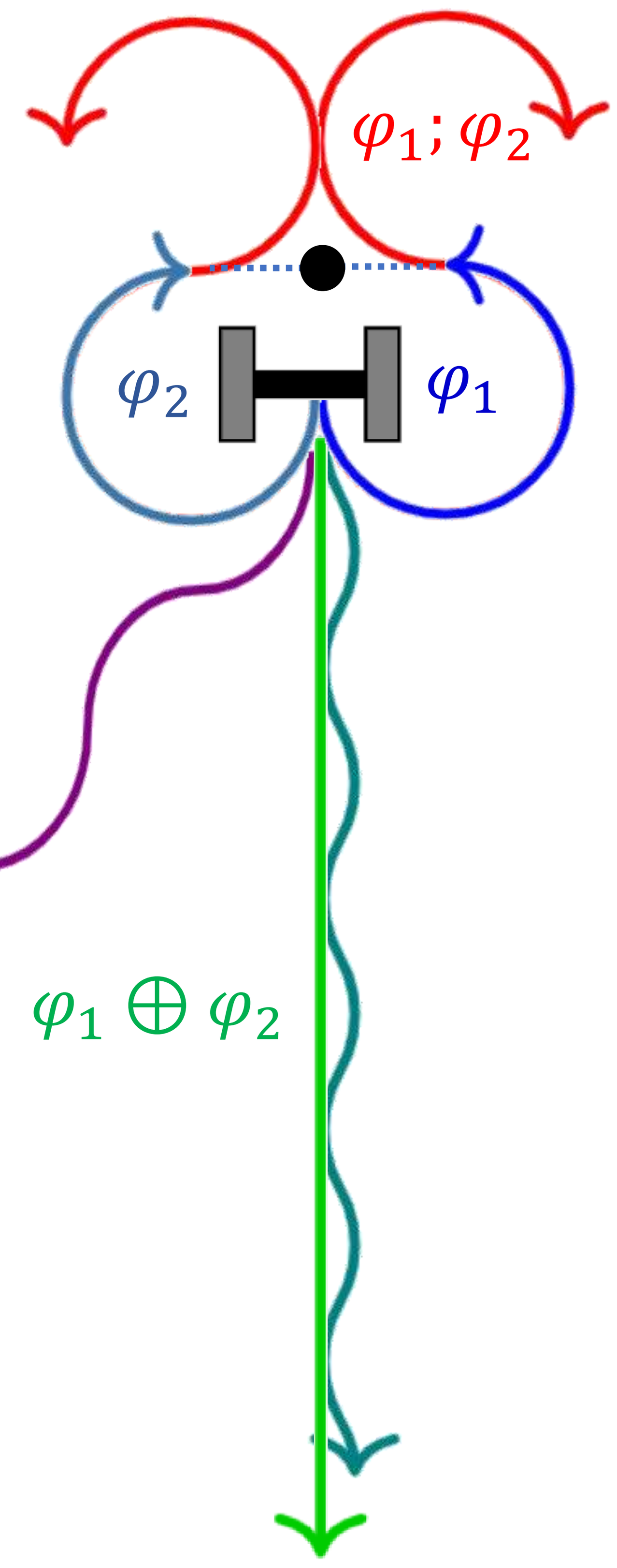
Mixture Languages

Oliver Richardson and Jialu Bao



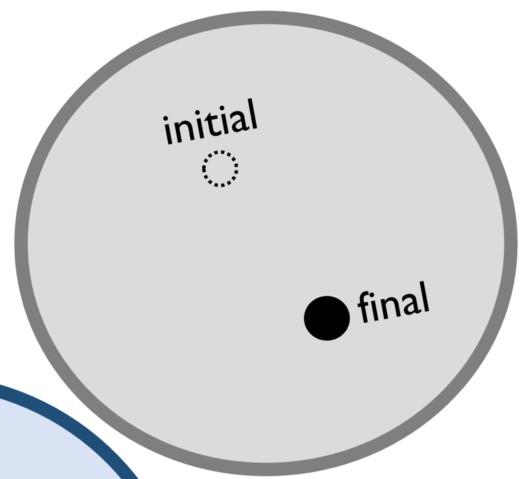
Syntax

$\mathcal{L}(\Phi) \ni \varphi, \varphi'$ ::= **skip** | ϕ | $\varphi; \varphi'$ | $\varphi \oplus \varphi'$ | $\varphi^{(s)}$

set of primitive commands · $\in \Phi$ primitive · sequencing · mixture · partial execution $s \in [0,1]$

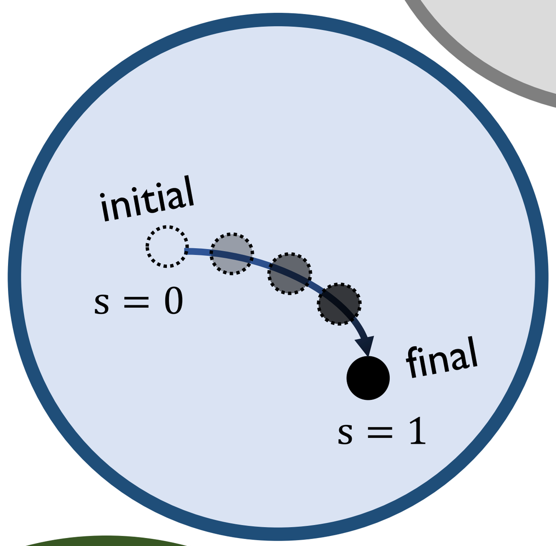
Operational Semantics

State space Θ with
- topology
- diff'ble structure
- Riemannian metric



State Transformer $[\![\varphi]\!]^* : \Theta \longrightarrow \Theta$ (initial, final)

$s = 1$ (select endpoint) $t = \infty$



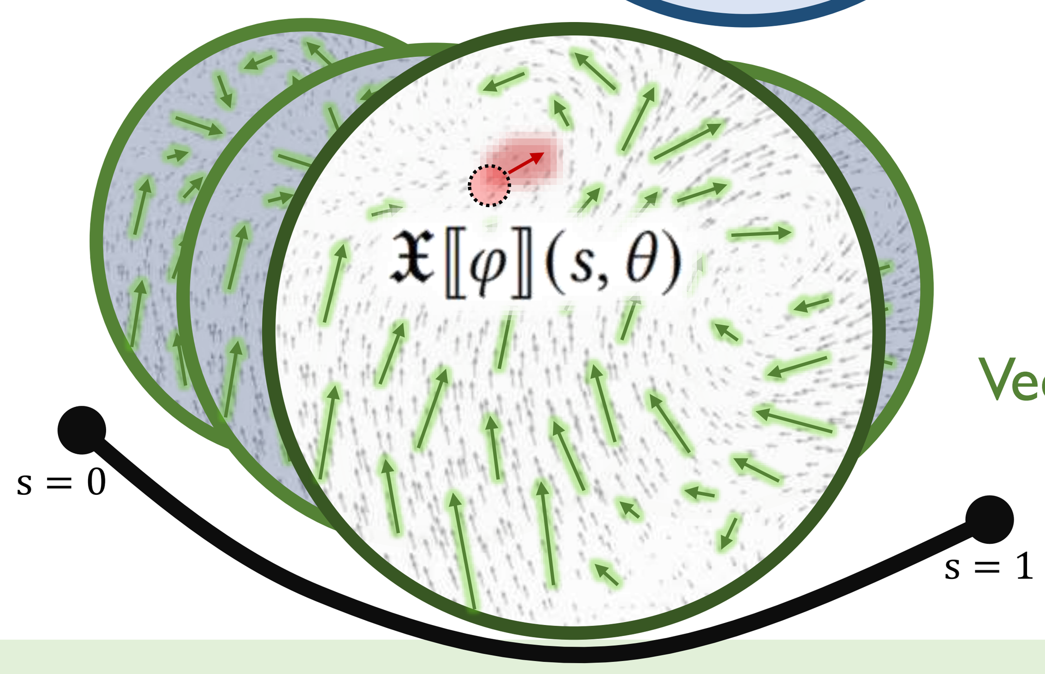
Path $\not{f}[\![\varphi]\!] : [0,1] \times \Theta \longrightarrow \Theta$ (fraction s, initial, final)

$s = 1 - e^{-t}$

$t = -\log(1-s)$

Trajectory: time $t \in [0, \infty]$

$\int$ (solve ODE) $\int$



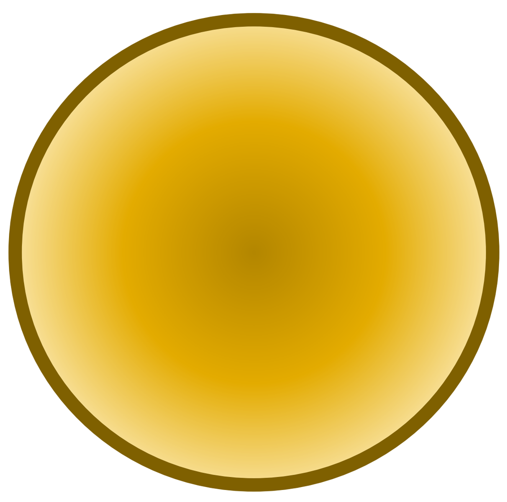
(s-dependent) Vector Field $\mathfrak{X}[\![\varphi]\!] : [0,1] \longrightarrow \mathfrak{X}(\Theta)$

$\frac{dt}{ds} = \frac{1}{1-s}$

(time-dependent) Vector Field

(take gradient) ∇

$\ell[\![\varphi]\!] : \Theta \longrightarrow \mathbb{R}$ Loss Function

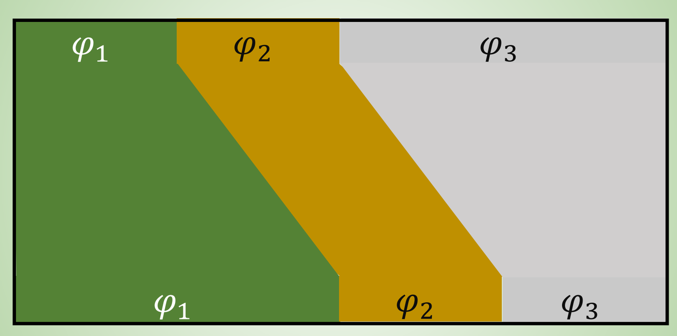


Inductive Construction

sequential execution
$$\mathfrak{X}[\![\varphi_1;\varphi_2]\!](s) := \begin{cases} 2\cdot\mathfrak{X}[\![\varphi_1]\!](2s) & s \leq \frac{1}{2} \\ 2\cdot\mathfrak{X}[\![\varphi_2]\!](2s-1) & s > \frac{1}{2} \end{cases}$$

not associative, but resulting paths are equivalent:

$(\varphi_1;\varphi_2);\varphi_3$



$\varphi_1;(\varphi_2;\varphi_3)$

partial execution ("clipping")
$$\mathfrak{X}[\![\varphi^{(c)}]\!](s) := \mathfrak{X}[\![\varphi]\!](cs)\cdot c$$

parallel execution ("mixture")
$$\mathfrak{X}[\![\varphi_1 \oplus \varphi_2]\!](s) := \mathfrak{X}[\![\varphi_1]\!](s) + \mathfrak{X}[\![\varphi_2]\!](s)$$

noop
$$\mathfrak{X}[\![\mathbf{skip}]\!](s) := \mathbf{0}$$

Key Example: Probabilistic Models

Relative entropy of belief p with respect to reality μ

$$D(\mu \,\|\, p) = \mathbb{E}_\mu\left[\log\frac{\mu}{p}\right]$$

Primitive Instructions

	Observe $p(Y\|X)$	Draw $y \sim p \mid X$
Loss	$D\big(\mu(X,Y) \,\big\|\, \mu(X)\cdot p(Y\|X)\big)$ p acts as belief	$D\big(\mu(X,Z)\cdot p(Y\|X) \,\big\|\, \mu(X,Y,Z)\big)$ p acts as reality
	⋮	⋮
Path	$\propto \mu\cdot\left(\frac{p(Y\|X)}{\mu(Y\|X)}\right)^s$ multiplicative interpolation	$\mu(X,Z)\big((1-s)\mu(Y\|X,Z) + (s)p(Y\|X)\big)$ additive interpolation
Endpoint	$\mu(X)p(Y\|X)\mu(Z\|X,Y)$ replace conditional distribution $\mu(Y\|X)$ with $p(Y\|X)$	$\mu(X,Z)p(Y\|X)$ re-sample Y given X, leaving Z alone

Probabilistic Dependency Graphs (and hence Bayesian Networks and Factor Graphs) are mixtures of **observe** commands!

Probabilistic programs are sequences of **draw** commands (and deterministic observe commands).