We thank all reviewers for their detailed feedback. We will be sure to address all questions and incorporate all suggestions from the reviewers in the final version of the paper. **Note: Reference numbers below refer to the main submission, not the supplementary version.** To reiterate and clarify, our contributions include the following:

1. We reduce contextual linear bandits with infinite action sets to a regression problem with an algorithm that is both practical and efficient.
2. We optimally adapt to an unknown level of misspecification, which is a non-trivial open problem [27]. Recall that previous works in contextual bandits required oracle knowledge of the misspecification level in order to tune the algorithm's parameters (e.g., the upper confidence bound in LinUCB). Let us emphasize that neither doubling tricks nor other methods were known to circumvent this requirement. The solution for static action sets heavily relies on elimination, and does not generalize to the contextual case [27].
3. We adapt to a compelling notion of sparsity defined by an average effective dimension.
4. We provide a novel view of corralling bandits and give an improved master algorithm.

**Questions common to multiple reviewers**

- *Optimization problem:* Several valid questions were raised regarding solving the optimization problem in Def. 7 and the associated computational cost. This is a convex optimization problem over a convex set with easily computable gradients. Finding an $\epsilon$-approximation takes $\mathcal{O}(\mathrm{Poly}(d)\log(1/\epsilon))$ time. (see "Relatively-Smooth Convex Optimization by First-Order Methods, and Applications" (Hu, Freund, Nesterov), Theorem 3.1). See also the optimal design example in Sect. 2.2 therein which, up to a linear term and a benign term in the Hessian, is equivalent to our problem. At every inner iteration, the solver needs to find $\mathrm{argmax}_{a\in\mathcal{A}_t}\langle a,\mu\rangle + \beta||a||^2_{H^{-1}}$, where $H^{-1}$ can be updated in $O(d^2)$ time, while finding the argmax is the same problem that the standard LinUCB algorithm is solving. We will add a formal proof and discussion along these lines to the paper.
- *Experiments:* We agree that experiments on real-world or synthetic data would be a bonus here, but we believe that our strong theoretical results stand for themselves.

**Reviewer 1**

- *Foster and Rakhlin show ... can you get $min(K,d)$ ?* The effective dimension we use in Theorem 13 (L. 249) is upper bounded by $K$, since the linear subspace spanned by the feature vectors of all arms trivially includes the action set. In fact, when $K$ is equal to the effective dimension, then the logdet barrier and the logbarrier coincide.

**Reviewer 2**

- *On logdet-barrier being a proper distribution:* This holds by definition, because the optimization problem in Def. 7 is over the probability simplex.
- *Adapting SquareCB to the misspecification setting is non-trivial:* Briefly, the optimal setting for the parameter $\gamma$ in SquareCB (see Theorem 5/6 in [21]) depends on $\varepsilon$. Any choice for $\gamma$ that ignores $\varepsilon$ leads to suboptimal regret (e.g., using the optimal choice for $\varepsilon = 0$ leads to regret scaling as $\varepsilon^2 T^{3/2}$ when $\varepsilon \neq 0$). Hence, the purpose of the master algorithm is to learn a near-optimal choice for $\gamma$ on-the-fly. See also Item 2 at the top of this page.
- *On assumptions in Theorem 10:* Theorem 10 is stated and proven for general function classes $f \in \mathcal{F}, f : \mathcal{X} \to \mathbb{R}^d$, we don't see any inconsistency with LL. 486-487 (note that the $\theta^\star_t$ notation in this proof is just shorthand for $f^\star(x_t)$).
- *Line 487, how to obtain the 2nd eq. from the 1st one:* This follows by adding and subtracting terms, and then applying the triangle inequality to term differences.

**Reviewer 3**

- *This is a little different ... the small deviation case of [22]:* These results are not comparable because their work only considers fixed action sets. We are not aware of a suitable definition of "small deviation" for the contextual case with changing action sets. We agree that it is an interesting direction for future research.
- *Regret bound still linear in $T$... :* There is a tight lower bound (see, e.g., [27]) showing that the $\varepsilon\sqrt{d}T$ term is generally unavoidable, even if $\varepsilon$ is known beforehand. Nonetheless, notice that our bounds rely on the *empirical* quantity $\varepsilon_T \le \varepsilon$ and, in practice, one may hope for an $\varepsilon_T$ of order $T^{-\alpha}$, for some positive $\alpha$, leading to no-regret results.

**Reviewer 4**

- *The statement of Theorem 12 contains an additional $\sqrt{d}$ ... seems to be a typo:* Indeed, thanks for spotting this!
- *On Assumption 1 in other settings other than linear, e.g., with kernels:* In a kernelized setting, one could use kernelized Online Gradient Descent as a regression oracle, which is dimension-independent (scaling instead with the RKHS norm) but has a $T^{\frac{1}{2}}$ regret rather than $d\log(T)$. On the other hand, one can always rely on the standard kernel online ridge regression regret bound, that replaces bound $d\log(T)$ by the log determinant of the kernel Gram matrix of the data, and then rely on the speed of eigenvalue decay of this matrix.
- *Misspecification in the case of universal kernels such as Gaussian:* With sufficiently small bandwidth, a universal kernel can be realizable, i.e. $\varepsilon = 0$. Yet, choosing small bandwidth comes at a cost of increasing sample complexity, and the optimal results for a particular problem instance may be obtained by trading off kernel bandwidth versus misspecification.