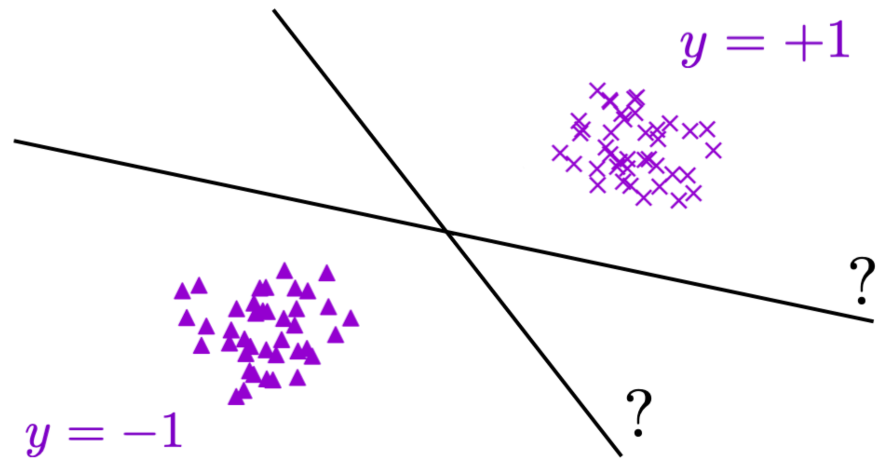
Implicit Bias of Mirror Flow on Separable Data

Scott Pesme (Inria Grenoble)
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':

$$\lim_{s\to\infty} L(s\beta^\star) = 0 \quad \text{for } \beta^\star \in S := \{\beta^\star \in \mathbb{R}^d, y_i\langle\beta^\star, x_i\rangle \geq 1, \forall i\}$$

set of vectors defining separating hyperplanes

For mirror flow, what is the directional limit $\lim_{t\to\infty} \frac{\beta_t}{\|\beta_t\|}$ of the iterates β_t ?

Many possible solutions in S: which one is preferred by the method? *(implicit regularisation)*

This question is important to understand the generalisation properties!

The gradient method: mirror flow

$$\mathrm{d}\nabla\phi(\beta_t) = -\nabla L(\beta_t)\mathrm{d}t$$

strictly convex potential

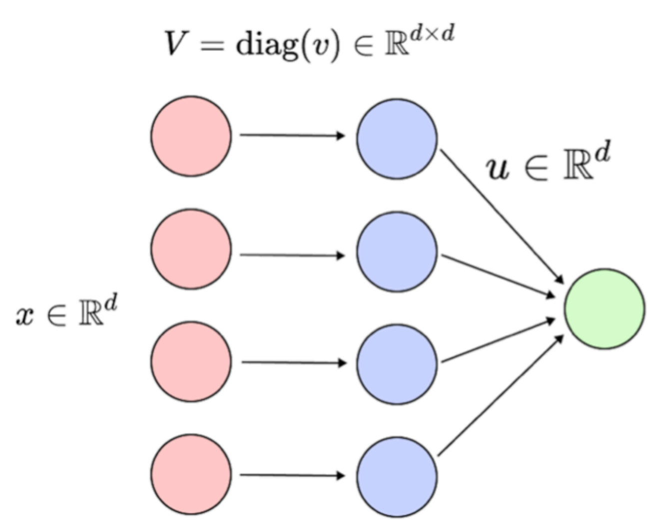
Motivation: reparametrisation $\beta = F(\theta)$, then under some (restrictive) conditions:

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

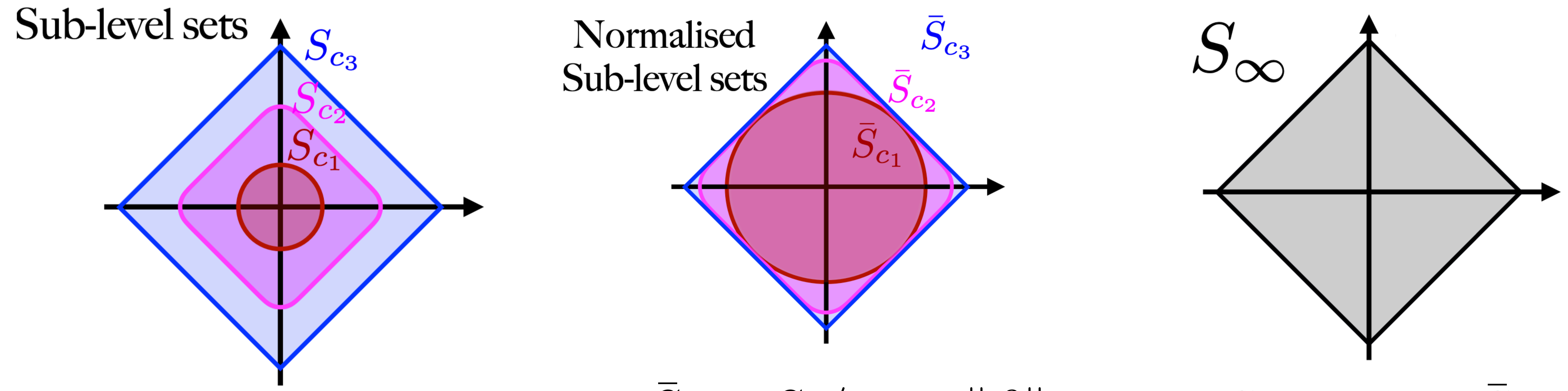
Example: $\beta = F(u, v) = u \odot v$ *"diagonal neural networks"*

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 : geometry of ϕ \`at infinity'



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

We say that ϕ admits a horizon function 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_∞
Asymmetric norm whose unit ball is S_∞

Main result: convergence and implicit bias

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

Theorem: if ϕ is tame, it admits a horizon ϕ_∞ and the mirror flow iterates β_t converge in direction towards the vector $\bar{\beta}_\infty$ satisfying the ϕ_∞-max margin problem:

$$\lim_{t\to\infty} \frac{\beta_t}{\|\beta_t\|} =: \bar{\beta}_\infty \propto \operatorname{argmin}\{\phi_\infty(\beta^\star) : \beta^\star \in S\}$$

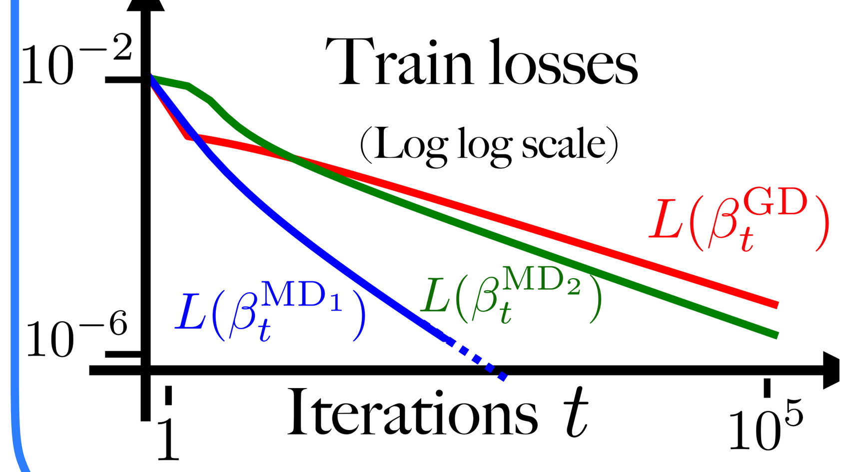
set of separating hyperplanes

Experiments

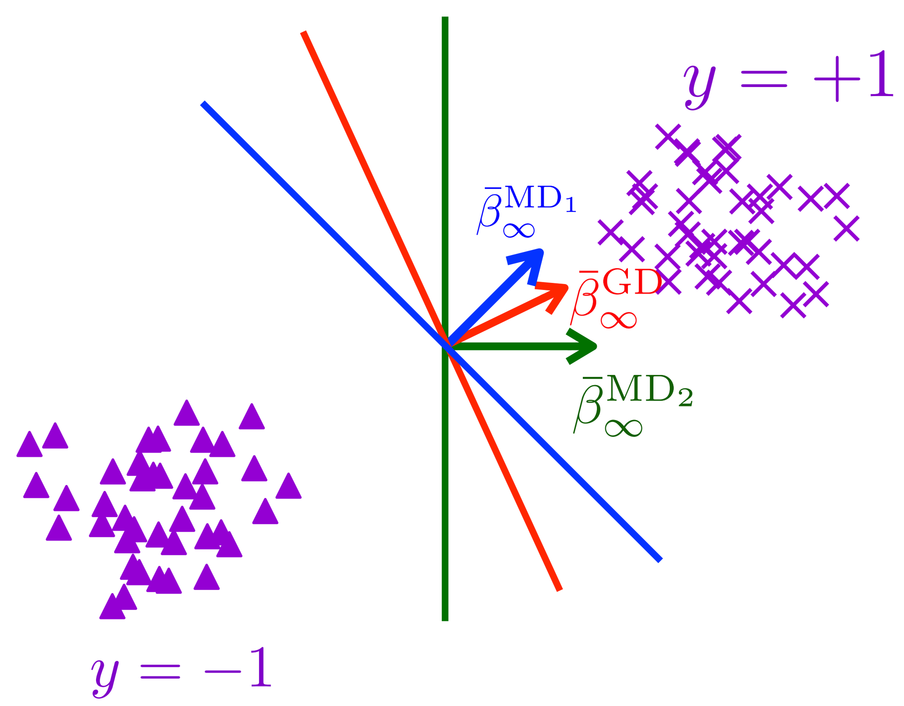
Mirror descent with:

$\phi^{\mathrm{GD}}(\beta) = \|\beta\|_2^2 \qquad \phi^{\mathrm{MD}_1}(\beta) = \sum_{i=1,2} \cosh(\beta_i)$

$\phi^{\mathrm{MD}_2}(\beta) = \sum_{i=1,2} \beta_i \operatorname{arcsinh}(\beta_i) - \sqrt{\beta_i^2 + 1}$



Limiting directions



ϕ_∞- max-margins