We thank all the reviewers for the detailed feedback.

R1 We thank the reviewer for appreciating our contributions. We will restructure the introduction as the reviewer suggested.

» *Comparison to [3]:* We enumerate the comprehensive list of differences of this work when compared to [3] for quick reference.

- We work with a strictly general path varaiational that promotes piecewise polynomial structure in the comparator sequence. The path variational in [3] promotes piecewise constant structures.
- By exploiting connections to regression splines, we formulate a more general restarting rule than [3].
- We demonstrate that zero padding (and many other padding approaches) prior to computing wavelet transform as done in [3] will not preserve the higher order total variation, thus lead to *far sub-optimal* results for the current problem. We then propose a novel packing scheme to alleviate this.
- We exploit the structure of CDJV wavelets and present a significantly more involved analysis to obtain *sharper* dynamic regret guarantees. Haar wavelets that worked in [3], did not work here.
- We characterise the optimality of our algorithm for the case of exact sparsity as done in section 5.2 which was not studied in [3]. Sharper dynamic regret guarantees for higher order discrete Sobolev and Holder classes are also obtained.
- We extend the framework to prediction in higher dimensions (Remark 6). We identify a class of loss functions other than squared error losses in which the dynamic regret guarantees of `Ada-VAW` still holds (Remark 7). Rationale behind both of these arguments can be found at the end of Appendix C.2.

R2 We thank the reviewer for appreciating the order optimality of our results. We agree that the readers can benefit from style of exposition that the reviewer suggested and we promise to incorporate it in the main paper. Please also see the comment to **R1** for a comparison to [3].

» *Comparator sequence:* We note that there are two popular notions of dynamic regret studied in literature as follows.

$$R_{\text{dyn-besbes}}\left(x_1, \ldots, x_n, f_1, \ldots, f_n\right) = \sum_{t=1}^{n} f_t(x_t) - f_t(u_t^*),$$

where u_t^* is the minimizer of $f_t(x)$ and

$$R_{\text{dyn-zinkevich}}\left(x_1, \ldots, x_n, f_1, \ldots, f_n, u_1, \ldots, u_n\right) = \sum_{t=1}^{n} f_t(x_t) - f_t(u_t),$$

where $u_1, \ldots, u_n$ is any arbitrary sequence. $R_{\text{dyn-zinkevich}}$ is the object of study in [1] where they consider designing algorithms with dynamic regret guarantees as a function of the path length of the comparator sequence. Of-course with this more general notion of regret, there is no notion of a fixed comparator sequence.

In our work, we consider $R_{\text{dyn-besbes}}$ as done in [2]. We note that when f_t are strongly convex, the comparator sequence $u_1^*, \ldots, u_n^*$ is unique and well defined. For our problem, $f_t(x) = (x - \boldsymbol{\theta}_{1:n}[t])^2$ and $\boldsymbol{\theta}_{1:n}$ is TV^k bounded as in Assumption A3.

R3 We thank the reviewer for appreciating the fine technical aspects of our work.

» *Other losses:* When the loss functions satisfy the conditions of Remark 7, we can still get dynamic regret guarantees for `Ada-VAW`. However, deriving dynamic regret bounds for more general losses with TV^k bounded comparators (or more generally, comparators that belong to Besov space) is a challenging future work.

» *Lower bound:* We would like to mention that Proposition 11 holds for all $k \geq 0$. There is a typo at line 799. It should be $\|D^{k+1}\boldsymbol{\theta}_{1:n}\|_0 \leq J_n$.

» *Other:* We will include the high-level proof ideas in Section 4.3 as the reviewer suggested.

References

[1] *Online convex programming and generalized infinitesimal gradient ascent*, Zinkevich, ICML 2003

[2] *Non-stationary stochastic optimization*, Besbes et al, In Operations Research 2015

[3] *Online Forecasting of Total-Variation-bounded Sequences*, Baby and Wang, NeurIPS 2019