We would like to sincerely thank all the reviewers for reading the paper carefully and their very valuable feedback.

General comments:

Presentation: For the camera-ready, we will de-densify the main body of the paper to improve clarity (by moving technical results to the Appendix), fix all typos and incorporate all suggested notation improvements.

Legendre symbol: Current name is unfortunate (conflict with number theory version), thus we will change it. Inequality $\mathcal{L}_X(a) > 0$ follows directly from the properties of the Legendre **Transform** and we will clarify it in the final version.

Reviewer 1:

Relation to previous work: We will add detailed discussion regarding [1], [2] (and other works on energy-minimization techniques) as well as [3] in the final version. As opposed to [1] and [2]: we consider different energy function, evaluate on kernels beyond Gaussian and arc-cosine classes, target extra applications beyond kernels and propose simpler OPT-NOMC algorithm. OPT-NOMC is not our main contribution, is not even our only NOMC version. ALG-NOMC is unrelated to energy-based methods. Lemma 3 follows from Weil's results (explicitly stated in the text) thus it is very interesting, but not surprising that it implied also by [3] which exploits them too, yet in a very different setting (recovering sparse multivariate trigonometric polynomials). Finally, **the main contribution of this work are our theoretical results on ND which preceed Sec. 4**. **L.596:** This is for continuous gradient since $\mathcal{M}$ is compact. General cases are handled by finite difference approach. We will clarify it in the final version (see also: comment in l.598-599.)

Reviewer 2:

Definition of $f_{\mathcal{Z}}$: $\mathcal{Z}$ refers to ordered subset of $\mathbb{R}^d$, thus we should have: $\mathcal{Z} \subseteq \mathbb{R}^d$, thank you for catching this typo.

Functions $f^{+/-}$ / $\mathrm{even}[f]^{+/-}$: $+/-$ corresponds to "increasing/decreasing" in $|u|$". Functions $f^{+/-}$ generally are not uniquely defined, but in most applications will simply relate to positive/negative part of the Taylor series (TS) for f (if TS exists, both parts are finite and TS does not contain odd-power terms). Prominent examples include Gaussian and Matern kernel, see: Table 1 where we also put formulae on f^+ and f^-. Here $f^{+/-}$ can be directly obtained from TS. Similar analysis is true for $\mathrm{even}[f]^{+/-}$ (see: Table 1). Now we only need to filter out from TS odd-power terms.

Discontinuous f: we need measurability, but no other properties are required.

Points a and b: **a** and **b**, these are simply: the lower and upper bound on f. The second formula for $p(\epsilon)$ is for the case of bounded f and the first one is for the case of unbounded f. **We will clarify all these in the final version.**

Reviewer 3:

Connection btw theory and NOMC: Thank you for the comment. OPT/ALG-NOMC algorithms were inspired by negative dependence results since some of the classic examples of negatively-dependent systems come from energy-based configurations (statistical Physics) and algebraic theory. We will properly describe this link in the final version.

ALG-NOMC: Eq. 8-9 uniquely define the ensemble of samples and consequently, completely determine the algorithm. Vectors from Ω are defined as: $(a_1, b_1, ..., a_p, b_p)$ (see: Eq. 9), where a_j, b_j are: real and imaginary part of $g_{c_1,...,c_r}(j-1)$ (see: Eq, 8 for the definition of $g_{c_1,...,c_r}$). Different vectors correspond to different $(c_1, ..., c_r) \in (\mathbb{F}_p)^r$. This compact construction does not require any optimization, thus it was not presented originally in the separate algorithmic box. However we do agree with the reviewer that, since the content is very technical, for the clarity of the exposition, it would benefit from separate algorithm box and more careful explanation. We will do it in the camera-ready version.

Lemma 3 in Sec. 4.2: The proof follows directly from the Weil conjecture and related construction is one of the flagship examples from algebraic number theory of the close-to-optimal ensemble (with respect to size, which is what Lemma 3 says) with Kabatjanskii-Levenstein Lemma establishing tight upper bound. In the camera-ready, we will clarify it and add all details since Weil conjecture and its implications deserve a separate paragraph.

Functions satisfying properties: $F1$-$F3$: Table 1 provides several examples of functions from classes $F1$-$F3$.

Hyperparameters: We chose $\eta = 1.0, \delta = 0.1, T = 50000$ (thus did not tune them) as stated at the beginning of Appendix B. In the main body in l.294 we explain that: "Additional experimental details are in the Appendix."

Downstream experiments: We run approximate GP regression from "*The Geometry of Random Features*" (Fig.8) for Gaussian kernel and $s/d = 1, 2, 3, 4$, obtaining: OPT-NOMC: **0.5**, **0.4**, **0.36**, **0.31**, BEST: **0.54**, **0.44**, **0.43**, **0.39** (average test RMSE), where BEST is the best result across methods considered there (will be added to final version).

Reviewer 4:

Isotropic distribution: An isotropic distribution $\mathcal{D}_{\mathrm{iso}} \in \mathcal{P}(\mathbb{R}^d)$ is defined as having pdf constant on every sphere centered at 0. We will explicitly state it in the final version. Thus, by definition, if $\mathbf{v} \sim \mathcal{D}_{\mathrm{iso}}$, then $\mathbf{v}$ can be rewritten as: $v = r * \mathbf{u}$, where $\mathbf{u} \sim \mathrm{Unif}(\mathcal{S}^{d-1})$ and r is an independent random variable defining how to renormalize length of $\mathbf{u}$ to get $\mathbf{v}$ (see also: *The Geometry of Random Features*, AISTATS 2018).

Time Complexity: As explained in l.261-265, apart from one-time extra cost of the optimization (and only for OPT-NOMC method, see: Sec. 7.8, where we run detailed ablation studies over d for that wall clock time), time complexity is **the same**. Cost of a true random rotation is cubic in d, but regular OMC method also applies random rotations.

Miscellaneous: Lemma 1,2 are new. **Misleading claim near Fig. 1**: We will clarify that claim is for large d. **Non-isotropic distributions:** it is an interesting topic, but beyond the scope of this work since OMCs were designed for isotropic distributions. **Colors in Fig. 1** indicate direction of the vector (red: head of the vector, blue: tail). **MSE:** The MSE (mean squared error) is defined where we use it the first time, i.e. at the beginning of Theorem 1. See also, our response to Reviewer 1 regarding relation to **previous work** (minimizing an energy function on the hyper-sphere).