The Assignment Flow

Christoph Schnörr

Ruben Hühnerbein, Stefania Petra, Fabrizio Savarino,
Alexander Zeilmann, Artjom Zern, Matthias Zisler

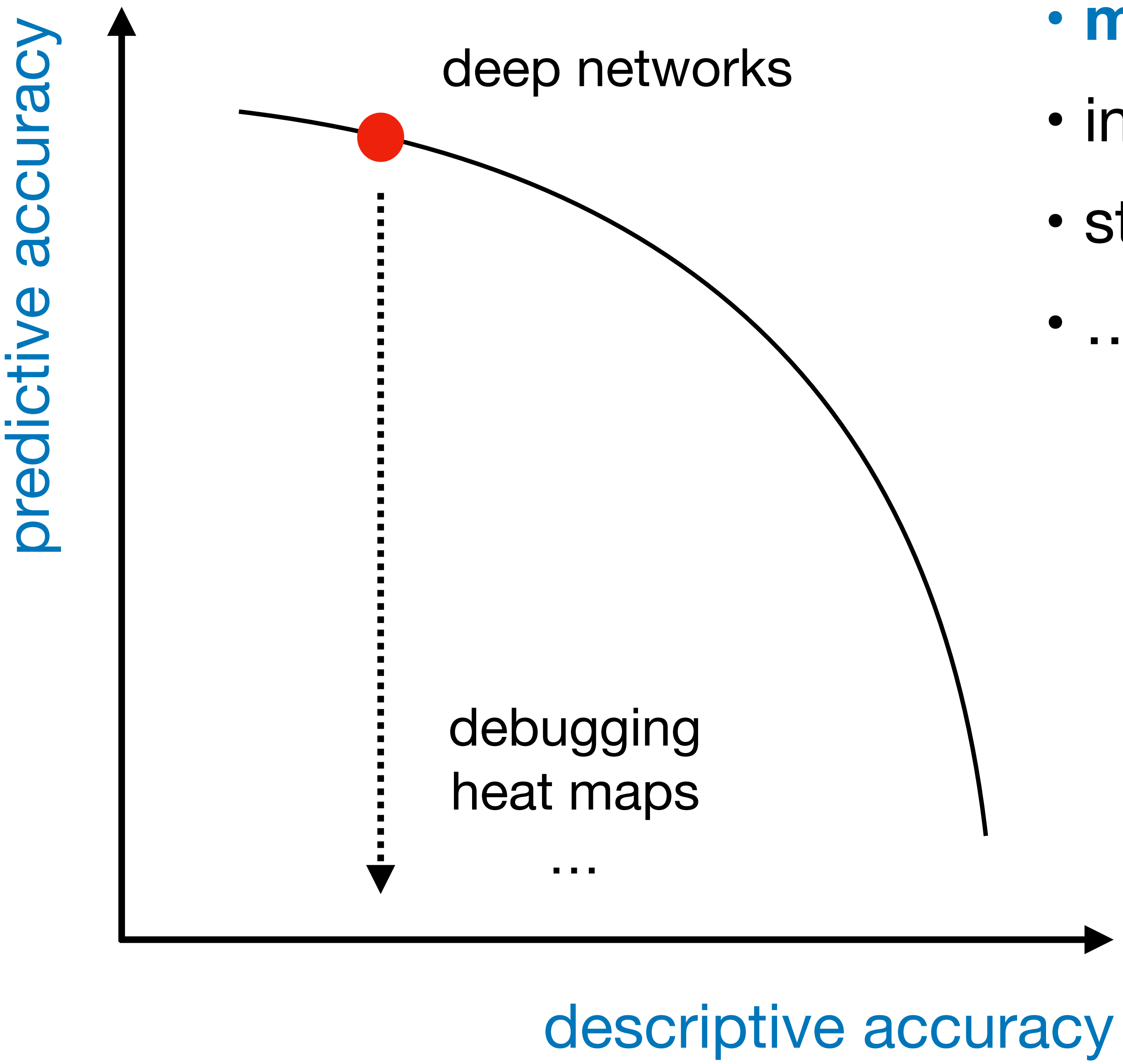
Image & Pattern Analysis Group
Heidelberg University

Mathematics of Imaging: W1
Feb 4-8, Paris

Machine learning and …

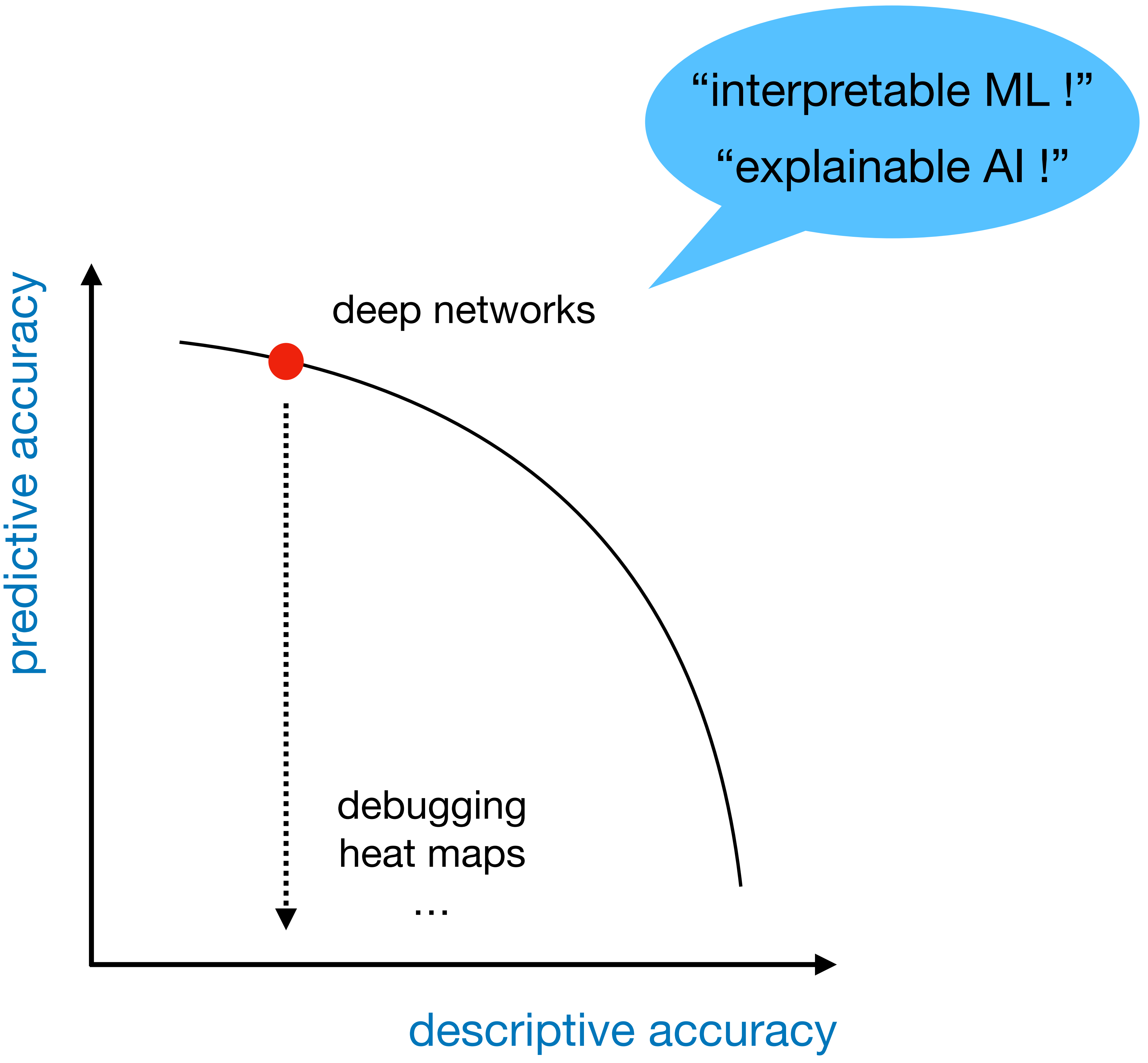
- signal & image proc.
- computer vision
- **math. imaging**
- inverse problems
- statistics
- …

Machine learning and …

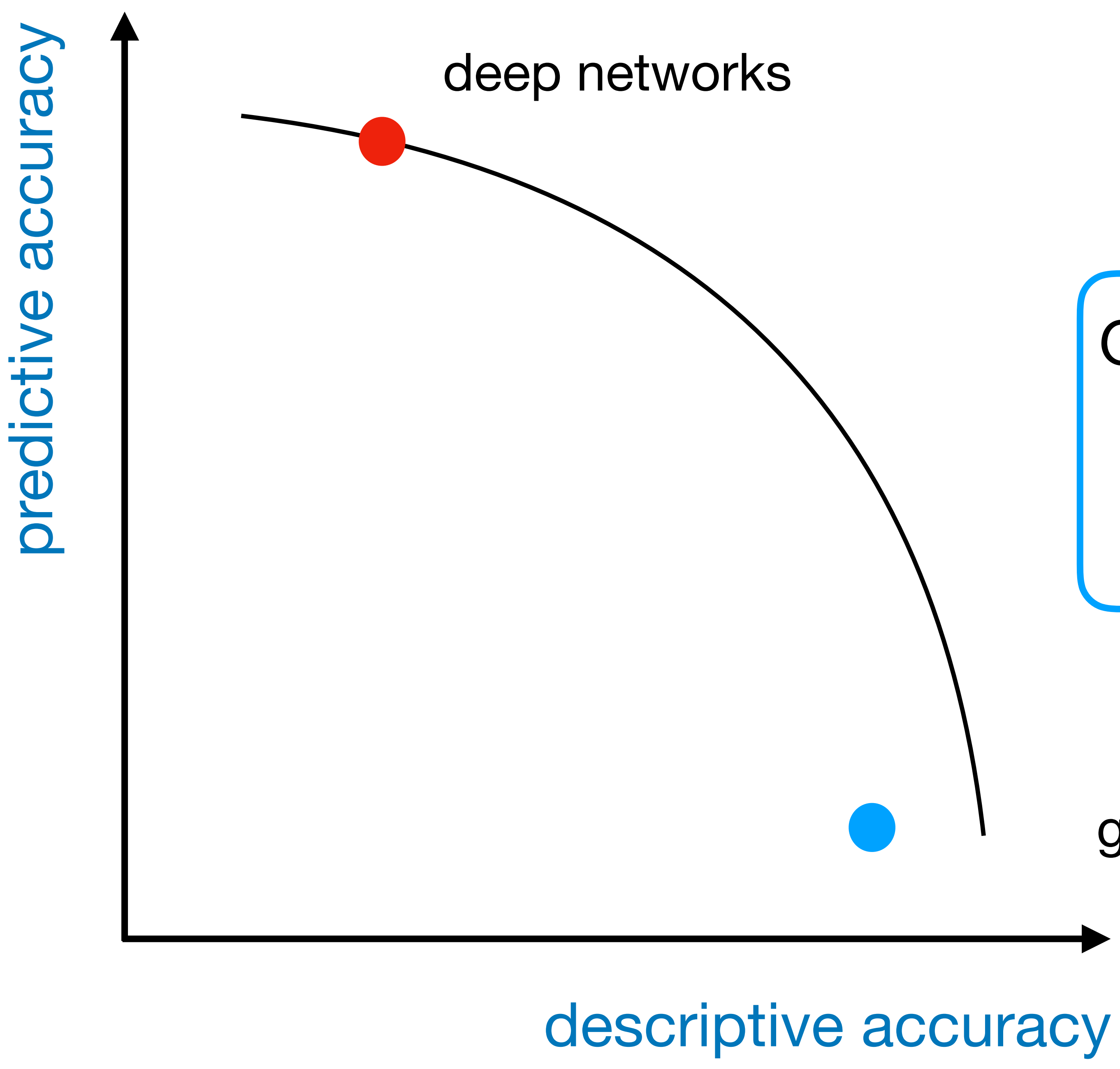


- signal & image proc.
- computer vision
- **math. imaging**
- inverse problems
- statistics
- …

Machine learning and …



Machine learning and …



Geman & Geman (T-PAMI 1984)

…

Kappes et al. (IJCV 2015)

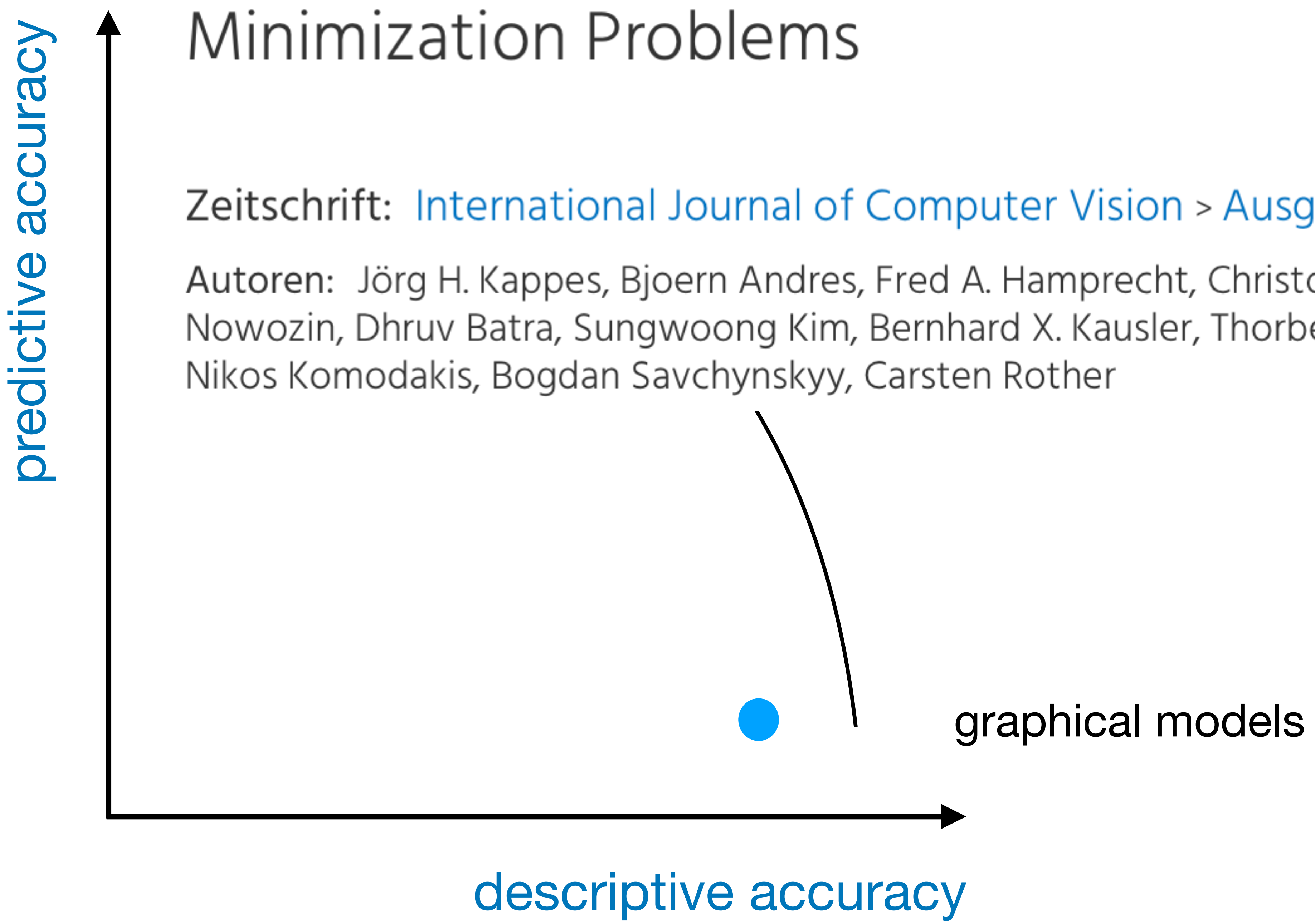
Machine learning and …

01.11.2015 | Ausgabe 2/2015

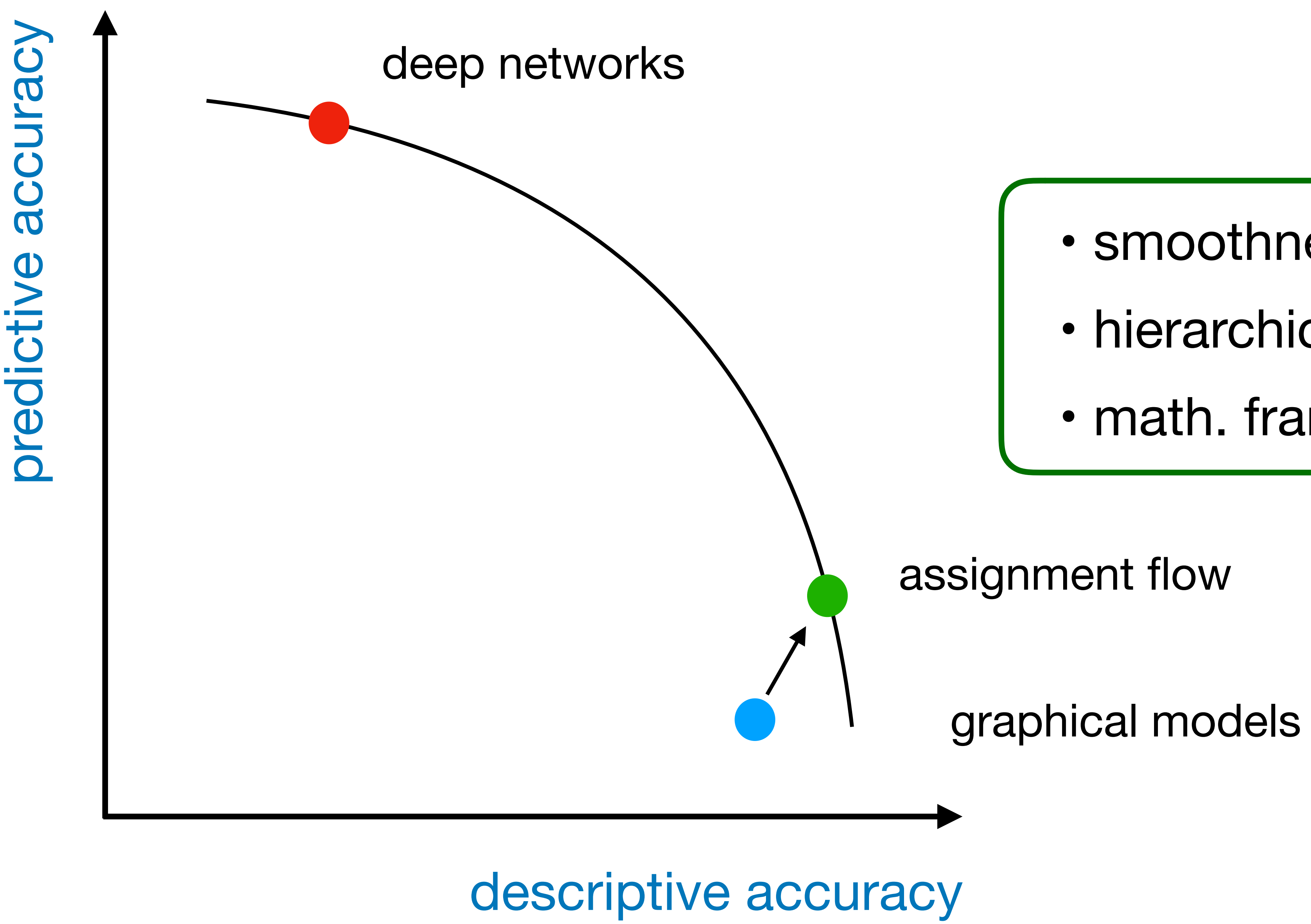
A Comparative Study of Modern Inference Techniques for Structured Discrete Energy Minimization Problems

Zeitschrift: International Journal of Computer Vision > Ausgabe 2/2015

Autoren: Jörg H. Kappes, Bjoern Andres, Fred A. Hamprecht, Christoph Schnörr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X. Kausler, Thorben Kröger, Jan Lellmann, Nikos Komodakis, Bogdan Savchynskyy, Carsten Rother



Machine learning and …



- smoothness
- hierarchical structure
- math. framework

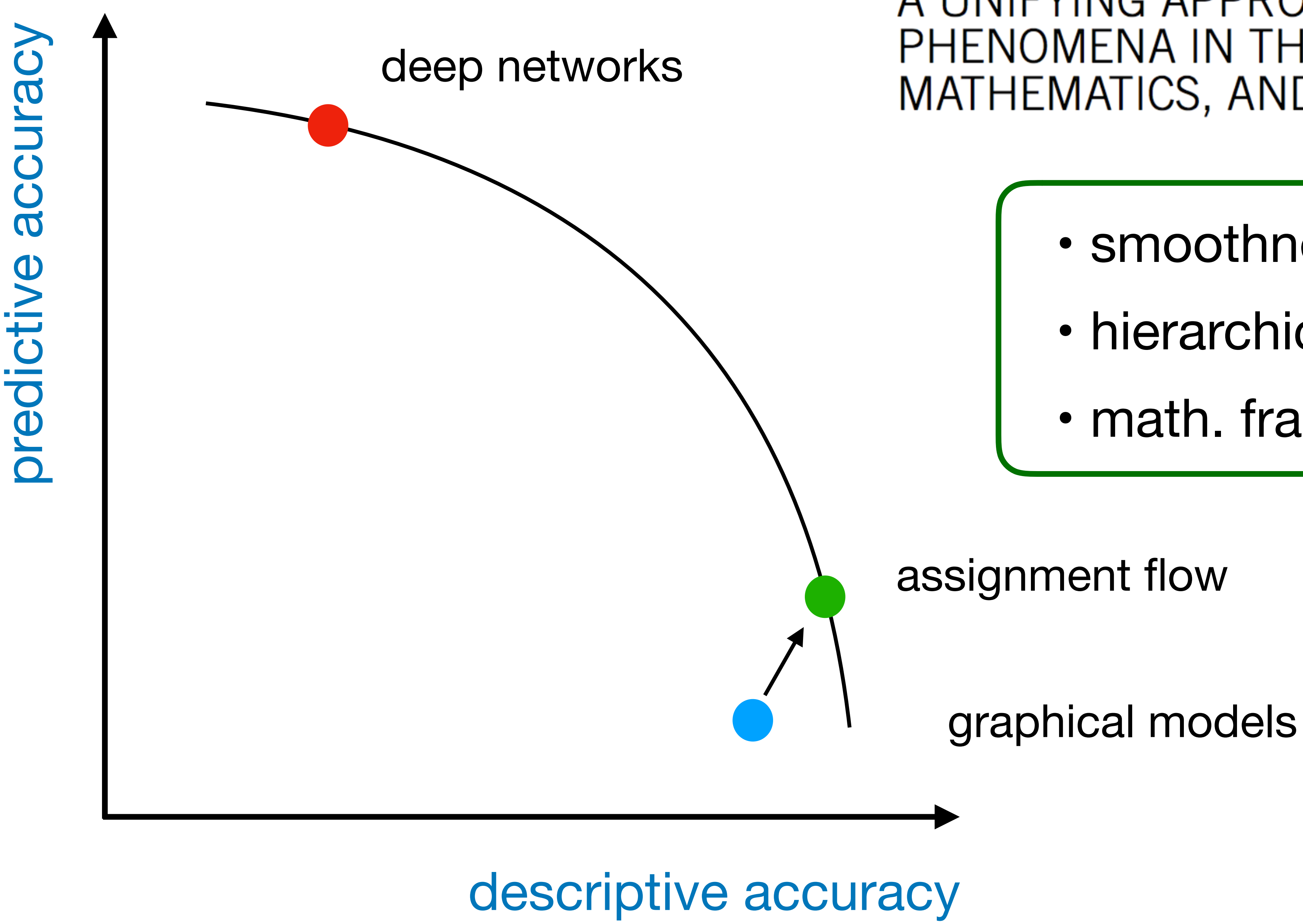
Machine learning and …

CLUSTER OF
EXCELLENCE
STRUCTURES

**UNIVERSITÄT
HEIDELBERG**
ZUKUNFT
SEIT 1386

STRUCTURES
A UNIFYING APPROACH TO EMERGENT PHENOMENA IN THE PHYSICAL WORLD, MATHEMATICS, AND COMPLEX DATA

- smoothness
- hierarchical structure
- math. framework



Outline

- set-up: assignment flow *JMIV'17*
  supervised labeling *SIIMS'18*

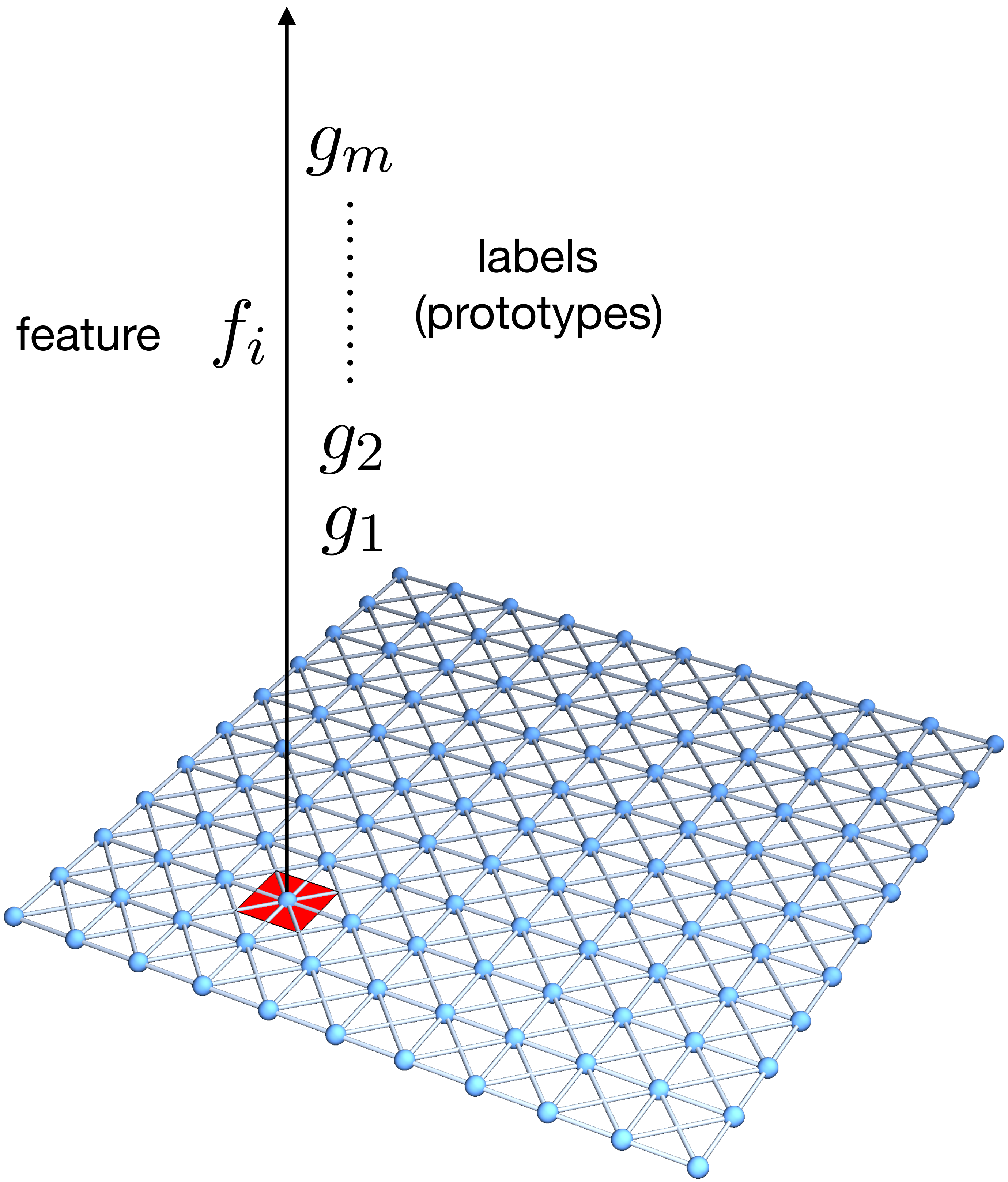
- unsupervised labeling
  - label evolution
  - label learning from scratch

- parameter estimation (control)

- outlook

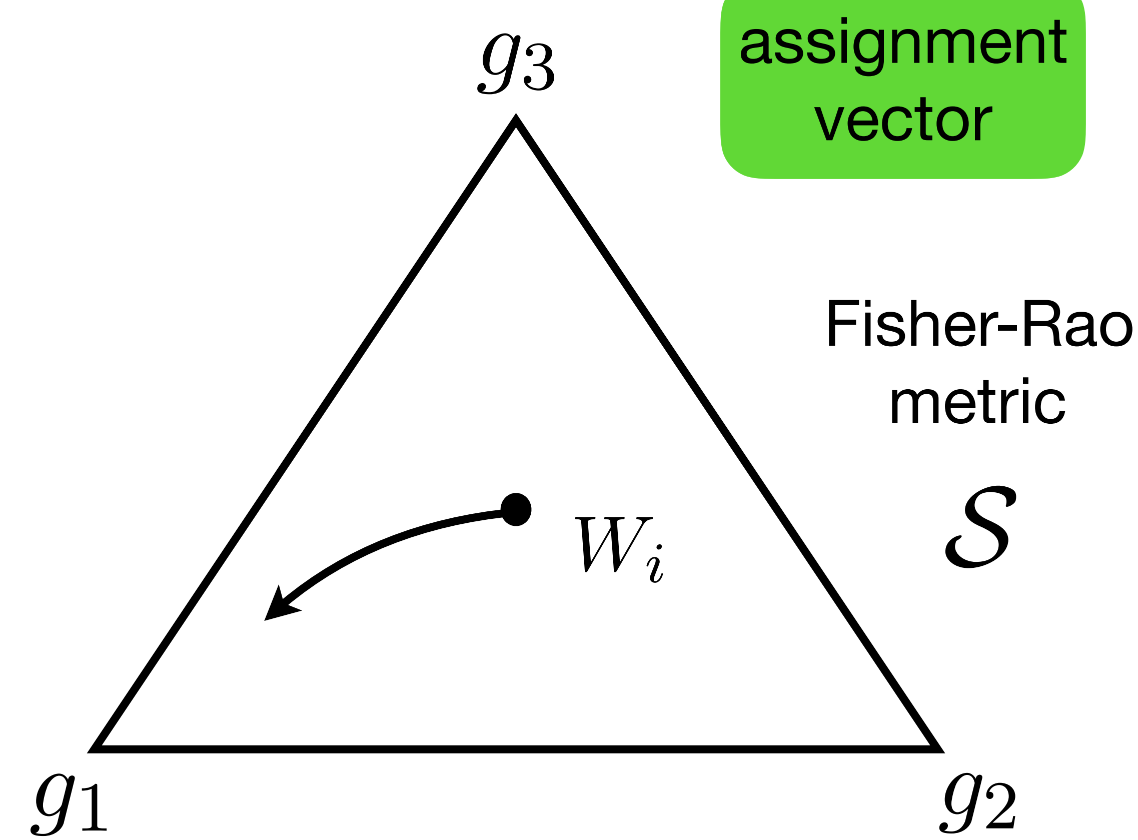
Set-up: assignment flow & supervised labeling

$$D_i = \big(d(f_i, g_1), \ldots, d(f_i, g_m)\big)$$ metric, distance vector (**data term**)



$$W_i = \big(\,\mathrm{Pr}(g_1|f_i), \ldots, \mathrm{Pr}(g_m|f_i)\big)$$

assignment vector



$f_i \in (\mathcal{X}, d)$

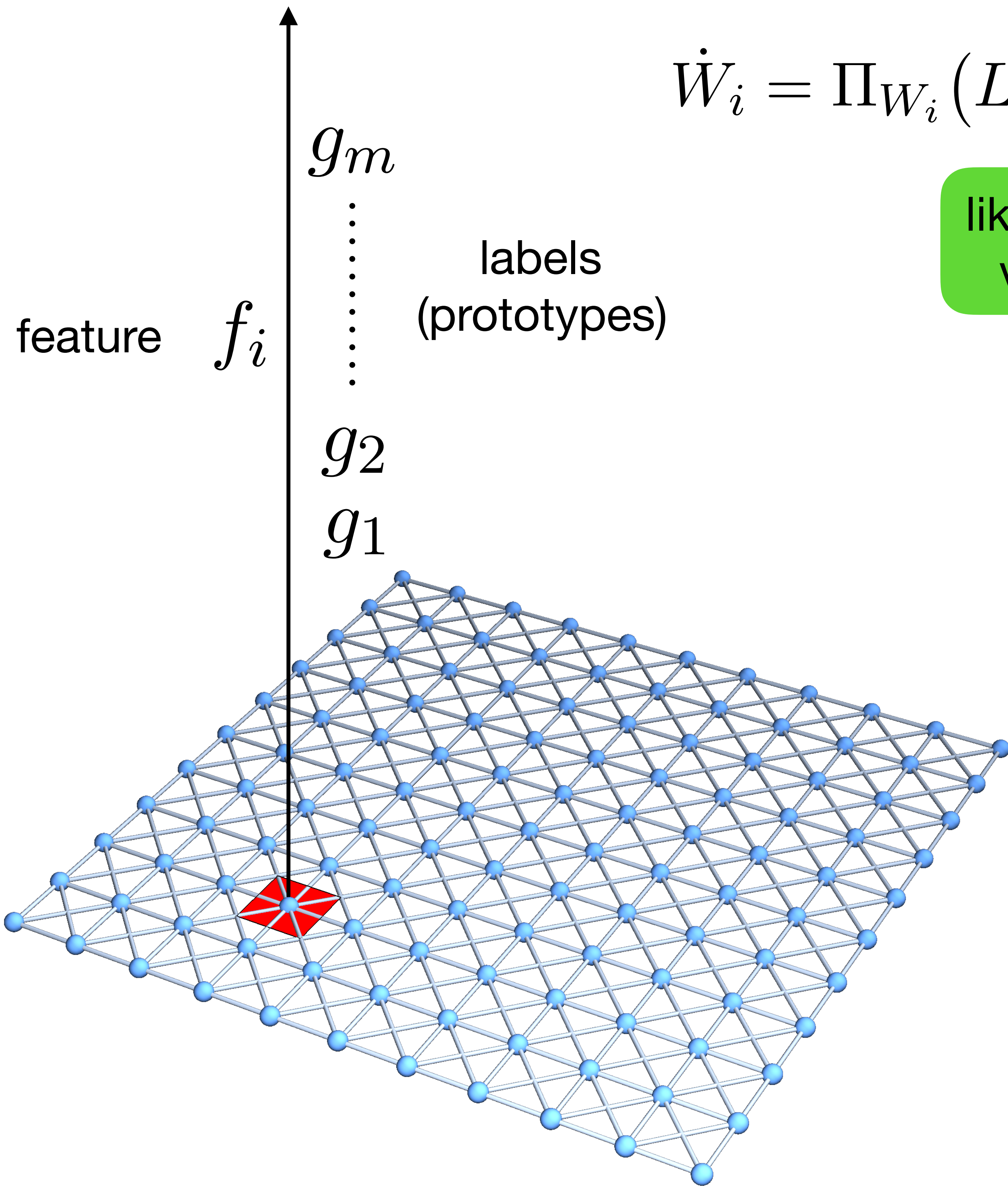
features metric space

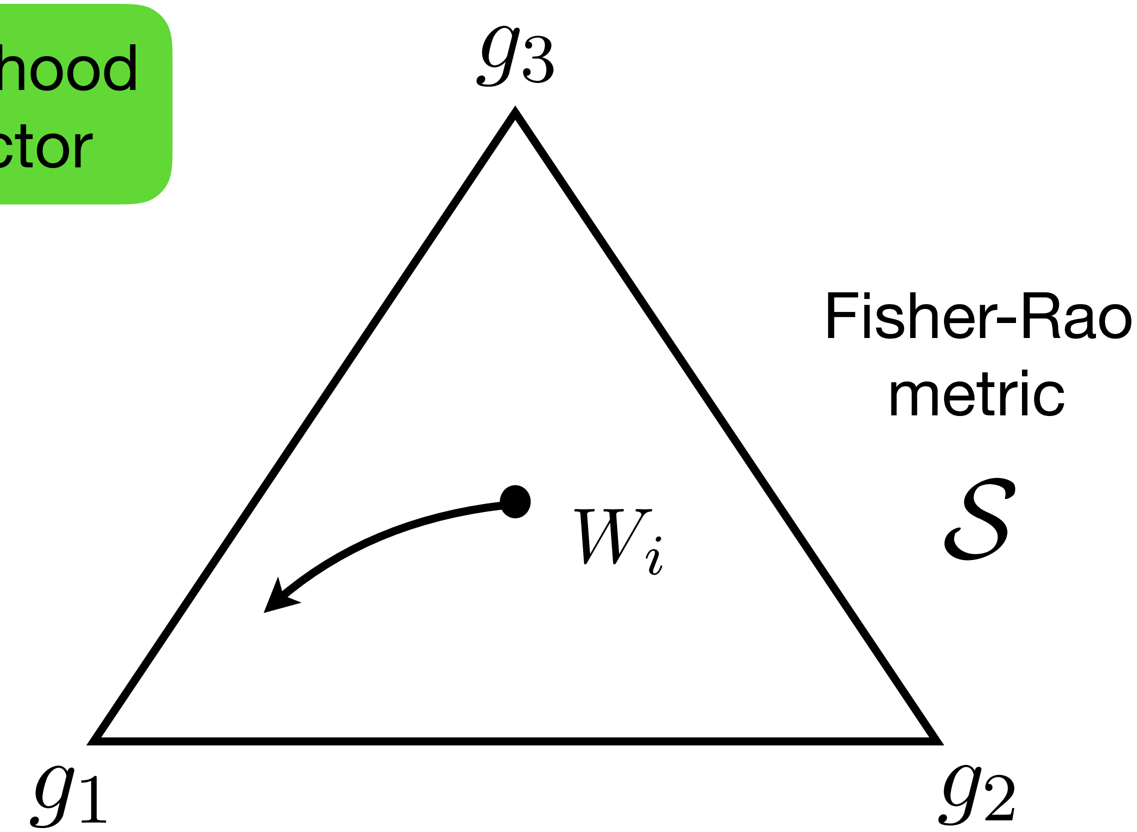
Set-up: assignment flow & supervised labeling

$$D_i = \big(d(f_i, g_1), \dots, d(f_i, g_m)\big)$$ metric, distance vector (**data term**)

$$\dot{W}_i = \Pi_{W_i}\big(L_i(W_i)\big) = \Pi_{W_i}\Big(\exp_{W_i}\big(-\frac{1}{\rho} D_i\big)\Big)$$

likelihood vector





$$f_i \in (\mathcal{X}, d)$$

features metric space

Set-up: assignment flow & supervised labeling

$$D_i = \big(d(f_i, g_1), \dots, d(f_i, g_m)\big)$$ metric, distance vector (**data term**)

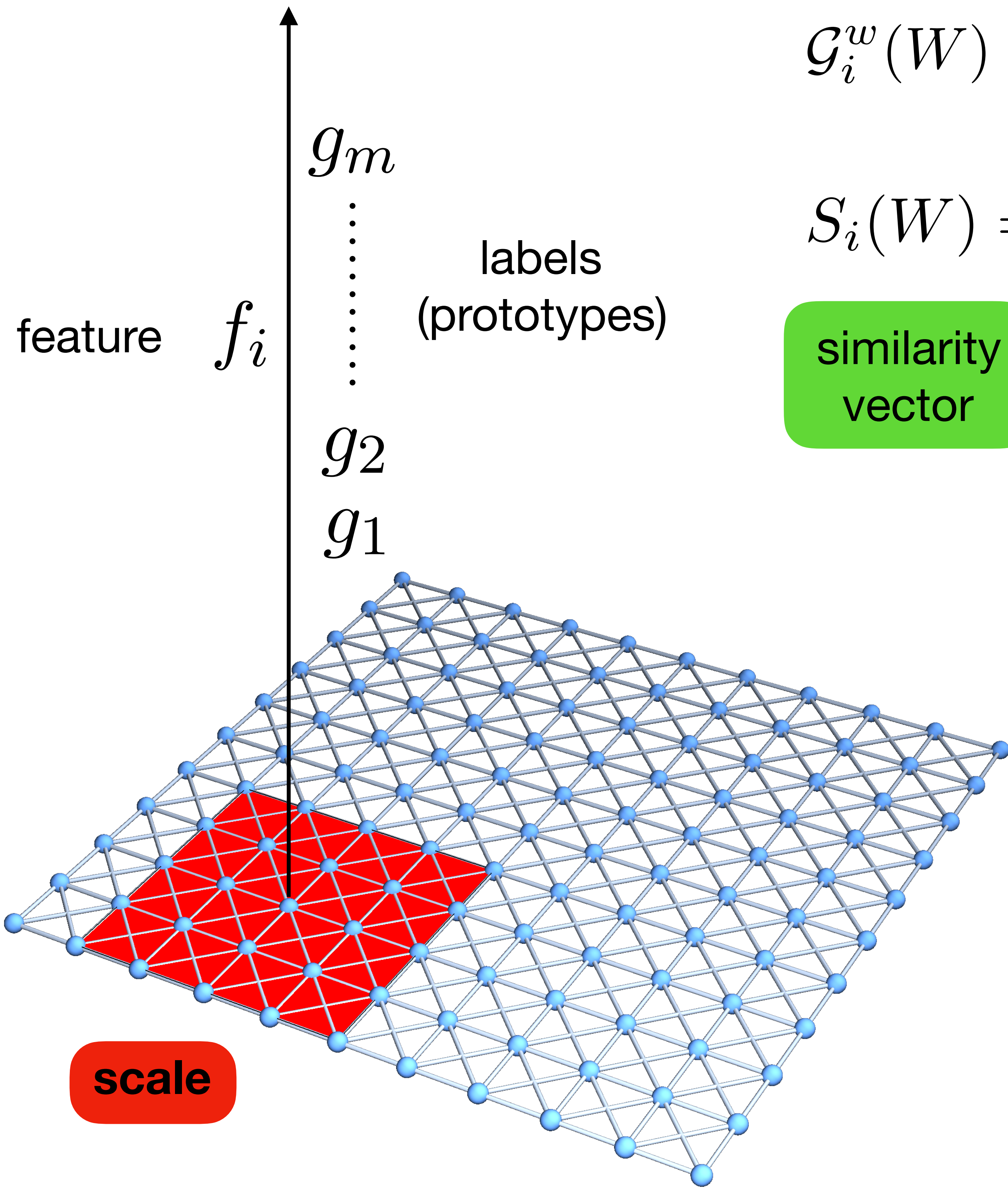
$$\mathcal{G}_i^w(W) = \mathrm{Exp}_{W_i}\Big(\sum_{k \in \mathcal{N}_i} w_{ik} \, \mathrm{Exp}_{W_i}^{-1}(W_k)\Big)$$

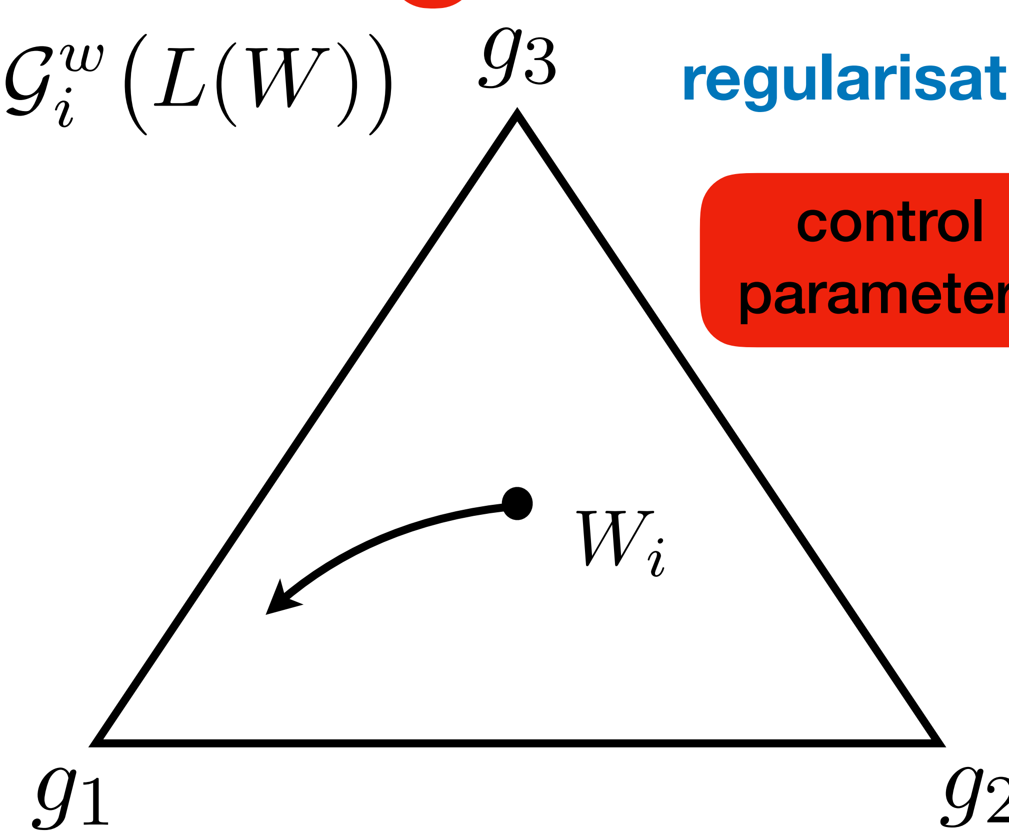
$$S_i(W) = \mathcal{G}_i^w\big(L(W)\big)$$

similarity vector

regularisation

control parameters





$$f_i \in (\mathcal{X}, d)$$

features metric space

Set-up: assignment flow & supervised labeling

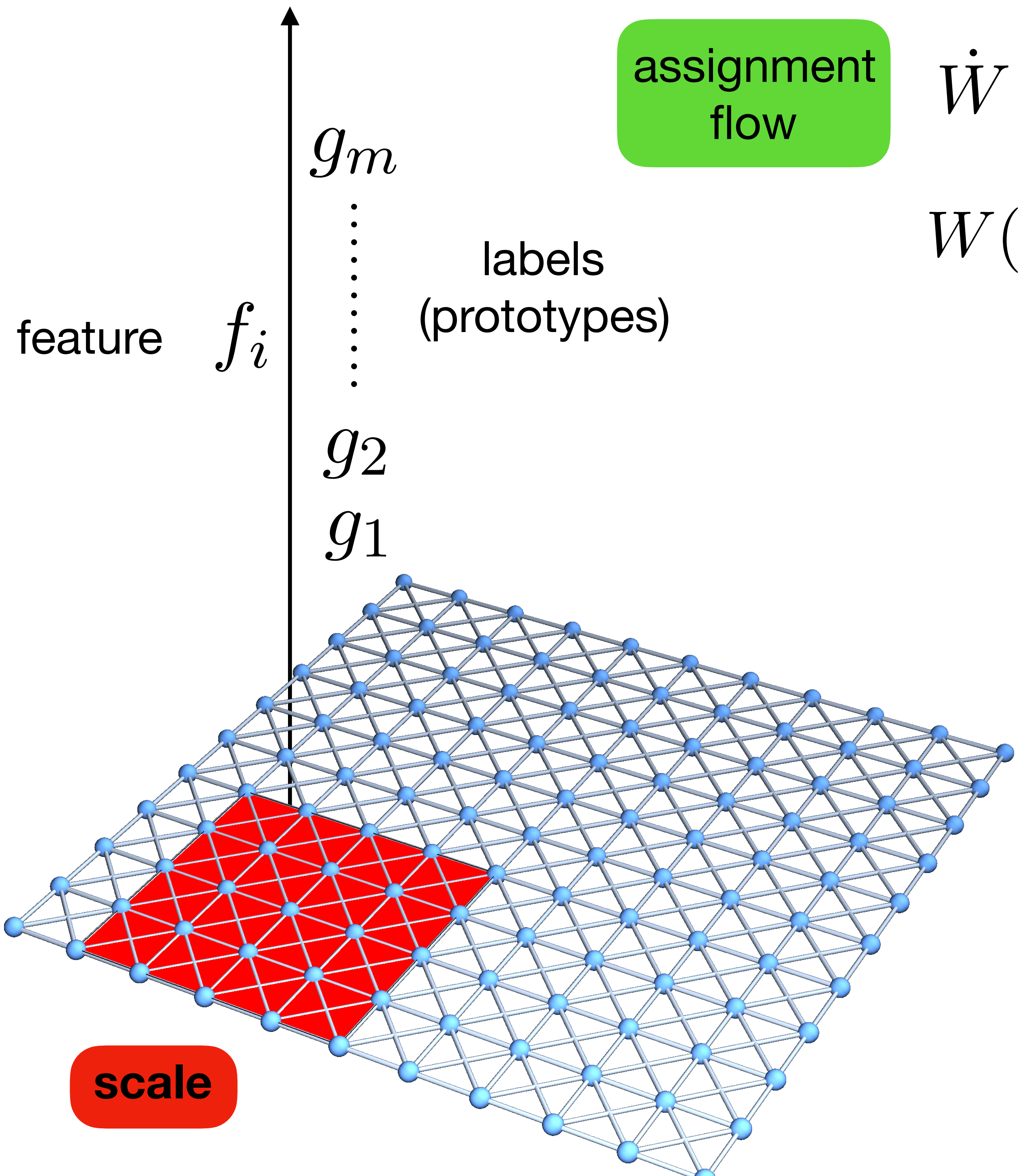
$$D_i = \big(d(f_i, g_1), \dots, d(f_i, g_m)\big)$$ metric, distance vector (**data term**)

assignment flow

$$\dot{W} = \Pi_W\big(S(W)\big), \quad W(0) = \mathbb{1}_{\mathcal{W}}$$

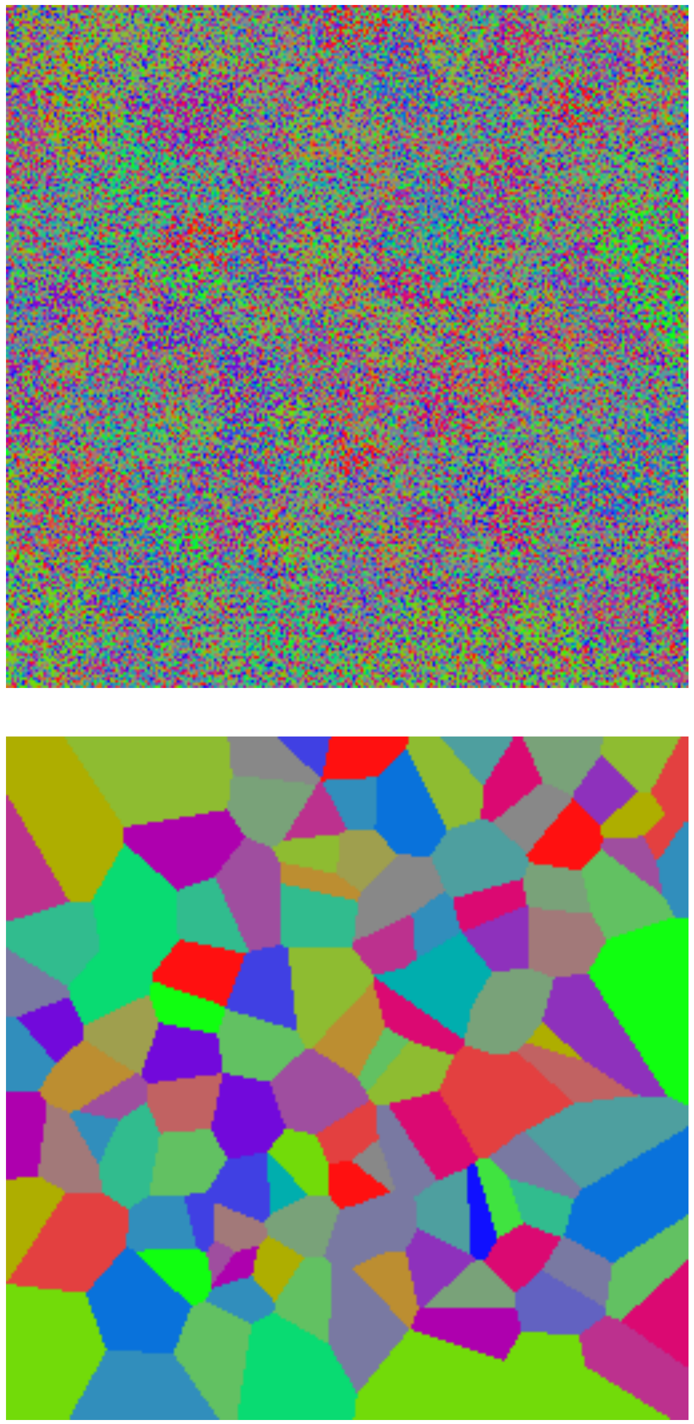
$$W(t) \in (\mathcal{W}, g_{FR})$$ assignment manifold

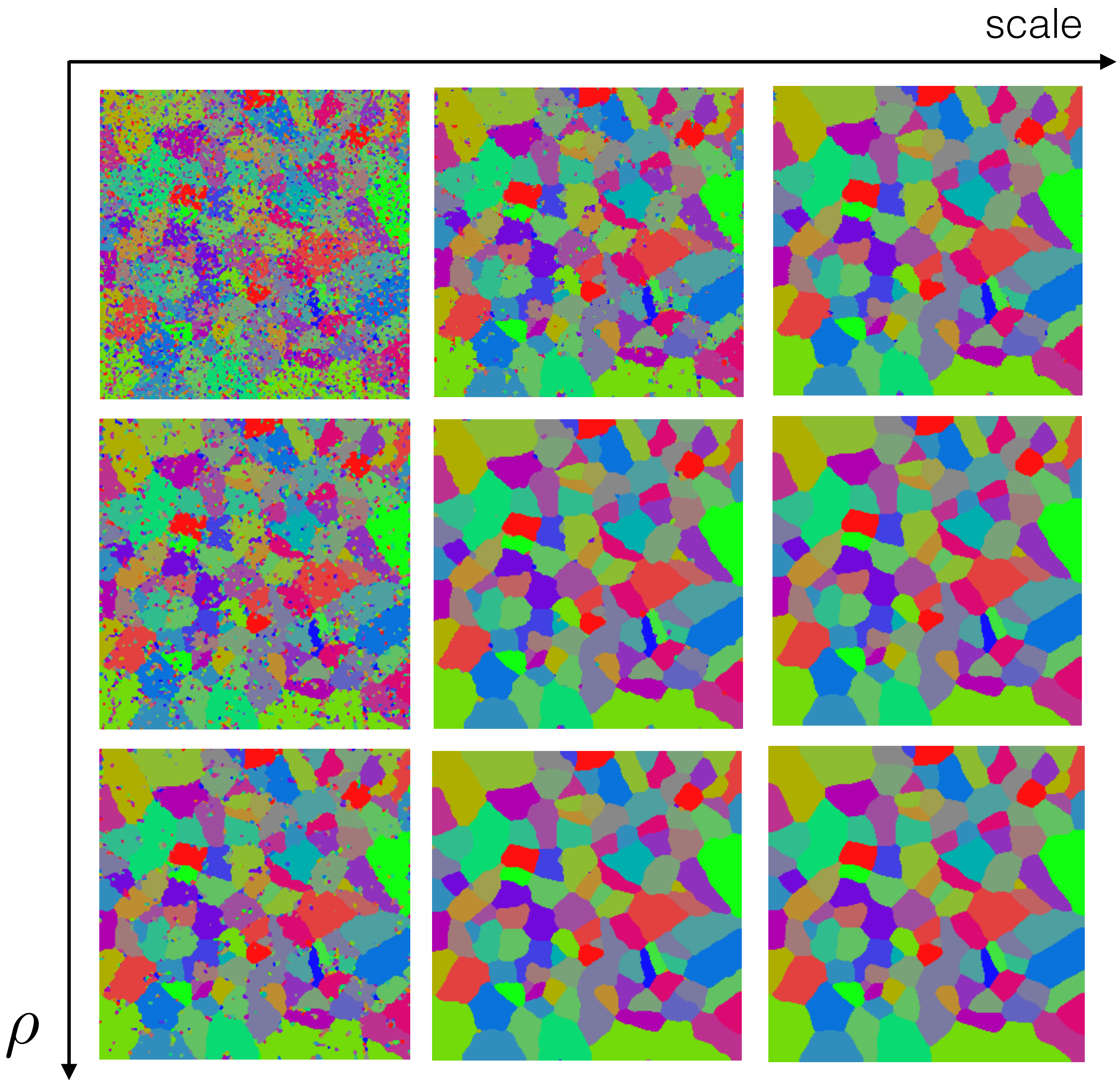
$$\mathcal{W} = \mathcal{S} \times \cdots \times \mathcal{S}$$



$$f_i \in (\mathcal{X}, d)$$ **features metric space**

Illustration





Illustration

local

3 x 3

5 x 5

Assignment flow: geometric integration

ODEs on manifolds, Lie group methods (*Iserles et al.'05, Hairer et al. '06*)

$$\lambda(v,p) = \Lambda(\exp_{\mathrm{G}}(v),p) \qquad (\lambda_* v)_p = \frac{\mathrm{d}}{\mathrm{d}t}\Lambda(\exp_{\mathrm{G}}(tv),p)\big|_{t=0}$$

$$\begin{tikzcd}
\mathfrak{g} \arrow[rr, "\boxed{\lambda_*}"] \arrow[d, "\exp_{\mathrm{G}}"'] \arrow[dr] & & \mathfrak{X}(M) \arrow[d, "\exp_M"] \\
\mathrm{G} \arrow[dr] & \boxed{\lambda} \arrow[r, bend right] & M \arrow[l] \arrow[ull, "\boxed{f}"'] \arrow[dl] \\
& \boxed{\Lambda} \arrow[ur, bend right] &
\end{tikzcd}$$

$$\boxed{\dot{y} = \big(\lambda_* f(t,y)\big)_y, \quad y(0) = p}$$

$$\Lambda \colon \mathrm{G} \times M \to M$$

$$y(t) = \lambda\big(v(t),p\big)$$

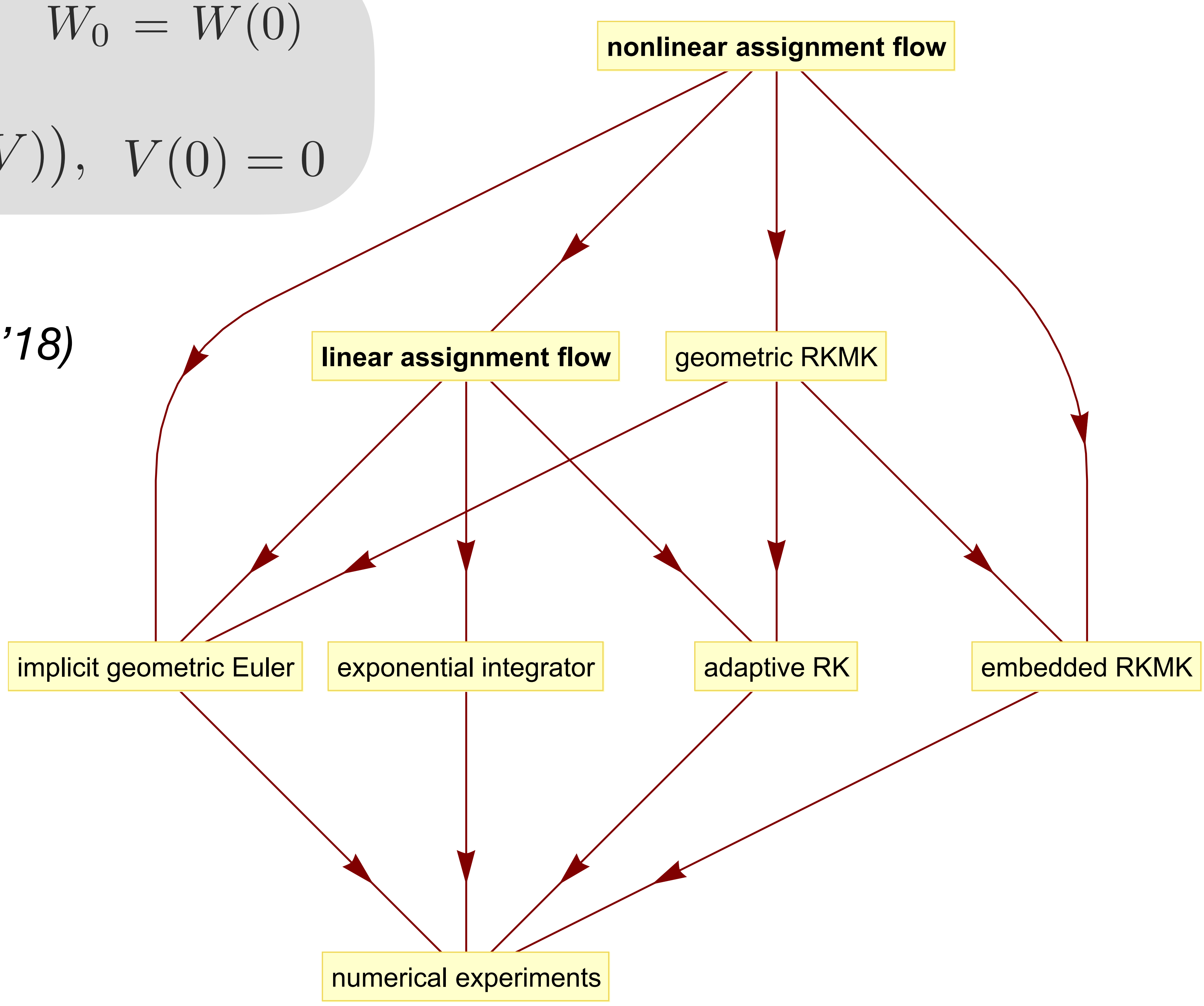
$$\dot{v} = (\mathrm{dexp}_{\mathrm{G}}^{-1})_v\big(f(t,\lambda(v,p))\big), \; v(0) = 0$$

Assignment flow: geometric integration

$$W(t) = \exp_{W_0}\big(V(t)\big) \qquad W_0 = W(0)$$

$$\dot{V} = \Pi_{\mathcal{T}_0} S\big(\exp_{W_0}(V)\big), \quad V(0) = 0$$

(Zeilmann et al, arXiv'18)



Properties (more general viewpoint)

- elementary state space — statistical manifold $(\mathcal{W}, g)$
- information geometry — dualistic structure (g, ∇, ∇^*)

  (Amari & Chentsov)

$$Zg(X,Y) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y)$$

- scale — small / cooperative & large / competitive
- distances D — adaptive distances $D(W),\ W \in \mathcal{W}$

Outline

- set-up: assignment flow
  supervised labeling

- unsupervised labeling
  - label evolution
  - label learning from scratch

(Zern et al., GCPR'18)

(submitted)

- parameter estimation (control)

- outlook

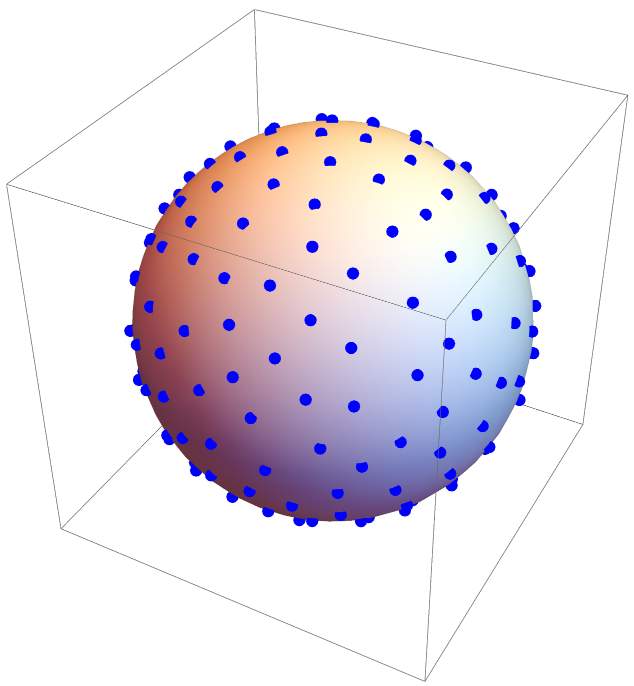
Label evolution

$$D_i = \big(d(f_i, g_1), \ldots, d(f_i, g_m)\big)$$ metric, distance vector (**data term**)

preprocessing

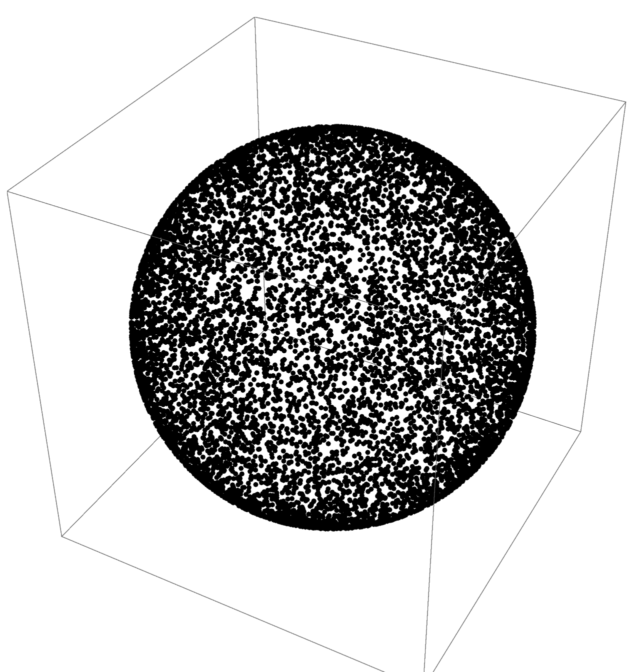
labels

g_j

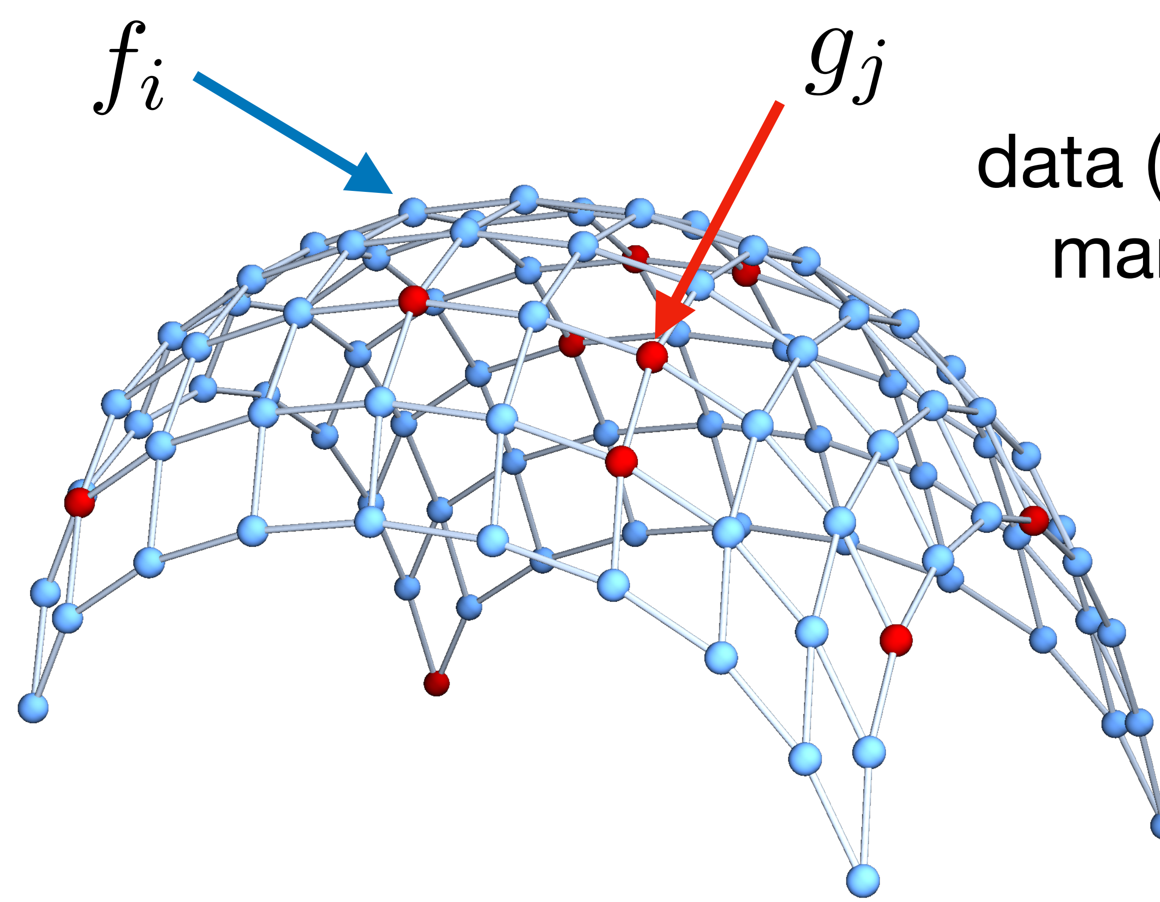


data

f_i



adapt online !



data (feature) manifold M

Label evolution

divergence measure

time scale

label flow ("*m*": label as Riemannian means)

$$\dot{m}_j(t) = -\alpha \sum_{i \in I} \nu_{j|i}\big(M(t)\big)\widehat{g}^{-1}\big(d_j D(f_i, m_j(t))\big), \quad m_j(0) = m_{j0}, \quad \alpha > 0$$

$$\nu_{j|i}(M) = \frac{L^{\sigma}_{ij}(W_i; M)}{\sum_{k \in I} L^{\sigma}_{kj}(W_k; M)}, \quad L^{\sigma}_{ij}(W_i; M) = \frac{W_{ij} e^{-\frac{1}{\sigma} D(f_i, m_j)}}{\sum_{l \in J} W_{il} e^{-\frac{1}{\sigma} D(f_i, m_l)}}, \quad \sigma > 0$$

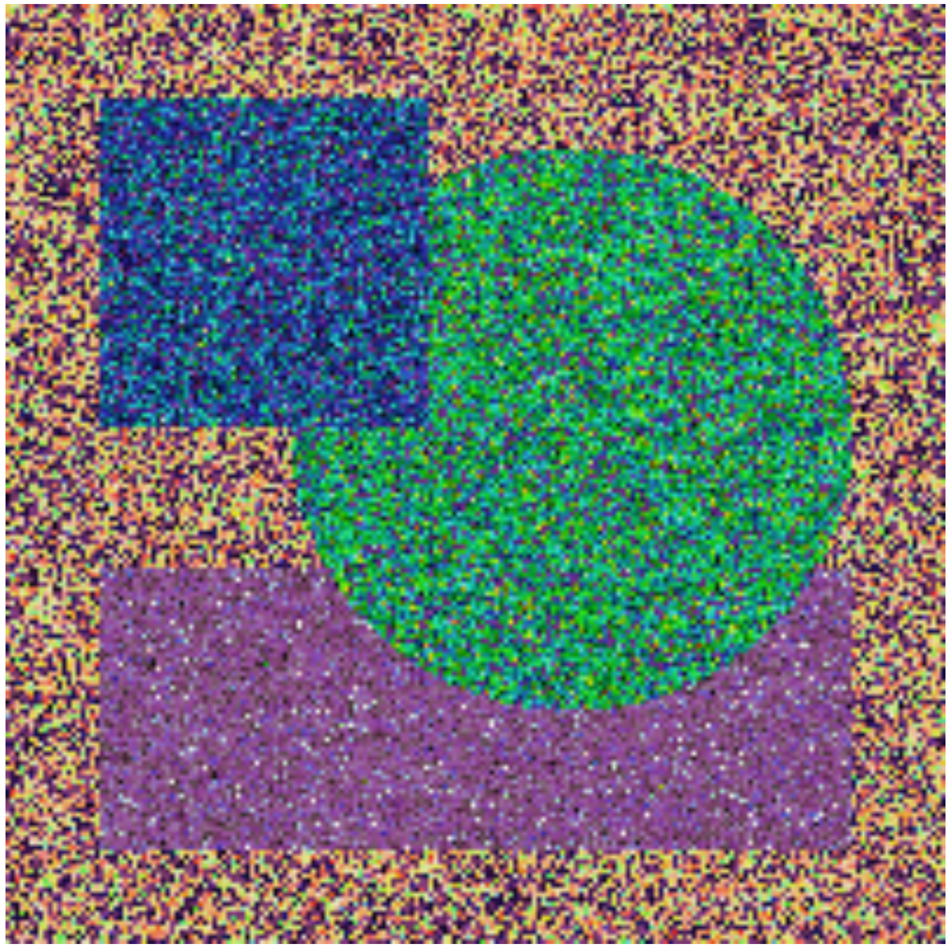
coupling spatial regularisation

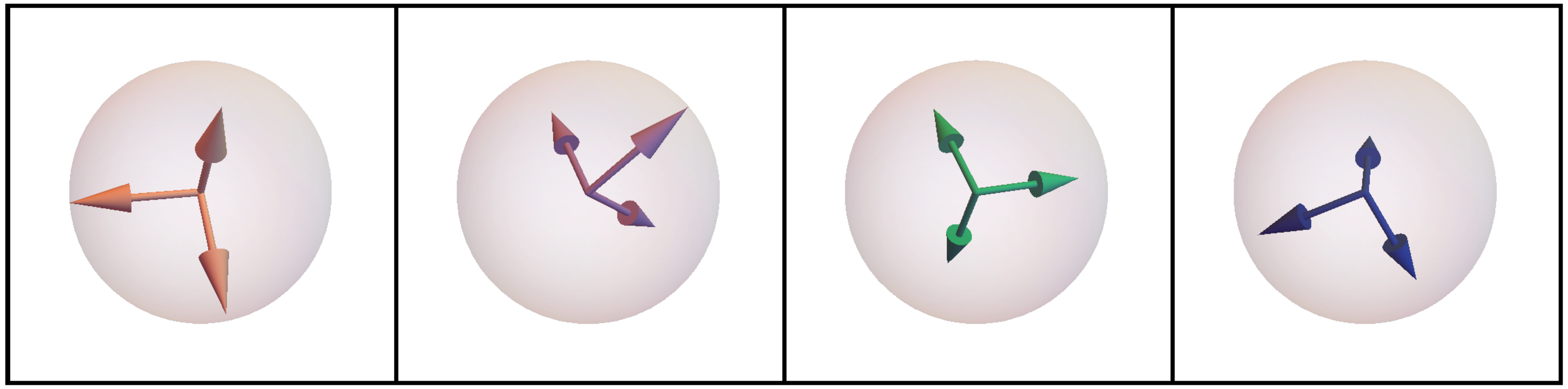
assignment flow

$$\dot{W}_i(t) = \Pi_{W_i(t)}\Big(S_i\big(W(t)\big)\Big), \quad W_i(0) = \mathbb{1}_{\mathcal{S}}, \quad i \in I$$

Label evolution

SO(3)-valued data

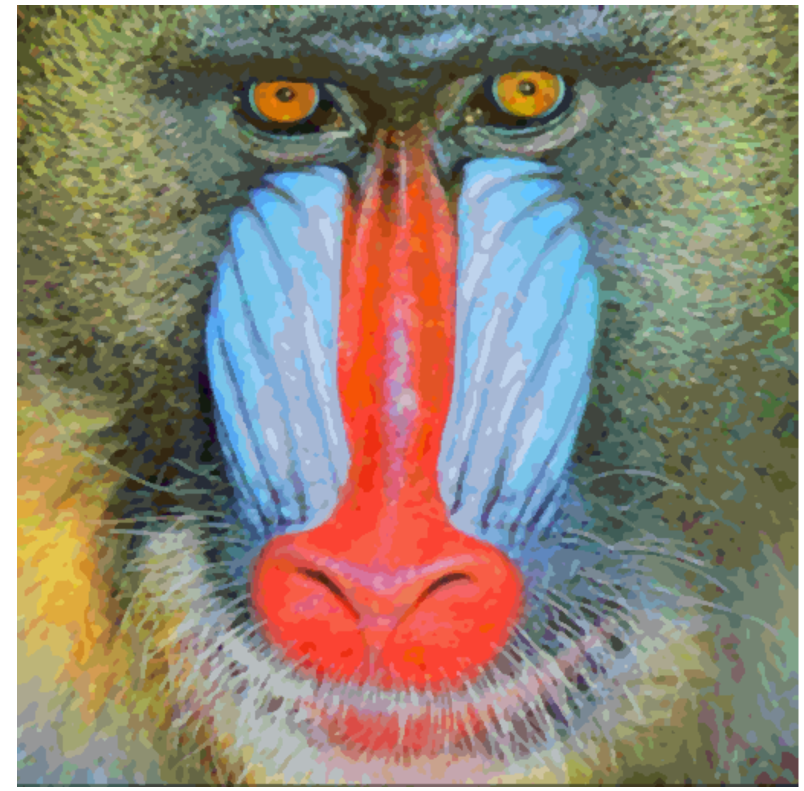




Label evolution

Euclidean color space

supervised: 200 labels



unsupervised: *few* labels

Label evolution

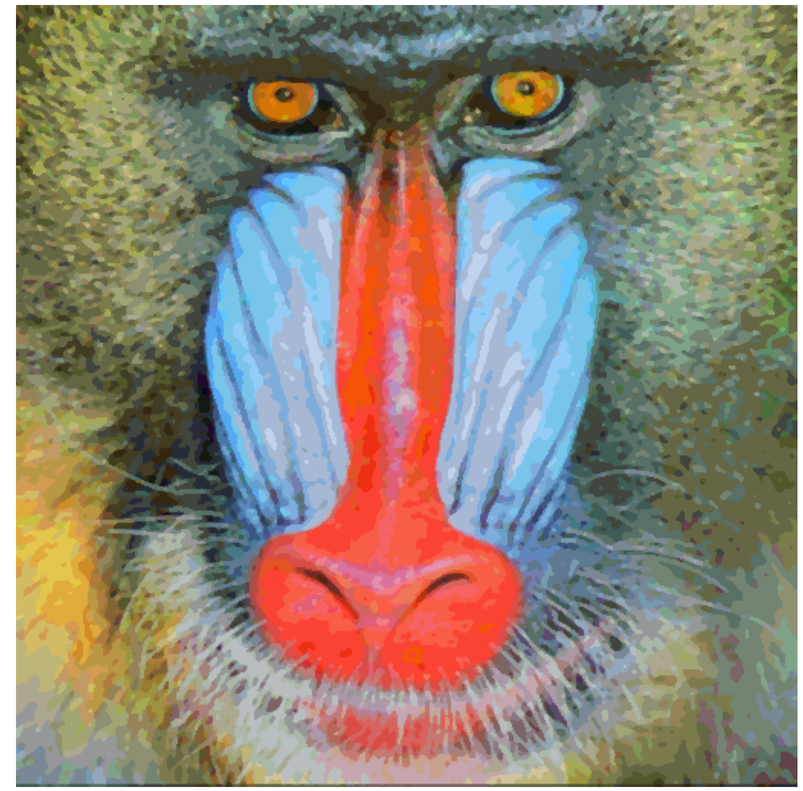
positive-def. manifold (dim = 120)

supervised: 200 labels

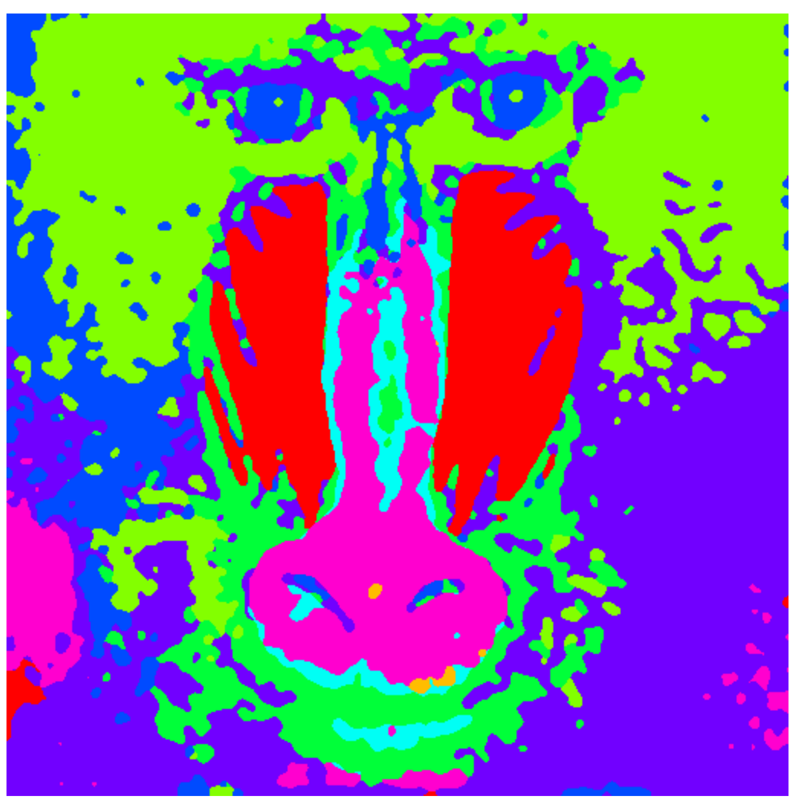
$$\mathcal{P}_d \ni F_i = \int h(x_i - y)\big[(f - \mathbb{E}_i[f]) \otimes (f - \mathbb{E}_i[f])\big](y)\,\mathrm{d}y$$

Label evolution

supervised: 200 labels



few labels



few labels

Outline

- set-up: assignment flow
  supervised labeling

- unsupervised labeling
  - label evolution *(Zern et al., GCPR'18)*
  - label learning from scratch *(submitted)*

- parameter estimation (control)

- outlook

Label learning from scratch

$$D_i = \big(d(f_i, g_1), \ldots, d(f_i, g_m)\big)$$ metric, distance vector (**data term**)

$$D = \big(d(f_i, f_k)\big)_{i,k \in I} \xrightarrow{?} W_i = \big(\operatorname{Pr}(g_1|f_i), \ldots, \operatorname{Pr}(g_m|f_i)\big)$$

each datum is a label ?

Label learning from scratch

$$D_i = \big(d(f_i, g_1), \dots, d(f_i, g_m)\big)$$ metric, distance vector (**data term**)

$$\downarrow$$

$$D = \big(d(f_i, f_k)\big)_{i,k \in I} \xrightarrow{?} W_i = \big(\Pr(g_1|f_i), \dots, \Pr(g_m|f_i)\big)$$

$\downarrow$

each datum is a label $\nearrow$ **?**

$$Q(f_i|f_k) = \frac{P(f_k|f_i)P(f_i)}{\sum_{l\in I} P(f_k|f_l)P(f_l)}, \qquad P(f_k) = \frac{1}{|I|}, \; k \in I$$

marginalize over "data labels" $\longrightarrow$ **self-affinity matrix**

$$\boxed{A_{ji}(W)} = \sum_{k\in I} Q(f_j|f_k)P(f_k|f_i) = \big(WC(W)^{-1}W^\top\big)_{ji}$$

symmetric, non-negative, doubly stochastic
***parametrised* by assigments**

Label learning from scratch

objective: spatially regularised ***data self-assignment***

$$\min_{W \in \mathcal{W}} E(W), \qquad E(W) = \langle D, A(W) \rangle \qquad \text{(generalizes } \langle D, W \rangle\text{)}$$

Label learning from scratch

objective: spatially regularised ***data self-assignment***

$$\min_{W \in \mathcal{W}} E(W), \qquad E(W) = \langle D, A(W) \rangle$$

(generalizes $\langle D, W \rangle$)

approach: redefine the likelihood vectors

$$L(W) = \exp_W \left(-\frac{1}{\rho} \nabla E(W) \right) \in \mathcal{W}$$

$$\dot{W} = \Pi_W \big(S(t) \big)$$

unsupervised

self-assignment flow

single parameter: **scale**

Label learning from scratch

objective: spatially regularised ***data self-assignment***

$$\min_{W \in \mathcal{W}} E(W), \qquad E(W) = \langle D, A(W) \rangle$$

(generalizes $\langle D, W \rangle$)

approach: redefine the likelihood vectors

$$L(W) = \exp_W\big(-\tfrac{1}{\rho}\nabla E(W)\big) \in \mathcal{W}$$

$$\dot{W} = \Pi_W\big(S(t)\big)$$

unsupervised

self-assignment flow

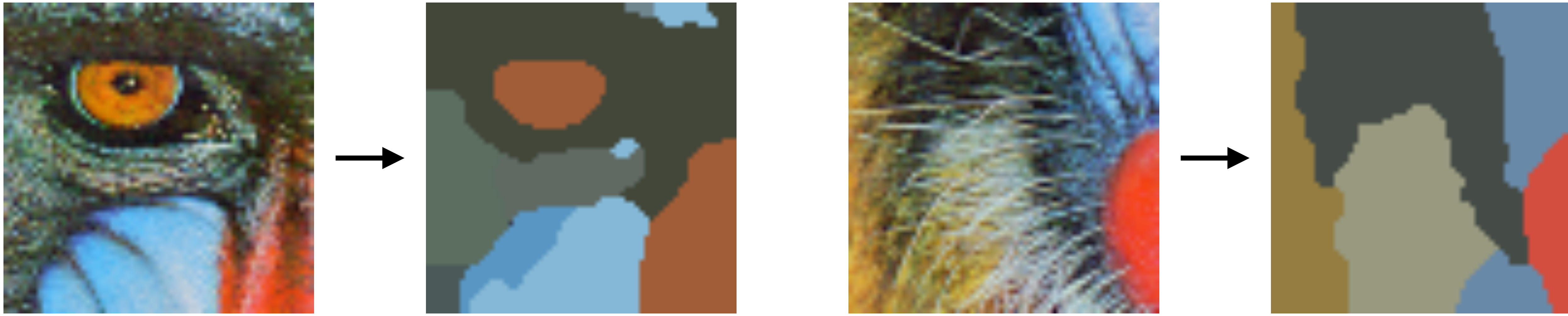
single parameter: **scale**

result: spatially regularised ***discrete optimal transport***

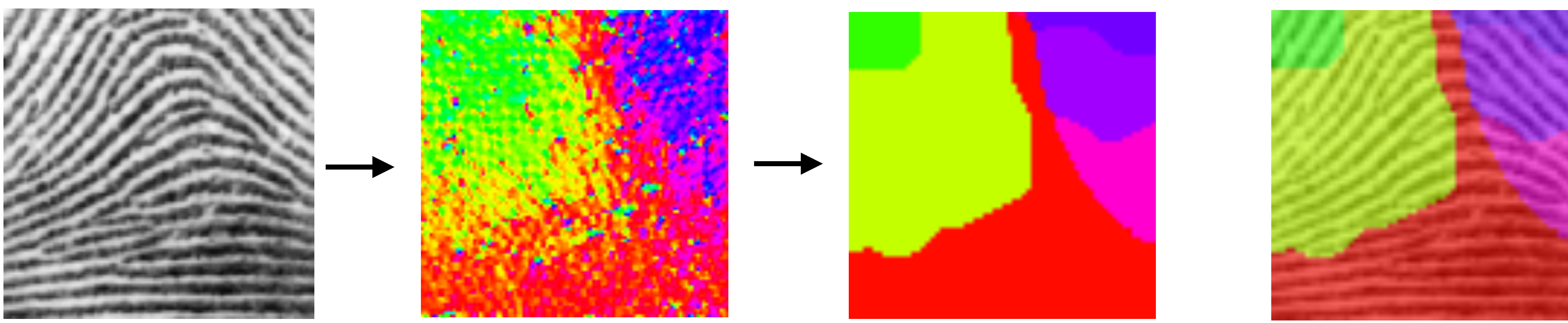
$A(W)$ approaches a **low-rank manifold**

labels g_k *and their number* emerge from data f_i as **latent variables**

Unsupervised self-assignment flow



S^1 - valued data



Outline

- set-up: assignment flow
  supervised labeling

- unsupervised labeling
  - label evolution
  - label learning from scratch

- $\boxed{\text{• parameter estimation (control)}}$ *(submitted)*

- outlook

Recall: regularisation & control parameters

$$D_i = \big(d(f_i, g_1), \dots, d(f_i, g_m)\big)$$

metric, distance vector (**data term**)

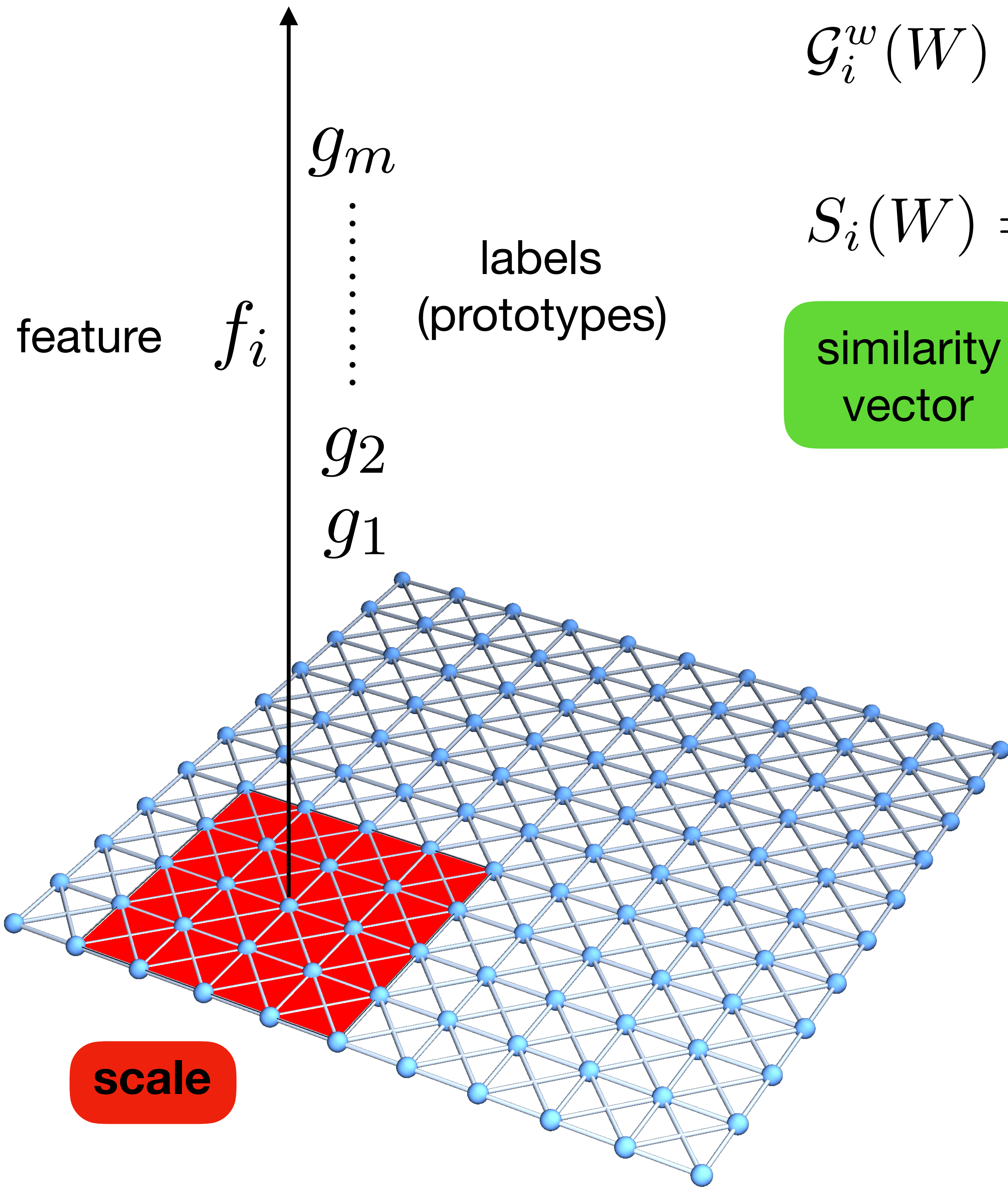
$$\mathcal{G}_i^w(W) = \mathrm{Exp}_{W_i}\Big(\sum_{k \in \mathcal{N}_i} w_{ik}\, \mathrm{Exp}_{W_i}^{-1}(W_k)\Big)$$

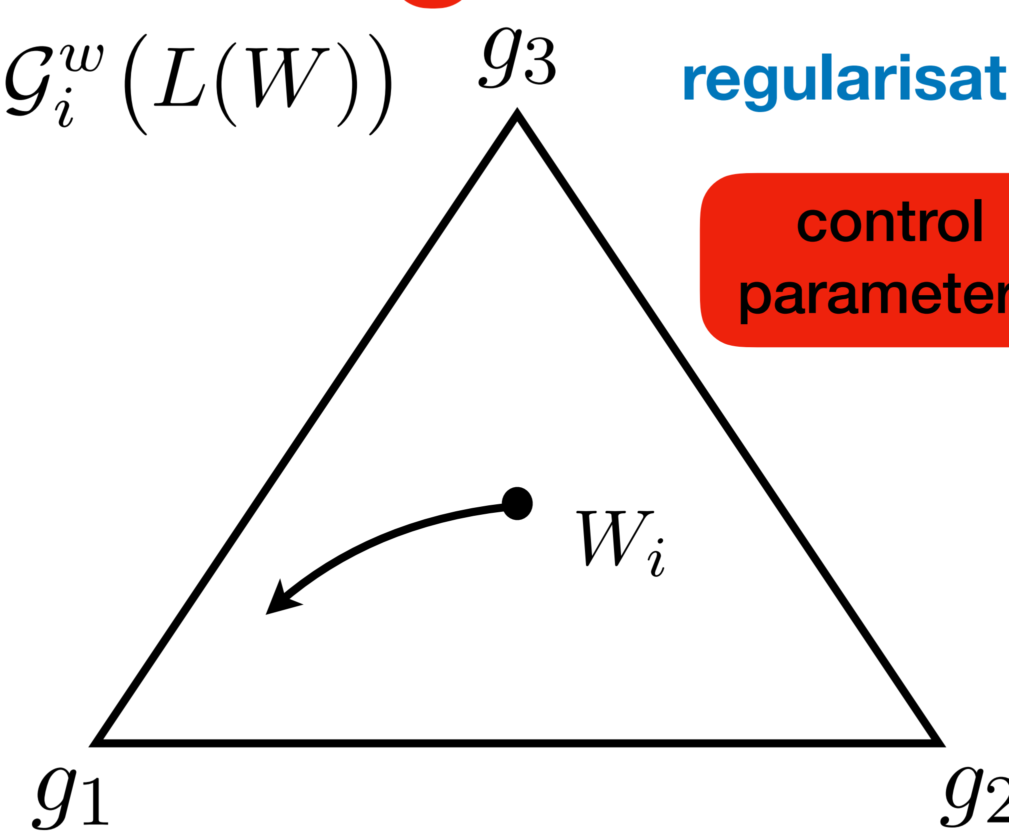
$$S_i(W) = \mathcal{G}_i^w\big(L(W)\big)$$

regularisation

control parameters

similarity vector



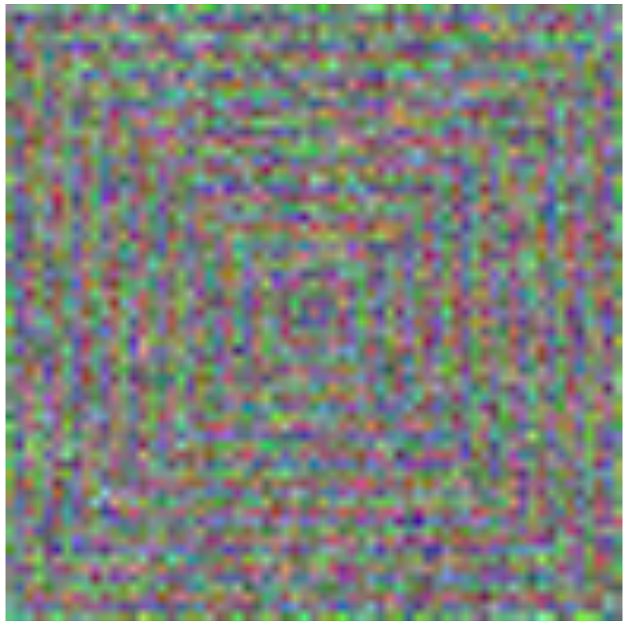


$$f_i \in (\mathcal{X}, d)$$

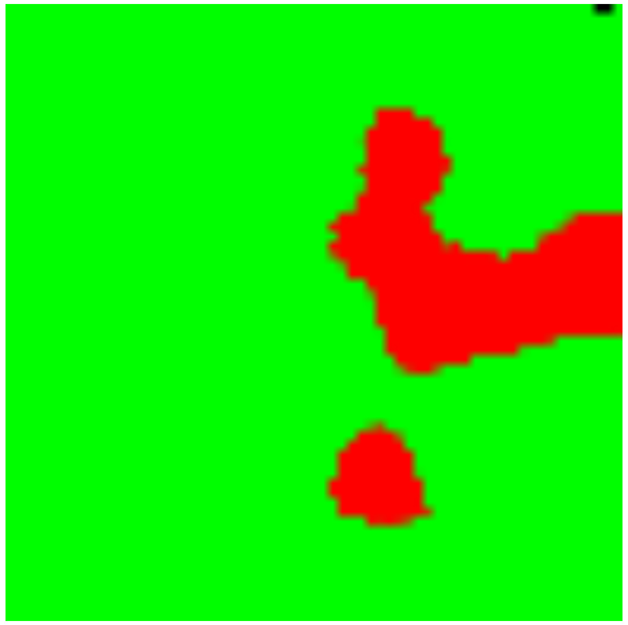
features
metric space

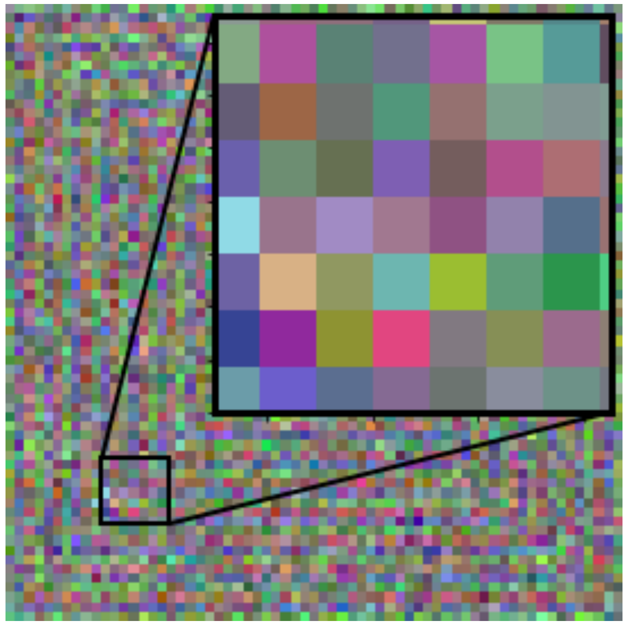
Motivation

all patch-weights $\Omega = \{w_{ik} : k \in \mathcal{N}_i,\ i \in I\} \in \mathcal{P}$ (= weight manifold)

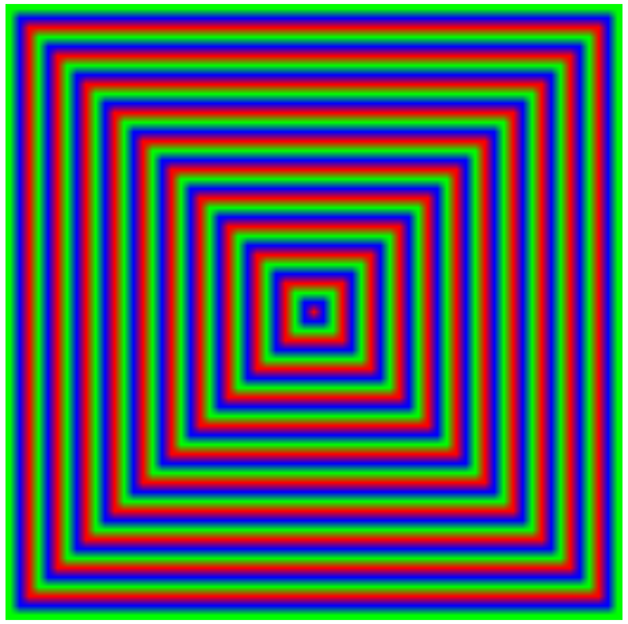


uniform weights





predicted weights
after *learning*



Approach

all patch-weights $\Omega = \{w_{ik} \colon k \in \mathcal{N}_i,\ i \in I\} \in \mathcal{P}$

linear

linear assignment flow

$$W(t) = \mathrm{Exp}_{W_0}\big(V(t)\big), \quad \dot{V} = \Pi_{W_0}\big(S(W_0) + dS_{W_0} V\big), \quad V(0) = 0$$

Approach

all patch-weights $\Omega = \{w_{ik} \colon k \in \mathcal{N}_i,\ i \in I\} \in \mathcal{P}$

linear

linear assignment flow

$$W(t) = \mathrm{Exp}_{W_0}\big(V(t)\big), \quad \dot{V} = \Pi_{W_0}\big(S(W_0) + dS_{W_0} V\big), \quad V(0) = 0$$

objective

ground-assignments (labelings)

$$E\big(V(T)\big) = D_{\mathrm{KL}}\big(W^*, \exp_{\mathbb{1}_{\mathcal{W}}}\big(V(T)\big)\big)$$

Approach

all patch-weights $\Omega = \{w_{ik} \colon k \in \mathcal{N}_i,\ i \in I\} \in \mathcal{P}$

linear

linear assignment flow

$$W(t) = \mathrm{Exp}_{W_0}\big(V(t)\big), \quad \dot{V} = \Pi_{W_0}\big(S(W_0) + dS_{W_0} V\big), \quad V(0) = 0$$

objective

ground-assignments (labelings)

$$E\big(V(T)\big) = D_{\mathrm{KL}}\big(W^*, \exp_{\mathbb{1}_{\mathcal{W}}}\big(V(T)\big)\big)$$

parameter estimation problem

training data for
parameter *prediction*

$$\min_{\Omega \in \mathcal{P}} \quad E\big(V(T)\big)$$

$$\text{s.t.} \quad \dot{V}(t) = f(V(t), \Omega), \quad t \in [0, T], \quad V(0) = 0_{|I| \times |J|}$$

Approach

all patch-weights $\Omega = \{w_{ik} : k \in \mathcal{N}_i,\ i \in I\} \in \mathcal{P}$

linear

linear assignment flow

$$W(t) = \mathrm{Exp}_{W_0}\big(V(t)\big), \quad \dot{V} = \Pi_{W_0}\big(S(W_0) + dS_{W_0} V\big), \quad V(0) = 0$$

objective

ground-assignments (labelings)

$$E\big(V(T)\big) = D_{\mathrm{KL}}\big(W^*, \exp_{\mathbb{1}_{\mathcal{W}}}\big(V(T)\big)\big)$$

parameter estimation problem

training data for parameter *prediction*

$$\min_{\Omega \in \mathcal{P}} \quad E\big(V(T)\big)$$

$$\text{s.t.} \quad \dot{V}(t) = f(V(t), \Omega), \quad t \in [0, T], \quad V(0) = 0_{|I| \times |J|}$$

parameter prediction

novel data

$$\widehat{w} : \mathcal{F}_i \to \mathcal{P}, \qquad (f_k)_{k \in \mathcal{N}_i} \to (w_{ik})_{k \in \mathcal{N}_i},$$

definition & meaning of *what the network learns*!

Parameter estimation algorithm

recurring structure

geometric integration (parameter manifold)

$$\dot{\Omega} = -\nabla_{\mathcal{P}} E\big(V(T,\Omega)\big) = -\Pi_{\Omega}\Big(\frac{\mathrm{d}}{\mathrm{d}\Omega} E\big(V(T,\Omega)\big)\Big), \quad \Omega(0) = \mathbb{1}_{\mathcal{P}}$$

optimize then discretise vs. discretise then optimise ?

Parameter estimation algorithm

recurring structure

gradient flow (weight parameter manifold)

$$\dot{\Omega} = -\nabla_{\mathcal{P}} E\big(V(T,\Omega)\big) = -\Pi_{\Omega}\Big(\frac{\mathrm{d}}{\mathrm{d}\Omega} E\big(V(T,\Omega)\big)\Big), \qquad \Omega(0) = \mathbb{1}_{\mathcal{P}}$$

Either way yields the *same* solution !

$$\begin{tikzcd}
E\big(V(T,\Omega)\big) \text{ s.t. } \dot{V} = f(V,\Omega) \arrow[r, "\text{diffentiate}"] \arrow[d, "\text{discretize}"'] & \textbf{adjoint system} \arrow[d, "\text{discretize}"] \\
\textbf{nonlinear program} \arrow[r, "\text{diffentiate}"'] & \textbf{sensitivity}
\end{tikzcd}$$

Key aspect: symplectic integrator for the joint system

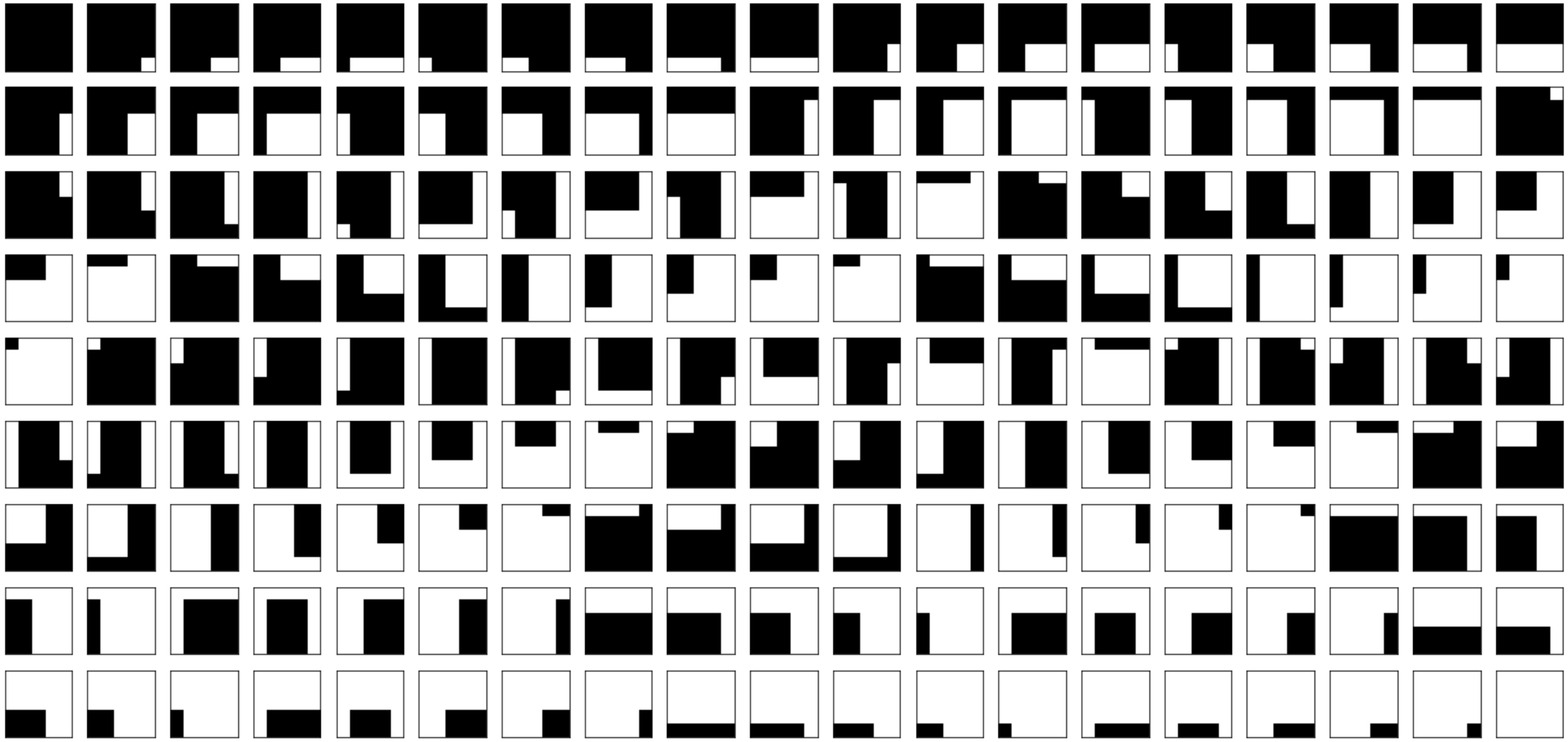
Adaptive regularization

image class:

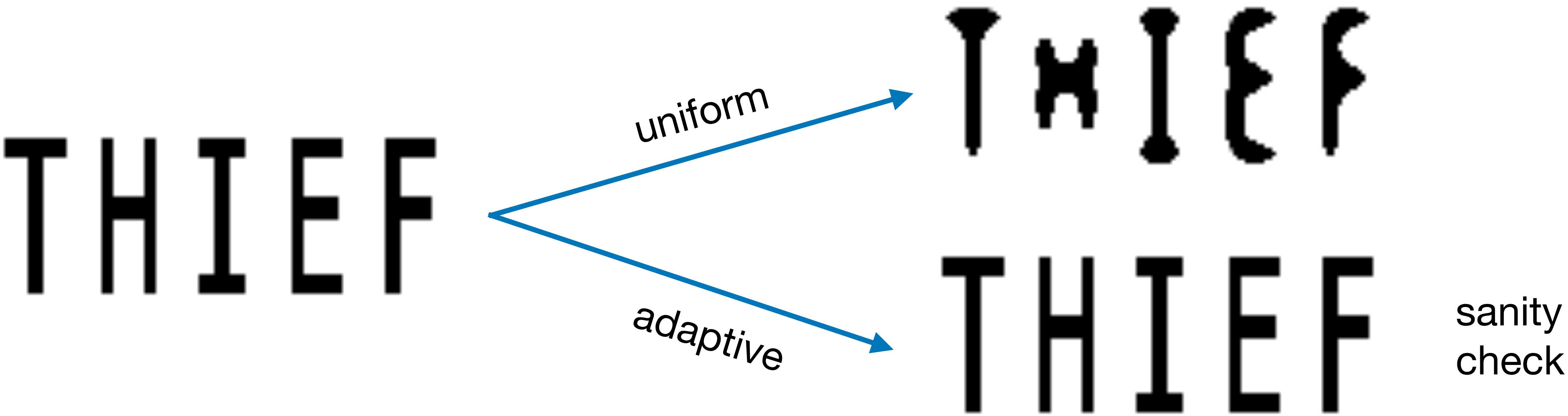
non-curvilinear letters

image features:

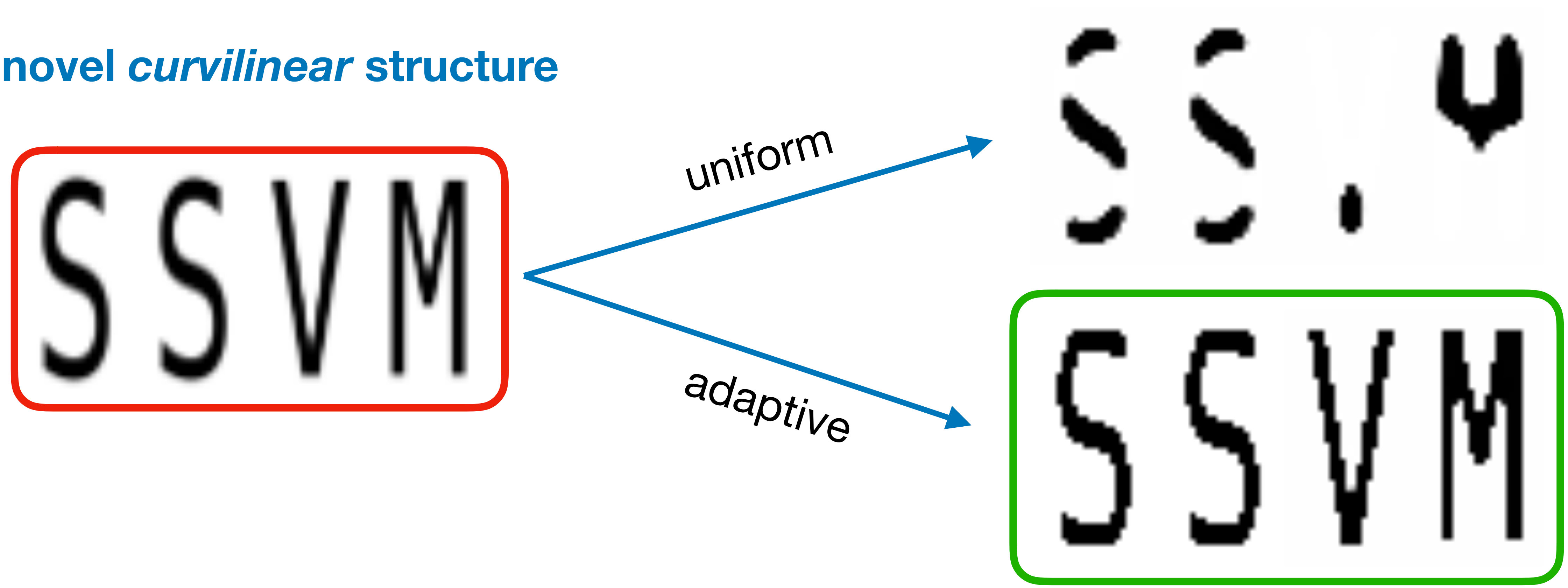
binary patches



Adaptive regularization

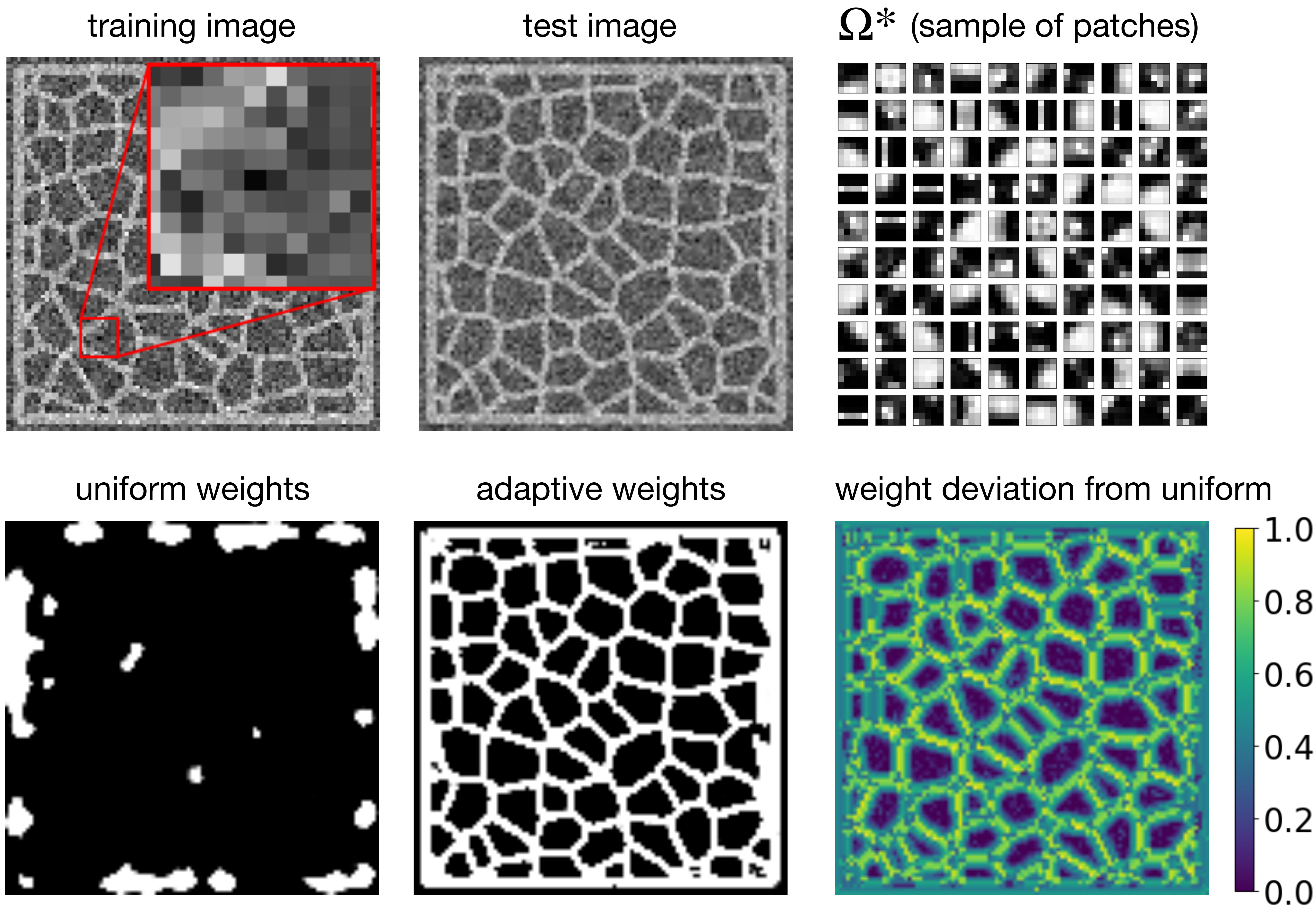


novel *curvilinear* structure



Adaptive regularization

noisy random voronoi images



Outlook

two-grid structure:	**data-driven**	**model-driven**
patch assignment	unsupervised	patch dictionary
pixel assignments	label learning	based prior statistics

dynamic adaptive parameter estimation

model reduction & optimal control $\qquad \widehat{w}\left(\mathscr{F}_{\mathcal{N}_i}, W_{\mathcal{N}_i}(t)\right)$

parameter estimation $\qquad \widehat{w}(\mathscr{F}_{\mathcal{N}_i})$

Publications

prior work: IPA group, Heidelberg

https://ipa.math.uni-heidelberg.de

geometric integration: arXiv:1810.06970, submitted to SI

label learning from scratch: submitted to conference
parameter estimation: full TRs: arXiv soon

synopsis: handbook: variational methods for nonlinear geometric data
and applications, ~ July'19

The Assignment Flow

Christoph Schnörr

Ruben Hühnerbein, Stefania Petra, Fabrizio Savarino,
Alexander Zeilmann, Artjom Zern, Matthias Zisler

Image & Pattern Analysis Group
Heidelberg University

Mathematics of Imaging: W1
Feb 4-8, Paris