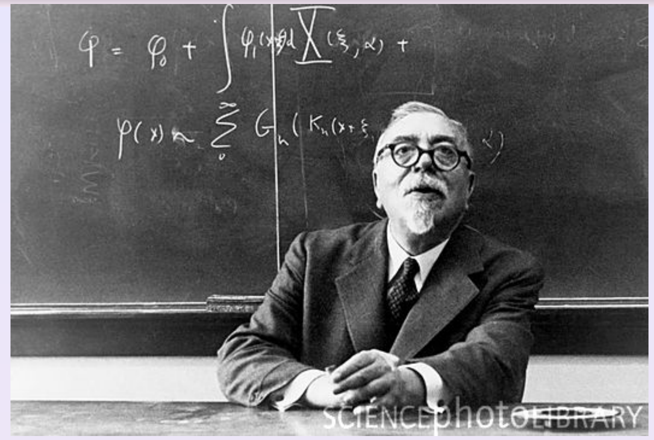
Turpike Control and Machine Learning

Enrique Zuazua

Chair of Dynamics, Control and Numerics
Department of Data Science
FAU
Erlangen, Germany
https://caa-avh.nat.fau.eu

Outline

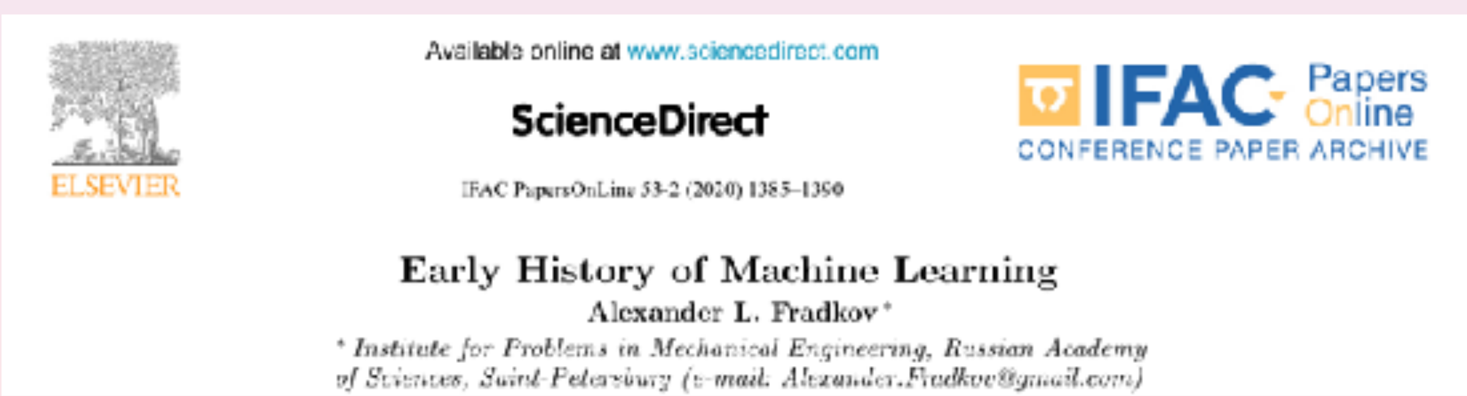
Historical perspective: Cybernetics



"Cybernétique" was proposed by the French physicist **A.-M. Ampère** in the XIX Century to design the nonexistent science of process controlling. This was quickly forgotten until 1948, when **Norbert Wiener** (1894–1964) chose **"Cybernetics"** as the title of his famous book.

Wiener defined Cybernetics as " the science of control and communication in animals and machines".[1]

In this way, he established the connection between Control Theory and Physiology and anticipated that, in a desirable future, engines would obey and imitate human beings.


Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 53-2 (2020) 1385–1390

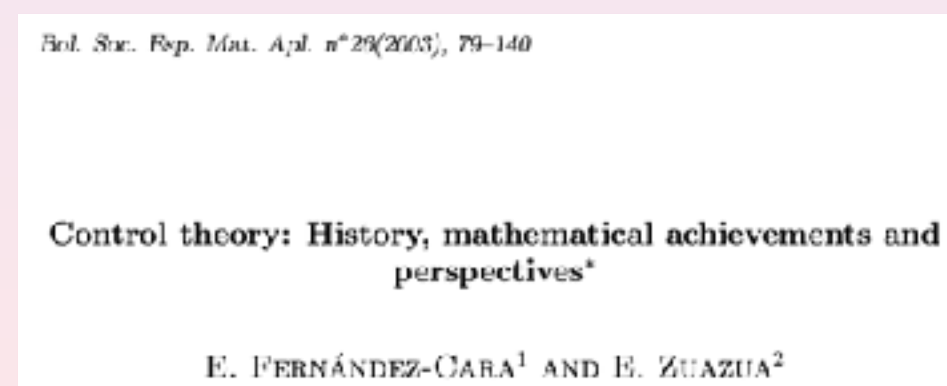
ELSEVIER

IFAC Papers Online CONFERENCE PAPER ARCHIVE

Early History of Machine Learning

Alexander L. Fradkov*

* Institute for Problems in Mechanical Engineering, Russian Academy of Sciences, Saint-Petersburg (e-mail: Alexander.Fradkov@gmail.com)


Bol. Soc. Esp. Mat. Apl. nº29(2003), 79–140

Control theory: History, mathematical achievements and perspectives*

E. Fernández-Cara[1] and E. Zuazua[2]

[1] "What we want is a machine that can learn from experience." Alan Turing, 1947.

Controllability

Let $n, m \in \mathbb{N}^*$ and $T > 0$ and consider the following linear finite-dimensional system

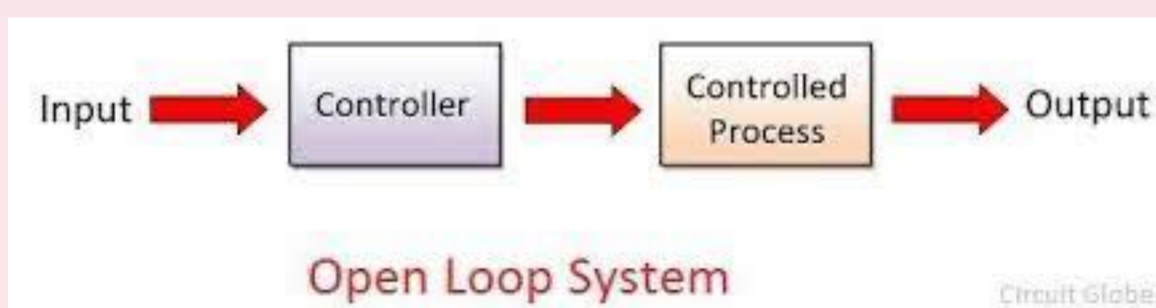
$$x'(t) = Ax(t) + Bu(t), \quad t \in (0, T); \quad x(0) = x^0. \tag{1}$$

In (1), A is a $n \times n$ real matrix, B is of dimensions $n \times m$ and x^0 is the initial sate of the sytem in $\mathbb{R}^n$. The function $x : [0, T] \longrightarrow \mathbb{R}^n$ represents the *state* and $u : [0, T] \longrightarrow \mathbb{R}^m$ the *control*.
¿Can we control the state x of n components with only m controls, even if $n >> m$ so that, for instance $x(T) = 0$?

Theorem

(1958, Rudolf Emil Kálmán (1930 – 2016)) System (1) is controllable iff

$$rank\,[B, AB, \cdots, A^{n-1}B] = n.$$



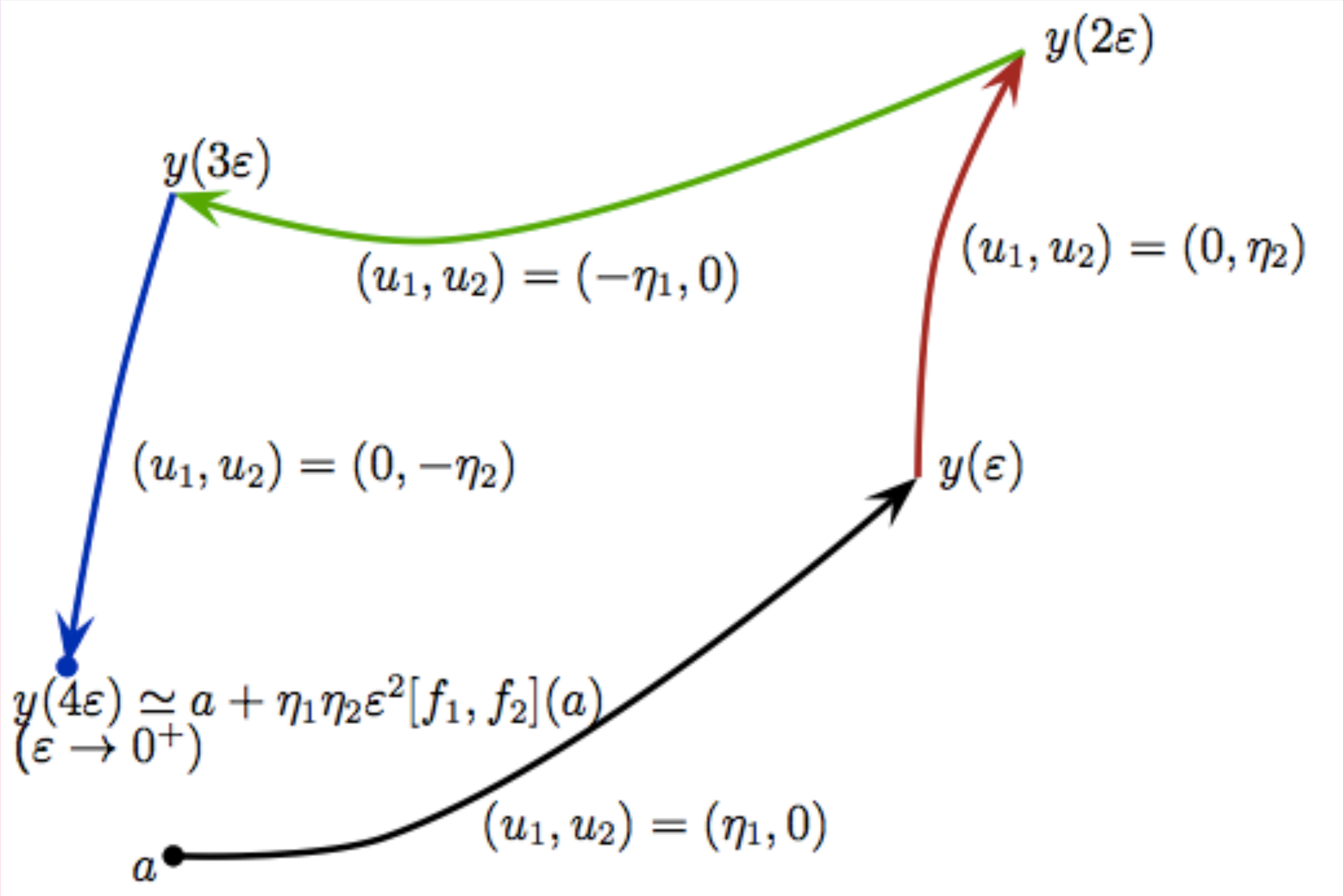
Sketch of the proof

From the variation of constants formula:

$$x(t) = e^{At}x^0 + \int_0^t e^{A(t-s)}Bu(s)ds = e^{At}x^0 + \int_0^t \sum_{k\geq 0} \frac{(t-s)^k}{k!} A^k Bu(s)ds.$$

By Cayley[2]-Hamilton's [3] theorem A^k, for $k \geq n$, is a linear combination of $I, A, ..., A^{n-1}$.

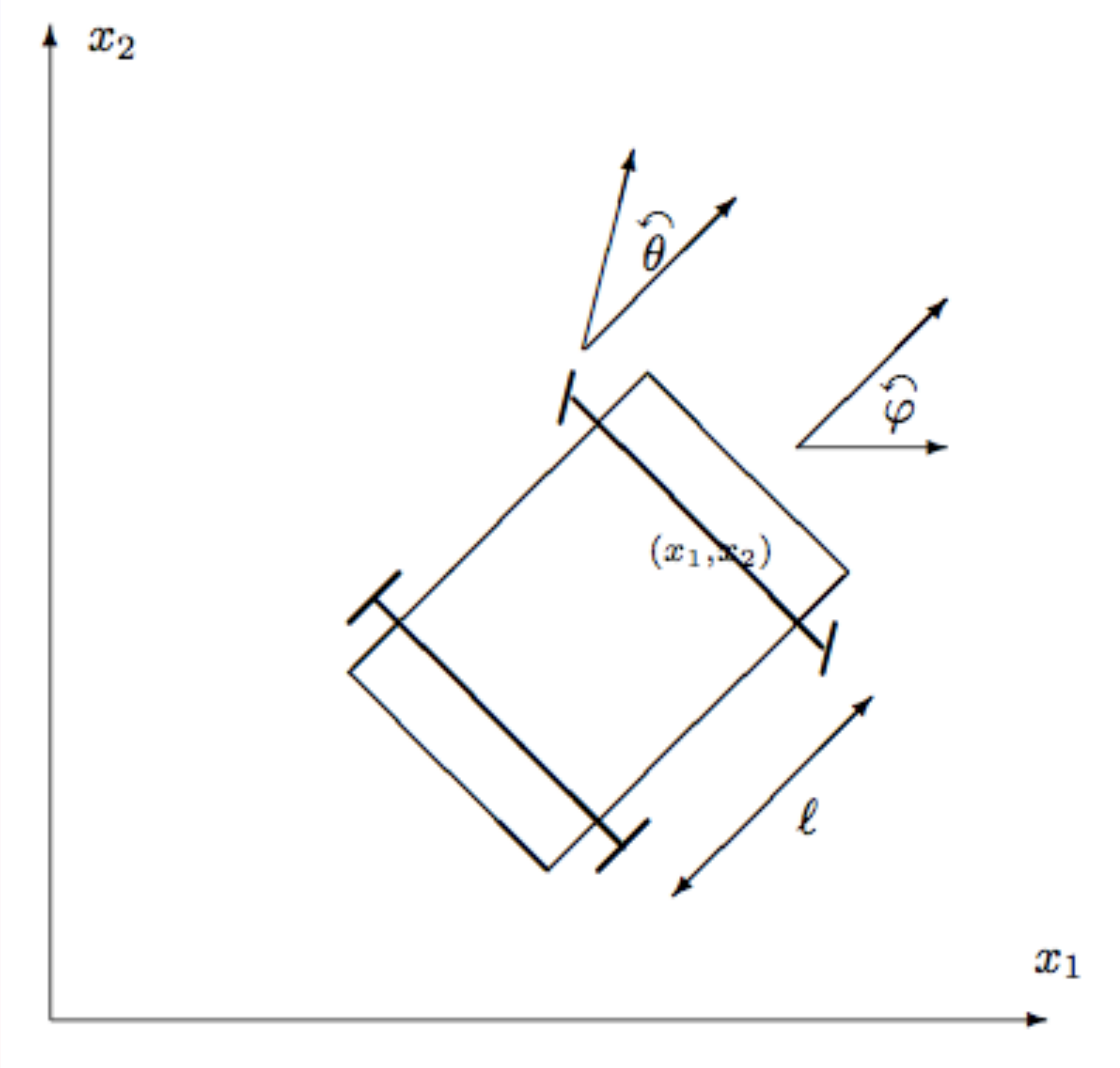
The Kalman rank condition assures that, manipulating the last term due to the variation of constants formula, out of a strategic choice of the control $u(t)$, the solution can be driven to any destination $x(T)$.



[2] Arthur Cayley (UK, 1821 - 1895)

[3] William Rowan Hamilton (Ireland, 1805 - 1865)

An example: Nelson's car.



Two controls suffice to control a four-dimensional dynamical system.

E. Sontag, *Mathematical control theory*, 2nd ed.,Texts in Applied Mathematics,vol.6, Springer-Verlag, NewYork,1998.

Duality, J.-L. Lions, SIREV, 1988

Consider the adjoint system

$$\begin{cases} -p' = A^* p, & t \in (0, T) \\ p(T) = p_T \end{cases}$$

and minimize

$$J(p_T) = \frac{1}{2} \int_0^T \mid B^* p \mid^2 dt + \langle x^0, p(0) \rangle$$

Then

$$u = B^* \widehat{p}$$

is the control of minimal L^2-norm.[4]
And the functional J is coercive iff the Kalman rank condition is satisfied.

The Kalman condition is equivalent to the Unique Continuation property

$$B^* p \equiv 0 \Rightarrow p^T \equiv 0.$$

The observability inequality plays a key role

$$||p^T||^2 \leq C \int_0^T \mid B^* \varphi \mid^2 dt$$

[4] This confirms Wiener's vision "..control and communication..."

Outline

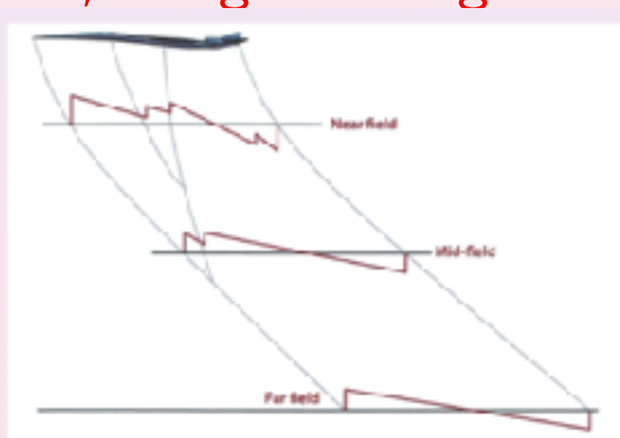
Sonic boom

Francisco Palacios, Boeing, Long Beach, California, Project Manager and Aerodynamics Engineer

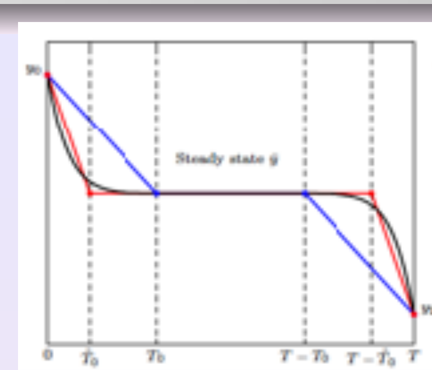
- Goal: the development of supersonic aircrafts, sufficiently quiet to be allowed to fly supersonically over land.
- The pressure signature created by the aircraft must be such that, when reaching ground, (a) it can barely be perceived by humans, and (b) it results in admissible disturbances to man-made structures.
- This leads to an inverse design or control problem in long time horizons.

Juan J. Alonso and Michael R. Colonno, Multidisciplinary Optimization with Applications to Sonic-Boom Minimization, Annu. Rev. Fluid Mech. 2012, 44:505 – 526.

Many other challenging problems of high societal impact raise similar issues: climate change, sustainable growth, chronically deseases, design of long lasting devices and infrastructures...



Origins



Although the idea goes back to John von Neumann in 1945, Lionel W. McKenzie traces the term to Robert Dorfman, Paul Samuelson, and Robert Solow's "Linear Programming and Economics Analysis" in 1958, referring to an American English word for a Highway:[5]
[6]

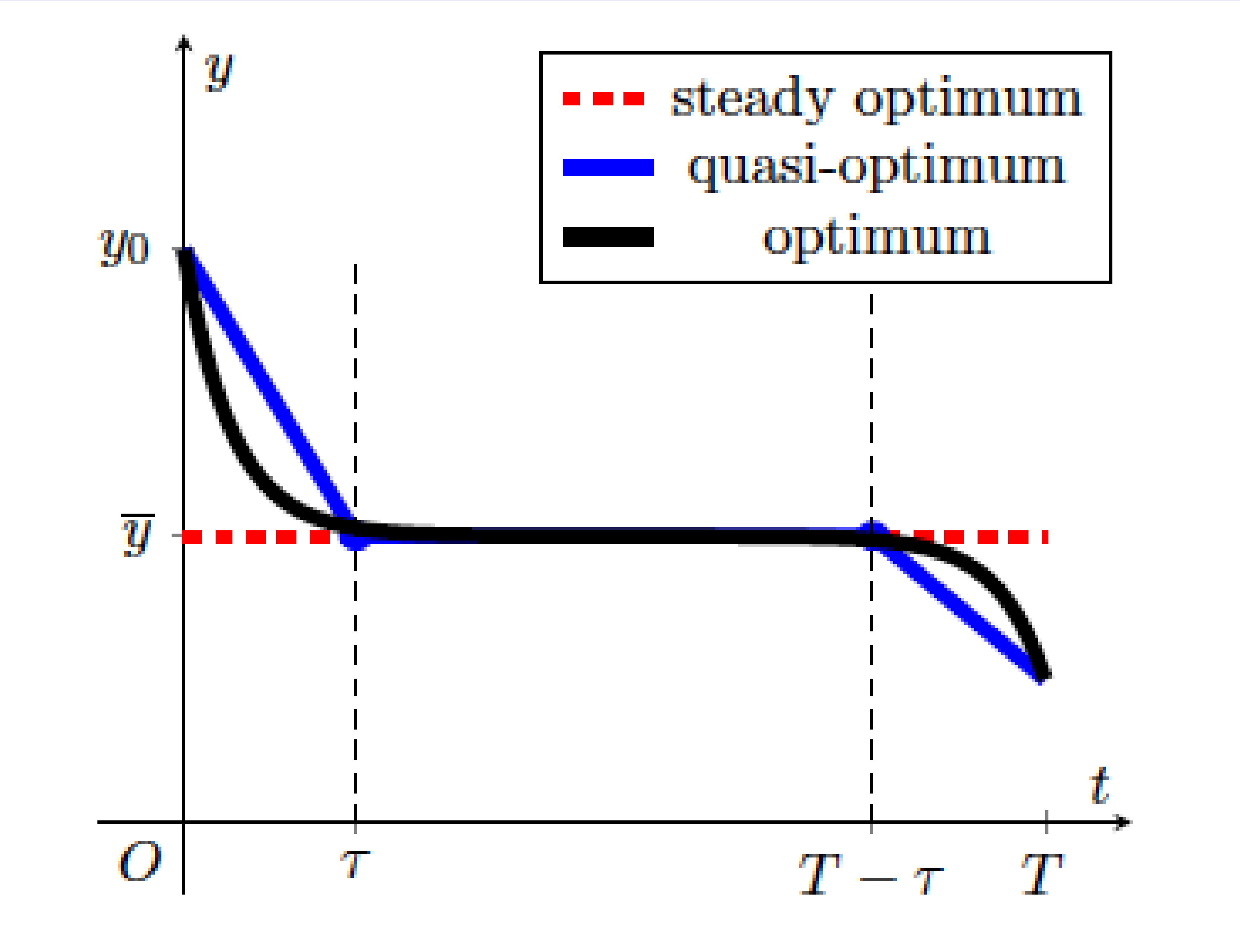
> *... There is a fastest route between any two points; and if the origin and destination are close together and far from the turnpike, the best route may not touch the turnpike. But if the origin and destination are far enough apart, it will always pay to get on to the turnpike and cover distance at the best rate of travel, even if this means adding a little mileage at either end.*

[5] A. J. Zaslavski, Springer, New York, 2006.
[6] L. Grüne, Automatica, 49, 725-734, 2013

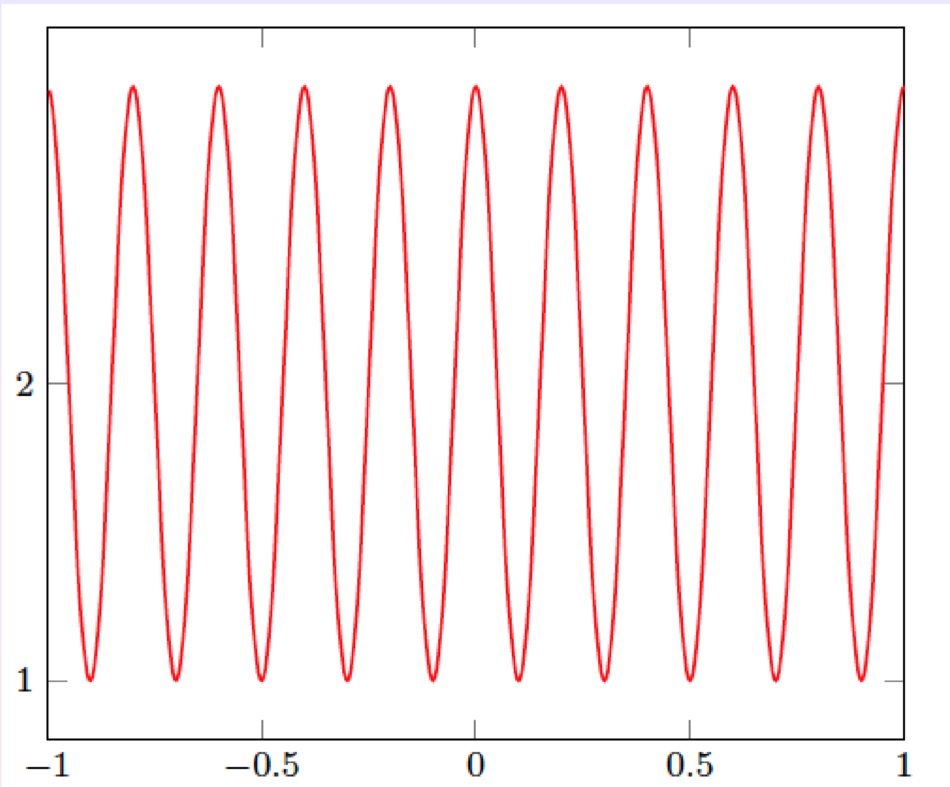
Substantiation and preliminary conclusion

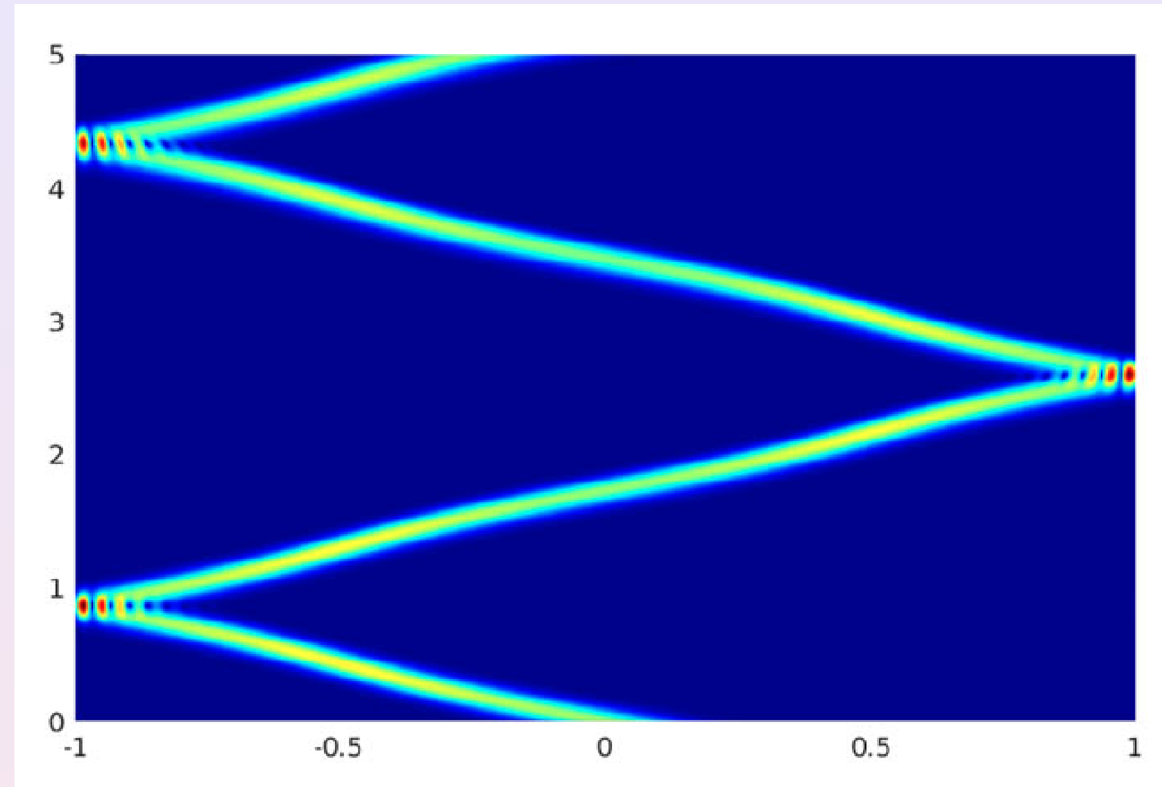
We implement turnpike (or nearby) strategies most often. And it is indeed a good idea to do it!

But this requires that the system under consideration to be controllable/stabilisable.



Wave propagation: Why do not we see the turnpike?



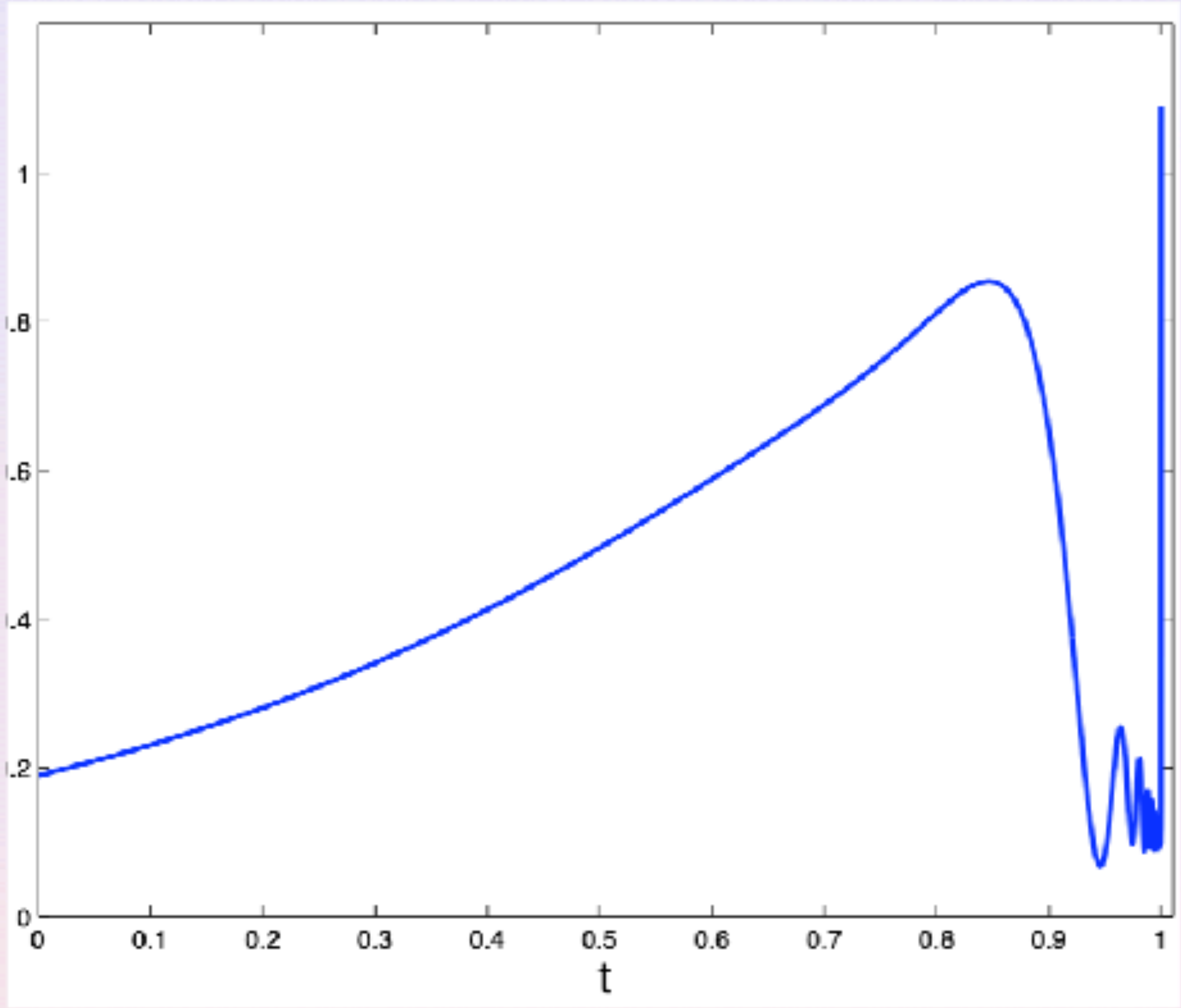


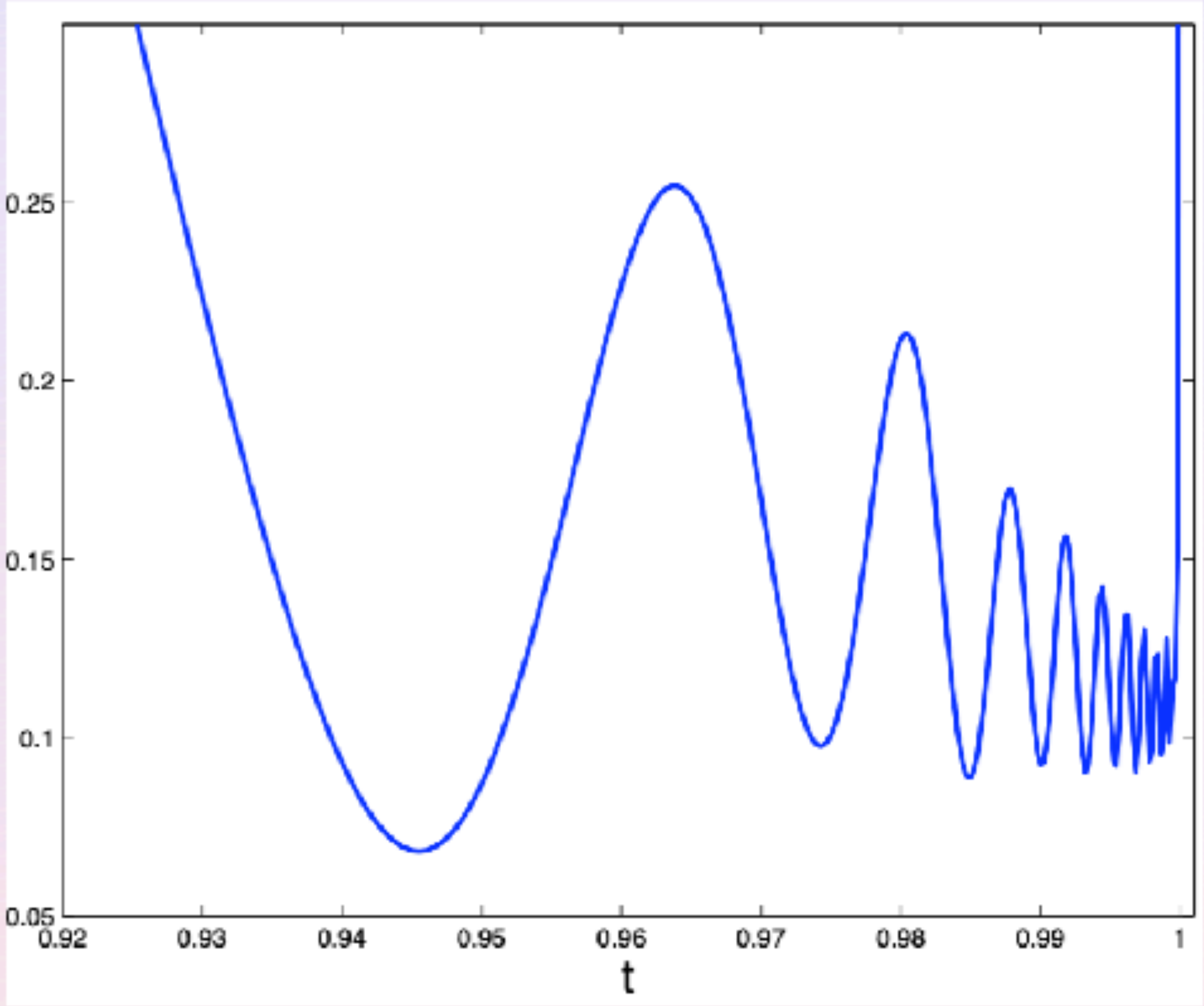
- Typical controls for the wave equation exhibit an oscillatory behaviour, and this independently of the length of the control time-horizon.
- This fact is intrinsically linked to the oscillatory nature of wave propagation.
- Waves are controlled through anti-waves, reproducing an oscillatory pattern.

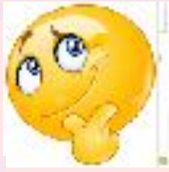
Heat and diffusion processes: Why do not we see the turnpike either?

Typical controls for the heat equation exhibit **unexpected** oscillatory and concentration effects. This was observed by R. Glowinski and J. L. Lions in the 80's in their works in the numerical analysis of controllability problems for heat and wave equations.

Why? Lazy controls?







Optimal controls are boundary traces of solutions of the **adjoint problem** through the optimality system or the **Pontryagin Maximum Principle**, and solutions of the adjoint heat equation

$$-p_t - \Delta p = 0$$

look precisely this way.

Large and oscillatory near $t = T$ they decay and get smoother when t gets down to $t = 0$. And this is independent of the time control horizon $[0, T]$.

For wave-like equations ontrols are given by the solutions of the adjoint system

$$p_{tt} - \Delta p = 0$$

that exhibit endless oscillations.

First conclusion:
Typical control problems for wave and heat equations do not seem to exhibit the turnpike property.
These are the controls of L^2-minimal norm. There are many other possibilities for successful control strategies (sparse controls by L^1-minimisation, bang-bang controls...)

May be the Turnpike Principle does not hold for for Partial Differential Equations (PDE), i. e. Infinite-Dimensional Dynamical Systems?

The control problem for diffusion : A closer look

Let $n \geq 1$ and $T > 0$, Ω be a simply connected, bounded domain of $\mathbb{R}^n$ with smooth boundary Γ, $Q = (0, T) \times \Omega$ and $\Sigma = (0, T) \times \Gamma$:

$$\begin{cases} y_t - \Delta y = f 1_\omega & \text{in} \quad Q \\ y = 0 & \text{on} \quad \Sigma \\ y(x,0) = y^0(x) & \text{in} \quad \Omega. \end{cases} \tag{3}$$

1_ω = the characteristic function of ω of Ω where the control is active.

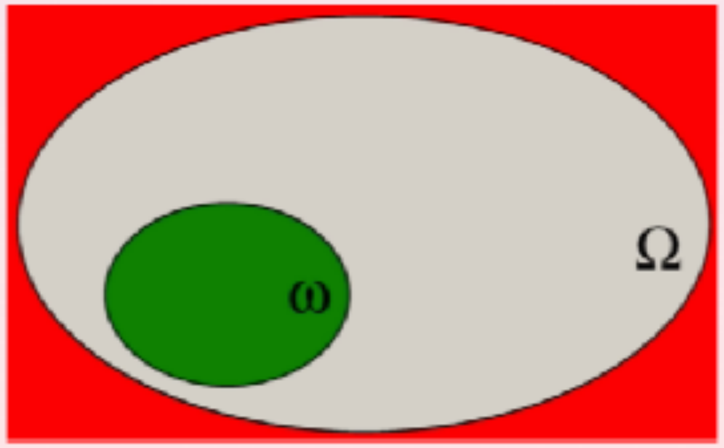
We know that $y^0 \in L^2(\Omega)$ and $f \in L^2(Q)$ so that (3) admits a unique solution

$$y \in C\left([0, T]\, ; L^2(\Omega)\right) \cap L^2\left(0, T; H_0^1(\Omega)\right).$$

$$y = y(x,t) = solution = state, \; f = f(x,t) = control$$

Goal: Drive the dynamics to equilibrium by means of a suitable choice of the control

$$y(\cdot, T) \equiv y^*(x).$$



We address this problem fro a classical optimal control / least square approach:

$$min\,\frac{1}{2}\left[\int_0^T\int_\omega |f|^2 dxdt + \int_\Omega |y(x,T)-y^*(x)|^2 dx\right].$$

According to Pontryagin's Maximum Principle the Optimality System (OS) reads

$$y_t - \Delta y = p1_\omega \text{ in } Q$$

$$-p_t - \Delta p = 0 \text{ in } Q$$

$$y = 0 \text{ on } \Sigma$$

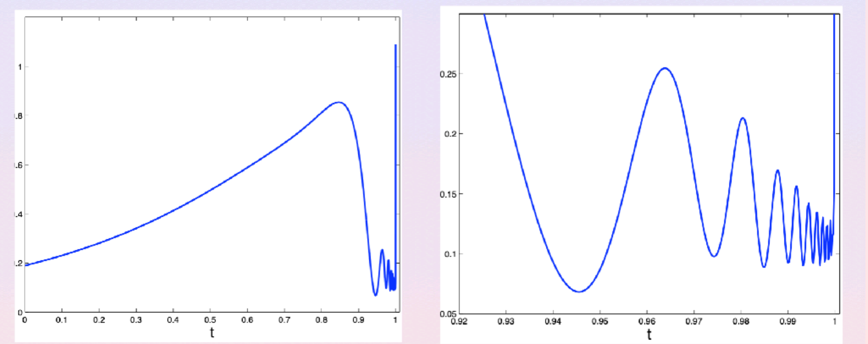
$$y(x,0) = y^0(x) \text{ in } \Omega$$

$$p(x,T) = y(x,T) - y^*(x) \text{ in } \Omega$$

$$p = 0 \text{ on } \Sigma.$$

And the optimal control is:

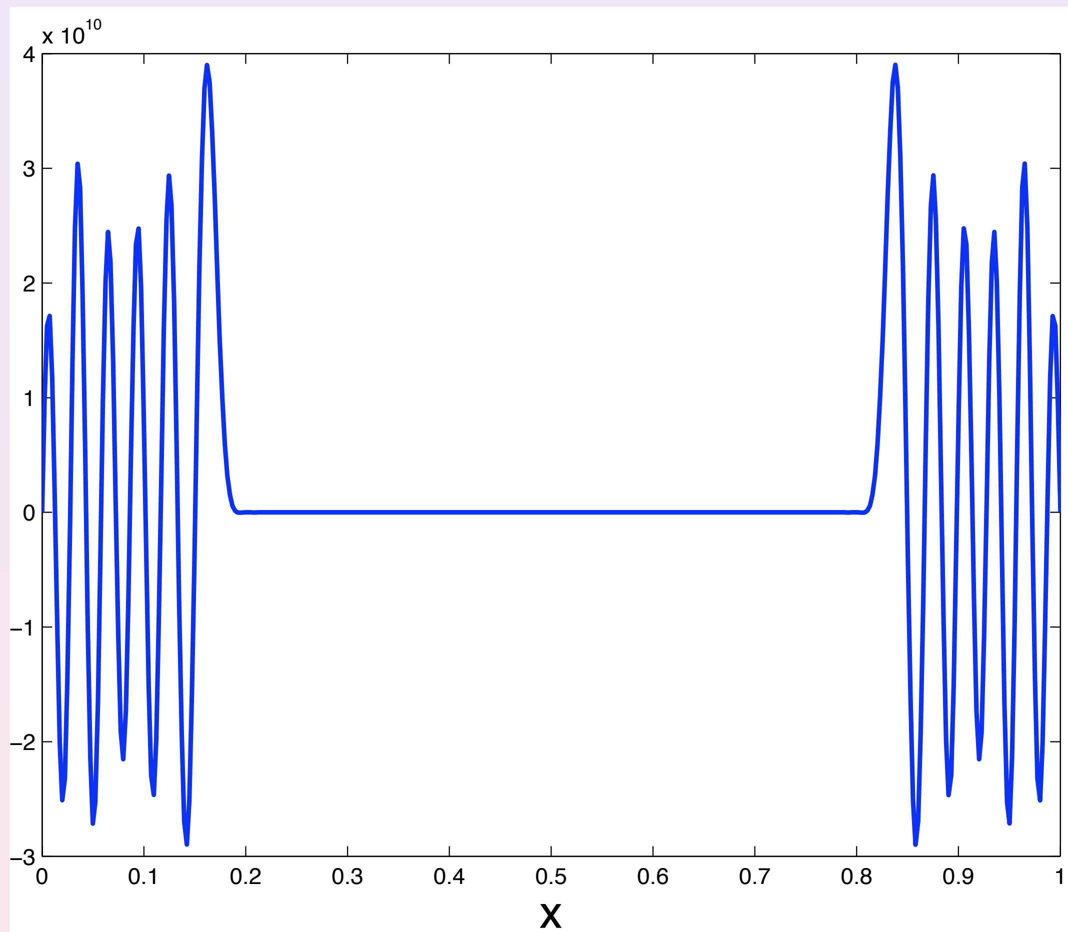
$$f(x,t) = p(x,t) \quad \text{in}\, \omega \times (0,T).$$



By duality (Fenchel-Rockafellar) the adjoint p at time $t = T$, p^T saturates the regularity properties required to assure the well-posedness of the functional:

$$\mathcal{H} = \{p^T : p(x,0) \in L^2(\Omega)\}$$

This is a huge space, allowing an exponential increase of Fourier coefficients at high frequencies. And, because of this, we observe the tendency of the control to concentrate all the action in the final time instant $t = T$, incompatible with turnpike effects[4]



Tychonoff's monster (1935)

Théorèmes d'unicité pour l'équation de la chaleur

A. Tychonoff (Moscou)

[4] A. Münch & E. Z., Inverse Problems, 2010

Remedy: Better balanced controls

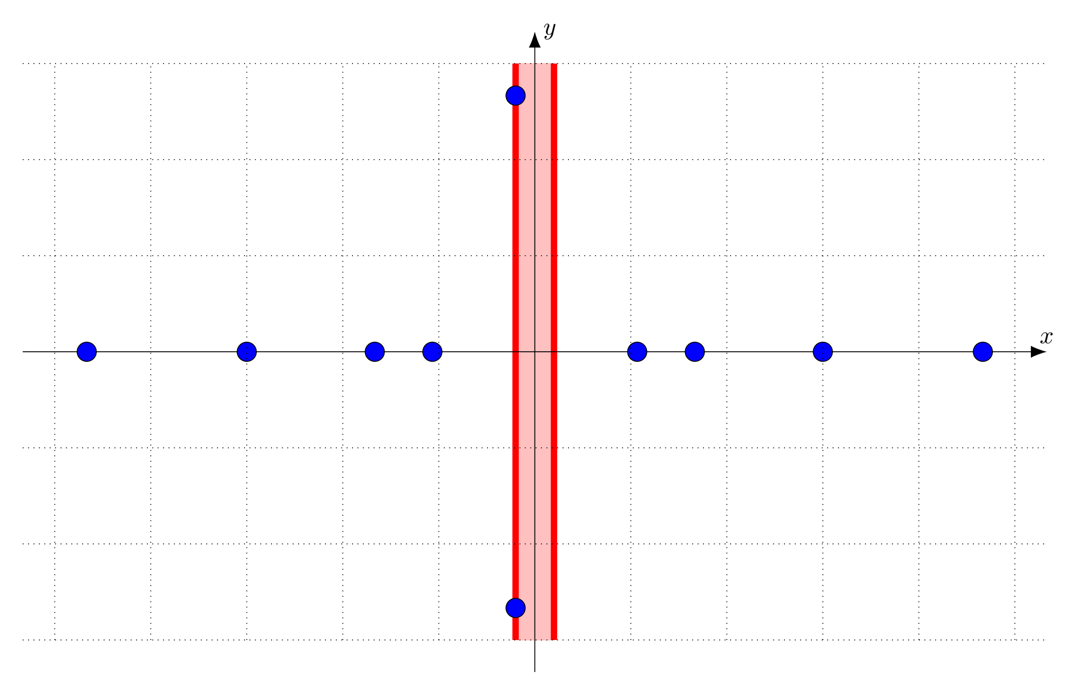
Let us now consider the control f minimising a compromise between the norm of the state and the control among the class of admissible controls:

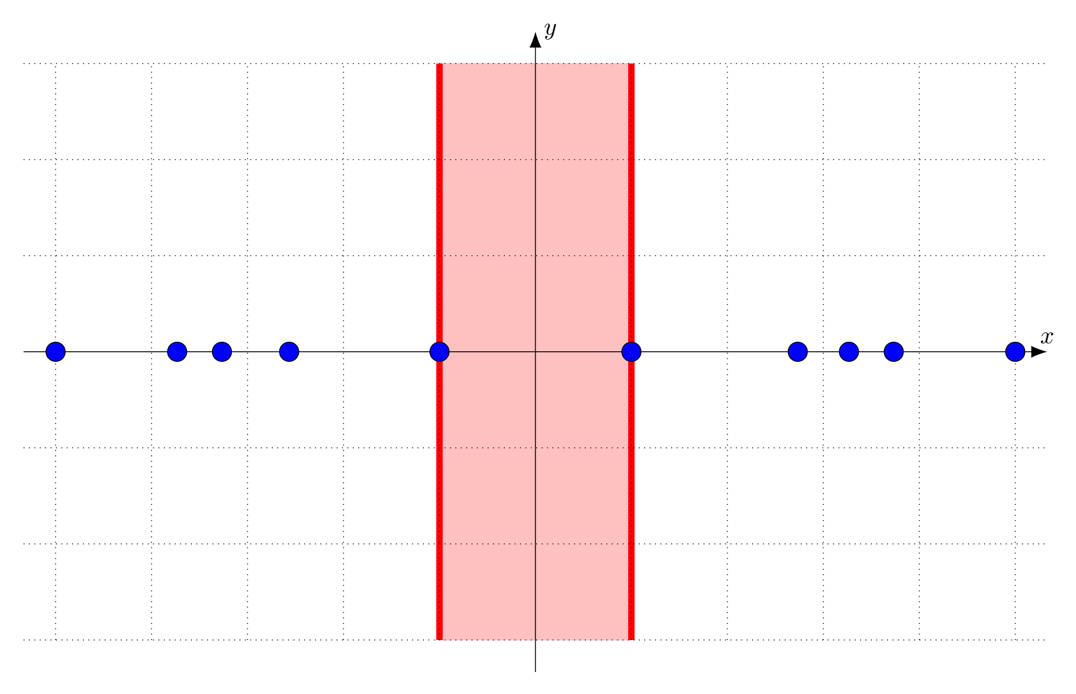
$$min \frac{1}{2}\left[\int_0^T\int_\Omega |y|^2 dxdt + \int_0^T\int_\omega |f|^2 dxdt + \int_\Omega |y(x,T) - y^*(x)|^2\right].$$

Then the Optimality System reads

$$y_t - \Delta y = -p1_\omega \text{ in } Q$$
$$-p_t - \Delta p = y \text{ in } Q$$
$$y = p = 0 \text{ on } \Sigma$$
$$y(x,0) = y^0(x) \text{ in } \Omega$$
$$p(x,T) = y(x,T) - y^*(x) \text{ in } \Omega$$

We now observe a coupling between p and y on the adjoint state equation![7]





[7] A. Porretta & E. Z., SIAM J. Cont. Optim., 2013.

New Optimality System Dynamics

What is the dynamic behaviour of solutions of the new fully coupled OS?
For the sake of simplicity, assume $\omega = \Omega$.

The dynamical system now reads

$$y_t - \Delta y = -p$$

$$p_t + \Delta p = -y$$

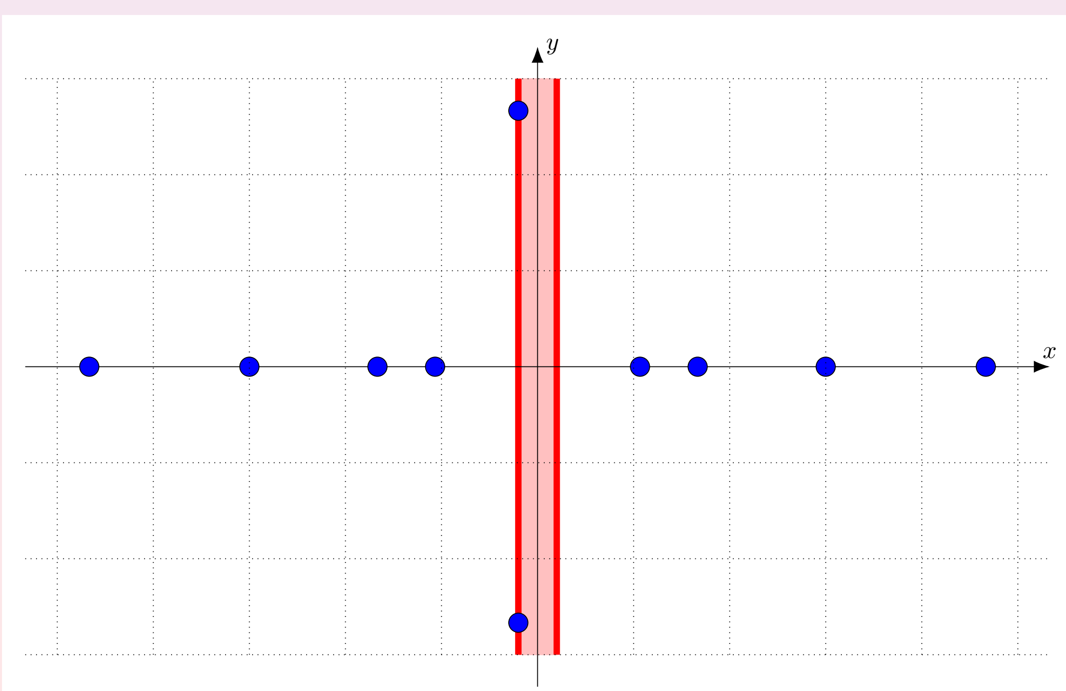
This is a forward-backward parabolic system.

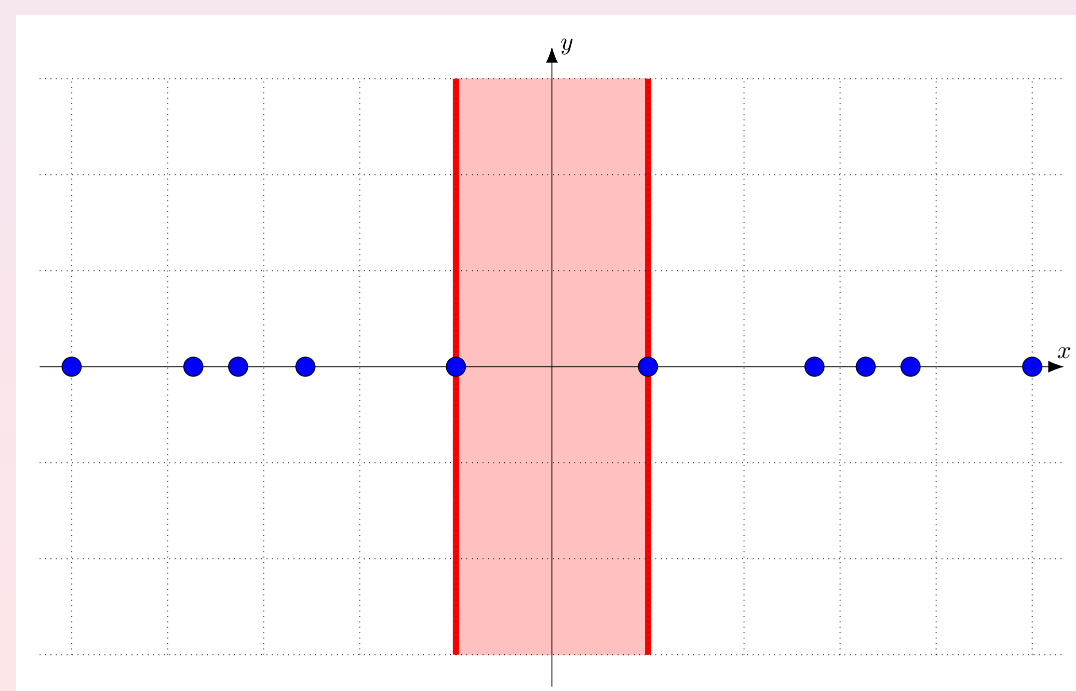
A spectral decomposition exhibits the characteristic values

$$\mu_j^{\pm} = \pm\sqrt{1 + \lambda_j^2}$$

where $(\lambda_j)_{j \geq 1}$ are the (positive) eigenvalues of $-\Delta$.

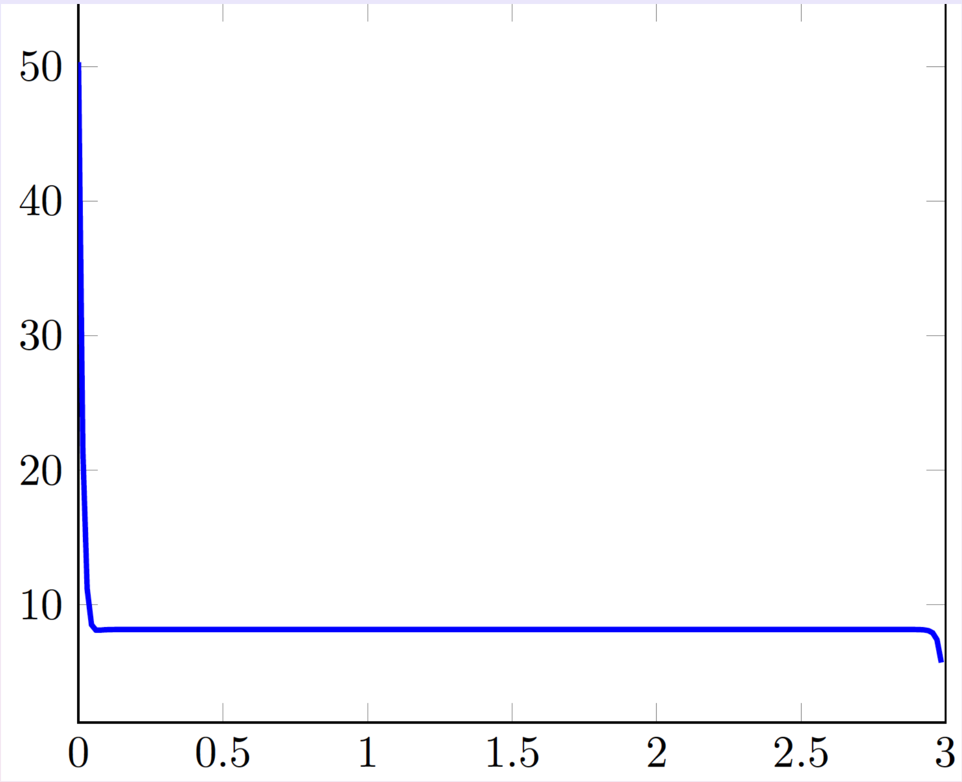
Thus, the system is the superposition of growing + diminishing real exponentials.





The turnpike property for the heat equation

New dynamics $=$ combination of exponentially stable and unstable branches $\equiv$ compatible with turnpike



The turnpike behaviour is ensured when $T \to \infty$ when the cost functional penalizes sufficiently state and control.

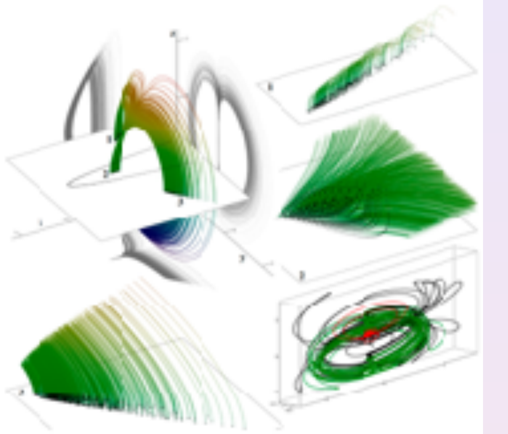
[Controllability] $+$ [Coercive in state $+$ control cost] $\to$ Turnpike

The same occurs for wave propagation[5]

[5] M. Gugat, E. Trélat, E. Zuazua, Systems and Control Letters, 90 (2016), 61-70.

Linear theory. Joint work with A. Porretta, SIAM J. Cont. Optim., 2013.

The same methods apply in the inifinite-dimensional context, covering in particular linear heat and wave equations



Consider the finite dimensional dynamics

$$\begin{cases} x_t + Ax = Bu \\ x(0) = x_0 \in \mathbb{R}^N \end{cases}$$

where $A \in M(N, N)$, $B \in M(N, M)$, with control $u \in L^2(0, T; \mathbb{R}^M)$.
Given a matrix $C \in M(N, N)$, and some $x^* \in \mathbb{R}^N$, consider the optimal control problem

$$\min_u \; J^T(u) = \frac{1}{2}\int_0^T (|u(t)|^2 + |C(x(t) - x^*)|^2)dt \, .$$

There exists a unique optimal control $u(t)$ in $L^2(0, T; \mathbb{R}^M)$, characterized by the optimality condition

$$u = -B^* p \, , \qquad \begin{cases} x_t + Ax = -BB^* p \\ x(0) = x_0 \end{cases} , \qquad \begin{cases} -p_t + A^* p = C^* C(x - x^*) \\ p(T) = 0 \end{cases} \tag{3}$$

The steady state control problem

The same problem can be formulated for the steady-state model

$$Ax = Bu.$$

Then there exists a unique minimum $\bar{u}$, and a unique optimal state $\bar{x}$, of the stationary control problem

$$\min_u J_s(u) = \frac{1}{2}(|u|^2 + |C(x - x^*)|^2) \tag{4}$$

which is nothing but a constrained minimization in $\mathbb{R}^N$.
The optimal control $\bar{u}$ and state $\bar{x}$ satisfy

$$\bar{u} = -B^*\bar{p}, \qquad A\bar{x} = B\bar{u}, \quad \text{and} \quad A^*\bar{p} = C^*C(\bar{x} - x^*).$$

We assume that

$$(A, B) \text{ is controllable,} \tag{5}$$

or, equivalently, that the matrices A, B satisfy the Kalman rank condition

$$Rank \left[B \ AB \ A^2 B \ldots \ A^{N-1} B \right] = N \ . \tag{6}$$

Concerning the cost functional, we assume that the matrix C is such that (void assumption when $C = Id$)

$$(A, C) \text{ is observable} \tag{7}$$

which means that the following algebraic condition holds:

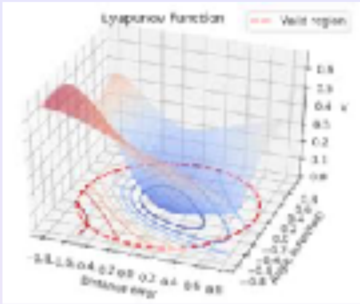
$$Rank \left[C \ CA \ CA^2 \ldots \ CA^{N-1} \right] = N \ . \tag{8}$$

$$x_t + Ax = Bu$$

$$J^T(u) = \frac{1}{2} \int_0^T (|u(t)|^2 + |C(x(t) - x^*)|^2) dt$$

$$\begin{cases} x_t + Ax = Bu \\ -p_t + A^* p = C^* Cx \end{cases}$$

Proofs

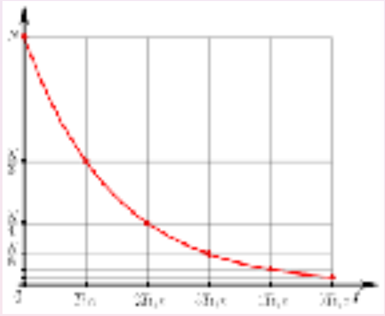


Proof # 1: Dissipativity

$$\frac{d}{dt}[(x-\bar{x})(p-\bar{p})] = -\left[B^*(p-\bar{p})|^2 + |C(x-\bar{x})|^2\right]$$

That is the starting point of a turnpike proof. Note however that it is much trickier than the classical Lyapunov stability: Two boundary layers at $t=0$ and $t=T$, moving time-horizon $[0,T]$...

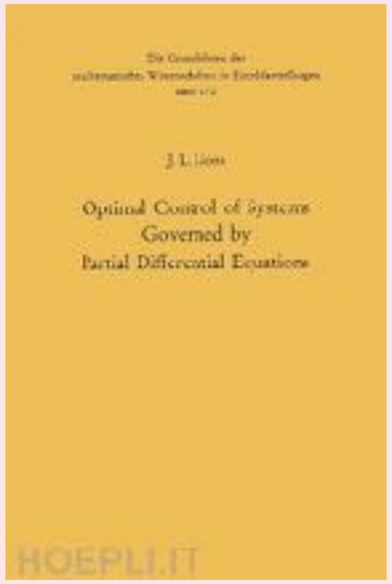
Proof #2 : Riccati



- Consider the Infinite Horizon Linear Quadratic Regulator (LQR) problem in $[0,\infty)$ with null target $x^* \equiv 0$.
- Employ Riccati feedback exponential stabilizator.
- Cut-it-off onto $[0,T]$.
- Correct the boundary layer at $t=T$ to match the terminal conditions .

Proof # 3: Singular perturbations Implement the change of variables $t \to sT$:

$$t \in [0,T] \Longleftrightarrow s \in [0,1].$$



System

$$x_t + Ax = Bu, \quad t \in [0,T]$$

becomes

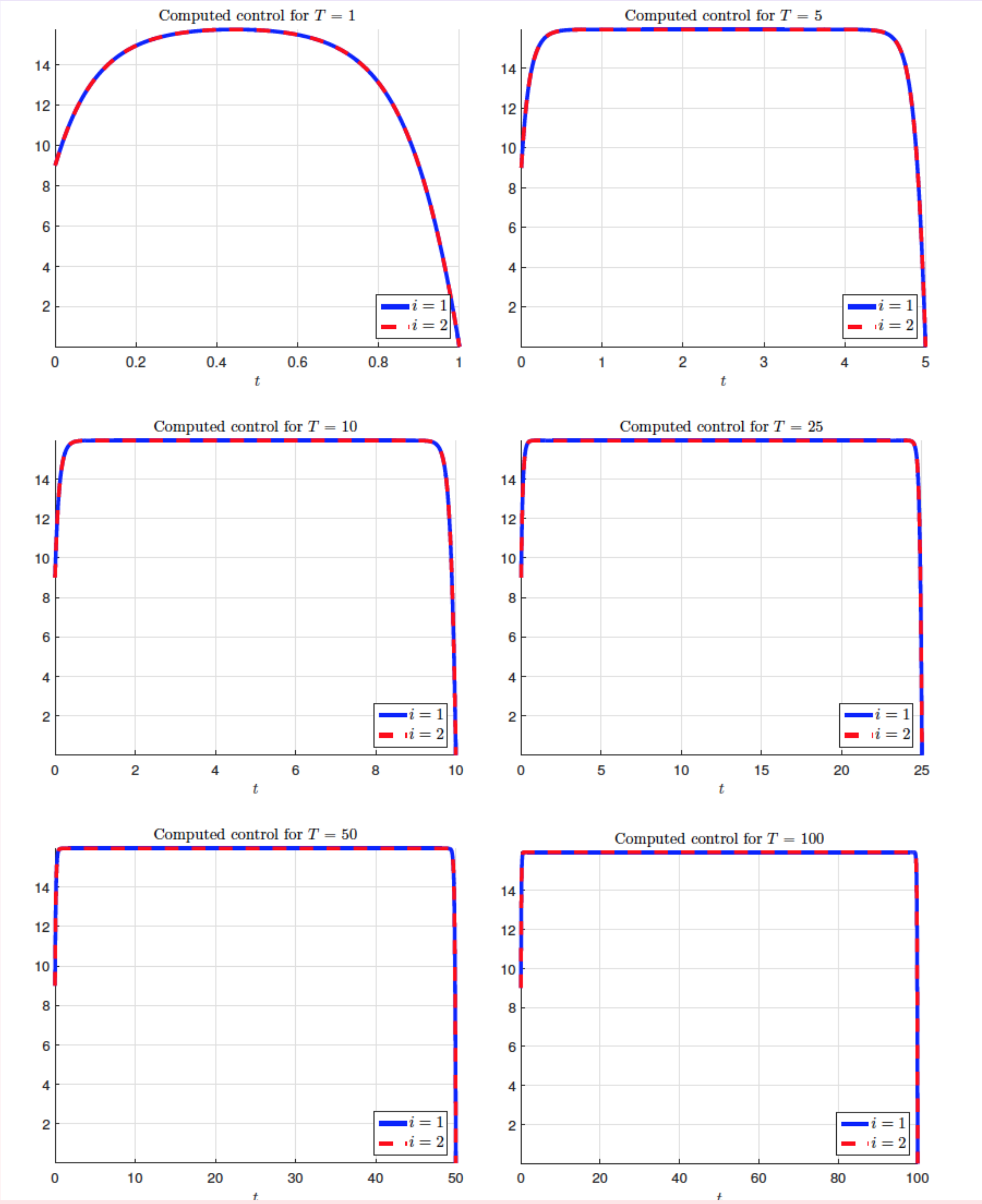
$$\frac{1}{T}x_s + Ax = Bu, \quad s \in [0,1]$$

As $T \to \infty$, $\varepsilon = 1/T \to 0$.

Nonlinear problems

A. Porretta & E. Z., Mathematical Paradigms of Climate Science, Springer INdAM Series, 2016.

The theory can be extended for semilinear PDEs provided the target is small employing linearization techniques.
But numerical simulations shows that the property is much more robust. [6]
Developing a complete theorem, for large targets and deformations, is a challenging problem.

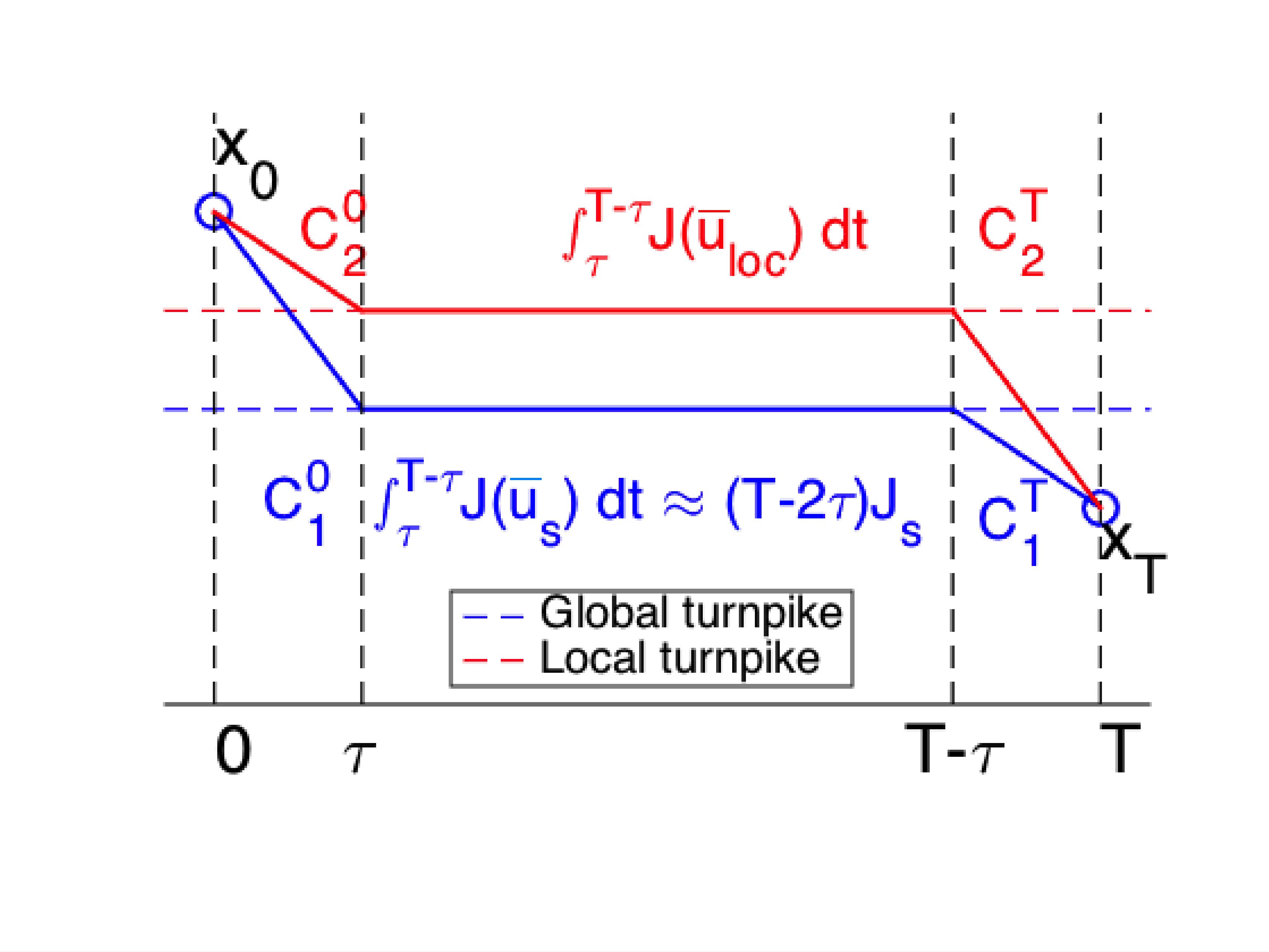


[6] Simulations by S. Volkwein for a cubic semilinear equation with large targets.

Heuristic explanation and practical use

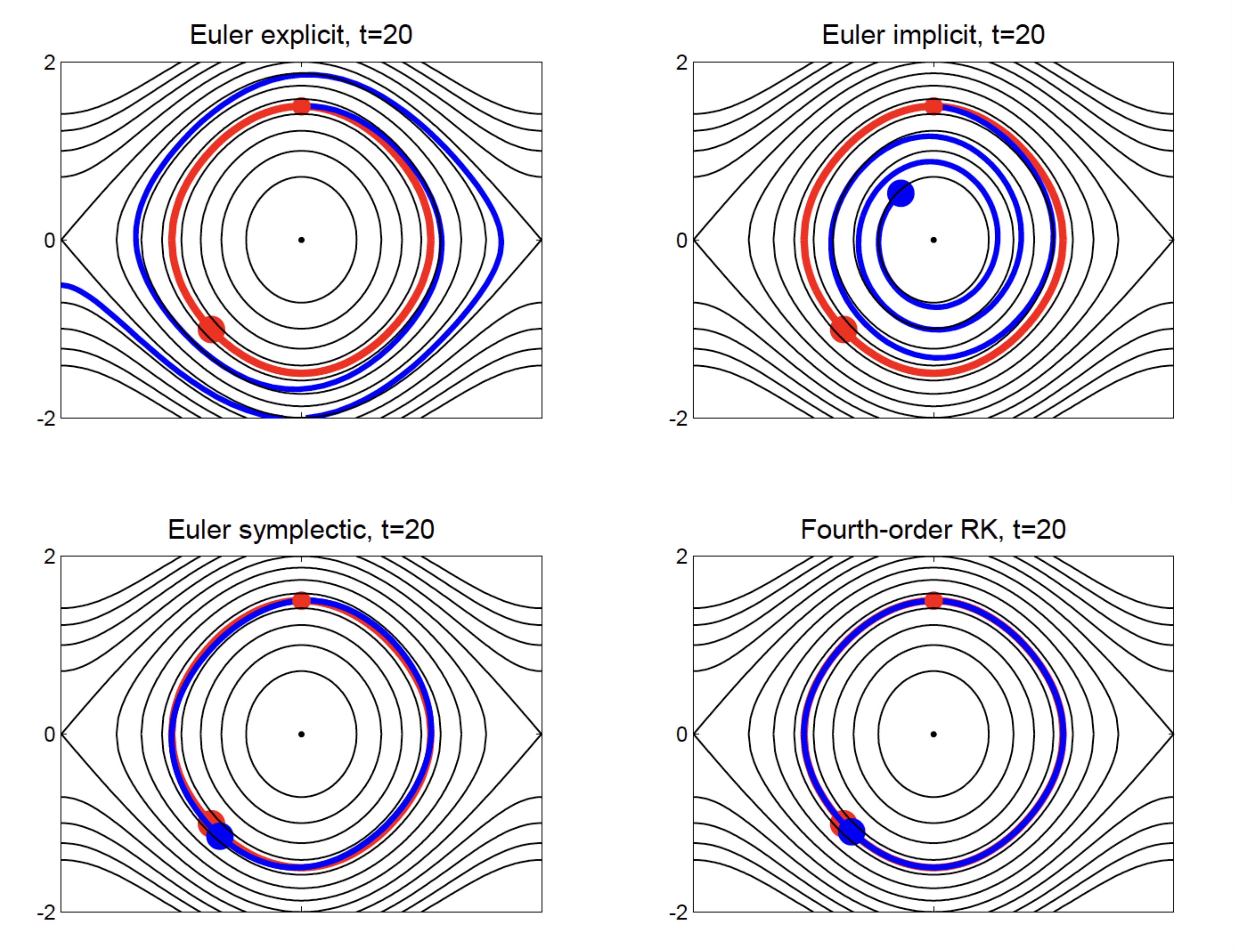
In applications and daily life we use a quasi-turnpike principle, which is very robust and ubiquitous, even in the context of multiple steady optima (local or global):

- Step 1: Compute the optimal steady optimal control and state,
- Step 2: Drive the system from the initial configuration to this steady state one;
- Step 3: Remain in this steady configuration as long as possible;
- Step 4: Exit this configuration if a terminal condition is to be met.



Warning! Long time numerics plays a key role: Geometric/Symplectic integration; Well balanced numerical schemes...

Numerical integration of the pendulum (A. Marica)



An open problem and biblio

Further extend the turnpike theory for nonlinear PDE, getting rid of the smallness condition on the target, which in numerical simulations seems to be unnecessary.

- A. Porretta, E. Z., SIAM J. Control. Optim., 51 (6) (2013), 4242-4273.
- A. Porretta, E. Z., Springer INdAM Series "Mathematical Paradigms of Climate Science", F. Ancona et al. eds, 15, 2016, 67-89.
- E. Trélat, E. Z., JDE, 218 (2015) , 81-114.
- M. Gugat, E. Trélat, E. Z., Systems and Control Letters, 90 (2016), 61-70.
- E. Z., Annual Reviews in Control, 44 (2017) 199-210.
- E.Trélat, C. Zhang, E. Z., SIAM J. Control Optim. **56** (2018), no. 2, 1222–1252.
- V. Hernández-Santamaria, M. Lazar, E.Z. Numerische Mathematik (2019) 141:455-493.
- D. Pighin, N. Sakamoto, E. Z., IEEE CDC Proceedings, Nice, 2019.
- G. Lance, E. Trélat, E. Z., Systems & Control Letters 142 (2020) 104733.
- J. Heiland, E. Z., arXiv:2007.13621, 2020.
- C. Esteve, H. Kouhkouh, D. Pighin, E. Z., arxiv.org/pdf/2006.10430, 2020.
- M. Gugat, M. Schuster and E. Z., SEMA/SIMAI Springer Series, 2020.

And further interesting work by collaborators: S. Zamorano (NS), M. Warma & S. Zamorano, Fractional heat,...

Our thanks to our FAU colleague Daniel Tenbrinck. He suggested to us to explore turnpike for Neural Networks.

Outline

Universal approximation theorem I

Math. Control Signals Systems (1989) 2: 303–314

Mathematics of Control, Signals, and Systems

Approximation by Superpositions of a Sigmoidal Function*

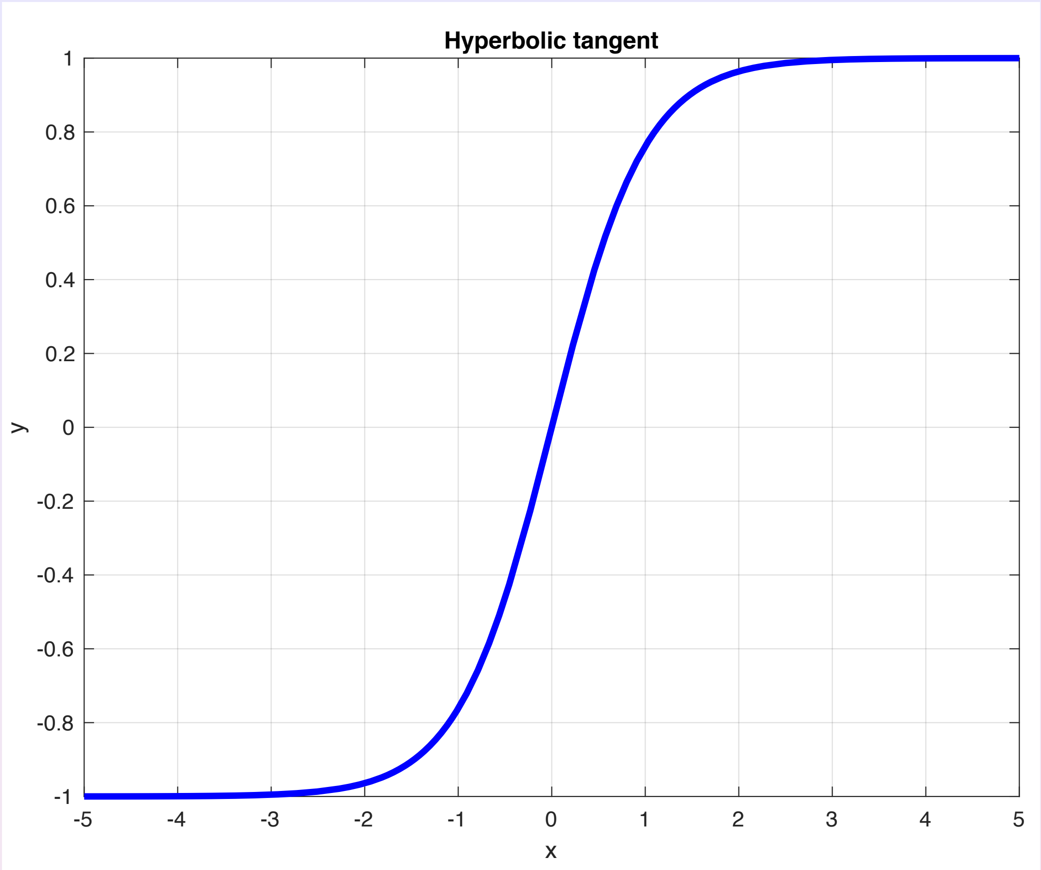
G. Cybenko†

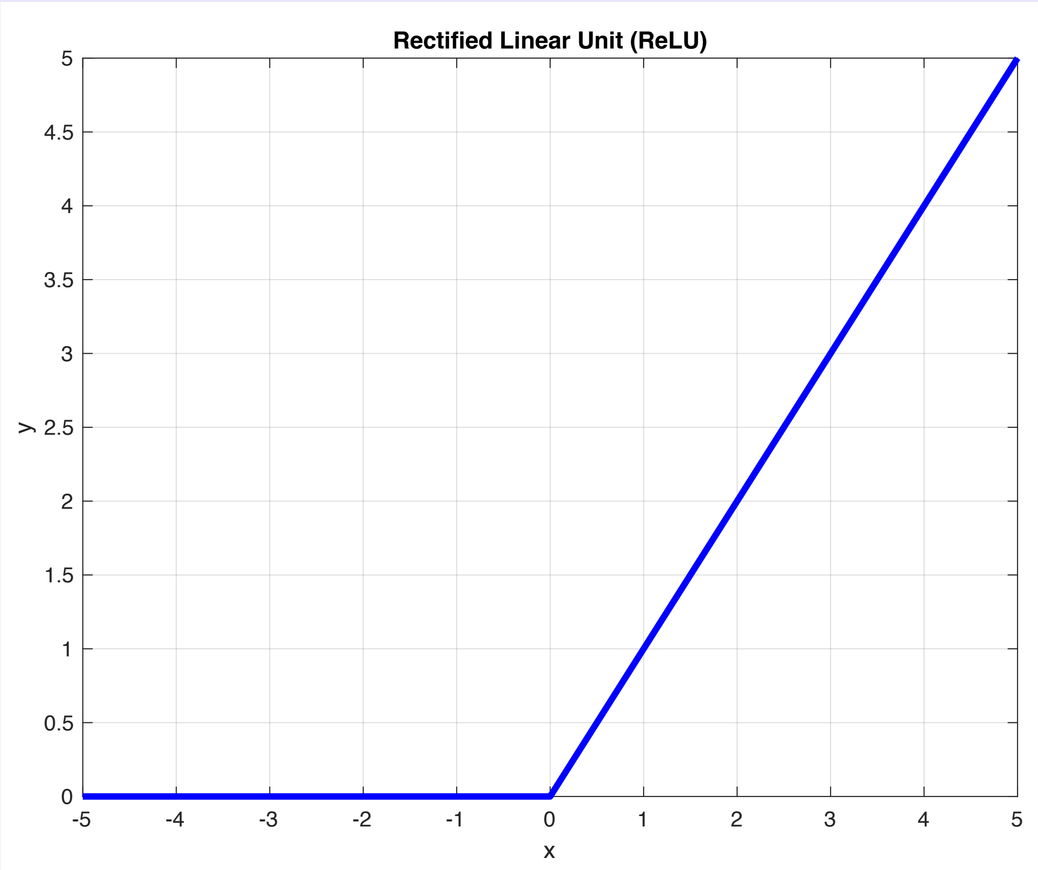
$$\sum_{j=1}^{N} \alpha_j \sigma(y_j^{\mathrm{T}} x + \theta_j), \quad (1)$$

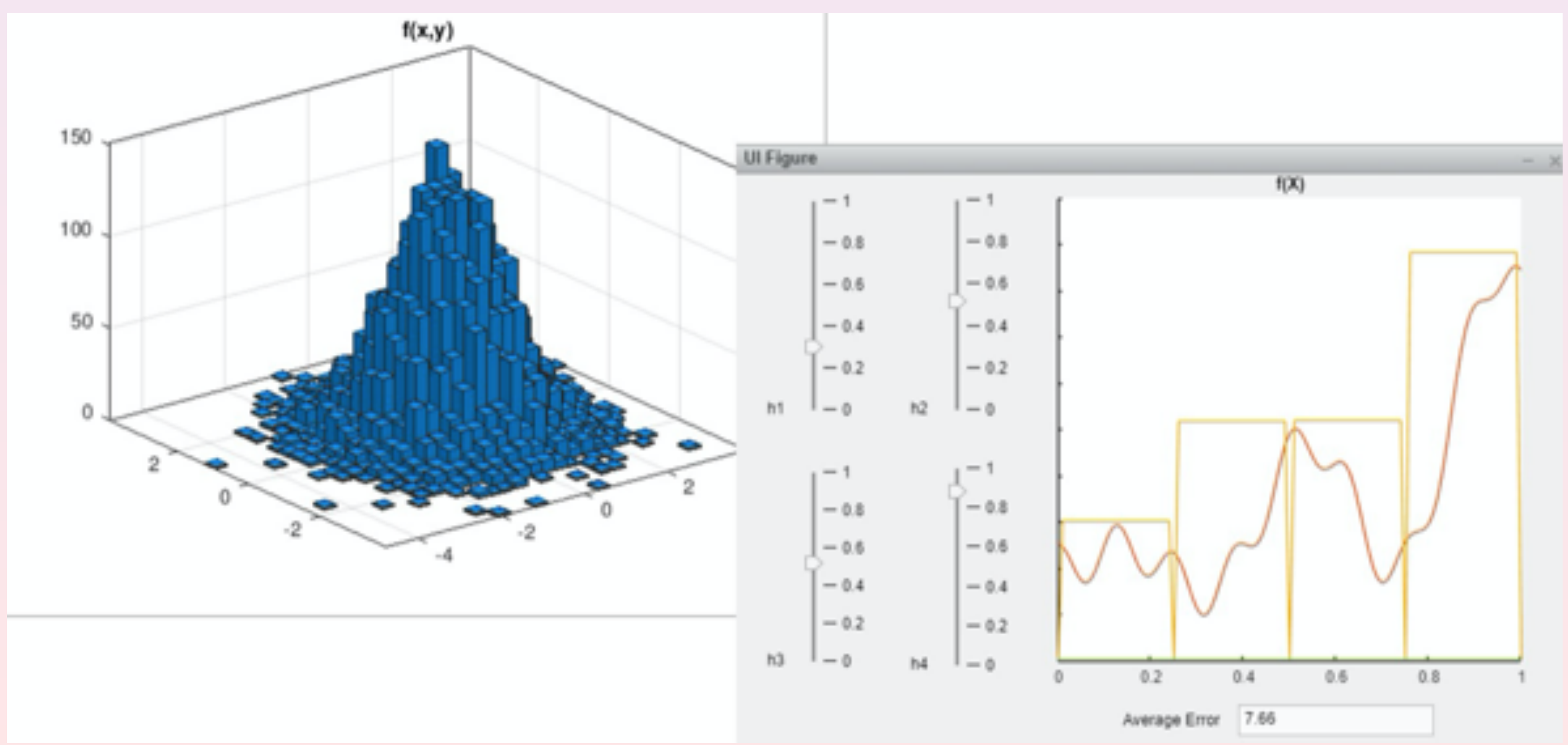
where $y_j \in \mathbb{R}^n$ and $\alpha_j, \theta \in \mathbb{R}$ are fixed. (y^{T} is the transpose of y so that $y^{\mathrm{T}}x$ is the inner product of y and x.) Here the univariate function σ depends heavily on the context of the application. Our major concern is with so-called sigmoidal σ's:

$$\sigma(t) \to \begin{cases} 1 & \text{as} \quad t \to +\infty, \\ 0 & \text{as} \quad t \to -\infty. \end{cases}$$

Universal approximation theorem II







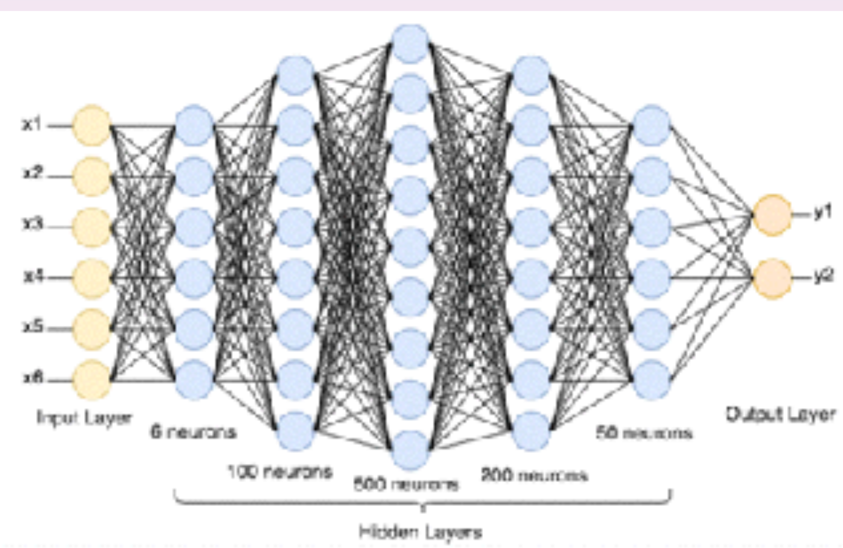
Supervised learning

Goal: Find an approximation of a function $f_\rho : \mathbb{R}^d \to \mathbb{R}^m$ from a dataset

$$\{\vec{x}_i, \vec{y}_i\}_{i=1}^N \subset \mathbb{R}^{d\times N} \times \mathbb{R}^{m\times N}$$

drawn from an unknown probability measure ρ on $\mathbb{R}^d \times \mathbb{R}^m$.

Classification: match points (images) to respective labels (cat, dog).
$\longrightarrow$ Popular method: **training a neural network**.



Residual neural networks

[1] K He, X Zhang, S Ren, J Sun, 2016: Deep residual learning for image recognition
[2] E. Weinan, 2017. A proposal on machine learning via dynamical systems.
[3] R. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud, 2018. Neural ordinary differential equations.
[4] E. Sontag, H. Sussmann, 1997, Complete controllability of continuous-time recurrent neural networks.

ResNets

$$\begin{cases} \mathbf{x}_i^{k+1} = \mathbf{x}_i^k + hW^k\sigma(A^k\mathbf{x}_i^k + b^k), & k \in \{0, \ldots, N_{layers} - 1\} \\ \mathbf{x}_i^0 = \tilde{\mathbf{x}}_i, \quad i = 1, ..., N \end{cases}$$

where $h = 1$, σ globally Lipschitz $\sigma(0) = 0$.

nODE

Layer $=$ timestep; $h = \frac{T}{N_{layers}}$ for given $T > 0$

$$\begin{cases} \dot{\mathbf{x}}_i(t) = W(t)\sigma(A(t)\mathbf{x}_i(t) + b(t)) & \text{for } t \in (0, T) \\ \mathbf{x}_i(0) = \vec{\mathbf{x}}_i, \quad i = 1, ..., N \end{cases}$$

The problem becomes then a giant simultaneous control problem in which each initial datum $\mathbf{x}_i(0)$ needs to the driven to the corresponding destination for all $i = 1, ..., N$ with the same controls:

- What happens when $T \to \infty$, i.e. in the deep, high number of layers regime?[8] [9]

[8] C. Esteve, B. Geshkovski, D. Pighin, E. Zuazua, Large-time asymptotics in deep learning, arXiv:2008.02491

[9] D. Ruiz-Balet & Zuazua, Neural ODE control for classification, approximation and transport, arXiv:2104.05278

Special features of the control of ResNets

- Nonlinearities are unusual in Mechanics: σ is flat in half of the phase space.
- We need to control many trajectories (one per item to be classified) with the same control! [10]

The very nature of the activation function σ allows actually to achieve this monster simultaneous control goal. The fact that σ leaves half of the phase space invariant while deforming the other one, allows to build dynamics that are not encountered in the classical ODE systems in mechanics and for which such kind of simultaneous control property is unlikely or even impossible.

[10] This would be impossible for instance, for the standard linear system $x' = Ax + Bu$.

Turnpike for ResNets

$T \to \infty \quad \sim \quad N_{\text{layers}} \to \infty.$

Turnpike refers to the fact that, in long time-horizons, optimal controls and trajectories are exponentially close to the optimal steady-state control and state in most of the time-horizon.

Supervised Learning* $\Longleftrightarrow$ **minimize**[11]

$$\boxed{\frac{1}{N}\int_0^T \|P\mathbf{x}(t) - \overline{y}\|^2 dt + \alpha \|u\|^2_{H^1(0,T;\mathbb{R}^{d_u})}.} \tag{SL*}$$

$\overline{y} := [\vec{y}_1, \ldots, \vec{y}_N]$, $u := [A, W, b]$ in (20) and $P : \mathbb{R}^d \to \mathbb{R}^m$.

Theorem (Turnpike): Under controllability assumptions, for any sufficiently large T, an optimal solution $(u_T, \mathbf{x}_T)$ to (SL*)–(20) satisfies

$$\|u_T\|_{H^1(0,T;\mathbb{R}^{d_u})} \leq C_1$$

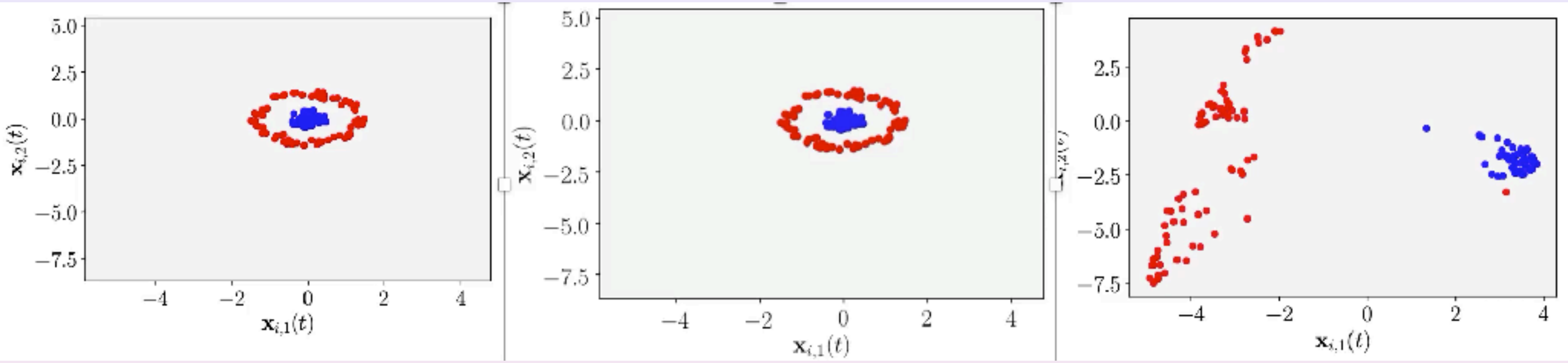
and

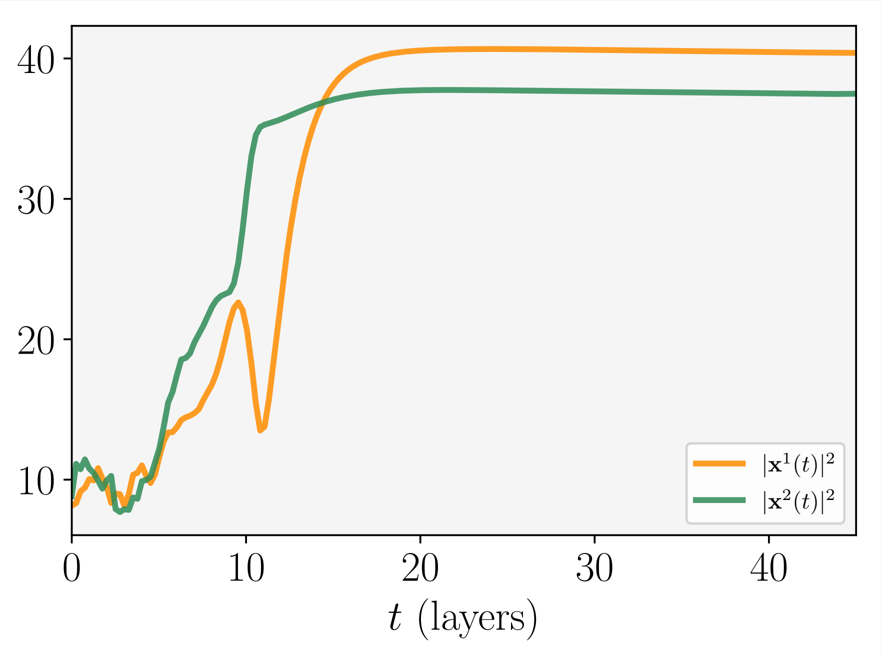
$$\|P\mathbf{x}_T(t) - \overline{y}\| \leq C_2 e^{-\mu t} \qquad \forall t \in [0, T]$$

for some $C_1 > 0$, $C_2 > 0$ and $\mu > 0$, all independent of T.

[11] Note that in this context we do not impose a perfect classification. We just expect that it will occur with high probability as an outcome of the optimar control problem

$$T \to \infty \quad \sim \quad N_{\text{layers}} \to \infty.$$

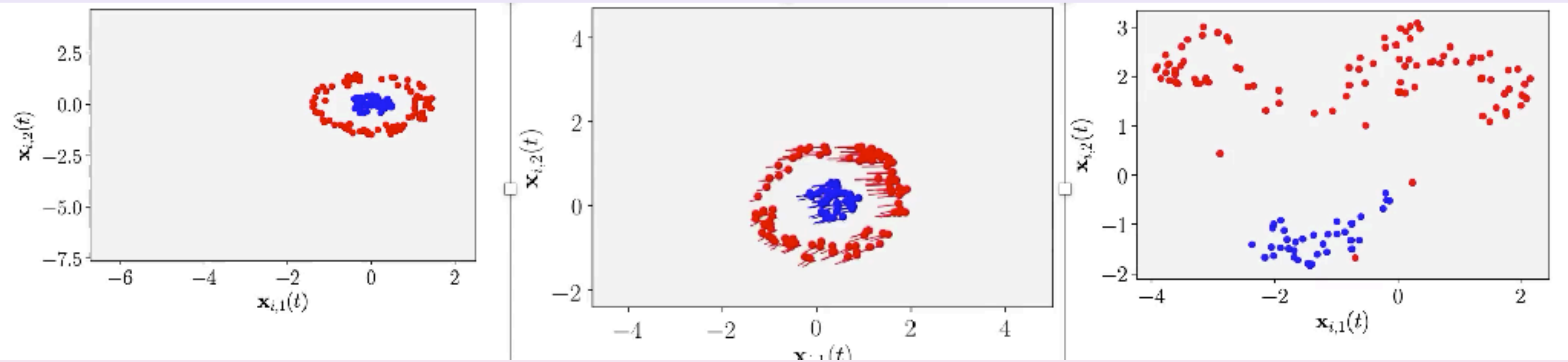


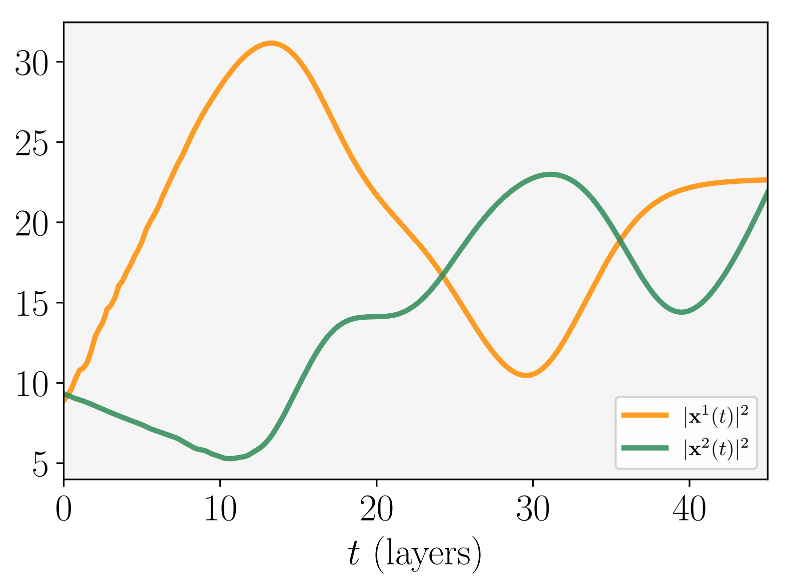


Classical SL problem?

$P : \mathbb{R}^d \to \mathbb{R}^m$, **minimize**

$$\frac{1}{N}\|P\mathbf{x}(T) - \overline{y}\|^2 + \alpha \|u\|^2_{H^1(0,T;\mathbb{R}^{d_u})}. \tag{SL}$$





Convergence of $\mathbf{x}(T)$ to $P^{-1}(\{\overline{y}\})$ when $T \to \infty$, but slow (no turnpike).

Classification by control

Theorem (Classification, Domènec Ruiz-Balet EZ, 2021)

[a] *Let* σ *be the ReLU.*
Let $d \geq 2$, *and* $N, M \geq 2$.
Let $\{x_i\}_{i=1}^N \subset \mathbb{R}^d$ *be data to be classified into disjoint open non-empty subsets* $S_m, m = 1, ..., M$ *with labels* $m = m(i), i = 1, ..., N$.
Then, for every $T > 0$, *there exist control functions* $A, W \in L^\infty\left((0,T); \mathbb{R}^{d\times d}\right)$ *and* $b \in L^\infty\left((0,T), \mathbb{R}^d\right)$ *such that the flow associated to the Neural ODE, when applied to all initial data* $\{x_i\}_{i=1}^N$, *classifies them simulatenously, i.e.*

$$\phi_T(x_i; A, W, b) \in S_{m_i}, \quad \forall i = 1, ..., N.$$

Furthermore,

- *Controls are piecewise constant with a maximal finite number of switches of the order of* $\mathcal{O}(N)$. *They also lie in* BV.
- *The control time* $T > 0$ *can be made arbitrarily small (scaling).*
- *The complexity of controls diminishes when initial data are structured in clusters.*
- *The complexity of controls also diminishes when the control requirement is relaxed so that not all data need to be classified.*
- *The targets* S_m *can be just* N *distinct points in the euclidean space.*

[a] Related results for smooth sigmoids using Lie bracket control techniques: A. Agrachev and A. Sarychev, arXiv:2008.12702, (2020).

Neural transport equations

The simultaneous control of the nODE

$$\begin{cases} \dot{x} = W(t)\boldsymbol{\sigma}(A(t)x + b(t)) \\ x(0) = x_i, \quad i = 1, ..., N \end{cases}$$

to arbitrary terminal states

$$x(T) = y_i, \quad i = 1, ..., N$$

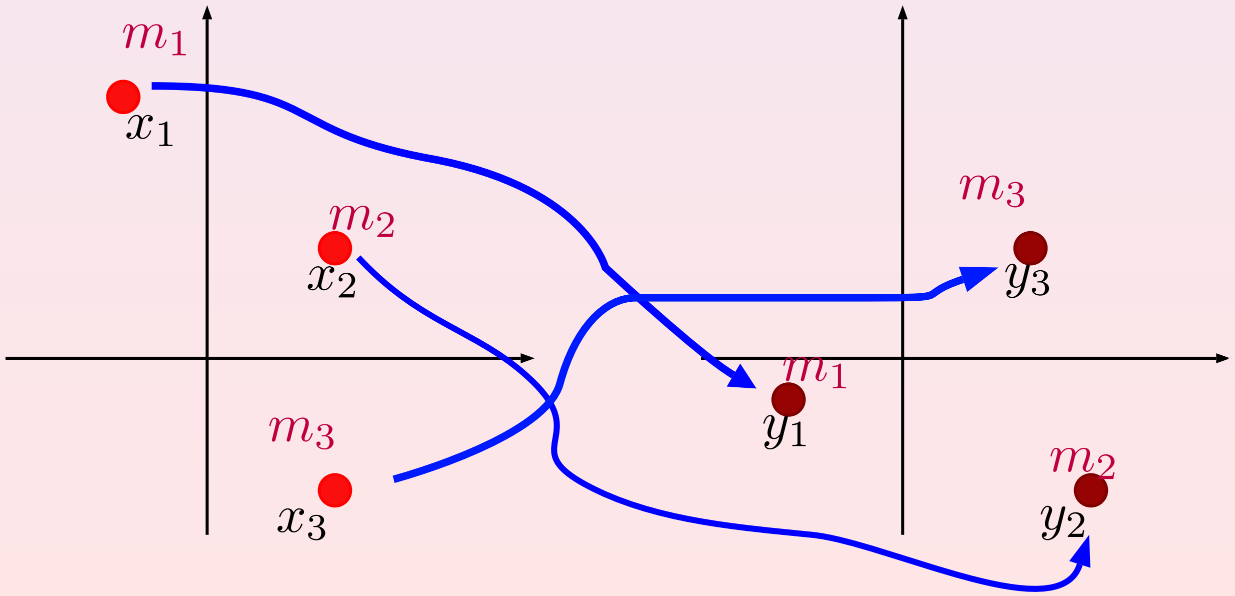
in terms of the transport equation, leads to the control of an atomic initial datum from

$$\rho(x, 0) = \sum_{i=1}^{N} m_i \delta_{x_i}$$

to the terminal one

$$\rho(x, T) = \sum_{i=1}^{N} m_i \delta_{y_i}.$$

But note that, even if the locations of the masses are transported, the amplitude of the masses do not vary.



Concluding remarks

An extraordinary and fertile field in the interplay between Dynamical Systems, Control, Machine Learning and applications

- Control and dynamical systems tools allow to explain the amazing efficiency of Neural Networks (NN) in some specific applications.
- Long-time / Turnpike control arise naturally in Deep Learning
- Interesting open questions:
  - How to deal with Neural ODEs that switch in dimension of the Euclidean phase space.
  - Are there results explaining how the clustering of data (number of separating interfaces needed) diminishes in higher dimensions?
  - How close is our piecewise constant control strategy from the optimal one (in the Pontryagin sense?)
  - How does our control strategy compare to those obtained in a purely NN setting?
  - How does the complexity of the controls diminish when we relax the classification criteria?
  - Links with Optimal Transport.
  - Other objectives: Unsupervised learning?

- My sincere thanks to collaborators: A. Porretta (Roma 2), E. Trélat (Paris Sorbonne), M. Gugat (FAU), D. Pighin (Innovalia), B. Geskhovski (UAM & Deusto), C. Esteve (UAM & Deusto), M. Schuster (FAU), M. Lazar (Dubrovnik), V. Hernández-Santamaria (Mx), N. Sakamoto (Nanzan), J. Heiland (Magdeburg), H. Kouhkouh (Padova), D. Ruiz-Balet (UAM

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 765579.

THANK YOU FOR THE INVITATION AND YOUR ATTENTION!