

Des règles de quadrature dans les RKHSs à base de DPPs

Ayoub BELHADJI¹

¹Univ Lyon, ENS de Lyon
Inria, CNRS, UCBL
LIP UMR 5668, Lyon, France
ayoub.belhadji@ens-lyon.fr

Résumé – On étudie des règles de quadrature à base de configurations de nœuds qui suivent la distribution d’un processus ponctuel déterminantal (DPP). Ces règles de quadrature sont adaptées à des fonctions qui appartiennent à des espaces d’Hilbert à noyau reproduisant. On montre que le taux de convergence dépend principalement des valeurs propres de l’opérateur d’intégration correspondant. En particulier, ce travail fournit une analyse unifiée de l’intégration numérique et offre un nouveau regard sur l’utilisation des DPPs dans ce domaine.

Abstract – We study quadrature rules based on configurations of nodes that follow the distribution of a determinantal point process (DPP). These quadrature rules are suitable for functions that belong to reproducing kernel Hilbert spaces (RKHS). We prove that the rate of convergence depends principally on the eigenvalues of the corresponding integration operator. In particular, this work furnishes a unified analysis of numerical integration and offers a new perspective on the use of DPPs in this field.

1 Introduction

Approximer des intégrales est une tâche récurrente dans plusieurs domaines des mathématiques appliquées [1] : traitement de signal, apprentissage automatique, statistiques Bayésiennes... En pratique, une telle approximation est réalisée à l’aide d’une règle de quadrature : une somme pondérée des évaluations de la fonction à intégrer. Dans ce travail, on s’intéresse aux règles de quadrature de la forme

$$\int_{\mathcal{X}} f(x)g(x)d\omega(x) \approx \sum_{i \in [N]} w_i f(x_i), \quad (1)$$

avec les nœuds x_i sont indépendants du choix des fonctions f et g , alors que les poids w_i peuvent dépendre de g . L’usage des processus ponctuels déterminantaux (DPPs) dans le cadre de l’intégration numérique a été proposé récemment par Bardenet et Hardy dans [2]. Les DPPs sont des modèles probabilistes de sous-ensembles aléatoires impliquant une répulsion entre les particules. Cette répulsion entre les nœuds permet de mieux couvrir le domaine d’intégration \mathcal{X} de telle sorte que l’approximation (1) soit de meilleure qualité comparée à une règle de quadrature impliquant des nœuds aléatoires indépendants. Ainsi, la règle de quadrature proposée dans [2] converge plus rapidement qu’une quadrature de type Monte Carlo pour des fonctions suffisamment régulières. Cependant, le domaine d’intégration dans [2] est limité à l’hypercube $[0, 1]^d$. Depuis la publication de ce travail, plusieurs articles [3, 4, 5] ont étudié des règles de quadrature pour des fonctions qui appartiennent à un espace d’Hilbert à noyau reproduisant (RKHS). Ce formalisme permet de travailler avec des fonctions régulières et

définis sur des domaines d’intégration variées.

Le présent article est une synthèse des résultats figurant dans ces articles. Il est structuré de la façon suivante : dans Section 2 on rappelle la construction des règles de quadrature proposée dans ces articles ; et dans Section 3, on rappelle les garanties théoriques des règles de quadratures proposées.

Notations On note par ω une mesure borélienne sur \mathcal{X} et on note par $\mathcal{L}_2(\omega)$ l’espace d’Hilbert des fonctions réelles définies sur \mathcal{X} et de carré intégrable par rapport à ω équipé d’un produit scalaire $\langle \cdot, \cdot \rangle_\omega$ et de la norme associée $\|\cdot\|_\omega$. On note par \mathcal{F} , l’RKHS associé au noyau $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ qu’on suppose continue et qui satisfait $\int_{\mathcal{X}} k(x, x)d\omega(x) < +\infty$. En plus, on suppose que k admet la décomposition Mercer

$$k(x, y) = \sum_{m \in \mathbb{N}^*} \sigma_m e_m(x) e_m(y), \quad (2)$$

où les σ_m et les e_m sont respectivement les valeurs propres et les fonctions propres de l’opérateur d’intégration Σ définie sur $\mathcal{L}_2(\omega)$ par

$$\Sigma f(\cdot) = \int_{\mathcal{X}} k(\cdot, y) f(y) d\omega(y). \quad (3)$$

On suppose que la suite $\sigma = (\sigma_m)_{m \in \mathbb{N}^*}$ est décroissante et de valeurs strictement positives et les fonctions propres correspondantes sont continues et normalisées : $\|\phi_m\|_\omega = 1$. En particulier, la famille $(\phi_m)_{m \in \mathbb{N}^*}$ est une base orthonormée de $\mathcal{L}_2(\omega)$ et tout élément $f \in \mathcal{F}$ satisfait la condition

$$\|f\|_{\mathcal{F}}^2 = \sum_{m \in \mathbb{N}^*} \frac{\langle f, e_m \rangle_\omega^2}{\sigma_m} < +\infty. \quad (4)$$

2 Une famille de règles de quadrature dans un RKHS

Dans cette section, on propose une construction universelle d'une règle de quadrature pour des fonctions de l'RKHS \mathcal{F} à l'aide des nœuds répulsifs. La construction des règles de quadrature est un champ de recherche très vaste qui ne peut être couvert convenablement dans ce travail. Cependant, on ne peut s'empêcher de remarquer que la construction d'une règle de quadrature se ramène à poser trois questions :

- comment choisir les nœuds x_1, \dots, x_N ?
- comment choisir les poids w_1, \dots, w_N ?
- quelle est l'erreur de l'approximation ?

La construction proposée répond à ces trois questions, mais dans le sens inverse. D'abord, on rappelle dans Section 2.1 comment les noyaux reproduisants peuvent être utilisés pour une quantification de l'erreur d'approximation d'une règle de quadrature. Ensuite, on propose dans Section 2.2 de choisir les poids pour une configuration de nœuds donnée. Enfin, on propose d'adopter des nœuds répulsifs dans Section 2.3.

2.1 L'analyse des quadratures à l'aide des noyaux

L'utilisation des noyaux dans le contexte de l'intégration numérique remonte aux travaux de Hickernell [6]. Dans ce travail, la qualité d'approximation d'une règle de quadrature est estimée à travers l'erreur d'intégration du pire cas dans la boule unité de l'RKHS \mathcal{F} associé au noyau $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$:

$$\sup_{\substack{f \in \mathcal{F} \\ \|f\|_{\mathcal{F}} \leq 1}} \left| \int_{\mathcal{X}} f(x)g(x)d\omega(x) - \sum_{i \in [N]} w_i f(x_i) \right|. \quad (5)$$

Cette quantité a une formule close

$$\left\| \mu_g - \sum_{i \in [N]} w_i k(x_i, \cdot) \right\|_{\mathcal{F}}, \quad (6)$$

avec $\mu_g = \Sigma g$ est le plongement de g dans l'RKHS \mathcal{F} . La formule (6) est égale à

$$\sqrt{\|\mu_g\|_{\mathcal{F}}^2 - 2\mathbf{w}^T \boldsymbol{\mu}_g(\mathbf{x}) + \mathbf{w}^T \mathbf{K}(\mathbf{x}) \mathbf{w}}, \quad (7)$$

avec $\mathbf{w} := (w_i)_{i \in [N]} \in \mathbb{R}^N$ et $\boldsymbol{\mu}_g(\mathbf{x}) := (\mu_g(x_i))_{i \in [N]} \in \mathbb{R}^N$ et

$$\mathbf{K}(\mathbf{x}) := (k(x_{i_1}, x_{i_2}))_{(i_1, i_2) \in [N] \times [N]} \in \mathbb{R}^{N \times N}. \quad (8)$$

En d'autres termes, la formule (7) ramène l'évaluation de (5), une quantité a priori intractable, à l'évaluation de $\|\mu_g\|_{\mathcal{F}}^2$, des $\mu_g(x_i)$ et des $k(x_{i_1}, x_{i_2})$, des quantités qu'on peut évaluer explicitement pour un grand nombre de noyaux k et de fonctions g . Cette formule a le mérite de ne pas dépendre d'une fonction test particulière $f \in \mathcal{F}$, ce qui rend la comparaison des règles de quadrature convenable et justifie l'utilisation des noyaux reproduisants dans le domaine de l'intégration numérique.

2.2 Les poids de la quadrature

Etant donnée une configuration de nœuds $\mathbf{x} = (x_1, \dots, x_N)$, on choisit le vecteur des poids $\hat{\mathbf{w}}^{\text{OKQ}}$ qui minimise (7) vue

comme une fonction de \mathbf{w} . Ainsi, la quadrature obtenue est appelée la *quadrature à noyau optimale* (OKQ)¹. En général, l'existence et l'unicité de ce vecteur est assuré quand la matrice

$$\mathbf{K}(\mathbf{x}) := (k(x_i, x_j))_{(i, j) \in [N]^2} \in \mathbb{R}^{N \times N}, \quad (9)$$

est inversible. Sous cette hypothèse, le vecteur des poids vérifie

$$\hat{\mathbf{w}}^{\text{OKQ}} = \mathbf{K}(\mathbf{x})^{-1} \boldsymbol{\mu}_g(\mathbf{x}), \quad (10)$$

avec $\boldsymbol{\mu}_g(\mathbf{x}) := (\mu_g(x_i))_{i \in [N]} \in \mathbb{R}^N$. En plus, on a

$$\sum_{i \in [N]} \hat{w}_i^{\text{OKQ}} k(x_i, \cdot) = \Pi_{\mathcal{T}(\mathbf{x})} \mu_g, \quad (11)$$

où $\Pi_{\mathcal{T}(\mathbf{x})} \mu_g$ est la projection orthogonale de μ_g sur le sous-espace $\mathcal{T}(\mathbf{x}) := \text{Span}(k(x_i, \cdot))_{i \in [N]}$ engendré par les translations de noyau. Finalement, on observe que la formule (10) implique que la fonction $\Pi_{\mathcal{T}(\mathbf{x})} \mu_g$ prend les mêmes valeurs que μ_g dans les nœuds x_1, \dots, x_N : la mixture est l'interpolante de μ_g par le noyau k . Figure 1 illustre l'importance du choix de la configuration des nœuds dans la qualité de l'interpolation : une configuration de nœuds répulsive offre une meilleure interpolation, et la OKQ correspondante aura de meilleures performances dans l'intégration numérique. Néanmoins, la construction d'une configuration de nœuds répulsifs peut poser des difficultés dans un domaine arbitraire \mathcal{X} et surtout en grande dimension. Cet obstacle peut être contourné en considérant une configuration aléatoire de nœuds définie dans Section 2.3.

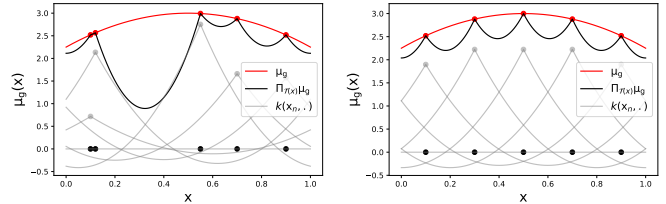


FIGURE 1 – Comparaison entre l'interpolante de μ_g à l'aide de deux configurations : une configuration non-répulsive (gauche) une configuration répulsive (droite).

2.3 Les nœuds de la quadrature

La configuration des nœuds $\mathbf{x} = (x_1, \dots, x_N)$ est une variable aléatoire à valeur dans \mathcal{X}^N qui suit la distribution de densité

$$p_{\kappa}(x_1, \dots, x_N) \propto \text{Det } \boldsymbol{\kappa}(\mathbf{x}), \quad (12)$$

relativement à la mesure produit $\omega^{\otimes N}$ défini dans \mathcal{X}^N , avec $\kappa : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ est un noyau p.s.d. qui vérifie la condition

$$\int_{\mathcal{X}} \kappa(x, x) d\omega(x) < +\infty. \quad (13)$$

Le noyau κ est appelé le noyau de répulsion qui peut être différent du noyau k de l'RKHS. La densité p_{κ} favorise les configurations t.q. $\text{Det } \boldsymbol{\kappa}(\mathbf{x})$ est grand, elle favorise ainsi les configurations les plus répulsives selon le noyau κ . En particulier, quand le noyau κ est de projection, cette distribution est équivalente à un DPP. Dans le cas général, on peut montrer que la distribution est une mixture de DPPs.

1. OKQ : optimal kernel quadrature.

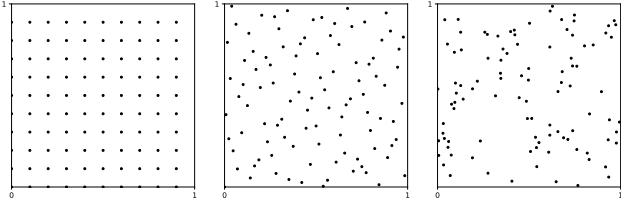


FIGURE 2 – Trois configurations de nœuds, avec $N = 100$, dans $\mathcal{X} = [0, 1]^2$: (gauche) grille uniforme, (centre) une réalisation du DPP de noyau κ_N , (droite) une réalisation de l'échantillonnage de N particules indépendantes de la loi uniforme.

Un exemple On considère $\mathcal{X} = [0, 1]^d$, ω est la mesure uniforme sur \mathcal{X} et \mathcal{F} est l'RKHS associé au noyau $k_{d,s}$

$$\forall x, y \in \mathcal{X}, \quad k_{d,s}(x, y) := \prod_{i=1}^d k_{1,s}(x_i, y_i), \quad (14)$$

avec $k_{1,s}(u, v)$ est défini pour $u, v \in [0, 1]$ par la formule

$$k_{1,s}(u, v) := 1 + \frac{(-1)^{s-1} (2\pi)^{2s}}{(2s)!} B_{2s}(\{u - v\}), \quad (15)$$

où B_{2s} est le polynôme de Bernoulli d'ordre $2s$. Ce noyau admet une décomposition Mercer (2) avec les e_m sont des fonctions trigonométriques et $\sigma_N = \mathcal{O}(\log^{2s(d-1)}(N)N^{-2s})$ [7]. Figure 2 compare une réalisation du DPP de noyau de répulsion $\kappa_N(x, y) := \sum_{m=1}^N e_m(x)e_m(y)$ où $N = 100$, avec la grille uniforme de N nœuds, et une configuration aléatoire obtenue comme l'échantillonnage de N particules indépendantes selon la distribution uniforme sur $[0, 1]^2$: la configuration obtenue comme une réalisation d'un DPP est moins régulière que la grille uniforme mais elle couvre $[0, 1]^2$ mieux que la configuration obtenue par échantillonnage indépendant. Cette dernière contient des rassemblements de points qui sont très rares à observer dans une configuration répulsive qui suit la loi d'un DPP.

3 Les garanties théoriques

Dans cette section, on présente les taux de convergence des règles de quadrature définis dans Section 2. On se focalise sur deux choix de noyaux de répulsion κ :

— le noyau de projection (la distribution PDPP²) :

$$\kappa(x, y) = \kappa_N(x, y) := \sum_{m=1}^N e_m(x)e_m(y), \quad (16)$$

— le noyau de l'RKHS (la distribution CVS³) : $\kappa = k$.

On commence par le résultat suivant.

Proposition 1 (Théorème 4 dans [4]). *Soit $g \in \mathcal{L}_2(\omega)$, on a*

$$\mathbb{E}_{\text{CVS}} \|\mu_g - \Pi_{\mathcal{T}(\mathbf{x})}\mu_g\|_{\mathcal{F}}^2 = \sum_{m \in \mathbb{N}^*} \langle g, e_m \rangle_{\omega}^2 \epsilon_m(N), \quad (17)$$

avec

$$\epsilon_m(N) = \sum_{\substack{|T|=N \\ m \notin T}} \prod_{t \in T} \sigma_t / \sum_{|T|=N} \prod_{t \in T} \sigma_t. \quad (18)$$

2. Projection DPP

3. Continuous volume sampling

Proposition 1 donne, pour $g \in \mathcal{L}_2(\omega)$, une formule close de $\mathbb{E}_{\text{CVS}} \|\mu_g - \Pi_{\mathcal{T}(\mathbf{x})}\mu_g\|_{\mathcal{F}}^2$ qui dépend principalement des coefficients de g dans la base orthonormée $(e_m)_{m \in \mathbb{N}^*}$. Cette formule met en relief l'importance des $\epsilon_m(N)$. On peut déjà observer qu'en appliquant (17) à $g = e_m$ on obtient

$$\epsilon_m(N) = \mathbb{E}_{\text{CVS}} \|\mu_{e_m} - \Pi_{\mathcal{T}(\mathbf{x})}\mu_{e_m}\|_{\mathcal{F}}^2. \quad (19)$$

En plus, d'après Proposition 5 dans [4] on a $\epsilon_m(N) = \mathcal{O}(\sigma_{N+1})$ dans plusieurs RKHSs. En particulier, en appliquant (17), on montre que $\mathbb{E}_{\text{CVS}} \|\mu_g - \Pi_{\mathcal{T}(\mathbf{x})}\mu_g\|_{\mathcal{F}}^2 = \mathcal{O}(\sigma_{N+1})$ pour tout $g \in \mathcal{L}_2(\omega)$. Remarquablement, ce taux de convergence est optimal comme indiqué par le résultat suivant.

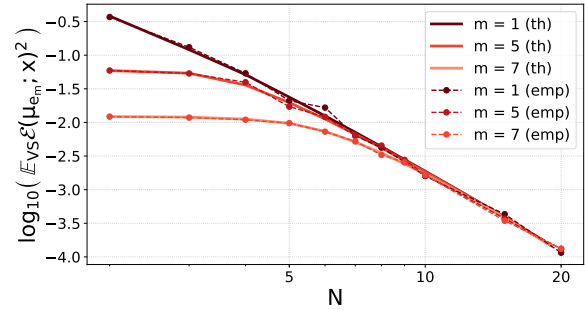


FIGURE 3 – La valeur théorique de $\epsilon_m(N)$ comparée à son estimation empirique réalisée à l'aide d'un algorithme MCMC.

Proposition 2 (Section 2.5. dans [4]). *Soit $\mathbf{x} = (x_1, \dots, x_N) \in \mathcal{X}^N$ t.q. $\text{Det } \mathbf{K}(\mathbf{x}) > 0$. On a*

$$\sup_{\substack{g \in \mathcal{L}_2(\omega) \\ \|g\|_{\omega} \leq 1}} \|\mu_g - \Pi_{\mathcal{T}(\mathbf{x})}\mu_g\|_{\mathcal{F}}^2 \geq \sigma_{N+1}. \quad (20)$$

En d'autres termes, la CVS offre une reconstruction presque optimale des μ_g . En pratique, il est possible d'échantillonner selon la distribution de la CVS grâce à un algorithme MCMC proposé dans [8]. Figure 3 compare les valeurs théoriques de $\epsilon_1(N)$, $\epsilon_5(N)$ and $\epsilon_7(N)$ à leurs estimations obtenus grâce à cet algorithme MCMC. L'RKHS adopté correspond au noyau défini dans l'exemple avec $d = 1$ et $s = 2$. On observe que la valeur de $\epsilon_m(N)$ passe par deux phases. Dans la première phase ($N \leq m$), la valeur de $\epsilon_m(N)$ est presque une constante égale à l'erreur initial σ_m . Dans la deuxième phase ($N > m$), la valeur de $\epsilon_m(N)$ tend vers 0 avec un taux de convergence $\mathcal{O}(\sigma_{N+1})$. Pour expliquer ce comportement, on observe que $\mu_{e_m} = \sqrt{\sigma_m} e_m^{\mathcal{F}}$ avec $e_m^{\mathcal{F}} := \sqrt{\sigma_m} e_m$ et que

$$\|e_m^{\mathcal{F}} - \Pi_{\mathcal{T}(\mathbf{x})}e_m^{\mathcal{F}}\|_{\mathcal{F}}^2 \leq \|e_m^{\mathcal{F}}\|_{\mathcal{F}}^2 = 1, \quad (21)$$

et on obtient

$$\epsilon_m(N) = \sigma_m \mathbb{E}_{\text{CVS}} \|e_m^{\mathcal{F}} - \Pi_{\mathcal{T}(\mathbf{x})}e_m^{\mathcal{F}}\|_{\mathcal{F}}^2 \leq \sigma_m. \quad (22)$$

Autrement dit, $\epsilon_m(N)$ est le produit de deux termes : σ_m qui ne dépend pas de N et $\mathbb{E}_{\text{CVS}} \|e_m^{\mathcal{F}} - \Pi_{\mathcal{T}(\mathbf{x})}e_m^{\mathcal{F}}\|_{\mathcal{F}}^2$ qui reflète l'erreur moyenne de reconstruction de la fonction propre $e_m^{\mathcal{F}}$ par sa projection dans le sous-espace $\mathcal{T}(\mathbf{x})$. C'est ce dernier terme qui converge vers 0 avec un taux $\mathcal{O}(\sigma_{N+1})$ à partir

de $N > m$. Autrement dit, le CVS permet de reconstruire les premières N fonctions propres $e_m^{\mathcal{F}}$ dans l'ordre : $e_m^{\mathcal{F}}$ est mieux reconstruite que $\phi_{m+1}^{\mathcal{F}}$ en moyen. L'algorithme MCMC utilisé pour réaliser cette simulation reste, néanmoins, une approximation de la distribution CVS et pour cette raison, on observe des déviations légères de l'estimation (emp) autour de la valeur théorique (th). Contrairement à la CVS, l'échantillonnage exact selon la distribution PDPP est possible à condition d'avoir une décomposition Mercer (2) qui soit explicite. En plus, on a le résultat suivant.

Proposition 3 (Théorème 5 dans [5]). *Sous la distribution PDPP, on a*

$$\forall g \in \mathcal{L}_2(\omega), \mathbb{E}_{\text{PDPP}} \left\| \mu_g - \Pi_{\mathcal{T}(\mathbf{x})} \mu_g \right\|_{\mathcal{F}}^2 \leq 4 \|g\|_{\omega}^2 r_N, \quad (23)$$

avec $r_N = \sum_{m \geq N+1} \sigma_m$.

Proposition 3 fournit une borne supérieure de l'espérance de l'erreur du pire cas de la OKQ, quand les nœuds suivent la distribution du DPP de noyau k_N . Le taux de convergence établie est de l'ordre de $\mathcal{O}(r_N)$. Néanmoins, les simulations numériques dans [3] suggèrent que ce taux de convergence est plutôt $\mathcal{O}(\sigma_{N+1})$. Figure 4 illustre ce taux de convergence dans le cas de l'RKHS considéré dans l'exemple avec $d = 2$ et $s = 1$. On compare le PDPP contre les nœuds d'une grille uniforme de même cardinalité en observant les

$$\epsilon_m(N) := \left\| \mu_{\phi_m} - \Pi_{\mathcal{T}(\mathbf{x})} \mu_{\phi_m} \right\|_{\mathcal{F}}^2,$$

moyennés sur 50 réalisations dans le cas de PDPP. On observe que sous la distribution PDPP, le taux de convergence est $\mathcal{O}(\sigma_N)$ alors que le taux de convergence pour la grille uniforme est $\mathcal{O}(N^{-s/2d})$ qui se dégrade fortement en grande dimension.

4 Conclusion

Dans ce travail, on a étudié la quadrature à noyau optimale avec des nœuds répulsifs. On a montré qu'avec un choix judicieux du noyau de répulsion, le taux de convergence de l'erreur du pire cas dans la boule unité d'un RKHS de cette quadrature est quasi optimal. En particulier, on a considéré deux distributions : la CVS et la PDPP. La première consiste à utiliser le noyau de l'RKHS comme noyau de répulsion, alors que la seconde correspond au noyau de projection associé aux premières fonctions propres. Finalement, les résultats théoriques ont été validés par des simulations numériques.

Références

- [1] Robert, C. P. and G. Casella (2004). Monte Carlo statistical methods. Springer
- [2] Bardenet, R. and Hardy, A. "Monte Carlo with determinantal point processes". In : The Annals of Applied Probability 30.1, pp. 368–417, 2020

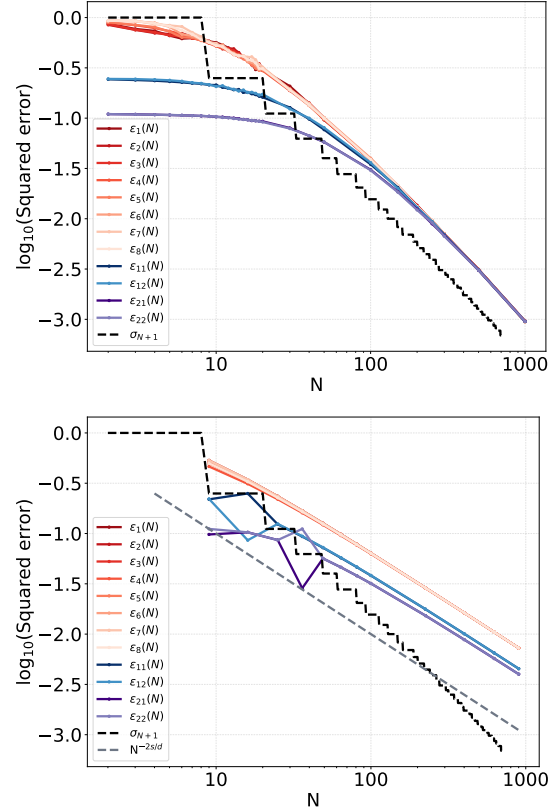


FIGURE 4 – Le terme d'erreur pour la OKQ sous la distribution PDPP (haut) une grille uniforme de $[0, 1]^2$. L'RKHS correspond au produit tensoriel des espaces de Sobolev en dimension $d = 2$ et de régularité $s = 1$.

- [3] Belhadji, A., Bardenet, R. and Chainais, P. *Kernel quadrature with DPPs*. In : Advances in Neural Information Processing Systems 32, pp. 12907–12917, 2019
- [4] Belhadji, A., Bardenet, R. and Chainais, P. *Kernel interpolation with continuous volume sampling*. In : Proceedings of the 37th International Conference on Machine Learning, pp. 725–735, 2020
- [5] Belhadji, A. (2021). *An analysis of Ermakov-Zolotukhin quadrature using kernels*. In : Proceedings of the 34th International Conference on Machine Learning, 2021
- [6] Hickernell, F. J. *A generalized discrepancy and quadrature error bound*. Mathematics of computation 67.221, pp. 299–322, 1998.
- [7] Bach, F. "On the equivalence between kernel quadrature rules and random feature expansions". In : The Journal of Machine Learning Research 18.1, pp. 714–751, 2017
- [8] Rezaei, A. and Gharan, S. O. A polynomial time MCMC method for sampling from continuous determinantal point processes. In International Conference on Machine Learning, pp. 5438–5447, 2019.