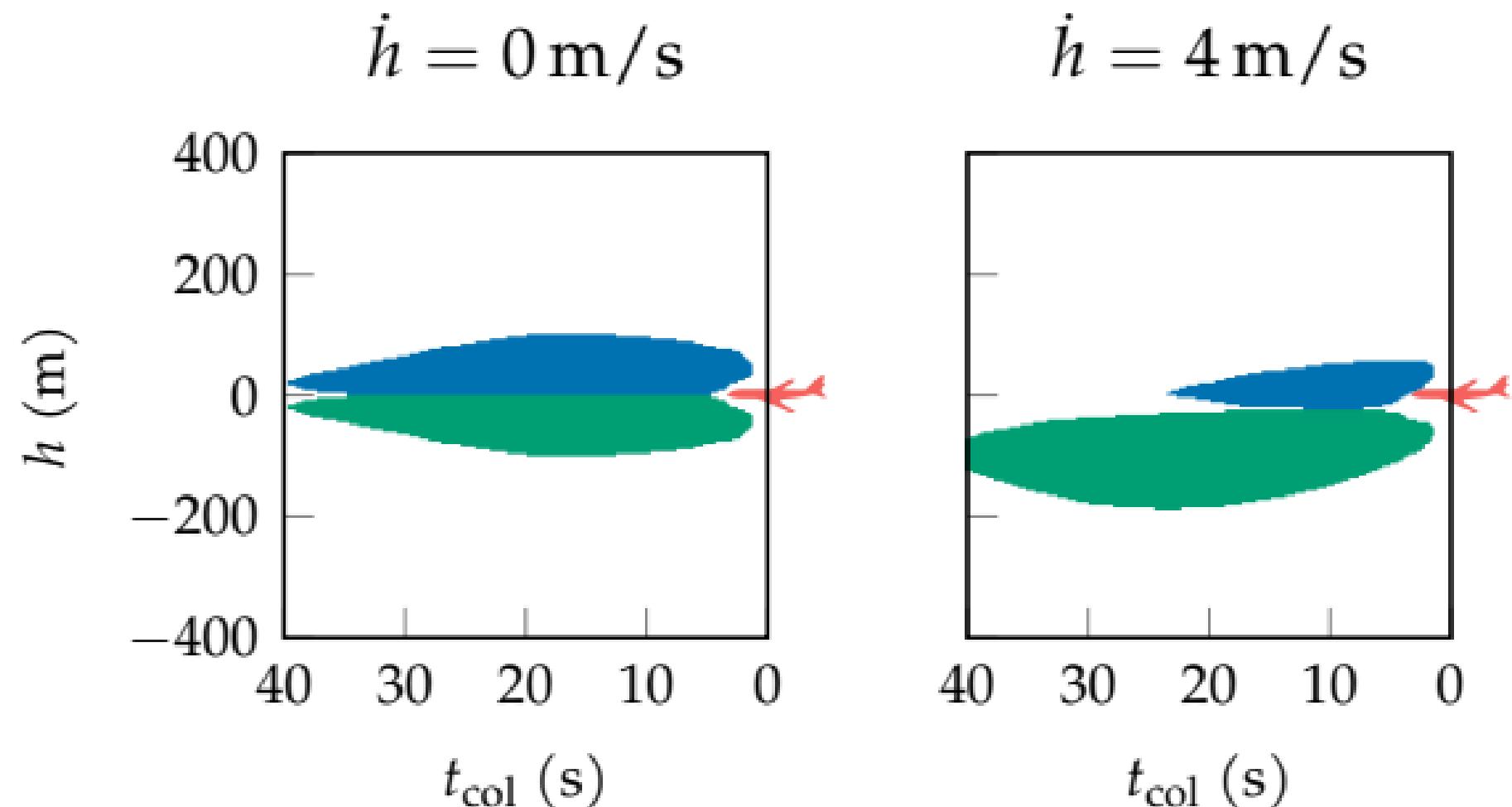


# Inverted Pendulum

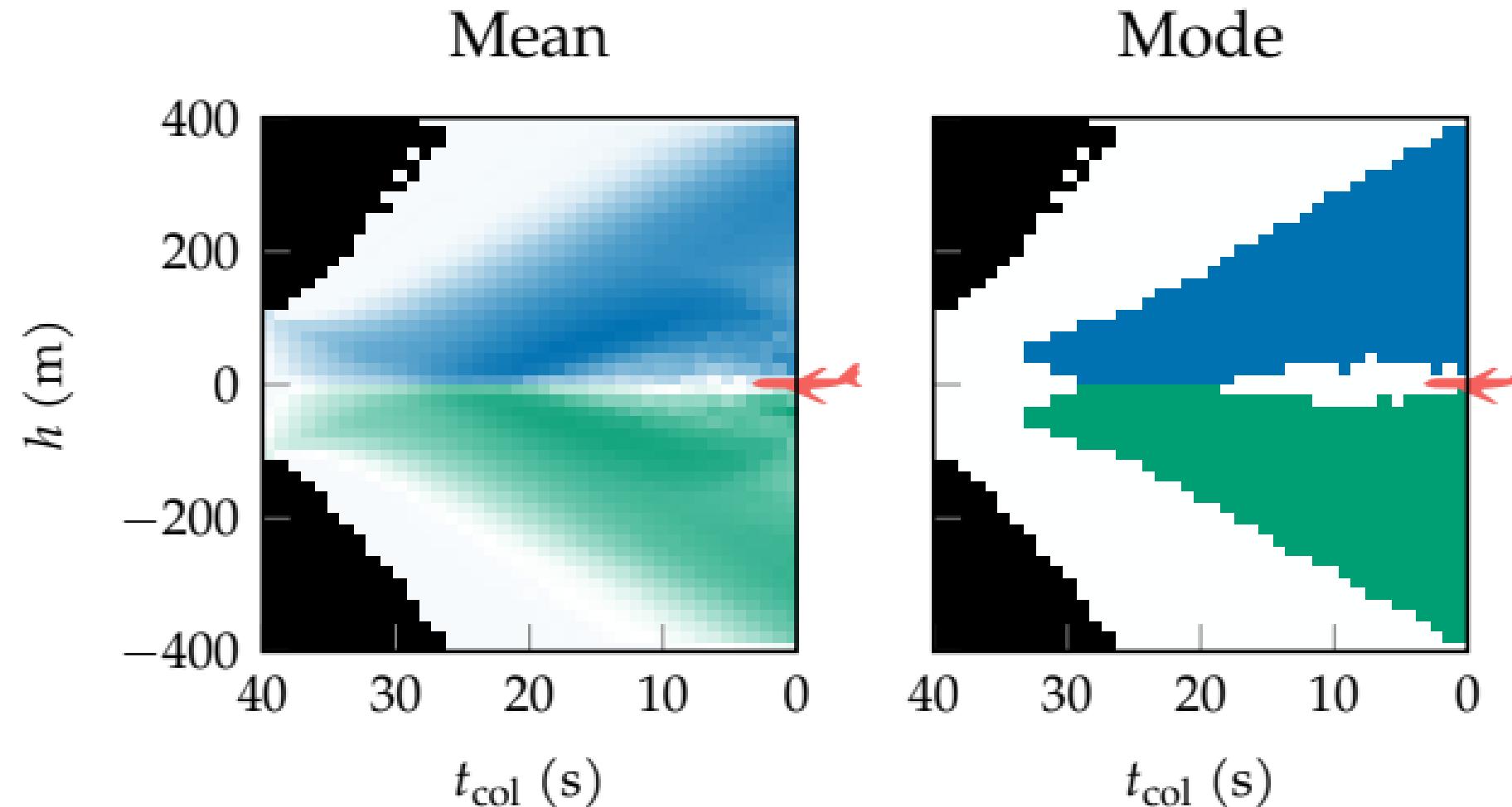


$\tau$ 

- no advisory
- descend
- climb

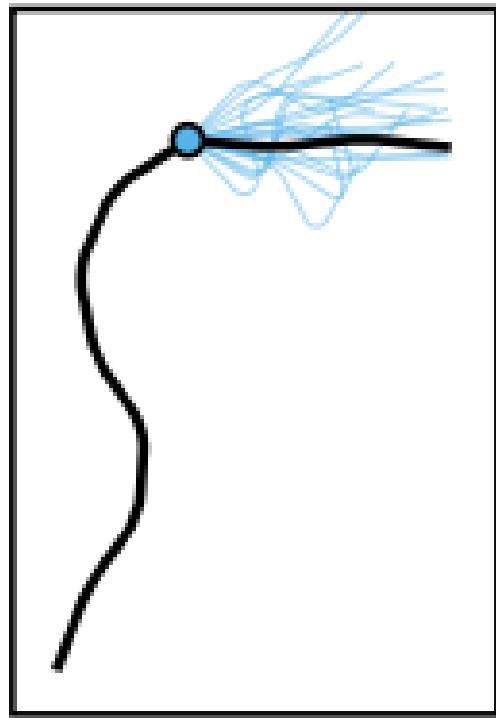
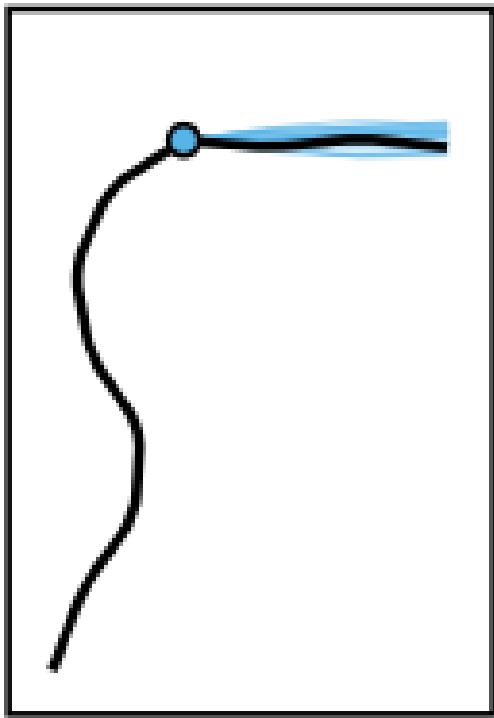


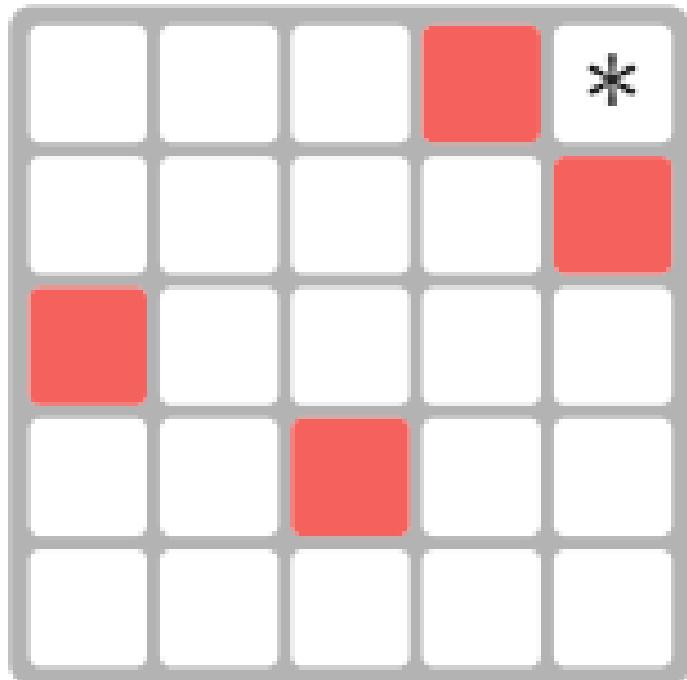
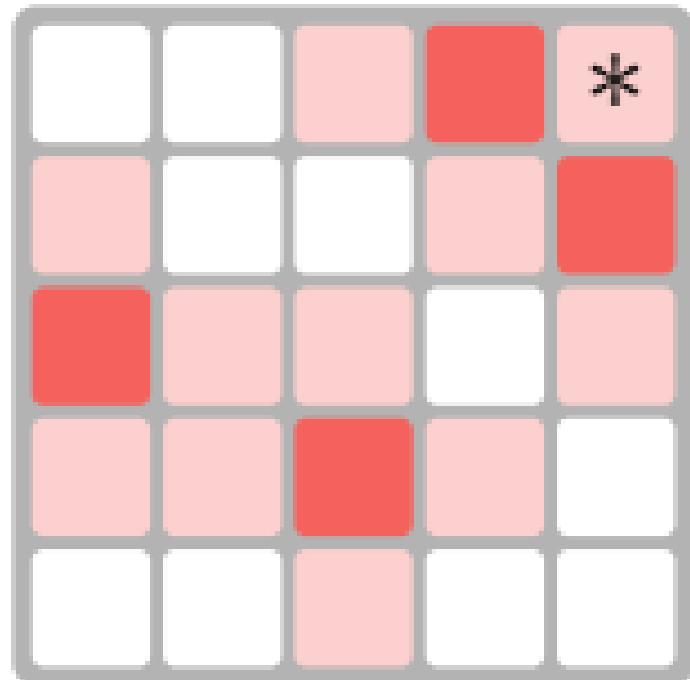
- no data
- no advisory
- descend
- climb

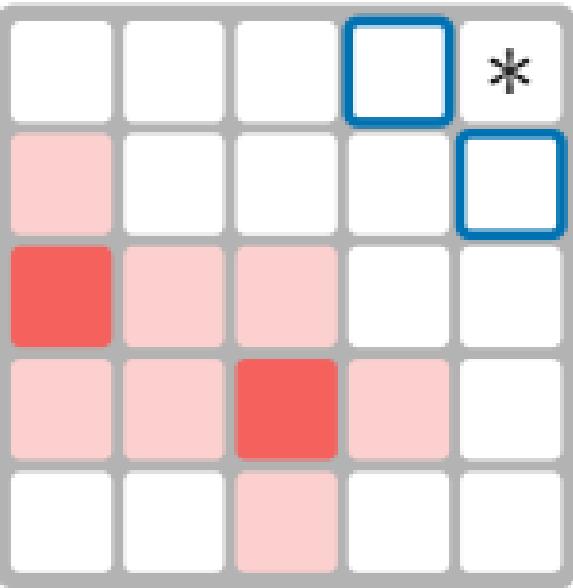
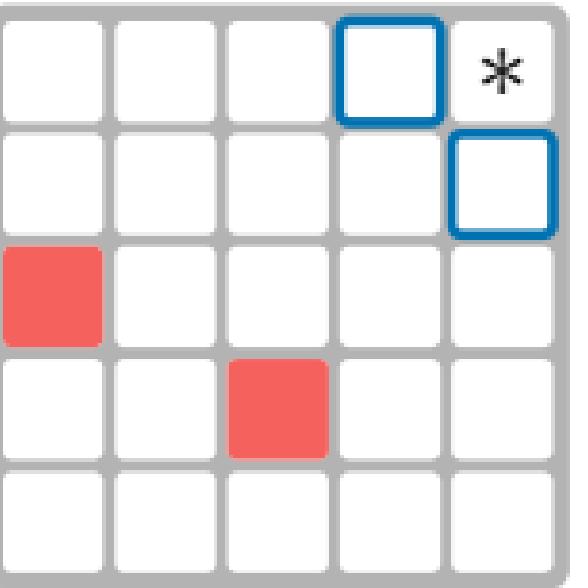
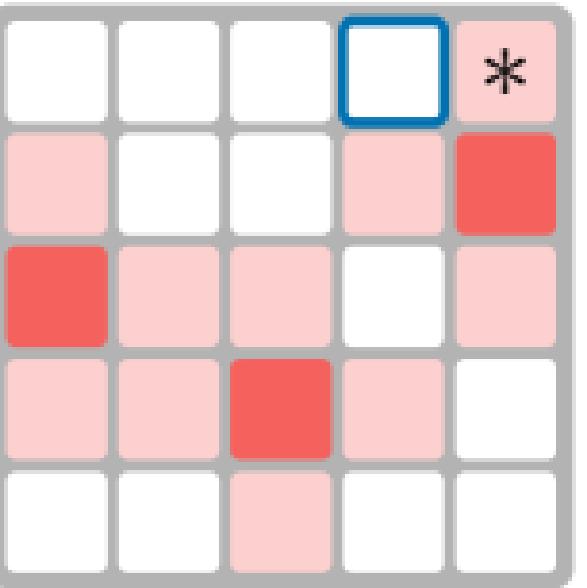
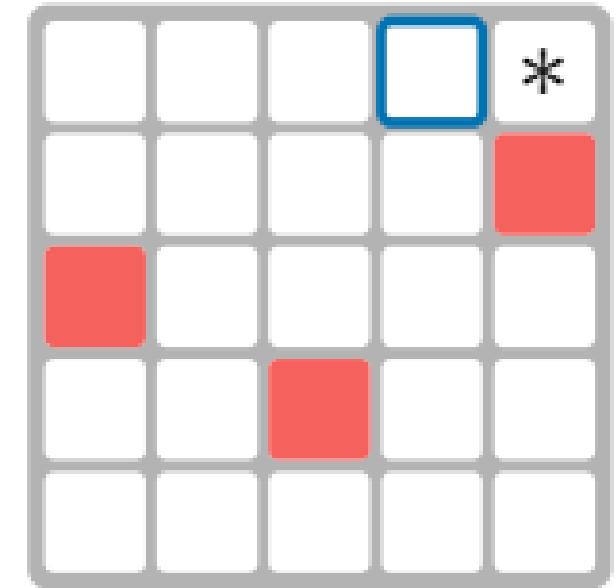


Low  
Sensitivity

High  
Sensitivity



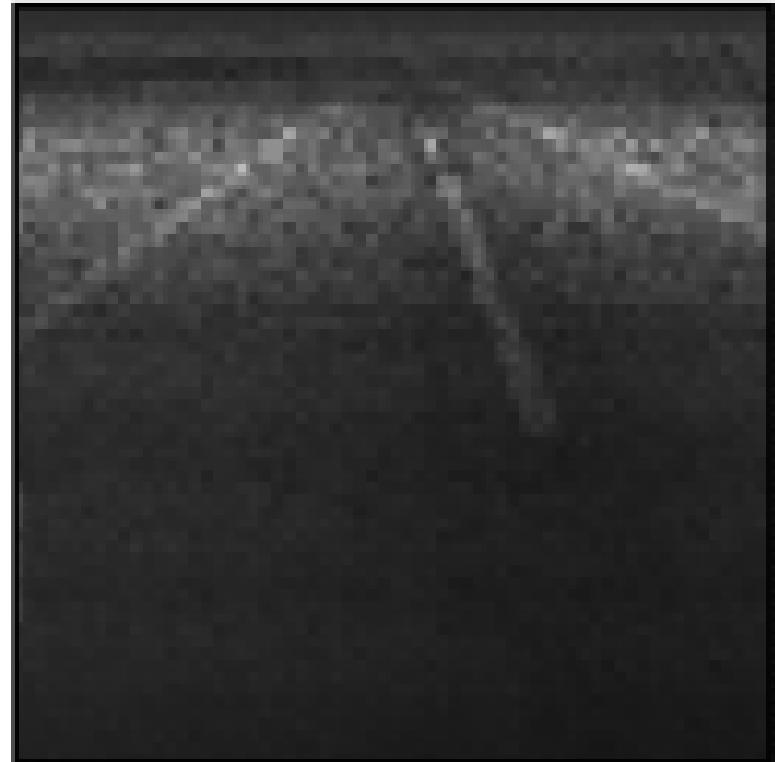
$s_t$  $P(s_{t+1})$ 

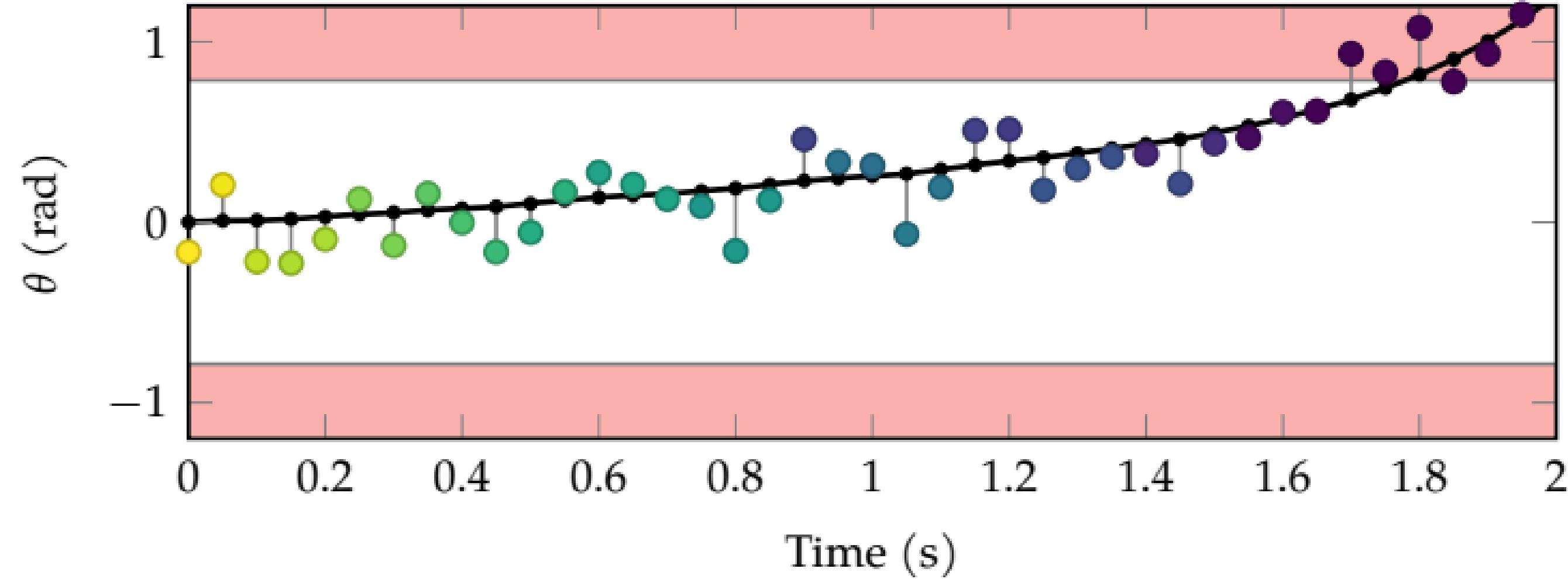
$s'_t$  $P(s'_{t+1})$  $s''_t$  $P(s''_{t+1})$ 

# Original Image

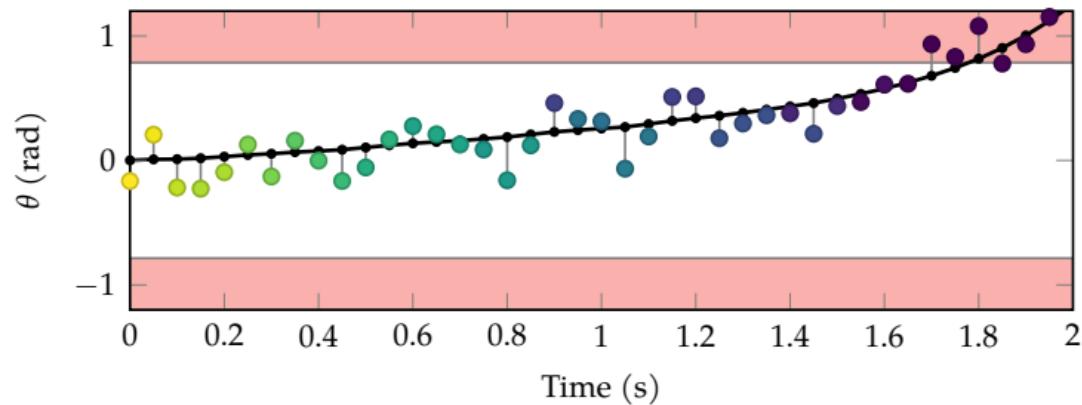


# Sensitivity Map

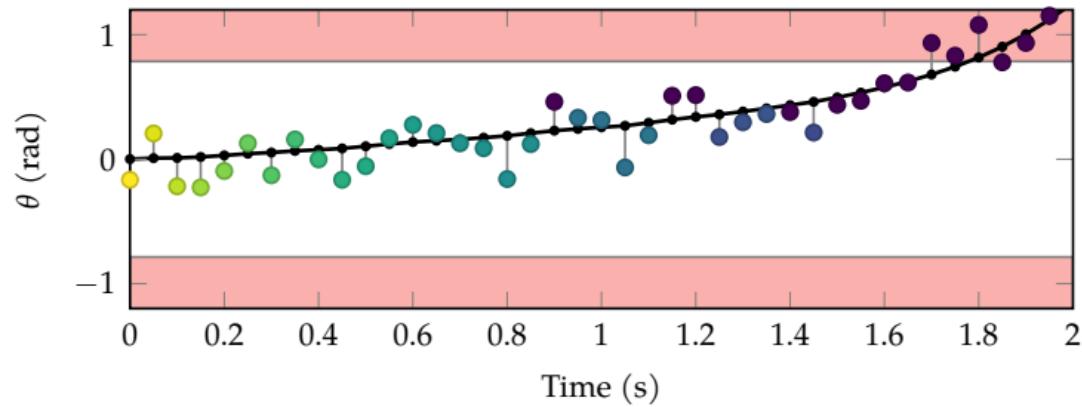




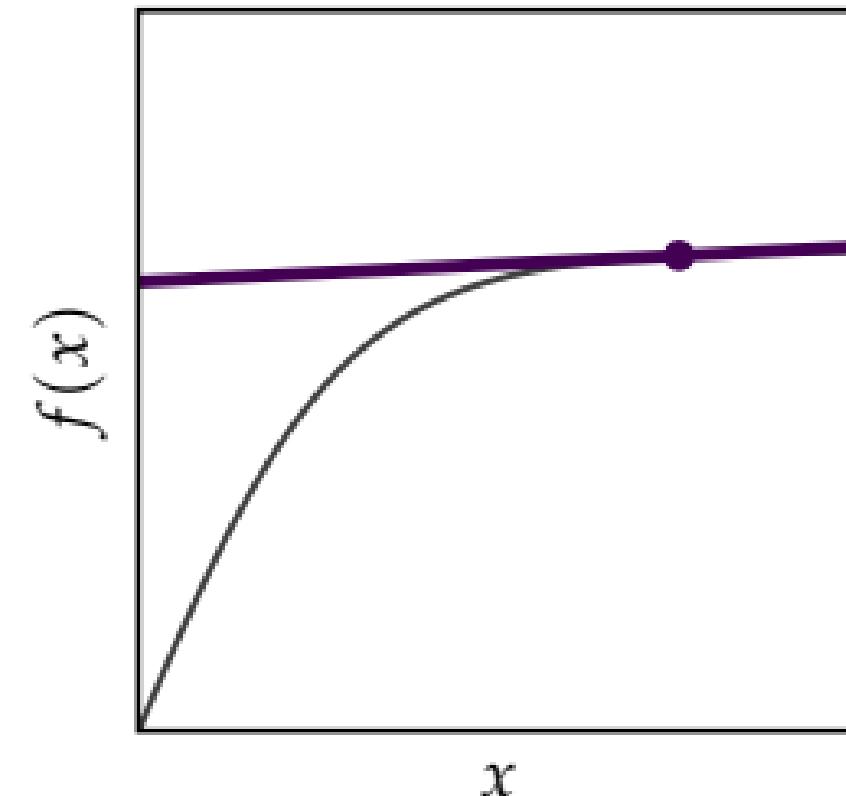
## Sensitivity



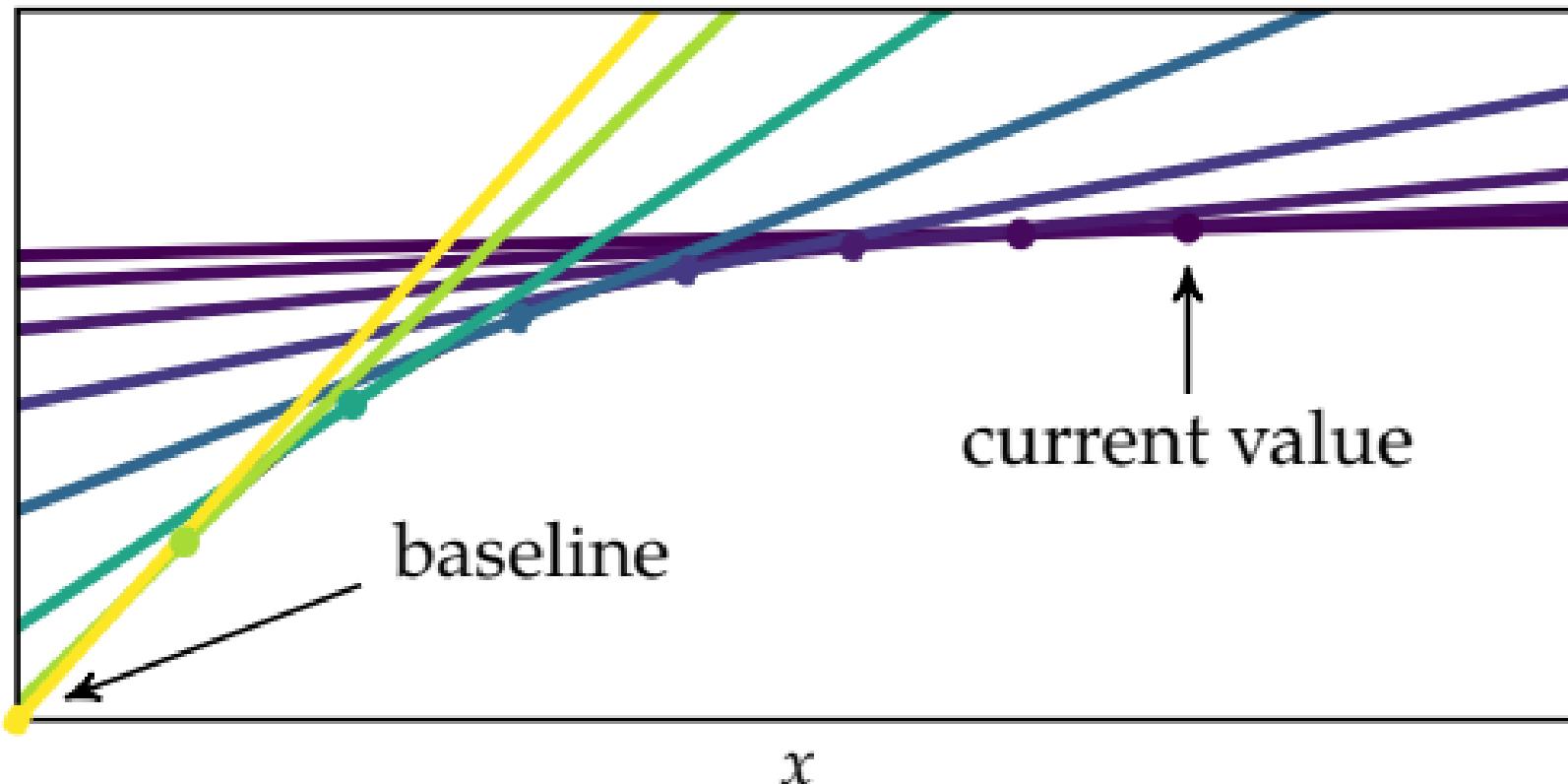
## Gradient Magnitude



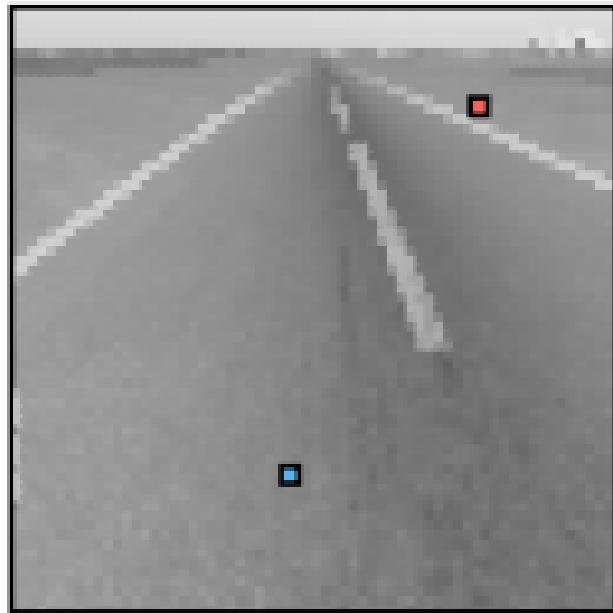
# Saturated Gradient



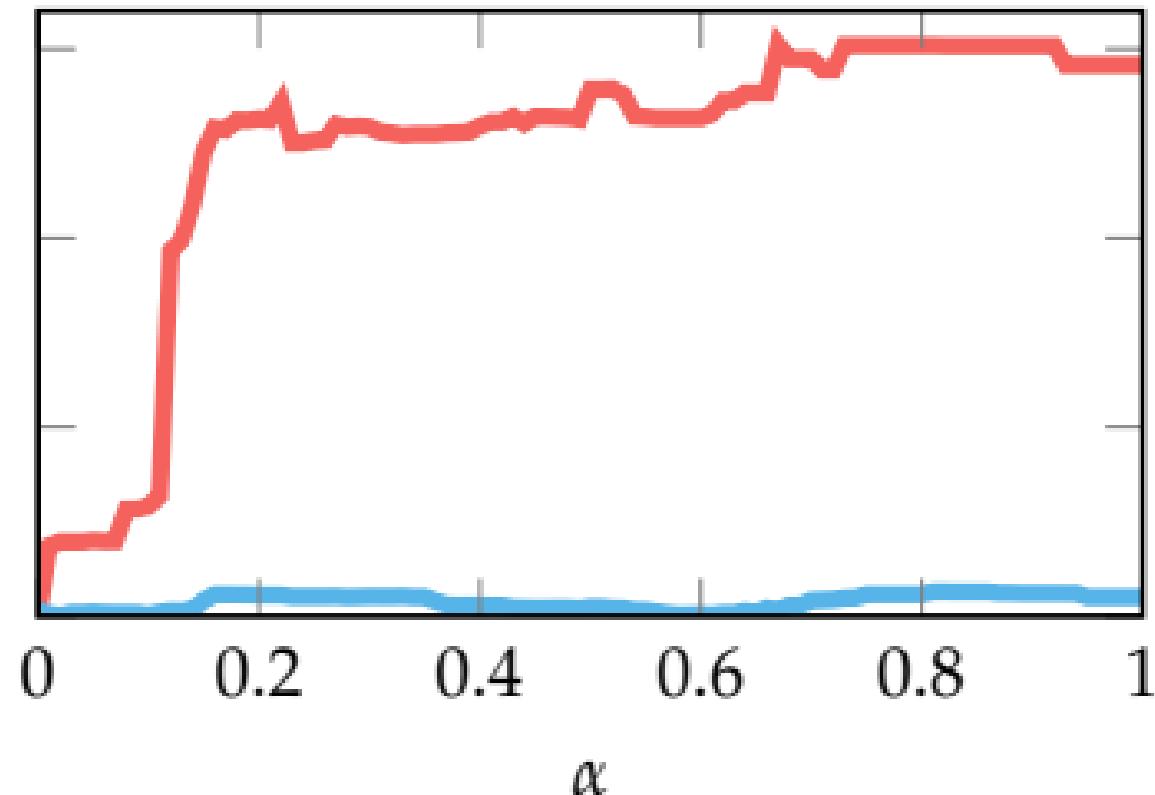
# Integrated Gradients

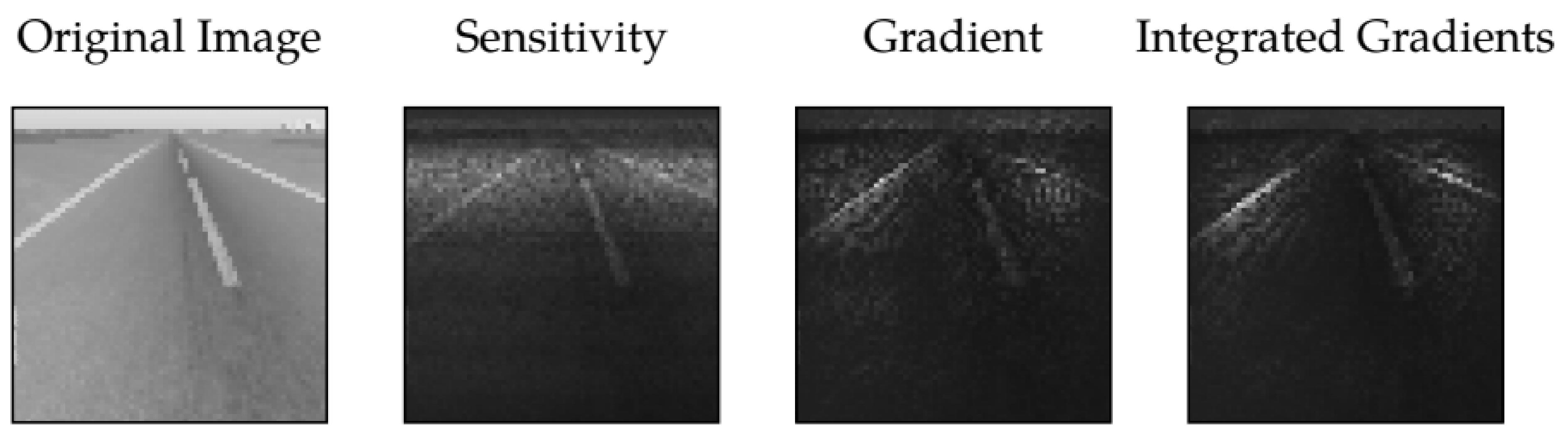


# Original Image

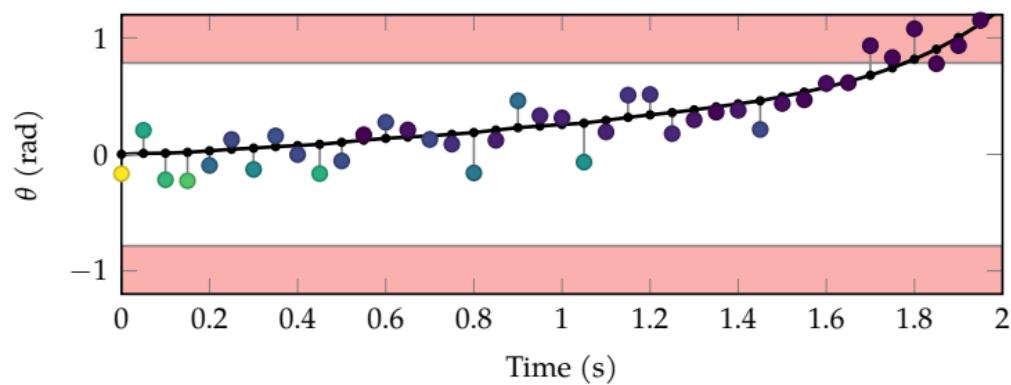


# Gradient Values

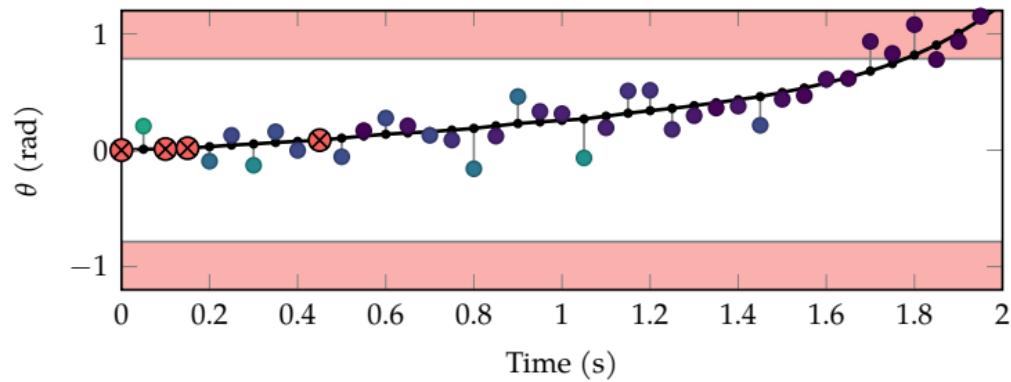




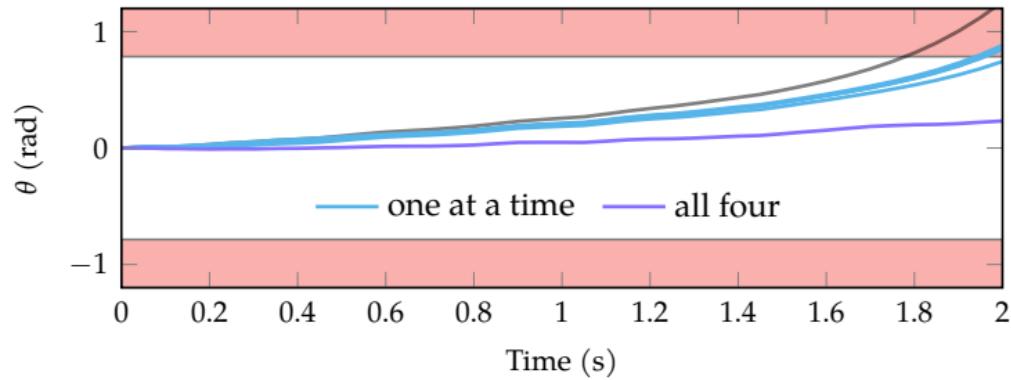
# Shapley Values



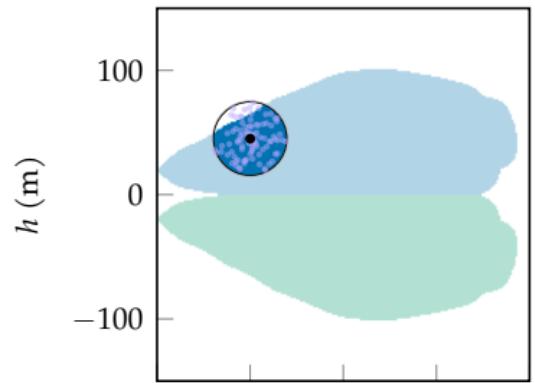
# Important Features



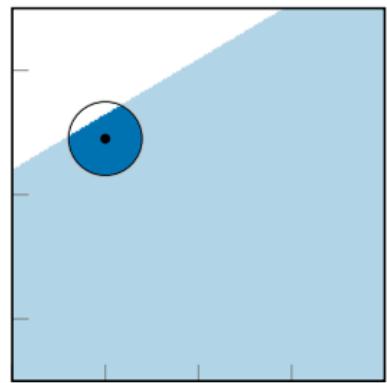
# Removing Important Features



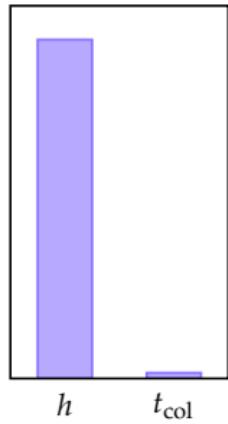
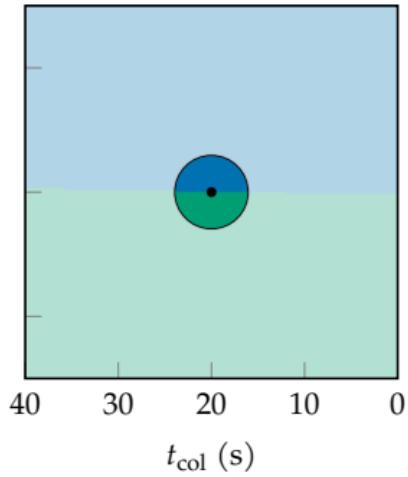
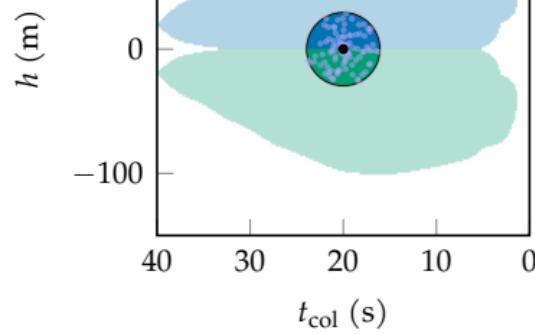
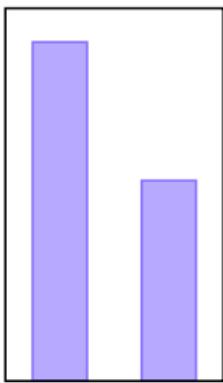
Original Policy

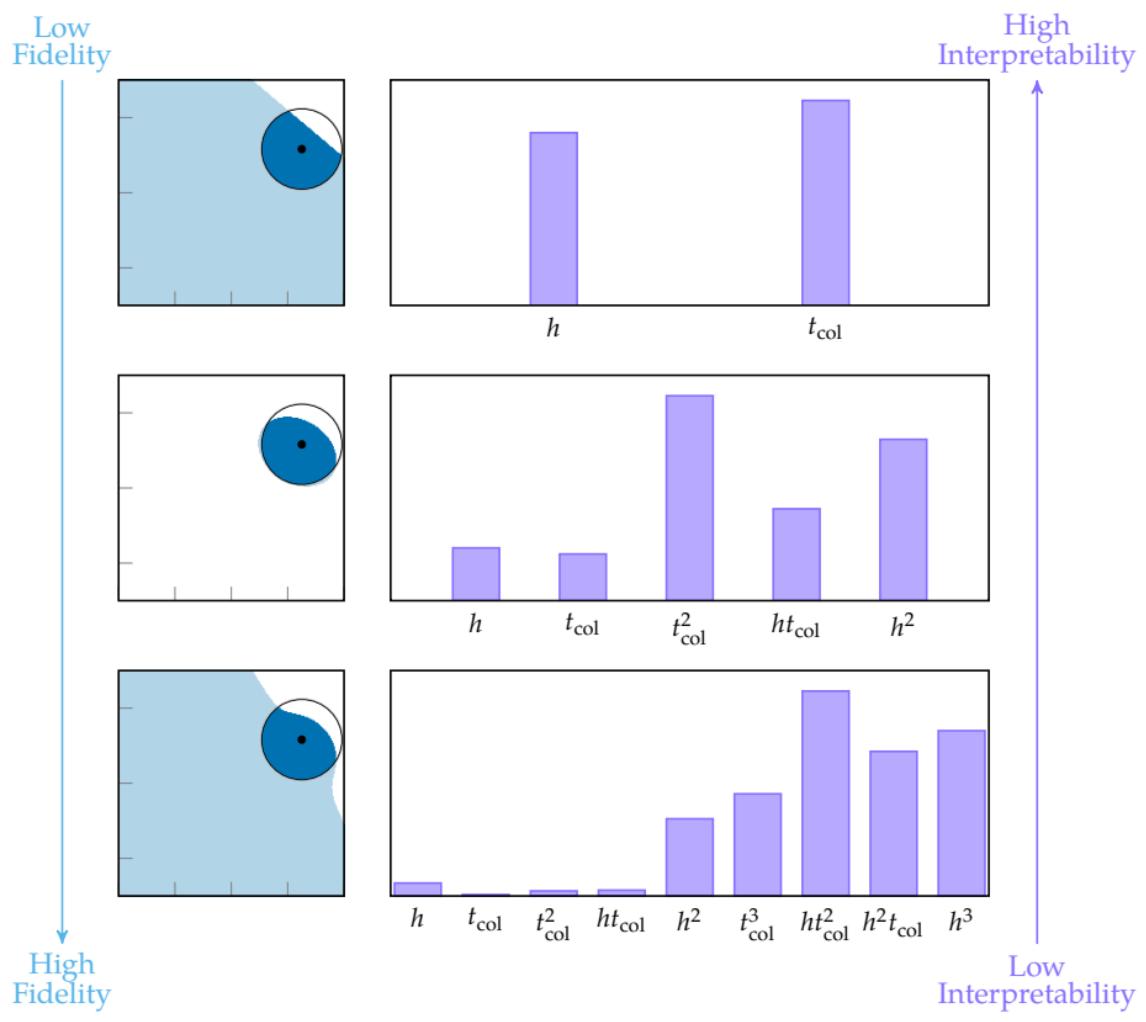


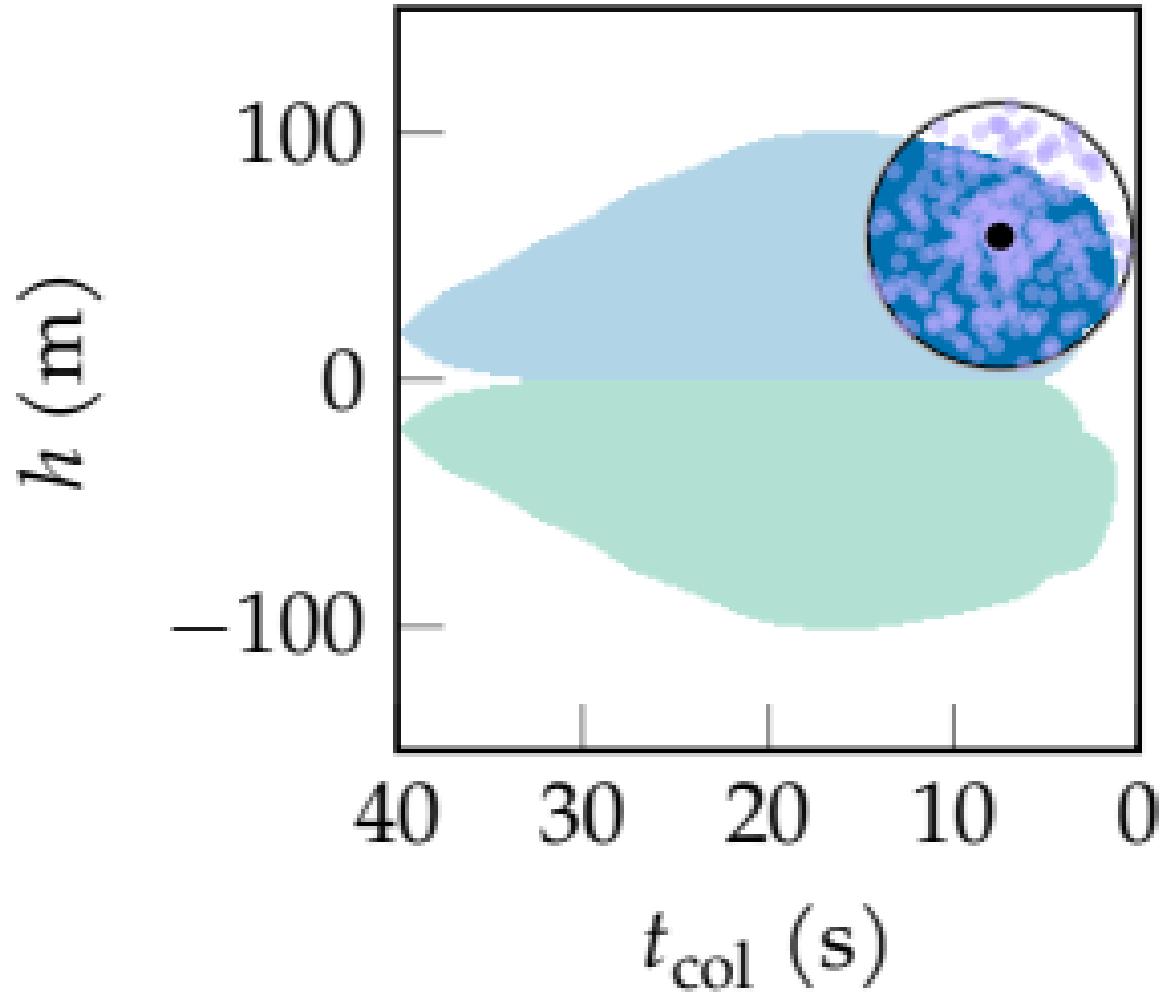
Linear Approximation



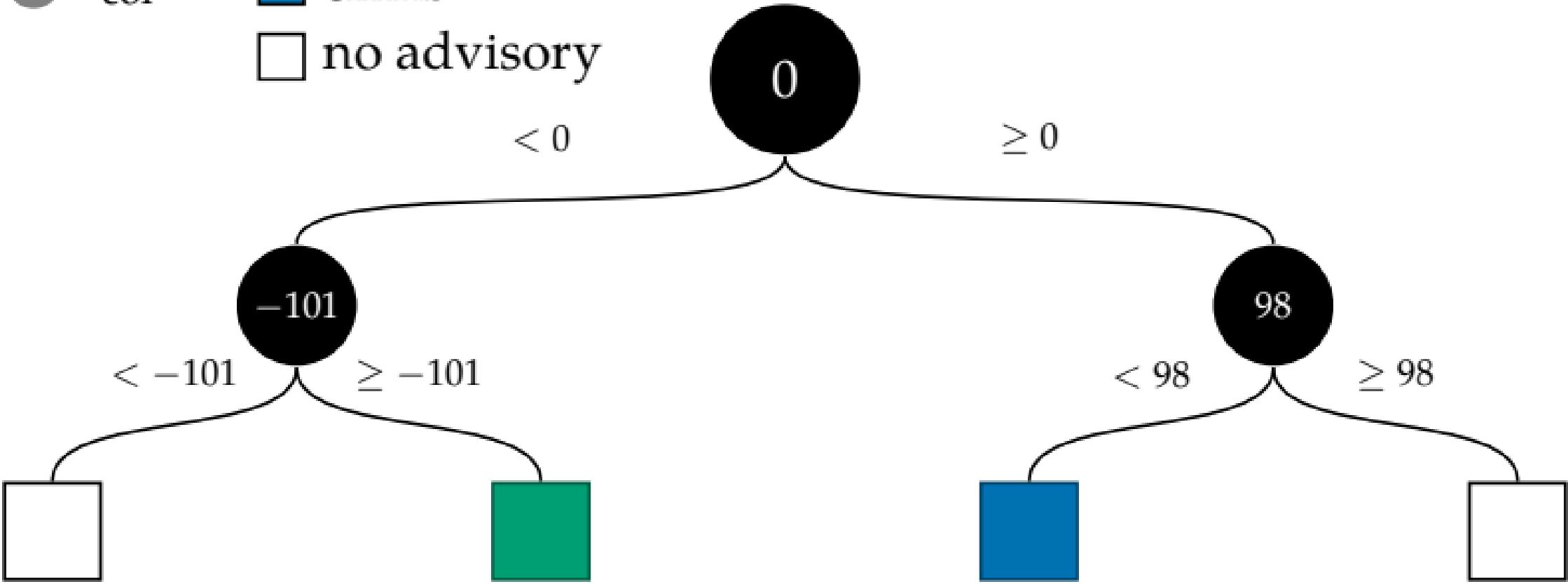
Feature Weights

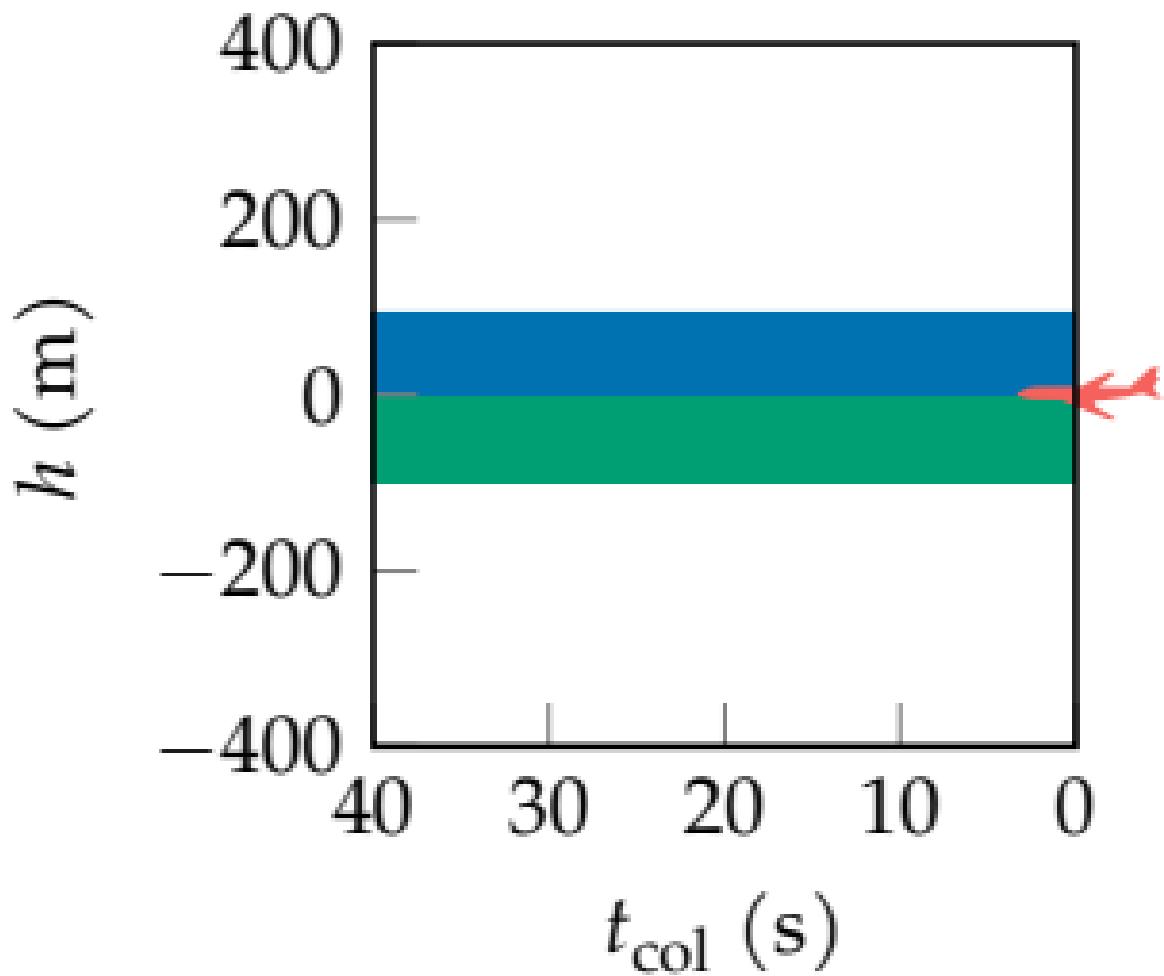




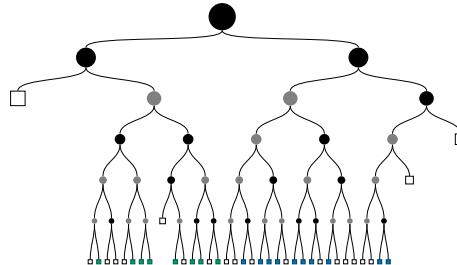
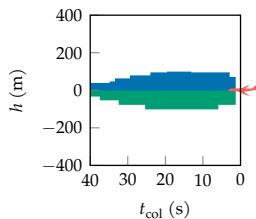
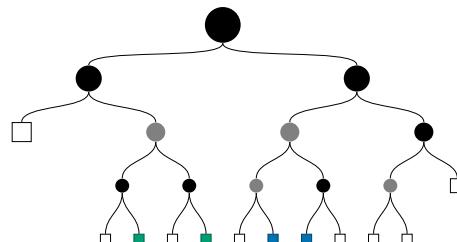
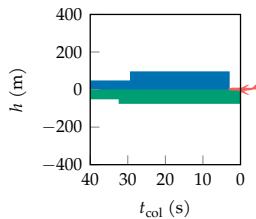
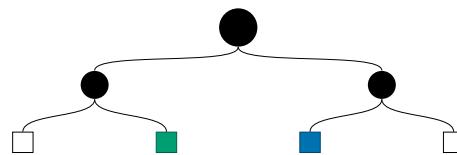
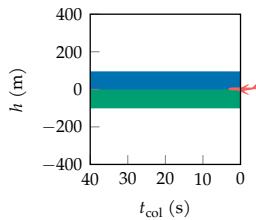


- $h$
- descend
- $t_{\text{col}}$
- climb
- no advisory

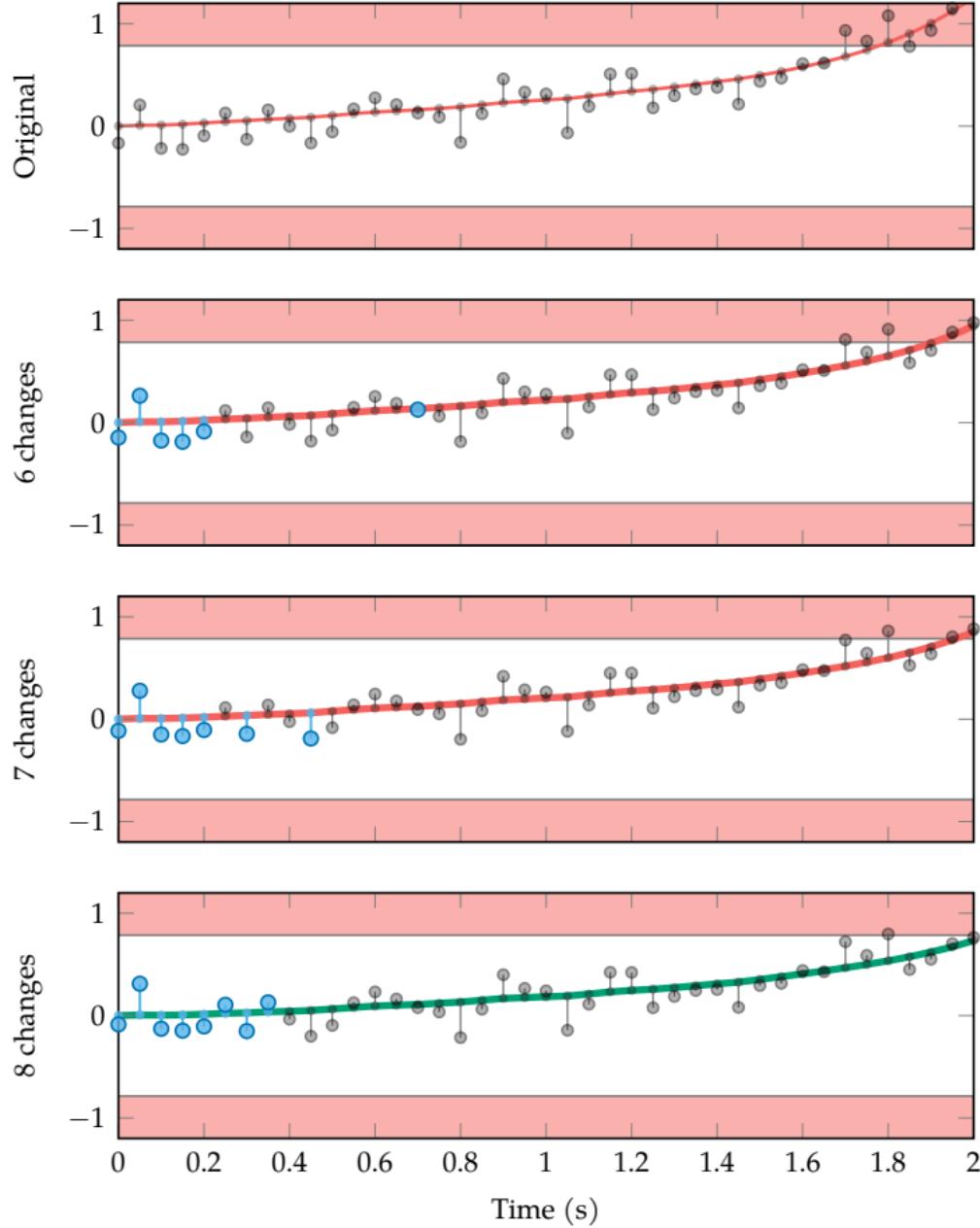


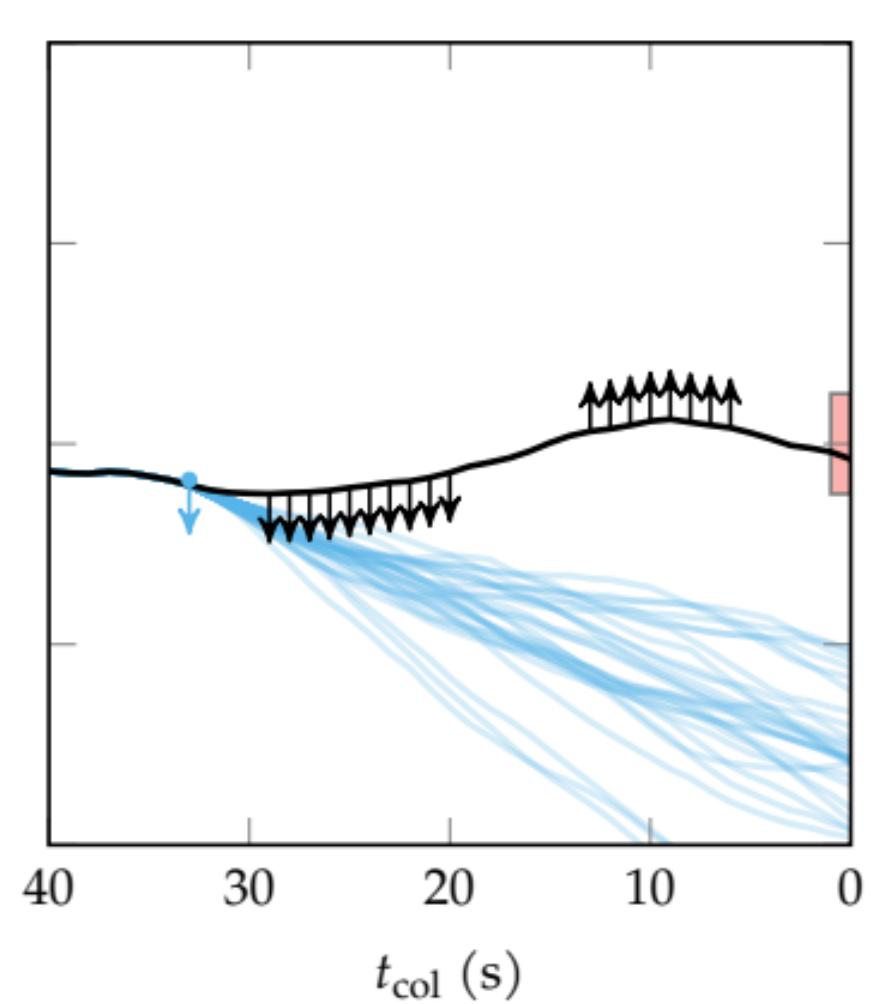
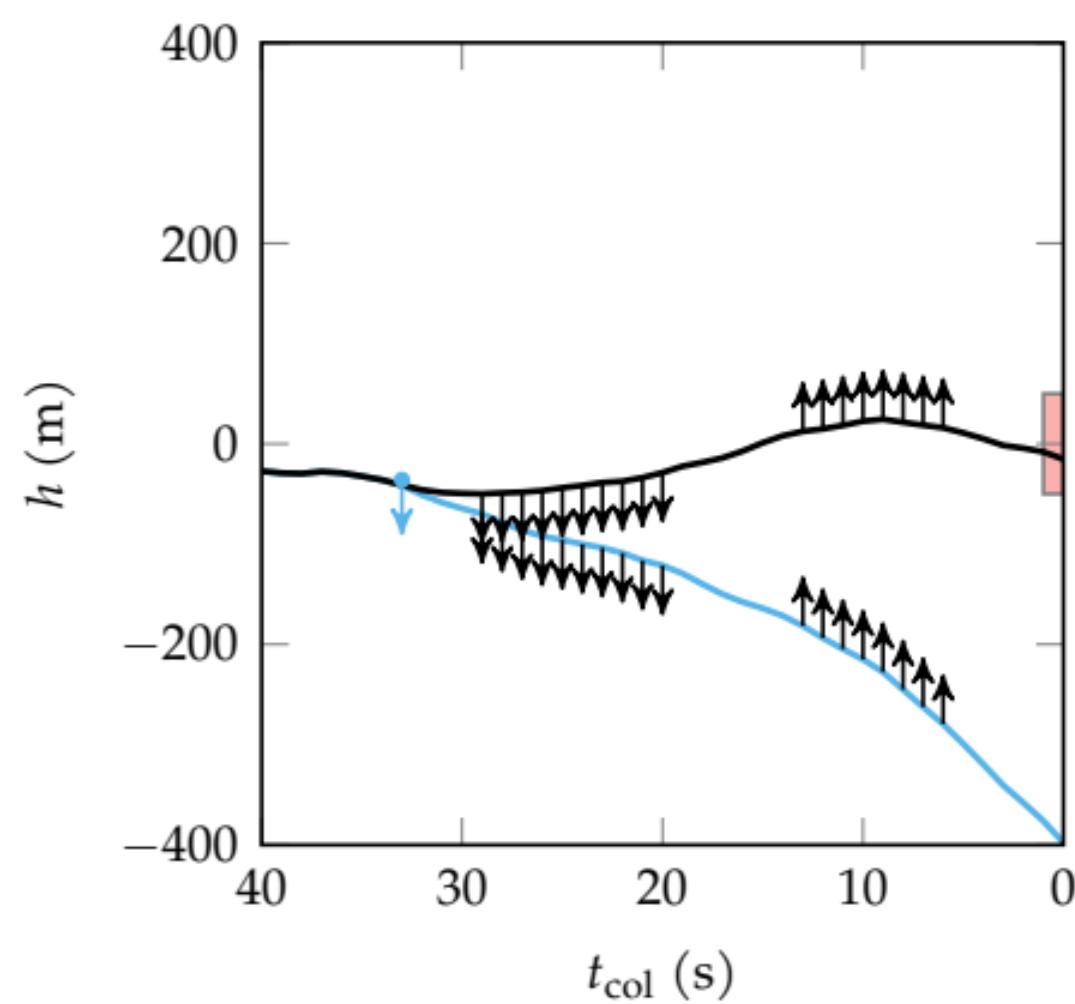


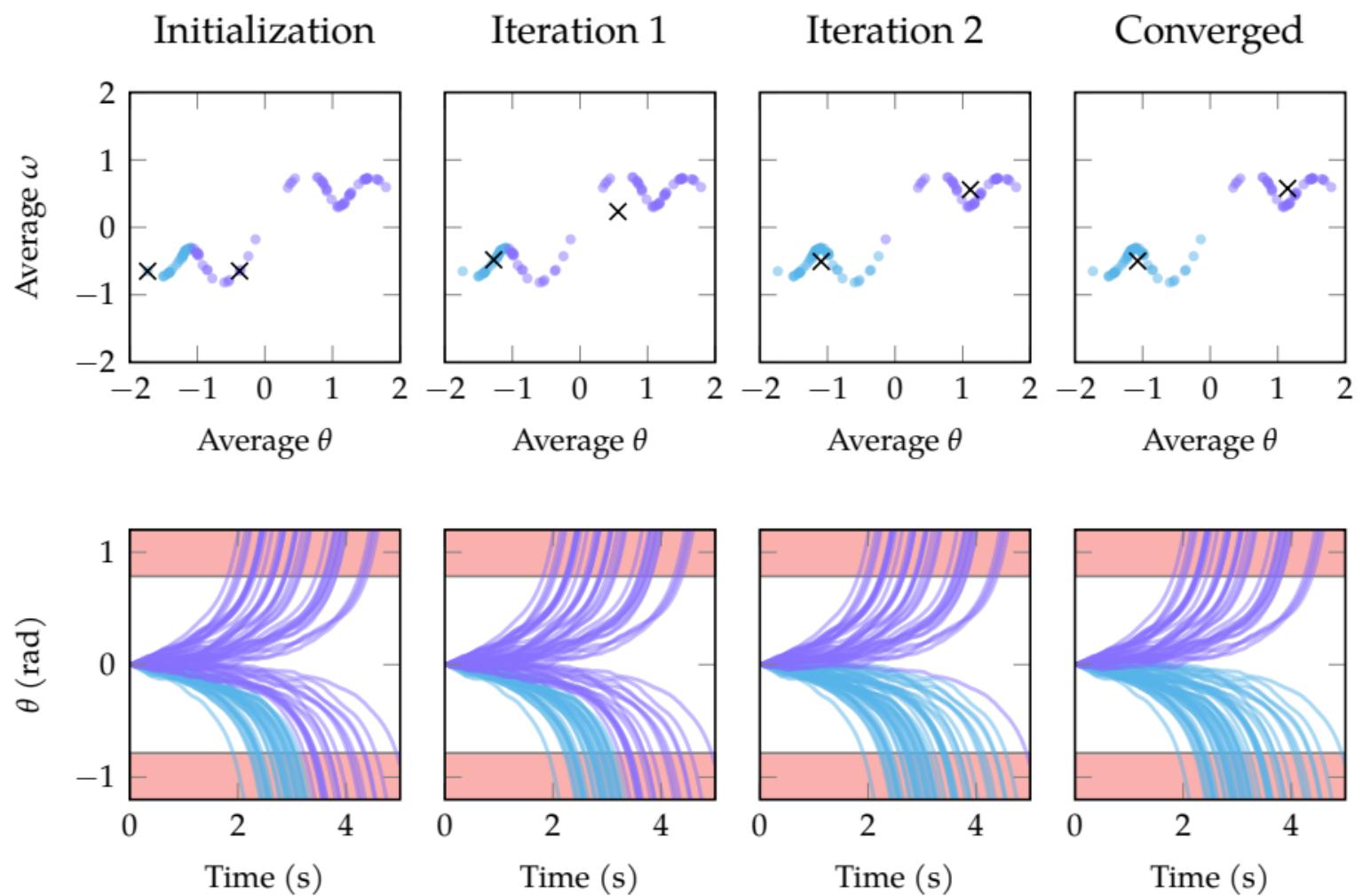
Low Fidelity High Interpretability

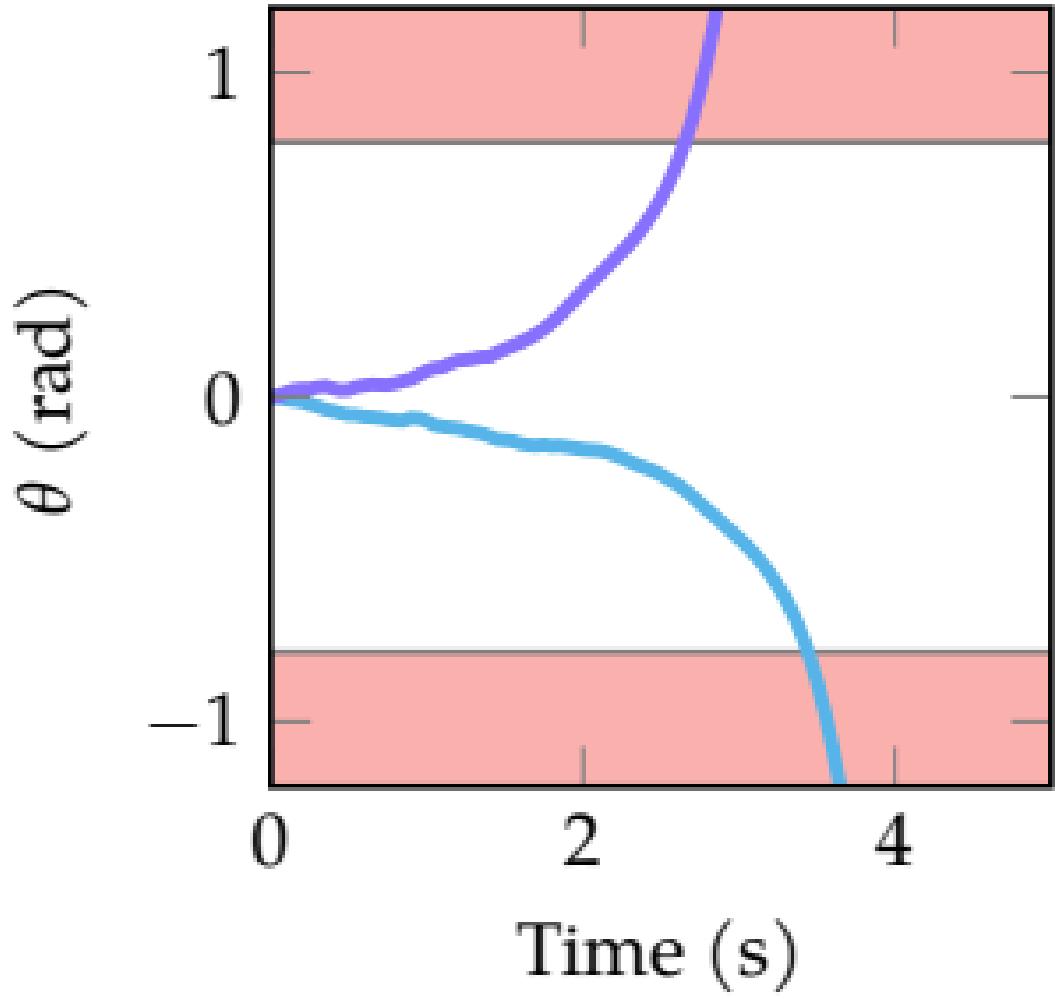


High Fidelity Low Interpretability

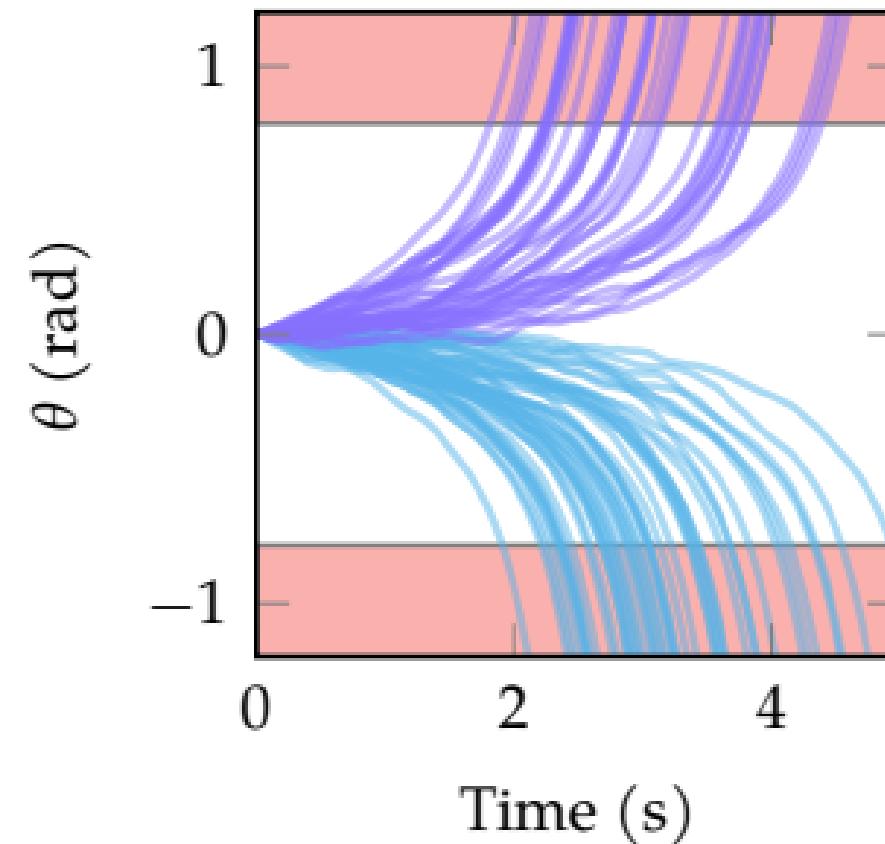




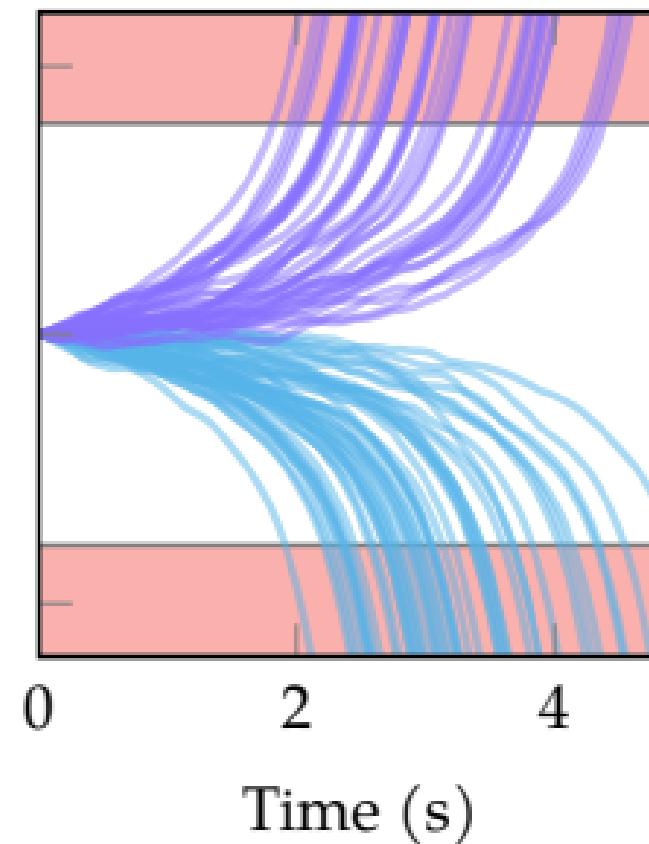




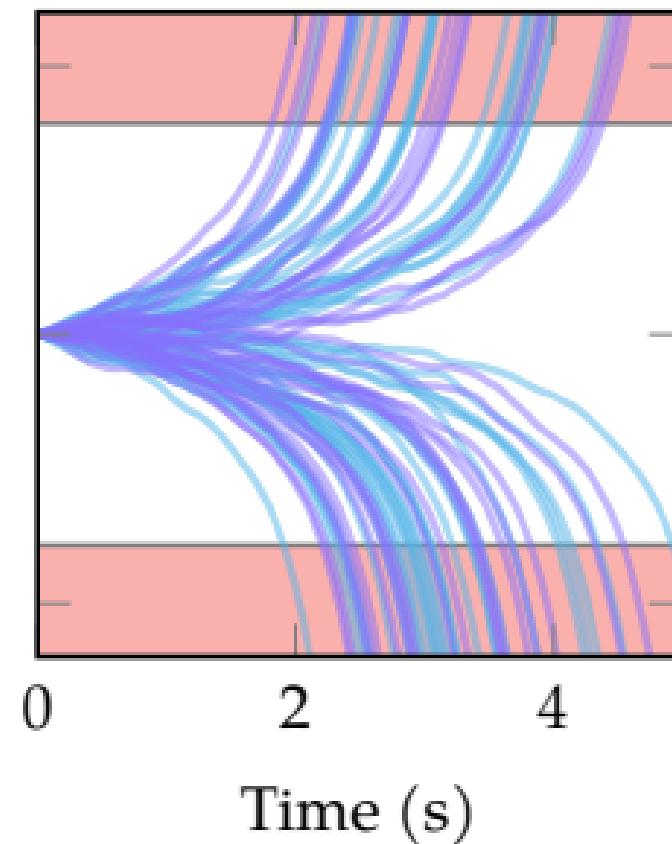
# State Features

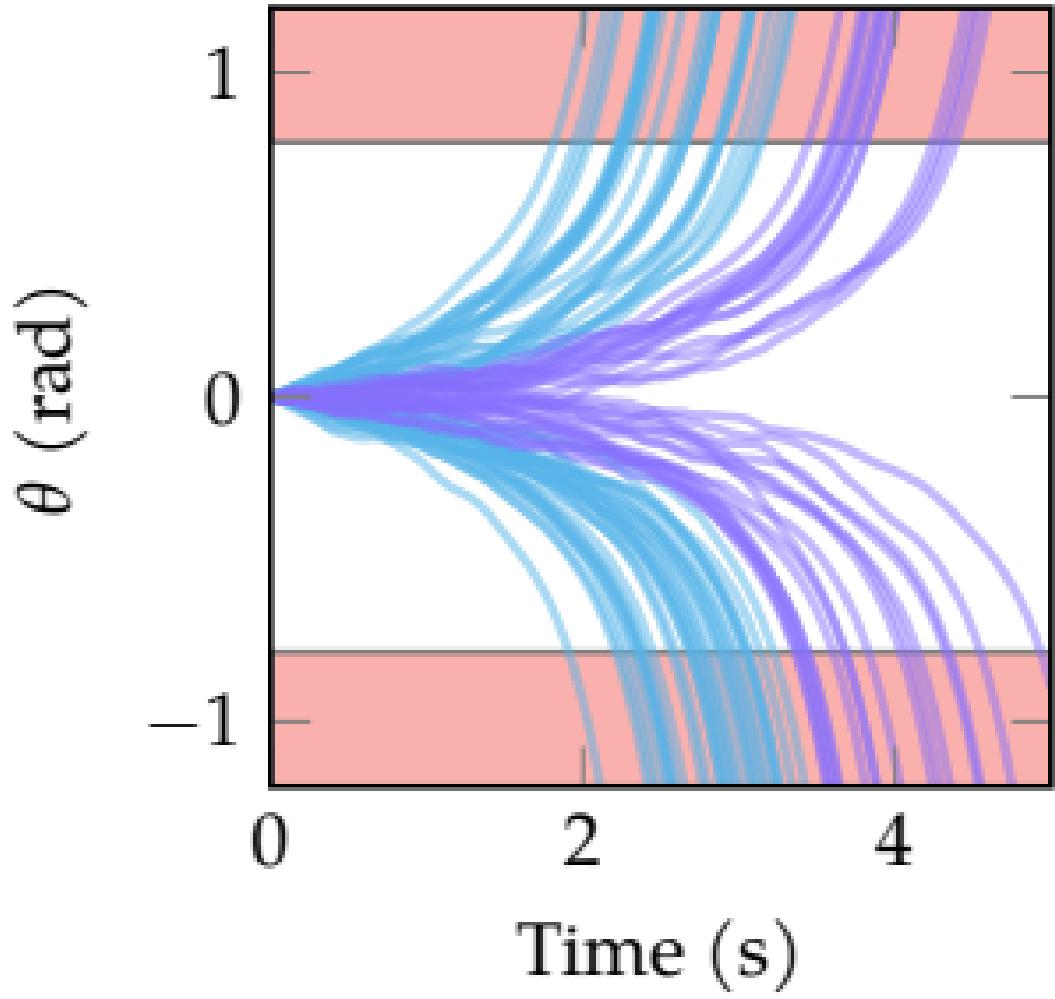


# Action Features

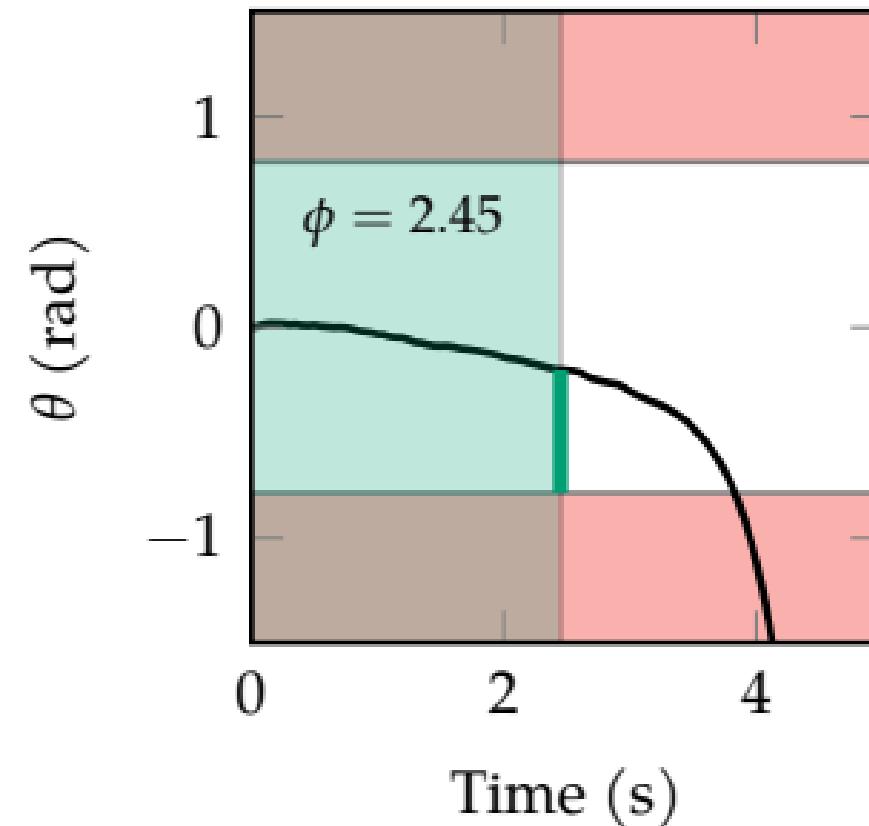


# Disturbance Features

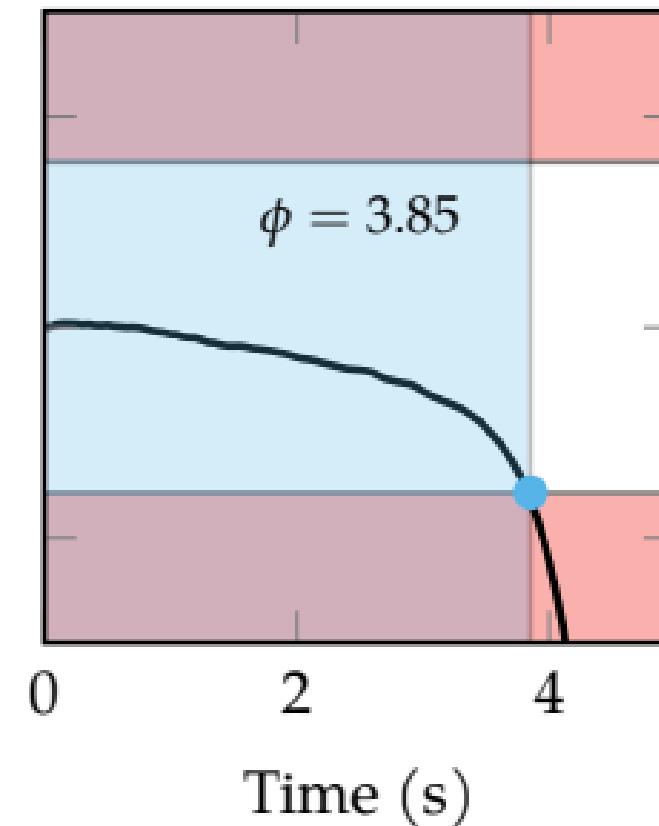




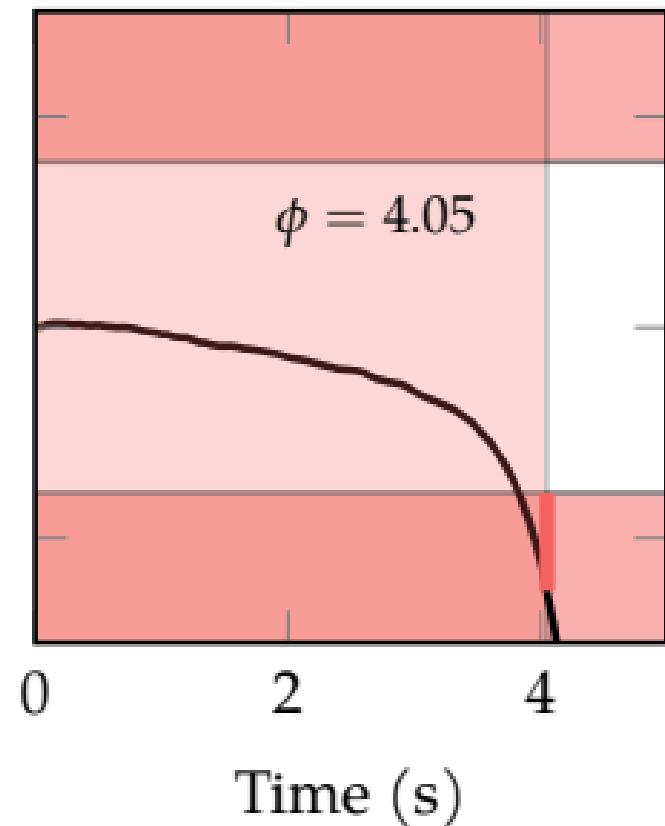
Satisfied

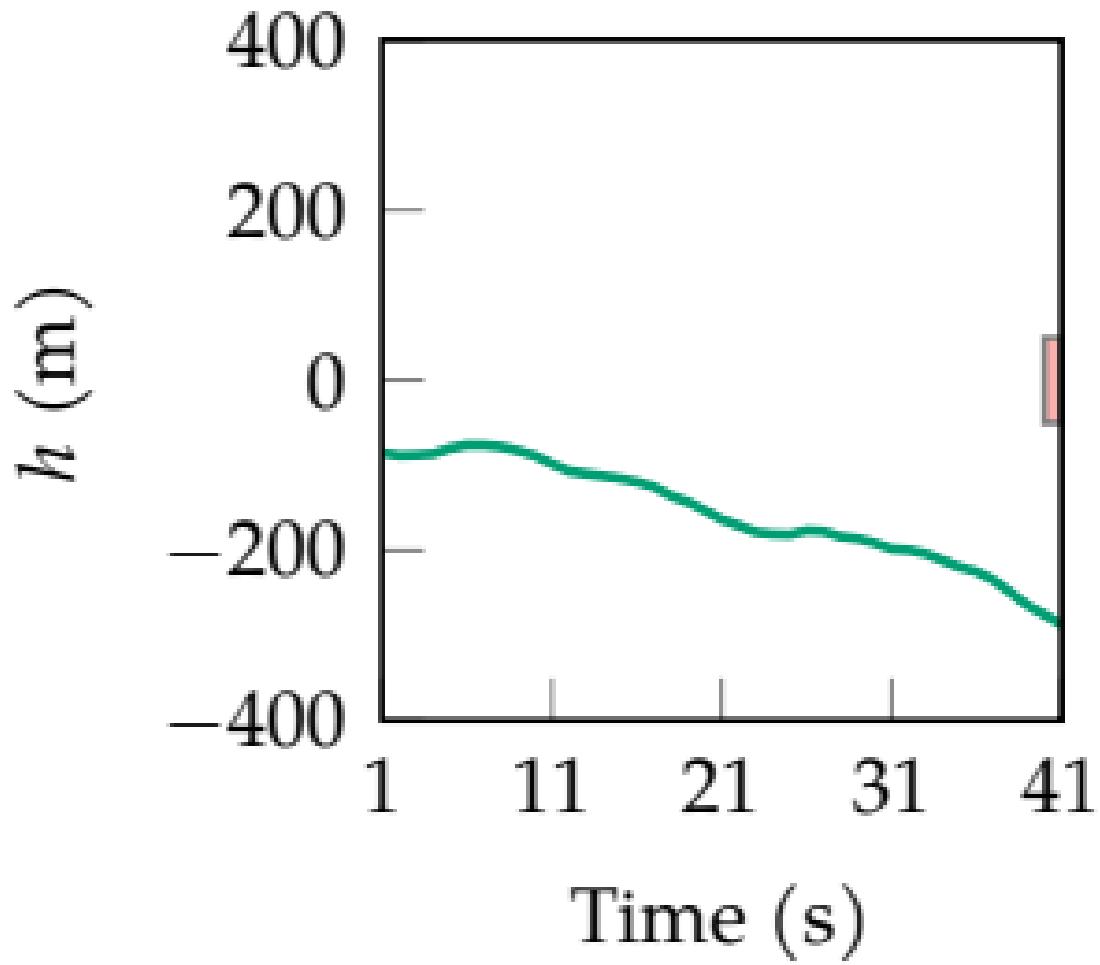


Marginally Satisfied



Not Satisfied







No Clouds



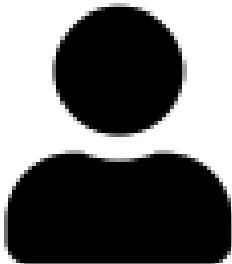
No Glare

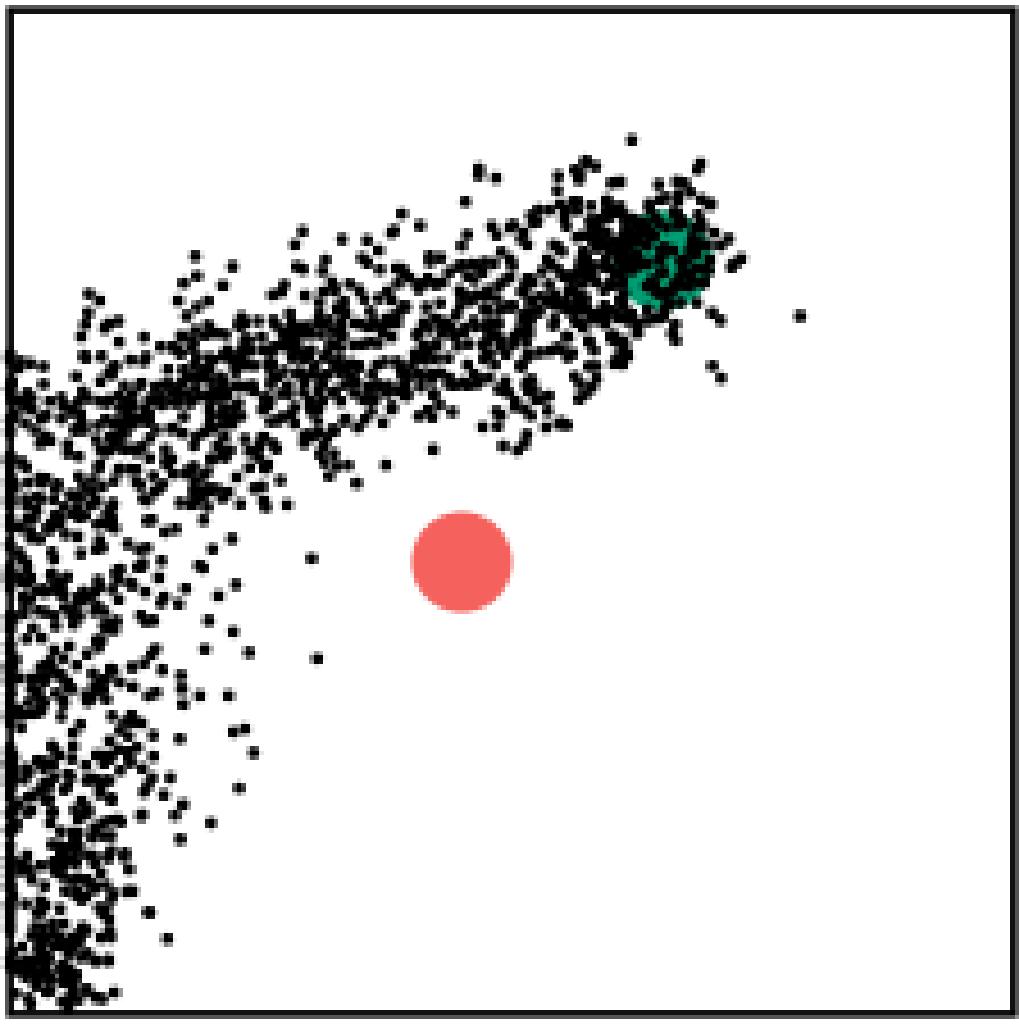


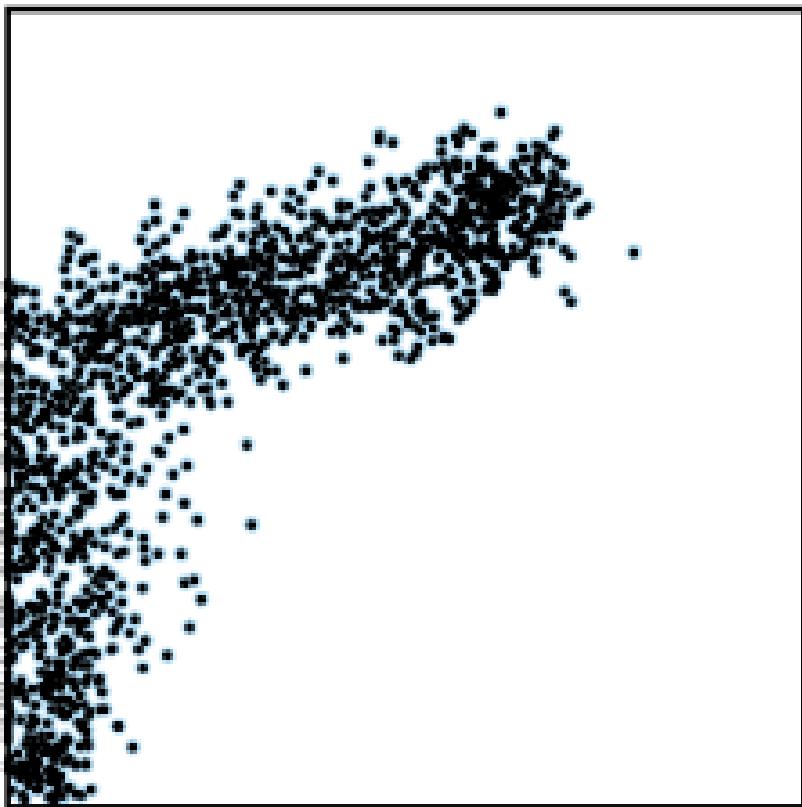
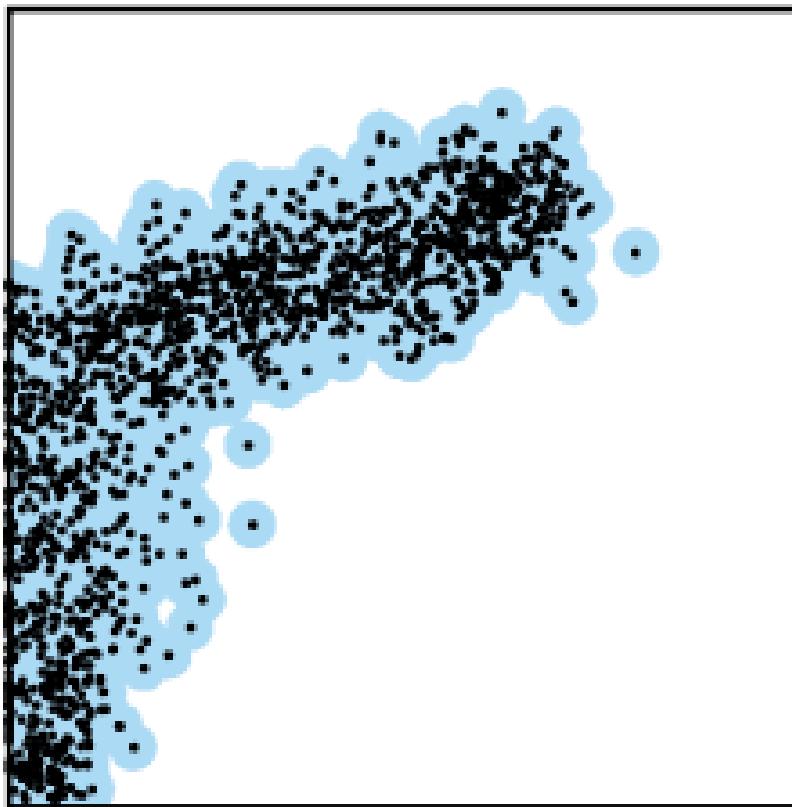
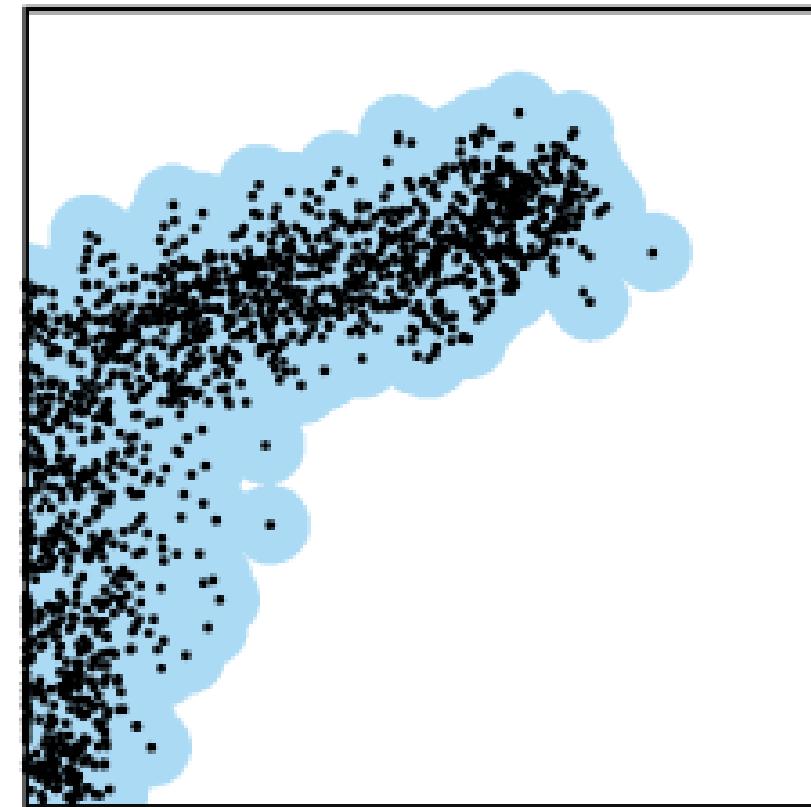
Daytime

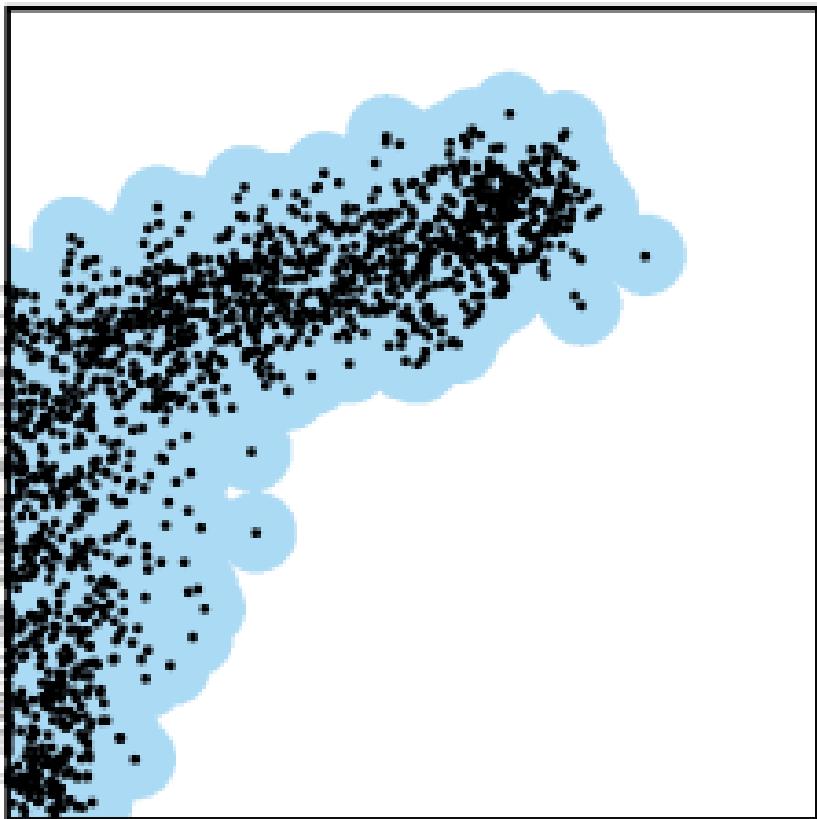
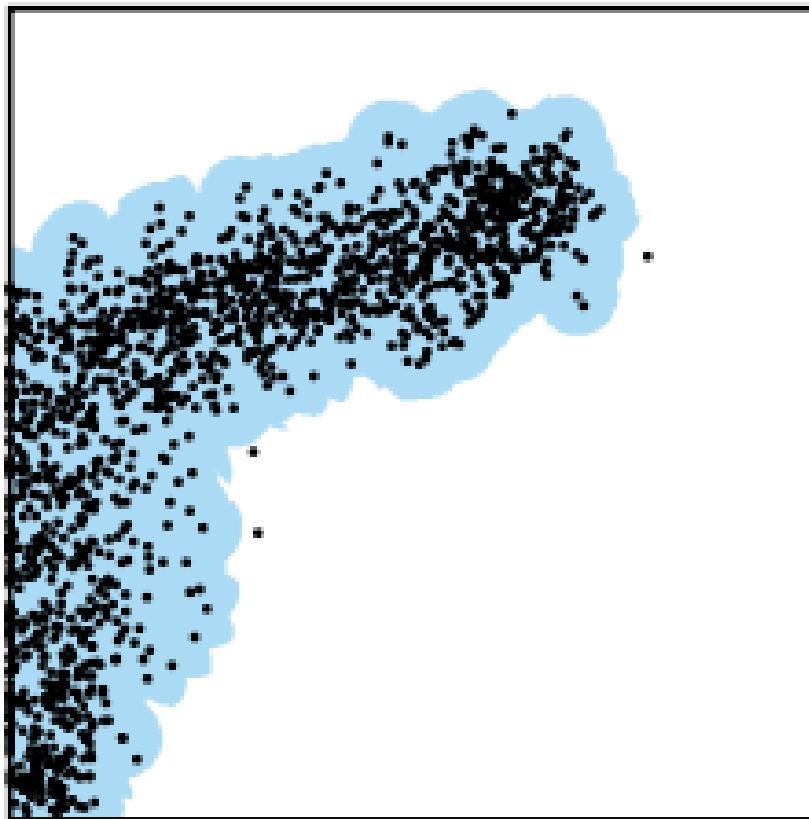
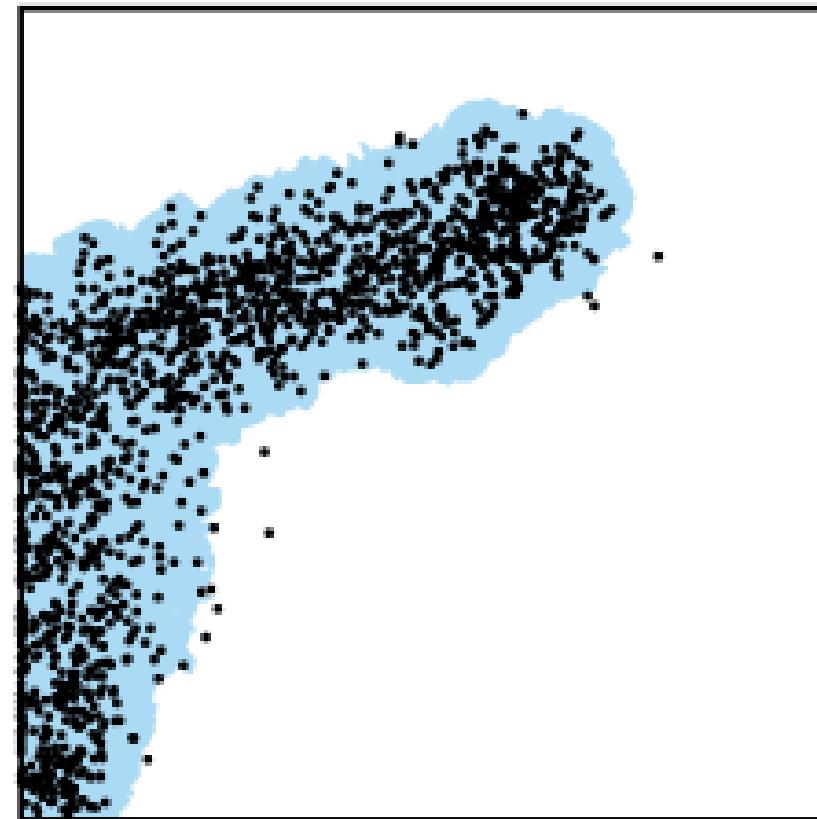


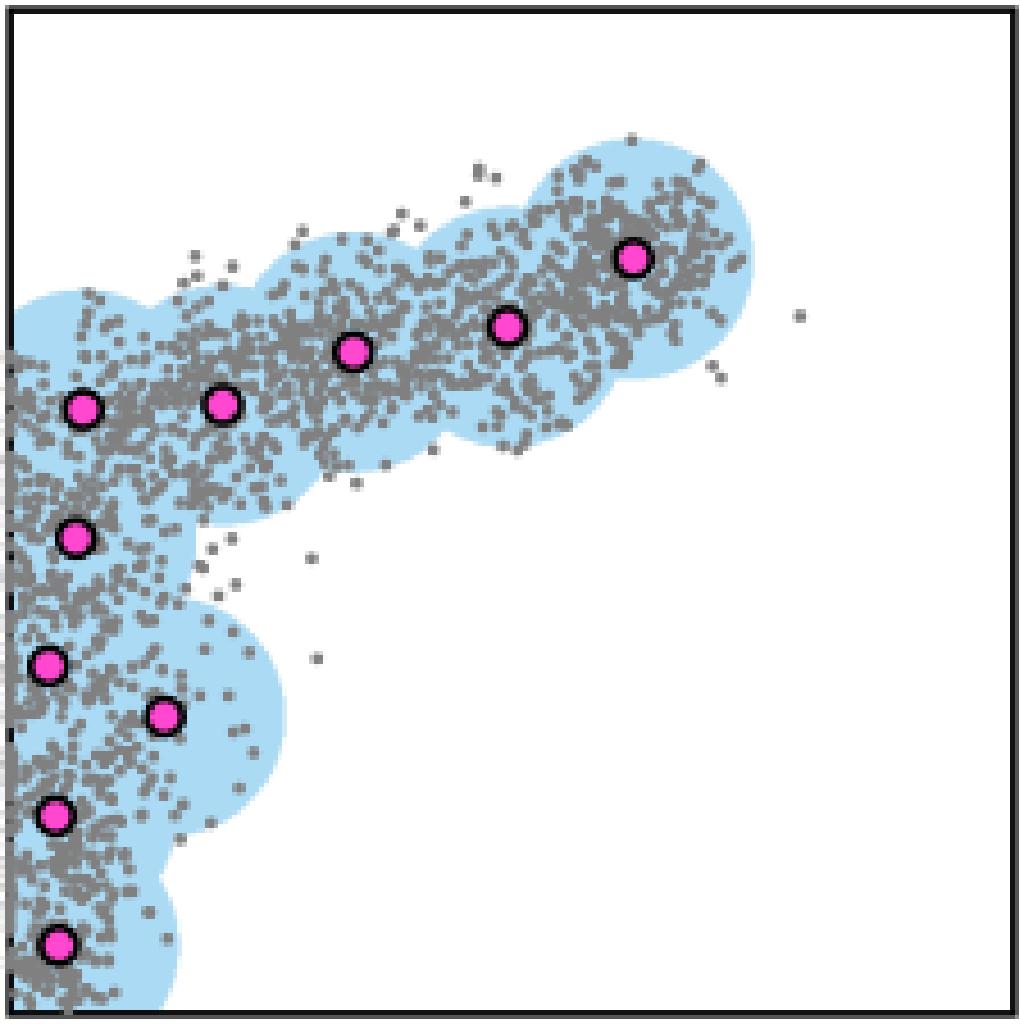
Taxiway A

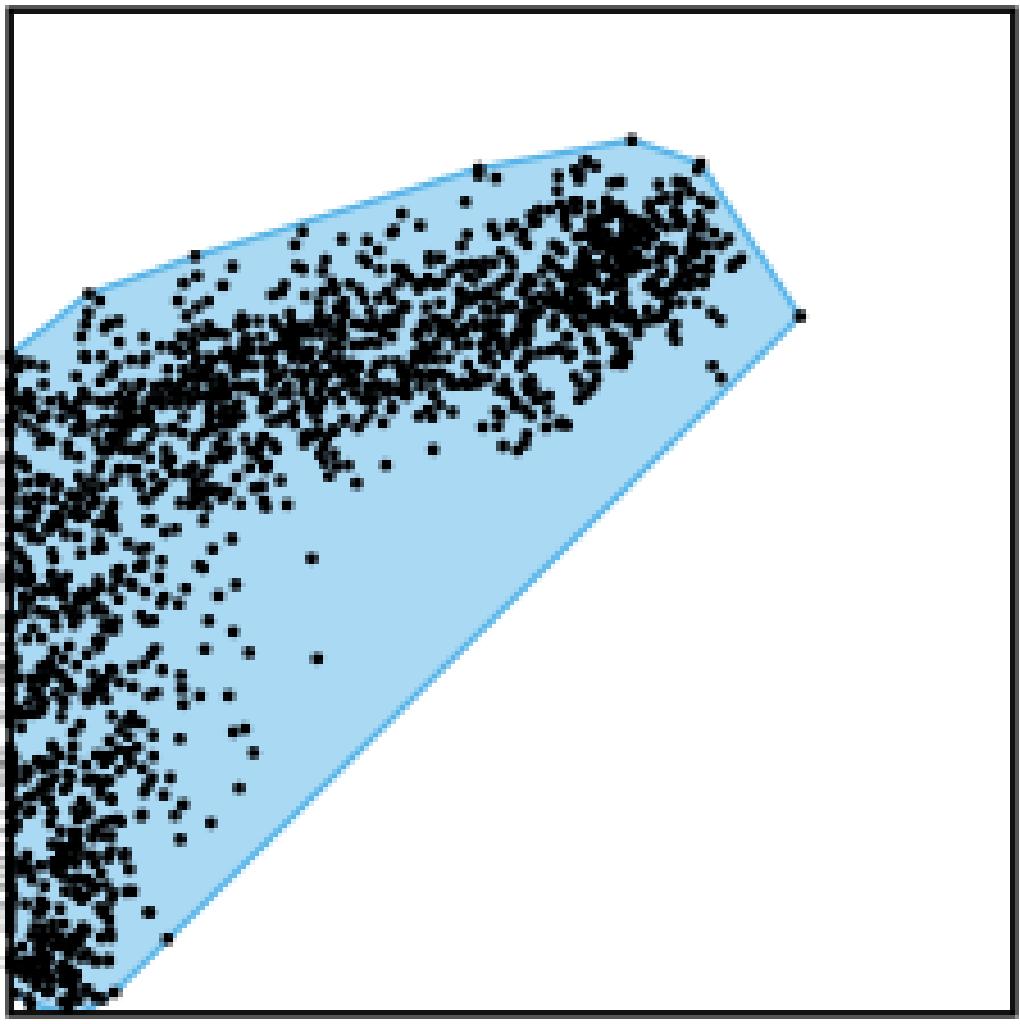




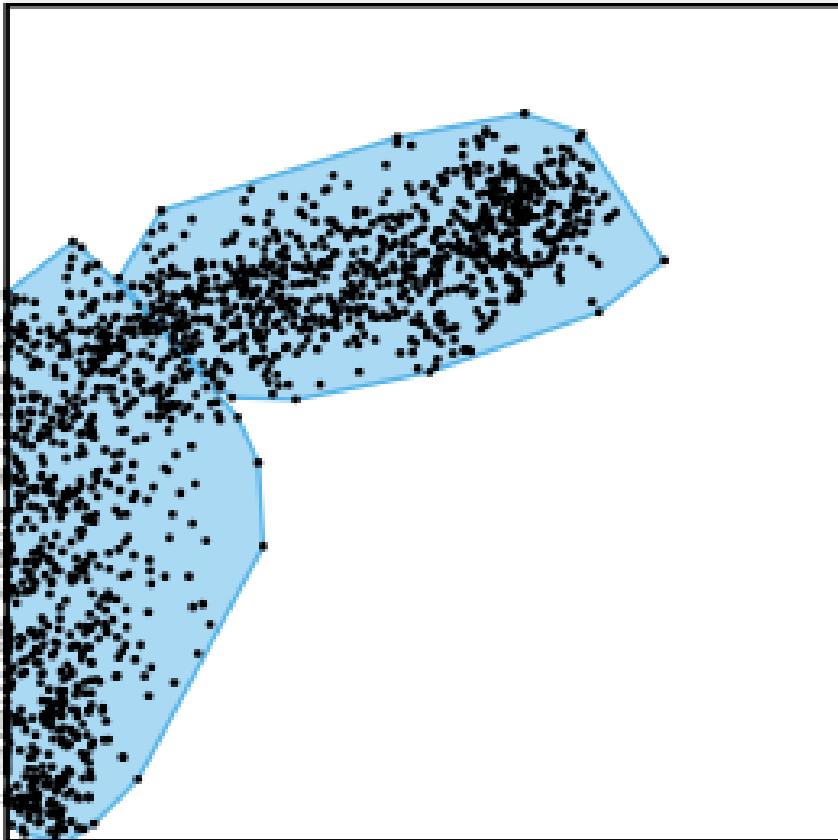
$\gamma = 0.1$  $\gamma = 0.3$  $\gamma = 0.5$ 

$k = 1$  $k = 2$  $k = 5$ 





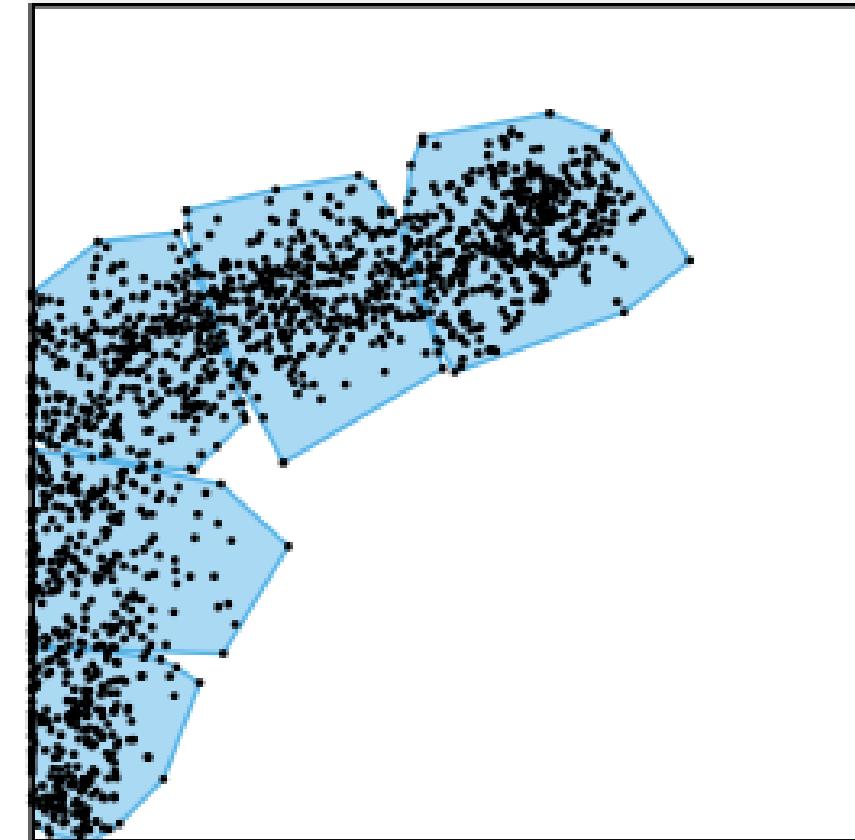
2 Clusters



3 Clusters



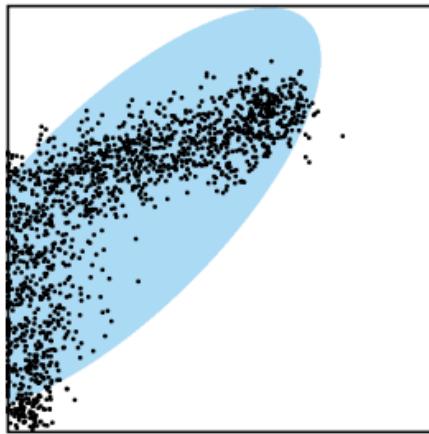
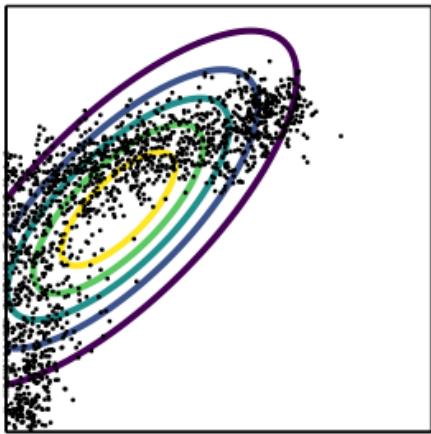
5 Clusters



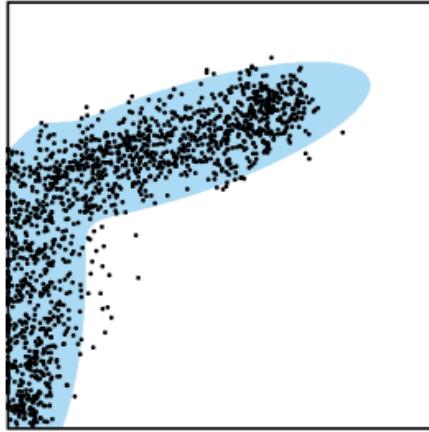
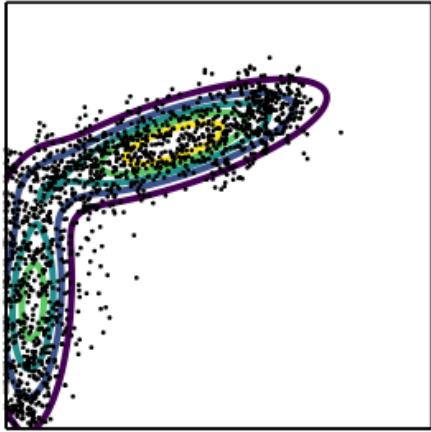
Distribution

Superlevel Set

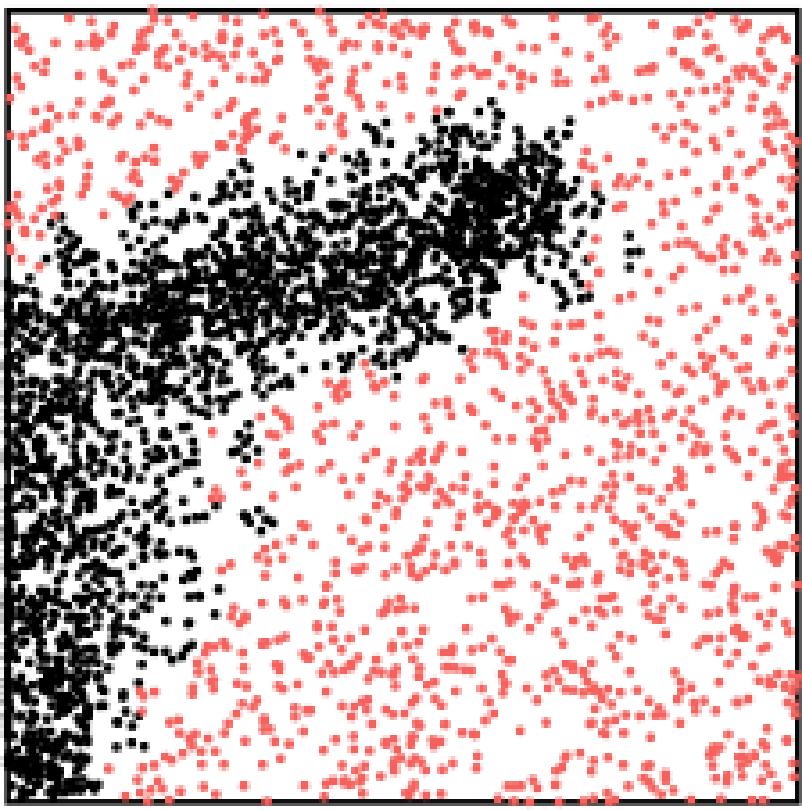
Gaussian



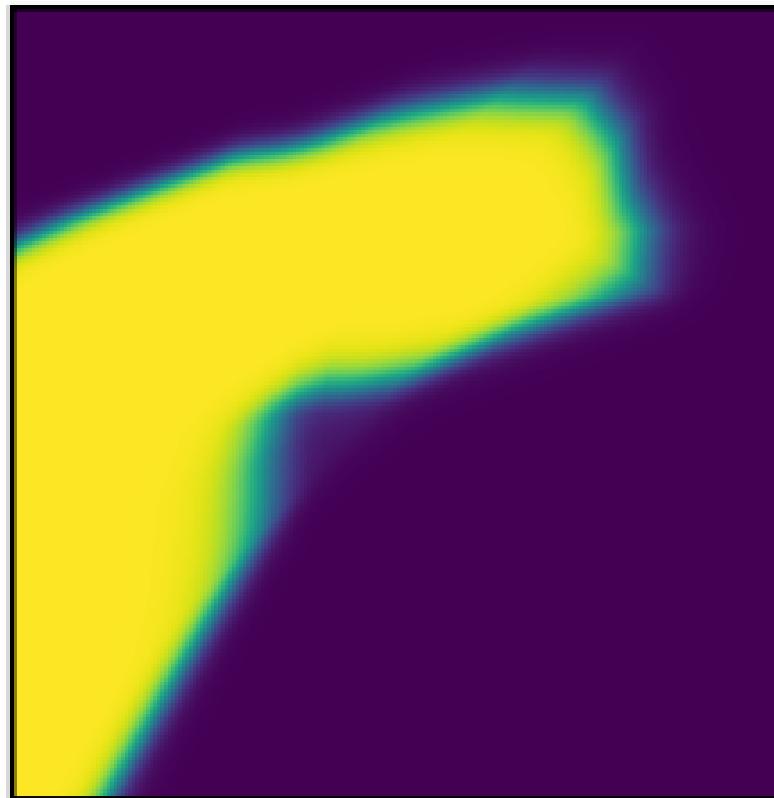
Gaussian Mixture



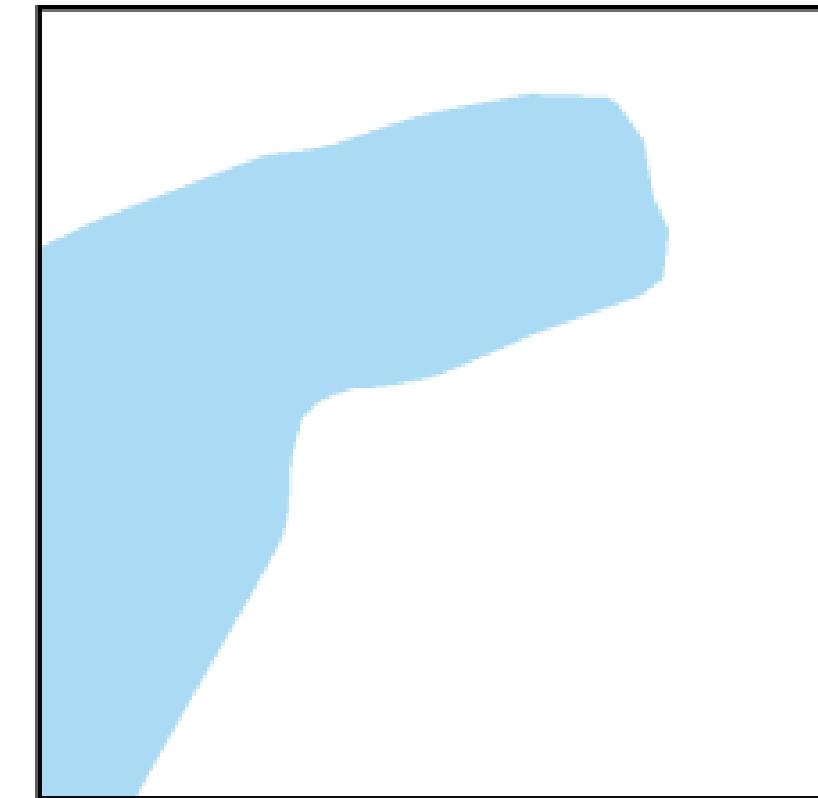
Training Data

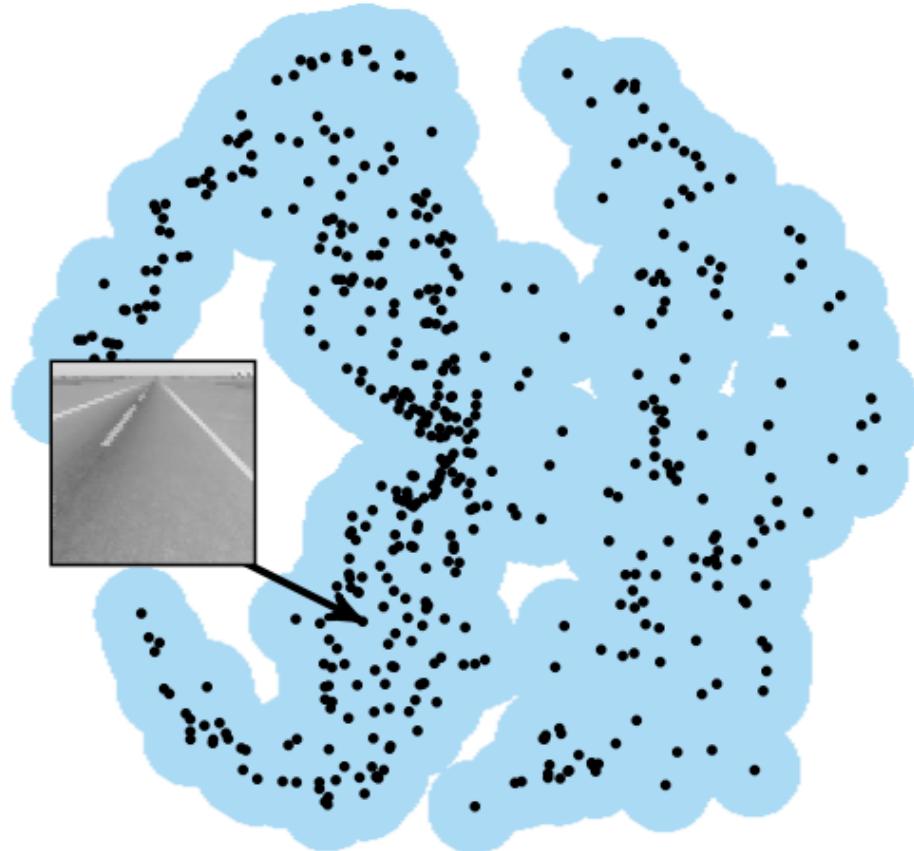


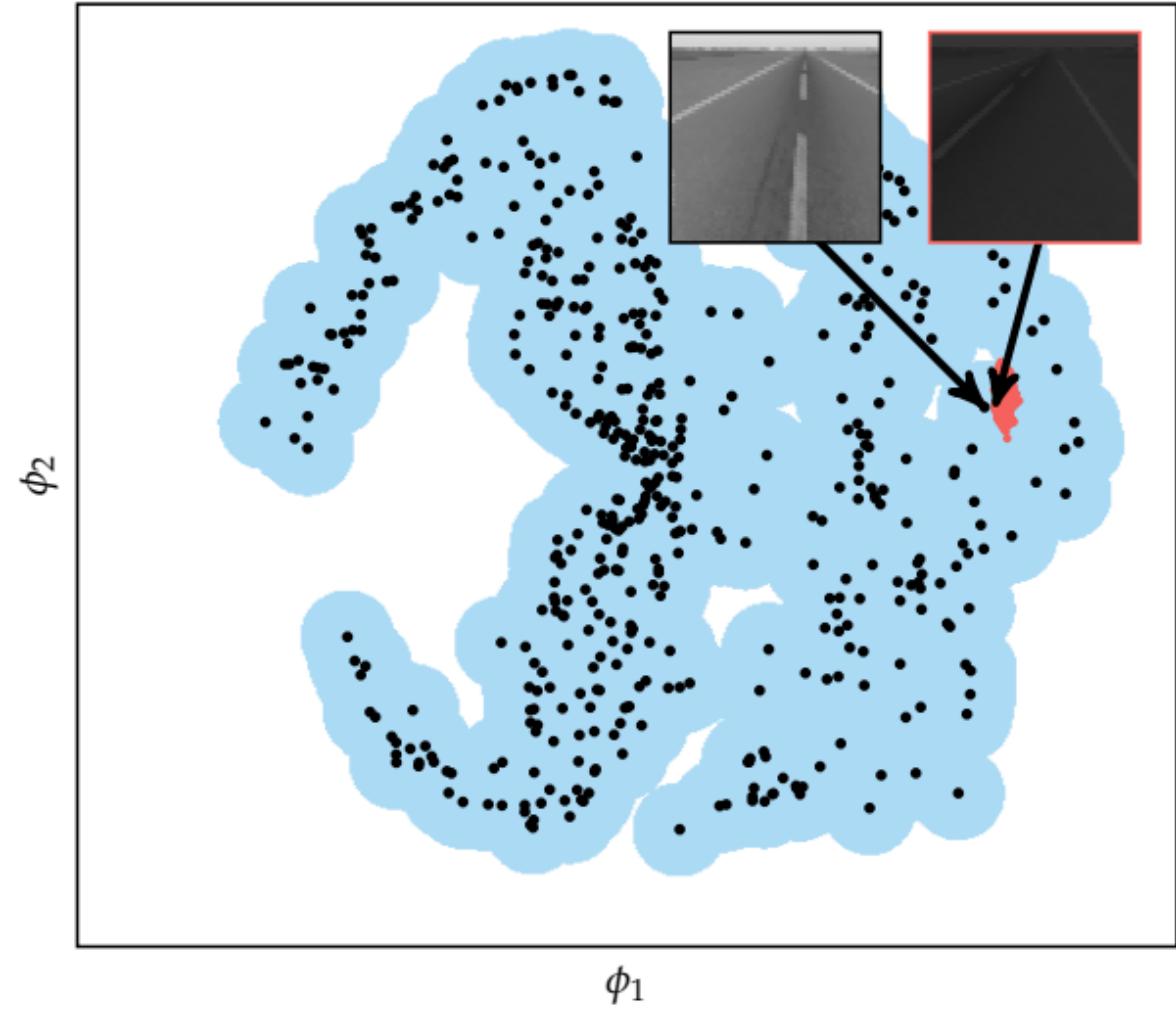
Classifier Probability



Superlevel Set



$\phi_2$  $\phi_1$



Low  
Output  
Uncertainty

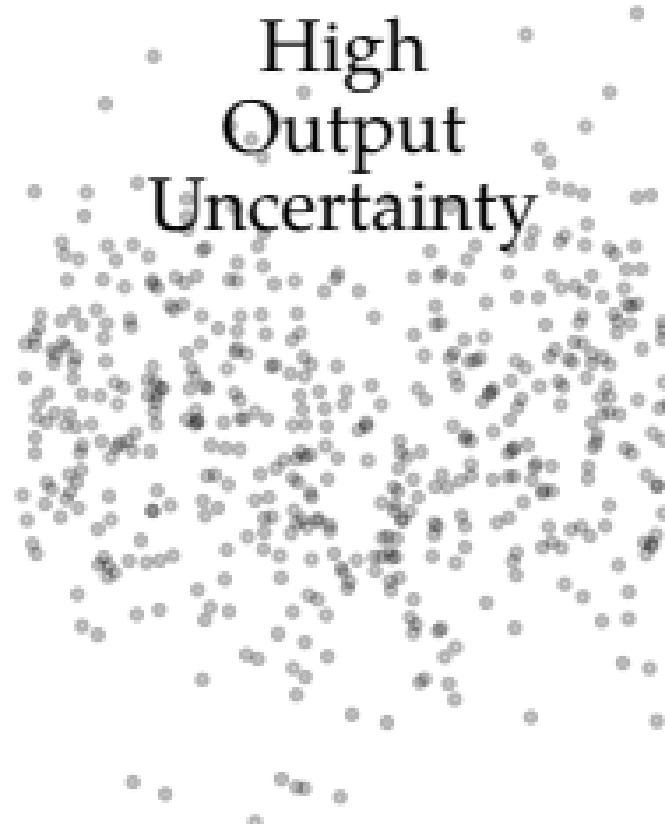
*y*

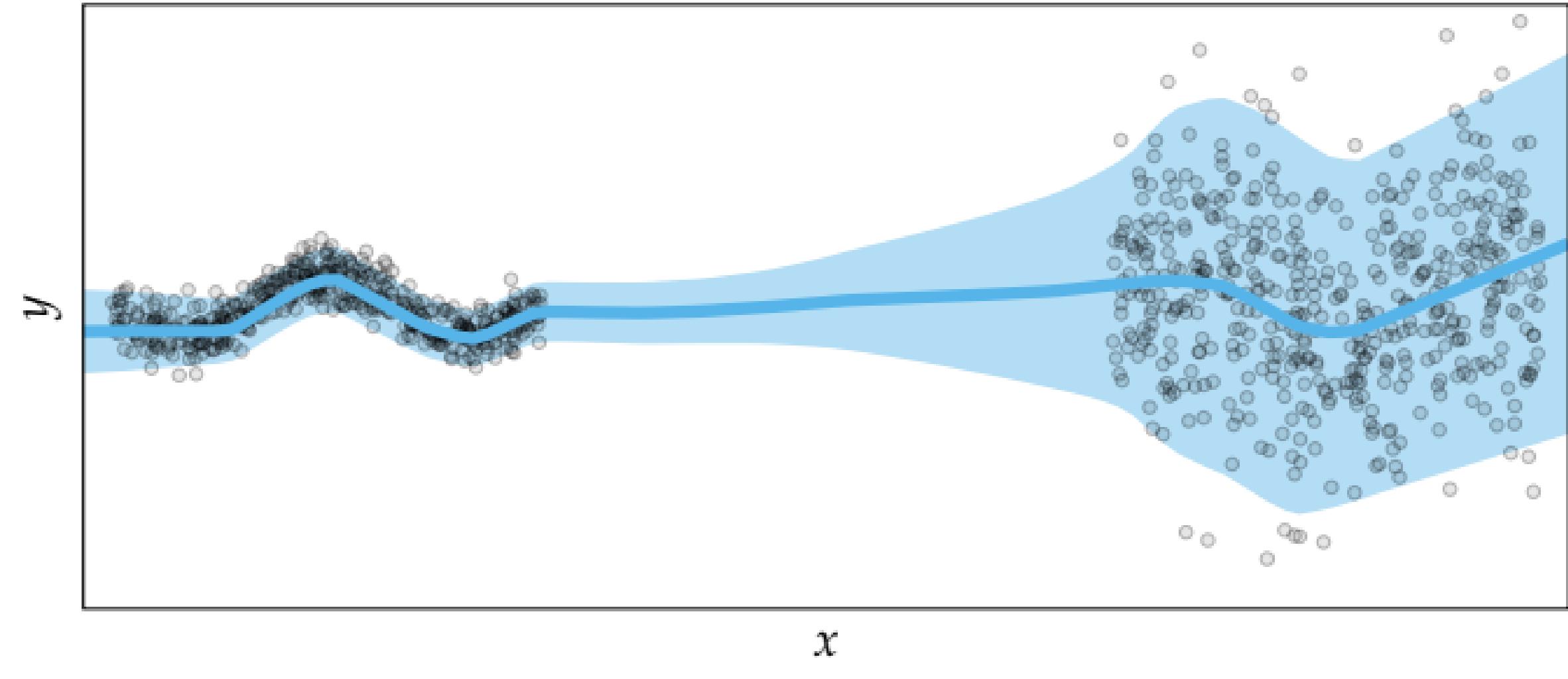


High  
Model  
Uncertainty

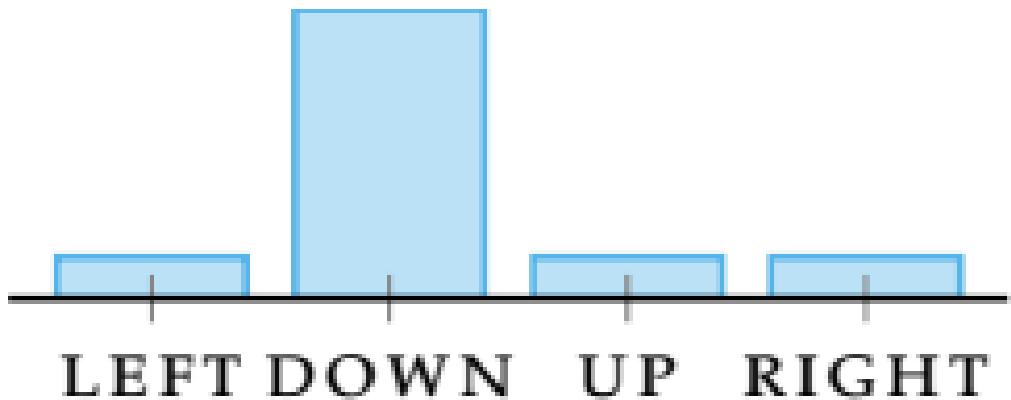
*x*

High  
Output  
Uncertainty

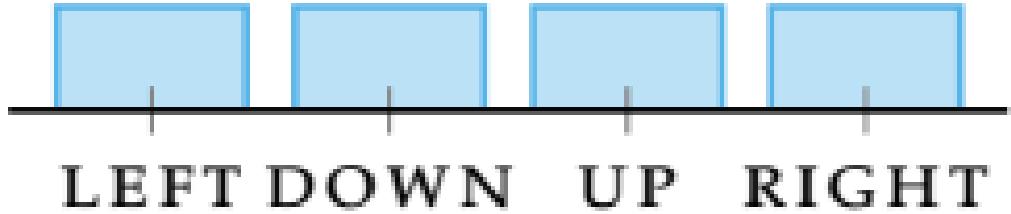


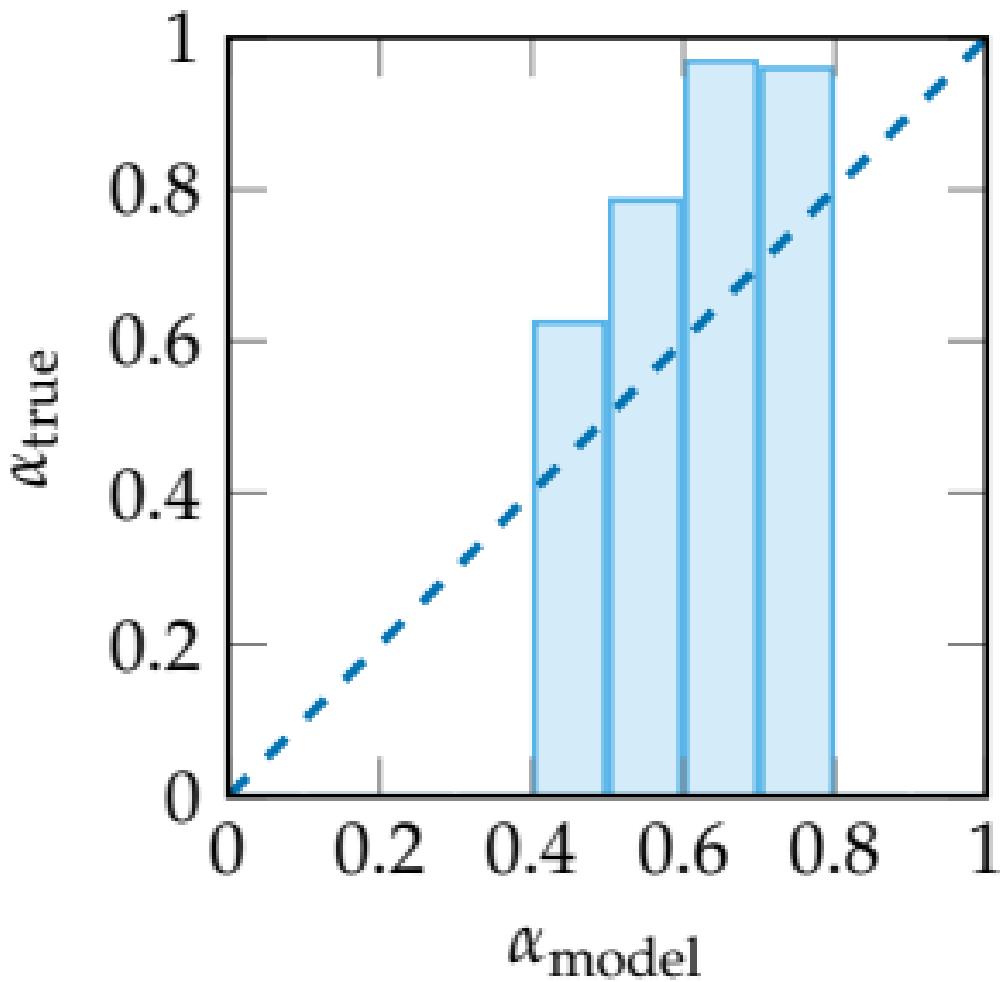


# Low Entropy

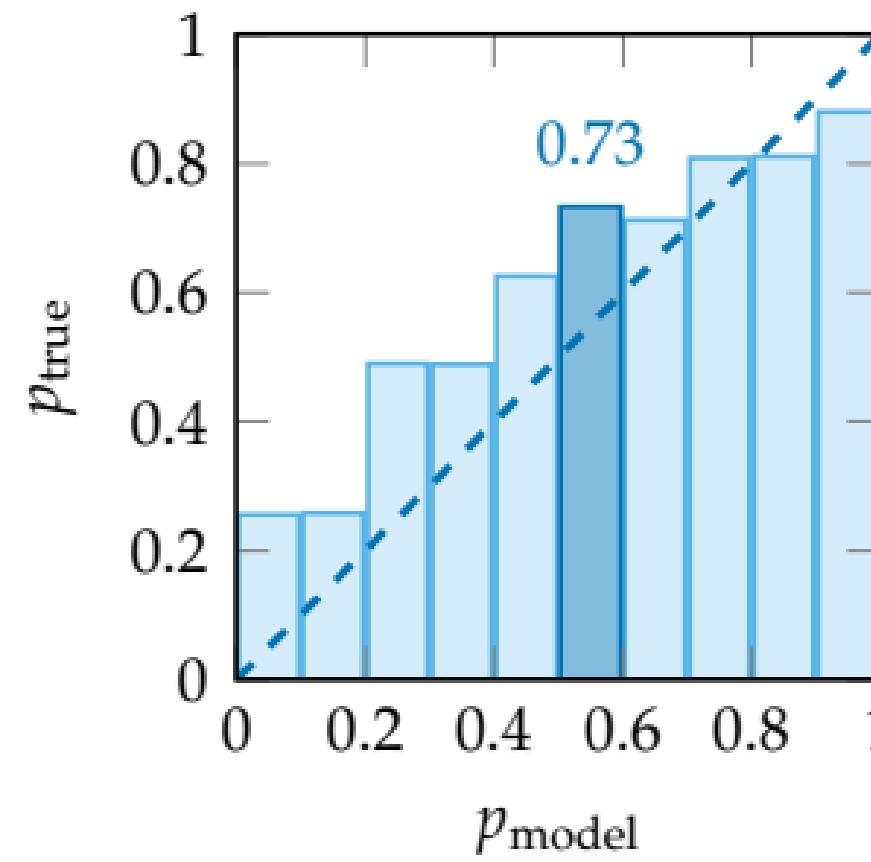


# High Entropy

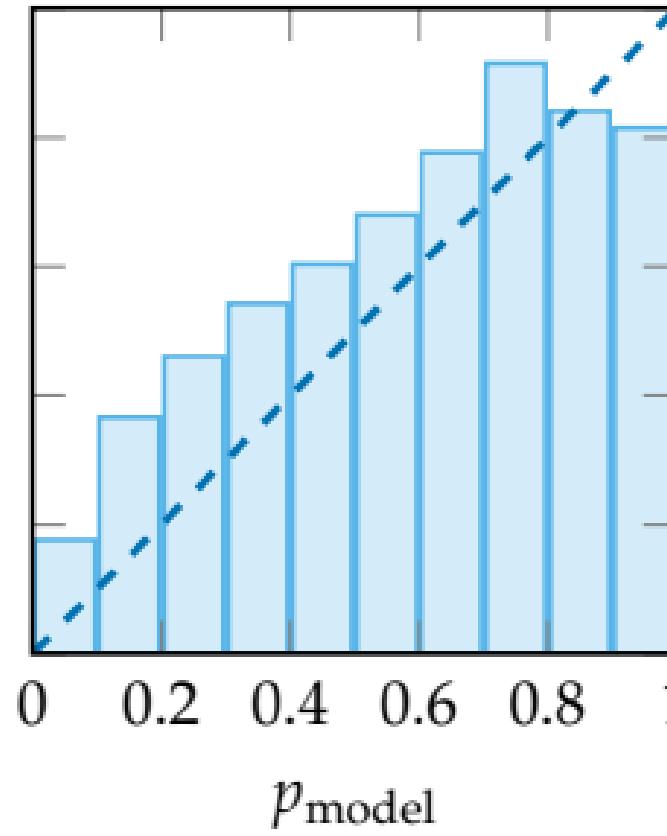




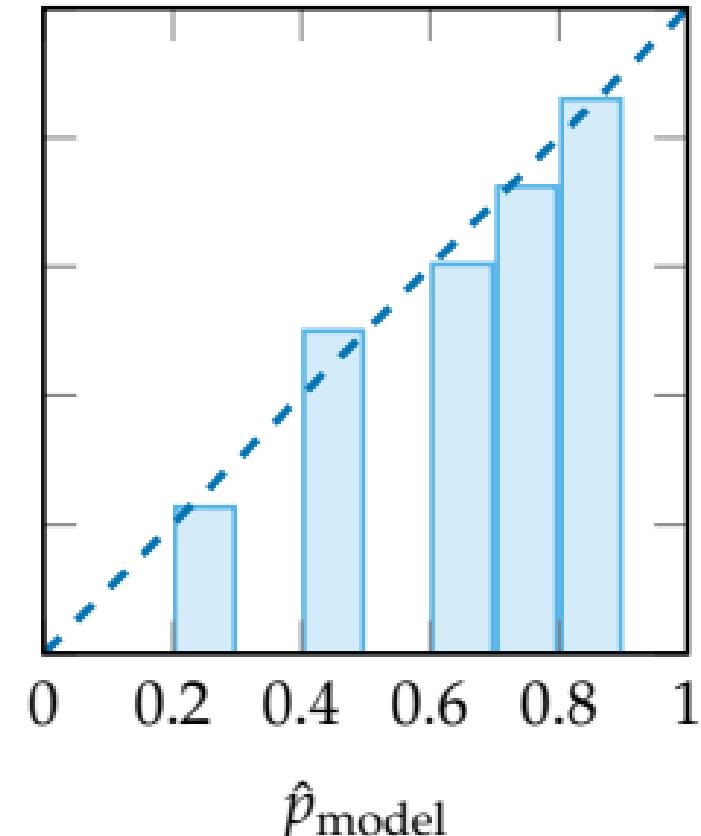
# Calibration Data

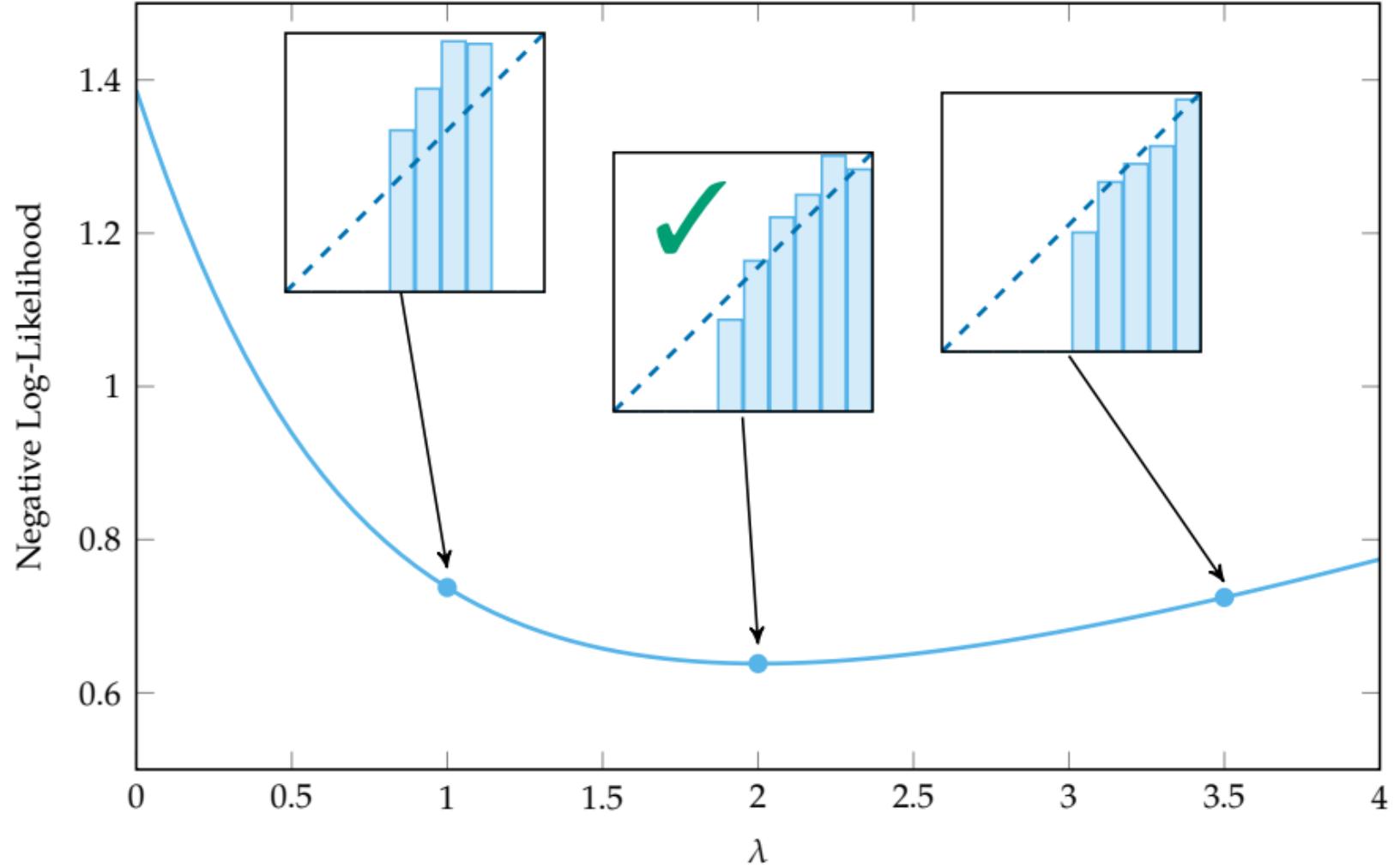


# Runtime Data

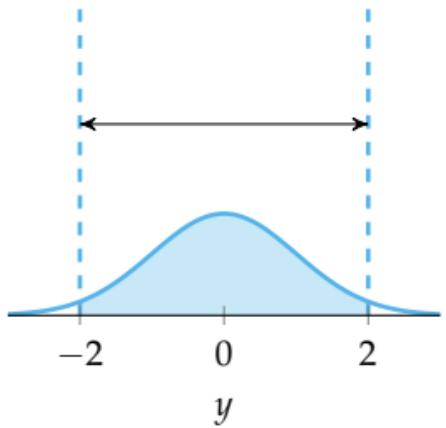


# Calibrated Data

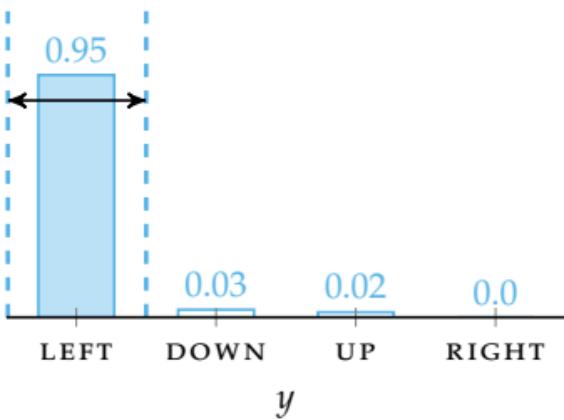
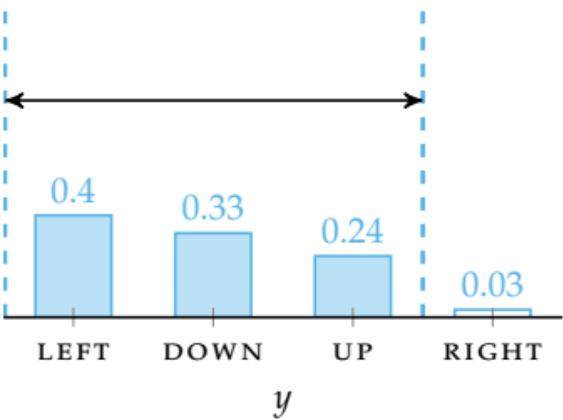
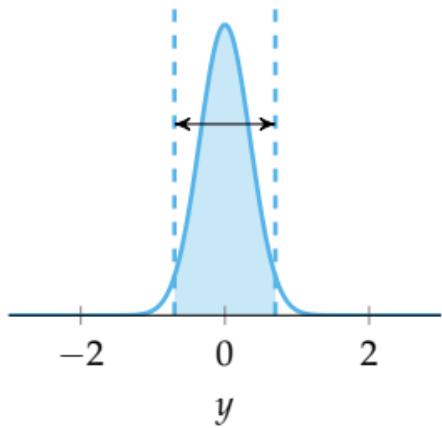




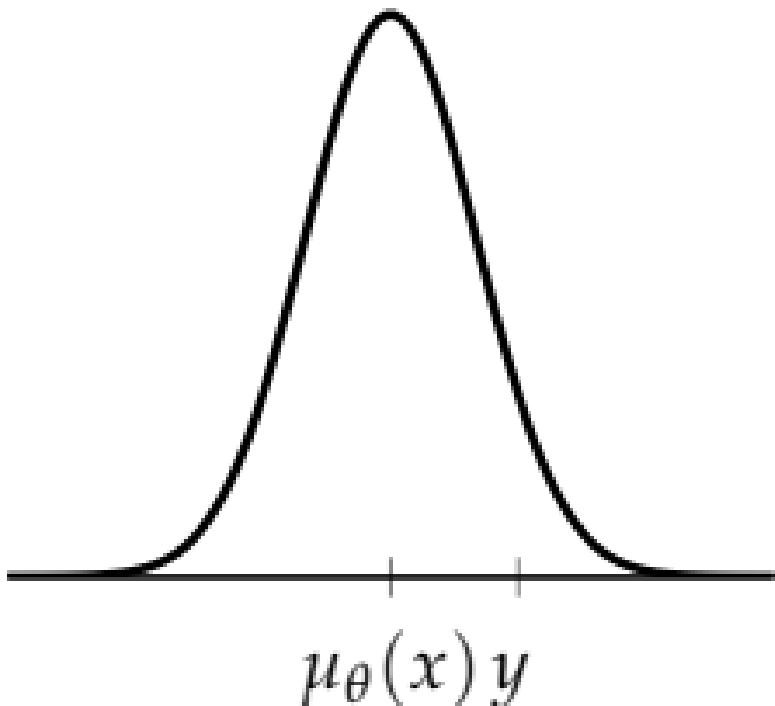
# High Uncertainty



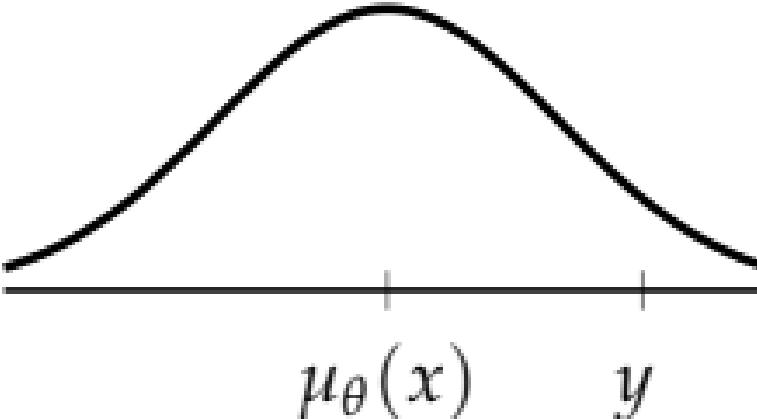
# Low Uncertainty



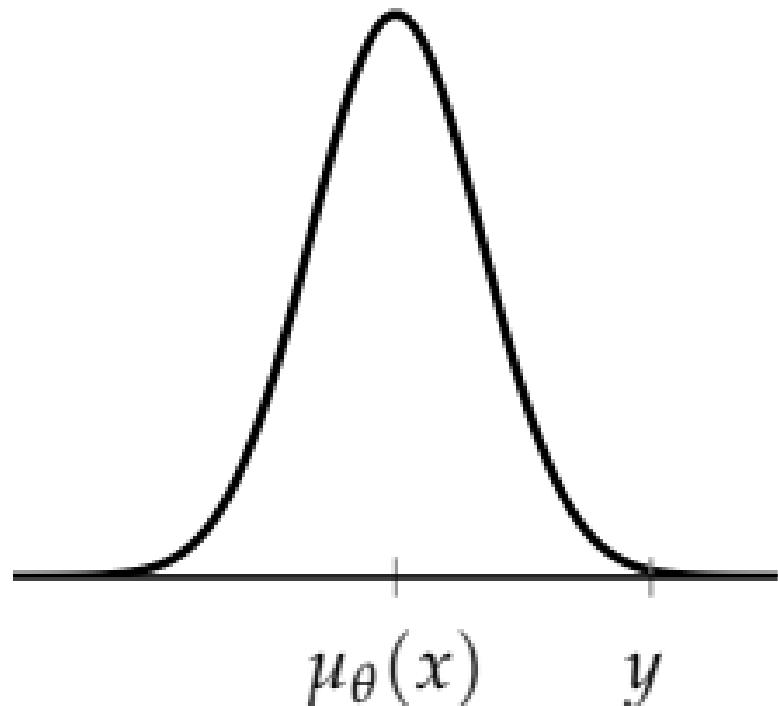
$$s(x, y) = 1.5$$



$$s(x, y) = 1.5$$



$$s(x, y) = 3.0$$



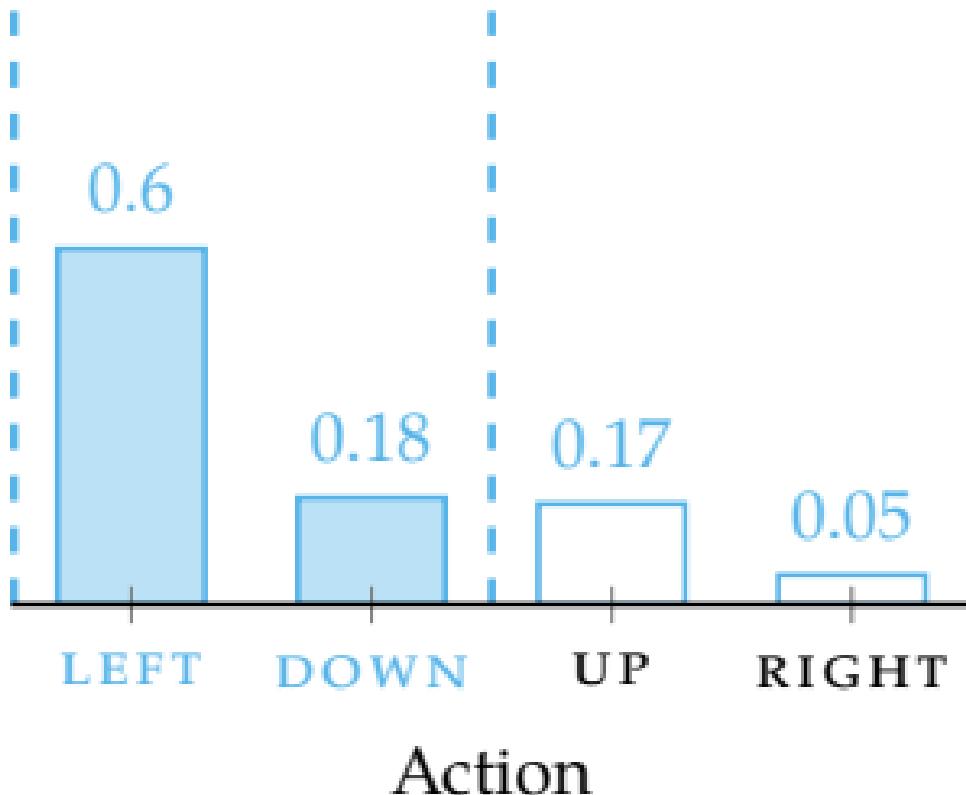
# Predicted Probabilities



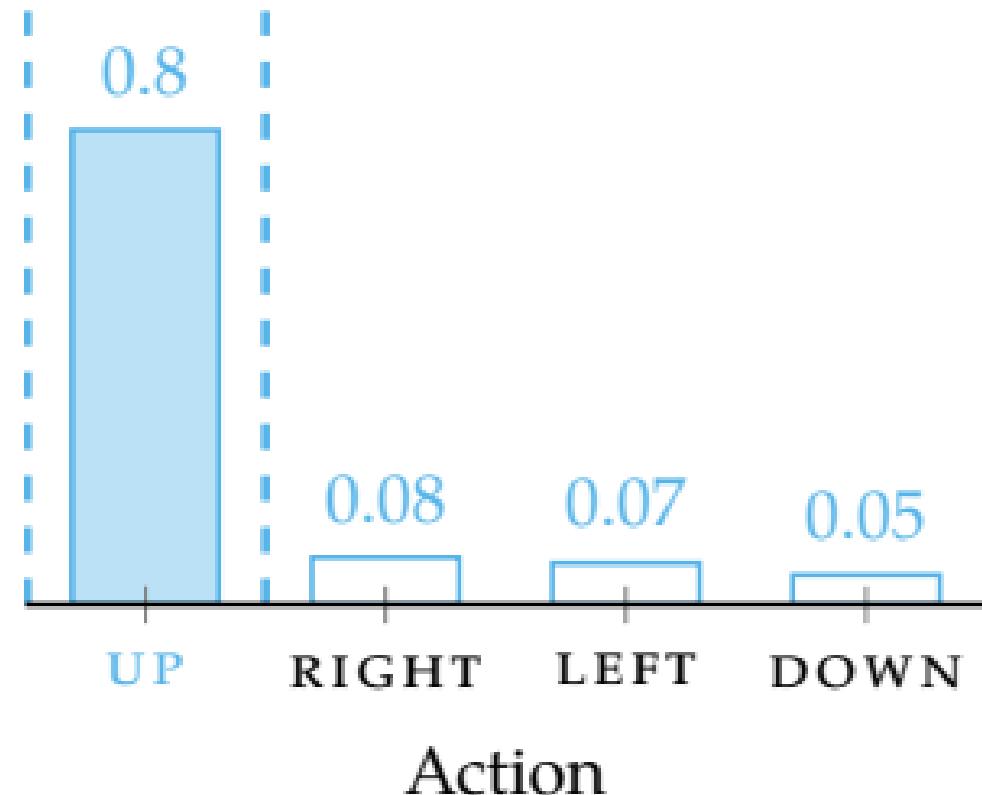
# Sorted Probabilities



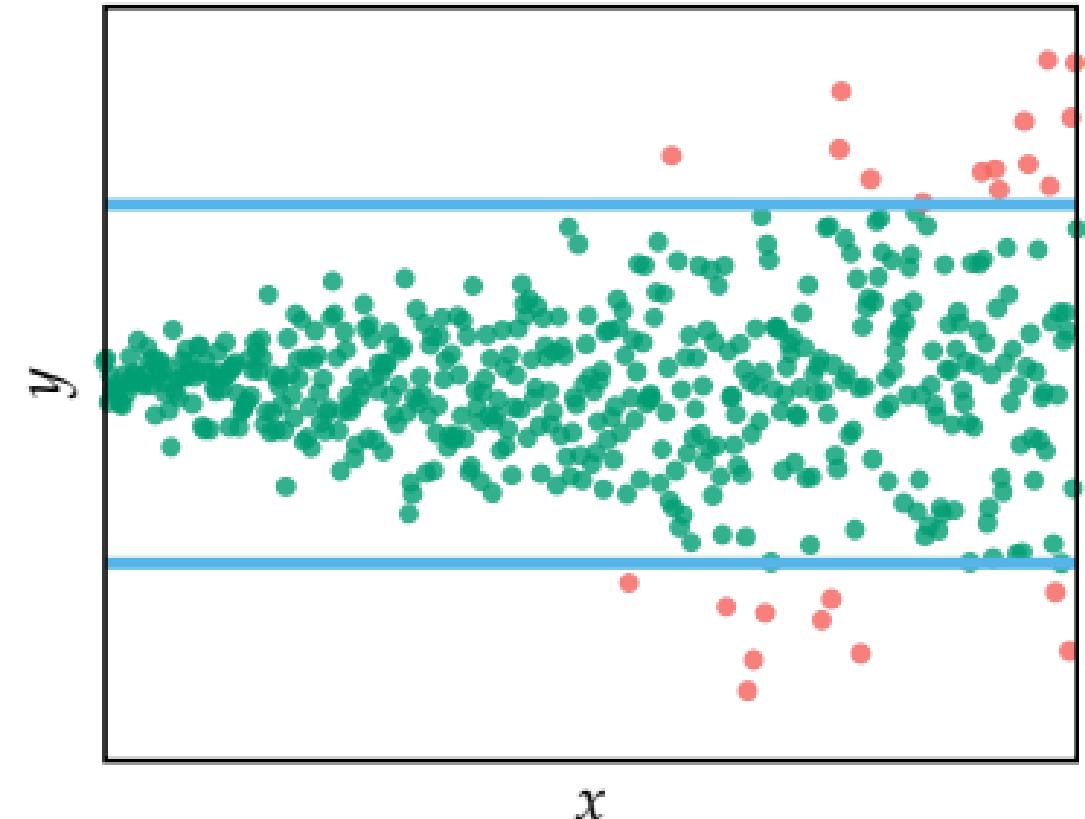
Prediction Set = {LEFT, DOWN}



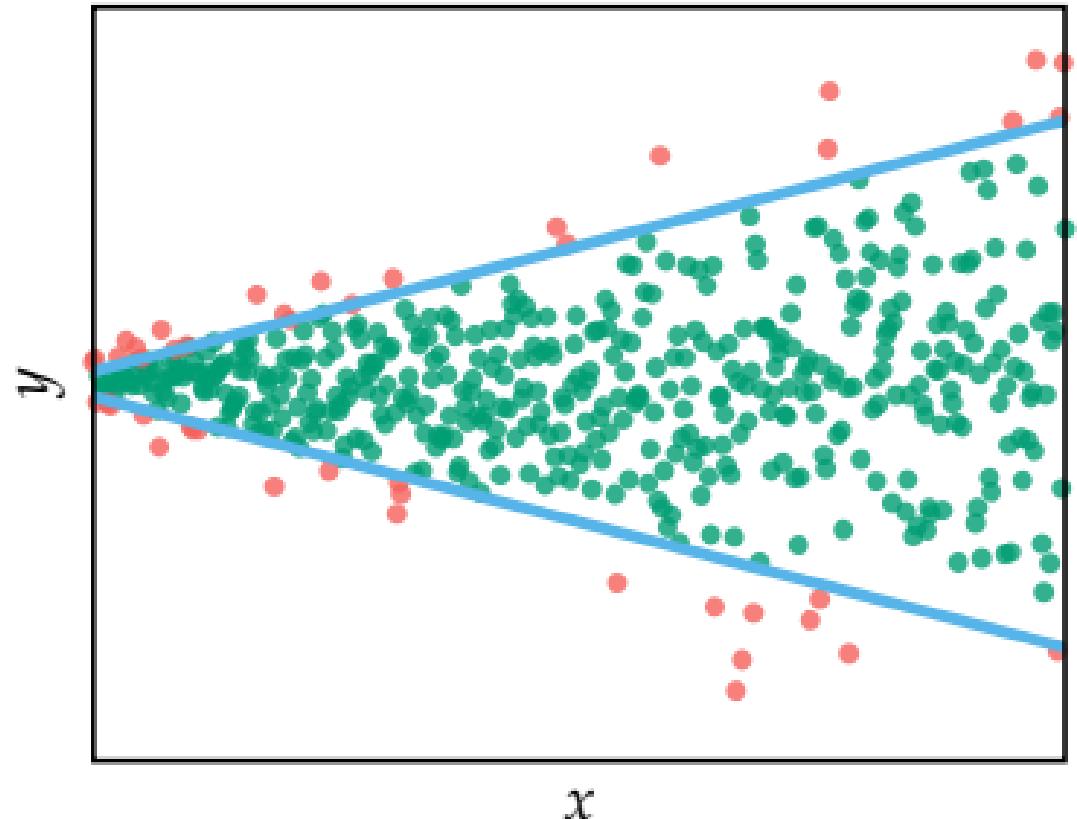
Prediction Set = {UP}

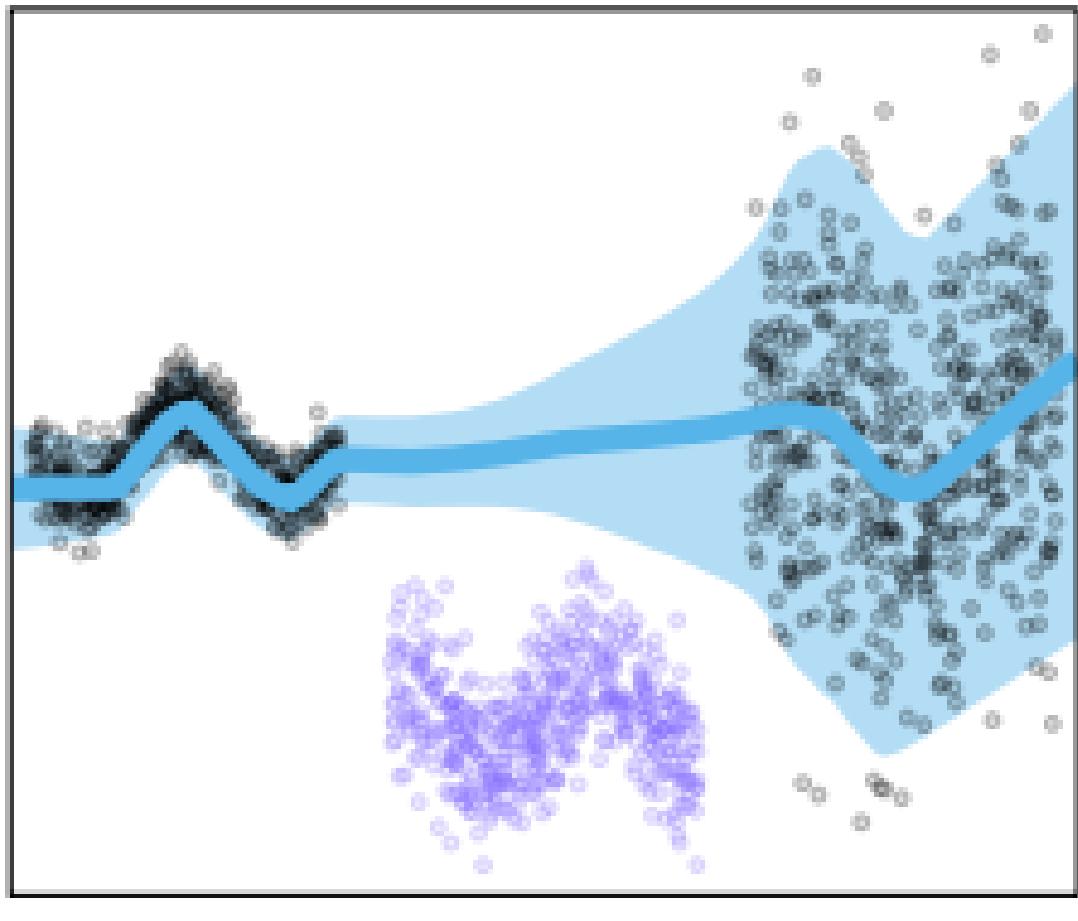


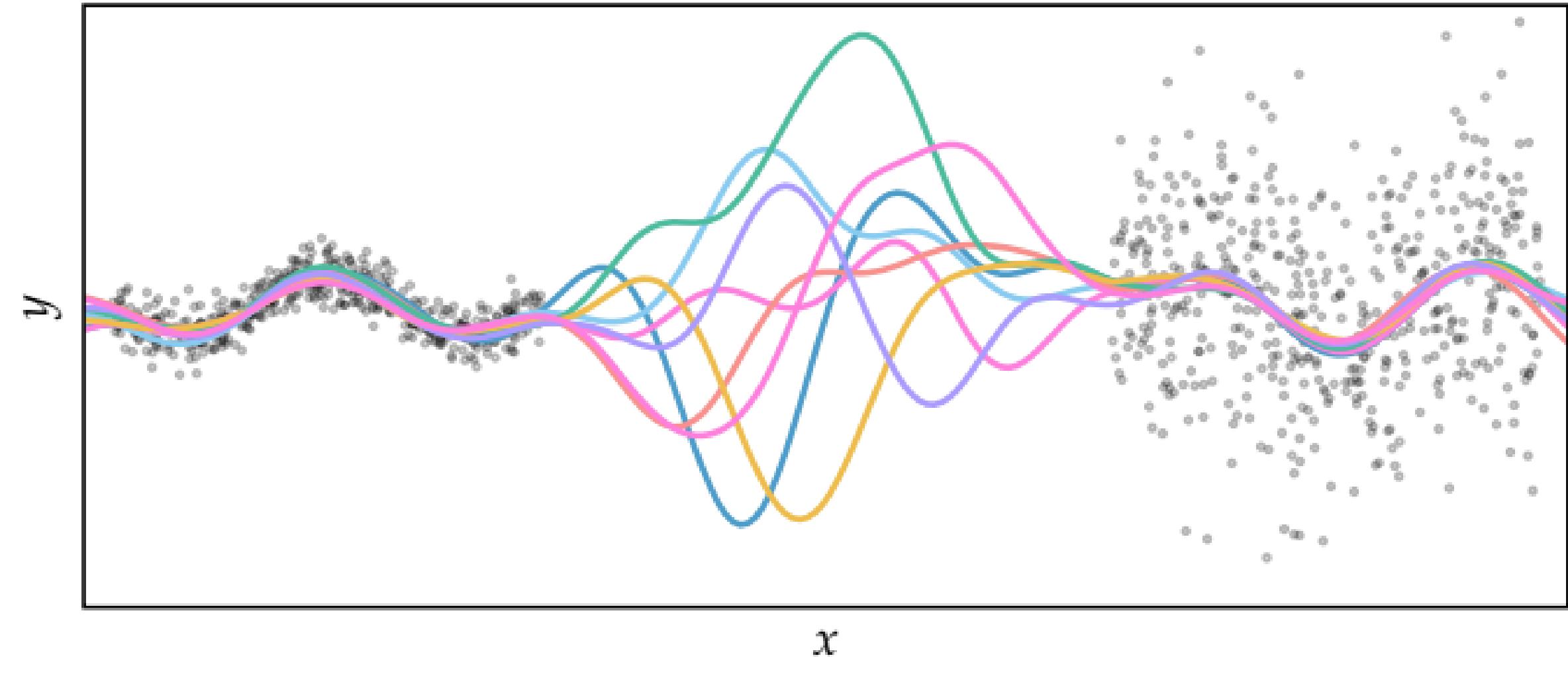
# Marginal Coverage



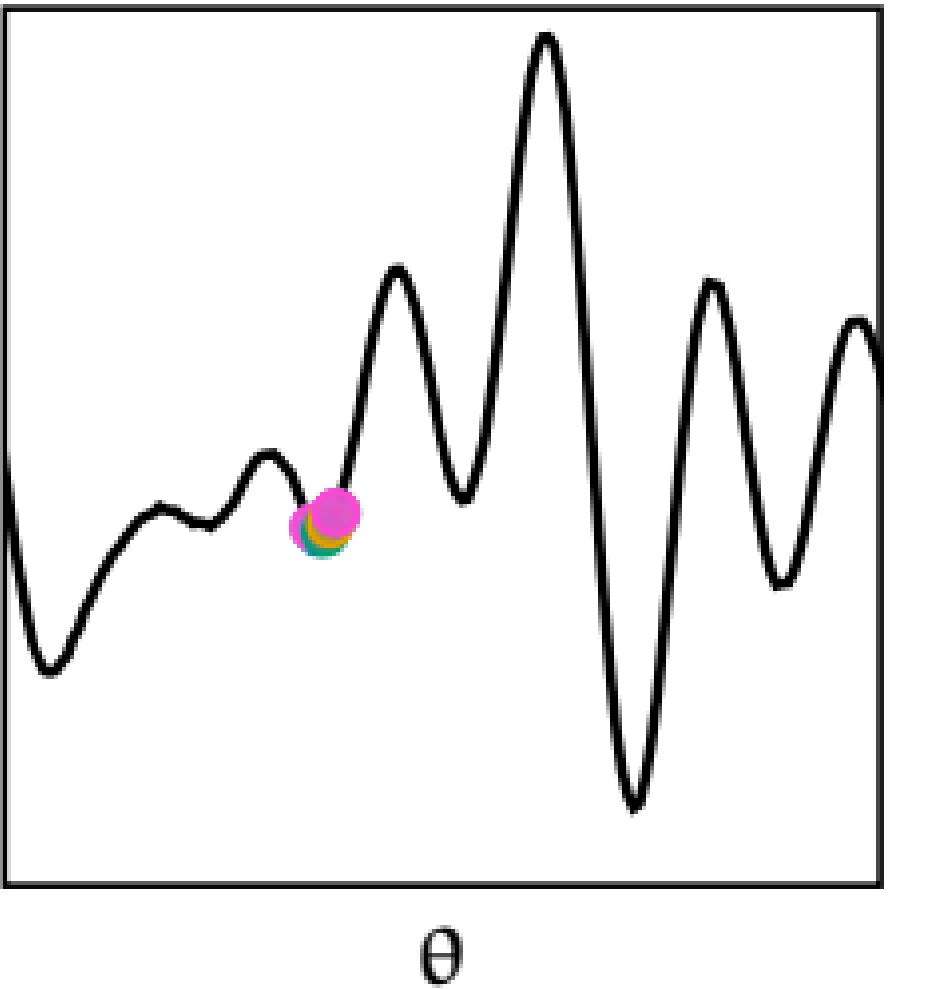
# Conditional Coverage



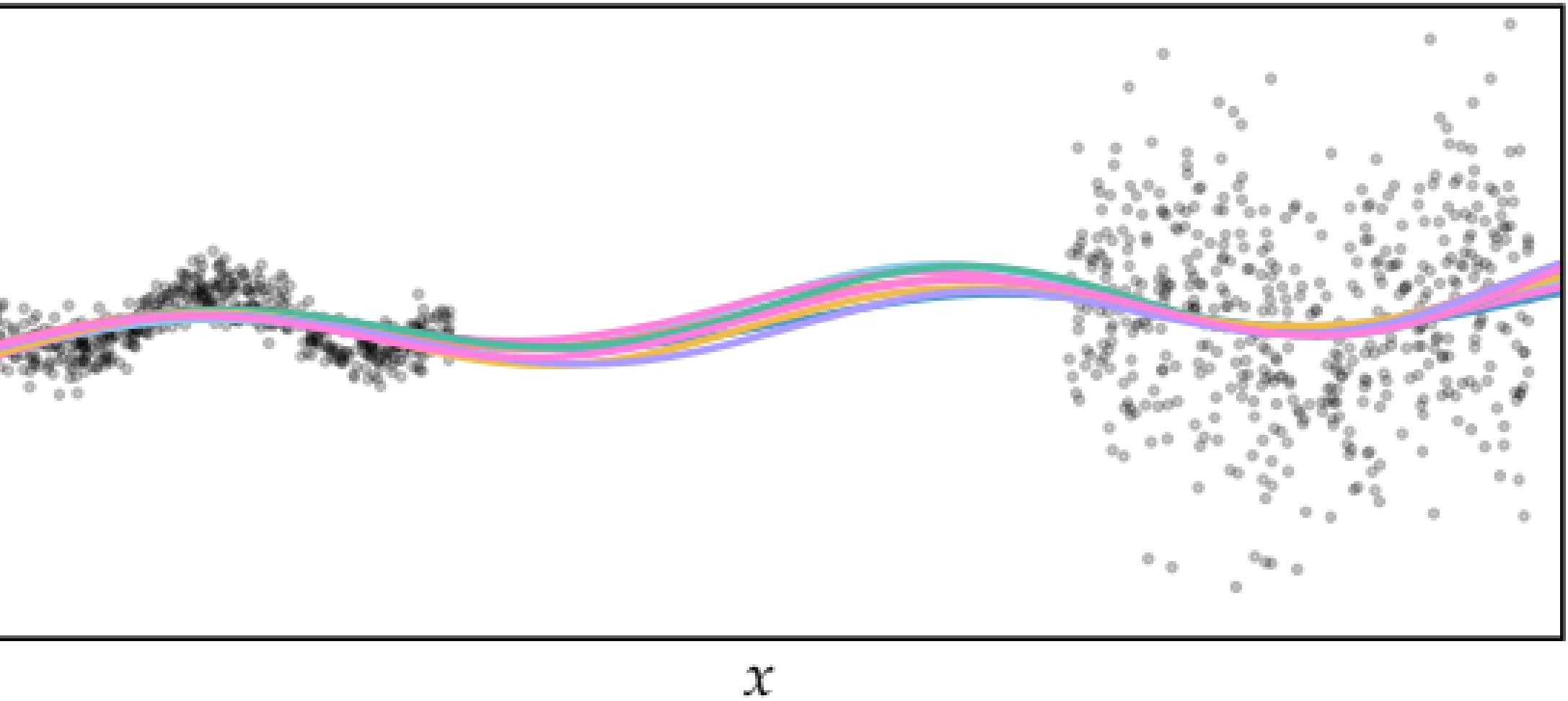
$y$  $x$ 



$$-\nu(\theta | D)$$



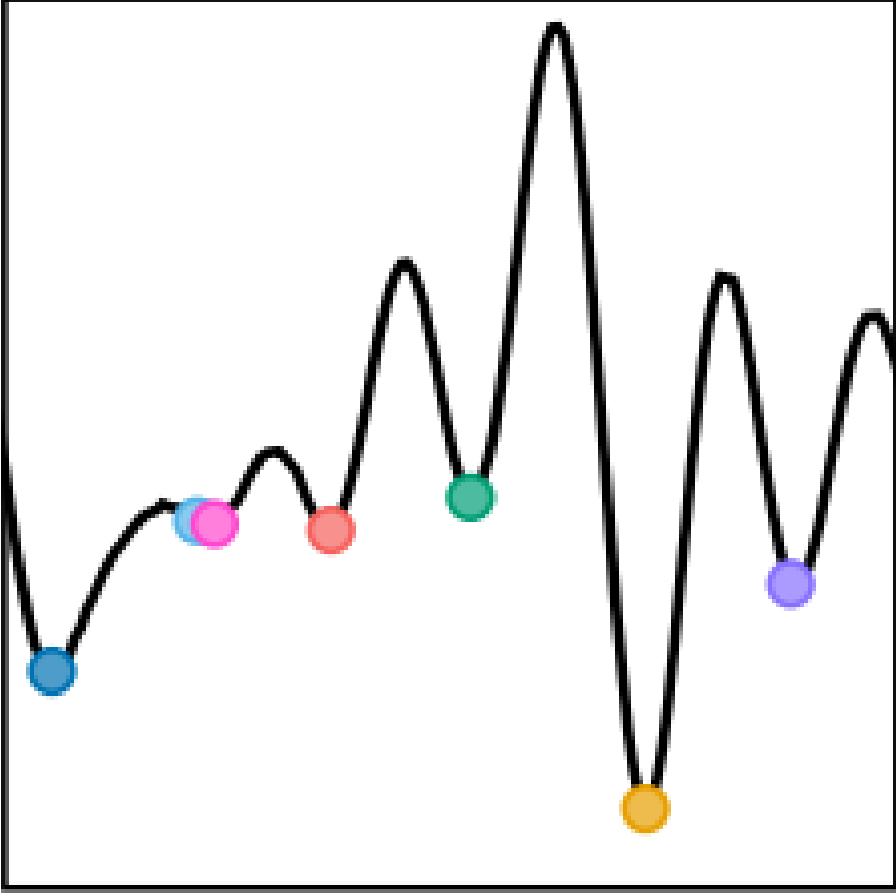
$y$

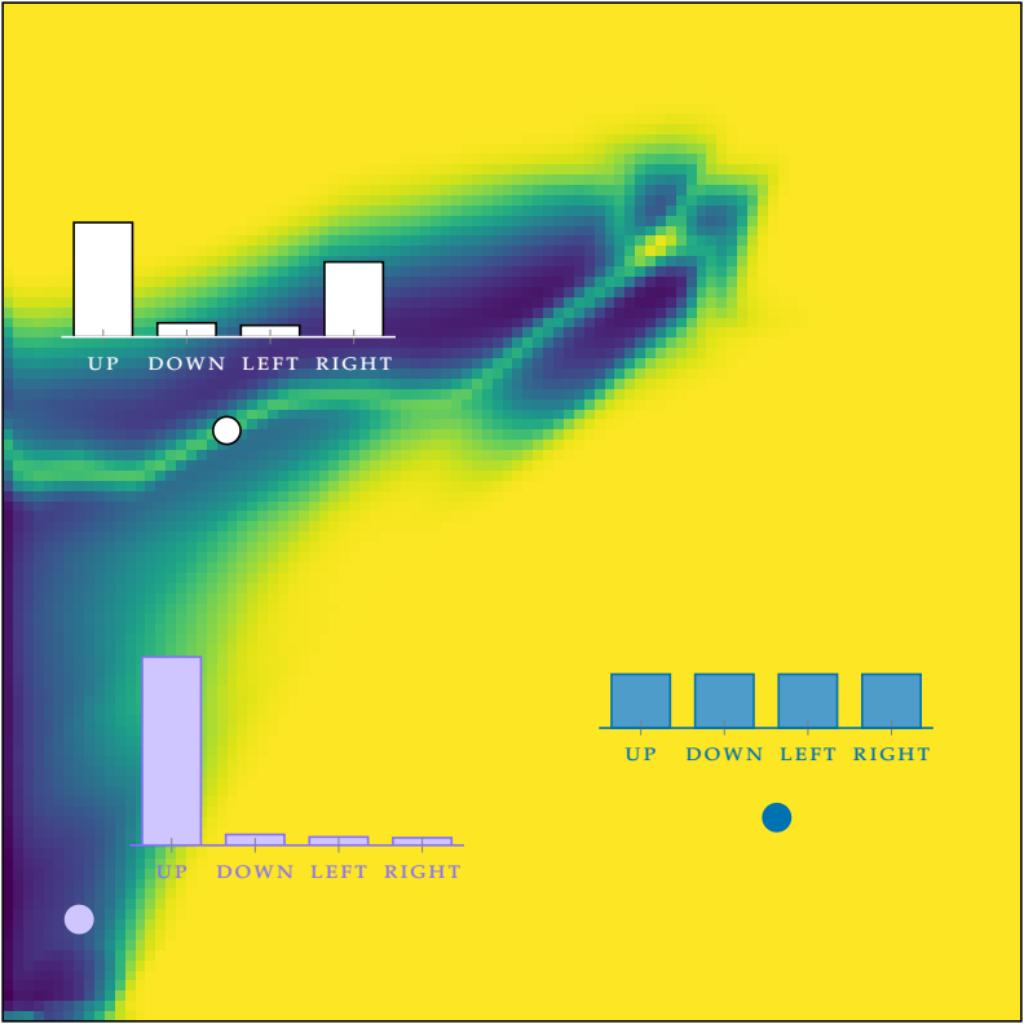


$x$

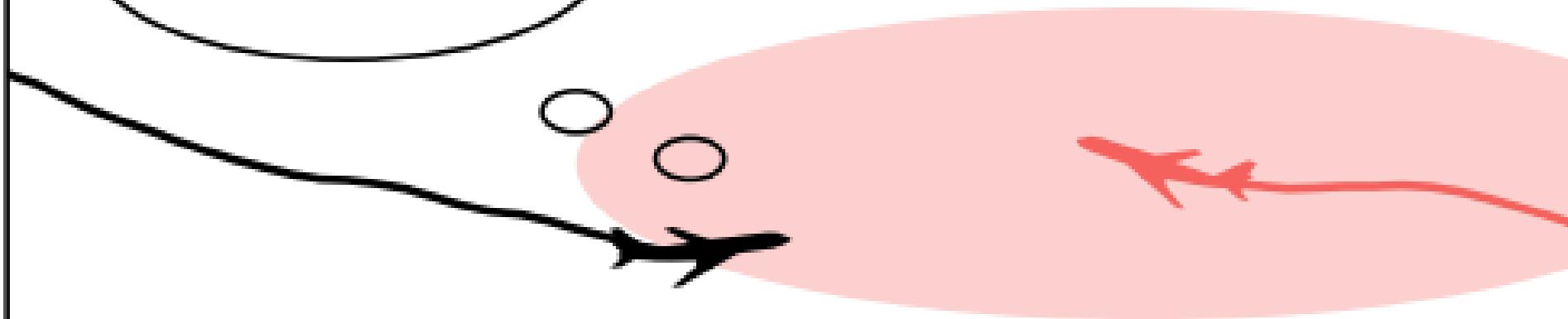
$$-p(\theta \mid D)$$

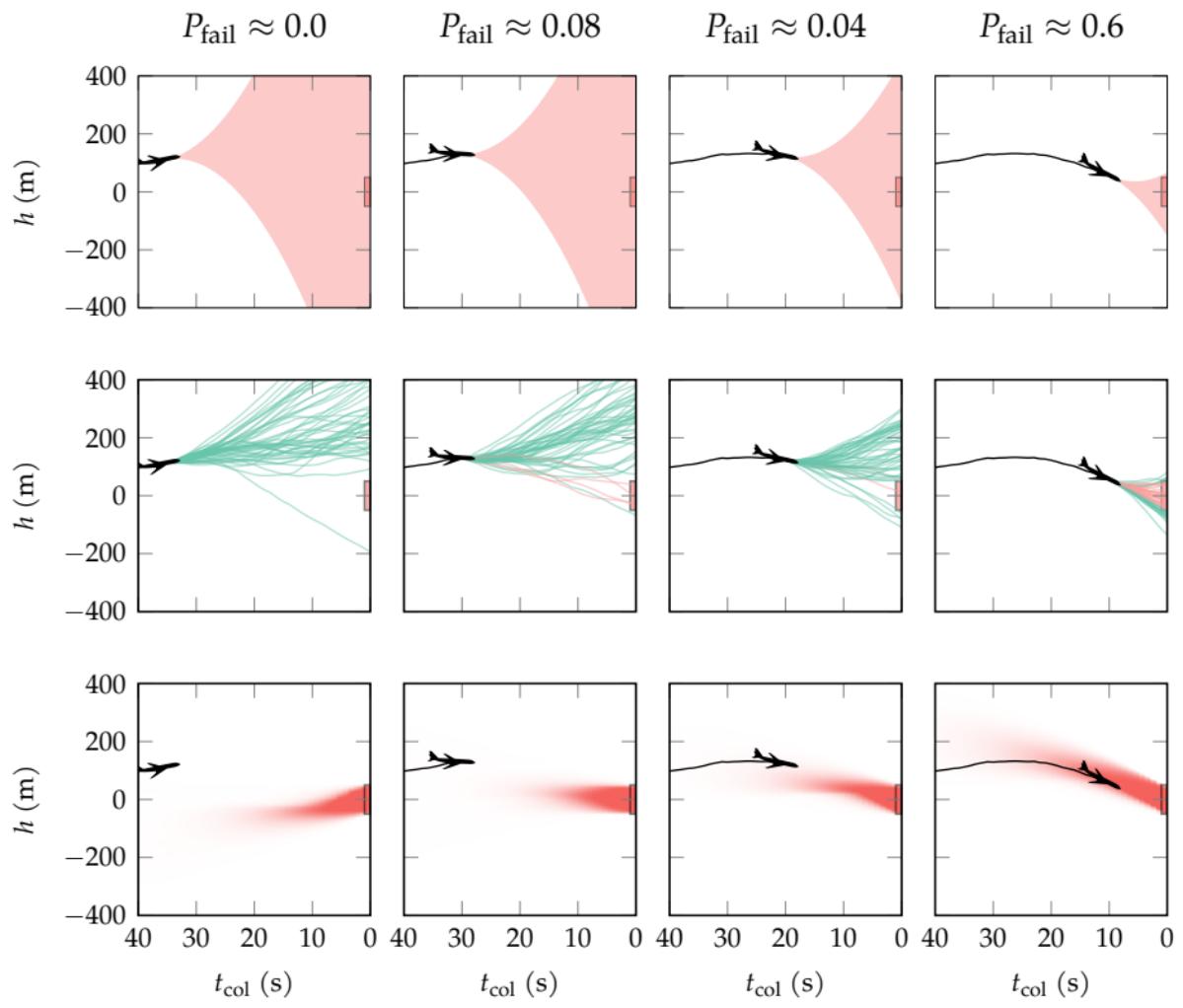
$\theta$

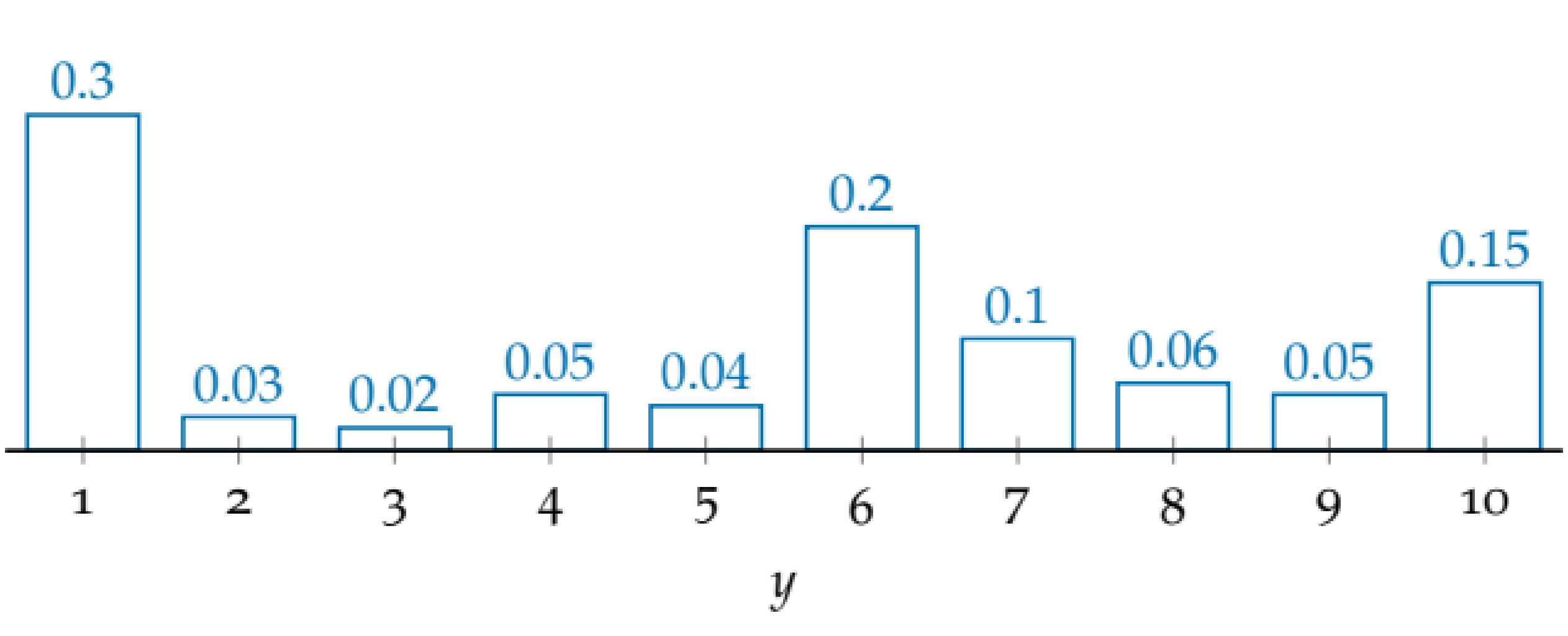


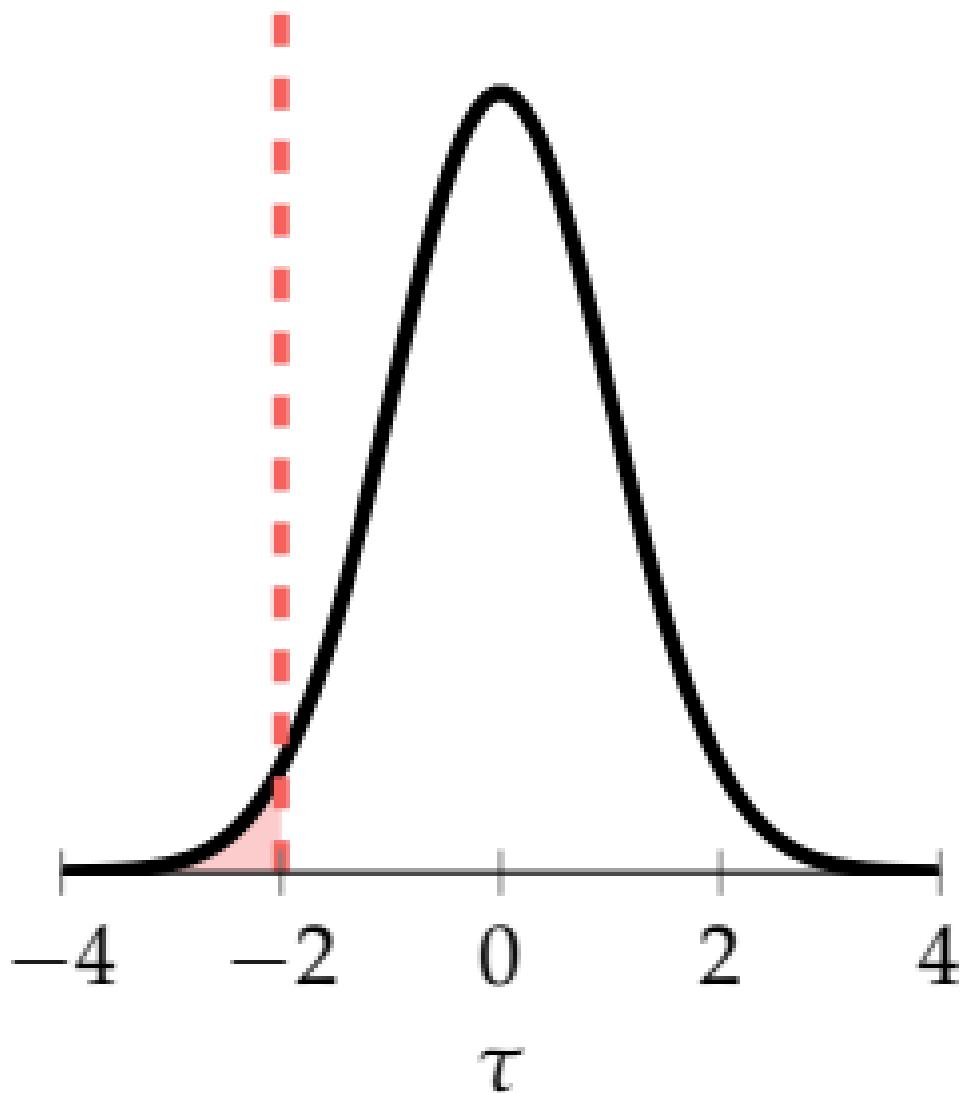


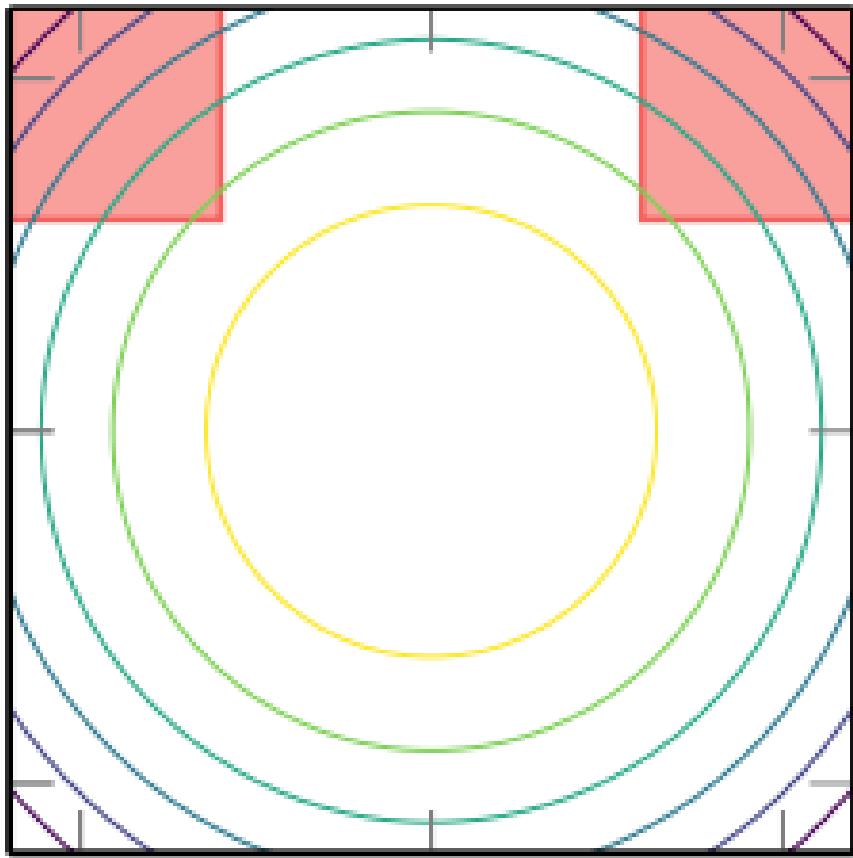
Warning!

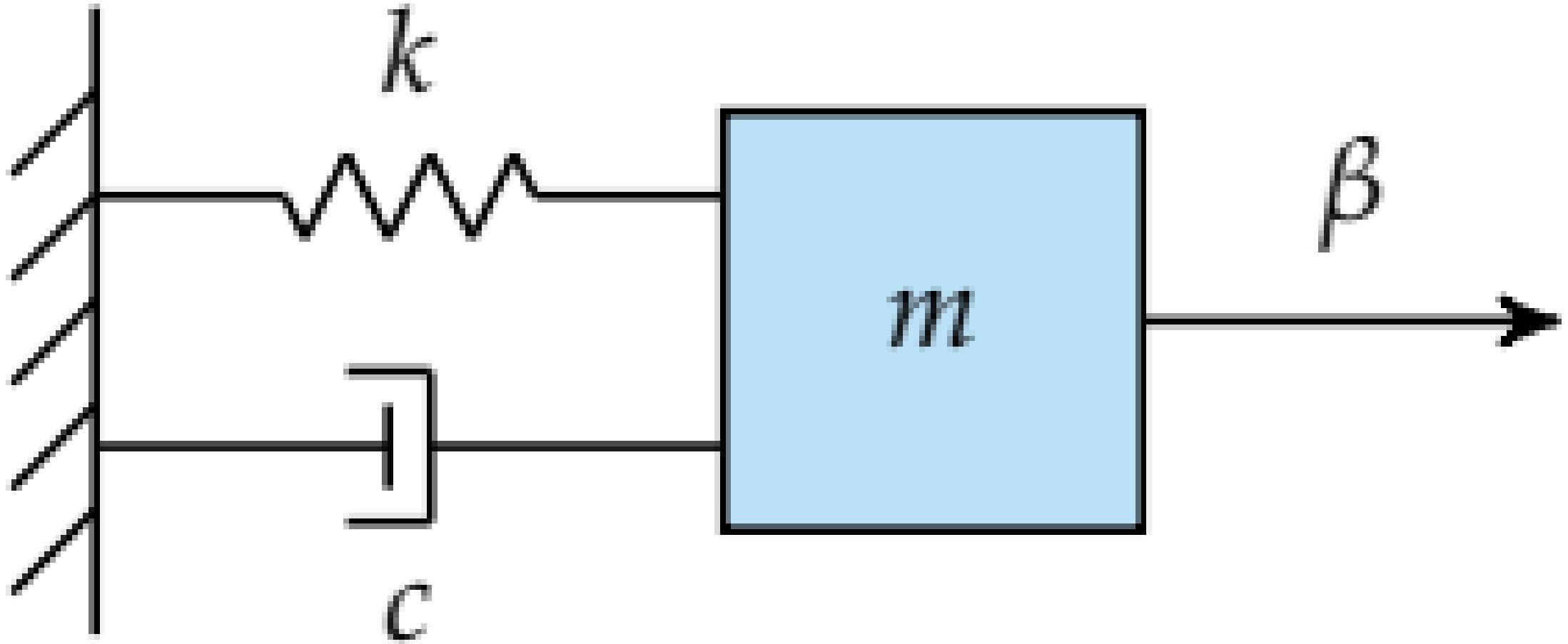


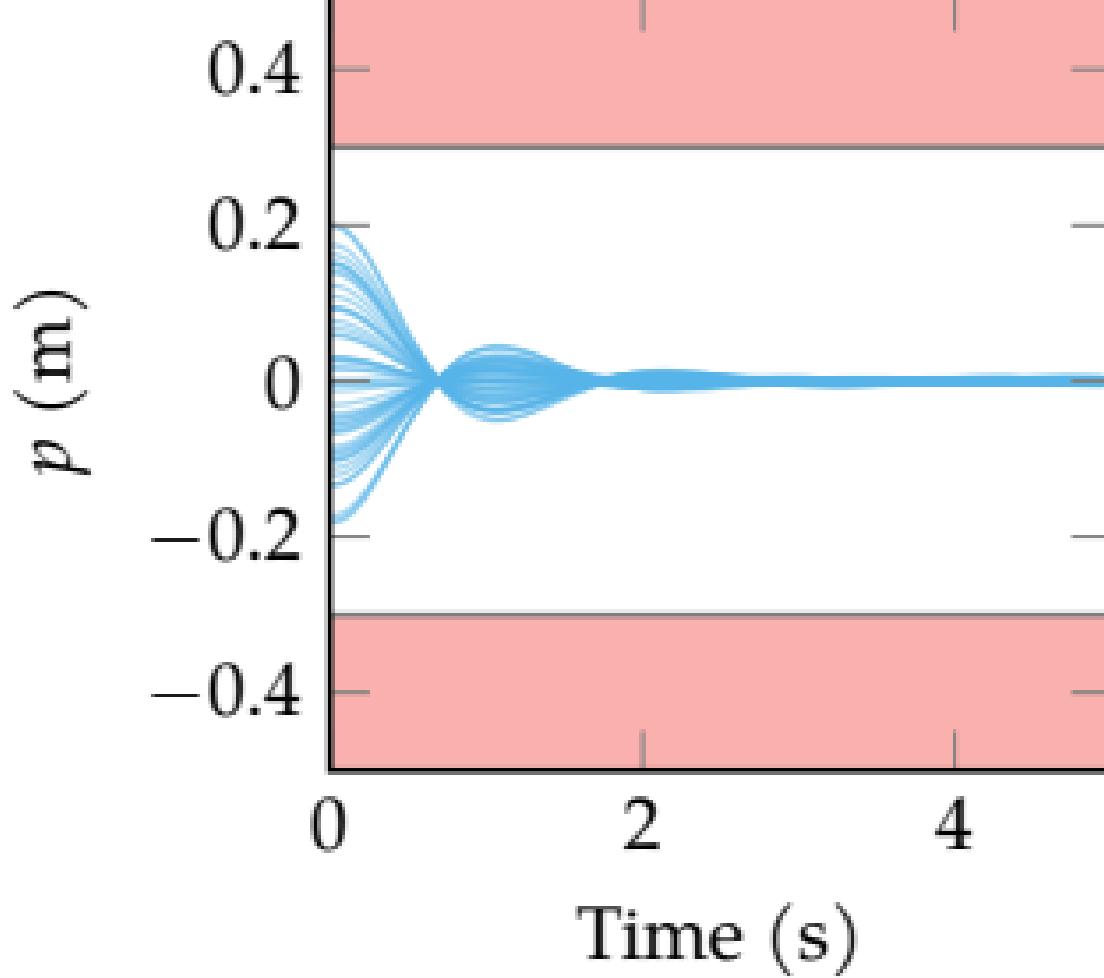


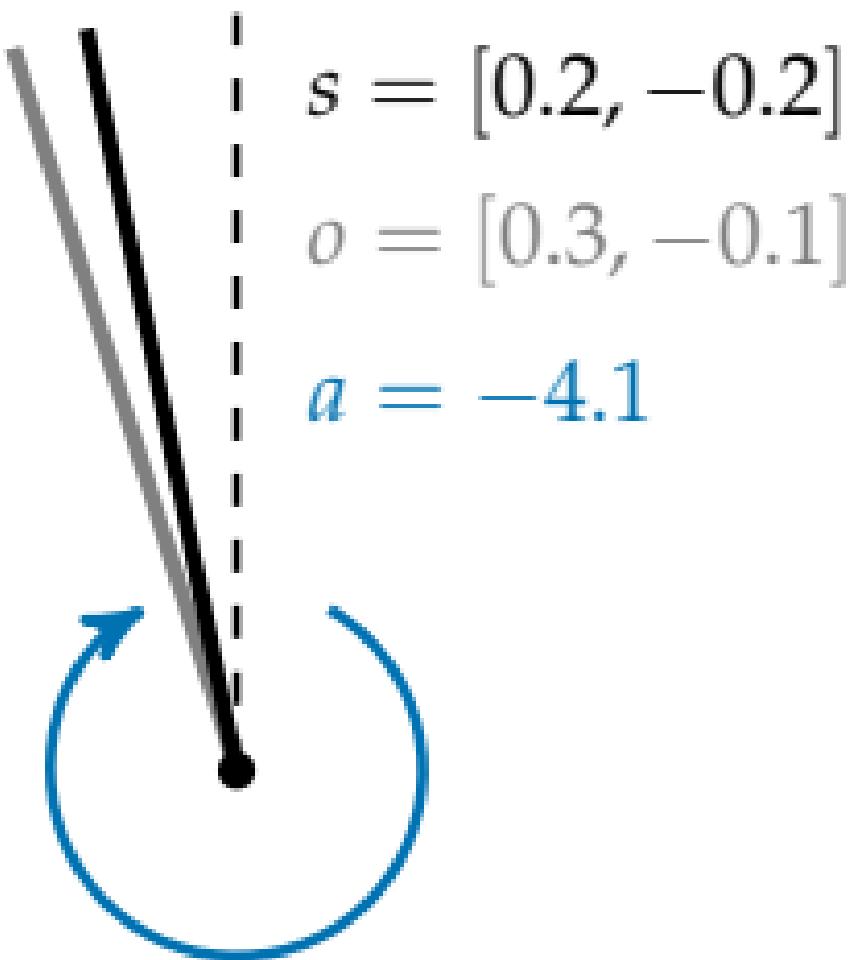


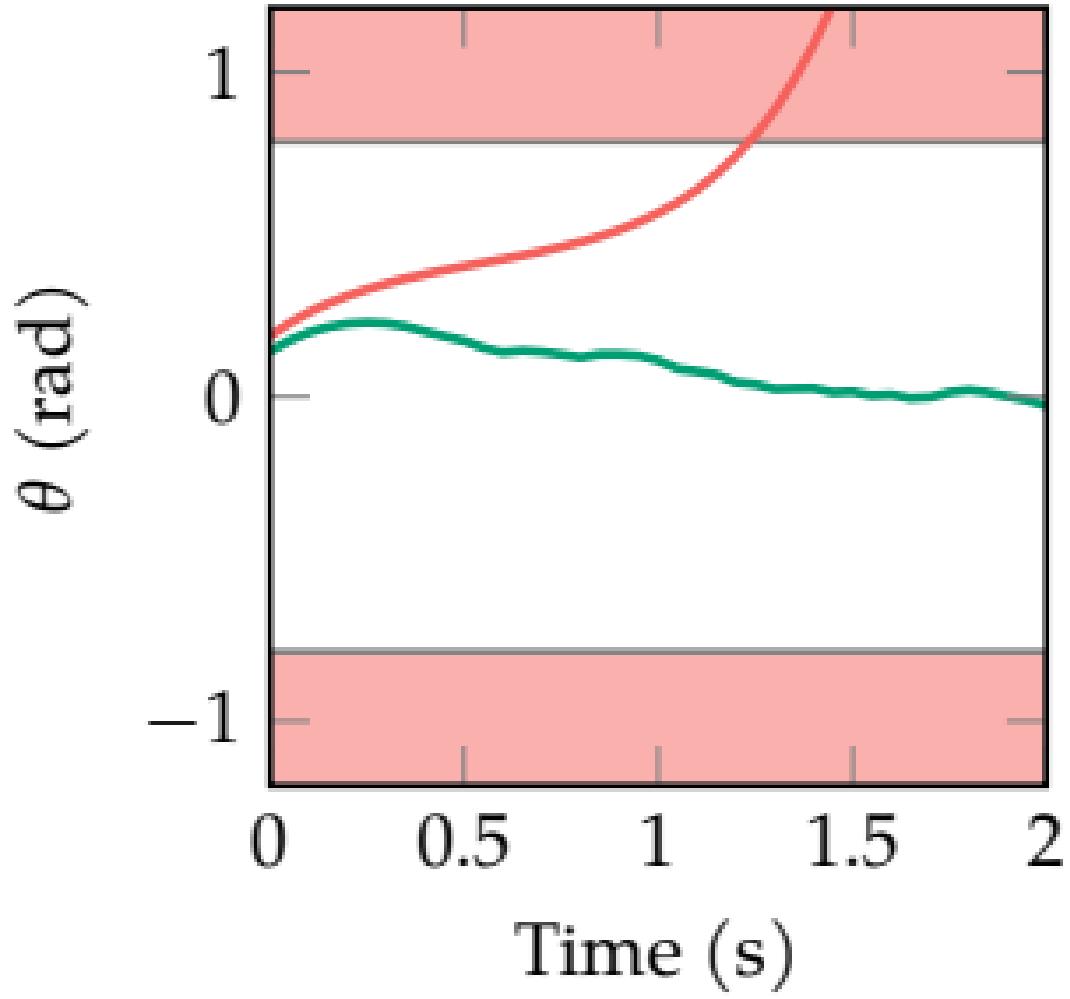


$s_2$  $s_1$ 







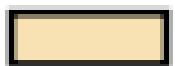




UP



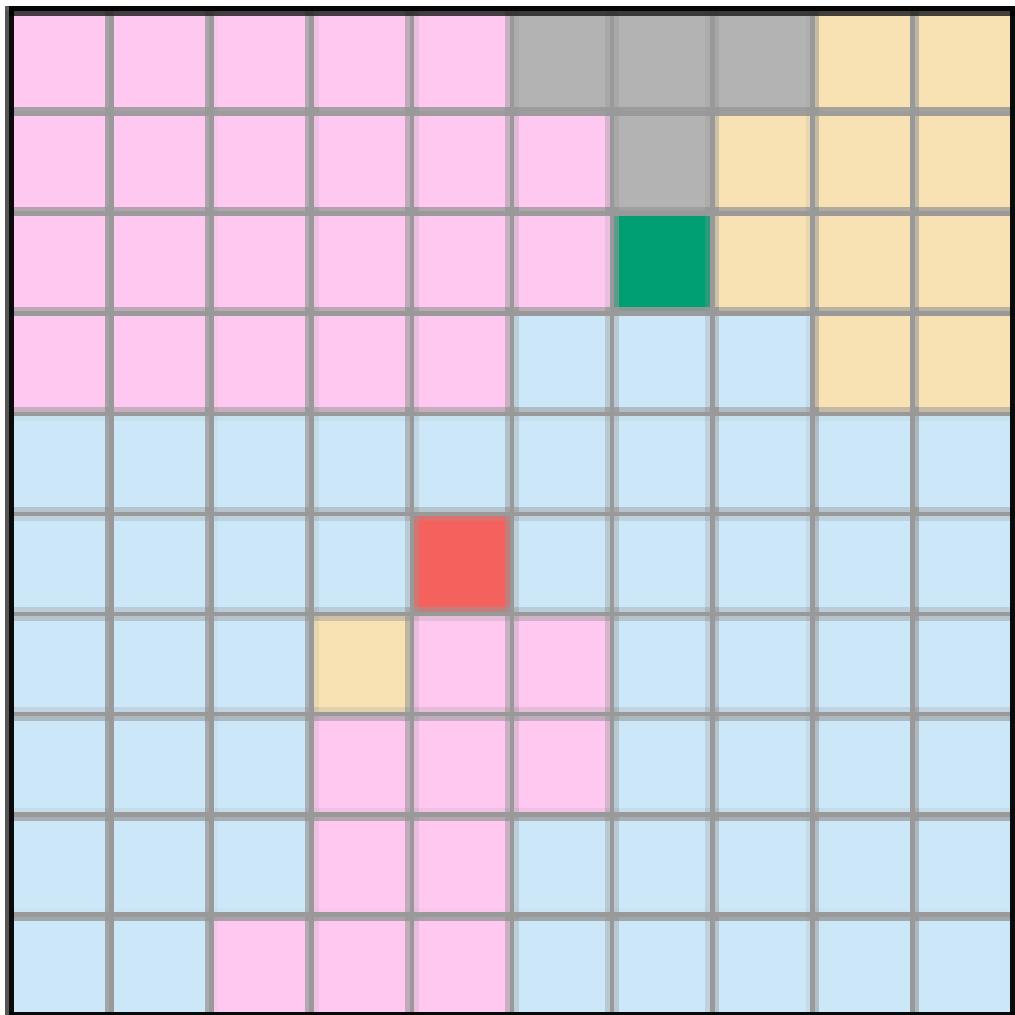
DOWN

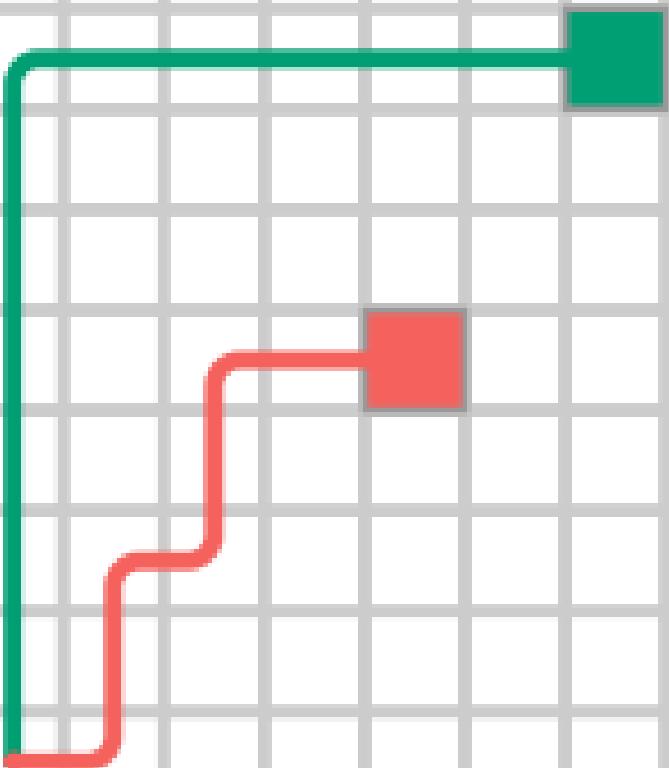


LEFT



RIGHT



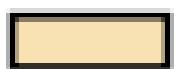




UP



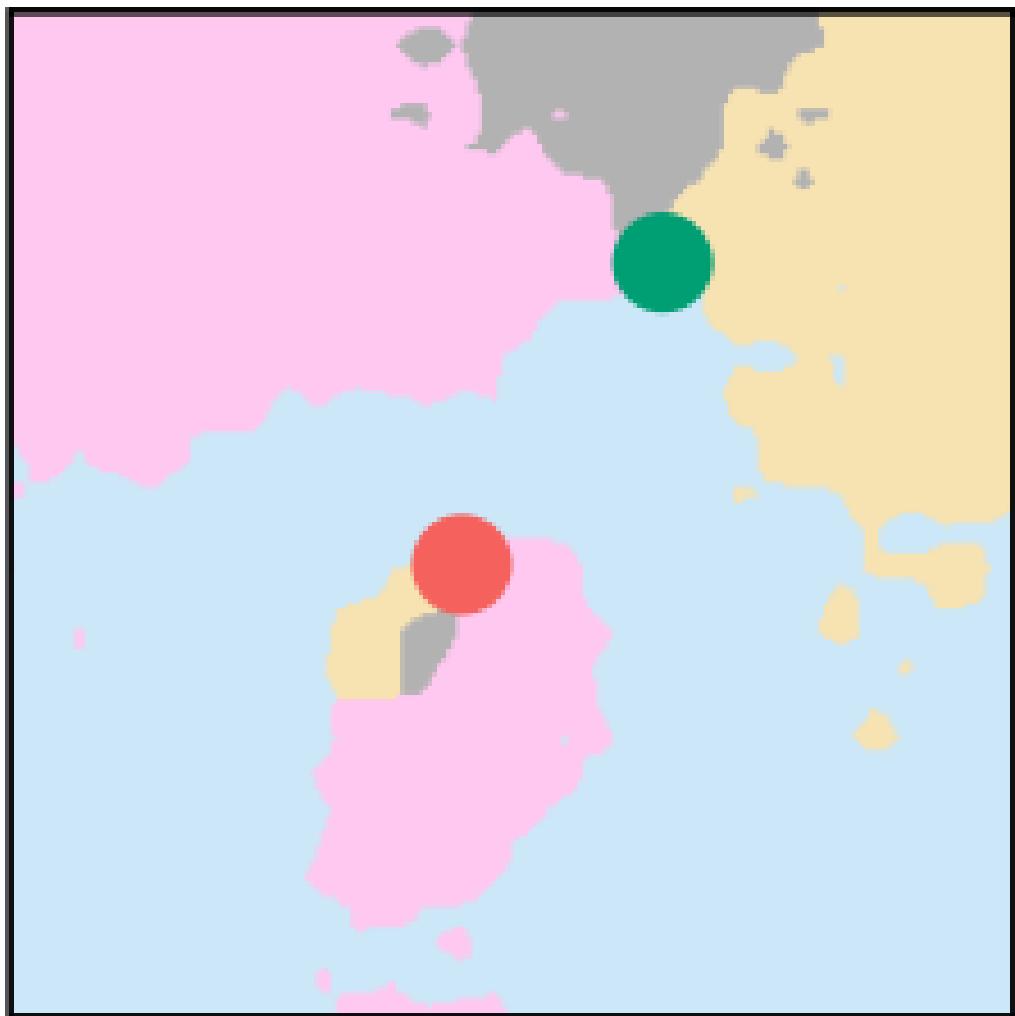
DOWN

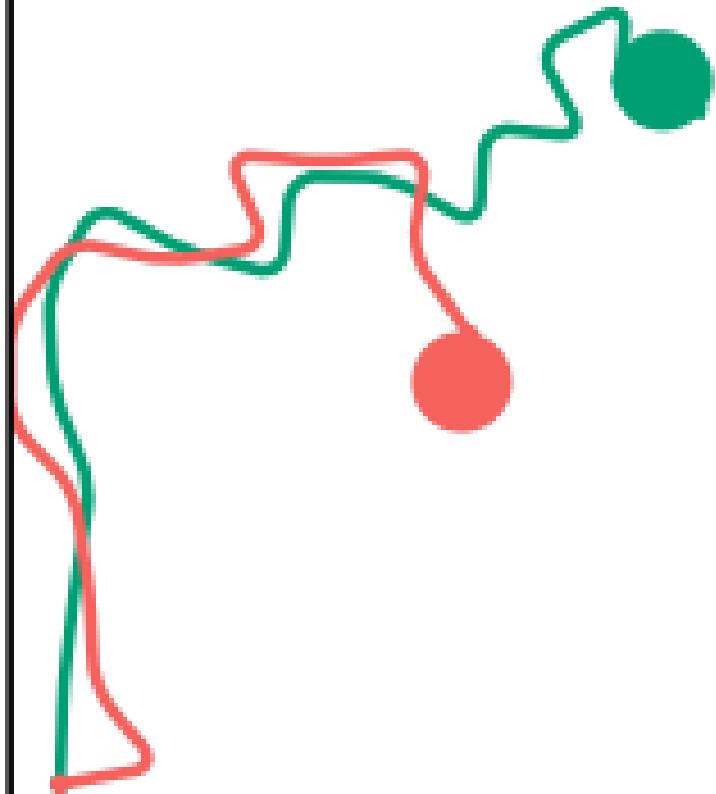


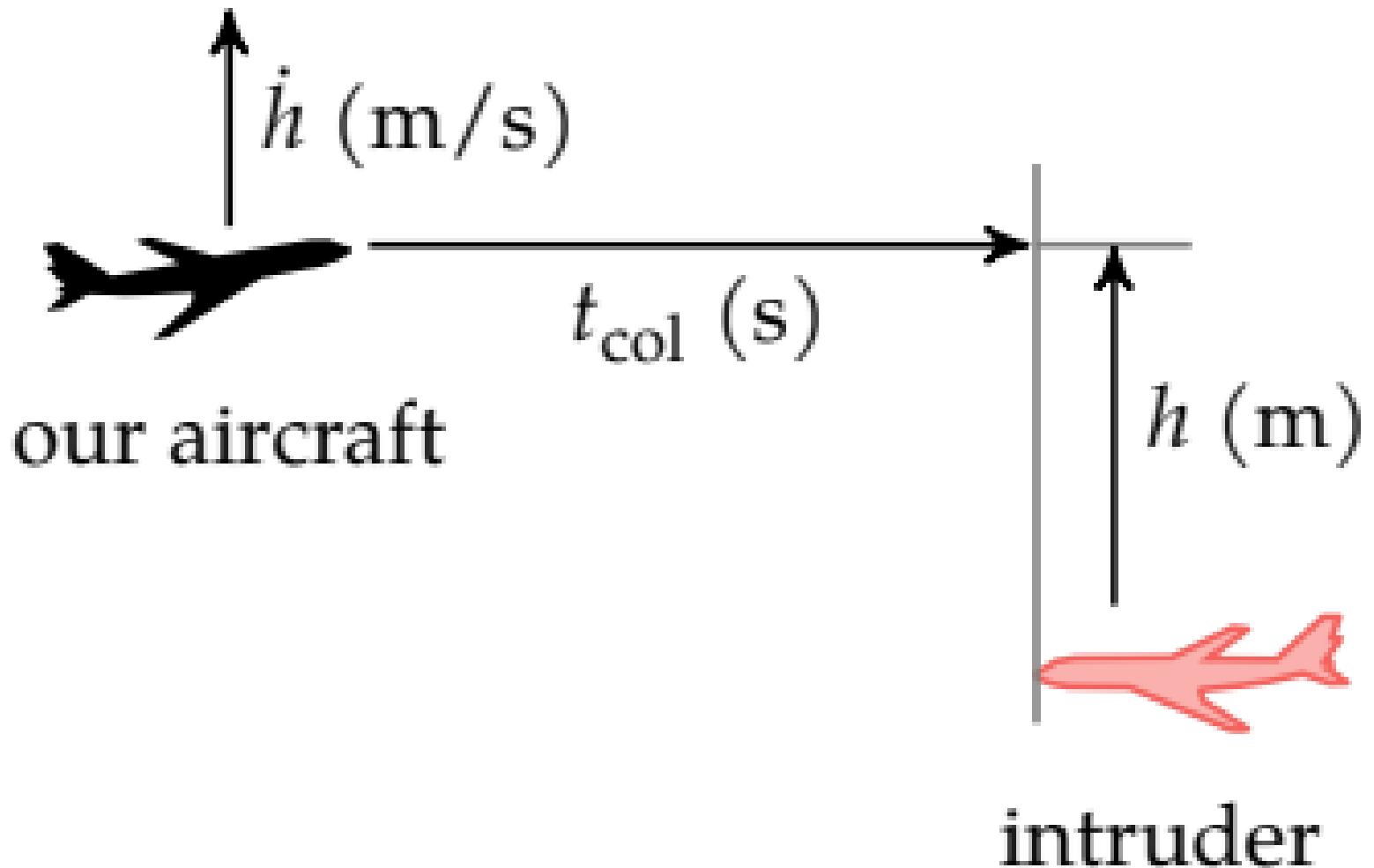
LEFT



RIGHT



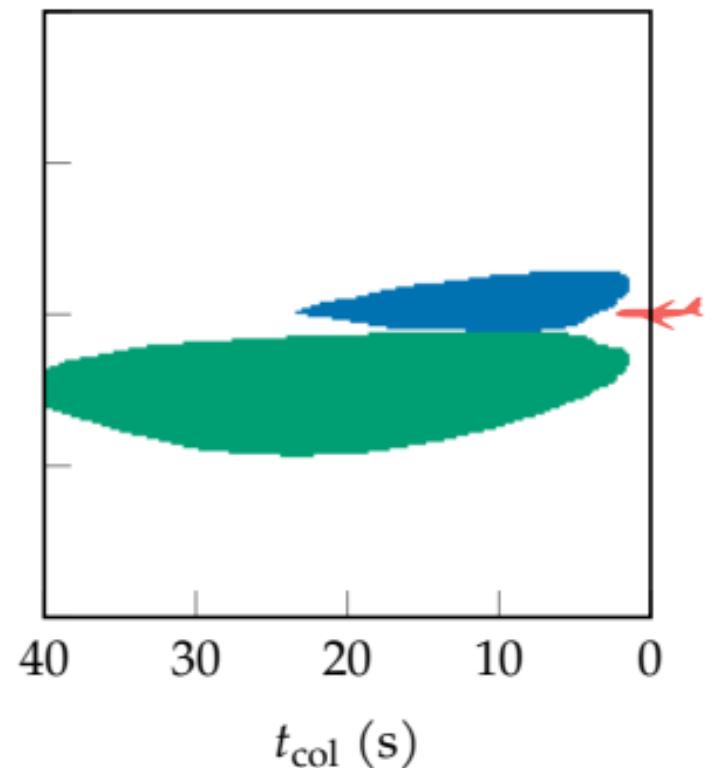
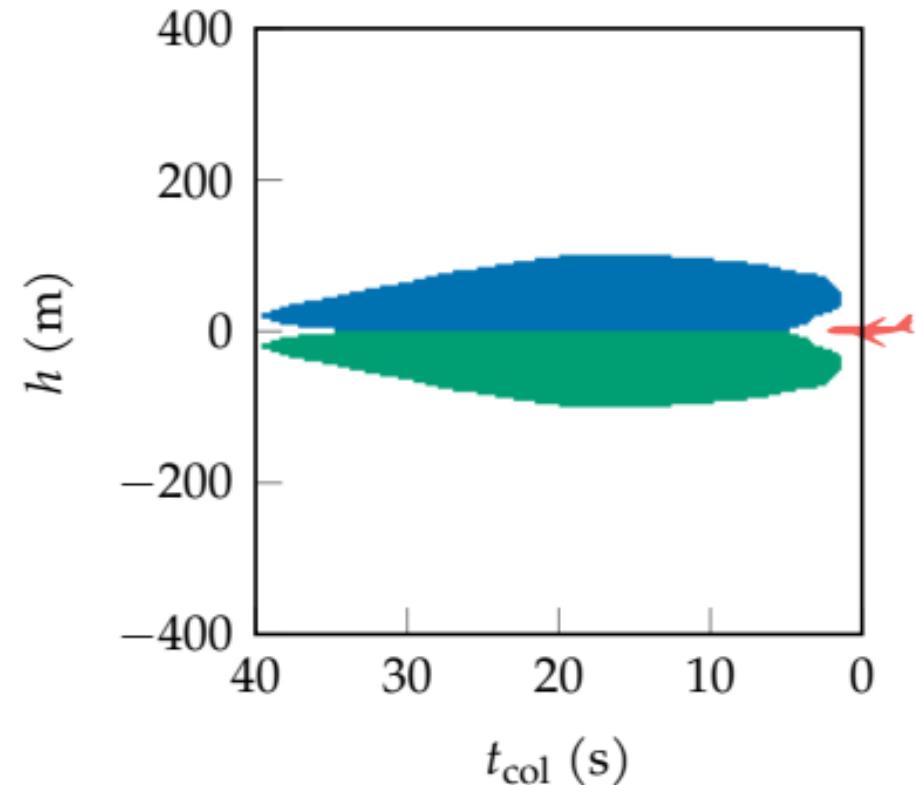




no advisory     descend     climb

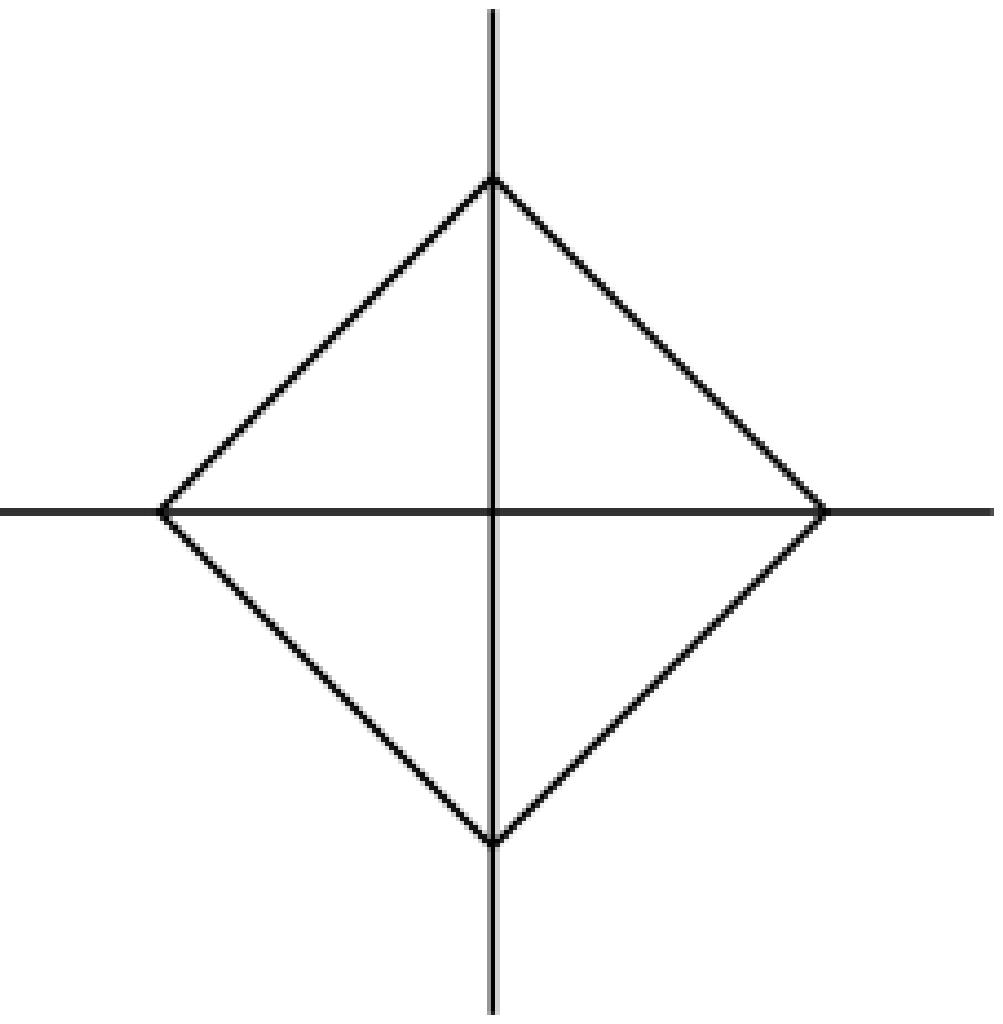
$$\dot{h} = 0 \text{ m/s}$$

$$\dot{h} = 4 \text{ m/s}$$



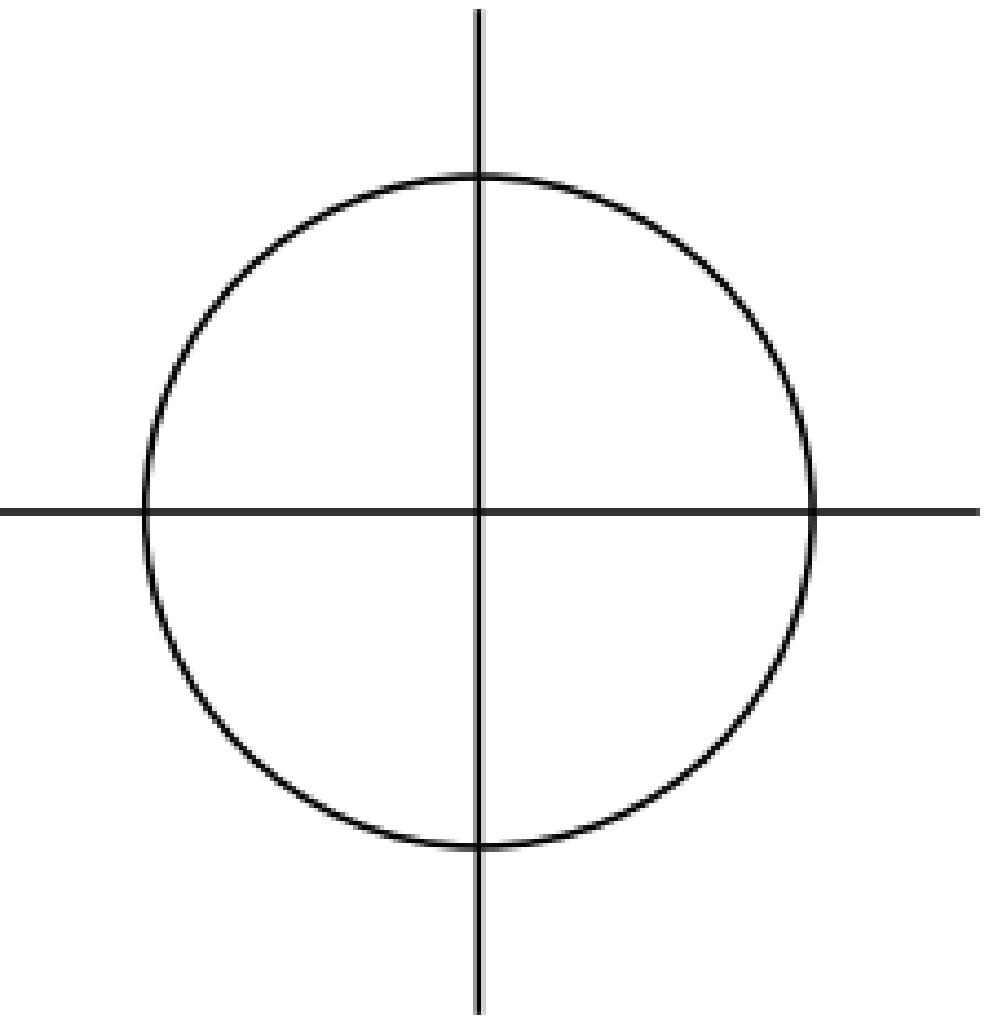
$$L_1: \|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

This metric is often referred to as the *taxicab norm*.



$$L_2: \|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

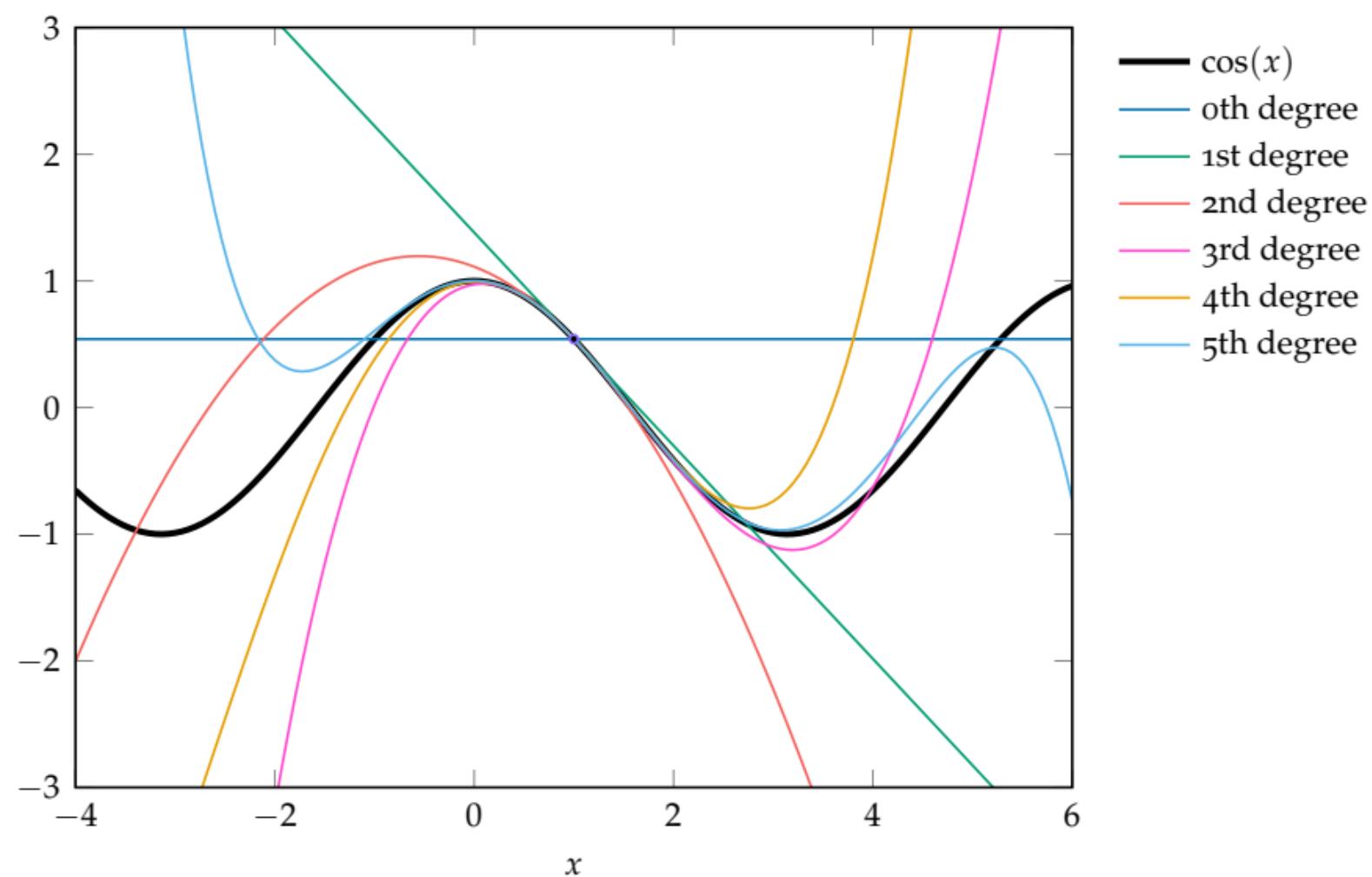
This metric is often referred to as the *Euclidean norm*.

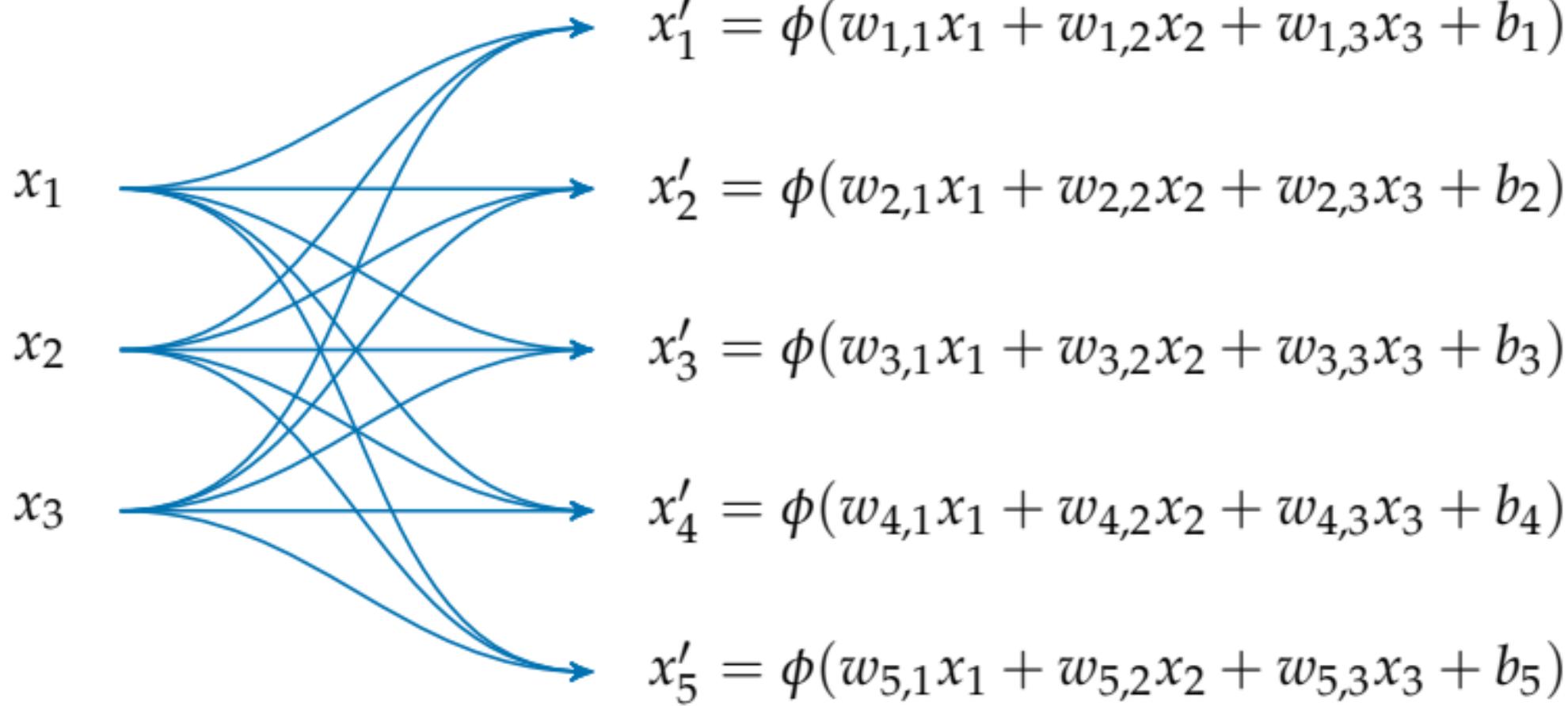


$$L_{\infty}: \|\mathbf{x}\|_{\infty} = \max(|x_1|, |x_2|, \dots, |x_n|)$$

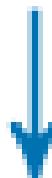
This metric is often referred to as the *max norm*, *Chebyshev norm*, or *chessboard norm*.

The latter name comes from the minimum number of moves that a king needs to move between two squares in chess.

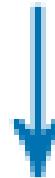





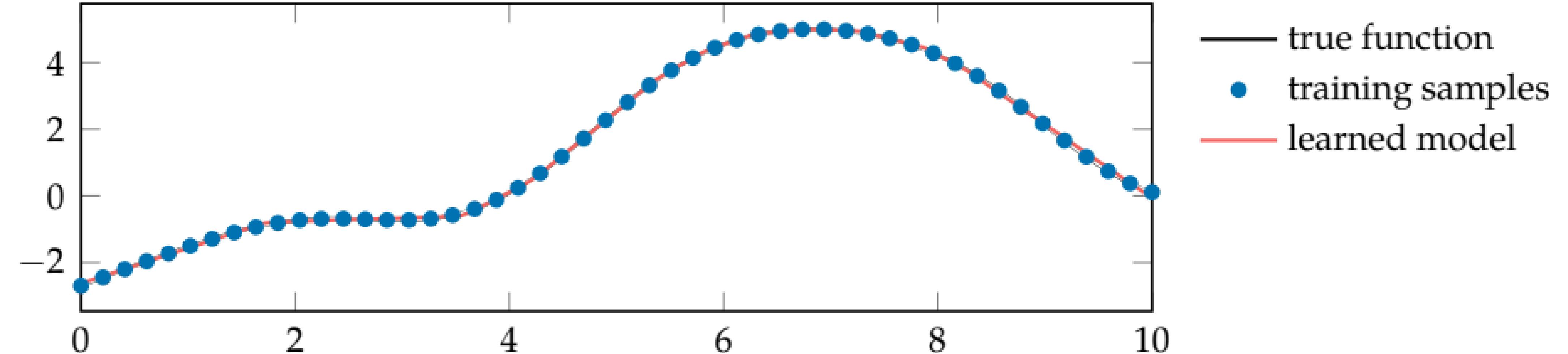
$$\mathbf{x} \in \mathbb{R}^3$$

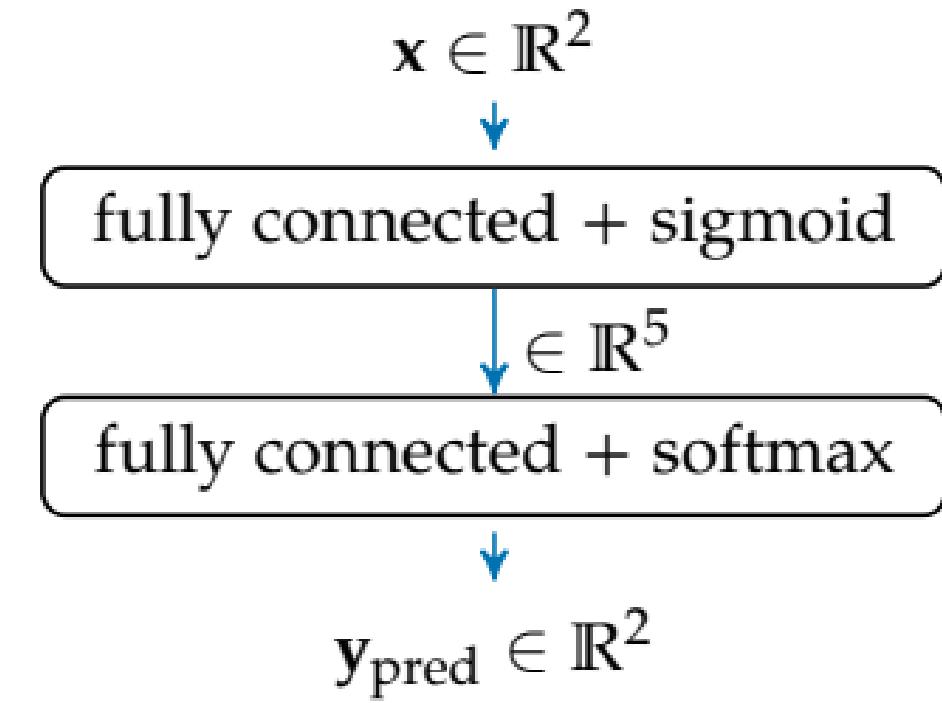
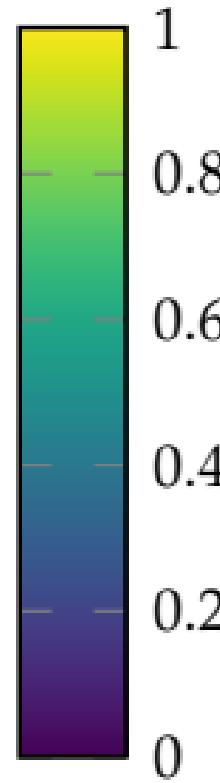
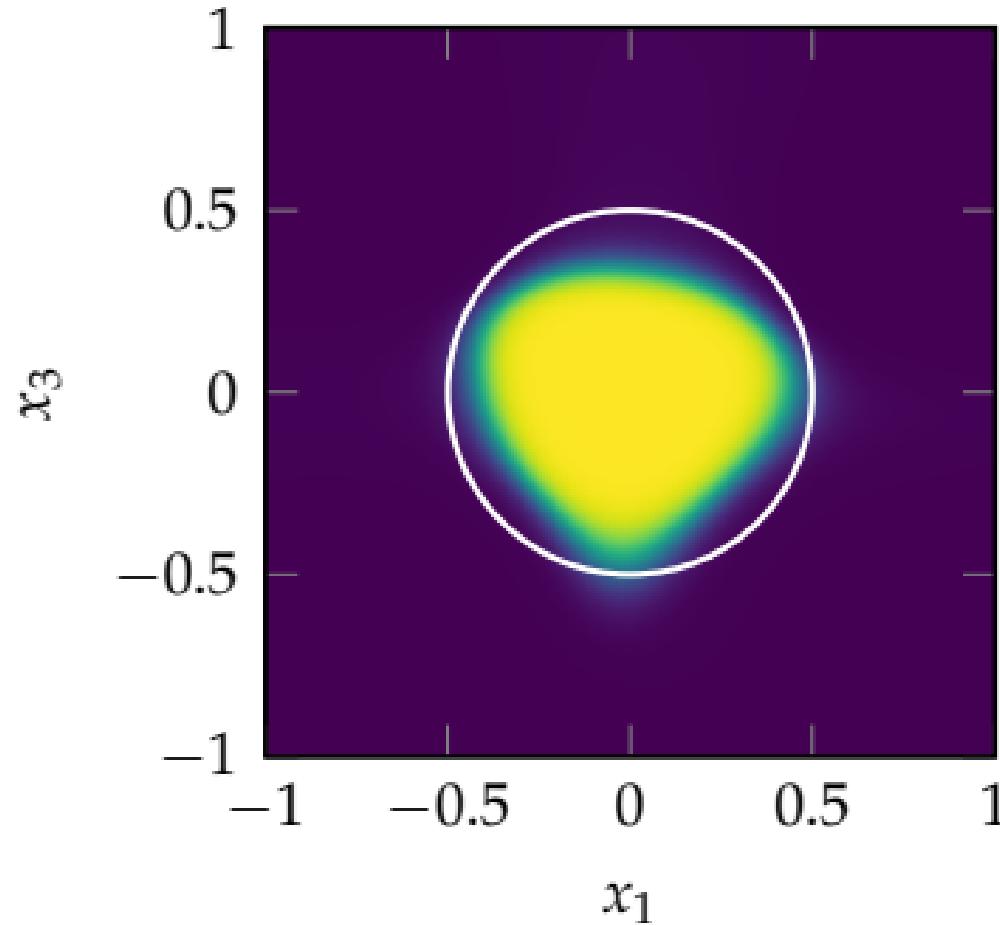


fully connected +  $\phi$



$$\mathbf{x}' \in \mathbb{R}^5$$





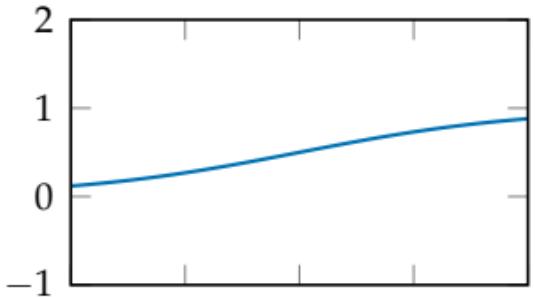
sigmoid

$$1/(1 + \exp(-x))$$

$\phi(x)$

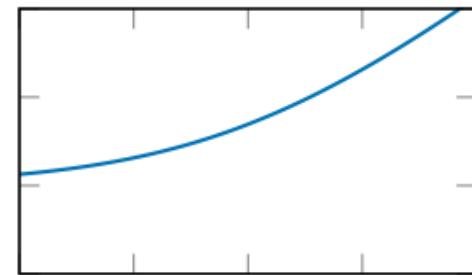
tanh

$$\tanh(x)$$



softplus

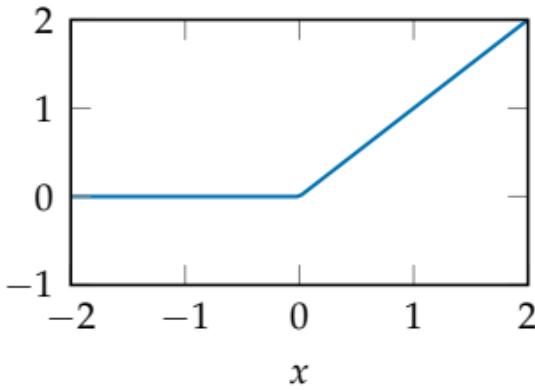
$$\log(1 + \exp(x))$$



relu

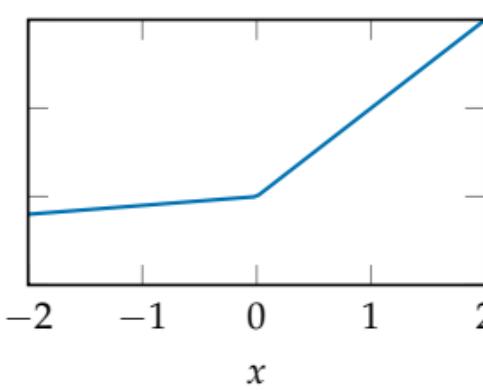
$$\max(0, x)$$

$\phi(x)$



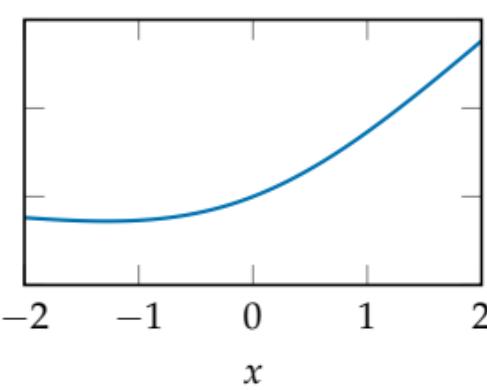
leaky relu

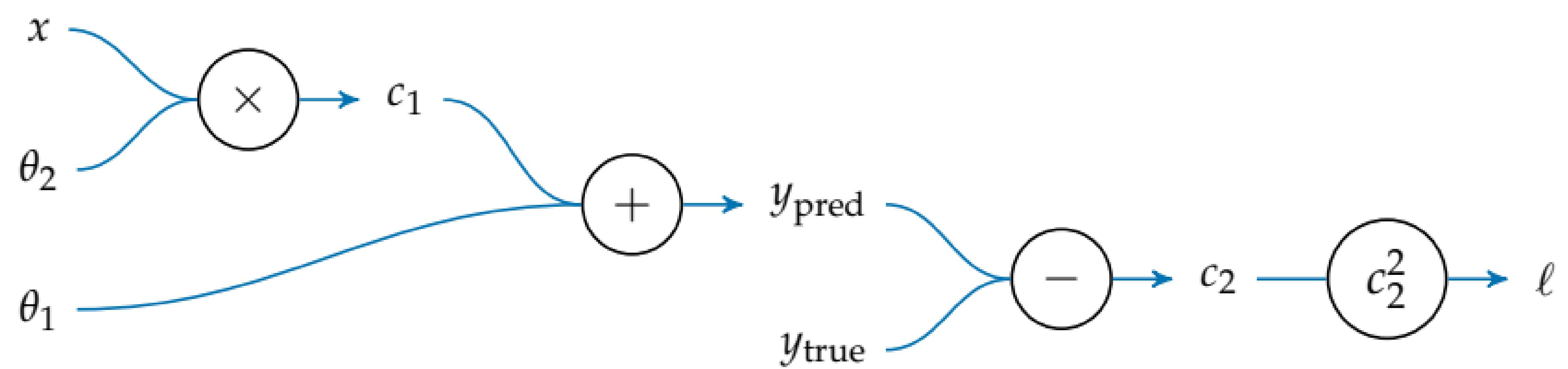
$$\max(\alpha x, x)$$

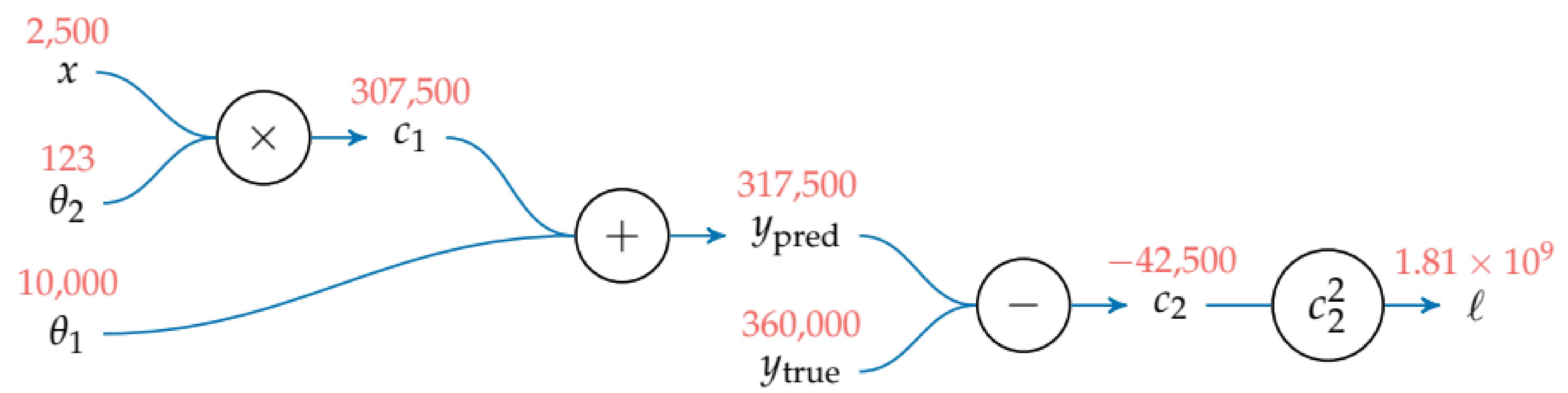


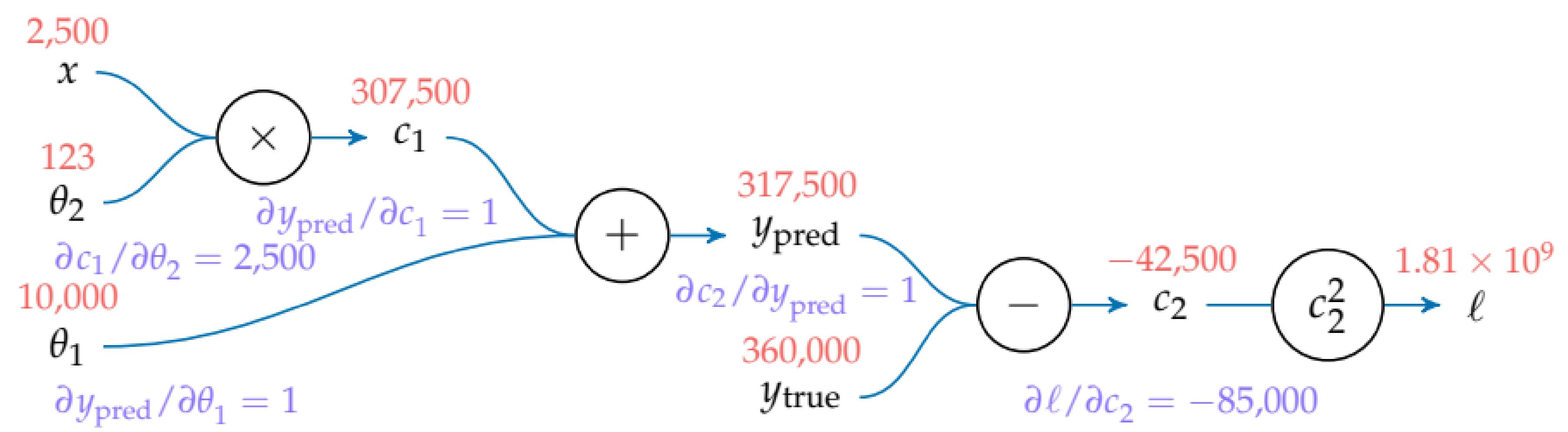
swish

$$x \operatorname{sigmoid}(x)$$









Any

