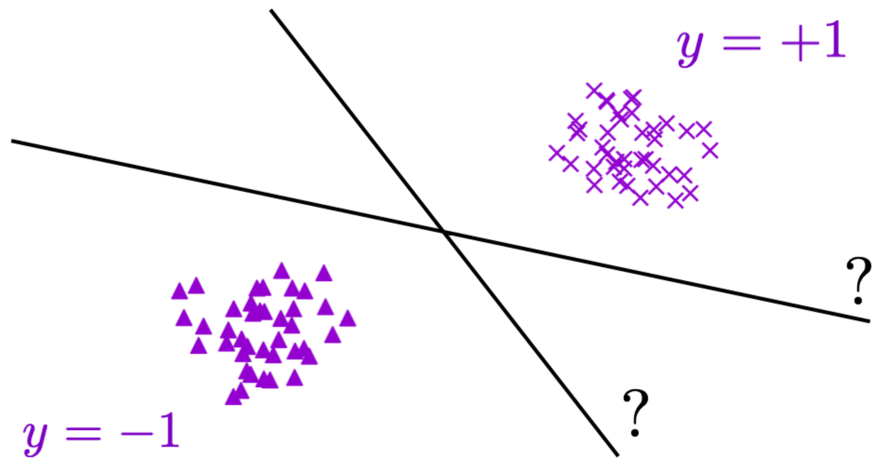
# Implicit Bias of Mirror Flow on Separable Data

Scott Pesme (EPFL)
Radu Dragomir (Télécom Paris)
Nicolas Flammarion (EPFL)

## Setup: logistic regression

$$\min_\beta L(\beta) = \sum_{i=1}^{n} \ln\left(1 + e^{-y_i\langle \beta, x_i\rangle}\right)$$




**Assumption:** linearly separable data → the loss is minimised *at infinity*

$$\lim_{s\to\infty} L(s\beta^*) = 0 \quad \text{for} \quad \beta^* \in S$$

*set of vectors defining separating hyperplanes*

**For a given algorithm, what is the directional limit $\lim_{t\to\infty} \frac{\beta_t}{\|\beta_t\|}$ of the iterates $\beta_t$ ?**

Many possible solutions in $S$: **which one is preferred by the method ?**

Is it one with good generalization properties ? *(implicit regularization)*

## The method: mirror flow

$$\dot\beta_t = -\nabla^2\phi(\beta_t)^{-1}\nabla L(\beta_t)$$

*strictly convex potential*

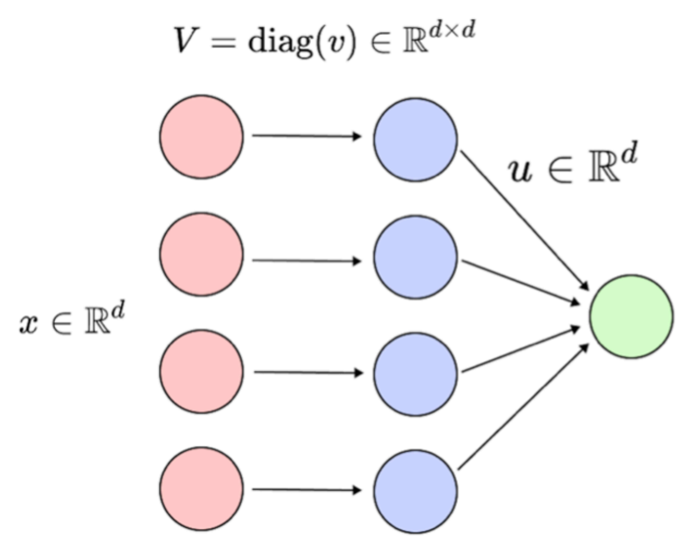
**Motivation:** *reparametrized* problems $\beta = F(\theta)$. **Under some conditions:**

**Gradient flow on $\theta \mapsto L(F(\theta))$ $\iff$ Mirror flow on $\beta \mapsto L(\beta)$**

**Example:** $\beta = F(u, v) = u \odot v$ *"diagonal neural networks"*

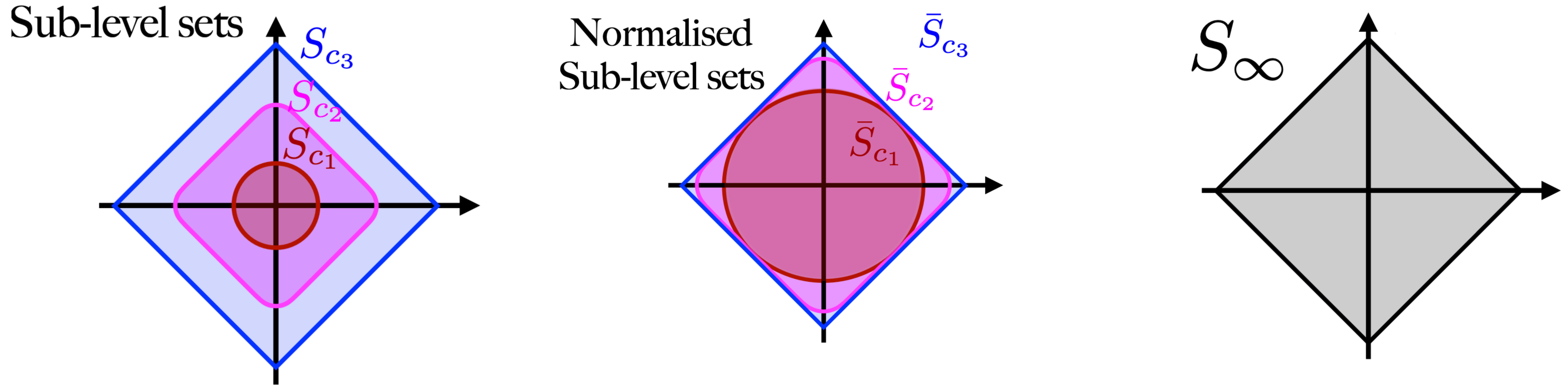
Gradient flow on $L(u \odot v)$ $\iff$ Mirror flow on $L(\beta)$ with hyperbolic potential

$$\phi(\beta) = \sum_{i=1}^{d}\left(\beta_i \operatorname{arcsinh}(\beta_i) - \sqrt{\beta_i^2 + 1}\right)$$




## Horizon function

**How to define the geometry of $\phi$ "at infinity" ?**




$S_c = \{\beta : \phi(\beta) \le c\}$ $\qquad$ $\bar S_c = S_c / \max_{\beta\in S_c}\|\beta\|$ $\qquad$ $S_\infty = \lim_{c\to\infty} \bar S_c$

We say that $\phi$ **admits a horizon** if $\lim_{c\to\infty} \bar S_c$ exists

**Horizon function:** $\phi_\infty(\beta) = \inf\{r > 0 : \frac{\beta}{r} \in S_\infty\}$

*Minkowski gauge of $S_\infty$: norm-like function whose level sets are $\propto S_\infty$*

**Theorem:** if $\phi$ is **tame**, it admits a horizon.

*e.g. polynomial, semialgebraic, subanalytic, log-exp…*

**Key property:** if $\|\beta_t\| \to \infty$ with $\frac{\beta_t}{\|\beta_t\|} \to \bar\beta$,

$$\frac{\nabla\phi(\beta_t)}{\|\nabla\phi(\beta_t)\|} \to \lambda \bar g \quad \text{with } \bar g \in \partial\phi_\infty(\bar\beta)$$

*$\phi_\infty$ is possibly nonsmooth*

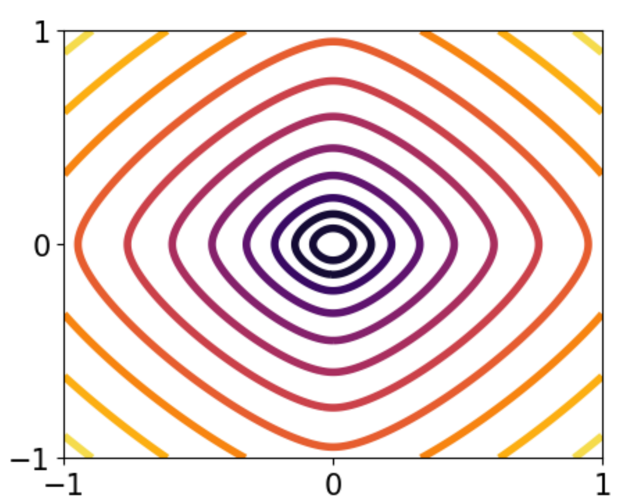
## Main result

**Theorem:** the mirror flow iterates $\beta_t$ converge **in direction** towards $\bar\beta$ satisfying the $\phi_\infty$-max margin problem

$$\bar\beta \propto \operatorname{argmin}\left\{\phi_\infty(\beta^*) : \beta^* \in S\right\}$$

*horizon function of mirror potential $\phi$* $\qquad$ *set of separating hyperplanes*

**Application:** hyperbolic potential




$$\phi_\infty \propto \|\cdot\|_1$$

Implicit bias towards **sparsity** in diagonal neural networks

(**known result**, more general proof)