Tilted Empirical Risk Minimization

Tian Li*
CMU

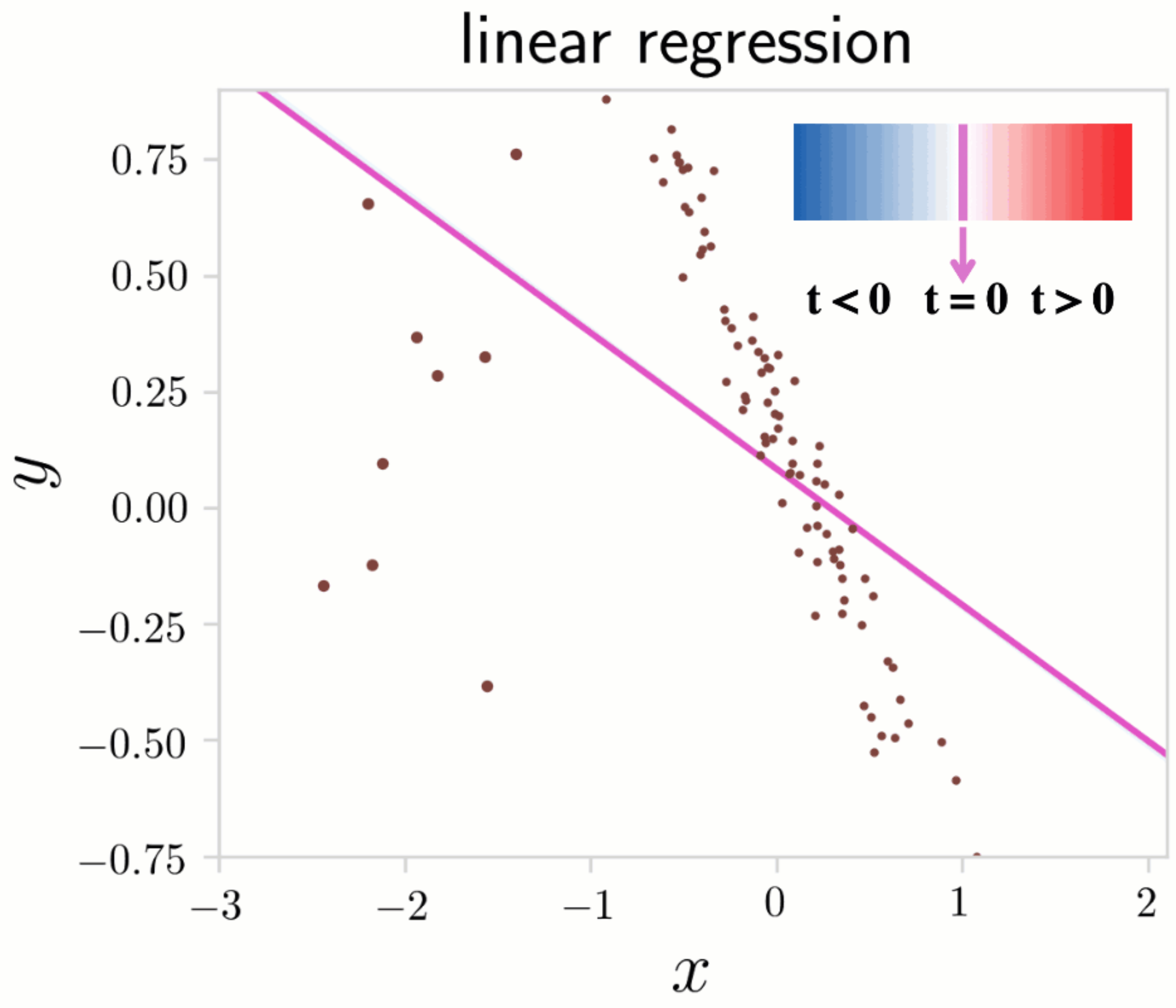
Ahmad Beirami*
Facebook AI

Maziar Sanjabi
Facebook AI

Virginia Smith
CMU

ICLR 2021

Tilted ERM (TERM) Objective



Empirical Risk Minimization

$$\min_{w} \frac{1}{n} \sum_{i=1}^{n} f(x_i; w)$$

Tilted ERM

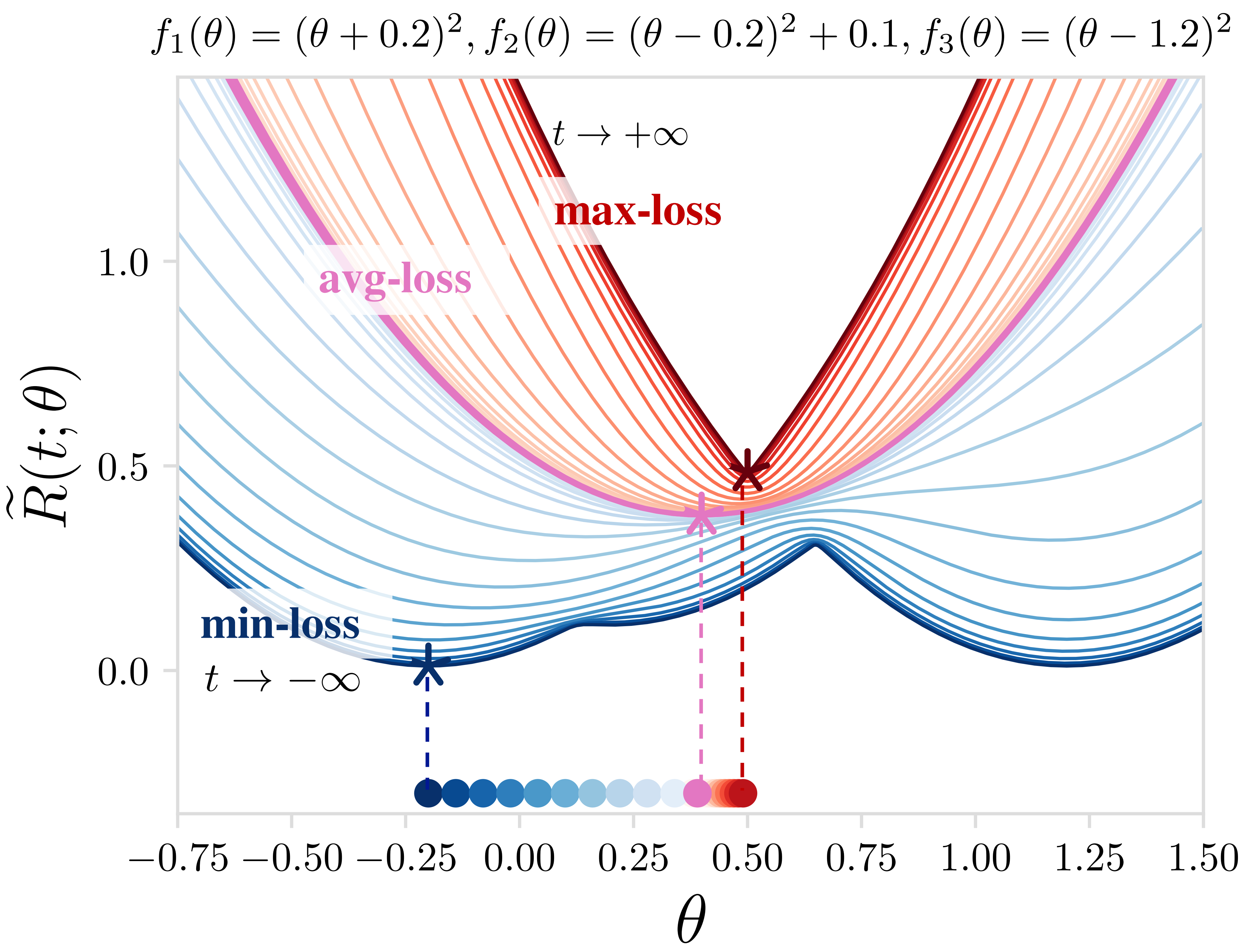
$$\min_{w} \frac{1}{t} \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{t f(x_i; w)} \right)$$

TERM can increase or decrease the influence of outliers to enable fairness or robustness

Tilted ERM (TERM) Objective

$$\tilde{R}(t;\theta) := \frac{1}{t}\log\left(\frac{1}{n}\sum_{i=1}^{n} e^{tf(x_i;w)}\right)$$

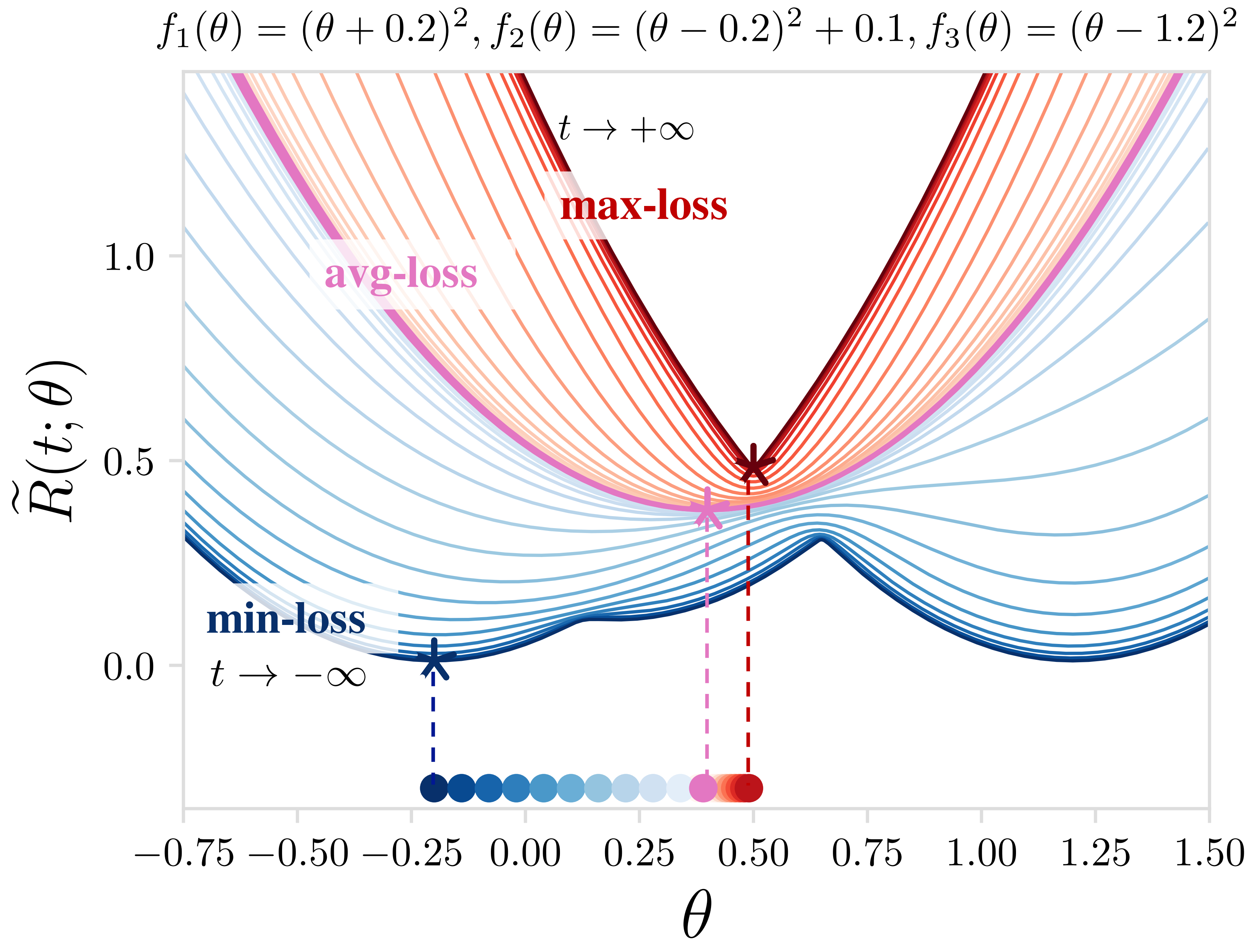
- recovers a family of objectives parameterized by t
- a smooth transition from **min-loss** to **avg-loss** to **max-loss**



Properties: Trade-off between average loss and max-/min-loss

positive t: as t increases, the average loss will increase, and the max-loss will decrease and the loss variance will decrease => better generalization

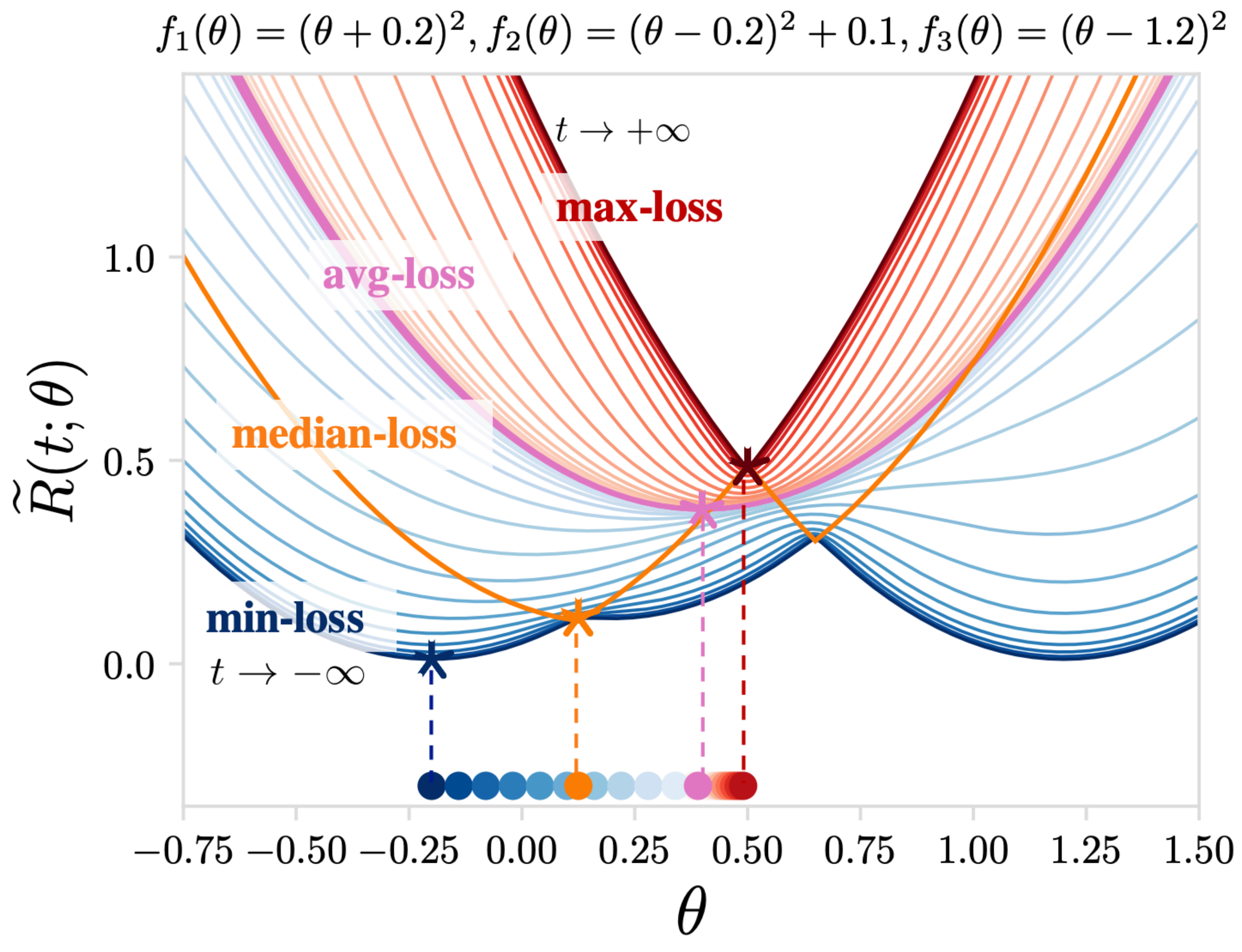
negative t: as t increases, the average loss will decrease, and the min-loss will increase

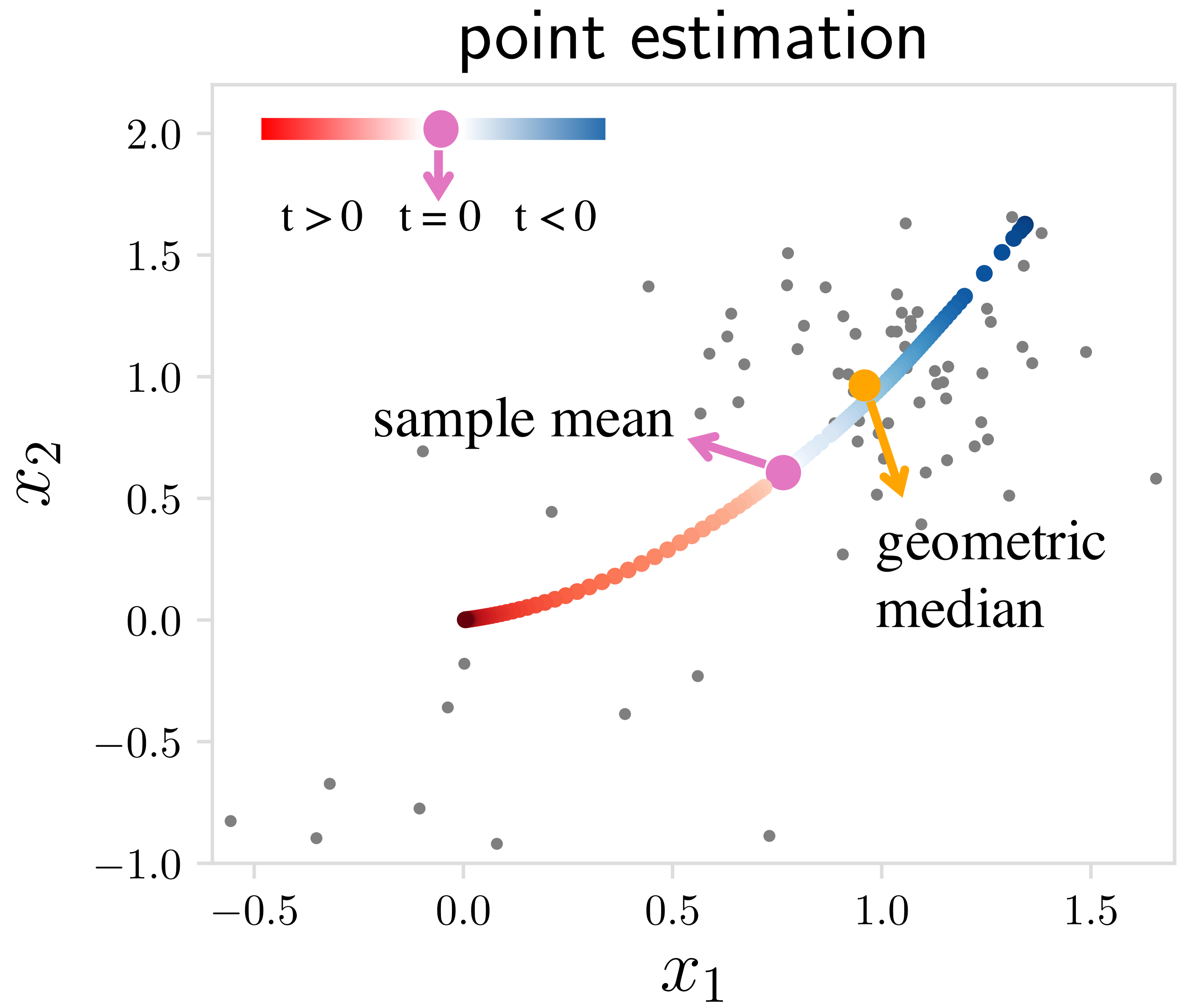


Properties: Approximation of quantile losses

k-th quantile losses: k-th largest individual loss from $\{f(x_i;\theta)\}_{i\in[N]}$

e.g., median loss ($k = N/2$)





TERM solutions can approximate k-loss solutions ($1 \leq k \leq N$)

TERM solvers

TERM can be solved with a simple modification to batch/stochastic ERM solvers

1) batch case

$$\nabla_\theta \tilde{R} = \sum_{i=1}^{N} w_i(t;\theta)\,\nabla_\theta f(x_i;\theta), \quad w_i(t;\theta) = \frac{e^{tf(x_i;\theta)}}{\sum_{j\in[N]} e^{tf(x_j;\theta)}}$$

✓ convergence rate scales linearly with t

2) stochastic case have some stochastic dynamics to estimate the normalizer of the weights

TERM is widely applicable to a broad range of ML problems

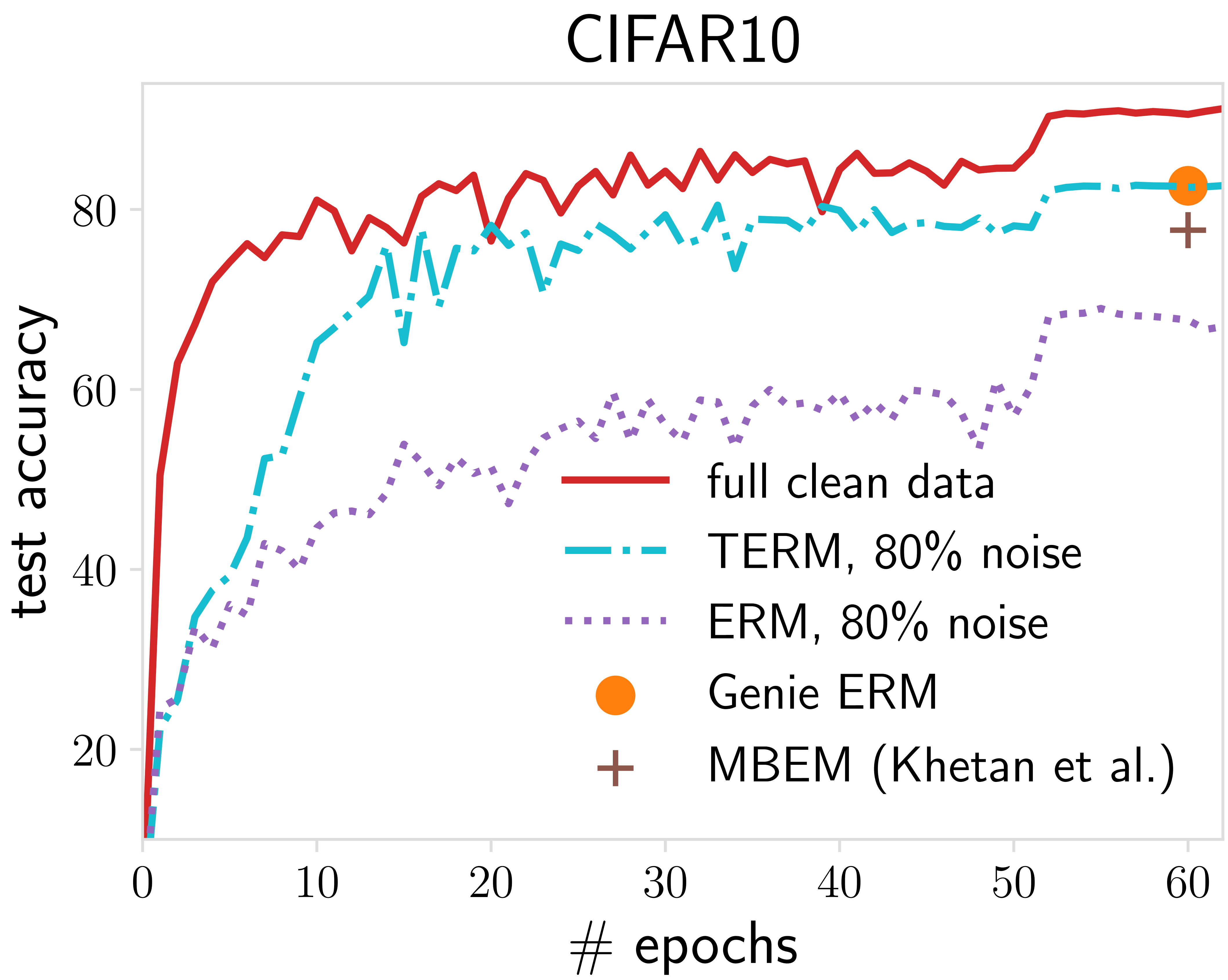
$t < 0$	Outlier Mitigation Robust Regression/Classification
$t > 0$	Class Imbalance Fair PCA Variance Reduction
$t_1 < 0$ $t_2 > 0$	Fairness + Robustness

, and many more

Competitive/Superior performance compared with application-specific approaches

E.g., TERM applied to Robust Classification ($t < 0$)

noisy annotators in for crowdsourcing



TERM is able to completely remove the noisy outliers, achieving the accuracy of Genie ERM

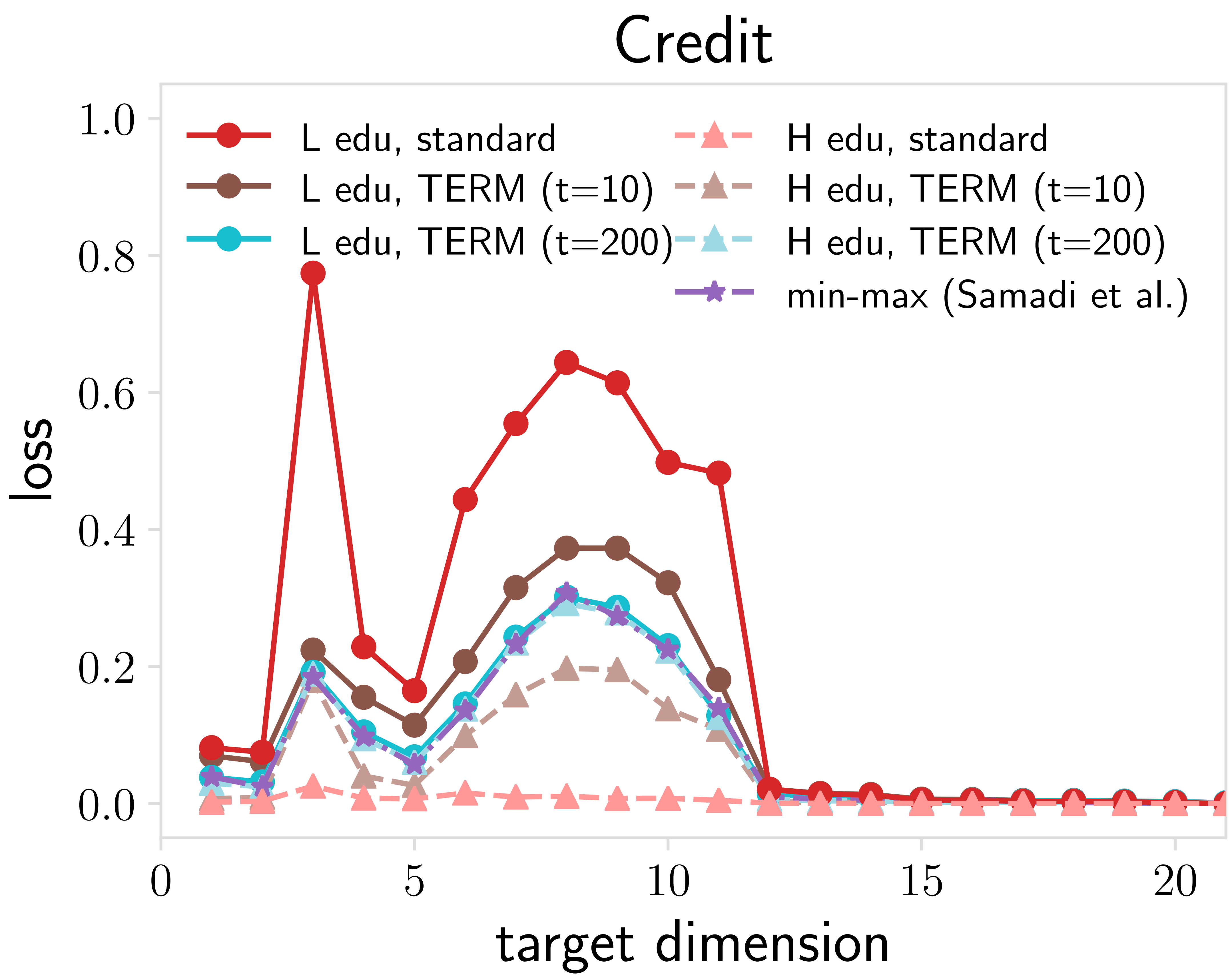
E.g., TERM applied to Fair PCA ($t > 0$)

Goal of fair PCA:

low-dimension features

$$loss(L; U) \approx loss(H; U)$$

two groups



TERM can recover the min-max solution with a large t

also offer more flexible tradeoffs between performance and fairness

Future Work

- Other applications and properties of the TERM framework
- Generalization guarantees of the TERM objective with respect to t
- Further connections with other risks (DRO, CVaR, IRM, etc)

Paper: OpenReview website

Code: *https://github.com/litian96/TERM*