CHAPTER 1

Introduction to nonlinear approximation

1.1. HISTORICAL ANALOGUES

1.1.1. Operator calculus

1.1.1.1. Heavisde study of telgraphist equation

In late nineteenth century, telegrapher's equations, a system of linear PDEs for current $I(x,t)$ and voltage $V(x,t)$

$$\frac{\partial}{\partial x}V(x,t) = -L\frac{\partial}{\partial t}I(x,t) - RI(x,t)$$

$$\frac{\partial}{\partial x}I(x,t) = -C\frac{\partial}{\partial t}C(x,t) - GV(x,t)$$

Heaviside avoided solution of the PDEs by reduction to an algebraic formulation historical formulation, e.g., for the ODE for $y(t)$

$$\frac{\mathrm{d}y}{\mathrm{d}t} + ay = b$$

Heaviside considered the associated algebraic problem for $Y(s)$

$$sY + aY = b \Rightarrow Y(s) = \frac{b}{a+s} \Rightarrow y(t) = \mathscr{L}^{-1}[Y(s)]$$

1.1.1.2. Development of mathematical theory of operator calculus

> Why should I refuse a good dinner simply because I don't understand the digestive processes involved? (Heaviside, ?)

Heaviside's formal framework (1890's) for solving ODEs was discounted since it lacked mathematical rigour.

- Russian mathematician 1920's established first results (Vladimirov)
- Theory of Distributions (Schwartz, 1950s)

1.2. BASIC APPROXIMATION THEORY

1.2.1. Problem definition

Consider function $f\colon \mathbb{R}^d \to \mathbb{R}$, $d \gg 1$ assumed large, f of unknown form, difficult to compute for general input. Seek $g\colon \mathbb{R}^n \to \mathbb{R}$, $T\colon \mathbb{R}^d \to \mathbb{R}^n$ such that

$$\|f - g \circ T\| < \varepsilon$$

for some $\varepsilon > 0$.

1.2.1.1. Linear approximation example

Choose a basis set (Monomials, Exponentials, Wavelets) $\{\phi_1, \phi_2, \ldots\}$ to approximation of $L^2(\mathbb{R})$ functions in Hibert space

$$g_n(t) = \sum_{j=1}^{n} (f, \phi_j)\phi_j = \sum_{j=1}^{n} c_j \phi_j$$

The approximation is convergent if

$$\lim_{n \to \infty} \|f - g \circ T\| = 0,$$

This assumes $c_j = (f, \phi_j)$ rapidly decrease.

THEOREM. *(Parseval) The Fourier transform is unitary. For $A, B: \mathbb{R} \to \mathbb{C}$, square integrable, 2π-periodic with Fourier series*

$$A(t) = \sum_{n=-\infty}^{\infty} a_n e^{int}, B(t) = \sum_{n=-\infty}^{\infty} b_n e^{int},$$

$$\sum_{n=-\infty}^{\infty} a_n \bar{b}_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(t) \bar{B}(t)\, \mathrm{d}t.$$

Bessel inequality:

$$\sum_{j=1}^{n} |(f, \phi_j)|^2 \leqslant \|f\|_2 .$$

Fourier coefficient decay: for $f \in C^{(k-1)}(\mathbb{R})$, $f^{(k-1)}$ absolutely continuous,

$$|c_n| \leqslant \min_{0 \leqslant j \leqslant k} \frac{\|f^{(j)}\|_1}{|n|^j}.$$

In practice: coefficients decay as

- $1/n$ for functions with discontinuities on a set of Lebesgue measure 0;
- $1/n^2$ for functions with discontinuous first derivative on a set of Lebesgue measure 0;
- $1/n^3$ for functions with discontinuous second derivative on a set of Lebesgue measure 0.

Fourier coefficients for analytic functions decay faster than any monomial power $c_n = o(n^{-p}), \forall p \in \mathbb{N}$, a property known as exponential convergence.

Denote such approximations by $\mathscr{L}$, and they are linear

$$\mathscr{L}(\alpha f + \beta g) = \alpha \mathscr{L}(f) + \beta \mathscr{L}(g)$$

-

1.2.1.2. Non-Linear approximation example

Choose a basis set (Monomials, Exponentials, Wavelets) $\{\phi_1, \phi_2, \ldots\}$ to approximation of $L^2(\mathbb{R})$ functions in Hibert space

$$g_n(t) = \sum_{j=1}^{n} c_j \phi_j$$

Let $\Phi_n = \{\varphi_{k(1)}, \varphi_{k(2)}, \ldots, \varphi_{k(n)}\}$ such

$$(f, \varphi_{k(1)}) \geqslant (f, \varphi_{k(2)}) \geqslant \cdots \geqslant (f, \varphi_{k(n)}).$$

Choose c_j=$(f, \varphi_{k(j)})$, and

$$g_n(t) = \sum_{j=1}^{n} c_j \phi_j .$$

Denote such approximations by $\mathscr{G}$, and they are non-linear.

1.2.2. Nonlinear approximation by composition

Consider function $f\colon \mathbb{R}^d \to \mathbb{R}$, $d \gg 1$ assumed large, f of unknown form, difficult to compute for general input. Seek $g\colon \mathbb{R}^n \to \mathbb{R}$, $T\colon \mathbb{R}^d \to \mathbb{R}^n$ such that

$$\|f - g \circ T\| < \varepsilon$$

for some $\varepsilon > 0$.

What questions do you ask?

Does T exist?. $\forall f, \varepsilon, \exists T$, such that $\|f - g \circ T\| < \varepsilon$

Can arbitrary ε be achieved?.

Can we construct T?.

→ By what procedure?

$$T = T_1 \circ T_2 \circ \ldots \circ T_J$$

with T_i simple modifications of identity (ReLU)

$$\min_{T_1, \ldots T_J} \|f - g \circ T_1 \circ T_2 \circ \ldots \circ T_J\|$$

$$\boldsymbol{T}_j(\boldsymbol{x}) = \eta(\boldsymbol{A}_j \boldsymbol{x} + \boldsymbol{b}_j)$$

$$\eta(t) = \begin{cases} 0 & t < 0 \\ t & t \geqslant 0 \end{cases}$$

→ At what cost?

How big is n?.

.