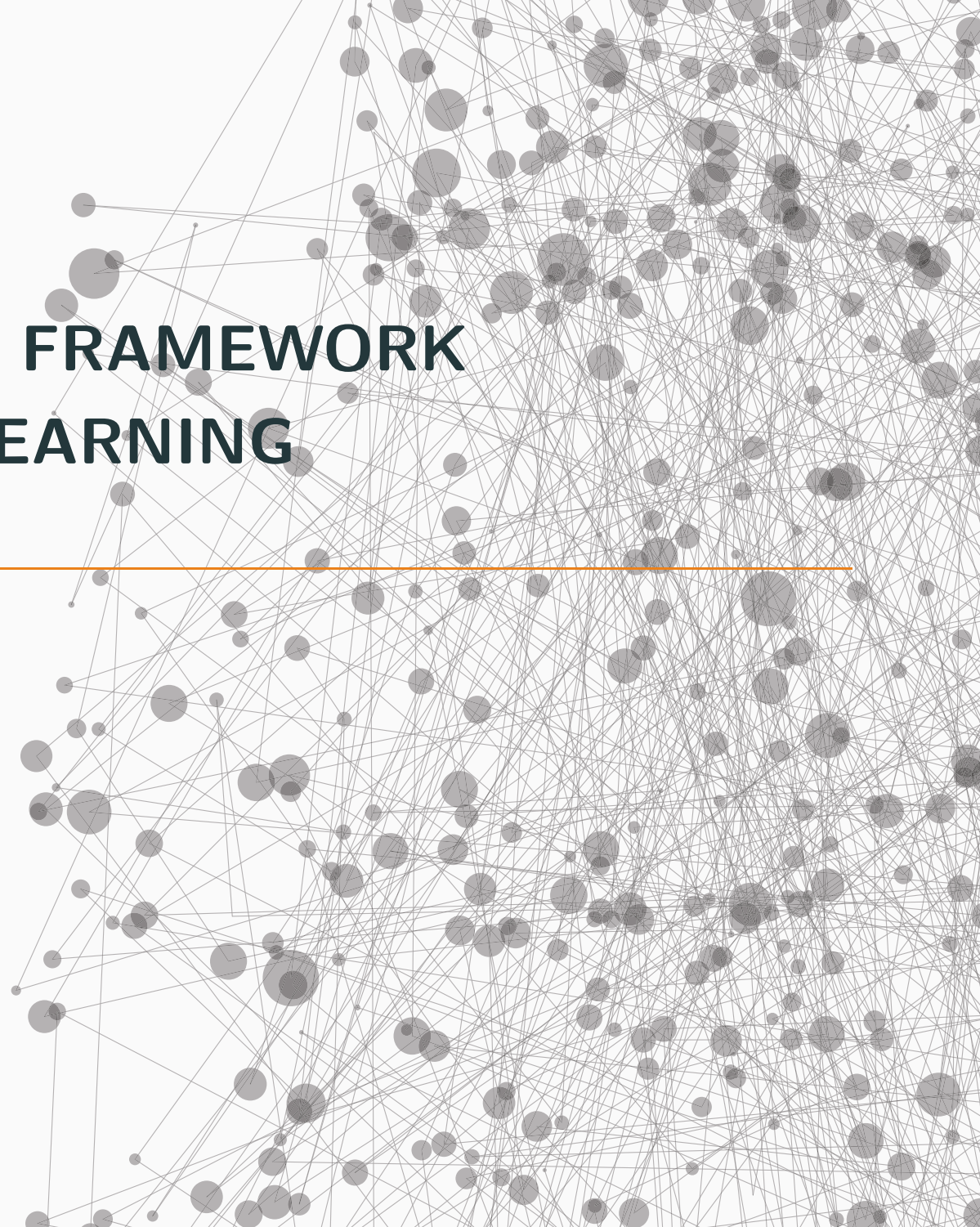


A VARIATIONAL FRAMEWORK FOR LAWS OF LEARNING

Alessandro Betti

October 28, 2020

SAILab, University of Siena

TABLE OF CONTENTS

1. Temporal Environments

2. Variational Laws

3. Vision

4. Causal approach to Vision

TEMPORAL ENVIRONMENTS

LEARNING IN TIME

- In many applications data come as a temporal signal: It is natural therefore to regard learning as an interaction with an environment.
- It allows us to explicitly write dynamical constraints on the learning agent.
- Temporal coordinates are represented in terms of a single real number: Functions in one dimension are much easier to handle than functions that operate on high dimensional feature spaces.

UNDERLYING MODEL

In classic supervised learning tasks we basically need to find the "rule" $x \mapsto f(x)$ that maps a feature vector $x \in X$ into a corresponding prediction $y \in Y$; we have at least two possibilities

- Choose an element f of an appropriate functional space $\mathcal{F}$ where the domain of the functions is X and the codomain is Y by specifying regularity properties[1].
- Use a parametric model $x \mapsto f(w, x)$, in which f is given and we only need to choose the appropriate parameters w in a finite dimensional space.

[1] Tomaso Poggio and Federico Girosi. "Networks for approximation and learning". In: *Proceedings of the IEEE* 78.9 (1990), pp. 1481–1497.

EMBEDDING A MODEL IN TIME

If we allow explicit dependence on time to fully capture the interaction with a *temporal environment* we can dualize the above two alternatives:

- Choose an element f of an appropriate functional space $\mathcal{F}_T$ where the domain of the functions is $X \times [0, T]$ and the codomain is Y, where $[0, T]$ is the temporal domain on which the agent is defined.
- Use a parametric model $x \mapsto f(w, x)$, in which f is given but where now the w is chosen in a functional space of maps $t \mapsto w(t)$.

EMBEDDING A MODEL IN TIME CONT.

While the first of the two approaches always require to solve a problem on a high dimensional space (how can we impose meaningful regularity properties without introducing a strong bias?) the last approach drastically simplify the problem.

However how should we choose the function f?

Recent successes in AI and well known theorem on the approximation capabilities of Neural Networks suggest this class of functions as a possible candidate.

SELECTION OF $t \mapsto w(t)$

We would like to select the trajectory of the parameters $w(t)$ based on the following principles:

1. Causality;
2. Temporal locality;
3. Optimization of a task-related *loss function*.

MACHANICS-BASED APPROACH

Classical mechanics-based dynamics based on a potential function with dissipation for the evolution model fit all this three requirements:

$$\begin{cases} \mu\ddot{w}(t) + \theta\dot{w}(t) + \nabla U(t, w(t)) = 0. \\ w(0) = w^0 \\ \dot{w}(0) = w^1 \end{cases}$$

Where U is a potential that describes the interaction with the environment trough its dependence on time[2].

[2] Alessandro Betti and Marco Gori. "The principle of least cognitive action". In: *Theoretical Computer Science* 633 (2016), pp. 83–99.

BATCH MODE

A perfect match with classical statistical machine learning can be drawn for batch mode learning with the following assumptions

- Viscous limit $\mu \to 0$;
- $U(t, \omega) := \sum_{k=1}^{\ell} v(\omega, x_k) \equiv U(\omega)$, where $v(\omega, \cdot)$ is a loss function that measure the goodness of the example-target pair x_k when the values of the parameters of the model are ω.

In this limit the dynamics of the weights reduce to (the continuous version) of the classical gradient descent algorithm.

SGD REDUCTION

On the other hand when we allow temporal dependencies on the potential (still in the viscous limit) we recover the computational scheme of Stochastic Gradient Descent[3]:

$$\dot{w}(t) = -\frac{1}{\theta}\nabla U(t, w(t)).$$

[3] Marco Gori, Marco Maggini, and Alessandro Rossi. "Neural network training as a dissipative process". In: *Neural Networks* 81 (2016), pp. 72–80.

LEARNING CONSISTENCY

When μ is fixed and when we allow explicit temporal dependence on the potential, what can we say about the quality of the weights that are developed following this ODE?

Under periodicity assumptions on the input signal $x(t)$ also $w(t)$ assumes a periodic behaviour so that on similar patterns the prediction of the model should be consistently similar[4].

[4] Giovanni Bellettini, Alessandro Betti, and Marco Gori. "Generalization in quasi-periodic environments". In: *arXiv preprint arXiv:1807.05343* (2018).

VARIATIONAL LAWS

VARIATIONAL APPROACHES

Is there a more systematic approach to select parameters trajectories?

- Variational calculus has proven to be particularly suitable for parsimonious description of natural phenomena and laws on general.

VARIATIONAL APPROACHES

Is there a more systematic approach to select parameters trajectories?

- Variational calculus has proven to be particularly suitable for parsimonious description of natural phenomena and laws on general.
- It allows us to go from static (the selection of values of w which get the gradient of the potential null) conditions to dynamic (the choice of an entire trajectory that because of regularity terms ends up being an ODE).

VARIATIONAL APPROACHES

Is there a more systematic approach to select parameters trajectories?

- Variational calculus has proven to be particularly suitable for parsimonious description of natural phenomena and laws on general.
- It allows us to go from static (the selection of values of w which get the gradient of the potential null) conditions to dynamic (the choice of an entire trajectory that because of regularity terms ends up being an ODE).

$$\nabla U = 0 \longrightarrow \delta(\text{something}) = 0.$$

CLASSICAL MECHANICS

Inspired by the above remarks we can look how classical mechanics is formulated in terms if calculus of variations.

Usually Hamilton's principle, once we define the functional

$$S(x) := \int_0^T \frac{1}{2}|\dot{x}|^2 - U(x) = \int_0^T L(x, \dot{x}),$$

is described as follows

Hamilton's principle

The solutions of the equations of Newtonian mechanic for a system described by a potential U coincides with the extremals of the functional S.

Unfortunately this statement is far from being precise as it fails to address in what kind of space of functions the functional S is defined, and it fails to specify what we mean by the extremal of a functional.

A more precise reformulation of Hamilton's principle could be

Hamilton's principle

The solution of the equations of Newtonian mechanic $\bar{x}(t)$ for a system described by a potential U coincides with the stationary (critical) point (when it is unique) of the functional S(x) defined over the set

$$X := \{x \in C^\infty([0, T]) : x(0) = \bar{x}(0), x(T) = \bar{x}(T)\}.$$

STILL NOT THERE YET...

However also this formulation has at least two major problems:

- The formulation does not include dissipation.
- In this formulation the causality principle is not met because of the way in which the functional space X is defined.

FIX DISSIPATION

The first of the two issues can be solved, still using the same framework by introducing an exponential weight in front of the Lagrangian:

$$S(x) \longrightarrow \int_0^T e^{\theta t} \left(\frac{1}{2} |\dot{x}(t)|^2 - U(x(t)) \right) \, dt.$$

Nonetheless the causality issue still remains.

BOUNDARY CONDITIONS VS CAUSALITY

The problem here is due to an inherent incompatibility between

Evolution problems and Variational problems for integral functionals.

Some different ideas to win back causality in this variational framework is therefore needed.

4. Limiti di problemi variazionali collegabili ad equazioni iperboliche

Nella ricerca di eventuali approssimazioni di problemi difficili ed instabili con problemi più facili e più stabili si può far rientrare l'idea di ottenere soluzioni di problemi di evoluzione come limite delle soluzioni di problemi di minimo. Un esempio interessante in cui quest'idea è stata applicata con successo si trova nel lavoro [8]. Una possibile variante del risultato di Ilmanen è data dalla seguente congettura.

CONGETTURA 1 – *Siano* $\varphi, \psi \in C_0^\infty(\mathbf{R}^n)$, $k > 1$ *intero; per ogni numero reale positivo* λ, $w_\lambda = w_\lambda(x_1, \dots, x_n, t)$ *fornisca il minimo del funzionale*

$$F_\lambda(u) = \int_{\mathbf{R}^n \times [0,+\infty[} e^{-\lambda t} \left[\left| \frac{\partial^2 u}{\partial t^2} \right|^2 + \lambda^2 |\nabla_x u|^2 + \lambda^2 u^{2k} \right] dx\, dt \tag{1}$$

nella classe delle u *soddisfacenti le condizioni iniziali*

$$u(x,0) = \varphi(x), \qquad \frac{\partial u}{\partial t}(x,0) = \psi(x).$$

Allora esiste il $\lim_{\lambda \to +\infty} w_\lambda(x,t) = w_0(x,t)$, *soddisfacente l'equazione*

$$\frac{\partial^2 w_0}{\partial t^2} = \Delta_x w_0 - k w_0^{2k-1}. \tag{2}$$

A NEW PRINCIPLE FOR MECHANICS

Recently[5] a two step reformulation of Classical Mechanics with dissipation based on the De Giorgi conjecture has been proposed:

- Fix ε and minimize the functional

$$F_\varepsilon(w) = \int_0^T e^{-t/\varepsilon} \left(\varepsilon^2 \frac{\mu}{2} |\ddot{w}(t)|^2 + \varepsilon \frac{\nu}{2} |\dot{w}(t)|^2 + U(w(t)) \right) dt,$$

  on the set $\{w \in H^2((0,T); \mathbf{R}^N) : w(0) = w^0, \dot{w}(0) = w^1\}$;

- take the limit of the minima as $\varepsilon \to 0$.

[5] Matthias Liero and Ulisse Stefanelli. "A new minimum principle for Lagrangian mechanics". In: *Journal of nonlinear science* 23.2 (2013), pp. 179–204.

RELEVANCE IN LEARNING i

This approach seems also particularly suitable for learning:

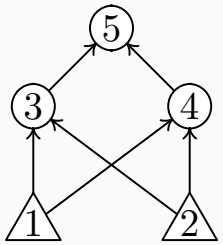
- All the previous comments on the use of Newton's law in learning applies also here;
- Moreover, with this approach, it is immediate to recover gradient dynamics ($\mu = 0$);
- At least when $T < \infty$ the approach can be followed also for time dependent potentials;
- It offers a very nice interpretations in terms of minima (cf. with the "$-$" in the Hamilton Principle);
- It is well posed with respect to the possibility of adding nonholonomous penalties of the form $P(t, w, \dot{w})$.

RELEVANCE IN LEARNING ii

- The variational framework also offers the possibility to define the learning problems for NN in a particularly satisfying way in terms of constraints.
  - In particular both the architecture and the interaction with the environment can be regarded as constraints to the variational problem.
  - Each constraint will produce a reaction just like it happens in the inclined plane problem in mechanics.

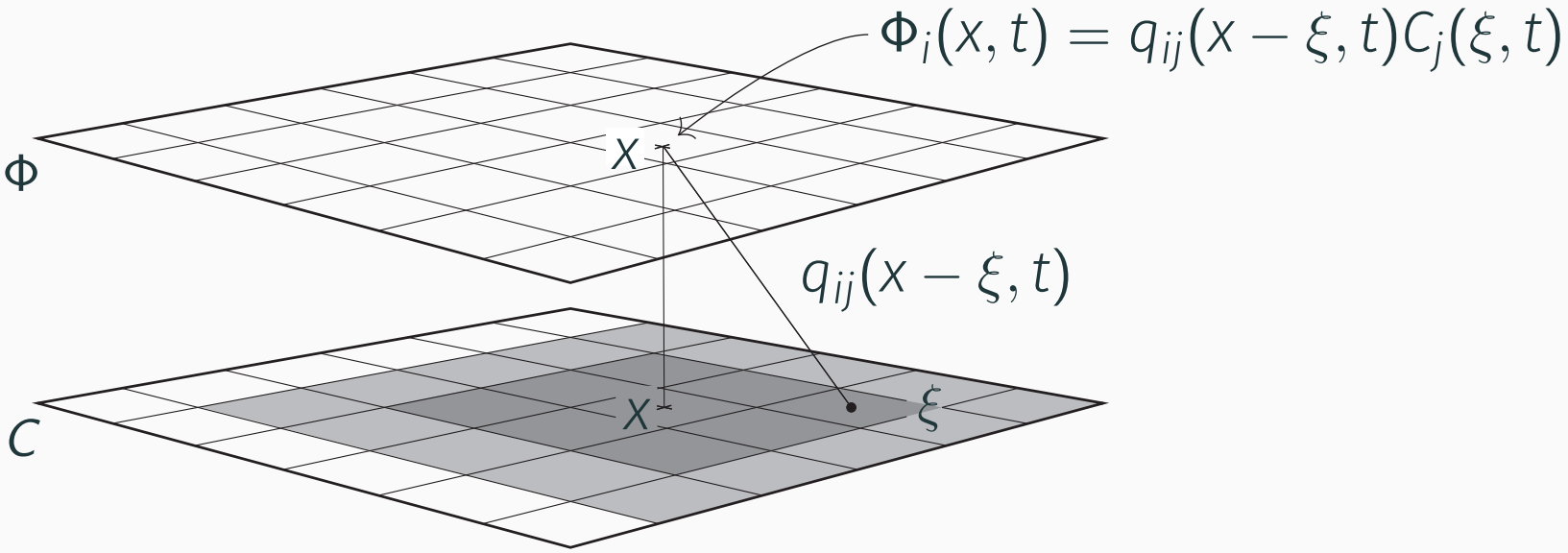
$$G^1 = x^1 - e^1, \quad G^2 = x^2 - e^2, \quad G^3 = x^3 - \sigma(w_{31}x^1 + w_{32}x^2);$$

$$G^4 = x^4 - \sigma(w_{41}x^1 + w_{42}x^2), \quad G^5 = x^5 - \sigma(w_{53}x^3 + w_{54}x^4).$$



VISION

VISUAL FEATURES EXTRACTION



The field q is the variable of our problem and Φ_i is the i-th convolutional feature map.

FUNCTIONAL CRITERION

In particular the functional index that we are interested in is

$$\begin{aligned}\mathcal{A}(q) =& \frac{1}{2}\left(\int_D d\mu\, \Phi_i\right)^2 - \frac{\lambda_C}{2}\int_D d\mu\, \Phi_i^2 \\ &+ \frac{\lambda_P}{2}\int_D dtdx\, h(t)(P_x q_{ij})^2 + \frac{\lambda_K}{2}\int_D dtdx\, h(t)(P_t q_{ij})^2 \\ &+ \frac{\lambda_M}{2}\int_D d\mu\, \left(\partial_t \Phi_i + v_j \partial_j \Phi_i\right)^2.\end{aligned}$$

Where $d\mu(x,t) = h(t)g(x)\,dx\,dt$, $v_j(x,t)$ is the velocity field (optical flow) and P_x and P_t suitable differential operators.

- Negative entropy
- Conditional entropy
- Regularization
- Motion invariance

VECTORIZATION

Now assume that we are working on a discrete retina $X^\sharp$, then we can rearrange the field variable of our problem $q_{ij}(x,t)$ into into a tensor with temporal dependence:

$$q_{ij}(x,t) \xrightarrow{\text{disc}} q_{ijx}(t) \xrightarrow{\text{vec}} q_k(t).$$

Where if we had n features, m image channels and a retina with ℓ^2 pixels, $q \in \mathbf{R}^{n\times m\times \ell^2}$.

$$\Gamma(q) = \int_0^T h(t)\left(\frac{\mu}{2}|\ddot{q}|^2 + \frac{\nu}{2}|\dot{q}|^2 + \gamma\dot{q}\cdot\ddot{q} + \frac{k}{2}|q|^2 + U(q,C)\right)dt + \lambda_M \mathcal{M}(q),$$

$$\mathcal{M}(q) := \int_0^T dt\, h(t)\left(\frac{1}{2}\dot{q}M^\natural(t)\dot{q} + qN^\natural(t)\dot{q} + \frac{1}{2}q(t)O^\natural(t)q(t)\right),$$

- Negative entropy
- Conditional entropy
- Regularization
- Motion invariance

EULER LAGRANGE

We have:

Theorem

The functional Γ*, admits a minimum on the set*

$$X = \{\, q \in H^2((0,T), \mathbf{R}^n) \mid q(0) = q^0, \dot{q}(0) = q^1 \,\}.$$

The Euler-Lagrange equation relative to the functional $\Gamma(q)$ defined on X are

$$\begin{cases} \hat{\mu}(t)q^{(4)}(t)+2\dot{\hat{\mu}}(t)q^{(3)}(t)+Z_2(t)\ddot{q}(t)+Z_1(t)\dot{q}(t)+Z_0(t)q(t)+\nabla_q \hat{U}(q,C) = 0. \\ \hat{\mu}\ddot{q}(T)+\hat{\gamma}\dot{q}(T) = 0; \\ -\hat{\mu}q^{(3)}(T)-\dot{\hat{\mu}}\ddot{q}(T)+(\hat{\nu}-\dot{\hat{\gamma}}+\lambda_M \hat{M}^{\natural})\dot{q}(T)+\lambda_M(\hat{N}^{\sharp})'q(T) = 0. \end{cases}$$

We adopted the convention $\hat{f}(t) = h(t)f(t)$.

CAUSAL APPROACH TO VISION

CAUSAL FORMULATION

A further development of this theory of vision is that of using the causal approach that follows from De Giorgi conjecture:

$$F_\varepsilon(q) := \int_0^T e^{-t/\varepsilon}\left(\varepsilon^2\frac{\rho}{2}|\ddot{q}|^2 + \varepsilon\frac{\nu}{2}|\dot{q}|^2 + \lambda_M\left(\frac{\varepsilon}{2}\dot{q}\cdot M^\natural\dot{q} + \varepsilon q\cdot N^\natural\dot{q} + \frac{1}{2}q\cdot O^\natural q\right) + U(q,C)\right)dt.$$

EULER EQUATIONS

A formal limit in the Euler Equations associated with this functional leads to the differential equation:

$$\rho\ddot{q} + (\nu + \lambda_M M^{\natural})\dot{q} + \lambda_M N^{\natural\prime} q + \lambda_M O^{\natural} q + \nabla U(q, C) = 0,$$

that is intended to be solved with the boundary conditions

$$q(0) = q^0, \qquad \dot{q}(0) = q^1,$$

with q^0 and q^1 assigned vectors.

EULER EQUATIONS CONT.

Compared with the previous formulation of the theory we have:

- Second order differential equations;
- No additional boundary conditions at $t = T$.

PLOTS

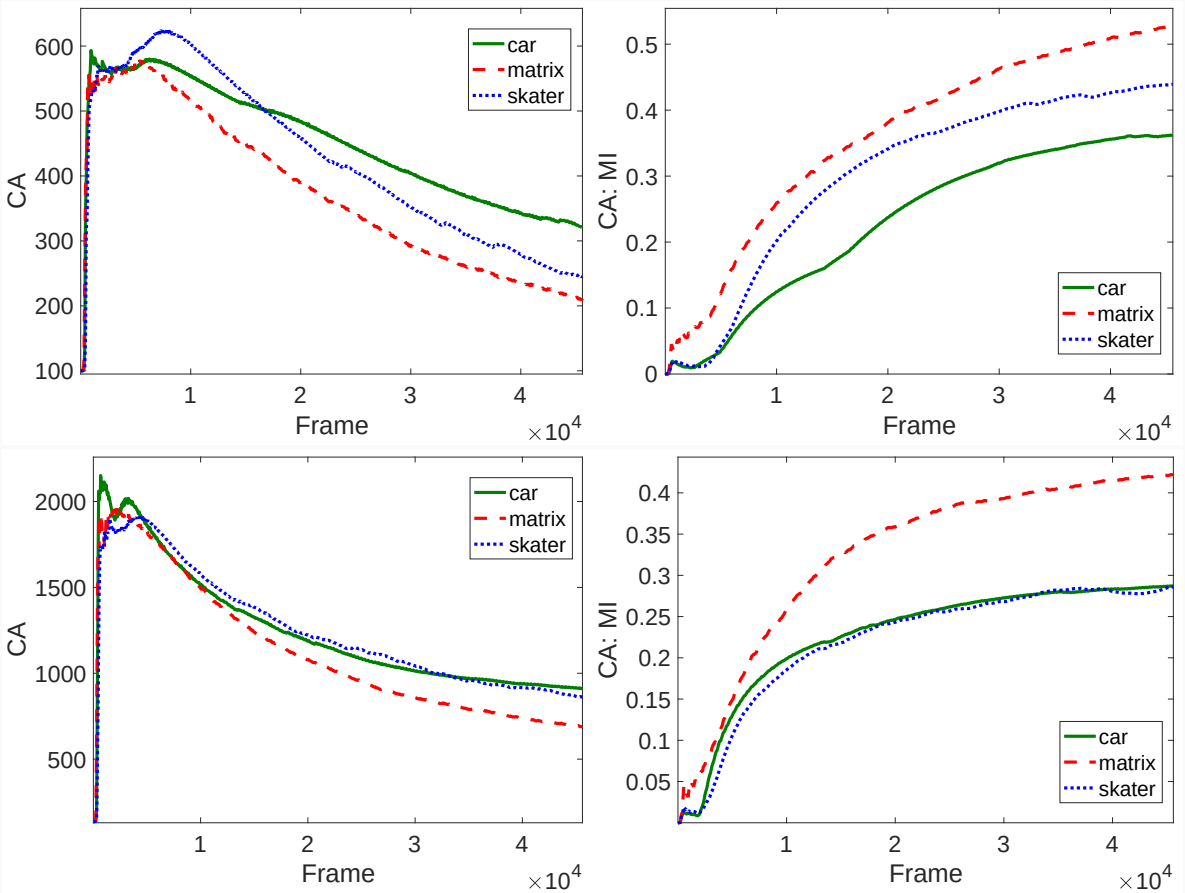


Figure 1: Different number of features and filter sizes (1st row: $n = 5, size = 5 \times 5$; 2nd row: $n = 11, size = 11 \times 11$) in 3 videos.

		$\lambda_M = 0$	10^{-8}	10^{-6}	10^{-4}
Skater	$\ell = 1$	.61±.11	.54 ± .11	.52 ± .07	.53 ± .08
	$\ell = 2$	.53 ± .12	**.62** ± .15	.60 ± .11	.43 ± .06
	$\ell = 3$	.56 ± .17	.58 ± .20	**.62** ± .10	.18 ± .16
Car	$\ell = 1$	.49 ± .05	.44 ± .02	.46 ± .04	.47 ± .04
	$\ell = 2$	.25 ± .26	.54 ± .10	**.65** ± .08	.46 ± .03
	$\ell = 3$	.26 ± .34	.45 ± .22	**.51** ± .11	.38 ± .20
Matrix	$\ell = 1$	.66 ± .01	.66 ± .02	**.67** ± .01	.63 ± .05
	$\ell = 2$	.55 ± .13	.56 ± .14	.43 ± 0	.45 ± .04
	$\ell = 3$	**.64** ± .03	.54 ± .11	.35 ± .07	.40 ± .01

		$\lambda_M = 10^{-2}$	1	10^2
Skater	$\ell = 1$	**.69** ± .07	.53 ± 0	.01 ± 0
	$\ell = 2$	.48 ± .06	.1 ± .1	.03 ± .01
	$\ell = 3$	.16 ± .17	.04 ± .02	.03 ± .02
Car	$\ell = 1$	**.66** ± .10	.60 ± .02	.01 ± 0
	$\ell = 2$	.63 ± .11	.18 ± .32	.03 ± .01
	$\ell = 3$	.24 ± .20	.09 ± .12	.04 ± .02
Matrix	$\ell = 1$	.59 ± .03	.44 ± 0	.23 ± .02
	$\ell = 2$	**.62** ± .02	.35 ± .19	.13 ± .08
	$\ell = 3$	.21 ± .07	.06 ± .03	.04 ± .02

Table 1: MI in different videos, up to 3 layers ($\ell = 1, 2, 3$), and for multiple λ_M of the motion-based term. All layers share the same λ_M.

FOR A CLOSER LOOK

- Alessandro Betti, Marco Gori, and Stefano Melacci. "Cognitive action laws: the case of visual features". In: *IEEE transactions on neural networks and learning systems* 31.3 (2019), pp. 938–949
- Alessandro Betti, Marco Gori, and Stefano Melacci. "Learning visual features under motion invariance". In: *Neural Networks* (2020)
- Matteo Tiezzi et al. *Focus of Attention Improves Information Transfer in Visual Features*. 2020. arXiv: 2006.09229 [cs.LG]

Thank you for listening!

REFERENCES

- Bellettini, Giovanni, Alessandro Betti, and Marco Gori. "Generalization in quasi-periodic environments". In: *arXiv preprint arXiv:1807.05343* (2018).
- Betti, Alessandro and Marco Gori. "The principle of least cognitive action". In: *Theoretical Computer Science* 633 (2016), pp. 83–99.
- Betti, Alessandro, Marco Gori, and Stefano Melacci. "Cognitive action laws: the case of visual features". In: *IEEE transactions on neural networks and learning systems* 31.3 (2019), pp. 938–949.
- – ."Learning visual features under motion invariance". In: *Neural Networks* (2020).
- Gori, Marco, Marco Maggini, and Alessandro Rossi. "Neural network training as a dissipative process". In: *Neural Networks* 81 (2016), pp. 72–80.


Liero, Matthias and Ulisse Stefanelli. “A new minimum principle for Lagrangian mechanics”. In: *Journal of nonlinear science* 23.2 (2013), pp. 179–204.


Poggio, Tomaso and Federico Girosi. “Networks for approximation and learning”. In: *Proceedings of the IEEE* 78.9 (1990), pp. 1481–1497.


Tiezzi, Matteo et al. *Focus of Attention Improves Information Transfer in Visual Features*. 2020. arXiv: 2006.09229 [cs.LG].

BACKUP SLIDES–SEQUENTIAL OPTIMIZATION i

Consider

$$P_\varepsilon^k : \min_{w \in \mathbf{X}_\varepsilon^k} J_\varepsilon^k(w), \quad k = 1, \dots, K$$

where J_ε^k are defined on a fixed temporal domains

$$J_\varepsilon^k(w) := \int_{t_{k-1}}^{t_k} \varpi_\varepsilon(t) \left(\varepsilon^2 \frac{\rho}{2} |\ddot{w}(t)|^2 + \varepsilon \frac{\nu}{2} |\dot{w}(t)|^2 + U(w(t), t) \right) dt$$

and the set $\mathbf{X}_\varepsilon^k \equiv \mathbf{X}^k$ with

$$\mathbf{X}^k := \begin{cases} \{w \in H^2(0, t_1) : w(0) = w^0, \dot{w}(0) = w^1\} & \text{if } k \\ \{w \in H^2(t_{k-1}, t_k) : w(t_{k-1}) = w_\varepsilon^k(t_{k-1}), \dot{w}(t_{k-1}) = \dot{w}_\varepsilon^k(t_{k-1})\} & \text{if } k \end{cases}$$

BACKUP SLIDES–SEQUENTIAL OPTIMIZATION ii

where w_ε^k is the solution of the problem P_ε^k. Also let

$$\bar{w}_\varepsilon(t) := \begin{cases} w^0 & \text{if } t = 0; \\ w_\varepsilon^k(t) & \text{if } t_{k-1} < t \leq t_k. \end{cases}$$

Then we conjecture that

$$\bar{w}_\varepsilon \to w,$$

where w solves

$$\begin{cases} \rho \ddot{w}(t) + \nu \dot{w}(t) + \nabla U(w(t), t) = 0; \\ w(0) = w^0, \qquad \dot{w}(0) = w^1. \end{cases}$$

BACKUP SLIDES–MUTUAL INFORMATION

Conditional entropy

$$S(Y \mid X, T, F) = -\int_{\Omega} \sum_{i=1}^{n} dP_{X,T,F}\ p_i \log p_i$$

$$= -\int_{D} d\mu(x,t)\ \sum_{i=1}^{n} \Phi_i(x,t) \log \Phi_i(x,t)$$

Entropy

$$S(Y) = -\sum_{i=1}^{n} \Pr(Y = y_i) \log \Pr(Y = y_i)$$

$$= -\sum_{i=1}^{n} \left(\int_{D} d\mu(x,t)\, \Phi_i(x,t)\right) \cdot \log\left(\int_{D} d\mu(x,t)\, \Phi_i(x,t)\right)$$

BACKUP SLIDES–FUNCTIONAL CRITERION

In particular the functional index that we are interested in is

$$\begin{aligned}\mathcal{A}(q) =& \frac{1}{2}\left(\int_D d\mu\, \Phi_i\right)^2 - \frac{\lambda_C}{2}\int_D d\mu\, \Phi_i^2 \\ &+ \frac{\lambda_1}{2}\int_D d\mu \left(\sum_{i=0}^{n-1}\Phi_i - 1\right)^2 - \lambda_0 \int_D d\mu\, \Phi_i \cdot [\Phi_i < 0] \\ &+ \frac{\lambda_P}{2}\int_D dtdx\, h(t)(P_x q_{ij})^2 + \frac{\lambda_K}{2}\int_D dtdx\, h(t)(P_t q_{ij})^2 \\ &+ \frac{\lambda_M}{2}\int_D d\mu \left(\partial_t \Phi_i + v_j \partial_j \Phi_i\right)^2.\end{aligned}$$

Where $d\mu(x,t) = h(t)g(x)\,dx\,dt$, $v_j(x,t)$ is the velocity field (optical flow) and P_x and P_t suitable differential operators.

- Negative entropy
- Conditional entropy
- Normalization
- Regularization
- Motion invariance

BACKUP SLIDES–THEOREM (A MORE PRECISE STATEMENT)

Theorem

Let $\mu = \alpha + \gamma_2^2$, $\nu = \beta + \gamma_1^2$ *and* $\gamma = \gamma_1\gamma_2$, *if* $\mu > \gamma_2^2$, $\nu > \gamma_1^2$, $k > 0$ *then the functional* Γ, *admits a minimum on the set*

$$X = \{\, q \in H^2((0,T), \mathbf{R}^n) \mid q(0) = q^0, \dot{q}(0) = q^1 \,\}.$$

BACKUP SLIDES–COEFFICIENTS

$$Z_0 = \hat{k} + \lambda_M \hat{O}^\natural - \lambda_M (\dot{\hat{N}}^\natural)';$$
$$Z_1 = \ddot{\hat{\gamma}} - \dot{\hat{\nu}} - \lambda_M (\dot{\hat{M}}^\natural + (\hat{N}^\natural)' - \hat{N}^\natural);$$
$$Z_2 = \ddot{\hat{\mu}} + \dot{\hat{\gamma}} - \hat{\nu} - \lambda_M \hat{M}^\natural.$$

BACKUP SLIDES–FOCUS OF ATTENTION

Suppose that we know the trajectory $a(t)$ of an attention mechanism, then the motion invariance term can be redefined as

$$\begin{aligned}&\frac{\lambda_M}{2}\int_0^T dt\, h(t)\left(\frac{d\Phi_i(a(t),t)}{dt}\right)^2\\ =&\frac{\lambda_M}{2}\int_0^T dt\, h(t)\big(\partial_t\Phi_i(a(t),t)+\nabla_x\Phi_i(a(t),t)\cdot\dot{a}(t)\big)^2,\end{aligned}$$

where $\dot{a}(t)$ is the velocity of the focus of attention.

BACKUP SLIDES–CHOICE OF F_M, F_N AND F_O.

$$F_\varepsilon(q) := \int_0^T e^{-t/\varepsilon}\Big(\varepsilon^2\frac{\rho}{2}|\ddot{q}|^2 + \varepsilon\frac{\nu}{2}|\dot{q}|^2 + \lambda_M\Big(\frac{f_M(\varepsilon)}{2}\dot{q}\cdot M^\natural\dot{q} + f_N(\varepsilon)q\cdot N^\natural$$
$$+ \frac{f_O(\varepsilon)}{2}q\cdot O^\natural q\Big) + U(q,C)\Big)\,dt.$$

A good choice for $f_M(\varepsilon)$ and $f_N(\varepsilon)$ is

$$f_M(\varepsilon) = f_N(\varepsilon) = \varepsilon.$$

Moreover since we can regard the term $q\cdot O^\natural q$ to be a *potential-like term* then we could make the reasonable choice

$$f_O(\varepsilon) \equiv 1.$$