We thank the reviewers for their appreciative and thoughtful feedback.

Reviewer 1. ***"However, the authors fail to bring the result to their impact of the current state of OT [...[Does having this closed form solution provide better ways to design algorithms, faster algorithms, etc?"*** As discussed in L59-L73, the first obvious impact of our contribution is that it provides the first example for which regularized (unbalanced) OT admits a closed form expression. These formulas provide a testbed for any theoretical conjecture that tries to understand better entropic OT, or any novel stochastic optimization algorithm designed to compute it faster. Additionally, our formulas offer a principled solution to alleviate the differentiability issues of the Bures metric that arise for singular matrices (L.129-130). Finally, one can foresee that applications relying on entropic OT might benefit from some local Gaussian approximations to use these closed form, in the spirit of sliced Wasserstein approaches. We will further emphasize these aspects. ***"Line 110: Please define what is a centered measure."*** A measure with 0 mean. We will clarify this.

Reviewer 2. ***"If the paper could show the formula for that case [TV] that would be supercool. Though maybe there isn't a closed-form for that."*** A glance at the prox-div operator of TV (https://arxiv.org/pdf/1607.05816.pdf) shows that after the first iteration, the (log) dual variable would be the pointwise projection of a quadratic function over the box $[-\frac{\lambda}{\varepsilon}, \frac{\lambda}{\varepsilon}]$ which is not obvious to convolve with a Gaussian kernel. ***"[Transport plan experiment] How many samples do you "need" before you stop making huge mistakes? (...) as the ambient dimension explodes"*** To answer this question we computed the distance between the ground truth (formulas of Thm 3) and the empirical moments (μ_n, Σ_n) of the transport plan using random inputs and fixed parameters $\sigma = 0.01, \gamma = 0.2, m_\alpha = 1, m_\beta = 1.1$. See Fig.1 on left.

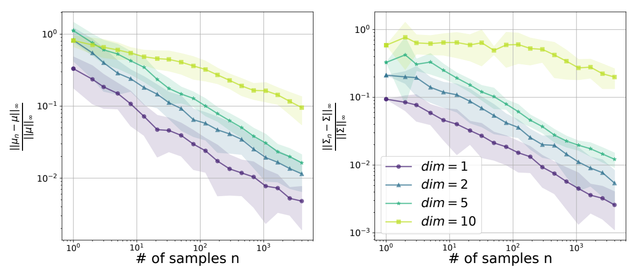

Figure 1: Large dimensions need more samples to approximate the moments of the unbalanced optimal transport plan.

Reviewer 3. ***""Figure 1 illustrates the convergence"... the convergence of what?"*** This convergence is in terms of number of samples from 2 Gaussian distributions: $\lim_{n\to+\infty} \mathrm{OT}_\sigma(\alpha_n, \beta_n) \to \mathrm{OT}_\sigma(\alpha, \beta)$. We will clarify this. ***"Figure 2 is also difficult to understand. Both source and target measures are sampled, and then Sinkhorn algorithm is used on a discretized measure?"*** Exactly, both measures are sampled, we run Sinkhorn to obtain an empirical transportation plan that we visualize by computing a histogram on a uniform 2D grid. ***"[On the proof of prop 2] Could the authors explain why the uniform bound is not necessary here?"*** Thank you for pointing this out, a uniform bound is indeed required and we will update our proof. From (42) and using Weyl's inequality, we can bound the smallest eigenvalue of $\mathbf{F}_n$ from below: $\forall n, \lambda_d(\mathbf{F}_n) \geq \frac{\sigma^2}{\lambda_1(\mathbf{A})}$ (where $\lambda_d(\mathbf{F})$ is the smallest eigenvalue of $\mathbf{F}$ and $\lambda_1(\mathbf{A})$ is the biggest eigenvalue of $\mathbf{A}$). Hence, the iterates live in $\mathcal{A} \stackrel{\text{def}}{=} \mathcal{S}_{++}^d \cap \{\mathbf{X} : \lambda_d(\mathbf{X}) \geq \frac{\sigma^2}{\lambda_1(\mathbf{A})}\}$. Finally, for all $\mathbf{X} \in \mathcal{A}$, $\|(\mathrm{Id} + \sigma^2 \mathbf{B}^{-\frac{1}{2}} \mathbf{X} \mathbf{B}^{-\frac{1}{2}})^{-1}\|_{\text{op}} = \frac{1}{\lambda_d(\mathrm{Id} + \sigma^2 \mathbf{B}^{-1/2}\mathbf{X}\mathbf{B}^{-1/2})} = \frac{1}{1 + \sigma^2 \lambda_d(\mathbf{B}^{-1/2}\mathbf{X}\mathbf{B}^{-1/2})} \leq \frac{1}{1 + \sigma^2 \lambda_d(\mathbf{B}^{-1})\lambda_d(\mathbf{X})} \leq \left(1 + \frac{\sigma^4}{\lambda_1(\mathbf{B})\lambda_1(\mathbf{A})}\right)^{-1}$. Which proves the uniform bound. ***"on l. 423, I think the authors could better explain why AB and C have same eigenvectors"*** Because AB is a quadratic polynomial of C (Eq 21). As explained in L157-158, C has positive eigenvalues, writing its EVD as: $C = QDQ^{-1}$ in eq (21) leads to: $Q(D^2 + \sigma^2 D)Q^{-1} = AB$ which is an EVD of AB. AB and C have the same eigenvectors Q. In the appendix, we wrote EVD of AB instead of C, we will correct this. ***"l. 201 "Moreover, the objective admits a lower bound if and only if ..." has not a clear meaning"*** This is the condition for the entropy KL to be finite. We will replace the previous statement. ***"Could the authors add a reference for the dual formulation of (9)? [...] is [1l] enough [...]?"*** We have cited [33], the context of [11] is more restrictive since they considered probability spaces (X, Y with reference measures having unit mass). ***"In Equation (6), [...] depends on the scalar product"*** the gradient is given for the Frobenius inner product. We will clarify this point. ***"- l. 149: The authors say that a sequence is contractive, but for which distance?"*** In the matrix operator norm, as per the proof of Proposition 2. We will explicitly state this.