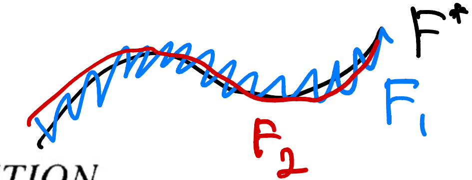
Lecture 2

UAT on Sobolev Spaces

Neural Networks, Vol. 3, pp. 551–560, 1990
Printed in the USA.

0893-6080/90 \$3.00 + .00



ORIGINAL CONTRIBUTION

Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks

KURT HORNIK, MAXWELL STINCHCOMBE, AND HALBERT WHITE

University of California, San Diego

(*Received* 11 *August 1989; revised and accepted* 31 *January* 1990)

$G = \sigma'$

$\sum_{|\alpha|<m} \|D^\alpha f\|_{L^1(\mathbb{R},\lambda)} < \infty$

THEOREM 3.1. *Let $G \neq 0$ belong to $S_1^m(\mathbb{R}, \lambda)$ for some integer $m \geq 0$. Then $\Sigma(G)$ is m-uniformly dense on compacta in $C_\downarrow^\infty(\mathbb{R}^r)$.* □

Ordinary Differential Equations (ODE)

- An ODE:

$$\dot{x}(t) = f(x(t)) \qquad x(0) = x_0 \in \mathbb{R}^d.$$

  - $\dot{x} = \frac{dx}{dt}$ time derivative.
  - $f : \mathbb{R}^d \to \mathbb{R}^d$ vector field.
  - x_0 is the initial condition
  - f independent of t $\Rightarrow$ time-homogeneous

- Time-inhomogeneous ODE

$$\dot{x}(t) = f(t, x(t)) \qquad x(0) = x_0 \in \mathbb{R}^d.$$

$$f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d.$$

- Transformation to time-homogeneous case.

Define $\dot{x}^0(t) = 1$, $x^0(0) = 0 \Rightarrow x^0(t) = t$

$$\begin{pmatrix} \dot{x}^0(t) \\ \dot{x}(t) \end{pmatrix} = \begin{pmatrix} 1 \\ f(x^0(t), x(t)) \end{pmatrix}, \quad \begin{pmatrix} x^0(0) \\ x(0) \end{pmatrix} = \begin{pmatrix} 0 \\ x_0 \end{pmatrix}$$

$$\underbrace{\dot{\tilde{x}}(t)} = \underbrace{\tilde{f}(\tilde{x}(t))} \qquad \underbrace{\tilde{x}(0)} = \underbrace{\tilde{x}_0}.$$

Examples

$f(t,x) = ax \qquad (d=1)$

$\dot{x}(t) = a\,x(t) \Rightarrow x(t) = e^{at} x_0$

$f(t,x) = Ax \qquad (x \in \mathbb{R}^d)$

$\dot{x}(t) = Ax(t) \Rightarrow x(t) = e^{At} x_0.$

$e^{At} = \sum_i \frac{(At)^i}{i!}$

Integral form.

$\int_0^t ds$

$\dot{x}(t) = f(t, x(t)), \qquad x(0) = x_0 \quad \to t \mapsto x(t)$ differentiable

$x(t) = x_0 + \int_0^t f(s, x(s))\, ds. \quad \to t \mapsto x(t)$ is absolutely continuous.

$x(t) = x(0) + \int_0^t g(s)\, ds$

Theorem (Picard-Lindelöf)

Let f be continuous in t and uniformly Lipschitz in x, i.e.
$\exists\, C > 0$ s.t.

$$\|f(t,x) - f(t,x')\| \le C\|x-x'\| \quad \forall x, x' \in \mathbb{R}^d, \ \forall t \in [0,T]$$

Then $\exists$ a unique solution to

$$x(t) = x(0) + \int_0^t f(s, x(s))\, ds.$$

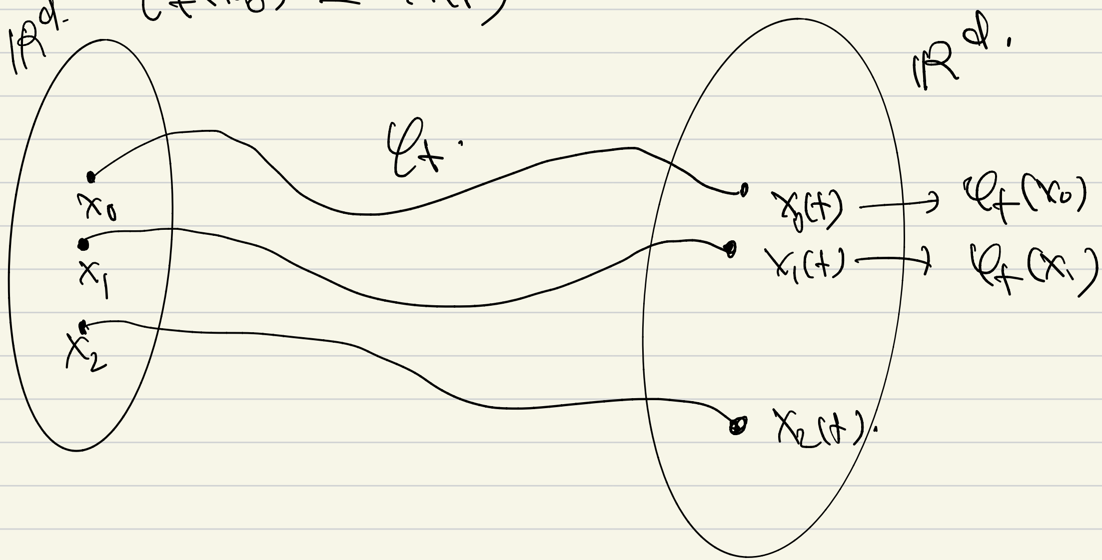
$$\begin{cases} \dot{x} = x^2. \\ \dot{x} = x^{1/3}. \end{cases} \qquad x(0) = 0.$$

Flow Map

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0. \quad \text{(time - homogeneous)}$$

Define the flow maps $\{\varphi_t : \mathbb{R}^d \to \mathbb{R}^d \mid t \in [0, T]\}$.

$$\varphi_t(x_0) = x(t)$$



Example

(1) $\dot{x}(t) = a x(t) \Rightarrow x(t) = e^{at} x(0)$.

$$\varphi_t(x) = e^{at} x.$$

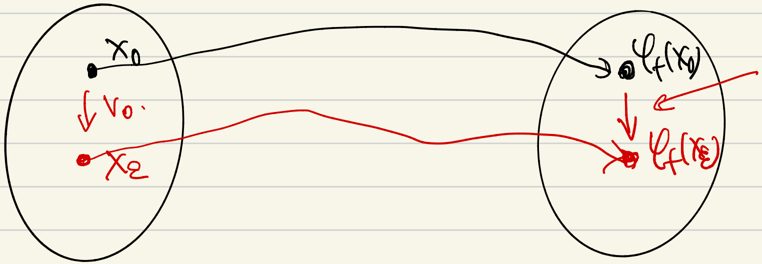
(2) $\dot{x}(t) = -x(t)^2 \qquad x(0) = x_0$.

$$-\frac{1}{x(t)^2}\dot{x}(t) = 1 \Rightarrow \frac{1}{x(t)} = t + C.$$

$$x(t) = \frac{1}{t + C} = \frac{1}{t + 1/x_0} = \frac{x_0}{1 + x_0 t}.$$

$$\varphi_t(x) = \frac{x}{1 + x t}.$$

Dependence on Initial Condition



$$\frac{\varphi_t(x_0+\varepsilon v_0)-\varphi_t(x_0)}{\varepsilon} \to \nabla \varphi_t(x_0)\cdot v_0 .$$

$$= v(t)$$

Theorem

Let f be C^1 and Lipschitz in x uniformly in t.

Let

$$\dot{x}(t) = f(t, x(t)), \quad x(0) = x_0$$

and $\{v(t)\}$ solve.

$$\dot{v}(t) = \nabla_x f(t, x(t))\, v(t), \quad v(0) = v_0 \in \mathbb{R}^d.$$

(Variational equation)

Then

$$\lim_{\varepsilon \to 0^+} \left\| \frac{\varphi_t(x_0+\varepsilon v_0) - \varphi_t(x_0)}{\varepsilon} - v(t) \right\| = 0.$$

uniformly in $t \in [0, T]$ and $\|v_0\| \le 1$

Corollary

$J(t) := \nabla_x \varphi_t(x_0)$ $\mathbb{R}^{d\times d}$. (fixed x_0)

Then $\dot{J}(t) = \nabla_x f(t, x(t))\, J(t)$ $\quad J(0) = I$.

"Proof"

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

$=$

$\frac{d}{dt}(\varphi_t(x_0)) = f(t, \varphi_t(x_0))$

∇x_0.

$\frac{d}{dt}(J(t)) = \nabla_x f(t, \varphi_t(x_0)) \cdot \nabla_{x_0} \varphi_t(x_0)$

$\Rightarrow \quad \dot{J}(t) = \nabla_x f(t, x(t))\, J(t)$. $\quad J(0) = I$

$\because \varphi_0(x_0) = x_0$

Example

① $\dot{x} = Ax \Rightarrow \varphi_t(x_0) = e^{At} x_0$

$f(t,x) = Ax$ $\quad \nabla_x \varphi_t(x_0) = e^{At}$ (indep of x_0)

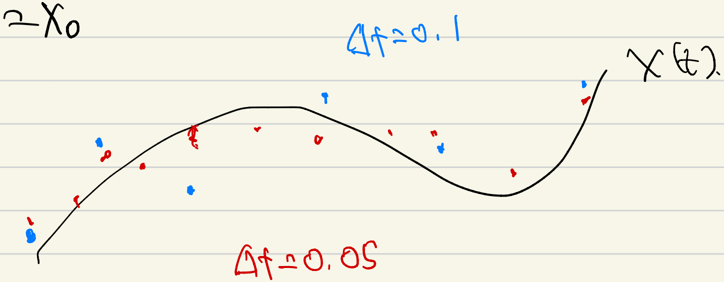
$\dot{J} = AJ \quad , \quad J(0) = I \Rightarrow J(t) = e^{At}.$

② $\dot{x} = -x^2 \qquad \varphi_t(x) = \dfrac{x}{1+tx}.$

Numerical Solution of ODEs

$\dot{x}(t) = f(t, x(t))$, $x(0) = x_0$

$x(t) \approx \hat{x}(k)$



Forward Euler Method.

$$\frac{\hat{x}(k+1) - \hat{x}(k)}{\Delta t} = f(k\Delta t, \hat{x}(k)) \qquad \hat{x}(0) = x_0.$$

$$\Rightarrow \quad \hat{x}(k+1) = \hat{x}(k) + \Delta t\, f(k\Delta t, \hat{x}(k))$$

Convergence

$$\max_{k \le K} \| \hat{x}(k) - x(k\Delta t) \| \longrightarrow 0 \quad \text{as} \quad \Delta t \to 0.$$

$$\le C\Delta t.$$

$K = \lfloor T/\Delta t \rfloor$

Calculus of Variations and Optimal Control

Finite dimensional optimization

$$\min_{x} R(x) \qquad R: \mathbb{R}^d \to \mathbb{R}.$$

Infinite dimensional optimization

$$\min_{\underline{x}} J[\underline{x}] \qquad J: \mathcal{X} \to \mathbb{R} \quad \text{(functional)}.$$

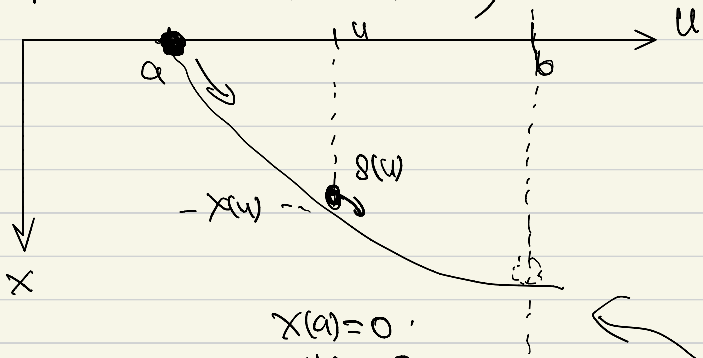
$$\underline{x} = \{ x(u) : a \le u \le b \} \qquad \mathcal{X} = C([a,b], \mathbb{R})$$

$x'(u) = \frac{dx}{du}$.

$$J[\underline{x}] = \int_a^b L(u, x(u), x'(u))\, du$$

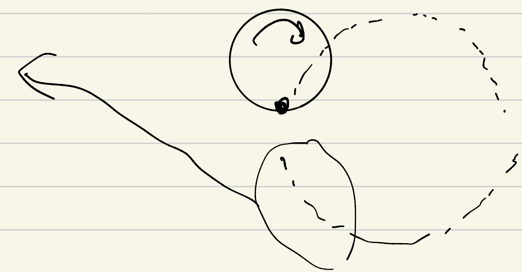
Lagrangian.

Example (Bradnistochrone)



$x(a) = 0$
$x(b) = ?$

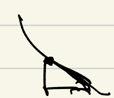
Cycloid.



Define speed at u : $s(u)$
height at u : $-x(u)$

Energy conservation : $\frac{1}{2} m s(u)^2 = m g x(u)$

$$\Rightarrow s(u) = \sqrt{2 g x(u)}$$



arc length $= \sqrt{1 + x'(u)^2}$

time $= \displaystyle\int_a^b \frac{\sqrt{1 + x'(u)^2}}{\sqrt{2 g x(u)}}\, du.$

$L(u, x(u), x'(u))$

Euler-Lagrange Equation

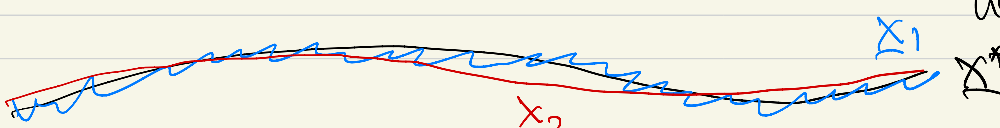
Let $\underline{x} \in C^1([a,b],\mathbb{R})$ be an extremum (weak) of J. Then

$$\partial_x L(u, x(u), x'(u)) = \frac{d}{du}\left(\partial_{x'} L(u, x(u), x'(u))\right)$$

Weak vs Strong Minimum.

Local minimum : $\exists\ \delta > 0$ s.t. $J[\underline{x}^*] \leq J[\underline{x}]$ for all $\underbrace{\|\underline{x} - \underline{x}^*\|}_{\text{what norm?}} \leq \delta$.

- Strong minimum if $\|\underline{x}^* - \underline{x}\|_0 = \sup_{u \in [a,b]} |x^*(u) - x(u)|$
- Weak minimum if $\|\underline{x}^* - \underline{x}\|_1 = \|\underline{x}^* - \underline{x}\|_0 + \sup_{u \in [a,b]} |x^{*\prime}(u) - x'(u)|$



weak min $\Rightarrow$ $J[\underline{x}^*] \leq J[\underline{x}_2]$

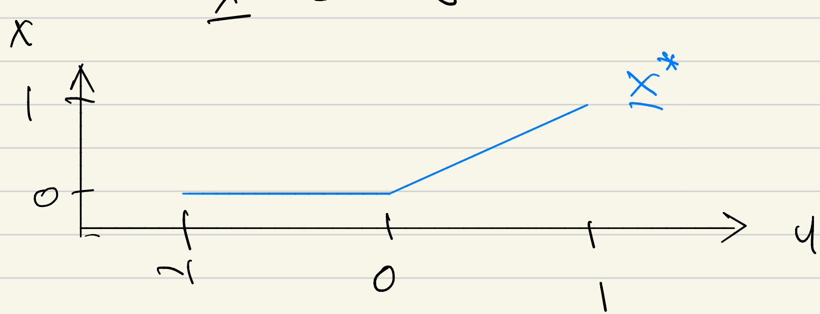
Strong min $\Rightarrow$ $J[\underline{x}^*] \leq J[\underline{x}_1], J[\underline{x}_2]$

<u>Example</u>

$$J[\underline{x}] = \int_{-1}^{1} (x(u))^2 (x'(u) - 1)^2 \, du.$$

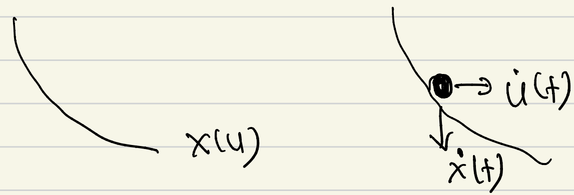
$$x(-1) = 0, \qquad x(1) = 1.$$

$$\min_{\underline{x}} J[\underline{x}].$$



$J[\underline{x}^*] = 0.$

Example (Brachistochrone continued)



$X(u) \longrightarrow x(t),\ u(t)$

Speed : $s(u)(t) = \sqrt{\dot{x}(t)^2 + \dot{u}(t)^2}$

$\frac{1}{2} m s(u)^2 = m g x(u) \Rightarrow 2g x(t) = \dot{x}(t)^2 + \dot{u}(t)^2.$

Define controls

$$\theta_1(t) = \frac{\dot{u}(t)}{\sqrt{2g x(t)}} \qquad \theta_2(t) = \frac{\dot{x}(t)}{\sqrt{2g x(t)}}$$

$$\theta_1(t)^2 + \theta_2(t)^2 = 1$$

$$\Rightarrow \begin{cases} \dot{u}(t) = \theta_1(t)\sqrt{2g x(t)} \\ \dot{x}(t) = \theta_2(t)\sqrt{2g x(t)} \end{cases} \qquad \theta_1(t)^2 + \theta_2(t)^2 = 1$$

$$(u(t_0), x(t_0)) = (a, 0) \qquad (u(t_1), x(t_1)) = (b, ?)$$

$$J[\underline{\theta}_1, \underline{\theta}_2] = \int_{t_0}^{t_1} dt = t_1 - t_0 .$$

Why generalization ?

Optimal Control Problem

Dynamics : $\dot{x}(t) = f(t, \underset{\text{state}}{x(t)}, \underset{\text{control}}{\theta(t)})$, $t \in [t_0, t_1]$, $x(t_0) = x_0$

for each t, $\theta(t) \in \Theta \subset \mathbb{R}^p$, Θ is closed.

- $f(t, x, \theta)$ is continuous in t, θ, $\forall x$
- $f(t, x, \theta)$ is continuously differentiable in x $\forall t, t, \theta$.

Not assumed

- f is differentiable wrt θ.
- $t \mapsto \theta(t)$ is regular.

Cost functional

$$S[\theta] = \underbrace{\int_{t_0}^{t_1} L(t, x(t), \theta(t))\, dt}_{\text{running cost}} + \underbrace{\Phi(x(t_1))}_{\text{terminal cost}}$$

$L : [t_0, t_1] \times \mathbb{R}^d \times \Theta \to \mathbb{R}$, $\Phi : \mathbb{R}^d \to \mathbb{R}$.

Bolza Problem

$$\inf_{\underline{\theta}} \; J[\underline{\theta}] \quad \text{subject to} \quad \dot{x}(t) = f(t, x(t), \theta(t))$$

$$x(0) = x_0 \;, \quad \theta(t) \in \boxed{H}$$

- Mayer Problem $\quad L \equiv 0$
- Lagrange Problem $\quad \Phi \equiv 0$.

Bolza $\rightarrow$ Mayer.

define $\quad \dot{x}^0(t) = L(t, x(t), \theta(t)) \;, \quad x^0(t_0) = 0$.

$$\Rightarrow \quad x^0(t) = \int_{t_0}^{t_1} L(t, x(t), \theta(t))\, dt$$

$$\tilde{x} = (x^0, x) \qquad \tilde{\Phi}(\tilde{x}) = \Phi(x) + x^0.$$

$$\tilde{f} = (L, f)$$

$$\min_{\underline{x}} \int_a^b L(u, x(u), x'(u))\, du.$$

$$\downarrow\ u \to t.$$

$$\min_{\underline{x}} \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))\, dt$$

$$\downarrow\ \dot{x}(t) = \theta(t).$$

$$\boxed{\begin{gathered}\min_{\underline{\theta}} \int_{t_0}^{t_1} L(t, x(t), \theta(t))\, dt. \\ \dot{x}(t) = \theta(t)\end{gathered}}$$

Lagrange problem