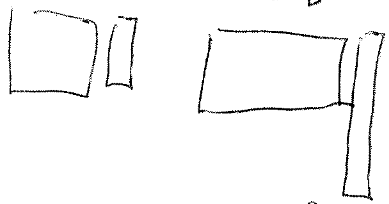
① 3/22/2023

Learning a Linear System

Suppose we want to minimize

Loss: $\mathcal{L}(\underset{\mathbb{R}^n}{x}, \underset{\mathbb{R}^q}{p}) = \frac{1}{2}\|x - \underset{\mathbb{R}^n}{y}\|^2 + \frac{1}{2}\|p\|^2$ for a known y

s.t. $f(x,p) = \underset{n\times n}{A}\,\underset{n}{x} + \underset{n\times q}{B}\,\underset{q}{p} = \underset{n}{0_n}$



lots of parameters

or

a few parameters

e.g. We have measurements $y_1, \dots, y_n$ that approximate $x_1, \dots, x_n$

or perhaps

$$\mathcal{L}(x,p) = y^T x + v^T p \quad \text{etc.}$$

Basic Loop

Start at x, p with $f(x,p) = 0$

① update p (e.g. follow $-\nabla_p \mathcal{L}$)

② compute $f(x,p)$ (not 0 now)

③ update x (f 0 again)

~~Repeat~~

④ compute $\mathcal{L}$

Repeat

(2)

Write as a graph

$$
\begin{tikzcd}
p \arrow[r, "df = B\,dp"] \arrow[rrr, bend right, "\dots + p^T dp"'] & f \arrow[r, "dx = -A^{-1} df"] & x \arrow[r, "d\mathcal{L} = (x-y)^T dx + \dots"] & \mathcal{L}
\end{tikzcd}
$$

Why?
Solve $Ax = -B\,dp$
$= -df$

$$\frac{d\mathcal{L}}{dp} = (x-y)^T \underbrace{(-A^{-1})B}_{\text{Matrix}\backslash\text{matrix}} + p^T \qquad \text{forward mode}$$

$$\nabla_p \mathcal{L} = -B^T \underbrace{(A^{-T})(x-y)}_{\text{Matrix}\backslash\text{vector}} + p \qquad \text{reverse mode}$$

You've seen this before: with one output
reverse is better because vector(matrix)(Matrix)

Generalize to non-linear problems

$$\text{Min } \mathcal{L}(\underset{n}{\underset{\uparrow}{x}}, \underset{q}{\underset{\uparrow}{p}}) \quad \text{s.t.} \quad f(\underset{n}{\underset{\uparrow}{x}}, \underset{q}{\underset{\uparrow}{p}}) = \underset{n}{\underset{\uparrow}{0}}$$

last example

For ease of notation write $f_x = \frac{\partial f}{\partial x}$ $n \times n$ matrix (was A)

$f_p = \partial f / \partial p$ $n \times q$ matrix (was B)

$\mathcal{L}_x = \frac{\partial \mathcal{L}}{\partial x}$ $1 \times n$ matrix (was $(x-q)^T$)

$\mathcal{L}_p = \frac{\partial \mathcal{L}}{\partial p}$ $1 \times p$ matrix (was p^T)

$$\begin{tikzcd} \mathring{p} \arrow[r, "f_p"] \arrow[rrr, bend right, "\mathcal{L}_p"'] & \mathring{f} \arrow[r, "-f_x^{-1}"] & \mathring{x} \arrow[r, "\mathcal{L}_x"] & \mathring{\mathcal{L}} \end{tikzcd}$$

$$\frac{d\mathcal{L}}{dp} = \mathcal{L}_x(-f_x^{-1})f_p + \mathcal{L}_p$$

$$\nabla_p \mathcal{L} = -f_p^T \underbrace{f_x^{-T} \mathcal{L}_x^T}_{\lambda^T} + \mathcal{L}_p^T$$

Solve $f_x^T \lambda^T = \mathcal{L}_x^T$ λ is $1 \times n$

$$n\underset{n}{\square}\ \underset{n}{\boxed{\lambda^T}} = \underset{n}{\boxed{\ }}$$

$$\cancel{\nabla_p \mathcal{L}}\ \frac{d\mathcal{L}}{dp} = \cancel{\ } \underset{1\times n}{-\lambda} \underset{n \times q}{f_p} + \underset{1 \times p}{\mathcal{L}_p}$$

$$\boxed{\nabla_p \mathcal{L} = \mathcal{L}_p^T - f_p^T \lambda^T}$$

Learning a differential equation

A vector x is a "discrete function" as much as a memory address

$i \to x[i]$ for $i = 1, \ldots, n$

A cont vector value function looks like

$t \to u(t)$ on $[t_0, T]$ $\quad u(t) \in \mathbb{R}^n$

Instead of

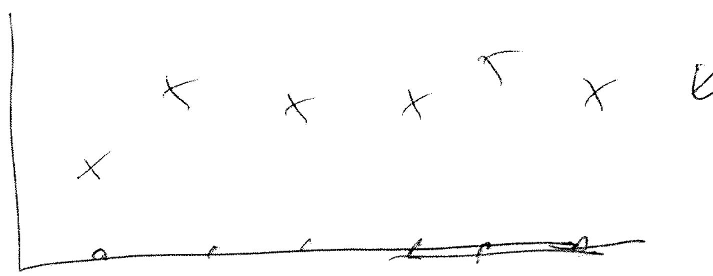
$$\mathcal{L}(x, p) = \frac{1}{2}\|x - y\|^2 + \frac{1}{2}\|p\|^2$$

we can have

$$\mathcal{L}(u, p) = \frac{1}{2}\int_{t_0}^{T} (u(t) - y(t))^2 \, dt + \frac{1}{2}\|p\|^2$$

or even

$$\sum (u(t_i) - y_i)^2$$



you know these values

what is p?

Need the analogs

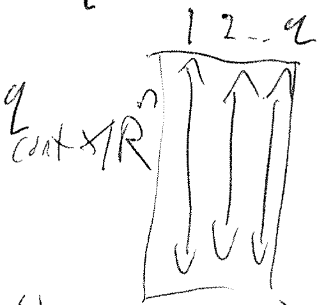
$$\begin{tikzcd} p \arrow[r, "F_p"] \arrow[rrr, bend right, "\mathcal{L}_p"'] & F \arrow[r, "-F_u^{-1}"] & u \arrow[r, "\mathcal{L}_u"] & \mathcal{L} \end{tikzcd}$$

$$F(u, p) = t \longrightarrow u'(t) - f(u, p, t)$$

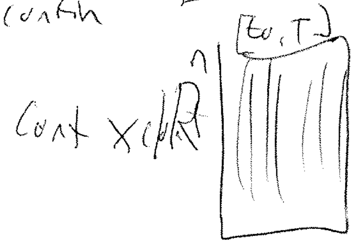
$$[t_0, T] \rightarrow \mathbb{R}$$

$p \in \mathbb{R}^q$

F_p : continuous $\times$ q



F_p : continu × continu



To be more general

let $\mathcal{L}(u,p) = \int_{t_0}^{T} g(u,p)\,dt$

$$= \int_{t_0}^{T} g(u(t),p)\,dt$$

scalar function of $p \in \mathbb{R}^q$

$\mathcal{L}_p = \frac{\partial \mathcal{L}}{\partial p} \to$ row vector

$$\mathcal{L}(u, p+\varepsilon e_i) - \mathcal{L}(u,p) \approx \varepsilon (\mathcal{L}_p)_i$$

and $\mathcal{L}_p(u,p) = \int_{t_0}^{T} g_p(u,p)\,dt$

differentiate inside the integral

$\mathcal{L}_u$: If u were just a scalar
$\mathcal{L}_u$ would be $1 \times n$
so now L_u is $1 \times$ (continuous $\times n$)

that is given t & i we can obtain

$$\mathcal{L}_u : (t_0, i) \longrightarrow \text{scalar}$$

$$\mathcal{L}(u + \varepsilon\delta(t-t_0)e^i, p) - \mathcal{L}(u,p)$$

What is an inner product on functions $u(t) : [t_0,T] \times \{1:n\} \to \mathbb{R}$

Well if $\langle u,v \rangle$ on $\mathbb{R}^n$ is $\sum u_i v_i$

& $\langle f,g \rangle$ on (t_0,T) is $\int f(t)g(t)\,dt$

$$\langle u(t), v(t) \rangle = \int_{t_0}^{T} u(t)^T v(t)\,dt$$

combines the best of both worlds

$$F(u,p) = t \to u'(t) - f(u,p,t)$$

to find an adjoint means to find an operator on $\lambda(t)$ s.t

$$\langle \lambda(t), \tfrac{d}{dp}(u'(t) - f(u,p,t)) \rangle = \langle (\ )^T \lambda(t), u(t) \rangle$$

$$\int \lambda(t)^T \tfrac{d}{dp}\left(u'(t) - f(u,p,t)\right) dt$$

Suppose $u_p(t_0) = 0$
$\lambda(T) = 0$

Let $s = u_p$

$s' = u'_p$ (prime $= \frac{d}{dt}$)

8

$$\int_{t_0}^{T} \lambda^T (s' - f_u s - f_p)\,dt$$

$$= \lambda^T s\Big]_{t_0}^{T} - \int_{t_0}^{T} (\lambda^T)' s\,dt$$

$$- \int_{t_0}^{T} (\lambda^T)(f_u s - f_p)\,dt$$

vanish

$\lambda(T) = 0$

$s(t_0) = 0$

$$\lambda^T = -\frac{df}{du}^T \lambda + \left(\frac{dy}{du}\right)^T$$

The adjoint is a difeq!