We thank the reviewers for their thorough and constructive reviews. As a general comment, we have now included the pointers given by R1 and R2 on general Bayesian quadrature (BQ), which we had indeed overlooked. In the manuscript, we meant « sequential BQ » as in (Huszar and Duvenaud, [20]). Using the suggested references, our approach is now introduced as randomized experimental designs for kernel quadrature, in the sense of Section 2.4.2 of (Briol et al., *Stat. Sci.* 2019); similarly, the DPPs used so far for numerical integration [5] are probabilistic relaxations of the classical Gaussian quadratures, themselves tightly connected to BQ (Karvonen and Särkkä, *MLSP* 2017). As requested by R2, we have also added deterministic grids and multivariate settings to our experiments, see Figure A and comments below.

R1: *optimal $\mathcal{O}(N^{-2s})$ rate for a deterministic point set can easily be established for BQ by taking a uniform grid [...] which rather raises the question of why one would want to use a random point set.*

Theorem 1 applies beyond the case of the uni-dimensional periodic Sobolev space, e.g., to Korobov spaces, to RKHSs on hyperspheres, to the space of band-limited functions restricted to an interval or to kernels defined over non-compact domains such as the Gaussian kernel on $\mathbb{R}^d$, etc. In particular, the last two examples correspond to exponentially decaying kernel eigenvalues. In these cases and unlike Sobolev, our bound is quite tight, as seen with the Gaussian kernel (see the notation paragraph for the general assumptions of Theorem 1). Beyond a new connection between DPPs and RHKSs, and as suggested by R1 and R2, we hope that the geometric arguments that we brought forward in our proofs can serve other approaches to BQ.

R2: *An obvious drawback of the use of uniform grids is that it suffers from the curse of dimensionality [...] whether the use of DPPs works for modestly large dimensional problems. This point might need a discussion*

We don't have conclusive theoretical arguments yet, but the way DPPs tie repulsive designs to the underlying RKHS may yield more meaningful bounds in $d > 1$ than fill-in distance arguments. In particular, we can expect explicit non-asymptotic bounds with smaller constants. Our manuscript shows that the key notion is the decay of the eigenvalues of the integration operator $\boldsymbol{\Sigma}$. The dependence of that spectrum on the dimension can be explicitly worked out in the Sobolev and Korobov cases, and in general for tensor product of RKHSs; see Appendix A in [3]. However, what this says about DPP-KQ will have to wait for tighter bounds on the quadrature error, which may be tough nuts; see next bullet. Our manuscript is only a first brick in that wall.

R1: *A rate $\mathcal{O}(N^{-2s})$ for the MSE was established for BMC in [BOGOS2019]*

We now highlight this result in the manuscript. As commented above, our generic bound is indeed not as tight in the Sobolev case. Meanwhile, our experiments suggest that the rate $\mathcal{O}(N^{-2s})$ holds for DPP-KQ, and that the bound is representative of the behavior of the error even for small N, while Bach's LVSQ (with $\lambda = 0$) needs to wait for large values of N for the error to actually fall down at that rate (see our Figure 1, and simulations in [3]). A potential way to tighten our bound when kernel eigenvalues decrease only polynomially, as in the Sobolev case, is discussed in Section 4.2. We are currently investigating this and trying to replace the term Nr_N by r_N in Theorem 1. We even conjecture a bound that only involves eigenvalue σ_{N+1}. This is illustrated in the new experiment in Figure A. This figure also illustrates multivariate integration and a uniform grid as required by R2. The worst-case error of DPPKQ for $g \equiv 1$ (blue) scales as σ_{N+1} (green), better than the sum r_N of all eigenvalues above σ_N (orange). We observed the same fast scaling for the Gaussian case when $d \geq 2$ (not shown). Such an improvement in our bound would propagate to all RKHSs; this generality is the strength of DPP-based experimental design.

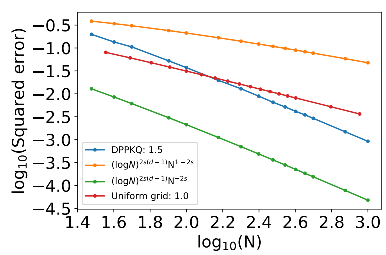

Figure A: Example of an additional experiment for the case of a multivariate Korobov space, $d = 2$, $s = 1$ and $g \equiv 1$.

R3: *not clear how to sample from the DPP if the eigenfunctions e_n's are inaccessible (...) same problem as in [3]*

We stress that even when the eigenfunctions e_n are accessible, it is not obvious how to sample from the regularized leverage-score distribution q_λ^* of [3]; see our Eqn (6). Indeed, the RHS of (6) is an infinite sum. On the contrary, our projection DPP only involves eigenfunctions up to index N, and the conditionals in the chain rule can thus be computed. However, we agree that when the eigendecomposition of the kernel is not available, exact sampling from the DPP seems out of reach. We would then rely on MCMC like, e.g., (Chafai and Ferré, Arxiv:1806.05985).

R3: *There is no intuition why a DPP with that particular repulsion kernel is better than other sampling schemes.*

We have added both geometric and probabilistic intuition to the text. We sketch here the latter. First, it is natural to take a repulsion kernel $\mathfrak{K}$ that is tied to the RKHS kernel k: the smoother the integrand is in one area, the more repulsive the quadrature nodes can be in that area without contributing much quadrature error. Second, while it is theoretically possible to define a DPP with repulsion kernel proportional to k [15], the resulting DPP would be a mixture of projection DPPs. The component with the highest weight in that mixture would be precisely the DPP we take in the paper. This intuitively suggests a variance reduction.

R3: *Explain the empirical results in Figure 1: what exactly is being plotted.*

For each number N of design points and each method, we draw 50 independent designs (x_i), compute the weights (w_i), and we report the average of $\|\mu_g - \sum_1^N w_i k(x_i, \cdot)\|^2$. In both Sections 5.1 and 5.2, μ_g is available in closed form. We have now clarified this in the manuscript, and added implementation details about sampling.