We thank the reviewers for their valuable feedback. We will incorporate the suggestions on the paper write-up and organization in the final version.

Reviewer 1

Q1) "Also, I find line 151 slightly odd, in the definition of ρ^*. This is not a minimum norm interpolant. Moreover you are defining ρ in terms of the entire function c. Can you expand on how you compute ρ^*?"
A1) This is the standard definition based on least square interpolation. The formula for ρ^* is $\rho^* = G^{-1}C$, where G is the gram matrix of mapped source, target data in $H_1 \otimes H_2$ and C is the matrix of cost evaluations at the given source, target data points. We shall clarify this appropriately in the final draft.

Reviewer 2

Q2) "I can't see many weaknesses, apart from the complexity matters : I would be curious to have some insights on this subject."
A2) We discuss the computational complexity in the supplementary material (lines $239 - 243$ and $246 - 247$). We shall include it in the main paper's final version.

Reviewer 3

Q3) "This paper proposed to utilize the kernel embedding method to reduce the dimensionality of the statistical optimal transport (OT) problem."
A3) This work does not propose to perform dimensionality reduction. It explores the novel idea of posing the OT problem as that of learning the kernel mean embedding of the optimal transport plan/map from the given samples.

Q4) "The proposed kernel embedding formulation of OT is based on the cross-covariance operator which ignores higher moments distributions."
A4) No, the higher order moments are not ignored. Unlike cross-covariance, which is a number, what we use here is the (kernelized) cross-covariance operator between two RKHS. Infact, whenever moments or expectations exist, all of them can be calculated from the operator using formula (3.16) in $[23]$. Instantiations of this formula are also used in our paper at lines 104, 119 etc.

Q5) "The dimensionality of canonical feature maps can be infinity. This paper did not carefully discuss this issue".
A5) Indeed the dimensionality of RKHS is infinite. However, thanks to Theorem 2, the fact is that the sample complexity is still finite. Moreover, the representer theorem (Theorem 3) guarantees finite parameterization for the optimal solution. To summarize, the issue of the infinite dimensionality of the RKHS is dealt with thoroughly via Theorems 2 and 3.

Q6) "Does the proposed method need to make a finite-dimensional approximation? If so, how to selec the number of terms? Does it depend on the dimensionality of the data?"
A6) Nowhere in the paper we make a finite-dimensional approximation nor do we perform dimensionality reduction. Also, we prove in Theorem 2 that the sample complexity is completely independent of dimensionality of the data.

Q7) "The objective functions in (3) and (4) are based on the expectation of functions defined in line 104 which are not empirically available."
A7) We agree that the objective in (3) cannot be computed. However, the objective in (4) can be computed straightforwardly using the data as it involves empirical estimates alone. The interesting result in Theorem 2 shows that at optimality, (4) converges to (3) at a rate that is $O(1/\sqrt{\min(m,n)})$ and is dimension-free.

Q8) "The algorithm discussed in Section 3.4 is based on a couple of simplification conditions. It will be helpful to discuss the approximation error as well as the computational complexity of the algorithm."
A8) Yes, discussion in Section 3.4 is only for special cases ($\epsilon_i = 0$). If we wish to choose hyper-parameters that do not satisfy these conditions, i.e., if $\epsilon_i \neq 0$, then instead one can always solve the convex problem (5) using existing off-the-shelf solvers. So there will be no "approximation error" if (5) is solved directly. In supplementary Section 3.4, we provide details of computational complexity (lines $239 - 243$ and $246 - 247$).

Q9) "Simulations on Gaussian OT may not be sufficient to demonstrate the effectiveness of the proposed method"
A9) We agree. This is the reason we also demonstrate performance on real-world benchmark problems in domain adaptation application (Section 5 and Table 1 in the main paper, and Section 5.2 in the supplementary material).

Q10) "how to construct Σ_1 and Σ_2 in the multivariate Gaussian example?"
A10) Additional details of experiments are present in the supplementary material. For instance, lines 279-281 in the supplementary material discusses how to construct Σ_1 and Σ_2.