

# Thesis manuscript

Ayoub Belhadji

May 5, 2020

# **1 Introduction**

## **1.1 Sensing on budget**

(3 pages)

## **1.2 Design problems in applied mathematics**

(3-6 pages)

## **1.3 Mathematical framework: Determinantal point processes**

(7-10 pages)

## **1.4 The layout**

(2 pages)

# **2 The column subset selection problem**

## **2.1 Related work**

## **2.2 The proposed algorithm**

## **2.3 Principal angles between subspaces**

## **2.4 Main results**

### **2.4.1 Low rank approximations**

### **2.4.2 Linear regression**

## **2.5 Numerical experiments**

## **2.6 Discussion**

(5-6 pages)

- Probabilistic inequalities
- A consequence for Rank Revealing QRs
- A consequence for graph signal reconstruction
- Connexions with optimal design problems
- The difference between sparse PCA and CSSP
- The approximation of the volume by random projections
- The limitations

## **2.7 Prologue: the infinite column subset selection problem**

(1-2 pages)

### 3 Kernel quadrature using DPPs

#### 3.1 Numerical integration problems

#### 3.2 Related work

The purpose of this section is to motivate the theoretical analysis of a class of quadratures based on projection DPPs that is naturally defined in the RKHS framework. The merit of this class of quadratures is its universality and applicability to a variety of settings. An exhaustive review of existing results on quadratures is outside the scope of this manuscript; nevertheless, it turns out that many landmark results on the field are connected to projection DPPs. For this reason, we review these classic results in the following.

##### 3.2.1 Quadratures on the real line

We start by a brief review on quadratures on the real line.

**Newton-Cotes quadrature** The approximation of the integral of a function  $f$  on some interval  $[a, b]$  using some evaluations of  $f$  can be tracked to the work of Newton. The idea of the eponymous quadrature is to consider a set of nodes  $\mathbf{x} = \{x_1, \dots, x_N\} \subset \mathcal{X}^N$ , and the corresponding interpolating polynomial of degree  $N - 1$ :

$$p_{N-1}(X) = \sum_{n \in [N]} f(x_n) \ell_{n,\mathbf{x}}(X), \quad (1)$$

where  $\ell_{n,\mathbf{x}}$  is defined by

$$\ell_{n,\mathbf{x}}(X) = \frac{\prod_{m \in [N], m \neq n} (X - x_m)}{\prod_{m \in [N], m \neq n} (x_n - x_m)}. \quad (2)$$

The idea of Newton was to consider the approximation

$$\int_a^b f(t) dt \approx \int_a^b p_{N-1}(t) dt \quad (3)$$

$$= \sum_{n \in [N]} f(x_n) \int_a^b \ell_{n,\mathbf{x}}(t) dt. \quad (4)$$

$$:= \sum_{n \in [N]} f(x_n) w_n. \quad (5)$$

The approximation (3) is exact if  $f$  is a polynomial of order less than  $N - 1$ . The scalars  $w_n$  are called Cotes numbers

**Gaussian quadratures** As we have seen previously, the Newton-Cotes quadrature is exact for polynomials of order less than  $N - 1$ . Gauss interested in the maximal order of exactness that could be achieved using the Newton-Cotes quadrature. This order depends on the design of nodes  $\mathbf{x}$  and is defined by

$$M(\mathbf{x}) = \sup\{m \in \mathbb{N}, \int_a^b t^m dt = \sum_{n \in [N]} w_n x_n^m\}. \quad (6)$$

Gauss proved that there exists a design  $\mathbf{x}_G$  such that  $M(\mathbf{x}_G) = 2N - 1$ , and that this order is maximal among all the possible designs. The proof of Gauss is based on arguments from continued fractions theory.

An alternative proof was given by Jacobi in [???], where he proved the existence of such a design,  $\mathbf{x}_G$ , that we call the Gauss nodes, using simple arguments of polynomial divisibility. Indeed, Jacobi shows that for any integer  $N' \in [N]$ , the Newton-Cotes quadrature has the maximal degree of exactness  $M(\mathbf{x}) = N - 1 + N'$  if and only if for every polynomial  $p$  of order less than  $N' - 1$ :

$$\int_a^b p(t) \pi_{\mathbf{x}}(t) dt = 0, \quad (7)$$

where  $\pi_{\mathbf{x}}$  is the polynomial

$$\pi_{\mathbf{x}}(X) = \prod_{n \in [N]} (X - x_n). \quad (8)$$

The case  $N' = N$  leads to the Gauss formula of maximal degree of exactness  $2N - 1$ . More importantly, this characterization of Gauss nodes, proved by Jacobi, highlights the importance of orthogonality between polynomials: the Gauss nodes are the roots of a polynomial of degree  $N$  that is orthogonal to all polynomials of degree less than  $N - 1$ . The corresponding nodes polynomials  $\pi_{\mathbf{x}}$  are nothing else but the scaled Legendre polynomials.

The notion of orthogonality between polynomials was pushed further by considering quadrature problems with respect to weighted measures:  $d\omega = w(t)dt$ .

$$\int_a^b f(t) w(t) dt \approx \sum_{n \in [N]} w_n f(x_n). \quad (9)$$

### 3.2.2 Quasi-Monte Carlo methods

The weights of a Quasi Monte Carlo quadrature are uniform  $w_i = \frac{1}{N}$ , while the nodes, defined on  $\mathcal{X} = [0, 1]^D$ , are deterministic, and "well-spread", such that the quadrature error is small for a certain class of functions.

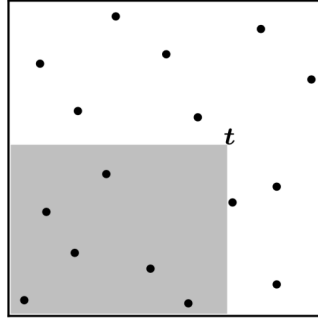


Figure 1

As we have seen previously, several quadrature rules are known for functions defined on the real line. The situation is far more complicated for integration problems in high-dimension. For instance, the uniform grid suffers from the "curse of the dimension". To precise the meaning of the curse in this context, we review some notions on low discrepancy sequences also known as quasi-Monte Carlo methods.

**The discrepancy functions** A discrepancy function of a design  $\mathbf{x} \subset \mathcal{X}^N$  quantifies the "spread" of the nodes of  $\mathbf{x}$ . It is defined with respect to a set of test sets  $\mathcal{S}$ .

For example, in  $\mathcal{X} = [0, 1]^d$ , a typical choice of  $\mathcal{S}$  is the set of the Cartesian products  $[\mathbf{0}, \mathbf{u}] = \prod_{\delta \in [d]} [0, u_\delta]$  with  $\mathbf{u} \in \mathcal{X}$ . In this case, the local discrepancy is defined for all  $\mathbf{u} \in \mathcal{X}$  by

$$\Delta_{\mathbf{x}}(\mathbf{u}) = \frac{1}{N} \sum_{n \in [N]} \mathbb{1}_{[\mathbf{0}, \mathbf{u}]}(x_n) - \prod_{\delta \in [d]} u_\delta. \quad (10)$$

The term  $\frac{1}{N} \sum_{n \in [N]} \mathbb{1}_{[\mathbf{0}, \mathbf{u}]}(x_n)$  in (10) counts the number of the nodes  $x_n$  that falls into  $[\mathbf{0}, \mathbf{u}]$

divided by the total number  $N$ ; while the second term  $\prod_{\delta \in [d]} u_\delta$  measures the volume of  $[\mathbf{0}, \mathbf{u}]$ :

$\Delta_{\mathbf{x}}$  measures the difference between the relative number of points that belongs to the interval  $[\mathbf{0}, \mathbf{u}]$  and its volume. A design of nodes  $\mathbf{x}$  would be "well-spread" if the discrepancy function  $\Delta_{\mathbf{x}}$  is low in  $\mathcal{X}$ . A way to quantify this "spreadness" is by taking the  $\|\cdot\|_p$  norm of the local discrepancy function on  $\mathcal{X}$ . Indeed, for  $p \in [1, \infty]$ ,  $\Delta_{\mathbf{x}} \in \mathbb{L}_p(\mathbb{R}^d)$  because  $[0, 1]^d$  is a compact set of  $\mathbb{R}^d$ ; and the  $\mathbb{L}_p$ -discrepancy of  $\mathbf{x}$  is defined by

$$\Delta_p(\mathbf{x}) = \left( \int_{[0, 1]^d} \Delta_{\mathbf{x}}(\mathbf{u})^p d\mathbf{u} \right)^{1/p}. \quad (11)$$

In particular, the  $\|\cdot\|_\infty$  norm of the local discrepancy is called the *star-discrepancy* and it is denoted  $\Delta_*(\mathbf{x})$ .

Figure 1 illustrates the concept of local discrepancy on  $[0, 1]^2$ : the two hypercubes  $[\mathbf{0}, \mathbf{u}_1]$  and  $[\mathbf{0}, \mathbf{u}_2]$  have the same volume, yet the corresponding local discrepancies with respect to the design  $\mathbf{x}$  are not equal as  $[\mathbf{0}, \mathbf{u}_1]$  contains more nodes from the design than  $[\mathbf{0}, \mathbf{u}_2]$ .

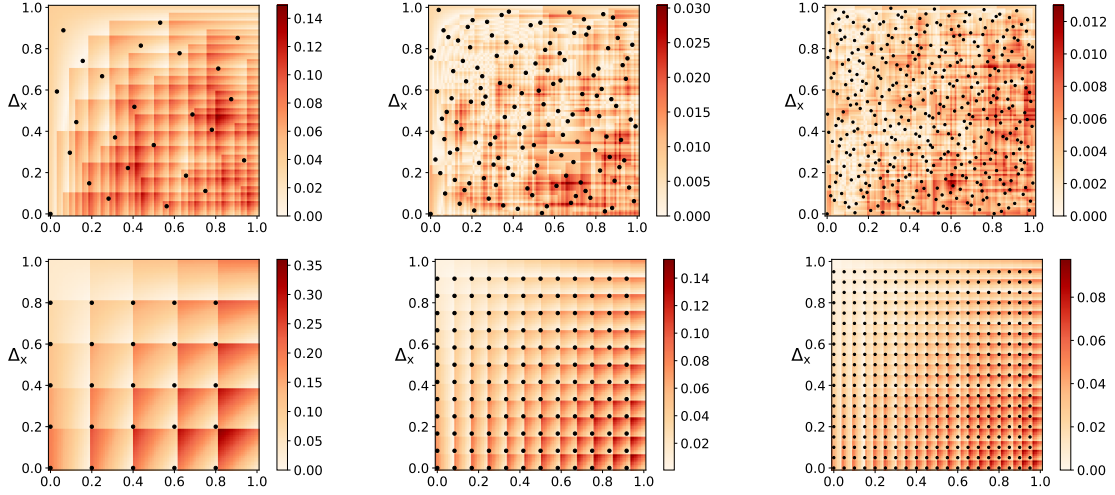


Figure 2: The local discrepancy  $\Delta_{\mathbf{x}}$  for two designs: in the top, the Halton sequence with  $N \in \{25, 144, 400\}$ ; in the bottom, the uniform grid with  $N \in \{25, 144, 400\}$ .

Another example is the hypersphere  $\mathcal{X} = \mathbb{S}^{d-1}$ , where commonly considered test sets are the spherical caps defined by

$$C(t, \mathbf{u}) = \{\mathbf{v} \in \mathbb{S}^{d-1}, \langle \mathbf{v}, \mathbf{u} \rangle \geq t\}, \quad (12)$$

**Koksma-Hlawka inequalities** The discrepancy functions for a design  $\mathbf{x}$  play an important role in the quantification of the integration error based on  $\mathbf{x}$ . Indeed, consider the quadrature:

$$\frac{1}{N} \sum_{n \in [N]} f(x_n), \quad (13)$$

as an approximation of  $\int_{[0,1]^d} f(\mathbf{u}) d\mathbf{u}$ . The relationship between the discrepancy functions and the integration error clear up considering the Koksma-Hlawka inequalities.

**Theorem 1.** *Let  $p, q \in [1, +\infty]$ , such that  $1/p + 1/q = 1$ . Consider  $f \in \mathbb{L}_q([0, 1]^d)$ . Then*

$$\left| \int_{[0,1]^d} f(\mathbf{u}) d\mathbf{u} - \frac{1}{N} \sum_{n \in [N]} f(x_n) \right| \leq \Delta_p(\mathbf{x}) \left\| \frac{\partial^d f}{\partial \mathbf{u}} \right\|_q. \quad (14)$$

The Koksma-Hlawka inequalities gives a decoupled upper bound on the integration error: the  $\mathbb{L}_p([0, 1]^d)$  discrepancy of the design  $\mathbf{x}$  does not depend on  $f$  and the  $q$  variation of the function  $f$  does not depend on the design  $\mathbf{x}$ . This decoupling allows to focus only on the discrepancy of the design.

Looking for designs having the lowest discrepancy was a topic of intense research. A very brief review of this line of research is provided in the following paragraph.

### Low discrepancy sequences

3.2.3	Kernel quadrature and Bayesian quadrature
3.2.4	DPPs for numerical integration
3.3	Main results
3.4	Sketch of the proofs
3.5	Numerical simulations
4	Kernel interpolation using Volume Sampling
5	Interpolation problems in signal processing
6	The Richest Hermite Kernel and integration problems in $\mathbb{R}^d$
7	Future work and conclusion
References	
A	Details about the simulations of Chapter 2
B	Proofs of the main results appearing in Chapter 2