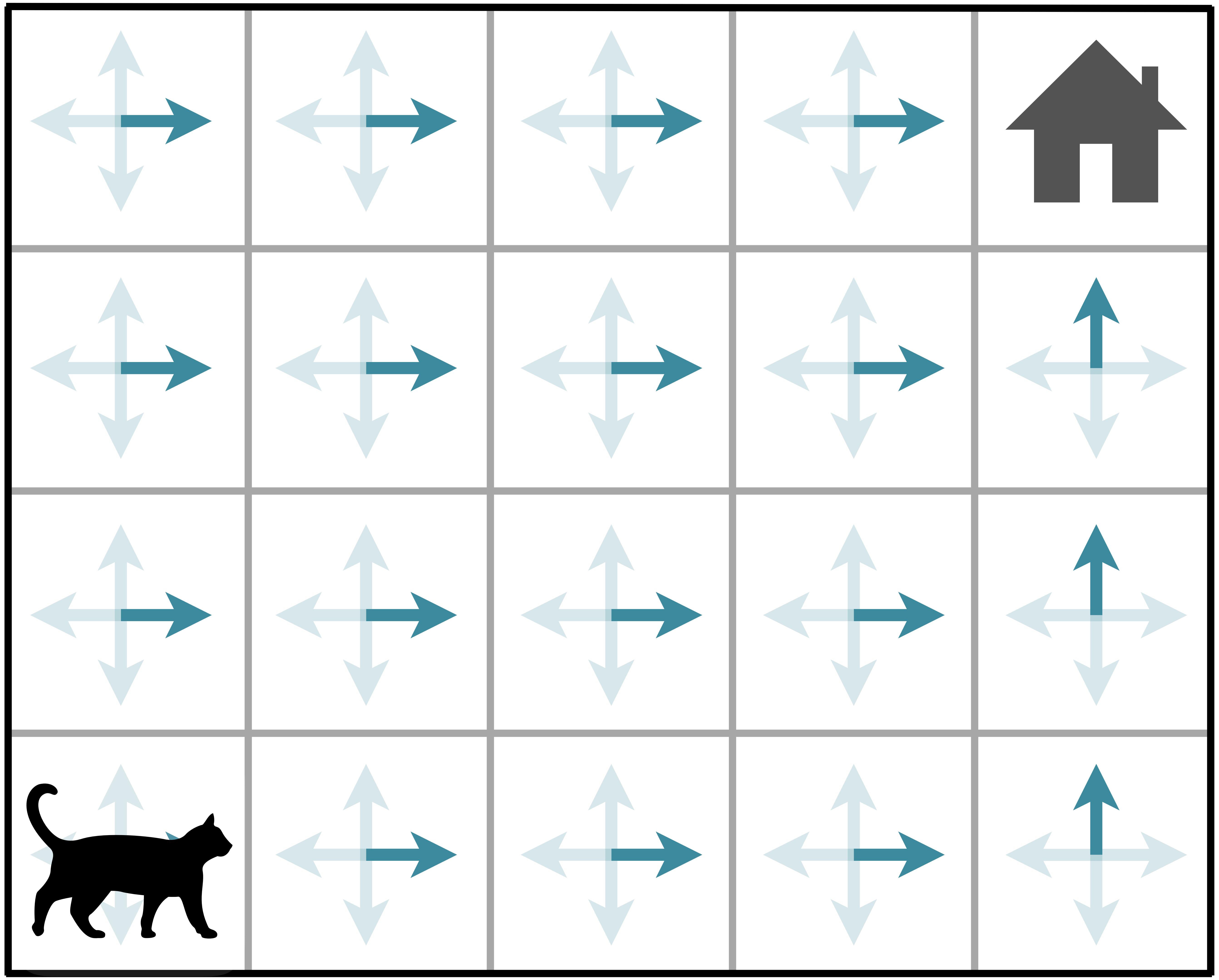
Importance Sampling Ratio Placement for Gradient-TD Methods

Andy Patterson



Roadmap

- **Importance Sampling (warm-up)**
- Off-policy TD(0) isr placement
- IS variance
- Gradient-TD placements

Importance Sampling

Sample: $x \sim b$

Estimate: $\mathbb{E}_\pi[X]$

Importance Sampling

Sample: $x \sim b$

Estimate: $\mathbb{E}_\pi[X]$

Importance Sampling

Sample: $x \sim b$

Estimate: $\mathbb{E}_{\pi}[X]$

Derivation of Importance Sampling

$$\mathbb{E}_\pi[X]$$

Derivation of Importance Sampling

$$\mathbb{E}_\pi[X] \doteq \sum_{x \in X} x\pi(x)$$

Derivation of Importance Sampling

$$\mathbb{E}_\pi[X] \doteq \sum_{x \in X} x\pi(x)$$

$$= \sum_{x \in X} x\pi(x)\frac{b(x)}{b(x)}$$

Derivation of Importance Sampling

$$\mathbb{E}_\pi[X] \doteq \sum_{x \in X} x\pi(x)$$

$$= \sum_{x \in X} x\pi(x)\frac{b(x)}{b(x)}$$

$$= \sum_{x \in X} x\frac{\pi(x)}{b(x)}b(x)$$

Derivation of Importance Sampling

$$\mathbb{E}_\pi[X] \doteq \sum_{x \in X} x\pi(x)$$

$$= \sum_{x \in X} x\pi(x)\frac{b(x)}{b(x)}$$

$$= \sum_{x \in X} x\boxed{\frac{\pi(x)}{b(x)}}b(x)$$

Importance sampling ratio

Derivation of Importance Sampling

$$\mathbb{E}_\pi[X] \doteq \sum_{x \in X} x\pi(x)$$

$$= \sum_{x \in X} x\pi(x)\frac{b(x)}{b(x)}$$

$$= \sum_{x \in X} x\rho(x)b(x)$$

Derivation of Importance Sampling

$$\mathbb{E}_{\pi}[X] = \sum_{x \in X} x\rho(x)b(x)$$

Derivation of Importance Sampling

$$\mathbb{E}_{\pi}[X] = \sum_{x \in X} x\rho(x) b(x)$$

Derivation of Importance Sampling

$$\mathbb{E}_{\pi}[X] = \sum_{x \in X} x \rho(x) b(x)$$

Derivation of Importance Sampling

$$\mathbb{E}_{\pi}[X] = \sum_{x \in X} x\rho(x)b(x)$$
$$= \mathbb{E}_b\left[X\rho(X)\right]$$

Derivation of Importance Sampling

$$\mathbb{E}_{\pi}[X] = \sum_{x \in X} x\rho(x)b(x)$$

$$= \mathbb{E}_{b}\left[X\rho(X)\right]$$

Roadmap

- **Importance Sampling**
- Off-policy TD(0) isr placement
- IS variance
- Gradient-TD placements

Roadmap

- Importance Sampling
- **Off-policy TD(0) isr placement**
- IS variance
- Gradient-TD placements

Off-Policy TD(0)

$$\delta = \rho(r + \gamma v') - v$$

$$w \leftarrow w + \alpha \delta x$$

Off-Policy TD(0)

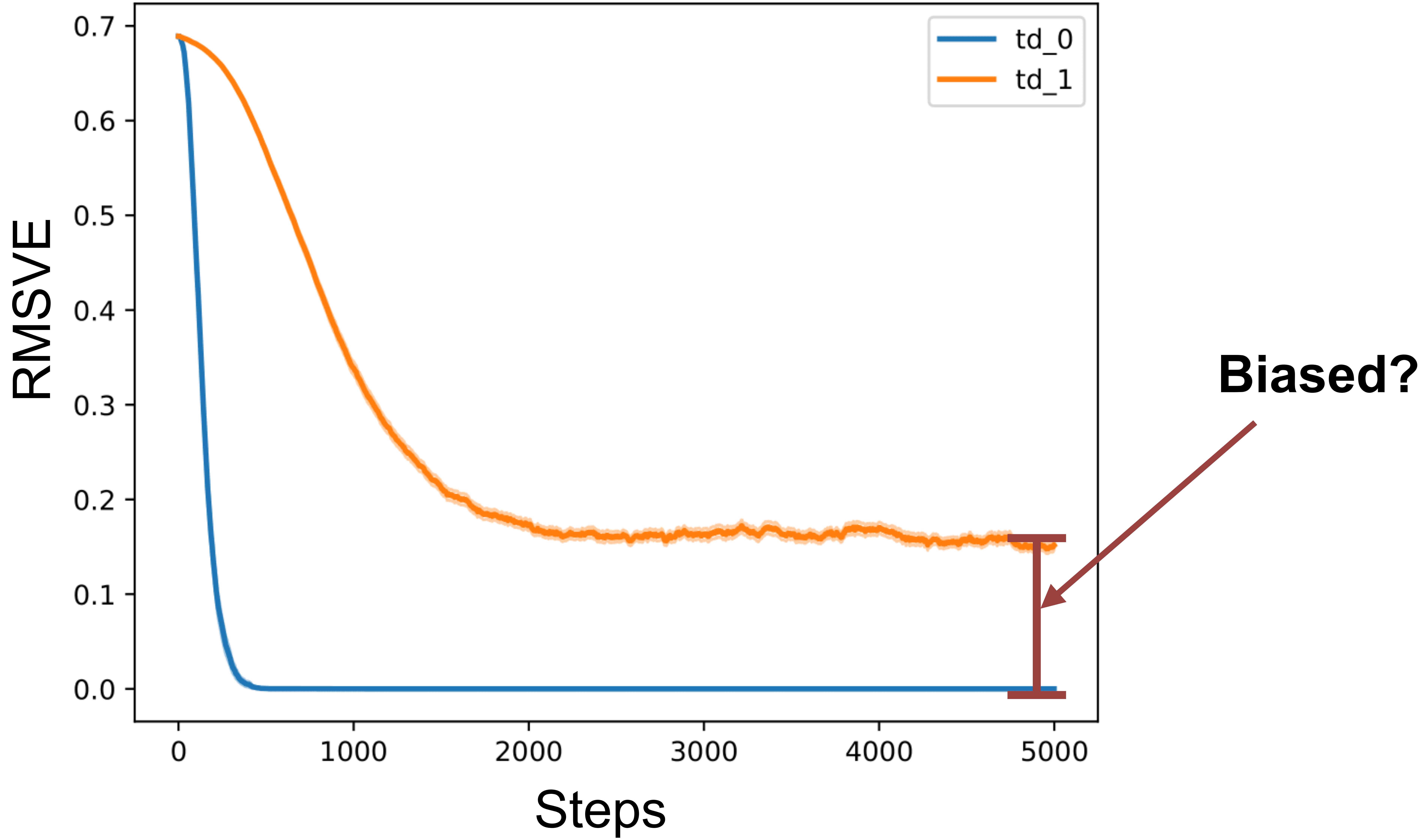
$\delta = \rho(r + \gamma v') - v$	$\delta^+ = \rho(r + \gamma v' - v)$

Off-Policy TD(0)

$\delta = \rho(r + \gamma v') - v$	$\delta^+ = \rho(r + \gamma v' - v)$
Precup, Sutton, Singh (2000)	**Maei (2011)**
Precup, Sutton, Dasgupta (2001)	**van Hasselt, Mahmood, Sutton (2014)**
	Mahmood, van Hasselt, Sutton (2014)

td_0: $\delta^{+} = \rho(r + \gamma v' - v)$

td_1: $\delta = \rho(r + \gamma v') - v$



$$\delta = \rho(r + \gamma v') - v$$

$$\delta^+ = \rho(r + \gamma v' - v)$$

$$\delta = \rho(r + \gamma v') - v \qquad \delta^{+} = \rho(r + \gamma v' - v)$$

$$\mathbb{E}_b[\delta] = \mathbb{E}_b\Big[\rho(r + \gamma v') - v\Big] \qquad \mathbb{E}_b[\delta^{+}] = \mathbb{E}_b\Big[\rho(r + \gamma v' - v)\Big]$$

$$\delta = \rho(r + \gamma v') - v \qquad\qquad \delta^+ = \rho(r + \gamma v' - v)$$

$$\begin{aligned}\mathbb{E}_b[\delta] &= \mathbb{E}_b\big[\rho(r + \gamma v') - v\big] \\ &= \mathbb{E}_b\big[\rho(r + \gamma v')\big] - \mathbb{E}_b[v]\end{aligned} \qquad \begin{aligned}\mathbb{E}_b[\delta^+] &= \mathbb{E}_b\big[\rho(r + \gamma v' - v)\big] \\ &= \mathbb{E}_b\big[\rho(r + \gamma v')\big] - \mathbb{E}_b[\rho v]\end{aligned}$$

$$\delta = \rho(r + \gamma v') - v \qquad\qquad \delta^+ = \rho(r + \gamma v' - v)$$

$$\begin{aligned}\mathbb{E}_b[\delta] &= \mathbb{E}_b\left[\rho(r + \gamma v') - v\right] \\ &= \boxed{\mathbb{E}_b\left[\rho(r + \gamma v')\right]} - \mathbb{E}_b[v]\end{aligned}$$

$$\begin{aligned}\mathbb{E}_b[\delta^+] &= \mathbb{E}_b\left[\rho(r + \gamma v' - v)\right] \\ &= \boxed{\mathbb{E}_b\left[\rho(r + \gamma v')\right]} - \mathbb{E}_b[\rho v]\end{aligned}$$

$$\delta = \rho(r + \gamma v') - v \qquad \delta^+ = \rho(r + \gamma v' - v)$$

$$\begin{aligned} \mathbb{E}_b[\delta] &= \mathbb{E}_b\left[\rho(r + \gamma v') - v\right] \\ &= \mathbb{E}_b\left[\rho(r + \gamma v')\right] - \mathbb{E}_b[v] \end{aligned}$$

$$\begin{aligned} \mathbb{E}_b[\delta^+] &= \mathbb{E}_b\left[\rho(r + \gamma v' - v)\right] \\ &= \mathbb{E}_b\left[\rho(r + \gamma v')\right] - \mathbb{E}_b[\rho v] \end{aligned}$$

$$\mathbb{E}_b[v] \qquad \mathbb{E}_b[\rho v]$$

$$\delta = \rho(r + \gamma v') - v \qquad\qquad \delta^+ = \rho(r + \gamma v' - v)$$

$$\begin{aligned}\mathbb{E}_b[\delta] &= \mathbb{E}_b\left[\rho(r + \gamma v') - v\right] \\ &= \mathbb{E}_b\left[\rho(r + \gamma v')\right] - \mathbb{E}_b[v]\end{aligned}$$

$$\begin{aligned}\mathbb{E}_b[\delta^+] &= \mathbb{E}_b\left[\rho(r + \gamma v' - v)\right] \\ &= \mathbb{E}_b\left[\rho(r + \gamma v')\right] - \mathbb{E}_b[\rho v]\end{aligned}$$

$$\mathbb{E}_b[v] \overset{?}{=} \mathbb{E}_b[\rho v]$$

$$\delta = \rho(r + \gamma v') - v \qquad \delta^+ = \rho(r + \gamma v' - v)$$

$$\begin{aligned}\mathbb{E}_b[\delta] &= \mathbb{E}_b\left[\rho(r + \gamma v') - v\right] \\ &= \mathbb{E}_b\left[\rho(r + \gamma v')\right] - \mathbb{E}_b[v]\end{aligned}$$

$$\begin{aligned}\mathbb{E}_b[\delta^+] &= \mathbb{E}_b\left[\rho(r + \gamma v' - v)\right] \\ &= \mathbb{E}_b\left[\rho(r + \gamma v')\right] - \mathbb{E}_b[\rho v]\end{aligned}$$

$$\begin{aligned}\mathbb{E}_b[v] &\overset{?}{=} \mathbb{E}_b[\rho v] \\ v &= v\mathbb{E}_b[\rho]\end{aligned}$$

$$\delta = \rho(r + \gamma v') - v \qquad \delta^+ = \rho(r + \gamma v' - v)$$

$$\mathbb{E}_b[\delta] = \mathbb{E}_b\big[\rho(r + \gamma v') - v\big]$$
$$= \mathbb{E}_b\big[\rho(r + \gamma v')\big] - \mathbb{E}_b[v]$$

$$\mathbb{E}_b[\delta^+] = \mathbb{E}_b\big[\rho(r + \gamma v' - v)\big]$$
$$= \mathbb{E}_b\big[\rho(r + \gamma v')\big] - \mathbb{E}_b[\rho v]$$

$$\mathbb{E}_b[v] \overset{?}{=} \mathbb{E}_b[\rho v]$$
$$v = v\,\mathbb{E}_b[\rho]$$

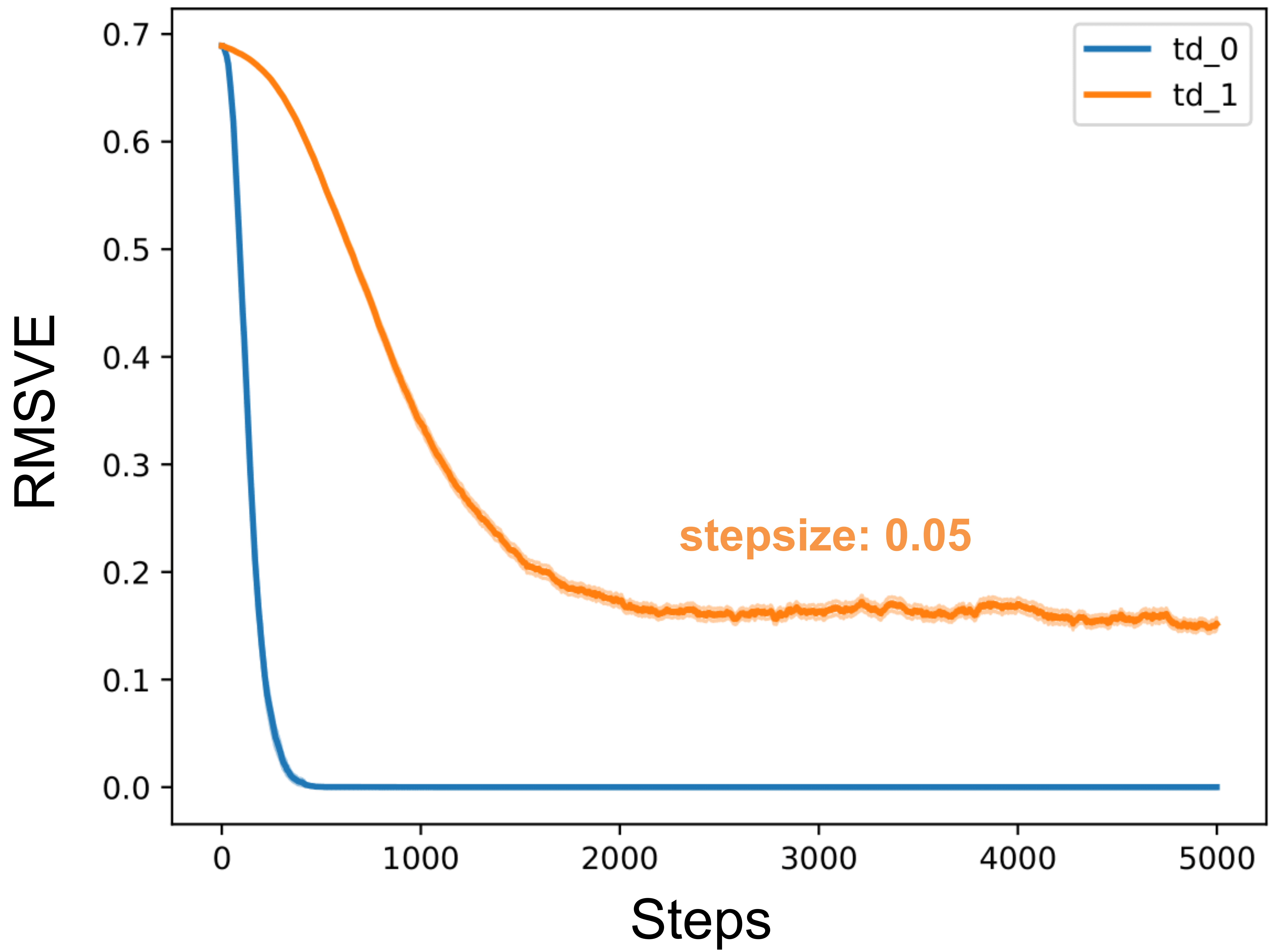
$$\mathbb{E}_b[\rho] = \sum \frac{\pi(x)}{b(x)} b(x)$$
$$= \sum \pi(x) = 1$$

$$\delta = \rho(r + \gamma v') - v \qquad \delta^+ = \rho(r + \gamma v' - v)$$

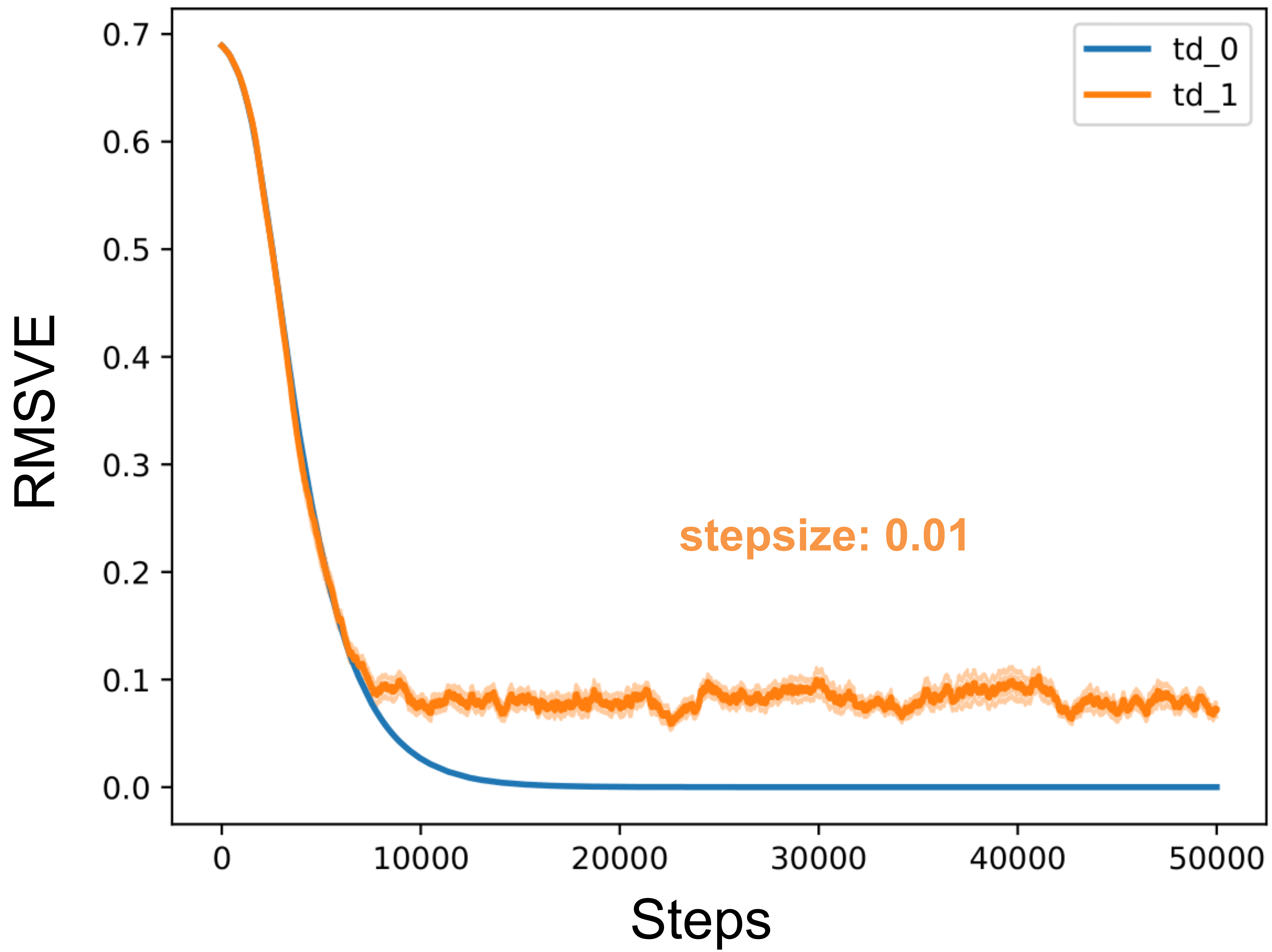
$$\begin{aligned} \mathbb{E}_b[\delta] &= \mathbb{E}_b\Big[\rho(r + \gamma v') - v\Big] \\ &= \mathbb{E}_b\Big[\rho(r + \gamma v')\Big] - \mathbb{E}_b[v] \end{aligned} \qquad \begin{aligned} \mathbb{E}_b[\delta^+] &= \mathbb{E}_b\Big[\rho(r + \gamma v' - v)\Big] \\ &= \mathbb{E}_b\Big[\rho(r + \gamma v')\Big] - \mathbb{E}_b[\rho v] \end{aligned}$$

$$\begin{aligned} \mathbb{E}_b[v] &\stackrel{?}{=} \mathbb{E}_b[\rho v] \\ v &= v\mathbb{E}_b[\rho] \\ v &= v \end{aligned}$$

5k steps

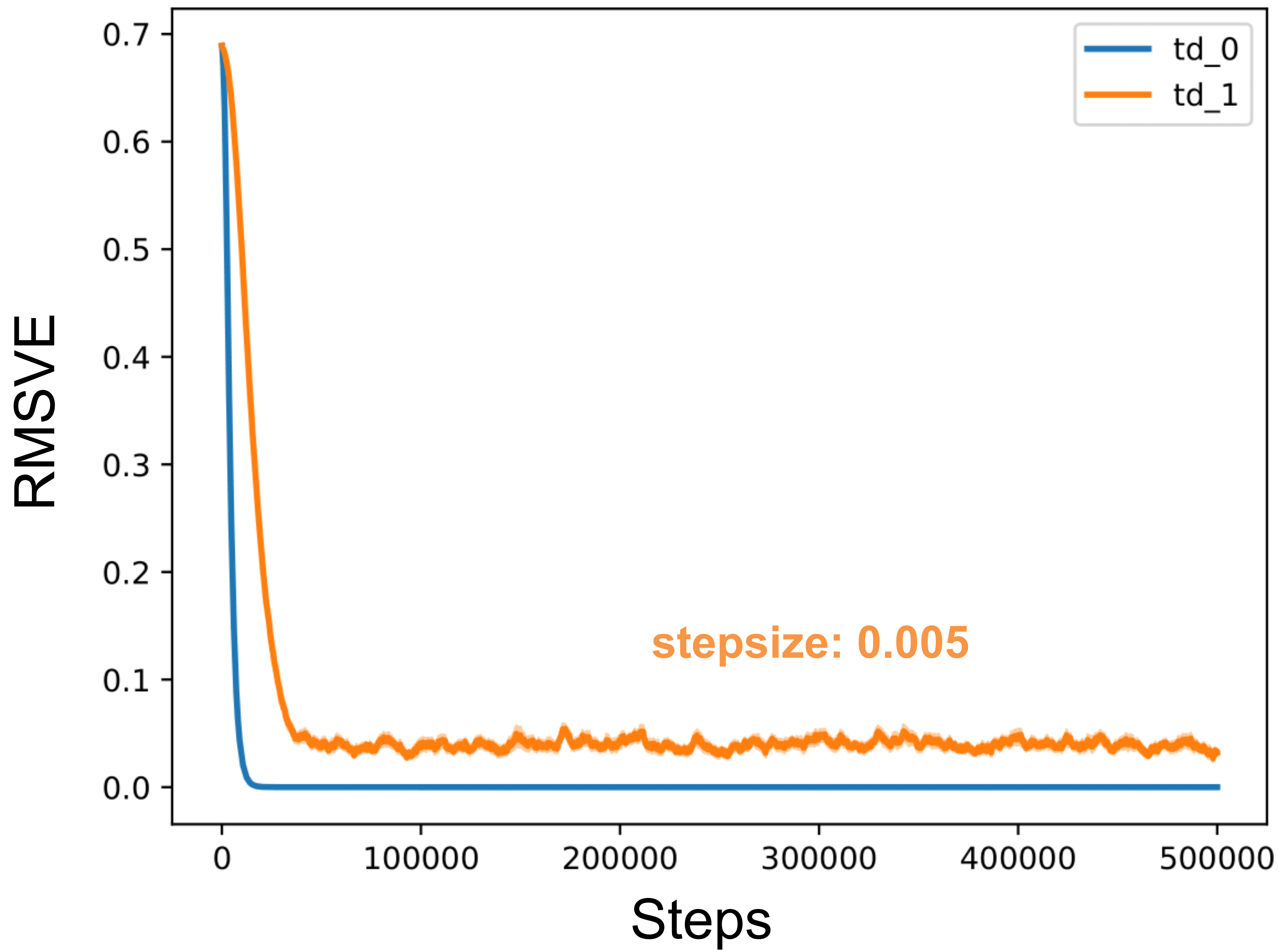

td_0
td_1
stepsize: 0.05
RMSVE
Steps
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1000
2000
3000
4000
5000

50k steps



0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
RMSVE
0
10000
20000
30000
40000
50000
Steps
td_0
td_1
stepsize: 0.01

500k steps



td_0
td_1
stepsize: 0.005
RMSVE
Steps
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
100000
200000
300000
400000
500000

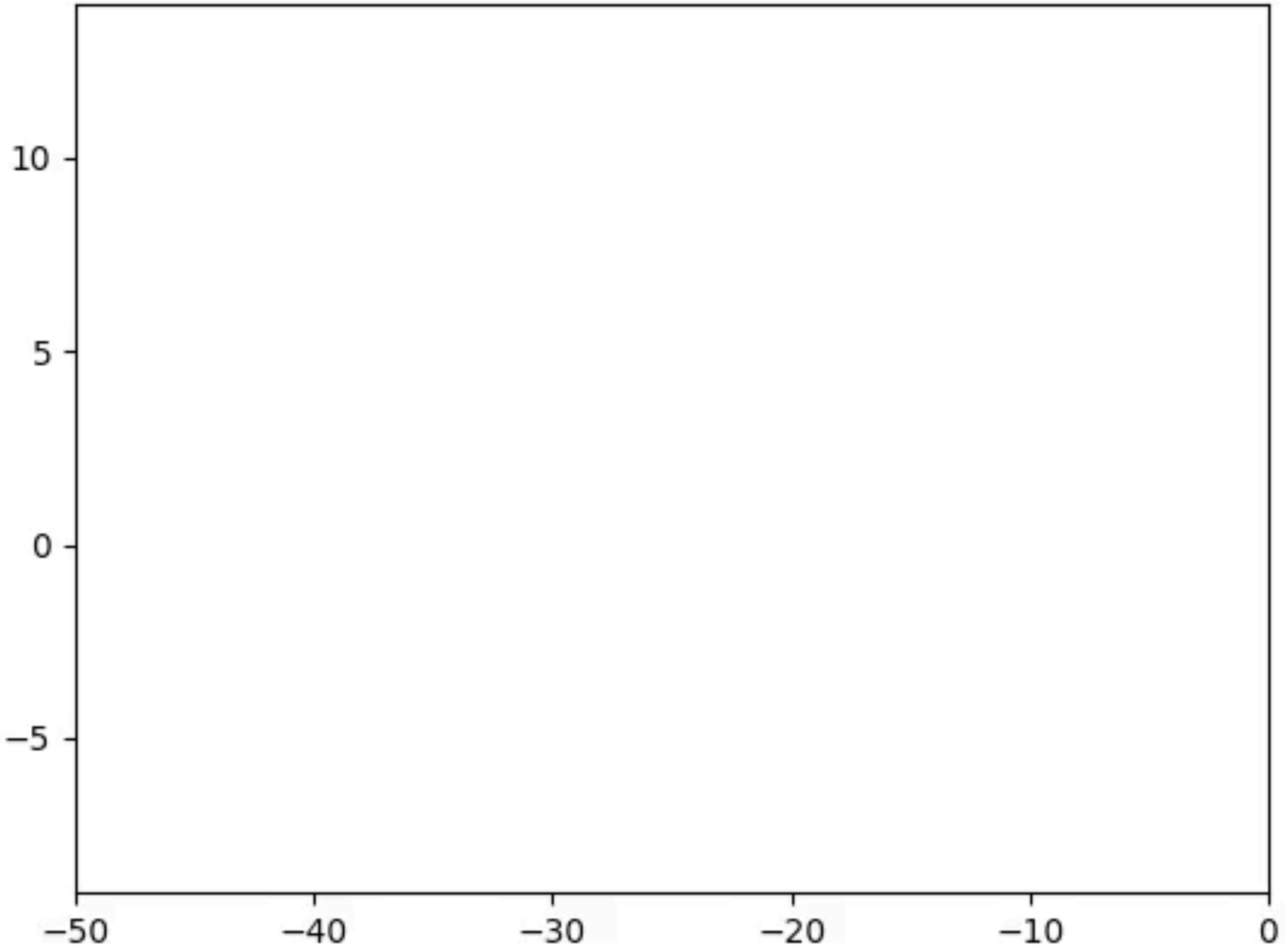
Roadmap

- Importance Sampling
- **Off-policy TD(0) isr placement**
- IS variance
- Gradient-TD placements

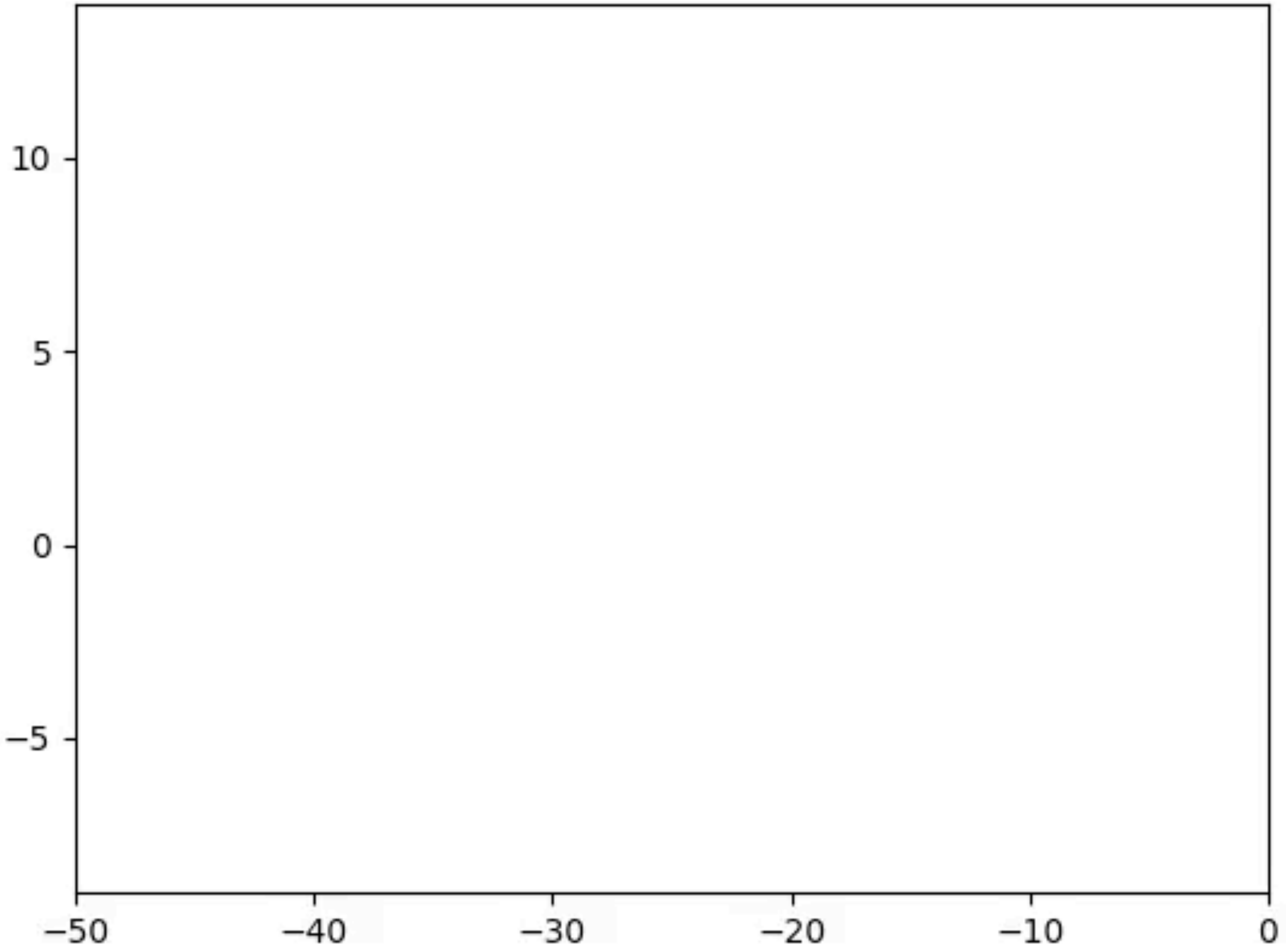
Roadmap

- Importance Sampling
- Off-policy TD(0) isr placement
- **IS variance**
- Gradient-TD placements

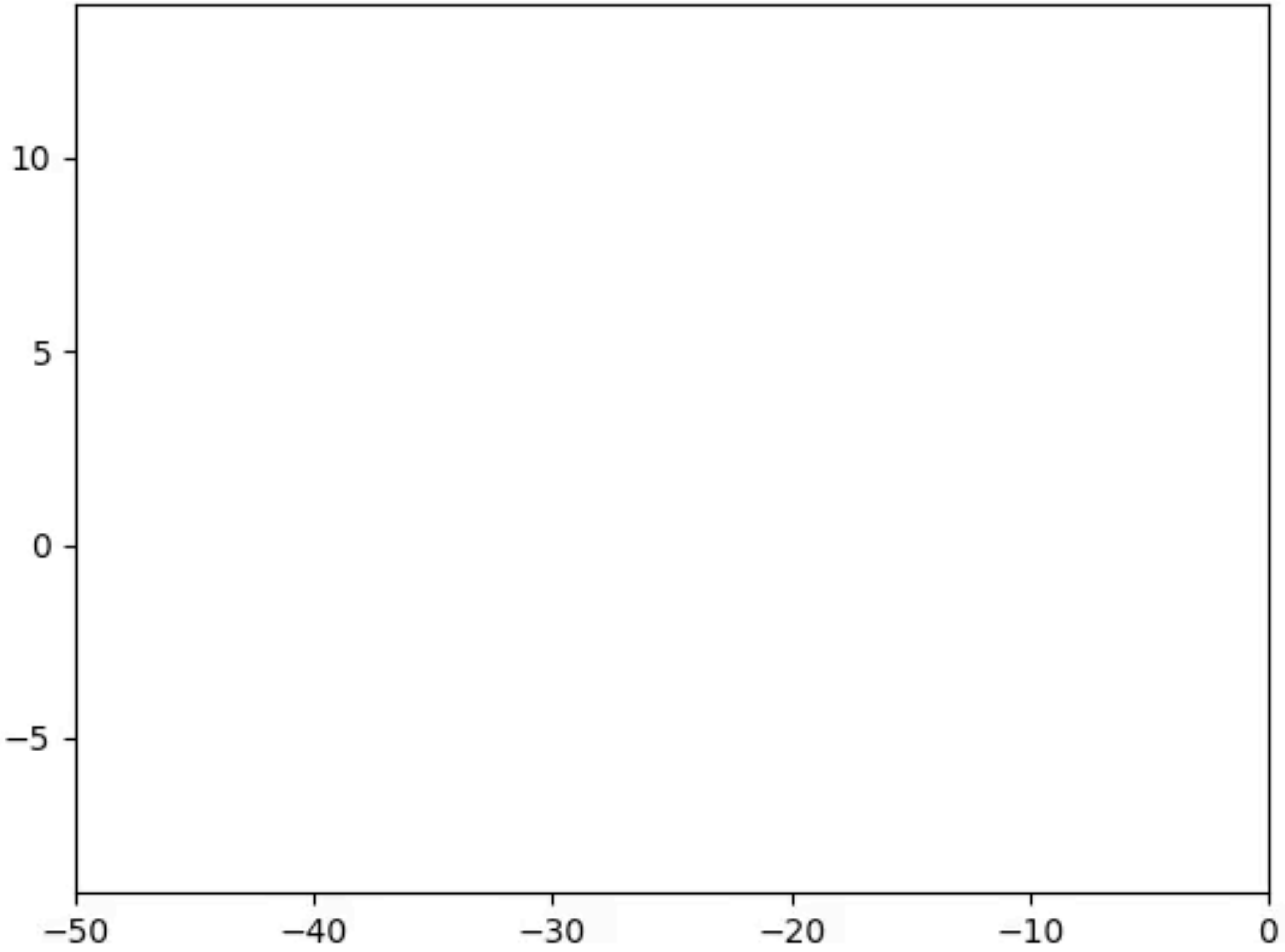
$x \sim b$



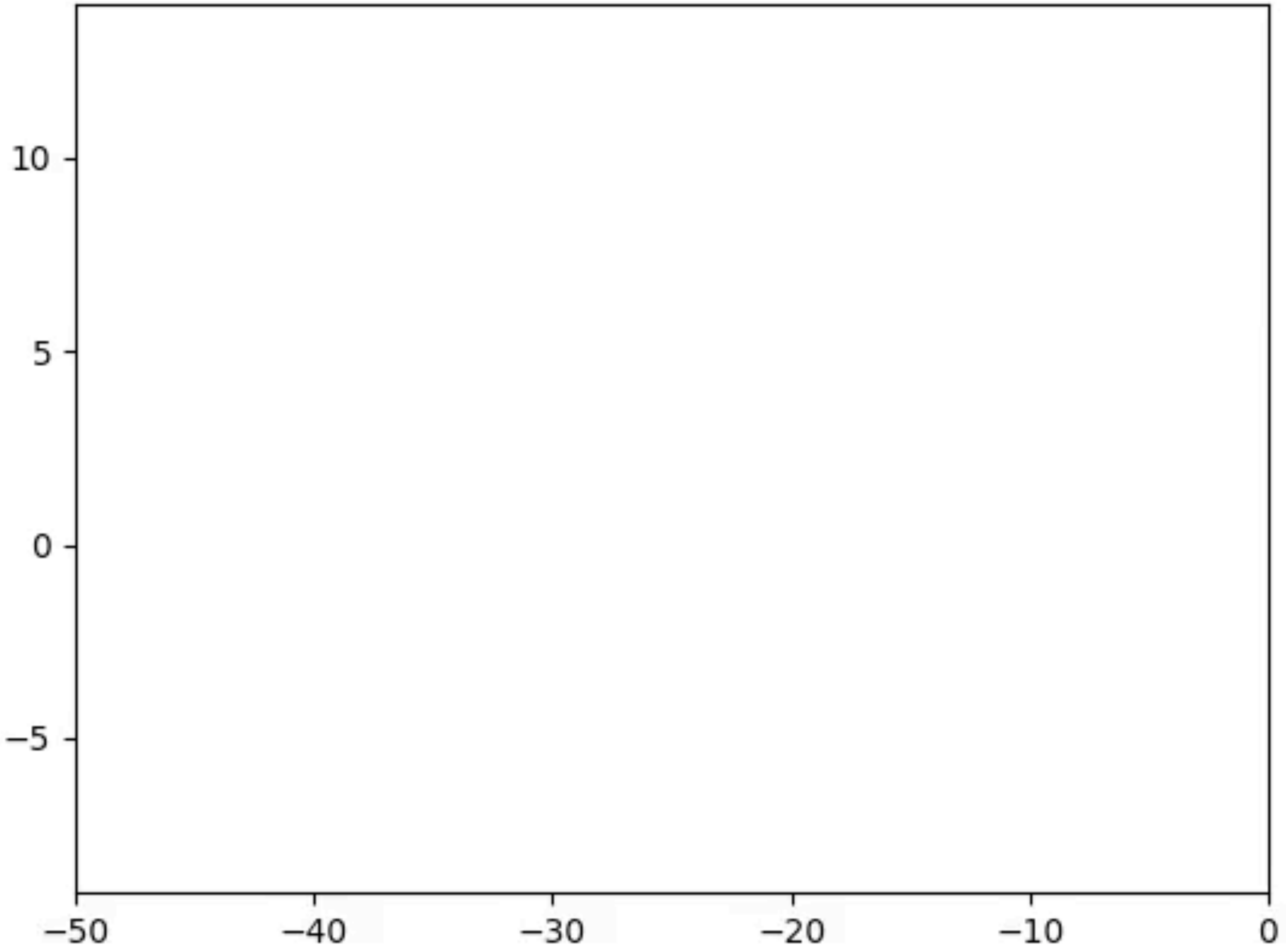
$\rho(x)x$



$x \sim b$



$\rho(x)x$



Control Variates

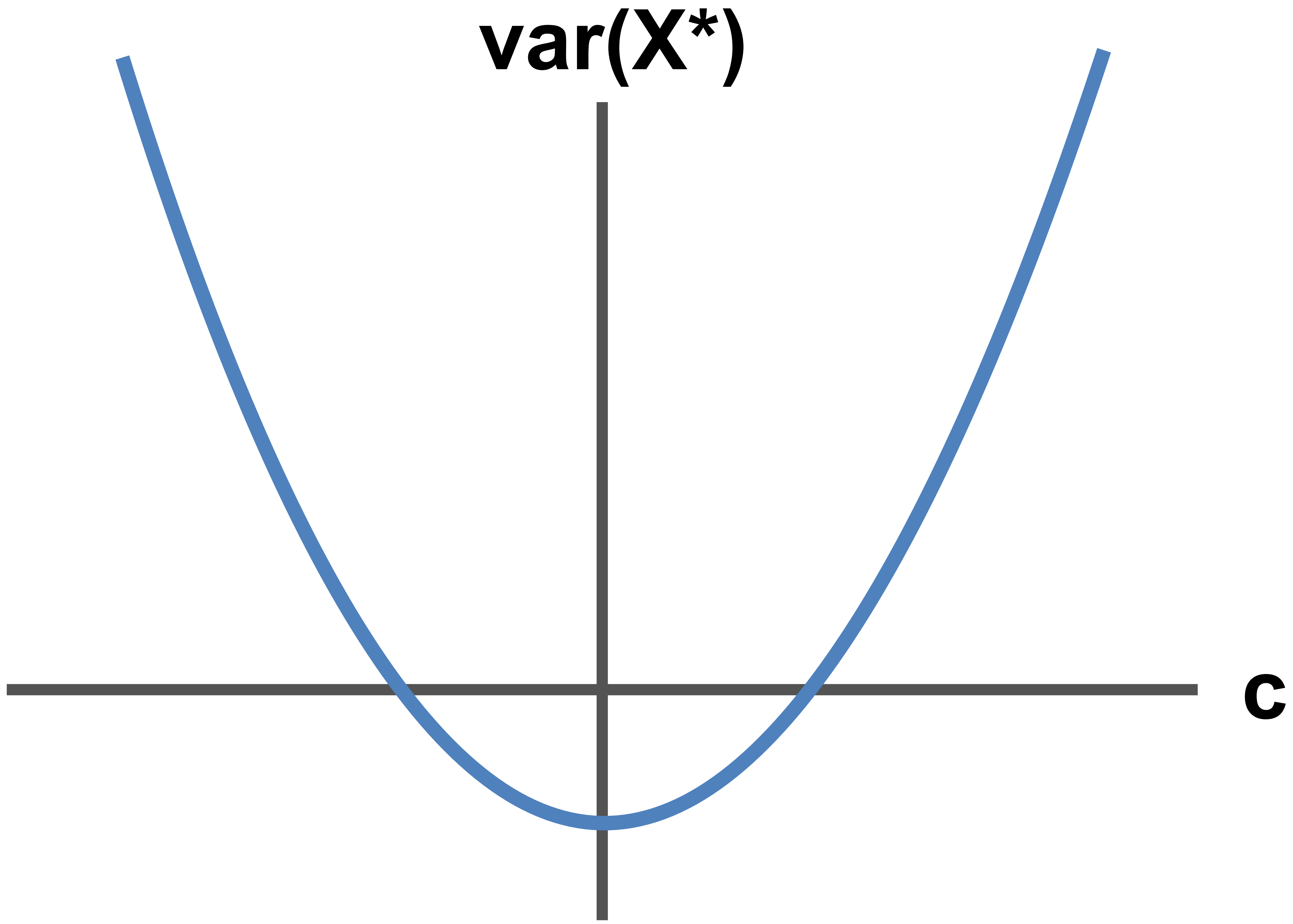
$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

Control Variates

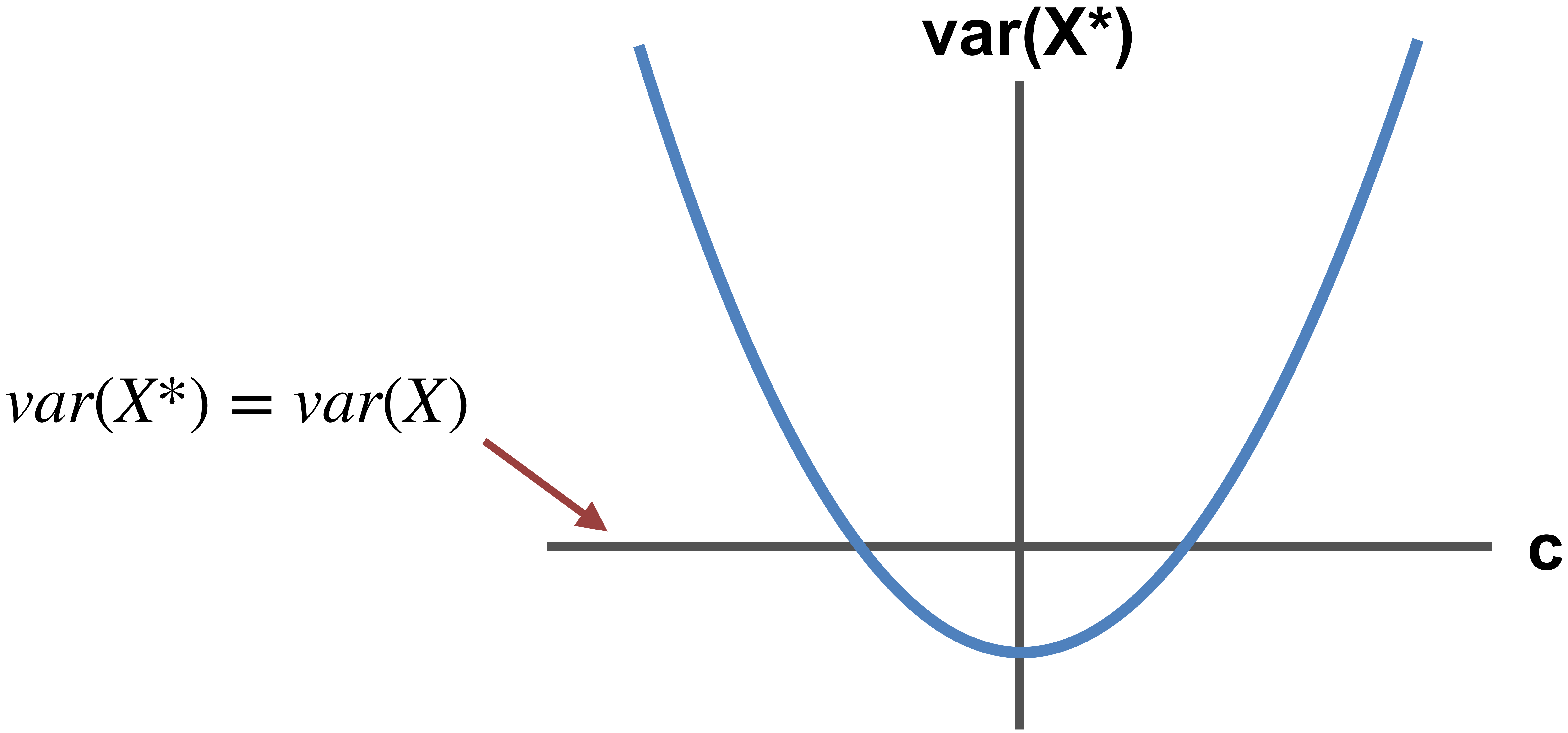
$$X^* = X + \underbrace{c(Y - \mathbb{E}_b[Y])}_{\text{Variance Control}}$$

Variance
Control

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$



$$X^* = X + c(Y - \mathbb{E}_b[Y])$$



$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\left(r + \gamma v'\right) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\left(r + \gamma v'\right) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$\delta^* = \delta + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right)$$

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\left(r + \gamma v'\right) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$\begin{aligned} \delta^* &= \delta + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \\ &= \delta + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \end{aligned}$$

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\left(r + \gamma v'\right) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$\begin{aligned} \delta^* &= \delta + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \\ &= \rho\left(r + \gamma v'\right) - v + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \end{aligned}$$

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\left(r + \gamma v'\right) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$\delta^* = \delta + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right)$$

$$= \rho\left(r + \gamma v'\right) - v + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right)$$

$$\mathbb{E}_\pi[v] = v$$

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\big(r + \gamma v'\big) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$\begin{aligned}\delta^* &= \delta + (-1)\big(\rho v - \mathbb{E}_b[\rho v]\big)\\ &= \rho\big(r + \gamma v'\big) - v + (-1)\big(\rho v - \mathbb{E}_b[\rho v]\big)\end{aligned}$$

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\big(r + \gamma v'\big) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$\begin{aligned}
\delta^* &= \delta + (-1)\big(\rho v - \mathbb{E}_b[\rho v]\big) \\
&= \rho\big(r + \gamma v'\big) - v + (-1)\big(\rho v - \mathbb{E}_b[\rho v]\big) \\
&= \rho\big(r + \gamma v'\big) - v + v - \rho v
\end{aligned}$$

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\left(r + \gamma v'\right) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$\begin{aligned}
\delta^* &= \delta + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \\
&= \rho\left(r + \gamma v'\right) - v + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \\
&= \rho\left(r + \gamma v'\right) - v + v - \rho v \\
&= \rho\left(r + \gamma v'\right) - \rho v
\end{aligned}$$

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\left(r + \gamma v'\right) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$\begin{aligned}
\delta^* &= \delta + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \\
&= \rho\left(r + \gamma v'\right) - v + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \\
&= \rho\left(r + \gamma v'\right) - v + v - \rho v \\
&= \rho\left(r + \gamma v'\right) - \rho v \\
&= \rho\left(r + \gamma v' - v\right)
\end{aligned}$$

$$X^* = X + c(Y - \mathbb{E}_b[Y])$$

$$X \doteq \rho\left(r + \gamma v'\right) - v$$

$$Y \doteq \rho v$$

$$c \doteq -1$$

$$\begin{aligned}
\delta^* &= \delta + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \\
&= \rho\left(r + \gamma v'\right) - v + (-1)\left(\rho v - \mathbb{E}_b[\rho v]\right) \\
&= \rho\left(r + \gamma v'\right) - v + v - \rho v \\
&= \rho\left(r + \gamma v'\right) - \rho v \\
&= \rho\left(r + \gamma v' - v\right) \\
&= \delta^+
\end{aligned}$$

Roadmap

- Importance Sampling
- Off-policy TD(0) isr placement
- **IS variance**
- Gradient-TD placements

Roadmap

- Importance Sampling
- Off-policy TD(0) isr placement
- IS variance
- **Gradient-TD placements**

Gradient-TD Update Equations

$$\delta = \rho_t\left[r_{t+1} + \gamma(w_t^\top x_{t+1}) - (w_t^\top x_t)\right]$$

$$z_t \leftarrow \rho_{t-1}(\gamma\lambda z_{t-1} + x_t)$$

$$h_{t+1} \leftarrow h_t + \alpha_h\left[\delta z_t - (h_t^\top x_t)x_t\right]$$

TDC

$$w_{t+1} \leftarrow w_t + \alpha\left[\delta z_t - \rho_t\gamma(1-\lambda)(h_t^\top z_t)x_{t+1}\right]$$

GTD2

$$w_{t+1} \leftarrow w_t + \alpha\left[(h_t^\top x_t)x_t - \rho_t\gamma(1-\lambda)(h_t^\top z_t)x_{t+1}\right]$$

Gradient-TD Update Equations

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1}) - (w_t^\top x_t) \right]$$

$$z_t \leftarrow \rho_{t-1}(\gamma \lambda z_{t-1} + x_t)$$

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - (h_t^\top x_t) x_t \right]$$

TDC

$$w_{t+1} \leftarrow w_t + \alpha \left[\delta z_t - \rho_t \gamma (1 - \lambda)(h_t^\top z_t) x_{t+1} \right]$$

GTD2

$$w_{t+1} \leftarrow w_t + \alpha \left[(h_t^\top x_t) x_t - \rho_t \gamma (1 - \lambda)(h_t^\top z_t) x_{t+1} \right]$$

Gradient-TD Update Equations

$$\delta = \boxed{\rho_t}\left[r_{t+1} + \gamma(w_t^\top x_{t+1}) - (w_t^\top x_t)\right]$$

$$z_t \leftarrow \rho_{t-1}(\gamma\lambda z_{t-1} + x_t)$$

$$h_{t+1} \leftarrow h_t + \alpha_h\left[\delta z_t - (h_t^\top x_t)x_t\right]$$

TDC

$$w_{t+1} \leftarrow w_t + \alpha\left[\delta z_t - \boxed{\rho_t}\gamma(1-\lambda)(h_t^\top z_t)x_{t+1}\right]$$

GTD2

$$w_{t+1} \leftarrow w_t + \alpha\left[(h_t^\top x_t)x_t - \boxed{\rho_t}\gamma(1-\lambda)(h_t^\top z_t)x_{t+1}\right]$$

Gradient-TD Update Equations

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1}) - (w_t^\top x_t) \right]$$

$$z_t \leftarrow \rho_{t-1} (\gamma \lambda z_{t-1} + x_t)$$

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - (h_t^\top x_t) x_t \right]$$

TDC

$$w_{t+1} \leftarrow w_t + \alpha \left[\delta z_t - \rho_t \gamma (1 - \lambda)(h_t^\top z_t) x_{t+1} \right]$$

GTD2

$$w_{t+1} \leftarrow w_t + \alpha \left[(h_t^\top x_t) x_t - \rho_t \gamma (1 - \lambda)(h_t^\top z_t) x_{t+1} \right]$$

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

GTD2

a: ∇_h
b: δ_h
c: ∇_w

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - (h_t^\top x_t) x_t \right]$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t \right]$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t \right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1}) - (w_t^\top x_t) \right]$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t\right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t \right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1}) \right] - (w_t^\top x_t)$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1}) - (w_t^\top x_t) \right]$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h\left[\delta z_t - \rho_t(h_t^\top x_t)x_t\right]$$

$$\delta = \rho_t\left[r_{t+1} + \gamma(w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

$$\delta = \rho_t\left[r_{t+1} + \gamma(w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \big[\delta z_t - \rho_t (h_t^\top x_t) x_t \big]$$

$$\delta = \rho_t \big[r_{t+1} + \gamma (w_t^\top x_{t+1}) \big] - (w_t^\top x_t)$$

$$\delta = \rho_t \big[r_{t+1} + \gamma (w_t^\top x_{t+1}) \big] - (w_t^\top x_t)$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

$$h_{t+1} \leftarrow h_t + \alpha_h \big[\delta z_t - (h_t^\top x_t) x_t \big]$$

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t \right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1}) \right] - (w_t^\top x_t)$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1}) \right] - (w_t^\top x_t)$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t \right]$$

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t\right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t\right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1}) - (w_t^\top x_t)\right]$$

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t\right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t\right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h\left[\delta z_t - \rho_t(h_t^\top x_t)x_t\right]$$

$$\delta = \rho_t\left[r_{t+1} + \gamma(w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

$$\delta = \rho_t\left[r_{t+1} + \gamma(w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

GTD2

a: ∇_h
b: δ_h
c: ∇_w

$$h_{t+1} \leftarrow h_t + \alpha_h\left[\delta z_t - \rho_t(h_t^\top x_t)x_t\right]$$

$$\delta = \rho_t\left[r_{t+1} + \gamma(w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

$$w_{t+1} \leftarrow w_t + \alpha\left[(h_t^\top x_t)x_t - \rho_t\gamma(1-\lambda)(h_t^\top z_t)x_{t+1}\right]$$

0: correct everything
1: correct as little as possible

TDC

a: ∇_h
b: δ_h
c: δ_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t\right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

GTD2

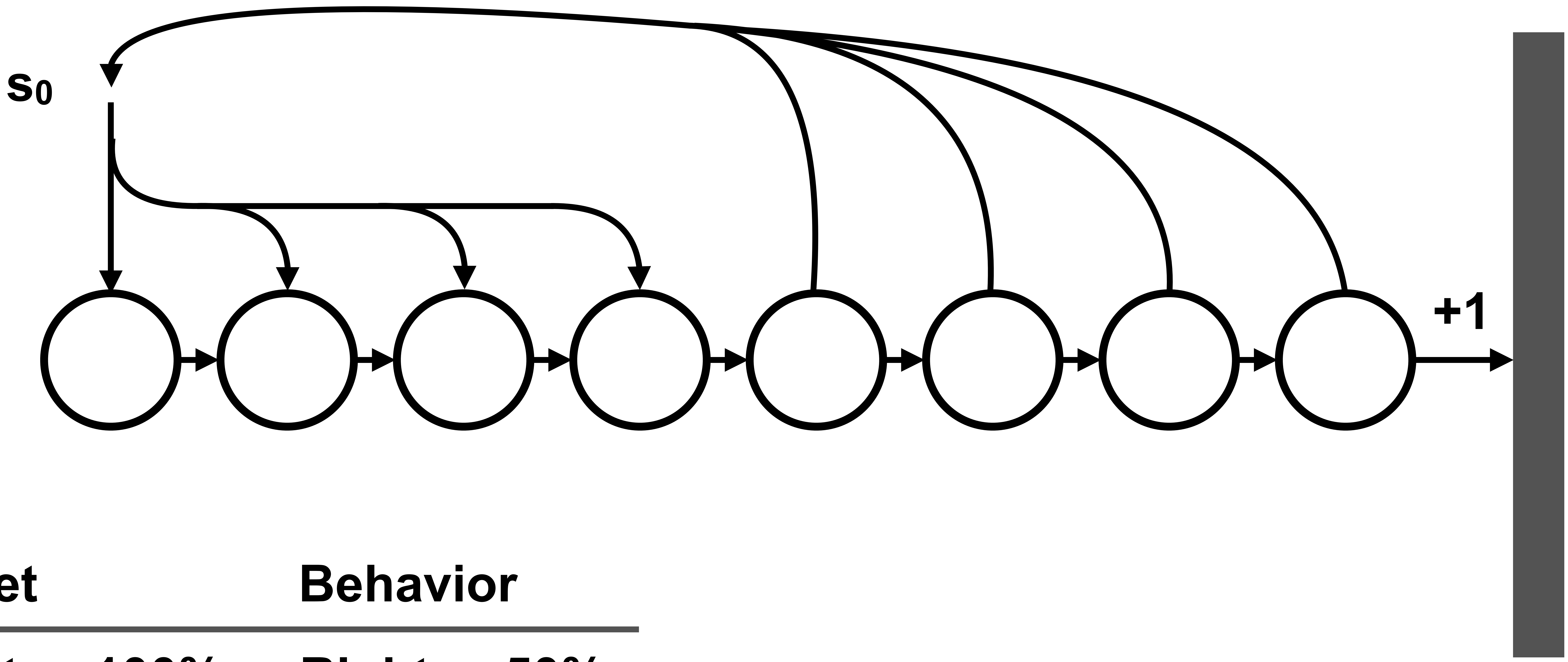
a: ∇_h
b: δ_h
c: ∇_w

$$h_{t+1} \leftarrow h_t + \alpha_h \left[\delta z_t - \rho_t (h_t^\top x_t) x_t\right]$$

$$\delta = \rho_t \left[r_{t+1} + \gamma (w_t^\top x_{t+1})\right] - (w_t^\top x_t)$$

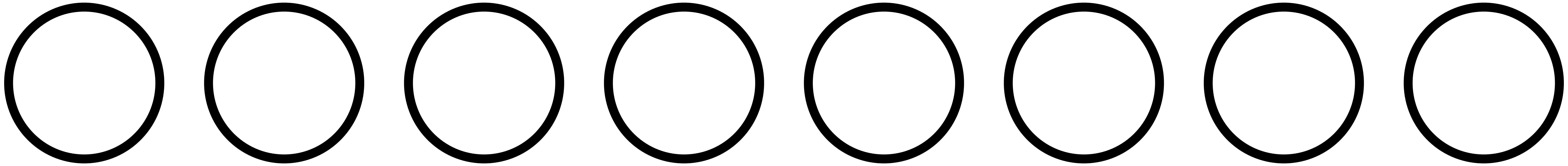
$$w_{t+1} \leftarrow w_t + \alpha \left[\rho_t (h_t^\top x_t) x_t - \rho_t \gamma (1 - \lambda)(h_t^\top z_t) x_{t+1}\right]$$

Collision



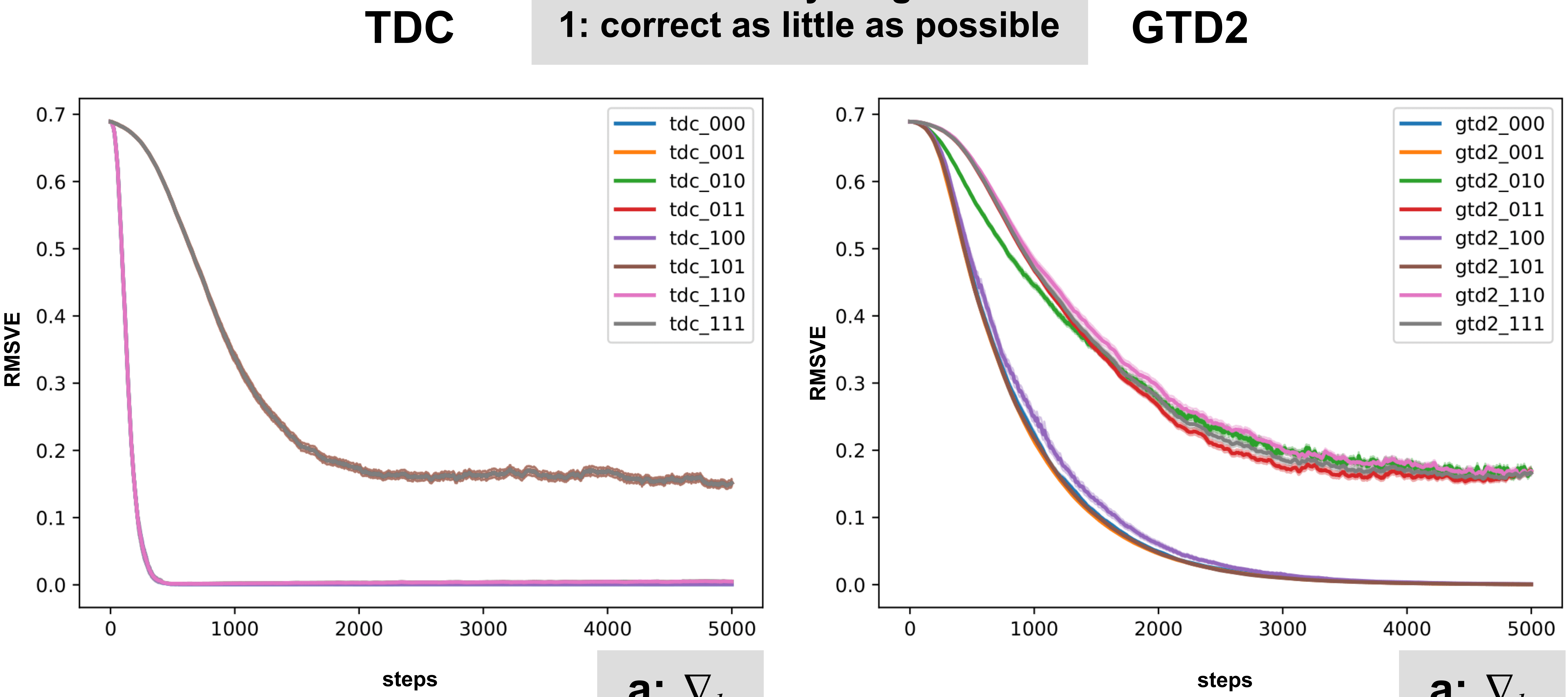
Target		Behavior	
Right:	100%	Right:	50%
Retreat:	0%	Retreat:	50%

Tabular Collision



1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	1	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	0	0	1	0	0	0
0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	1

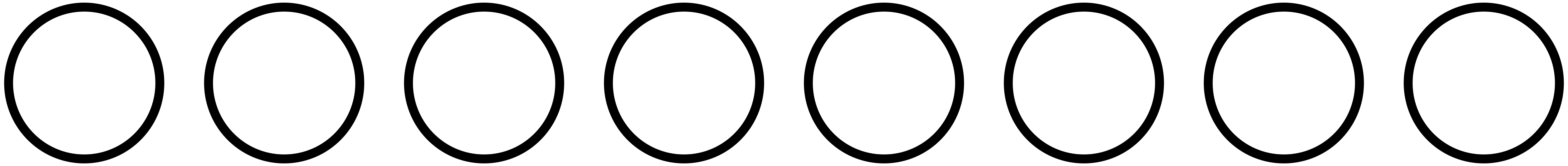
0: correct everything
1: correct as little as possible



a: ∇_h
b: δ_h
c: δ_w

a: ∇_h
b: δ_h
c: ∇_w

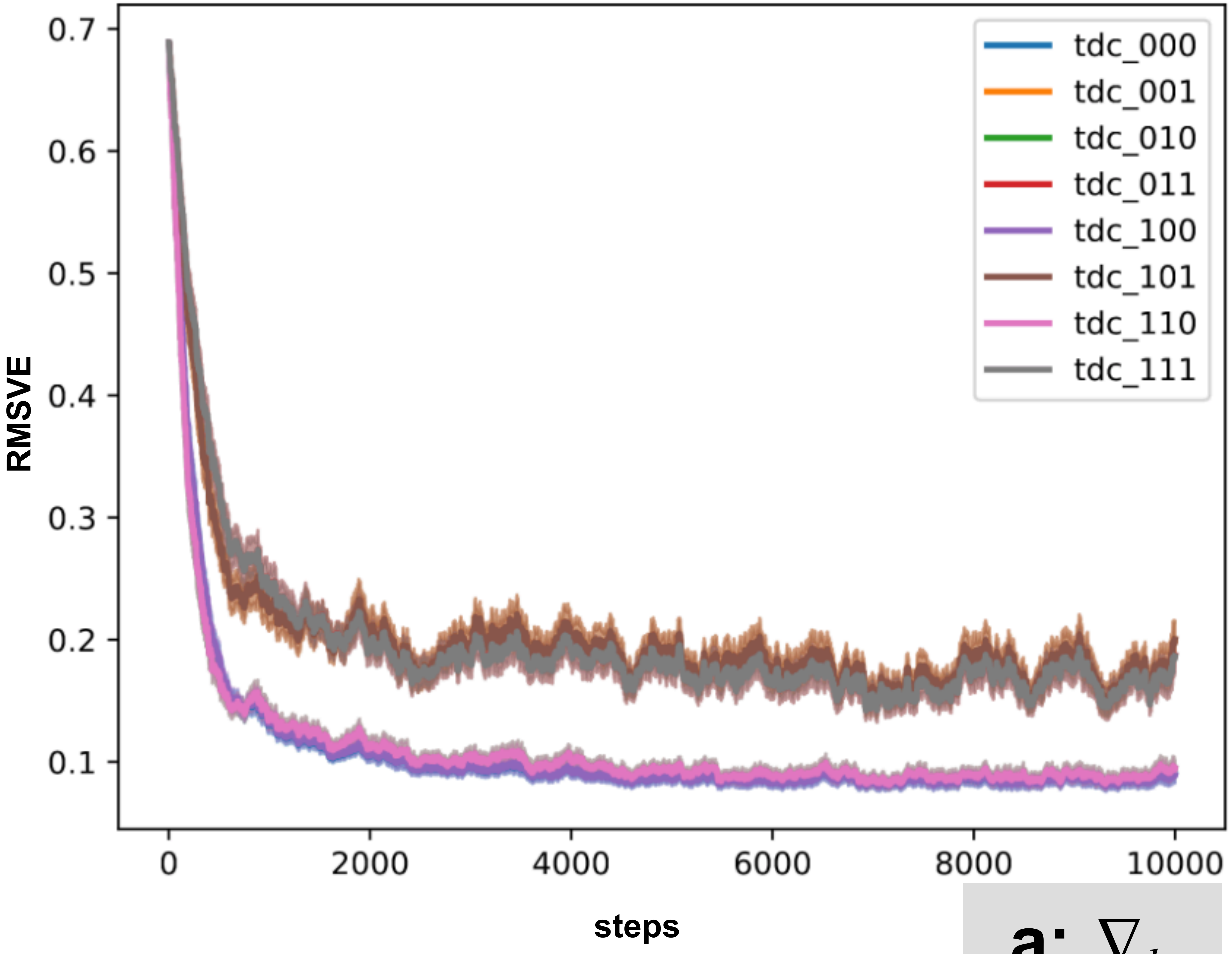
Binary Encoder Collision



1	0	0	1	0	1	1	1
0	0	1	0	1	0	1	0
1	1	1	0	0	1	0	1
1	1	0	1	0	0	0	0
0	0	1	1	1	0	1	1
0	1	0	0	1	1	0	0

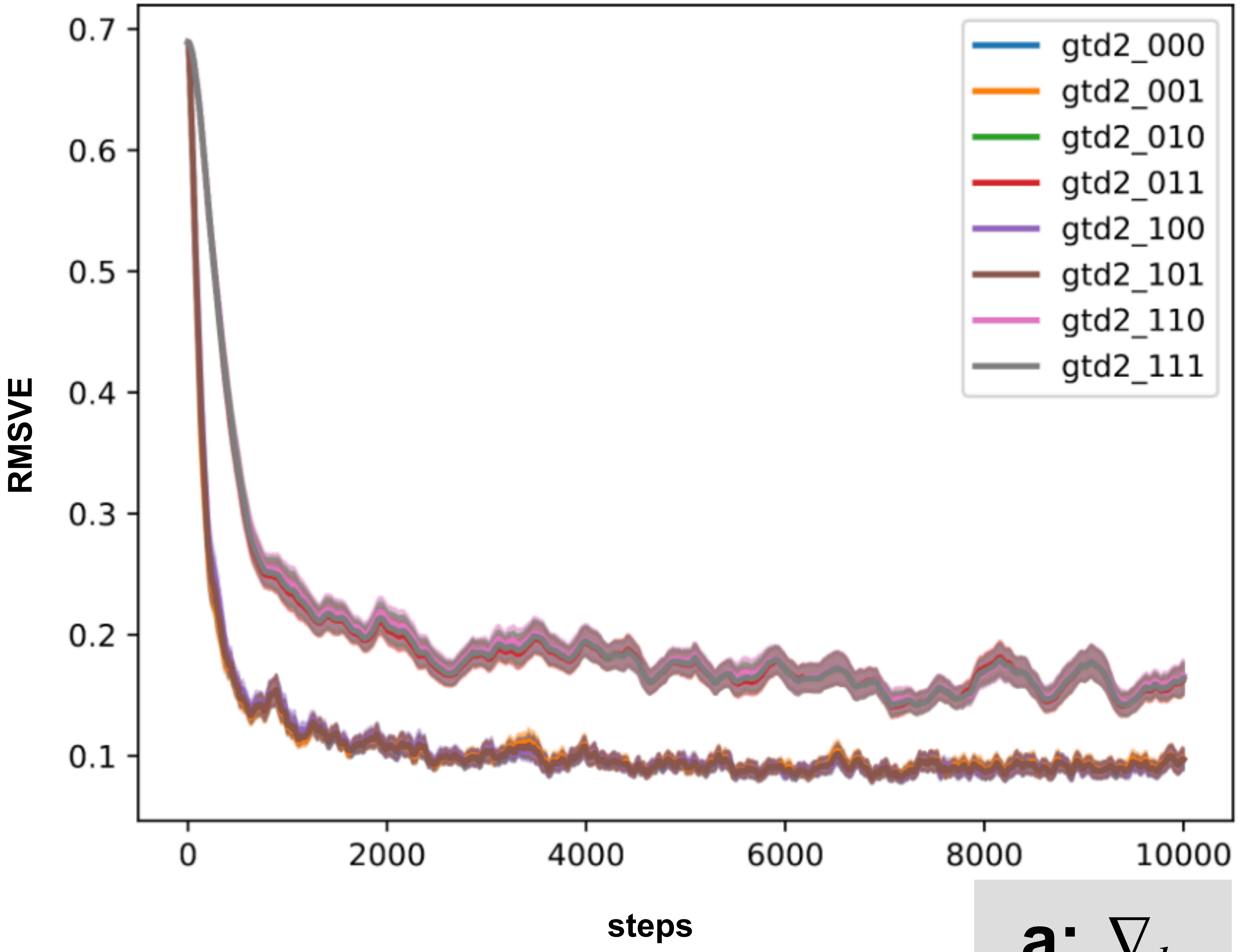
0: correct everything
1: correct as little as possible

TDC



a: ∇_h
b: δ_h
c: δ_w

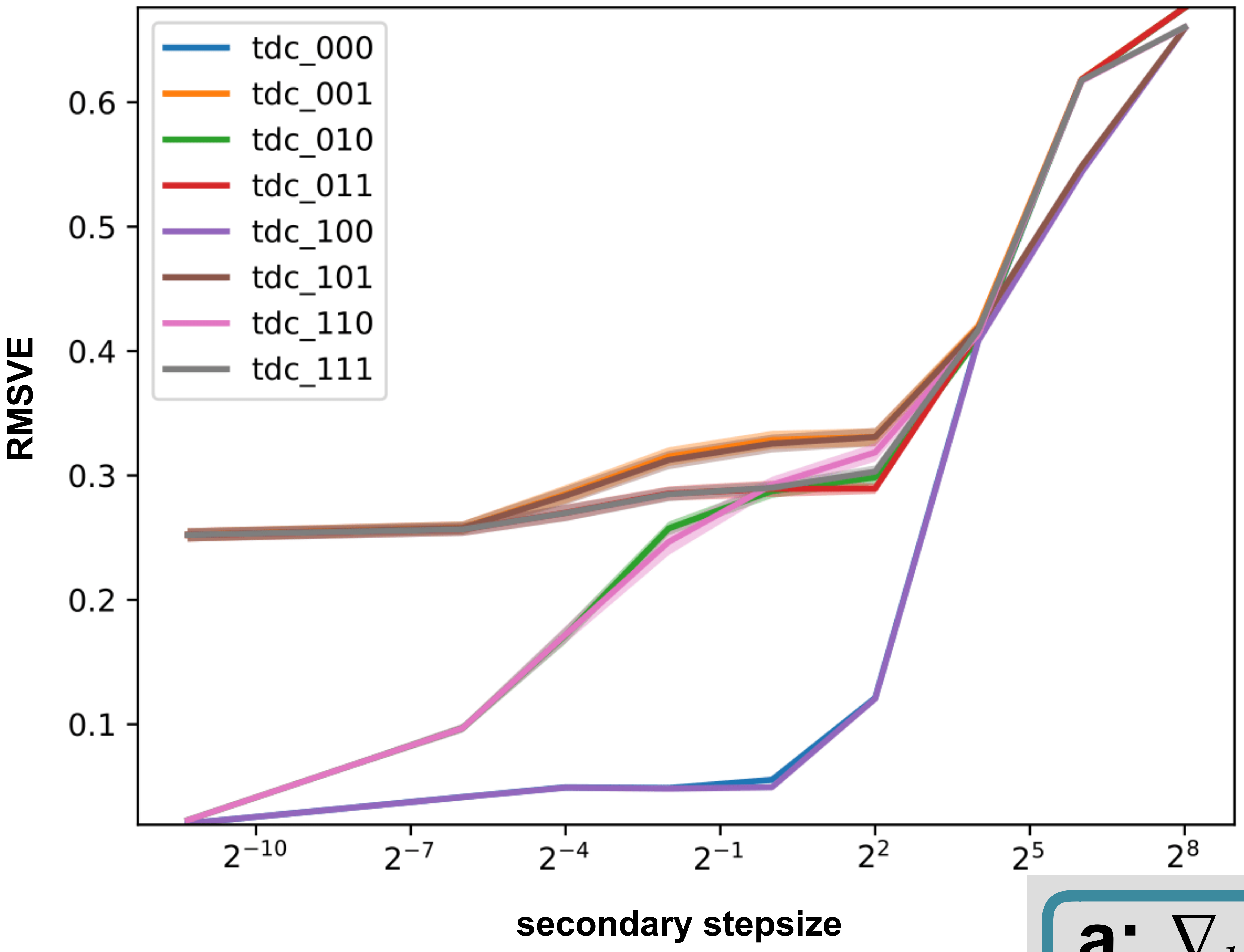
GTD2



a: ∇_h
b: δ_h
c: ∇_w

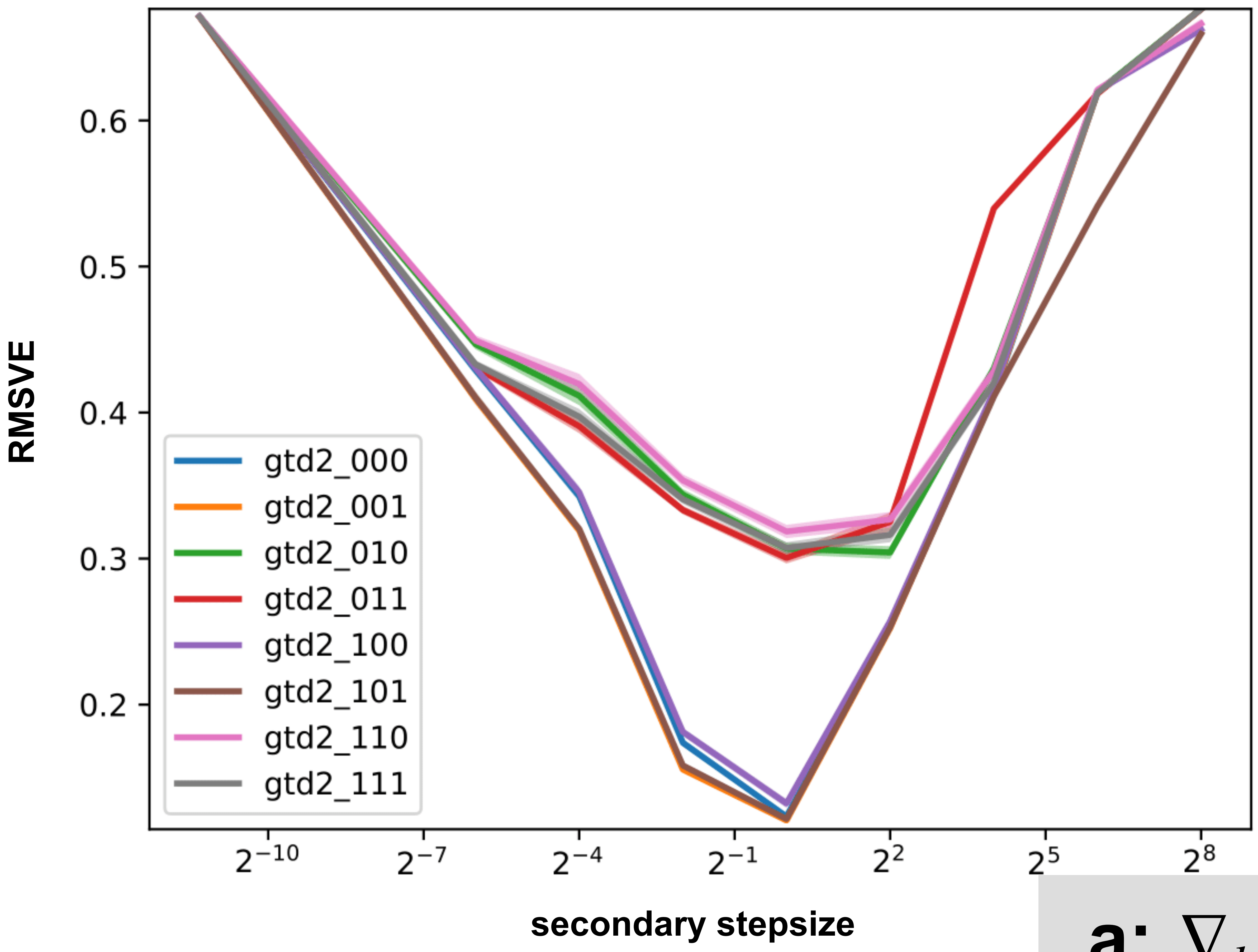
0: correct everything
1: correct as little as possible

TDC



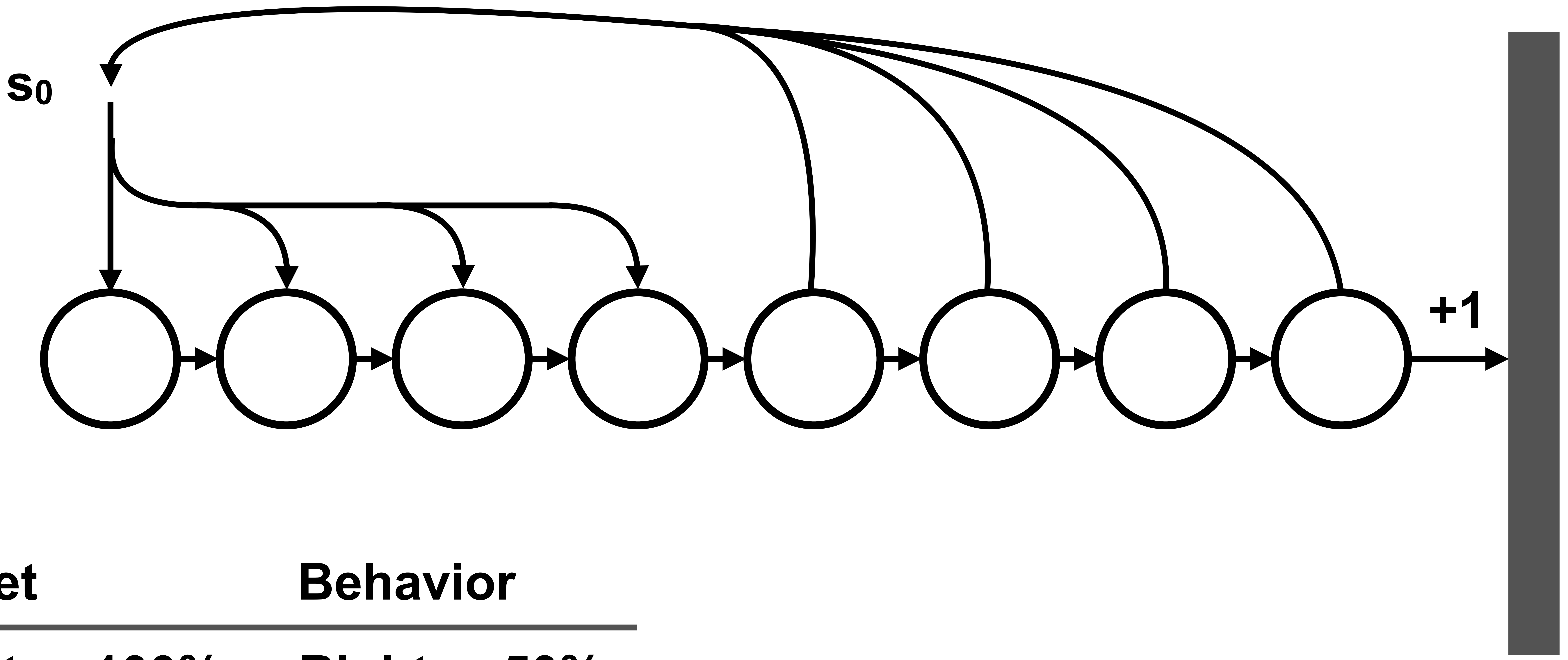
a: ∇_h
b: δ_h
c: δ_w

GTD2



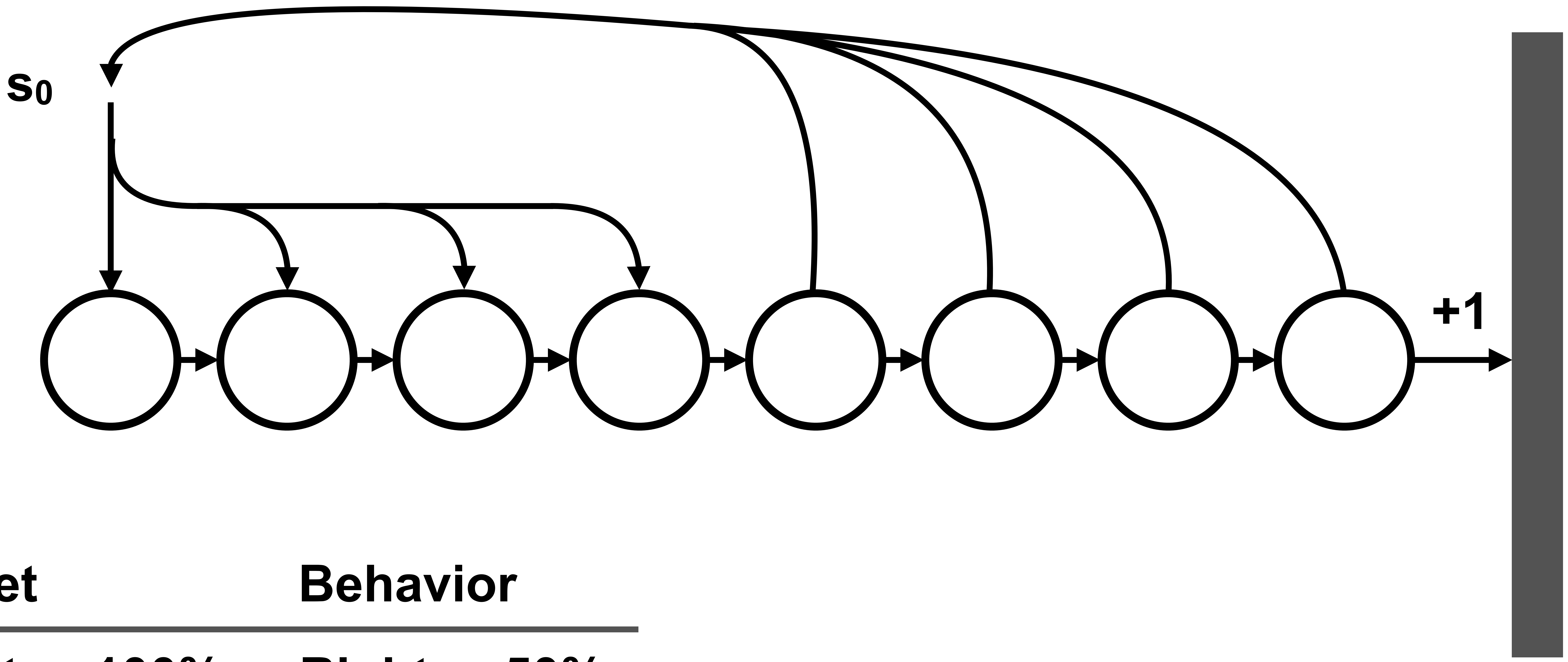
a: ∇_h
b: δ_h
c: ∇_w

Collision



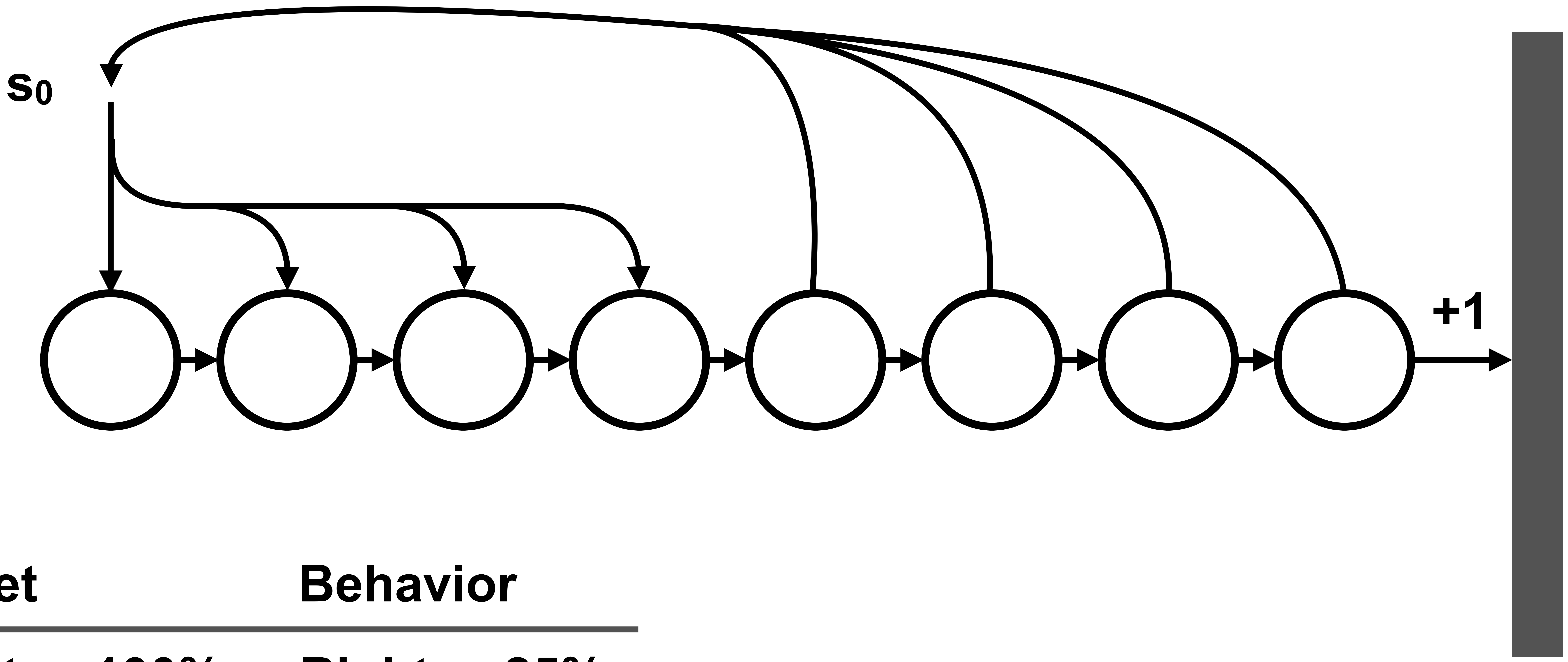
Target		Behavior	
Right:	100%	Right:	50%
Retreat:	0%	Retreat:	50%

Collision



Target		Behavior	
Right:	100%	Right:	50%
Retreat:	0%	Retreat:	50%

Collision

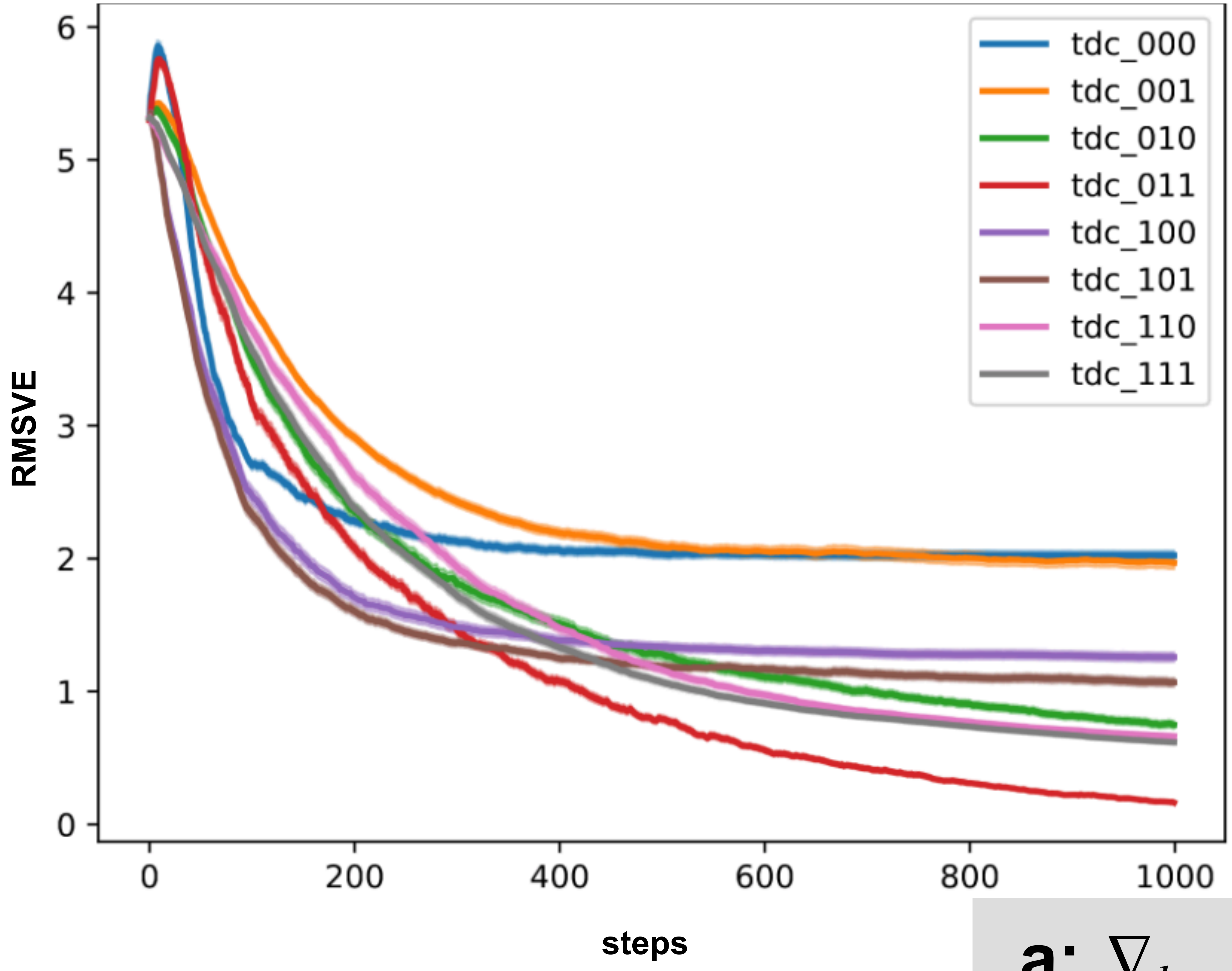


Target		Behavior	
Right:	100%	Right:	25%
Retreat:	0%	Retreat:	75%


0: correct everything
1: correct as little as possible
TDC
GTD2
tdc_000
tdc_001
tdc_010
tdc_011
tdc_100
tdc_101
tdc_110
tdc_111
RMSVE
steps
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5000
10000
15000
20000
25000
30000
35000
40000
gtd2_000
gtd2_001
gtd2_010
gtd2_011
gtd2_100
gtd2_101
gtd2_110
gtd2_111
a: ∇_h
b: δ_h
c: δ_w
a: ∇_h
b: δ_h
c: ∇_w

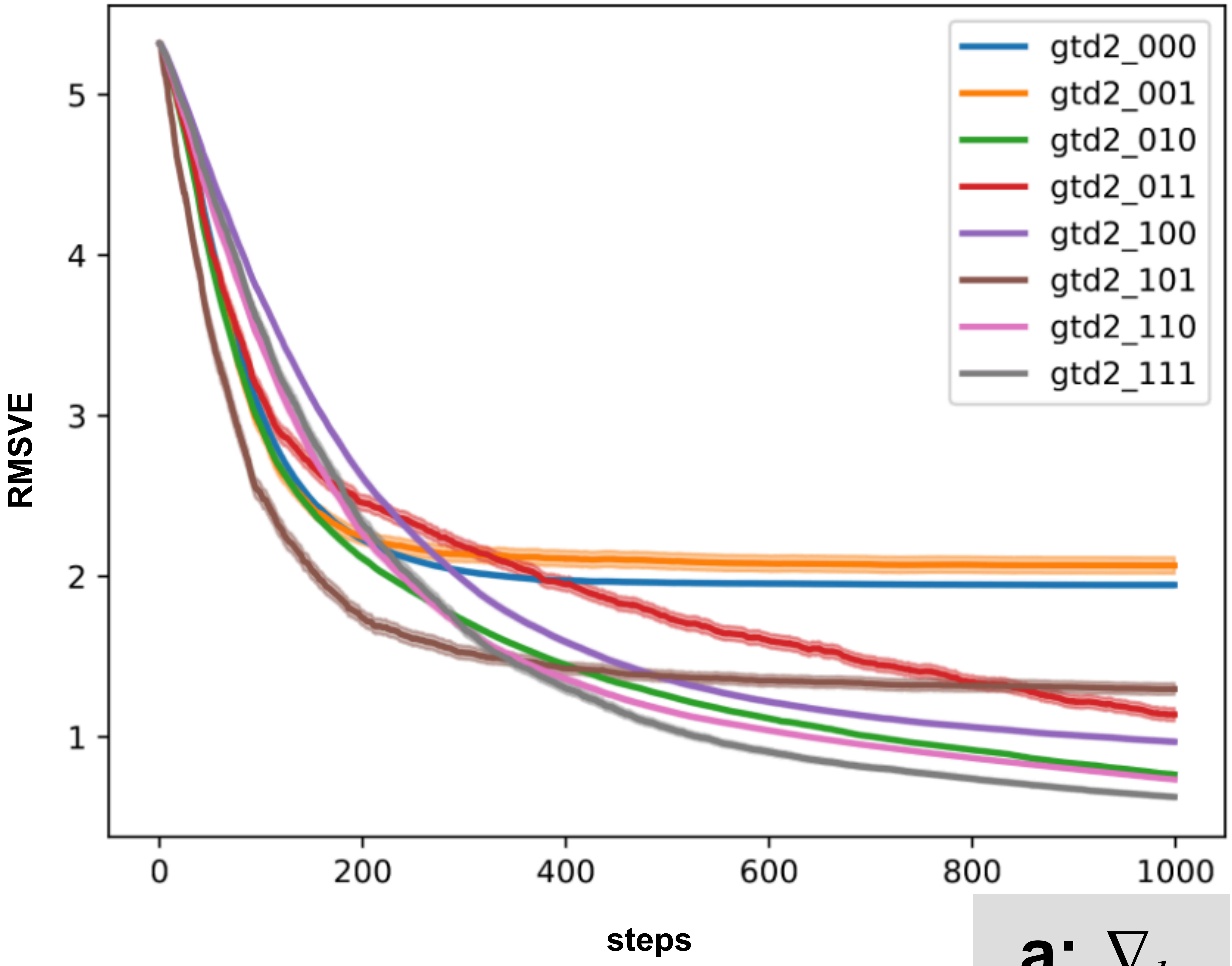
0: correct everything
1: correct as little as possible

TDC



a: ∇_h
b: δ_h
c: δ_w

GTD2



a: ∇_h
b: δ_h
c: ∇_w

Baird's Counterexample

Baird, L. C. (1995).

Thanks for your time