THE PRINCIPLE OF COGNITIVE ACTION

THE CASE OF VISUAL FEATURES

Science is like sex: sometimes something useful
come out, but that is not the reason we are doing it

– Richard Feynman

Marco Gori
Dipartimento di Ingegneria dell'Informazione e Scienze Matematiche

Università di Siena

ACDL 2018

Acknowledgments

Stefano Melacci, Marco Maggini, Alessandro Betti, Giovanni Bellettini,
Paolo Nistri - UNISI

Alessandro Rossi, FBK, Trento

Marcello Sanguineti - University of Genoa
Giorgio Gnecco - IMT Lucca

Tomaso Poggio - MIT

Fabio Roli, University of Cagliari

Marcello Pelillo - Ca' Foscari (University of Venice)

Delfim Torres - University of Aveiro

Outline

Part I - Cognitive Action Laws

- Natural laws of learning
- Cognitive Action Laws of learning
- Causality and the control of information overloading
- Energy balance
- Links with stochastic gradient

Outline (con't)

Part II - A new framework for vision

- The role of time and motion invariance
- CAL of convolutional visual features (shallow model)
- Focus of attention as a necessary computational issue
- Discretization in the retina of "shallow models"
- Deep convolutional nets

Machine Learning

A CONSTRAINT-BASED APPROACH

MK
MORGAN KAUFMANN

Marco Gori

NATURAL LAWS OF LEARNING

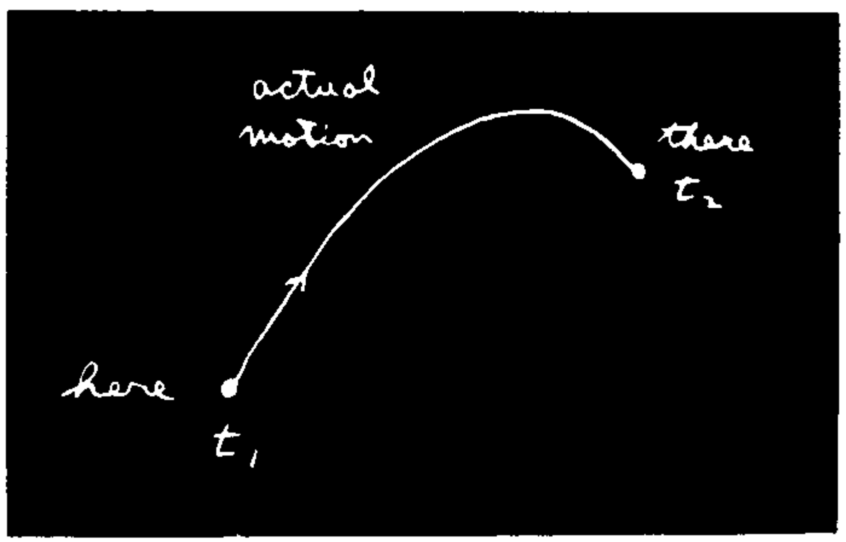
MOTIVATIONS
from perceptual tasks

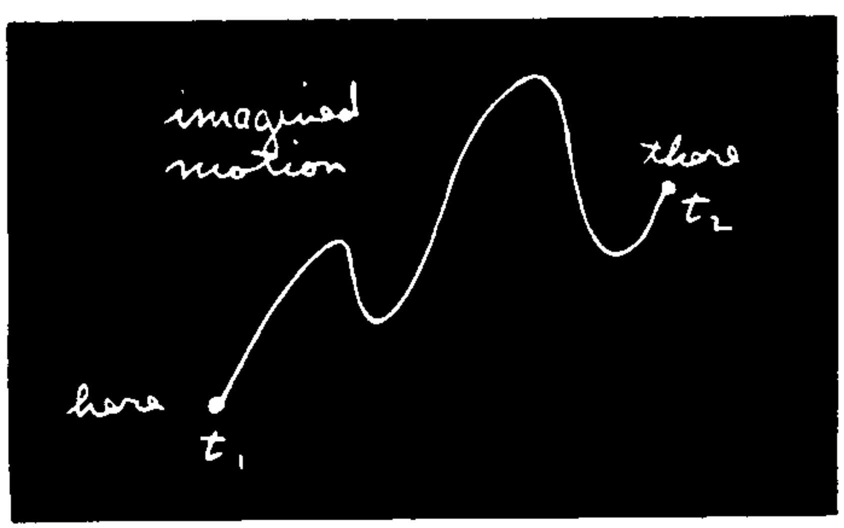
In computer vision we have been facing problems harder than those offered by Nature!

Perception in the framework of computational laws of nature

- Time isn't only the iteration index of on-line algorithms
- Perceptual information is indexed by time!
- The weights of a neural network can be thought of as Lagrangian coordinates
- Environmental interactions can be modeled by constraints on the tasks being developed

Least Action in Mechanics





$$S = \int \left[\frac{m}{2}\left(\frac{dx}{dt}\right)^2 - V(x)\right] dt$$

$$\int_{t_1}^{t_2} \left[\frac{1}{2}\, m \left(\frac{dx}{dt}\right)^2 - mgx\right] dt.$$

which path?

Stationarity of Action

variation

$$S = \int_{t_1}^{t_2} \left[\frac{m}{2} \left(\frac{d\underline{x}}{dt} + \frac{d\eta}{dt} \right)^2 - V(\underline{x} + \eta) \right] dt$$

$$\left(\frac{d\underline{x}}{dt}\right)^2 + 2\,\frac{d\underline{x}}{dt}\,\frac{d\eta}{dt} + \cancel{\left(\frac{d\eta}{dt}\right)^2}$$

$$S = \int_{t_1}^{t_2} \left[\frac{m}{2} \left(\frac{d\underline{x}}{dt} \right)^2 - V(\underline{x}) + m\,\frac{d\underline{x}}{dt}\,\frac{d\eta}{dt} - \eta V'(\underline{x}) + (\text{second and } \cancel{\text{higher}} \text{ order}) \right] dt$$

$$\delta S = m\,\cancel{\frac{d\underline{x}}{dt}\,\eta(t)}\,\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{d}{dt}\left(m\,\frac{d\underline{x}}{dt} \right) \eta(t)\,dt - \int_{t_1}^{t_2} V'(\underline{x})\,\eta(t)\,dt$$

known boundaries … or transversality conditions

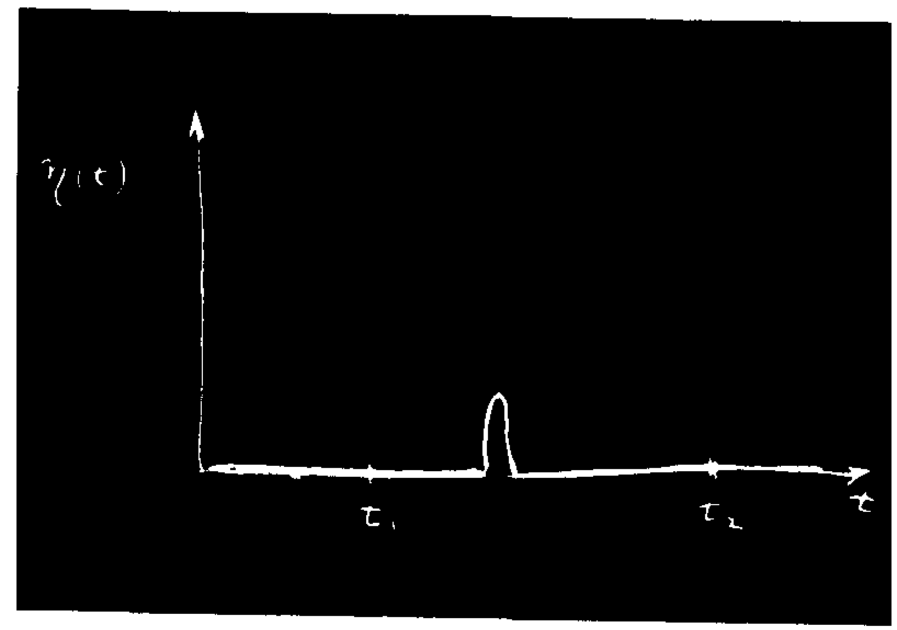
$$\delta S = \int_{t_1}^{t_2} \left[-m\,\frac{d^2\underline{x}}{dt^2} - V'(\underline{x}) \right] \eta(t)\,dt$$

Stationarity of Action (con't)

$$\delta S = \int_{t_1}^{t_2} \left[-m \frac{d^2 \underline{x}}{dt^2} - V'(\underline{x}) \right] \eta(t)\, dt.$$

$$\int F(t)\, \eta(t)\, dt = 0.$$

Fundamental Lemma of variational calculus



$$\left[-m \frac{d^2 \underline{x}}{dt^2} - V'(\underline{x}) \right] = 0$$

Euler-Lagrange equations

Kinetic Energy

how fast are you moving/learning?

from mechanics

$$K(\dot{w}) = \frac{1}{2}\sum_{i=1}^{m} m_i \dot{w}_i^2$$

how fast are weights moving?

A Kinetic energy

with high-order differential operators

$$T_i(t) = \sum_{j=0}^{\ell} \alpha_{i,j}(t) \frac{d^j}{dt^j}$$

$$K(t, Tw) = \frac{1}{2}\psi \sum_{i=1}^{n} m_i (T_i w_i)^2$$

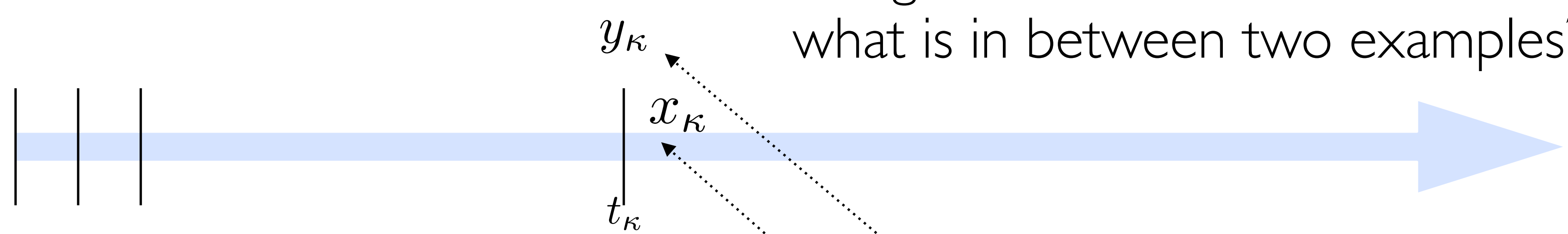
dissipation function: why do we need it?

order higher than two will be proven to be of crucial importance for "cognitive laws"!

Potential energy

as a loss function

a degenerate form of time:
what is in between two examples?



supervised learning

$$\int_{t_0}^{t_1} V(t, w(t)) \quad = \int_{t_0}^{t_1} \sum_{\kappa=1}^{\ell} (f(w(t), x(t)) - y_\kappa)\delta(t - t_\kappa)$$

$$E(w) = \sum_{\kappa=1}^{\ell} (y_\kappa - \mathsf{f}(w, x_\kappa))^2$$

$$x_\kappa = x(t_\kappa)$$

Potential energy
as a loss function (con't)

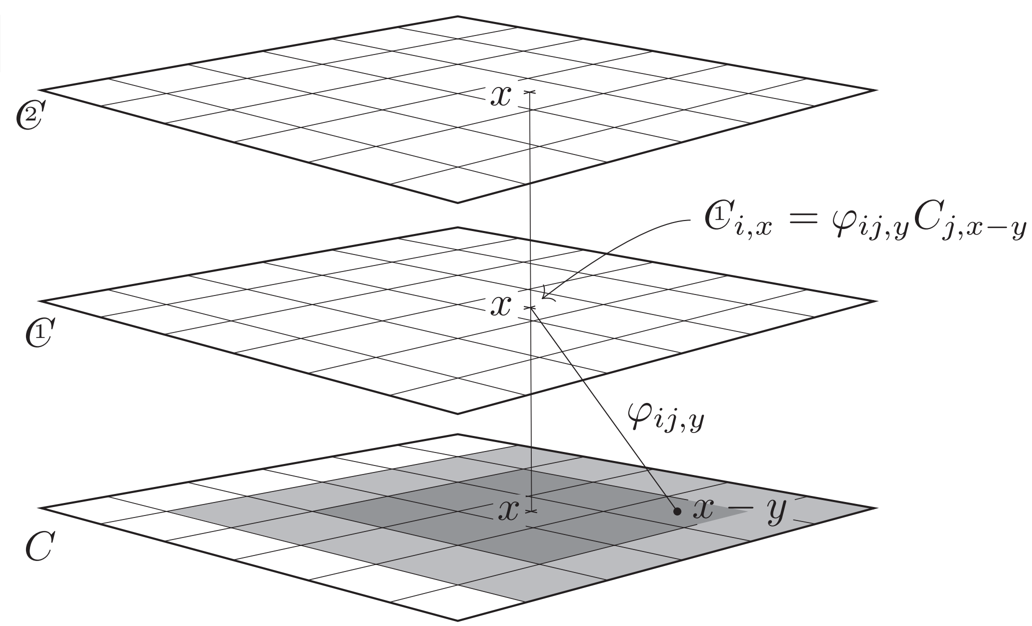
visual environment

$D = X \times [0\,..\,T]$

retina

frame

unsupervised learning



$$\mathbb{C}^{1}_i(x,t) = \frac{1}{n} + \sum_{j=0}^{m-1} \int_X dy\; \varphi_{ij}(x,y,t) C_j(y,t) = \frac{1}{n} + (\varphi_t \times C_t)_i(x)$$

time has a physical meaning here!

Potential energy

as a loss function (con't)

$$V(t, w(t)) = U(w, u)$$

unsupervised learning

$$D = X \times [0 \,..\, T]$$

convolutional filter weights

$$S(Y \mid X, T, F) = -\int_D d\mu(x,t) \; \sum_{i=1}^{n} \mathbb{C}_i(x,t) \log \mathbb{C}_i(x,t)$$

$$S(Y) = -\sum_{i=1}^{n} \Big(\int_D d\mu(x,t)\, \mathbb{C}_i(x,t) \Big) \log \Big(\int_D d\mu(x,t)\, \mathbb{C}_i(x,t) \Big)$$

$$I(Y; X, T, F) = S(Y) - S(Y \mid X, T, F).$$

Cognitive Action

special case $+$ $\gamma V(t, w)$

$$F_m(t, w, \dot{w}) = \sum_{i=1}^{m} F(t, w_i, \dot{w}_i) = \sum_{i=1}^{m} K(t, \dot{w}_i) - V(t, w)$$

$$F(t, w, \dot{w}) = \psi(t) F_m(t, w, \dot{w})$$

dissipation function $\psi(t) = e^{\theta t}$

$$A(w) = \int_{t_0}^{t_1} F(t, w, Tw)\, dt$$

An action for visual features

$$A(w) = \int_0^T e^{\theta t}\left(\frac{\mu}{2}|\ddot w|^2 + \frac{\nu}{2}|\dot w|^2 + \gamma \dot w \ddot w + \frac{k}{2}|w|^2 + U(w,u)\right)dt$$

Theorem (Betti et al, arxiv 2018)

coercivity conditions

$\mu > \gamma_2^2, \quad \nu > \gamma_1^2, \quad k > 0$

$\gamma = \gamma_1 \gamma_2$

minimum under initial conditions $w(0) = w^0, \ \ \dot w = w^1$

intuition …

Natural laws of cognition

Natural Learning Theory $\rightsquigarrow$ Mechanics	Remarks
$w_i \rightsquigarrow q_i$	Weights are interpreted as generalized coordinates.
$\dot{w}_i \rightsquigarrow \dot{q}_i$	Weights variations are interpreted as generalized velocities.
$\upsilon_i \rightsquigarrow p_i$	The conjugate momentum to the weights is defined by using the machinery of Legendre transforms.
$A(w) \rightsquigarrow S(q)$	The cognitive action is the dual of the action in mechanics.
$F(t, w, \dot{w}) \rightsquigarrow L(t, q, \dot{q})$	The Lagrangian F is associated with the classic Lagrangian L in mechanics.
$H(t, w, \upsilon) \rightsquigarrow H(t, q, p)$	When using w and υ, we can define the Hamiltonian, just like in mechanics.

COGNITIVE ACTION LAWS

time is not
the iteration index of traditional
machine learning algorithms

… e.g. recurrent nets

Euler-Lagrange equations

$$\xi^{(\kappa)}(t_0) = \xi^{(\kappa)}(t_1) = 0, \qquad \kappa = 0, \ldots, \ell - 1.$$

conditions on the boundaries on $w^{(\kappa)}$

Then we have

$$\begin{aligned}
\delta A &= A[\check{w}] - A[w] \\
&= \int_{t_0}^{t_1} [F(t, w + \epsilon\xi, Tw + \epsilon T\xi) - F(t, w, Tw)]\, dt \\
&= \epsilon \cdot \int_{t_0}^{t_1} [F_w \cdot \xi + F_{Tw} \cdot T\xi]\, dt + \mathcal{O}(\epsilon^2).
\end{aligned}$$

$$\int_{t_0}^{t_1} F_{Tw} \cdot T\xi\, dt = \int_{t_0}^{t_1} (T^\star F_{Tw}) \cdot \xi\, dt. \qquad \delta A = \epsilon \cdot \int_{t_0}^{t_1} [F_w + T^\star F_{Tw}] \cdot \xi\, dt$$

$$F_w + T^\star F_{Tw} = 0$$

Cognitive Action Laws as Euler-Lagrange equations

Cauchy and boundary conditions: causality issues

$$w^{(\kappa)}(t_0) \text{ and } w^{(\kappa)}(t_1) \text{ with } \kappa = 0, \ldots, \ell - 1.$$

$$F_{w_i} + T_i^{\star} F_{T_i w_i} = 0 \qquad i = 1, \ldots, n.$$

$$T_i^{\star}(\psi T_i w_i) + \gamma V_{w_i} = 0 \qquad i = 1, \ldots, n.$$

REDUCTION TO MECHANICS

Let us assume that $\forall i = 1, \ldots, n: \quad T_i = T = D$ be, and let $\gamma = -1$ be. Then we consider the Lagrangian

$$F(t, w, Tw) = \frac{1}{2}\psi \sum_{i=1}^{n} m_i (Dw_i)^2 - \psi V(w)$$

Betti & Gori, TCS-2016

where $\psi = e^{\theta t}$. Then the Euler–Lagrange equations (see Theorem 3.1) become

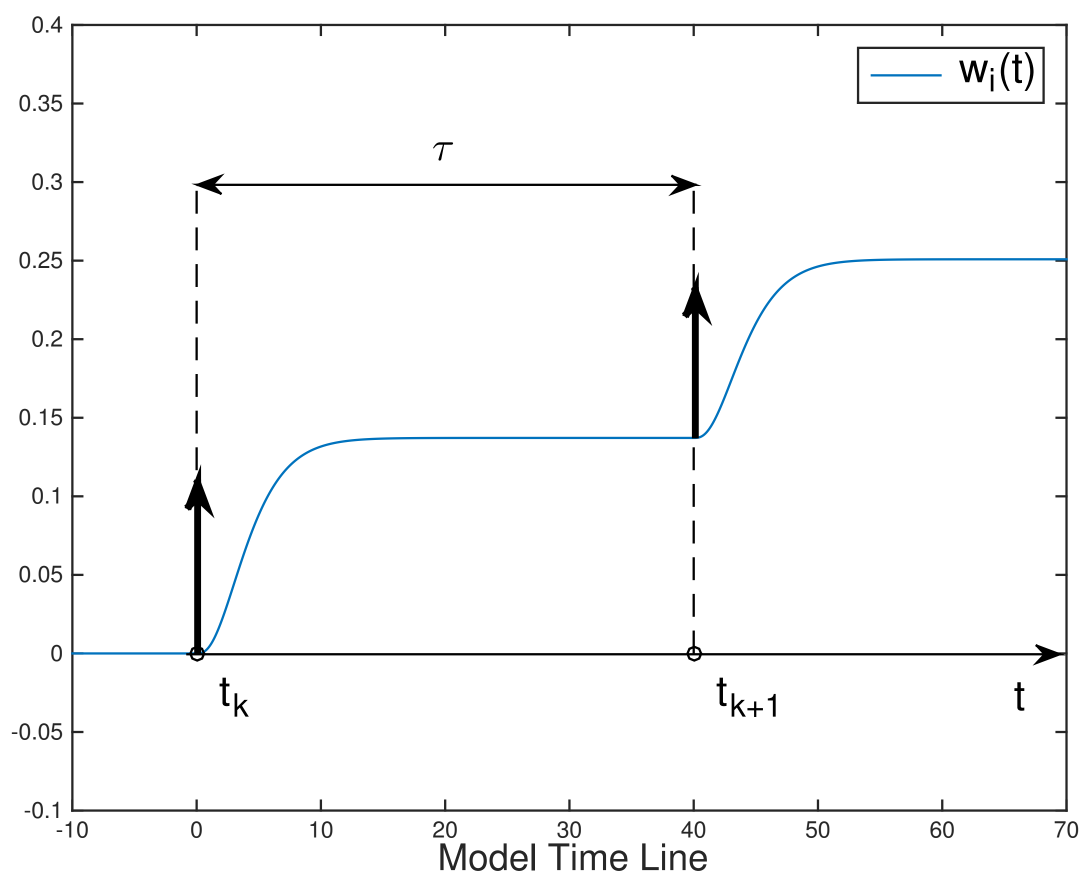
$$D^2 w_i + \theta D w_i + V_{w_i} = 0 \qquad i = 1, \ldots, n,$$

strong dissipation corresponds with gradient descent!

Supervised Learning

$$\psi(t) = e^{\theta t}; \quad \frac{\dot{\psi}}{\psi} = \theta$$

$$t \in (t_\kappa, t_{\kappa+1}) : \quad m_i \ddot{w}_i(t) + \theta \dot{w}_i(t) + \sum_{i<\kappa} V_{w_i}(t_\kappa)\delta(t - t_i) = 0$$



strong dissipation

Consistent boundary condition

$\dot{w}(t_1) = 0$

Fourth-order Euler-Lagrange equations

Interpretation as a minimum
Consistent boundary condition under stability (sign flip issue)

Let us consider the case $T = \alpha_o + \alpha_1 D + \alpha_2 D^2$ and $\psi(t) = e^{\theta t}$.

From Proposition 3.1, we have $T^\star = \alpha_o - \alpha_1 D + \alpha_2 D^2$ and we can easily see that the EL-equations becomes

$$D^4 w_i + \beta_3 D^3 w_i + \beta_2 D^2 w_i + \beta_1 D w_i + \beta_o w_i + \frac{\gamma}{\alpha_2^2 m_i} V_{w_i} = 0$$

stability and "sign flip"

where

$$\beta_o := \frac{\alpha_o \alpha_2 \theta^2 - \alpha_o \alpha_1 \theta + \alpha_o^2}{\alpha_2^2}$$

$$\beta_1 := \frac{\alpha_1 \alpha_2 \theta^2 + (2\alpha_o \alpha_2 - \alpha_1^2)\theta}{\alpha_2^2}$$

$$\beta_2 := \frac{\alpha_2^2 \theta^2 + \alpha_1 \alpha_2 \theta + 2\alpha_o \alpha_2 - \alpha_1^2}{\alpha_2^2}$$

$$\beta_3 := 2\theta.$$

CAUSALITY AND CONTROL OF INFORMATION OVERLOADING

ACDL 2018

Causality Issues

think of

- Backpropagation
- Bellman's dynamic programming
- Optimal control
- … physical problems typically involves boundary conditions

An action for visual features

$$A(w) = \int_0^T e^{\theta t}\left(\frac{\mu}{2}|\ddot{w}|^2 + \frac{\nu}{2}|\dot{w}|^2 + \gamma\dot{w}\ddot{w} + \frac{k}{2}|w|^2 + U(w,u)\right)dt$$

Euler-Lagrange Equations

$$\mu w^{(4)} + 2\theta\mu w^{(3)} + (\theta^2\mu + \theta\gamma - \nu)w^{(2)} + (\theta^2\gamma - \theta\nu)w^{(1)} + kw + \nabla_w U(w,u) = 0$$

$$w(0) = w^0, \quad \dot{w} = w^1, \quad \ddot{w}(0) = w^2, \quad w^{(3)}(0) = w^3$$

initial conditions

boundary conditions

$$\hat{\mu}\ddot{w}(T) + \hat{\gamma}\dot{w}(T) = 0$$

$$\hat{\mu}w^{(3)}(T) + \dot{\hat{\mu}}\ddot{q}(T) + (\dot{\hat{\gamma}} - \hat{\nu})\dot{w}(T) = 0$$

consistency needed

ACDL 2018

Controlling information overloading

… during learning

- Filtering the input - blurring in newborns, it’s not only a matter of choosing the learning rate properly! What if humans are given a video with 300 frames/sec?

- Reset of system dynamics - saccadic movements

  they correspond with the presence of null input

$$w(0), \quad \dot{w}(0) = 0$$

$$w^{(1)}(T) = 0, \; w^{(2)}(T) = 0, \; w^{(3)}(T) = 0$$

ENERGY BALANCE

Resembling mechanics …

$$K(\dot{w}) = \frac{1}{2}\sum_{i=1}^{m} m_i \dot{w}_i^2$$

$$F_m(t, w, \dot{w}) = \sum_{i=1}^{m} K(t, \dot{w}_i) - V(t, w)$$

$$F(t, w, \dot{w}) := \psi(t)\, F_m(t, w, \dot{w}) \qquad A(w) = \frac{1}{T}\int_0^T F(t, w, \dot{w})\, dt$$

$$m_i \ddot{w}_i + \frac{\dot{\psi}}{\psi}\dot{w}_i + V_{w_i} = 0$$

$$\lim_{t \to 0^+} \dot{w}_i(t) = 0, \qquad \lim_{t \to T^-} \dot{w}_i(t) = 0$$ looking for constant solutions

Energy balance

$$m_i\ddot{w}_i + \frac{\dot\psi}{\psi}\dot{w}_i + V_{w_i} = 0 \qquad \psi(t) = e^{\theta t}; \quad \frac{\dot\psi}{\psi} = \theta$$

$$\sum_{i=1}^m m_i\ddot{w}_i\dot{w}_i + \frac{\dot\psi}{\psi}\sum_{i=1}^m \dot{w}_i^2 + \sum_{i=1}^m V'_{w_i}\dot{w}_i = 0.$$

$$DK(\dot w(t)) = D\Big(1/2\textstyle\sum_{i=1}^m m_i\dot{w}_i^2\Big) = \textstyle\sum_{i=1}^m m_i\ddot{w}_i\dot{w}_i$$

$$DV(t,w(t)) - \partial_t V(t,w(t)) = \sum_{i=1}^m V'_{w_i}\dot{w}_i$$

$$\int_0^T DK + \sum_{i=1}^m \int_0^T \frac{\dot\psi}{\psi}\dot{w}_i^2 dt + \int_0^T DV - \int_0^T \partial_t V(t,w(t))dt = 0$$

Energy balance (con't)

$$\int_0^T DK + \sum_{i=1}^m \int_0^T \frac{\dot\psi}{\psi}\dot w_i^2 dt + \int_0^T DV - \int_0^T \partial_t V(t, w(t))dt = 0$$

$$U(t) := V(t, w(t)) + K(\dot w(t))$$ internal energy

$$Z(T) := \sum_{i=1}^m \int_0^T \frac{\dot\psi}{\psi}\dot w_i^2 dt$$ dissipated energy

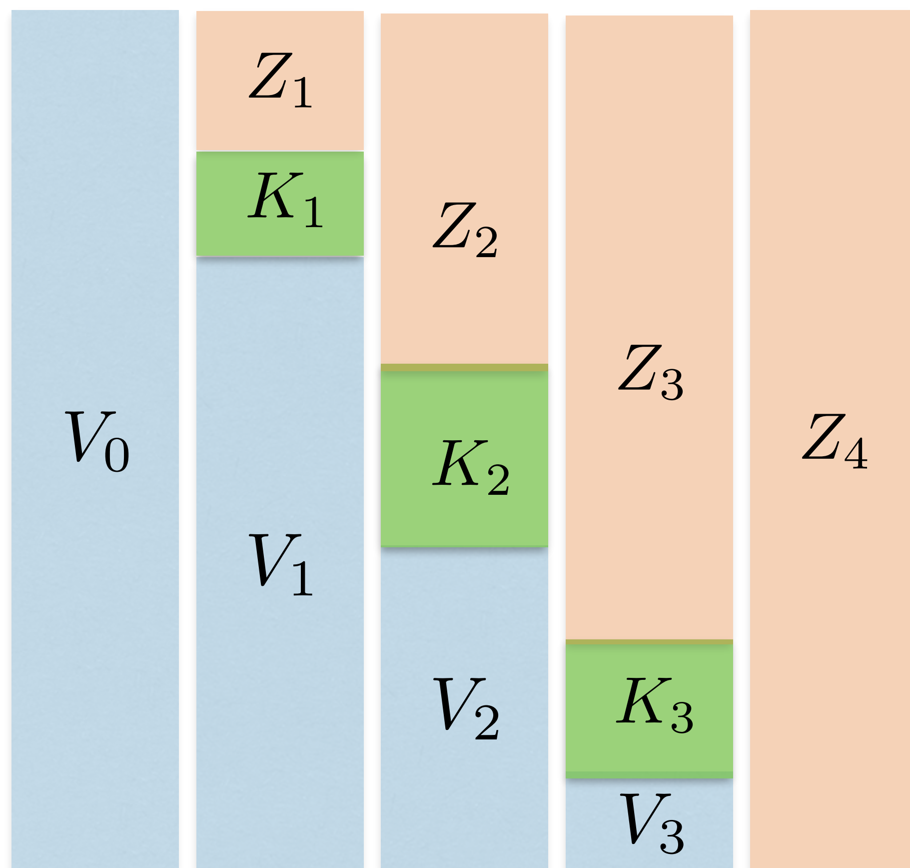
$$E(T) = \int_0^T \partial_t V(t, w(t))dt$$ environmental energy

$$\boxed{Z + \Delta U - E = 0}$$

General Hamiltonian analysis in
Betti & Gori, TCS 2016

A Qualitative Picture

$$\mathrm{V}(0) = \mathrm{V}(\mathrm{T}) + \mathrm{Z}(\mathrm{T})$$



this is why we need an increasing monotone

$$\psi(t) = e^{\theta t}; \quad \frac{\dot{\psi}}{\psi} = \theta$$

Learning yields ordered configurations which require dissipation of energy!

LINKS WITH STOCHASTIC GRADIENT

Back to Gradient Descent:
Strong Dissipation

$$m_i \ddot{w}_i + \frac{\dot{\rho}}{\rho}\dot{w}_i + V'_{w_i} = \cancel{m_i \ddot{w}_i} + \theta \dot{w}_i + V'_{w_i} = 0$$

strong dissipation

$$\dot{w}_i = -\frac{m_i}{\theta} V'_{w_i}$$

$$\frac{w_i(\kappa+1) - w_i(\kappa)}{\tau} = -\frac{m_i}{\theta} V'_{w_i}$$

$$w_i(\kappa+1) = w_i(\kappa) - \frac{m_i \tau}{\theta} V'_{w_i}$$

Back to Gradient Descent:

Strong Dissipation

$$\mu w^{(4)} + 2\theta\mu w^{(3)} + (\theta^2\mu + \theta\gamma - \nu)w^{(2)} + (\theta^2\gamma - \theta\nu)w^{(1)} + kw + \nabla_w U(w, u) = 0$$

$$\mu = \nu \equiv 0 \text{ and } \gamma = 1/\theta^2$$

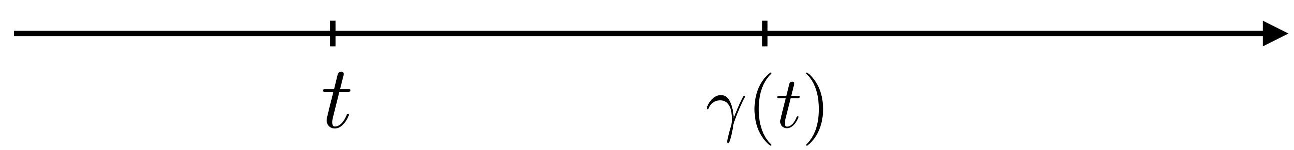
$$\xcancel{\theta^{-1} w^{(2)}} + w^{(1)} + kw + \nabla_w U = 0$$

$$\theta \to \infty$$

Convergence analysis
based on energy balance

Definition 1 *An environment is* quasi-periodic *in* $[0,T]$ *of order* $p \in \mathbb{R}$ *if there exist* $\alpha > 0$, $\epsilon > 0$ *and a positive differentiable function* $\tau : [0,T] \to (0,\infty]$ *with* $\tau(t) > 0$ *such that* $\gamma(t) := t + \tau(t)$, *satisfies* $\gamma'(t) > 1$ *and*

$$\forall t \in [0,T] \quad \textit{we have} \quad \|x(t) - x(\gamma(t))\| \leq \frac{\epsilon}{(\alpha+t)^p}. \tag{11}$$



$$\lim_{t\to\infty} w(t) = \overline{w}$$

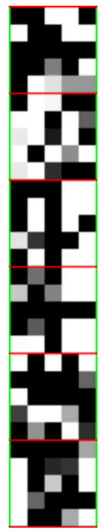
G. Bellettini et al, Generalization in quasi-periodic environments - arXiv, 2018

NEXT:

NEW CONVOLUTIONAL NETS


Video


Motion


Filters


Feature 1

{
  "status": "day",
  "rho": 0.0002999700082000345,
  "action_cur": 832.4277954101562,
  "mi_real_full": 0.00017303228378295898,
  "motion_full": 0,
  "mi_real": 0.00017303228378295898,
  "mi": 0.00010585784912109375,
  "ce": 1.0000181198120117,
  "minus_ge": -0.0001239776611328125,
  "motion": 0,
  "norm_q": 52.456607818603516,
  "norm_q_mixed": 1.5736880687455823e-9,
  "norm_q_dot": 5.2456615440288346e-11,
  "norm_q_dot_dot": 4.7210324538582427e-8,
  "norm_q_dot_dot_dot": 0.0000047209064177
  "eps1": 960,
  "eps2": 960,
  "eps3": 960
}


Feature 2


Feature 3


Feature 4


Feature 5


Feature 6