Lecture 4 - 2/5

- Polynomial Reproduction
- Nonlinear stability
- Constrained Optimization

Our main hammer will be constrained opt

$$\min_{\theta} F(x;\theta)$$

$$\text{s.t } G(x;\theta)$$

Typical Cases

- F is a reconstruction, G are physics
  "Physics-Constrained / PDE-constrained opt"

- F is arbitrary, G is property we actually care about / "wish-list"

  "Feasibility problem"

Equality constrained quadratic programming

For non-linear Replace w/ Hessian + Jacobian

Consider
$$\min_{x} \quad \frac{1}{2} x^T A x + b^T x + c$$
$$\text{s.t} \quad Bx = c$$

To solve, add Lagrange multiplier λ

$$\min_{x,\lambda} \quad \underbrace{\frac{1}{2} x^T A x + b^T x + c + \lambda^T (Bx - c)}_{L(x,\lambda)}$$

The Karush Kuhn Tucker (KKT) conditions state that the minimizer satisfies

$$\partial_x L = \partial_\lambda L = 0$$

$$\partial_\lambda L = \quad Bx - C \quad = 0$$
$$\partial_x L = \quad Ax + b + B^T \lambda \quad = 0$$

Yielding the saddle point problem

$$\begin{array}{c} ① \\ ② \end{array} \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -b \\ c \end{pmatrix}$$

Mult top by BA^{-1}

$$B\cancel{A^{-1}A}\,x + BA^{-1}B^T\lambda = -BA^{-1}b$$

Subtract from ②

$$-BA^{-1}B^T\lambda = c + BA^{-1}b$$

Let $S = BA^{-1}B^T$ be defined as the <u>Schur complement</u>

$$\lambda = -S^{-1}(c + BA^{-1}b)$$

$$x = A^{-1}(-B^T\lambda - b)$$

We will use this repeatedly as a tool to enforce properties we'd like as equality constraints

A note on training

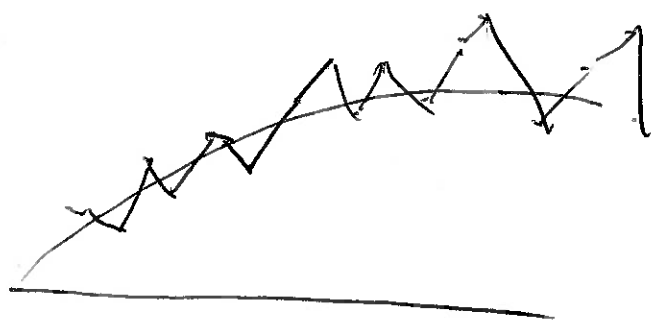
- In the HW we consider matching derivatives (this is often called force matching)

$\hat{u} \to$ target

$$\mathcal{L} = \left\| \frac{\hat{u}_j^{n+1} - \hat{u}_j^{n}}{k} + \frac{\sum_{k=-\alpha}^{\beta} S_{j+k}\, u^n_{j+k}}{h^{|\alpha|}} \right\|^2$$

to approximate an operator D^α

- This is prone to over fitting for noisy realistic data



- Some strategies to mitigate
  - explicitly model noise as MLE
  - Tikhinov / weight decay
  - filtering / pre processing

  ← we'll talk about these later

- Better to fit to the solution itself rather than its derivative

② Reduced space constrained optimization

Sub res. back into $\mathcal{L}_e$ to optimize directly over the space of FD solns.

$$u_j^n = Q(u_j^{n-1})\, u_j^{n-1}$$ is the (potentially nonlinear) update operator

$$\mathcal{L} = \frac{1}{2} \sum_{n,j} \left| Q(u_j^{n-1}) \circ \dots \circ Q(u_j^0)\, u_j^0 - \hat{u}_j^n \right|^2$$

- For nonlinear / explicit schemes, need to backprop through a linear solve, or Newton / iterative nonlinear solve

③ Lagrange Multipliers

$$\mathcal{L} = \mathcal{L}_R + \lambda^T r, \quad \lambda \in \mathbb{R}^n$$ ← Lagrange mult

$$\min_{u^n, r^n, \theta} \frac{1}{2} \sum_{n,j} |u_j^n - \hat{u}_j^n|^2 + \sum_n \lambda_n (r_{n,j})$$

Process For 1, ... # iterations

1. $\delta_\lambda \mathcal{L} = 0 \Rightarrow$ solve forward problem

Define the reconstruction loss and physics residual

$$\mathcal{L}_R = \frac{1}{2} \sum_{n,j} | u_j^n - \hat{u}_j^n |^2$$

$D_\theta^\alpha \leftarrow$ parameterized stencil

$$r_{n,j} = u_j^{n+1} - u_j^n - k D_\theta^\alpha u_j^n$$

Pose the constrained fit to the solution

$$\min_{u_j^n, \theta} \mathcal{L}_R$$

$$\text{s.t} \quad r_{n,j} = 0 \quad \text{for all } n,j$$

3 ways to do this

① Penalty (worst!)

$$\min_{u_j^n, \theta} \mathcal{L}_R + \lambda \sum_{n,j} r_{n,j}^2$$

- Take a scalar penalty param λ. For big λ, $r_{n,j}$ is small.
- No way to pick λ a priori (Too big unstable, too small big error)
- Small residuals translate to large soln errors (think of the $O(k^2)$ errors we saw in explicit Euler)

2. $\delta_{u^n} \mathcal{L} = 0$ solve adjoint problem

$$\left(J_{u}(r)\right)^T \lambda = -\left(u_j^n - \hat{u}_j^n\right)$$

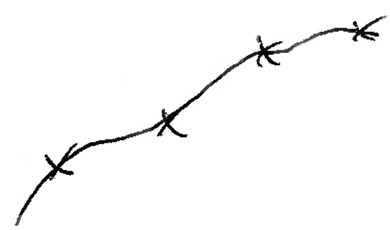
3. Update θ w/ gradient opt. using current guess for u, λ

- Backprop doesnt need to go through fwd + adj solve, best for nonlinear + implicit

For linear problems 2 + 3 are the same

Mini batching

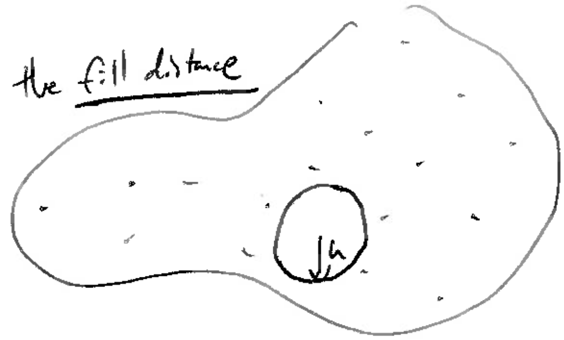
- These are <u>big</u> gradient calculations
- Typically our dataset consists of <u>many</u> solns
- <u>Stochastic gradient descent</u>
    - Pick random soln $u_{j,d}^n$
    - Pick subset of time series $u_{j,d}^n, \ldots, u_{j,d}^{n+m}$
    - Train using $\hat{u}_{j,d}^n$ as IC



Polynomial Reproduction and least squares

Consider points $X = \{x_1, \dots x_N\} \subseteq \Omega \subseteq \mathbb{R}^d$

For scattered data characterize the <u>fill distance</u>



$$h_{X,\Omega} = \sup_{x \in \Omega} \min_{1 \le j \le N} \|x - x_j\|_2$$

Biggest ball in Ω w/o data in it

<u>Separation Distance</u>

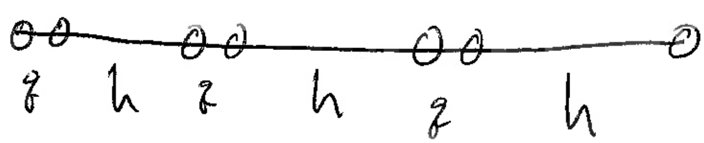
$$q_X := \frac{1}{2} \min_{i \neq j} \|x_i - x_j\|_2$$

<u>def</u> Data sites are quasi-uniform (w/ constant c_{qu})

if $$q_X \le h_{X,\Omega} \le c_{qu} q_X$$

For our FD stencils so far

$$h_{X,\Omega} = h, \quad q_X = h \Rightarrow c_{qu} = 1$$



Introduce moving least squares, generalized FDM

For $x \in \Omega$ $\quad S_{f,X}(x) = p^*(x)$

$$\min_{p} \sum_{i=1}^{N} \left[f(x_i) - p(x_i) \right]^2 w(x, x_i) : p \in \Pi_m(\mathbb{R}^d)$$

We'll assume $w(x,y) = \Phi_\delta(x-y)$

As a simple example where $m=0$

$$\min_{c(x)} \sum_{i=1}^{N} \frac{1}{2}\left(f(x_i) - c(x)\right)^2 \Phi_\delta(x-x_i)$$

Taking derivative and setting to zero

$$0 = \cancel{\frac{2}{2}} \sum_{i=1}^{N} \left(f(x_i) - c(x)\right) \Phi_\delta(x-x_i)$$

$$c(x) = \frac{\sum_i \Phi_\delta(x-x_i) f(x_i)}{\sum_i \Phi_\delta(x-x_i)}$$

and we recover Kernel density estimation

aka the Shepard Interpolant

Thm Solving the moving least squares problem is equivalent to solving

$$s_{f,\bar{X}}(x) = \sum_i a_i^*(x) f(x_i)$$

$$a_i^* = \operatorname{argmin} \frac{1}{2} \sum_i a_i(x)^2 \frac{1}{\Phi_\delta(x-x_i)}$$

(DP) dual problem $\quad$ $$\text{s.t.} \quad \sum a_j(x) p(x_j) = p(x)$$

Rmk We can thus search for a minimal norm stencil that has reproduction properties

Pf Set some notation $\to$ lin alg heavy

N - # nodes
$M = \dim(\Pi_m(\mathbb{R}^d))$

$a \in \mathbb{R}^N$

$W = \operatorname{diag}(\Phi_\delta(x-x_i)) \in \mathbb{R}^{N \times N}$

$P(x) \in \mathbb{R}^M$

$P(x_i) \in \mathbb{R}^{M \times N}$ $\leftarrow$ will write as P and distinguish against $P(x)$

$u := u(x_i) \in \mathbb{R}^N$

$c \in \mathbb{R}^M$

Pf Primal Problem

$$\min_{c \in \mathbb{R}^m} \underbrace{\frac{1}{2}(u - cP)^T W (u - cP)}_{L}$$

Expanding

$$L = \frac{1}{2} c^T P W P^T c - u W P^T c + \underset{\text{stuff indep of } c}{K}$$

Recall

$$\frac{\partial}{\partial x} \frac{1}{2} x^T A x = \frac{1}{2}(A + A^T) x$$

$$\frac{\partial}{\partial x} y^T x = y$$

$$\frac{\partial L}{\partial c} = P W P^T c - W P^T u = 0$$

$$\Rightarrow c = (P W P^T)^{-1} W P^T u$$

$$S_{f,X}(x) = c \cdot P(x)$$

$$= \boxed{P(x)^T (P W P^T)^{-1} W P^T u}$$

Dual Problem

$$S(x) = a(x) \cdot u(x_i)$$

$$\min_{a \in \mathbb{R}^N} \frac{1}{2} a(x)^T W^{-1} a(x)$$

$$\text{s.t.} \quad a(x)^T P = P(x)$$

Applying Schur complement formula from before

$$\text{KKT} \rightarrow \quad \begin{pmatrix} W^{-1} & P^T \\ P & 0 \end{pmatrix} \begin{pmatrix} a(x) \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ P(x) \end{pmatrix}$$

$$S = PWP^T$$

$$\lambda = -S^{-1} P(x)$$

$$a(x) = -WP^T \lambda$$

$$= WP^T S^{-1} P(x)$$

$$S(x) = u^T a(x)$$

$$= u^T W P^T S^{-1} P(x)$$

$$= \boxed{P^T(x) (PWP^T)^{-1} PWu}$$

$\blacksquare$

Corr Replace $a(x)^T P = P(x)$ constraint

with $a(x)^T P = \cancel{D^\alpha P}$

$= D^\alpha P(x)$ constraint

to obtain differential operators / stencils

Q? Is polynomial reproduction set non-empty?

Thm (Wendland "Scattered Data Approx" Thm 4.7)

Suppose $\Omega \subseteq \mathbb{R}^d$ is compact and satisfies an interior cone condition w/ $\theta \in (0, \pi/2)$ and radius r. Then there exists $a^*(x)$ for any x such that

① $\sum_j a_j^*(x)\, p(x_j) = p(x) \quad \forall p \in \Pi_m(\mathbb{R}^d)$

② $\sum_j |a_j^*(x)| \leq \tilde{C}_1$

③ $a_j^*(x) = 0$ if $\|x - x_j\|_2 > \tilde{C}_2 h_{x,\Omega}$

For $\tilde{C}_1$ and $\tilde{C}_2$ indep of $h_{x,\Omega}$ that can be ~~solved for~~ explicitly derived