High Dimensional Approximation - Background and Sources

Wolfgang Dahmen

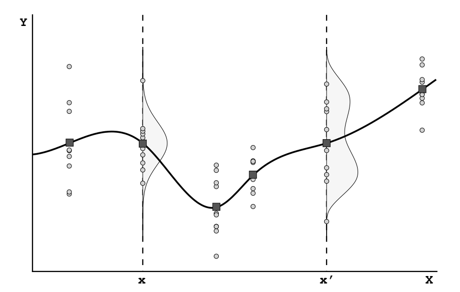
Seminar: USC, High Dimensional Approximation, Feb 13, 2008

Outline

Learning Theory - Regression Problem

ρ unknown measure on $Z := X \times Y$

$$f_\rho(x) := \int_Y y \, d\rho(y|x) = \mathbb{E}(y|x)$$



$$\mathcal{E}(f) := \int_Z (y - f(x))^2 d\rho \quad \rightsquigarrow \mathcal{E}(f) = \mathcal{E}(f_\rho) + \|f - f_\rho\|^2_{L_2(X,\rho_X)}$$

Concepts - Obstructions

Relevant concepts

Nonparametric estimation, concentration inequalities, nonlinear approximation

Solution strategies

- Adaptive partitioning
- Complexity regularization (model selection)

$\dim X$ large - Curse of dimensionality:
Are there ways around it?

Ameliorating the curse of dimensionality

- Dimensionwise decompositions - ANOVA-type schemes
- Kernel methods, neural networks
- Sparse grids, hyperbolic cross approximation
- Kronnecker-product approximation
- Dimension reduction - "learning" embedded manifolds

Recovery schemes:

- Greedy algorithms
- Procedural recovery (sparse occupancy trees, Sprecher's alg.)

A higher level of difficulty:

- Learning on Banach spaces ($\dim X = \infty$)
- Learning implicitly given functions - e.g. solutions of stochastic PDEs

Dynamical System Input

$$\frac{\partial \psi}{\partial t} + D(\psi, x) = P(\psi, x)$$

- ψ 3D prognostic dependent variable, e.g. temperature, pressure, moisture, etc.
- x 3D dependent variable, e.g. latitude, longitude, height,
- D model dynamics, PDE of motion, thermodynamcs, balance laws, etc.
- P model physics, long, short range athmospheric radiation , turbulence, convection, clouds, interactions with land, chemistry, etc. so complicated even as simplified parametrized versions – based on solving deterministic equations

Alternative: Learning

Instead of computing the forcing terms by solving deterministic equations, taking most of the time, one tries to "learn" P from aquired data

Problem:
Given $Z = \{z_i = (x_i, y_i) \in X \times Y \subset \mathbb{R}^{d \times d'} : i = 1, \ldots, N\}$
find $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ with $f(x_i) = y_i$, $i = 1, \ldots, N$

Possible strategy: Sparse occupancy trees

Question: reasonable error bounds?- concentration of measure phenomenon

Finance - high dimensional integration

In the US mortgages last 30 years and may be repaid each month, which gives $12 \times 30 = 360$ repayment possibilities $\rightsquigarrow$
Computation of 360-dimensional expected value

$$\int_0^1 \cdots \int_0^1 f(x_1, \ldots, x_{360}) dx_1 \cdots dx_{360}$$

Note: Quadrature rule with k nodes in $[0, 1]$ requires k^{360} point evaluations...

Electronic Structure Calculation

Goal: Numerical simulation of molecular phenomena in chemistry, molecular biology, semiconductor devices, material sciences...

"Ab-Initio" Calculations based on first principles in quantum mechanics (ignoring relativistic effects and using the Born-Oppenheimer approximate Model) $\rightsquigarrow$

Quantum mechanical postulates:

- System of N identicle (non-relativistic) particles with spin s_i described by a state function

$$\psi(x_1, s_1; \dots; x_N, s_N), \quad \psi : \mathbb{R}^{3N} \otimes S^N \to \mathbb{C}, \quad \langle \psi, \psi \rangle = 1$$

- ψ satisfies (stat.) Schrödinger equation with Hamiltonian H

$$H\psi = E_0\psi, \qquad E_0 = \min_{\langle \psi,\psi \rangle = 1} \langle H\psi, \psi \rangle$$

Born-Oppenheim:

$$H = \sum_{i=1}^{N} \Big\{ -\frac{1}{2}\Delta_i - \sum_{j=1}^{M} \frac{z_j}{|x_i - R_j|} + \frac{1}{2}\sum_{j \neq i} \frac{1}{|x_i - x_j|} \Big\}$$

z_j = charge of jth nucleus at position R_j

Typical Applications

- Simulation of porous media flow, contamination prediction, well protection
- Understanding heterogeneous materials like concrete

Classical diffusion equation:

$$-\operatorname{div}(A\nabla u) = f \quad \text{in} \quad D \subset \mathbb{R}^d, \quad u\,|_{\partial D} = 0, \ (d = 2, 3) \quad (1)$$

$A = A(x)$ describes diffusivity of the material

Problem: In heterogeneous porous media the small scales of the material make it impossible to describe all details by A and to resolve them numerically

Stochastic Model

Idea: view A as a random field ($A = aI$ scalar) about which (coarsely sampled) measurements provide uncertain information:

$$a = a(\cdot, \omega) : \omega \to L_\infty(D) =: X, \quad \omega \in \Omega$$

where

$$(\Omega, \Sigma, \rho) \quad \text{probability space on data space} \quad X$$

Proposition:

When $a(\cdot, \omega)$ stays bounded away from zero ρ-a.s. then (1) is well posed, i.e. there exists a unique $u(\cdot, \omega) : \Omega \to H_0^1(D)$ which is a weak solution of (1).

Transformation into Paremeter Dependent PDE

Typical goal: determine $\overline{u} = \mathbb{E}_\Omega(u)$

Possible strategy – Ansatz:

$$a(x,\omega) = \mathbb{E}_\Omega(a)(x) + \sum_{m=1}^{\infty} a_m(x) y_m(\omega)$$

specification of $a_m(x), y_m(\omega)$ via "Karhunen-Loewe-expansion".... lots of stochastic assumptions ...$\rightsquigarrow$

$$-\operatorname{div}\left(a_M(x; y_1, \ldots, y_M)\nabla u_M(x)\right) = f(x) \quad x \in D \quad u\,|_{\partial D} = 0 \qquad (2)$$

Issues and Objectives

- Solve (2) by numerical methods
- balance discretization error and truncation error due to M
- Compute function $u(x; y_1, \ldots, y_M)$ of $d + M$ variables
- Number of variables M in y becomes a discretization parameter

An Instance of “Manifold Learning”

Optimal control, shape optimization $\rightsquigarrow$ parameter dependent PDEs

$$F(u;\mathbf{y}) = f \quad \rightsquigarrow \quad u = u(\cdot,\mathbf{y}) \in \mathcal{H}, \quad \mathbf{y} \in Y \quad \rightsquigarrow$$

Manifold: $\mathcal{M} := \{u(\cdot;\mathbf{y}) : \mathbf{y} \in Y\} \subset \mathcal{H}$

Analogously: replace $\mathcal{H}$ by $\mathcal{H}_N \rightsquigarrow \quad \mathcal{M} \subset \mathbb{R}^N$

$$u_N(x;\mathbf{y}) = \sum_{i=1}^{N} u^i(\mathbf{y})\phi_i(x), \quad u_N(\cdot;\mathbf{y}) \leftrightarrow (u^1(\mathbf{y}),\ldots,u^N(\mathbf{y})) \in \mathbb{R}^N$$

Objective: Assess this manifold with complexity $<< N$

Reduced Order Methods: Maday, Patera,.....

Summary of Key Issues

Enimies:

- Complexity of neighborhood search, strong dependence on particular norm
- Curse of dimensionality, exponential dependence on d

Remedies ?

- Greedy techniques with problem adapted dictionaries
- Dimension reduction techniques (compressed sensing)
- Sparsity preserving recovery techniques, anisotropy

A New Paradigm in Signal Processing

Classical model:

- bandlimited signals
- sampling at Nyquist rate $x_i = x(t_i)$

Compressed Sensing (CS)

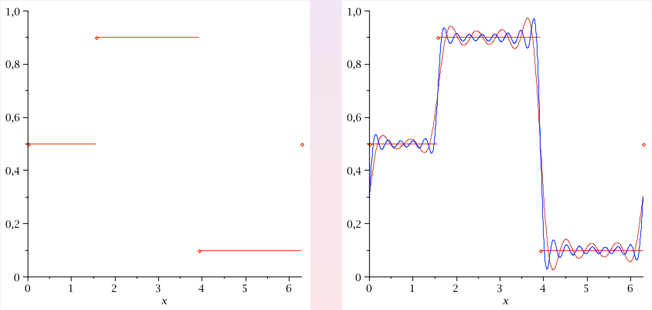
- Sparsity model:
  $x \in \mathbb{R}^N, \quad x = \Psi z, \quad \Psi \in \mathbb{R}^{N \times N}, \quad \#\mathrm{supp} z = k \ll N$
- Change notion of sampling
  $x \to \phi^j \cdot x, j = 1, \ldots, n, \quad n \ll N$
- Rate $\sim$ information content $k \sim n$

Goal:

- Minimize *a-priori* the number of measurements from complex signals $x \in \mathbb{R}^N$
- while retaining "essential" information

A Simple Effect

$$y_\ell = \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{-i\ell t}dt \approx \frac{1}{N}\sum_{j=0}^{N-1} \underbrace{f(2\pi j/N)}_{=:x_j}\,\underbrace{e^{-i\ell 2\pi j/N}}_{=:\phi_{\ell,j}} = (\Phi x)_\ell$$



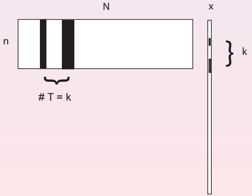
Exact reconstruction: $f = \mathrm{argmin}\,\{\|g'\|_1 : \hat{g}(\ell) = \hat{f}(\ell),\ |\ell| \le k\}$

Key Task

Question:

How to design data-independent linear functionals $\phi_i, i = 1, \ldots, n << N$, such that one can still recover "substantial" information on x from $\phi_i \cdot x$, $i = 1, \ldots, n$

Formally: $y_i = \phi_i \cdot x \quad \rightsquigarrow \quad y = \Phi x, \quad \Phi \in \mathbb{R}^{n \times N}$



Decoding Concepts

Main Issue:

Recovery of sparsity

- ℓ_1-minimization (Donoho, Candes/Romberg/Tao...):

$$x^* = \mathrm{argmin}_{\Phi z = y} \|z\|_{\ell_1}$$

- Greedy algorithms (Gilbert/Tropp, Cohen/D/DeVore, Temlyakov...)

Goal:

Use such concepts in the other contexts (Guermond/Popov)

Greedy agorithms - Curse of dimensionality

$\mathcal{H}$ Hilbert space, $\mathcal{D} \subset \mathcal{H}$ dictionary, $\|g\| = 1, g \in \mathcal{D}$, $f \in \mathcal{H}$

- $r_0 = f$, $f_0 = 0$
- given f_{k-1} determine $g_k := \mathrm{argmax}_{g \in \mathcal{D}} \langle r_{k-1}, g \rangle$ and set

$$f_k := P_{\mathrm{span}\{g_1,\dots,g_k\}} f$$

Theorem:

If $f \in \mathcal{L}_1(\mathcal{D})$, where $\|f\|_{\mathcal{L}_1} := \inf\{\sum_{g\in\mathcal{D}} |c_g| : f = \sum_{g\in\mathcal{D}} c_g g\}$, then

$$\|f - f_k\| \leq k^{-1/2} \|f\|_{\mathcal{L}_1}$$

Issue:

Can one get "problem dependent" $\mathcal{D}$, e.g. reduced bases...