We thank the reviewers for the feedback. We reply to the outstanding points below.

REVIEWER 1.

R. *It's weak novelty-wise, the effectiveness of task-conditional parameterizations have already shown promise in other works.* A. We are not aware of other papers (nor the reviewer is providing any reference) proposing an efficient meta-learning method addressing heterogeneous environments of tasks which is computationally efficient (i.e. based on a convex method) and supported by learning guarantees. Note that the method we propose and analyze is based on similar ideas already used in literature, but it is a new method designed by us. All the related works we know are cited in the paper, but we are happy to discuss any other related paper we missed. R. *The abstract suffers from clarity problems.* A. Note that we described fine tuning and biased regularization as meta-learning strategies (not algorithms). As conventional in literature, we denoted by task's target vector the minimum norm weight vector minimizing the true risk associated with that task. We will clarify this. R. *In the introduction there isn't much motivation on meta-learning.* A. We decided to briefly recall the well-known intuition on which standard meta-learning is based on in order to mainly focus on the less known conditional setting, the main topic of the paper. R. *You fail to explain how your method differs from the other conditional meta-learning methods.* As explained in the paper, the related works we know consider different formulations of the problem and they do not provide a complete theoretical analysis for the methods they propose. Because of this, drawing a complete comparison between our method and the preexisting literature is not always easy. We will add more details about this comparison where possible.

REVIEWER 2.

R. *The proposed method is restricted to the family of linear regression models.* A. Note that a feature map can be also added in the within-task problem. In our experimental settings and in many other scenarios in literature (such as [5,13,14,22] in our paper), the use of linear models with the right feature map revealed to be sufficiently effective. Extending our method and the corresponding analysis to the non-convex setting is for sure an interesting direction we leave for future research. R. *The choice of the hyper-parameters λ and γ depends on unknown quantities.* A. In practice, we validated the hyper-parameters. We are confident about the fact that a more efficient parameter-free version of our method can be developed by using arguments based on coin betting algorithms (see [12, 30] in our paper). We will investigate this in the future. This however would not change the main message that we want to convey in the paper: developing a meta-learning method for heterogeneous tasks and providing theoretical guarantees for it. R. *One may choose to augment the training data with the side information and use a constant initialization. Is there an optimal split?* A. In experiments we observed that even though we augmented the training datapoints, the standard unconditional approach performed worse than the conditional one. We guess that the independence between the side information and the training dataset can be removed, by using alternative generalization arguments, but this requires further investigation. The optimal splitting size is an interesting point, still open. R. *Regarding Eq. (25) and (26).* A. Eq. (25) and Eq. (26) are the state-of-the-art bounds for optimal unconditional meta-learning and independent task learning (ITL). Both these bounds are deduced from the general bound in Thm. 4 for our conditional meta-learning approach. This means that our method includes also unconditional meta-learning and ITL. We investigate the advantage of using the conditional approach with respect to the unconditional one and ITL in Sec. 3.

REVIEWER 5.

R. *A limitation of the existing meta-learning is the lack of scalability to apply real-world scenarios.* A. As described in the last appendix with experimental details, our fine tuning method scales linearly with the dimension of the features space and the dimension of the image space of the side information's feature map. To develop even more scalable methods is for sure an interesting future direction, but, for the moment, the priority was to develop a theoretically grounded method, something which is missing in literature. R. *About relation to prior work.* A. See reply to reviewer 1.

REVIEWER 6.

R. *It could cross the mind of the reader whether the proposed approach is worth considering if the performance is not compared with other approaches.* A. Please note that the main question we addressed in this work was to understand when the conditional meta-learning approach results to be advantageous with respect to the standard unconditional one and independent task learning. We thus think that adding to the comparison other methods would complicate the message of the paper.