Universal ODEs : non-trivial $\underbrace{\text{differential algebraic eq°}}_{\text{DAE ?}}$

DAE : system of eq° that either diff eq° and algebraic eq°

- the derivatives of some variable are not explicitly given
- the system is subject to algebraic constraints

A) General Form (Implicit) :

$$F\left(t, x(t), \frac{dx(t)}{dt}\right) = 0_{\mathbb{R}^n}$$

$F : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \longmapsto \mathbb{R}^n$

A systeme is a DAE if $\frac{\partial F}{\partial \dot{x}}$ is singular i.e. $\frac{\partial F}{\partial \dot{x}} = 0$ where $\frac{\partial F}{\partial \dot{x}}$ is the Jacobian. If $\frac{\partial F}{\partial \dot{x}} \neq 0 \Rightarrow$ ODE.

B) Semi-Explicit Form (Index 1)

We set
- $x \in \mathbb{R}^n$, diff variable
- $z \in \mathbb{R}^m$, alge variable

$$\begin{cases} \dfrac{dx(t)}{dt} = f(t, x, z) \\ 0 = g(t, x, z) \end{cases}$$

This system is of "Index 1" if the Jacobian $\frac{\partial g}{\partial z} \neq 0$
this condition allows the alge variables to be theoretically determined by the diff variab.

C) Differentiation Index

Index: measure between a DAE and ODE

Diff Index

Diff Index : minimum number of times certain eq in the syst must be diff with respect to t to express $\frac{dx(t)}{dt}$ and $\frac{dz(t)}{dt}$ as functions of x, z, t.

example :
$$\begin{cases} \dfrac{dx(t)}{dt} = z \\ x = \sin(t) \quad \text{(Alge constraint)} \end{cases}$$

$$\frac{dx(t)}{dt} = \cos(t) \Rightarrow \cos(t) = z \quad \text{et} \quad \frac{dz(t)}{dt} = -\sin(t) \quad \checkmark$$

Since we diff the constraint twice to find the dynamics of all variables : Index 2 DAE

N.B. : Index ≥ 2 numerically unstable DAEs
↳ need to use Index reduction tech.

Th : Index Reduction by Diff

Given a index $r > 1$, there exists a set of hidden constraints obtained by diff the original algebraic eq $r-1$ times. Replacing the original constraints with these derivatives yields an Index-1 DAE or an ODE

- Equivalent Dynamics : The reduced system shares the same sol as the original, provided the initial condi. are const. for all deri of the constraints

- The Drift Phenomenon: Numerical integration of the reduced system does not naturally satisfy $g(t, x) = 0$, it satisfies $\frac{d^r g}{d^r t} = 0$
⚠ small truncation errors lead to a violation of the original physical law

· Hessenberg Form: Most high-index systems in physics appear in Hessenberg form, where variables are tiered by their distance from the alge constraint

example:

$$\begin{cases} \frac{dx(t)}{dt} = u \\ \frac{dy(t)}{dt} = v \\ \frac{du(t)}{dt} = -\lambda x \\ \frac{dv(t)}{dt} = -\lambda y - g \\ x^2 + y^2 = L^2 \quad \text{constraint} \end{cases}$$

Where λ is the tension. We can see that we have to diff 3 times to find an expression for $\frac{d\lambda}{dt}$.

a) $\frac{d}{dt}\left(x^2 + y^2 = L^2\right) \Rightarrow 2x \frac{dx(t)}{dt} + 2y \frac{dy(t)}{dt} = 0$

$$\Rightarrow \quad xu + yv = 0 \quad , \text{ index } 2.$$

b) $\frac{d}{dt}\left(xu + yv = 0\right) \Rightarrow \frac{dx(t)}{dt} u + x \frac{du(t)}{dt} + \frac{dy(t)}{dt} v + \frac{dv(t)}{dt} y = 0$

We substitute : $u^2 + x(-\lambda x) + v^2 + y(-\lambda y - g) = 0$

$$\Leftrightarrow u^2 + v^2 - \lambda(x^2 + y^2) - gy = 0$$

$$\Leftrightarrow u^2 + v^2 - \lambda L^2 - gy = 0$$

$$\Leftrightarrow \quad \lambda = \frac{u^2 + v^2 - gy}{L^2} \quad \text{index } 1.$$

We can diff λ to get $\frac{d\lambda}{dt}$ if we want a ODE (index 0).

D) Consistency of initial Conditions

The initial conditions must satisfy the constraints

For example, a semi-explicit system, the set of a consist initial cond:

$$\Sigma = \{ (x_0, z_0) \mid g(t_0, x_0, z_0) = 0 \}$$

Starting outside this manifold Σ will cause numerical solvers to fail immediately

E) The Mass Matrix Formulation

DAEs are often written using a Mass Matrix $M(x,t)$

$$M(x,t)\,\dot{x} = f(x,t)$$

- if M is the id matrix I, it's an ODE
- if $\det(M) = 0$, it's a DAE

Now we can dive into UDEs:

A) Defs:

It's a dynamical syst where part of the vector field is def by a universal approximator (a Neural Network: NN)

This syst is defined as:

$$\frac{dx}{dt} = f(t, x, \theta) + NN_\phi(x, \pi)$$

where:

- $f(t, x, \theta)$: the known part of the physics
- $NN_\phi(x, \pi)$: A neural network with weights ϕ that approx the unknown term.
- π : external inputs or auxilary para

B) The Universal Approx Theorem

"A feed forward NN with at least one hidden layer and a non-linear function can approximate any continuous funct $g(x)$ to any desired degree of accuracy $\varepsilon > 0$ on a compact subset of $\mathbb{R}^n$"

C) Training UDEs : Adjoint Sensitivity Analysis.

· Calculating the grad of a loss function $\mathcal{L}$ with respect to the NN weight ϕ

Th : Adjoint Sensitivity

To compute the gradient $\mathcal{L} = \int_{t_0}^{t} g(x(t), \phi)\, dt$ we solve the original ODE forward time and a secondary Adj Eq° backward in time :

$$\frac{d\lambda}{dt} = -\lambda^T \frac{\partial f}{\partial x}$$

$$\Rightarrow \frac{d\mathcal{L}}{d\phi} = \int_{t_0}^{t_f} \lambda^T \frac{\partial f}{\partial \phi}\, dt$$

Now we can set our model : Gompertzian Growth

↳ phenomenological law describing a growth rate that decays expo as the volume increases accounting for nutrient scarcity and metabolic stress

The dinamics are governed by the ODE:

$$\begin{cases} \dfrac{dN}{dt} = r N \ln\left(\dfrac{K}{N}\right) \\ N(t_0) = N_0 \end{cases}$$

Where $N(t) \in \mathbb{R}^+$: tumor cell pop or volume of cell pop
$r \in \mathbb{R}^+$: proliferation rate
$K \in \mathbb{R}^+$: Carrying capacity of the tissue

To account for unmodeled bio processes (e.g. vascular resis or immunotherapy response) we embed a NN :

$$\frac{dN}{dt} = f(t, N, \theta) = rN\ln\left(\frac{K}{N}\right) + NN_\theta(N)$$

to recover the hidden dynamics we min Loss Function $J(\theta)$:
L^2 Loss Funct :

$$J(\theta) = \sum^{N} \|N(t_i, \theta) - \hat{N}_i\|^2 - \alpha\|\theta\|^2$$

Tikhonov Regu
Weight Decay

If we use finite diff to compute $\nabla_\theta J$
it will cost alot $\Rightarrow$ Adjoint Method $O(1)$ with respect nb para

$$J = \int_{t_0}^{t_f} L(t, N, \theta)\, dt$$

were L is Dirac delta distr (we can also choose hight-w Gaussians centered at t_i)

a) We want to find the sensibility of J with respect to θ
we define the Lagrangian $\mathcal{L}$ by intro a time-dep f $\lambda(t)$
(it's the adjoint state)

$$\mathcal{L}(\lambda, N, \theta) = \int_{t_0}^{t_f} L(t, N, \theta)\, dt - \underbrace{\int_{t_0}^{t_f} \lambda(t)^T \left(\frac{dN}{dt} - f(t, N, \theta)\right) dt}_{0 \text{ if it's a ODE}}$$

$\Rightarrow \mathcal{L} = J$ on the path. We want to find the gradient
we look at the total var $\delta\mathcal{L}$ with respect to $\delta\theta$ (perturba.)

b) First we look at $\int_{t_0}^{t_f} \lambda(t)^T \frac{dN}{dt}\, dt$ Int per parts :

So that we can set t_0 $\lambda(t)$ or dt, Int per parts:

$$\int_{t_0}^{t_f} \lambda(t)^T \frac{dN}{dt}\, dt = [\lambda(t)^T N(t)]_{t_0}^{t_f} - \int_{t_0}^{t_f} \frac{d\lambda^T(t)}{dt} N(t)\, dt$$

So:

$$\mathcal{L} = \int_{t_0}^{t_f} \left(L(t, N, \theta) + \lambda^T f(t, N, \theta) + \frac{d\lambda^T(t)}{dt} N(t) \right) dt - \lambda(t_f)^T N(t_f)$$

$$+ \lambda(t_0)^T N(t_0)$$

If we pert $\delta\theta$ the state N changes by δN $\Rightarrow$ grad dep only on θ

$$\delta_N \mathcal{L} = \underbrace{\int_{t_0}^{t_f} \left(\frac{\partial L}{\partial N} + \lambda^T \frac{\partial f(t,N,\theta)}{\partial N} + \frac{d\lambda^T(t)}{dt} \right) \delta N\, dt}_{\text{we set } = 0} - \lambda^T(t_f)\, \delta N(t_f) + \lambda(t_0)^T \delta N(t_0) = 0$$

$$\Rightarrow \frac{d\lambda^T(t)}{dt} = -\lambda^T(t) \frac{\partial f}{\partial N} - \frac{\partial L}{\partial N}$$

$$\Rightarrow \frac{d\lambda}{dt} = -\left(\frac{\partial f}{\partial N}\right)^T \lambda - \left(\frac{\partial L}{\partial N}\right)^T$$

we set $\delta N(t_f) = 0$ (need to resolve EDO) ($N(t_0)$ is fixed)

we set transversality cond: $\lambda(t_f) = 0$

So:

$$\frac{dJ}{d\theta} = \frac{\partial \mathcal{L}}{\partial \theta} = \int_{t_0}^{t_f} \left(\frac{\partial L}{\partial \theta} + \lambda(t)^T \frac{\partial f}{\partial \theta} \right) dt$$

$$= \int_{t_0}^{t_f} \left(\frac{\partial L}{\partial \theta} + \lambda(t)^T \frac{\partial NN_\theta(N)}{\partial \theta} \right) dt$$

Regularity? For the pb to be well-posed (Hadamard conditions) we need a guarantee of a unique local sol

- $rN \ln\left(\frac{K}{N}\right) \in \mathcal{C}^1(]0, K[)$
- $NN_\theta(N)$ if we use ReLu ($\max(0, x)$) $\in \mathcal{C}^0$ not $\mathcal{C}^1$

  $\hookrightarrow$ non-Lip. and we need our Jacobian

  $\Rightarrow$ we use tanh $\in \mathcal{C}^\infty$ and $\exists$ Jacobian.

  $\Rightarrow$ Adjoint Method!

- And we need to contain NN for simulation of pharma/bio dynamics => control our data (train only for r and K
  => freeze r et K / train NN θ

- $\frac{\partial Loss}{\partial \theta}$? => Use RG + backprop but $O(T)$ huge Time
  ↳ We use our Adjoint ODE + lib. $O(1)$
    ↳ need to explain the lib on report.