ERRATUM: "ASYMPTOTIC DISTRIBUTION AND CONVERGENCE RATES OF STOCHASTIC ALGORITHMS FOR ENTROPIC OPTIMAL TRANSPORTATION BETWEEN PROBABILITY MEASURES"

BERNARD BERCU AND JÉRÉMIE BIGOT

Université de Bordeaux
Institut de Mathématiques de Bordeaux et CNRS (UMR 5251)

This erratum offers a correction of paper “Asymptotic distribution and convergence rates of stochastic algorithms for entropic optimal transportation between probability measures” by B. Bercu and J. Bigot posted on Arxiv at: https://arxiv.org/abs/1812.09150v5, and that has been published in Annals of Statistics 49(2): 968-987 (2021) [1].

In Appendix A of paper [1], inequality (A.4) in Lemma A.1 states that

\[
\rho_{A_\varepsilon}(v^*) \geq \frac{1}{\varepsilon} \min_{1 \leq j \leq J} \nu_j.
\]

However, this lower bound for \(\rho_{A_\varepsilon}(v^*)\) is not correct. Using arguments in the proof of Lemma A.1, the right expression for this lower bound is

\[
\rho_{A_\varepsilon}(v^*) \geq \frac{1}{\varepsilon} \mathbb{E}\left[ \min_{1 \leq j \leq J} \pi_j(X, v^*) \right]
\]

where \(v^*\) is the optimal dual Kantorovich potential and \(\pi_j(X, v^*)\) defined by (3.3) in paper [1], satisfies for all \(1 \leq j \leq J\), \(\nu_j = \mathbb{E}[\pi_j(X, v^*)]\).

We thank Lenaïc Chizat and Alex Delalande for pointing this out to us. Changing this lower bound does not affect the results of the paper as they have been obtained for a fixed value of \(\varepsilon > 0\). Moreover, the proofs of the paper do not use the lower bound (A.4).

PROOF OF THE LOWER BOUND IN LEMMA A.1

For any \(v \in \mathbb{R}^J\), let \(A_\varepsilon(v)\) be the negative semi-definite matrix defined by

\[
A_\varepsilon(v) = -\frac{1}{\varepsilon} \mathbb{E}\left[ A_\varepsilon(X, v) \right]
\]

where, for all \(x \in \mathcal{X}\),

\[
A_\varepsilon(x, v) = \text{diag}(\pi(x, v)) - \pi(x, v)\pi(x, v)^T,
\]

with \(\pi_j(x, v)\) defined by (3.3) in paper [1]. We denote by \(\lambda_j^\varepsilon(v), \ldots, \lambda_J^\varepsilon(v)\) the real eigenvalues of the matrix \(A_\varepsilon(v)\) that is of rank \(J - 1\) with \(\lambda_j^\varepsilon(v) = 0\) and \(\lambda_j^\varepsilon(v) < 0\) for all \(1 \leq j \leq J - 1\). We also recall that \(\mathbf{v}_J = \frac{1}{\sqrt{J}} \mathbf{1}_J\) is the unit eigenvector associated to \(\lambda_J^\varepsilon(v)\) and we denote by \(\langle \mathbf{v}_J \rangle\) the one-dimensional subspace of \(\mathbb{R}^J\) spanned by \(\mathbf{v}_J\). Hereafter, choose \(v = v^*\) and let

\[
\rho_{A_\varepsilon}(v^*) = -\max_{1 \leq j \leq J - 1} \left\{ \lambda_j^\varepsilon(v^*) \right\} = \min_{1 \leq j \leq J - 1} \left\{ -\lambda_j^\varepsilon(v^*) \right\}.
\]
We clearly have
\[ \rho_{\lambda_1}(\nu^*) = \min_{\{\nu \in (v_J)^\perp : \|\nu\|=1\}} -u^TA_{\varepsilon}(\nu^*)u = \frac{1}{\varepsilon} \min_{\{\nu \in (v_J)^\perp : \|\nu\|=1\}} u^T\mathbb{E}[A_{\varepsilon}(X, \nu^*)]u. \]

However, it follows from Theorem 6 and inequality (5.11) in [5] that the second smallest eigenvalue of \( A_{\varepsilon}(x, \nu^*) \) is lower bounded by \( \min_{1 \leq j \leq J} \pi_j(x, \nu^*) \) for all \( x \in \mathcal{X} \). Moreover, \( A_{\varepsilon}(x, \nu^*) \) is a positive semi-definite matrix of rank \( J-1 \), and its eigenvector associated to its smallest eigenvalue (equal to zero) is \( v_J \) for any \( x \in \mathcal{X} \). Consequently, for all unit vector \( u \in (v_J)^\perp \),
\[ u^T\mathbb{E}[A_{\varepsilon}(X, \nu^*)]u \geq \mathbb{E}\left[ \min_{1 \leq j \leq J} \pi_j(X, \nu^*) \right] \]
which leads to
\[ \rho_{\lambda_1}(\nu^*) \geq \frac{1}{\varepsilon} \mathbb{E}\left[ \min_{1 \leq j \leq J} \pi_j(X, \nu^*) \right]. \]

ILLUSTRATIVE EXAMPLE

We conclude this erratum by providing an illustrative example suggested by Alex Delalande for which one may obtain explicit expressions of both \( \rho_{\lambda_1}(\nu^*) \) and the lower bound (A.4) as functions of \( \varepsilon \) that allows to study their convergence as \( \varepsilon \) goes to zero. Assume that the dimension \( d = 1 \), that \( \mu \) is the uniform distribution on \([-1/2, 1/2]\) and that
\[ \nu = \frac{1}{2} \delta_{y_1} + \frac{1}{2} \delta_{y_2} \quad \text{with} \quad y_1 = -1, y_2 = 1. \]
Since \( \mu([-\infty, 0]) = \mu([0, +\infty[) = \frac{1}{2} \), it follows by symmetry of \( \mu \) and \( \nu \) that
\[ \nu^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \]
Hence, if one considers the cost function \( c(x, y) = -xy \), one obtains that
\[ \pi_1(x, \nu^*) = \frac{e^{-x/\varepsilon}}{e^{-x/\varepsilon} + e^{x/\varepsilon}} \quad \text{and} \quad \pi_2(x, \nu^*) = \frac{e^{x/\varepsilon}}{e^{-x/\varepsilon} + e^{x/\varepsilon}}. \]
Consequently,
\[ A_{\varepsilon}(x, \nu^*) = \begin{pmatrix} \pi_1(x, \nu^*)(1 - \pi_1(x, \nu^*)) & -\pi_1(x, \nu^*)\pi_2(x, \nu^*) \\ -\pi_1(x, \nu^*)\pi_2(x, \nu^*) & \pi_2(x, \nu^*)(1 - \pi_2(x, \nu^*)) \end{pmatrix}. \]

Hence, we obtain that
\[ \mathbb{E}[A_{\varepsilon}(X, \nu^*)] = \varepsilon a_{\varepsilon} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \]
where
\[ a_{\varepsilon} = \frac{1}{\varepsilon} \int_{-1/2}^{1/2} \frac{1}{(e^{-x/\varepsilon} + e^{x/\varepsilon})^2} \, dx = \int_{-1/2}^{1/2} \frac{1}{(e^{-y} + e^y)^2} \, dy = \frac{1}{2} \left[ \frac{1}{1 + e^{2y}} \right]_{-1/2}^{1/2} = \frac{1}{2} \left( \frac{1}{1 + e^{1/\varepsilon}} - \frac{1}{1 + e^{-1/\varepsilon}} \right). \]
Therefore, we find that
\[ A_{\varepsilon}(\nu^*) = -a_{\varepsilon} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \]
and $\rho_{A_{\varepsilon}}(v^{*}) = 2a_{\varepsilon}$. In addition,
\[
\lim_{\varepsilon \to 0} a_{\varepsilon} = \frac{1}{2} \quad \text{and} \quad \lim_{\varepsilon \to 0} \rho_{A_{\varepsilon}}(v^{*}) = 1.
\]
Finally, we have
\[
\mathbb{E}[\min(\pi_{1}(X, v^{*}), \pi_{2}(X, v^{*}))] = \mathbb{E}[\pi_{2}(X, v^{*})I_{X<0} + \pi_{1}(X, v^{*})I_{X>0}] = \varepsilon b_{\varepsilon}
\]
where
\[
b_{\varepsilon} = \frac{1}{\varepsilon} \left( \int_{-1/2}^{0} e^{x/\varepsilon} dx + \int_{0}^{1/2} e^{-x/\varepsilon} dx \right)
\]
\[
= \int_{-1/2}^{0} e^{y} e^{-y + e^{y}} dy + \int_{0}^{1/2} e^{-y} e^{-y + e^{y}} dy
\]
\[
= \left[ \frac{1}{2} \log(1 + e^{2y}) \right]_{-1/2}^{0} + \left[ y - \frac{1}{2} \log(1 + e^{2y}) \right]_{0}^{1/2} = \log 2 - \log(1 + e^{-1/\varepsilon}).
\]
We have
\[
\lim_{\varepsilon \to 0} b_{\varepsilon} = \log 2.
\]
Moreover,
\[
\rho_{A_{\varepsilon}}(v^{*}) \geq \mathbb{E}[\min(\pi_{1}(X, v^{*}), \pi_{2}(X, v^{*}))] \iff 2a_{\varepsilon} \geq b_{\varepsilon}
\]
which is of course true for all $\varepsilon > 0$. Therefore, the above calculation shows that $\rho_{A_{\varepsilon}}(v^{*})$ and the lower bound (A.4) remain finite as $\varepsilon$ goes to zero. These findings are in agreement with recent research works [3][Theorem 3.2] and [2][Proposition 5.1] on the computation of lower bounds for $\rho_{A_{\varepsilon}}(v^{*})$ in the semi-discrete setting when $\mu$ is an absolutely continuous probability measure with an upper and lower bounded density $f_{\mu}$. Using the fact that $\lim_{\varepsilon \to 0} \rho_{A_{\varepsilon}}(v^{*}) = 1$, we can guess from the above computations that the un-regularized semi-dual OT functional $H_{0}(v)$, defined in Equation (2.10) of paper [1], should be twice differentiable at its optimiser $v^{*}$ with Hessian given by
\[
A_{0}(v^{*}) = -\frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.
\]
To confirm this guess, we can use [4][Theorem 1.3] with $d = 1$ that ensures that, for $i \neq j \in \{1, 2\}$,
\[
\frac{\partial^{2}}{\partial v_{i} \partial v_{j}} H_{0}(v^{*}) = \int_{T^{-1}(y_{i}) \cap T^{-1}(y_{j})} \frac{f_{\mu}(x)}{\partial x} c(x, y_{i}) - \frac{\partial}{\partial x} c(x, y_{j}) \right] \mathcal{H}^{d-1}(x),
\]
where $T$ is the unique optimal mapping from $\mu$ to $\nu$, $T^{-1}(y_{i})$ and $T^{-1}(y_{j})$ are the so-called Laguerre cells, and $\mathcal{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure. For this illustrative example, one has that
\[
T^{-1}(-1) = ] - \infty, 0 ] \quad \text{and} \quad T^{-1}(1) = ] 0, + \infty [.
\]
Therefore, one has that $T^{-1}(y_{i}) \cap T^{-1}(y_{j}) = \{0\}$. Moreover, as $\frac{\partial}{\partial x} c(x, y) = -y$, we obtain that, for $i \neq j \in \{1, 2\}$,
\[
\left| \frac{\partial}{\partial x} c(x, y_{i}) - \frac{\partial}{\partial x} c(x, y_{j}) \right| = |1 - (-1)| = 2.
\]
Hence, since \( f_\mu(0) = 1, \mathcal{H}^0(\{0\}) = 1 \), we have that, for \( i \neq j \),
\[
\frac{\partial^2}{\partial v_i \partial v_j} H_0(v^*) = \frac{1}{2}
\]
from which it follows that
\[
\nabla^2 H_0(v^*) = A_0(v^*) = -\frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.
\]

References

[1] Bercu, B., and Bigot, J. Asymptotic distribution and convergence rates of stochastic algorithms for entropic optimal transportation between probability measures. *The Annals of Statistics* **49**, 2 (2021), 968 – 987.

[2] Chizat, L., Delalande, A., and Vaskevicius, T. Sharper exponential convergence rates for sinkhorn’s algorithm in continuous settings, 2024.

[3] Delalande, A. Nearly tight convergence bounds for semi-discrete entropic optimal transport. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics* (28–30 Mar 2022), G. Camps-Valls, F. J. R. Ruiz, and I. Valera, Eds., vol. 151 of *Proceedings of Machine Learning Research*, PMLR, pp. 1619–1642.

[4] Kitagawa, J., Mérigot, Q., and Thibert, B. Convergence of a newton algorithm for semi-discrete optimal transport. *Journal of the European Mathematical Society* **21**, 9 (2019).

[5] Steerneman, T., and van Perlo-ten Kleij, F. Properties of the matrix \( A - XY^* \). *Linear Algebra and its Applications* **410**, 1 (2005), 70–86.