Response to Reviewer

D.B. Rokhlin, O.V. Gurtovaya

We would like to thank the Reviewer for the careful reading of our manuscript and for the constructive comments, which have helped us to improve the quality and clarity of the paper. Below we provide a point-by-point response to the suggestions.

Comment 1: *VAW is used both as the name of proposed algorithm and as a acronym for Vovk-Azoury-Warmuth algorithm. I suggest changing the name of algorithm or to not use VAW as a short naming for Vovk-Azoury-Warmuth method.*

Response: Regarding the name of the algorithm (VAW^2: "VAW squared"), we would prefer to retain it. This notation has already been used in the defended PhD thesis of the second author (O.V. Gurtovaya), and we aim to maintain consistency with this primary source.

However, to address a possible confusion, we have revised the manuscript to enforce a clear distinction. Throughout the text, we now consistently use the qualifier "standard" when referring to the original Vovk-Azoury-Warmuth algorithm (e.g., "the standard VAW method is used...").

Comment 2: *In the abstract: ... Pls check, also "method methods".*

Response: We apologize for this typo. We have corrected "method methods" to "methods".

Comment 3: *In the Introduction part: I suggest adding more information about functional approximation strategy. This is a general suggestion: since journal do not expect that every reader is a specialist in online methods, RKHS etc, pls add more information in text, at least briefly.*

Response: We have accepted this suggestion. To make the paper more accessible to a broader audience, we have revised the Introduction (Section 1) in two ways:

1. We added a brief intuitive explanation of RKHS methods in the first paragraph, clarifying how they enable linear algorithms to solve non-linear problems.

2. We expanded the description of the functional approximation strategy, explicitly mentioning that this class of methods encompasses techniques such as the Nyström method and Random Fourier Features (RFF). We also clarified that our work specifically focuses on the RFF approach, which constructs an explicit feature map into a finite-dimensional space.

Comment 4: *The Introduction lacks information about this setting. It could be both in the form of related works regarding this setting or the mention that this is a new bound and there is not previous known algorithms for this setting.*

Response: We have revised the Introduction to clarify the novelty of our theoretical results relative to the existing literature. Specifically:

- As previously, we acknowledge that the $O(T^{1/2})$ regret bound for the single-kernel case was established by Vovk (2006) [22].

- We emphasize that our main contribution is extending this result to the *multi-kernel* setting using *computationally feasible* algorithms based on Random Fourier Features. This allows us to avoid the linear growth of the dictionary size inherent in the standard kernel methods analyzed in classical works, achieving a bound of $O(T^{1/2} \ln T)$ with fixed computational complexity per iteration.

- We highlight that, unlike general gradient-based multi-kernel frameworks that typically require the global Lipschitz condition (which is restrictive for the unbounded square loss), our analysis leverages the structure of the square loss to establish bounds without this assumption.

Comment 5: *Preliminaries: I really think that it would be beneficial to add a little more intuition behind RKHS and notation. For example, it is not clear what is $\alpha(\theta)$, $\phi(x, \theta)$ and how they related to k (is k defined by them?). Missed reference (??) at the last sentence of this Section.*

Response: We have revised Section 2 to to provide more intuition:

- We clarified that the integral representation can be viewed as a continuous generalization of a linear model, where $\phi(x, \theta)$ serves as a basic feature and $\alpha(\theta)$ acts as a weight function.

- We explicitly interpreted the kernel function $k(x, y)$ as the expected product (or correlation) of the random features extracted from x and y, thereby clarifying its relationship to the feature map.

- We removed the sentence containing the missing reference mentioned by the Reviewer.

Comment 6: *Main results: I struggle to find the definition of Y from (10), (11), (16) etc. Probably it is in some references, but I believe that all definitions that are used in main results should be properly defined in the text. Same hold for R. Pls checks other parts of the paper.*

Response: We have added an explicit paragraph immediately preceding Theorem 1 in Section 3. Specifically:

- We introduced Y as an upper bound on the absolute value of the outcomes ($|y_t| \le Y$).

- We specified that the regret bounds are established with respect to functions f in the RKHS with norm bounded by R ($||f||_{\mathcal{H}} \le R$).

We also corrected several typos.

Sincerely,
D.B. Rokhlin, O.V. Gurtovaya