# CLASS NOTES

# Models, Algorithms and Data: Introduction to computing

# 2019

Petros Koumoutsakos, Jens Honore Walther, Julija Zavadlav

(Last update: March 29, 2019)

## IMPORTANT DISCLAIMERS

Much of the material (ideas, definitions, concepts, examples, etc) in these notes is taken for teaching purposes, from several references:

- Numerical Analysis by R. L. Burden and J. D. Faires
- The Nature of Mathematical Modeling by N. Gershenfeld
- A First Course in Numerical methods by U. M. Ascher and C. Greif

These notes are only informally distributed and intended ONLY as study aid for the final exam of ETHZ students that were registered for the course Models, Algorithms and Data (MAD): Introduction to computing 2019. The notes have been checked, however they may still contain errors so use with care.

# LECTURE 6 Interpolation and extrapolation III: Radial basis functions

## 6.1 Orthogonal Functions

In a polynomial fit, the coefficients at each order are intimately connected. Dropping a high-order term gives a completely different function; it is necessary to re-fit to find a lower-order approximation. For many applications, it is convenient to have a model that allows successive terms to be added to improve the agreement without changing the coefficients that have been already found.

**Example 6.1**

If we store an image in such a way, then the fidelity with which it is displayed can be varied by changing the number of terms used. The fidelity can then be based on the available display, bandwidth and processor.

A classical approach to allow for successive corrections goes via **orthogonal functions**. Orthogonal functions are the functional analog to orthogonal vectors. Using the definition of the inner product on function spaces we can define the notion of orthonormal in exact correspondence

$$\boxed{\langle \phi_i(x)\phi_j(x)\rangle = \int_{-\infty}^{\infty} \phi_i(x)\phi_j(x)\,dx = \delta_{ij}} \,. \tag{6.1.1}$$

With the Kronecker delta $\delta_{ij}$. As $\phi_i$ is assumed to be a basis we can approximate any function using the usual expansion

$$y(x) = \sum_{i=1}^{M} \alpha_i \phi_i(x) \tag{6.1.2}$$

The orthogonality of the basis functions allows us to derive an explicit expression for the coefficients

$$\int_{-\infty}^{\infty} y(x)\phi_j(x)\,dx = \int_{-\infty}^{\infty} \phi_j(x) \sum_{i=1}^{M} \alpha_i \phi_i(x)\,dx \tag{6.1.3}$$

$$= \sum_{i=1}^{M} \alpha_i \int_{-\infty}^{\infty} \phi_i(x)\phi_j(x)\,dx \tag{6.1.4}$$

$$= \sum_{i=1}^{M} \alpha_i \delta_{ij} = \alpha_j \tag{6.1.5}$$

When a set of experimental observations $\{x_i, y_i\}_{i=1,...,N}$ is available instead of a functional form $y(x)$, then it is not possible to directly evaluate the integrals in 6.1.3. We can still use orthonormal functions but they must be orthonormal with respect of the probability density $p(x)$ of the measurements. This means we now choose the $\phi_i$'s such that:

$$\langle \phi_i(x)\phi_j(x)\rangle_p = \int_{-\infty}^{\infty} \phi_i(x)\phi_j(x)p(x)dx = \delta_{ij}$$

In this case we again find a closed form for the coefficients

$$\alpha_i = \langle y(x)\phi_i(x)\rangle = \int_{-\infty}^{\infty} y(x)\phi_i(x)p(x)dx.$$

**Example 6.2**

We approximate the probability density $p(x)$ as sums of Dirac delta functions:

$$p(x) \approx \frac{1}{N}\sum_{n=1}^{N} \delta(x - x_n),$$

where Dirac delta functions have the property

$$\int_{-\infty}^{\infty} f(x)\delta(x - x_n)dx = f(x_n).$$

We therefore have

$$\boxed{\alpha_i \approx \frac{1}{N}\sum_{n=1}^{N} y_n\phi_i(x_n)}. \tag{6.1.6}$$

## Gram-Schmidt orthogonalisation

If we are given a complete set of functions $g_i(x)$ then an orthonormal set of functions $\phi_i(x)$ can be created via the Gram-Schmidt orthogonalization

$$\begin{aligned}\tilde{\phi}_n(x) &= g_n(x) - \sum_{i=1}^{N-1} \phi_i(x) \int_{-\infty}^{\infty} \phi_i(x) g_n(x)\mathrm{d}x \\ \phi_n(x) &= \frac{\tilde{\phi}_n(x)}{\|\tilde{\phi}_n(x)\|}\end{aligned} \tag{6.1.7}$$

Let us make this explicit by performing some steps. We start with an arbitrary vector from our complete set of functions and normalize it

$$\phi_1(x) = \frac{g_1(x)}{[\int g_1(x)g_1(x)dx]^{1/2}}$$

Let us now create a second element which is orthogonal to this first element

$$\widetilde{\phi}_2(x) = g_2(x) - \phi_1(x)\int_{-\infty}^{\infty} \phi_1(x) g_2(x) dx$$

We now normalize this element via

$$\phi_2(x) = \frac{\widetilde{\phi}_2(x)}{[\int_{-\infty}^{\infty} \widetilde{\phi_2}(x)\widetilde{\phi_2}(x)dx]^{1/2}}$$

Let us now add another vector to our orthonormal set

$$\widetilde{\phi}_3(x) = g_3 - \phi_1 \int \phi_1 g_3 dx - \phi_2 \int \phi_2 g_3 dx$$

again normalizing

$$\phi_3 = \frac{\widetilde{\phi}_3(x)}{\left[\int \widetilde{\phi}_3 \widetilde{\phi_3} dx\right]^{\frac{1}{2}}}$$

and so on...

We can do the same in the case of experimental data, but replace the scalar product by the one weighted by the distribution underlying our dataset

$$\begin{aligned} \tilde{\phi}_n(x) &= g_n(x) - \sum_{i=1}^{N-1} \phi_i(x) \int_{-\infty}^{\infty} \phi_i(x) g_n(x) p(x) \mathrm{d}x \\ \phi_n(x) &= \frac{\tilde{\phi}_n(x)}{\|\tilde{\phi}_n(x)\|} = \frac{\tilde{\phi}_n(x)}{\sqrt{\int_{-\infty}^{\infty} \tilde{\phi}_i(x)\tilde{\phi}(x) p(x)\mathrm{d}x}} \end{aligned} \tag{6.1.8}$$

In contrast to our previous methods in this case the functions are fit to data by evaluating experimental expectations, rather than doing an explicit search for the fit coefficients.

**Example 6.3**

Basis functions can be orthogonalised with respect to known distributions. Some well known examples include:

- Hermite polynomials which are obtained by assuming $p(x) \sim e^{-x^2}$
- Laguerre polynomials where we assume $p(x) \sim e^{-x}$
- Chebyshev polynomials which are orthogonal with respect to $p(x) = (1 - x^2)^{-1/2})$

## 6.2 Radial Basis Function

In polynomials the only way to improve the fit is to add more high order terms which eventually diverge even more quickly. High order functions even when passing through the data are useless for interpolation or extrapolation. In particular this is true for functions with discontinuities or sharp peaks. Radial Basis functions (RBF) offer a sensible alternative.

Here we choose a set of identical basis functions $\phi$, which depend on the distance from a set of centers $c_i$ and on parameters $a_i$, which do not necessarily have to enter linearly:

$$y(x) = \sum_{i=1}^{M} \phi(|x - c_i|; a_i) \tag{6.2.1}$$

**Advantages:**

- Extra terms added without increased divergence (since all basis functions identical)
- Centers can be placed where needed

If the centers are fixed and the coefficients enter linearly then we recover the form we are used from the first chapter $y(x) = \sum_{i=1}^{M} a_i \phi(|x - c_i|)$

**Issues:**

1. **Ambiguity in choice of $\phi$**

$$\begin{aligned} \text{linear functions}: &\quad \phi(r) = r \\ \text{power functions}: &\quad \phi(r) = r^n \\ \text{Gaussian functions}: &\quad \phi(r) = e^{-r^2} \qquad \text{and many more...} \end{aligned}$$

   It can be shown that diverging functions have better convergence properties than local ones.

2. **Choice of $c_i$**: The centers can be chosen

   (a) Random or at preset positions

   (b) Data based, where we further have to destinguish between fixed or moving centers

3. **Linear or non-linear coefficients**: For non-linear coefficients, iterations are necessary. For linear we need to solve a system as we did for Least Squares.

**Exam checklist**

Functional representations with basis functions:

- Why/when to use orthogonal functions for interpolations
- How to construct them through the Gram-Schmidt orthogonalisation
- What are radial basis functions