Loss of performances in Figures 6, 8 (Reviewer #1). After investigation, those losses are due to hyperparameter choices for the non-convex WDA and OTDA problems. When appropriately selected for each model (decimation factor), we obtain running time gain of same orders without compromising performances. In the final version, we will add to the supplementary new figures related to the regularization path computations and resulting accuracies.

Comparison with other solvers (Reviewers #1 and #3). We have considered experiments with Greenkhorn algorithm but the implementation in POT library and our custom Python version of Matlab Altschuler's Greenkhorn code were not competitive with Sinkhorn. Hence, for both versions, Screenkhorn is more competitive than Greenkhorn. The computation time gain reaches an order of 30 when comparing our method with Greenkhorn while Screenkhorn is almost 2 times faster than Sinkhorn,We will provide this comparison and discussion in the final version.

On the use of constrained L-BFGS (Reviewers #2 and #3). Our proposed screened dual problem given in (3) or (6) involves explicit box constraints on $e^{u_i^{sc}}$ and $e^{v_i^{sc}}$ (see Proposition 1). Hence, it is a constrained smooth optimization problem, and standard Sinkhorn's alternating minimization can not be applied. This appears more clearly while writing its optimality conditions. We resort to L-BFGS-B to solve our constrained convex optimization problem, but any efficient solver (e.g. proximal based method or Newton method) can be used. Notice that as for the Sinkhorn algorithm, our Screenkhorn can be accelerated using a GPU implementation of L-BFGS-B [2].

Main concerns of Reviewer #2. *Concern 1.* The bound in Proposition 3 is similar, up to the additive term ω_κ (a discussion about ω_κ is provided in below), to the ones found in the literature; in particular for the Sinkhorn algorithm (see Lemma 2 in [1]) and for the Greenkhorn algorithm (see Corollary 3.3 in [4]). More formally, letting $\{(u^k, v^k)\}_{k\geq 1}$ denote the iterates returned by the Sinkhorn or the Greenkhorn algorithm, they have $\Psi(u^k, v^k) - \Psi(u^\star, v^\star) = \mathcal{O}(RE^k)$ where $E^k = ||B(u^k, v^k)\mathbf{1} - \mu||_1 + ||B(u^k, v^k)^\top \mathbf{1} - \nu||_1$, and $R = C_{max}/\eta + \log(n) - 2\log(c_{\mu\nu})$ which comes from an upper bound for the ℓ_∞-norm of the optimal pair solution $(u^\star, v^\star)$ of Sinkhorn divergence. In our case, supposing that $n = m$ and acknowledging that $\log(1/K_{min}^2) = 2C_{max}/\eta$, we have $R = C_{max}/\eta - 3.5\log(c_{\mu\nu})$. Additionally, we give in Proposition 2 a bound on E^k that becomes small as the sample budget increases.

Concern 2. The new formulation (3) has the form of $(\kappa\mu, \nu/\kappa)$-scaling problem under constraints on the variables u and v and the problem is not invariant anymore. This differs significantly from the standard scaling-problems [3], though the sought transportation map P takes a matrix-scaling form. We further emphasize that κ plays a key role (that we will emphasize in the final version) in our screening strategy for the dual of Sinkhorn divergence. Indeed, without κ, e^u and e^v can have inversely related scale that may lead in, for instance e^u being too large and e^v being too small, situation in which the screening test would apply only to coefficients of e^u or e^v and not for both of them. In addition note that given n, the bounds in Propositions 2 and 3 are derived using the following control of the parameter ε, which is induced by the screening test's construction (4), $c_{\mu\nu}^{1/4}/\sqrt{n} \leq \varepsilon \leq 1/\sqrt{nK_{\min}}$.

Concern 3. An explicit form of ω_κ (with $\omega_1 = 0$) is given in L449 of the paper. In the setting of $n = m$ and using the upper bounds of $||u^{sc}||_\infty$ and $||v^{sc}||_\infty$ in L447, we derive the following bound: $\omega_k \lesssim R'((1-\kappa)||\mu^{sc}||_1 + (1 - \kappa^{-1})||\nu^{sc}||_1)$ where $R' = C_{\max}/\eta - 0.5\log(n) - 0.5\log(c_{\mu\nu})$. A control for the ℓ_1-norms of the screened marginals μ^{sc} and ν^{sc} are given in Equations (18) and (19) in Lemma 3. Using the bound of the term ω_κ, we will clarify the bound in Proposition 3 for the final version of the paper.

Minor comments (all Reviewers). The final version of the paper will include all suggested modifications.

References

[1] P. Dvurechensky, A. Gasnikov, and A. Kroshnin. Computational optimal transport: Complexity by accelerated gradient descent is better than by Sinkhorn's algorithm. volume 80 of *PMLR*, pages 1367–1376, 2018.

[2] Y. Fei, G. Rong, B. Wang, and W. Wang. Parallel L-BFGS-B algorithm on GPU. *Computers & Graphics*, 40:1 – 9, 2014.

[3] B. Kalantari and L.Khachiyan. On the complexity of nonnegative-matrix scaling. *Linear Algebra and its Applications*, 240:87 – 103, 1996.

[4] T. Lin, N. Ho, and M. Jordan. On efficient optimal transport: An analysis of greedy and accelerated mirror descent algorithms. volume 97 of *PMLR*, pages 3982–3991, 2019.