



HAL
open science

Shannon-and von neumann-entropy regularizations of linear and semidefinite programs

Saroj Prasad Chhatoi, Jean B Lasserre

► To cite this version:

Saroj Prasad Chhatoi, Jean B Lasserre. Shannon-and von neumann-entropy regularizations of linear and semidefinite programs. 2025. <hal-05010676v2>

HAL Id: hal-05010676

<https://laas.hal.science/hal-05010676v2>

Preprint submitted on 3 Apr 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

SHANNON- AND VON NEUMANN-ENTROPY REGULARIZATIONS OF LINEAR AND SEMIDEFINITE PROGRAMS

S.P. CHHATOI AND JB. LASSERRE

ABSTRACT. We consider the LP in standard form $\min \{ \mathbf{c}^T \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}; \mathbf{x} \geq 0 \}$ and inspired by ε -entropy regularization in Optimal Transport, we introduce its ε -regularization $\min \{ \mathbf{c}^T \mathbf{x} + \varepsilon f(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{b}; \mathbf{x} \geq 0 \}$ via the (convex) Boltzmann-Shannon entropy $f(\mathbf{x}) := \sum_i x_i \ln x_i$. We also provide a similar regularization for the semidefinite program $\min \{ \text{Tr}(\mathbf{C} \cdot \mathbf{X}) : \mathcal{A}(\mathbf{X}) = \mathbf{b}; \mathbf{X} \succeq 0 \}$ but with now the so-called Von Neumann entropy, as in Quantum Optimal Transport. Importantly, both entropies are not barriers of the LP and SDP cones respectively. We show that for both LP and SDP, the resulting regularized problem admits an equivalent unconstrained convex problem $\max_{\boldsymbol{\lambda} \in \mathbb{R}^m} G_\varepsilon(\boldsymbol{\lambda})$ for an explicit concave differentiable function G_ε in dual variables $\boldsymbol{\lambda} \in \mathbb{R}^m$. As ε goes to zero, its optimal value converges to the optimal value of the initial LP. In the limit, one also recovers the unique optimal solution of the LP that minimizes the Shannon entropy among all LP-minimizers. While it resembles the log-barrier formulation of interior point algorithms, it has a distinguishing advantage. Namely, $G_\varepsilon(\boldsymbol{\lambda})$ is obtained in closed form (via an explicit minimizer $\mathbf{x}(\boldsymbol{\lambda}) > 0$) for every arbitrary $\boldsymbol{\lambda} \in \mathbb{R}^m$, hence with *no* need to restrict $\boldsymbol{\lambda}$ to the feasible set of the dual LP. This explicit form of G_ε is crucial for its maximization over the whole \mathbb{R}^m .

1. INTRODUCTION

In this paper we introduce the Boltzmann-Shannon entropy-regularization of linear programs (LPs) and the von Neumann entropy-regularization of semidefinite programs (SDPs), and describe and comment some of their nice features. The main remarkable feature is that with $\varepsilon > 0$ fixed, arbitrary, in both LP and SDP cases, one has to solve an unconstrained convex optimization problem whose reward function is explicit, strictly concave and differentiable. Hence basic optimization algorithms can be applied.

Namely, Let \mathbf{P} and \mathbf{Q} be a linear program (LP) and \mathbf{Q} be a semidefinite program (SDP) in standard form:

$$(1.1) \quad \mathbf{P} : \quad \tau = \min \{ \mathbf{c}^T \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}; \mathbf{x} \geq 0 \}$$

$$(1.2) \quad \mathbf{Q} : \quad \rho = \min \{ \text{Tr}(\mathbf{C} \cdot \mathbf{X}) : \mathcal{A}(\mathbf{X}) = \mathbf{b}; \mathbf{X} \succeq 0 \},$$

where $\mathbf{c} \in \mathbb{R}^d$, $\mathbf{C} \in \mathcal{S}^n$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times d}$, and $\mathcal{A} : \mathcal{S}^n \rightarrow \mathbb{R}^m$ is a linear map (with \mathcal{S}^n being the space of real $(n \times n)$ symmetric matrices).

Their respective ε -regularization \mathbf{P}_ε and \mathbf{Q}_ε (for $\varepsilon > 0$) via Shannon- and von Neumann-entropy¹, read:

$$(1.3) \quad \mathbf{P}_\varepsilon : \tau_\varepsilon = \min_{\mathbf{x}} \{ \mathbf{c}^T \mathbf{x} + \varepsilon \sum_{i=1}^d x_i \ln(x_i) : \mathbf{A}\mathbf{x} = \mathbf{b}; \mathbf{x} \geq 0 \}$$

$$(1.4) \quad \mathbf{Q}_\varepsilon : \rho_\varepsilon = \min_{\mathbf{X}} \{ \text{Tr}(\mathbf{C} \cdot \mathbf{X}) + \varepsilon \text{Tr}(\mathbf{X} \cdot \ln(\mathbf{X})) : \mathcal{A}(\mathbf{X}) = \mathbf{b}; \mathbf{X} \succeq 0 \}.$$

Notice that \mathbf{P}_ε resembles the log-barrier method

$$(1.5) \quad \min_{\mathbf{x}} \{ \mathbf{c}^T \mathbf{x} - \varepsilon \sum_{i=1}^d \ln(x_i) : \mathbf{A}\mathbf{x} = \mathbf{b}; \mathbf{x} \geq 0 \}$$

to solve (1.1) via interior point methods, and iterating with a sequence of parameters $\varepsilon_k \downarrow 0$ as $k \rightarrow \infty$. Similarly, the log-barrier method for solving (1.2) reads

$$(1.6) \quad \min_{\mathbf{X}} \{ \text{Tr}(\mathbf{C} \cdot \mathbf{X}) - \varepsilon \ln \det(\mathbf{X}) : \mathcal{A}(\mathbf{X}) = \mathbf{b}; \mathbf{X} \succeq 0 \}.$$

In fact, the function $-\sum_i \ln(x_i)$ is also an entropy function initially used by Boltzmann. However in the context of LP and SDP, and to the best of our knowledge, in its initial version, the log-barrier was not referred to as an entropy regularization. It was more conceived as a “repulsive barrier” to prevent \mathbf{x} (resp. \mathbf{X}) from approaching the boundary of the convex cone \mathbb{R}_+^d (resp. \mathcal{S}_+^n). The emphasis was on its self-concordance property of utmost importance for efficiency of interior point methods. More recently, an *entropic* variant was proposed in [3] where its properties are related to that of log-concave densities in probability. In the discussion in [3, Section 2.1] Hildebrand’s canonical barrier (a variant) is also related to the differential entropy of a Gaussian probability measure.

But we emphasize that in contrast with the log-barrier, the Boltzmann-Shannon entropy $\mathbf{x} \mapsto \sum_i x_i \ln(x_i)$ is *not* a barrier for the positive orthant because $u \ln(u) \rightarrow 0$ whenever $u \rightarrow 0$. This is why the term ε -regularization is more appropriate. Indeed without the entropy (i.e., $\varepsilon = 0$), an optimal solution $\mathbf{x}^*(0) \in \mathbb{R}^d$ of the LP is a vertex of the polytope hence with at least $d - m$ zero coordinates. With $\varepsilon > 0$ the associated optimal solution $\mathbf{x}^*(\varepsilon)$ has *no* zero coordinate and is a regularization of the vertex to a positive vector with small coordinates $x_i^*(\varepsilon)$ if $x_i^*(0) = 0$ and $\varepsilon > 0$ is small. Similarly for SDPs, $\mathbf{X}^*(0)$ is not full rank whereas $\mathbf{X}^*(\varepsilon)$ is full rank, but with some very small eigenvalues if $\varepsilon > 0$ is small.

Concerning SDPs, note that the classical log-barrier $-\ln(\det(\mathbf{X}))$ is also not comparable with the Von Neumann entropy $\text{Tr}(\mathbf{X} \cdot \ln(\mathbf{X}))$ in (1.4), as again and as in the scalar case, the latter is *not* a barrier for the SDP cone.

Depending on the context (thermodynamics, information theory, probability, etc..) the entropy $\sum_i x_i \ln x_i$ is given different names, like e.g., Boltzmann, Gibbs, Shannon). For an account on and historical references to entropy, the interested reader is referred to e.g., [7, 8, 4, 9] and the many references therein.

To the best of our knowledge, we are not aware of prior works using the Boltzmann-Shannon entropy to solve or approximate general LPs or SDPs (possibly because they are not *barriers* of the respective LP and SDP cones). An exception is the LP formulation of (commutative) optimal transport and the SDP formulation of

¹Observe that $\text{Tr}(\ln(\mathbf{X}))$ is the analogue for positive semidefinite matrices of $\sum_i \ln(x_i)$ for $\mathbf{x} \geq 0$. Indeed $\text{Tr}(\ln(\mathbf{X})) = \sum_i \ln(\sigma_i)$ where $(\sigma_i)_i$ are the eigenvalues of \mathbf{X} .

non-commutative quantum optimal transport. The former (resp. the latter) is a particular case of (1.3) (resp. (1.4)) and resulted in the celebrated Sinkhorn algorithm. For more details on the topic of optimal transport and entropy, the interested reader is referred to e.g. [1, 2, 5, 11, 4, 7, 8, 9]. It is worth noting that the LP formulation of optimal transport can be solved by Kuhn's Hungarian method as a more efficient alternative to the simplex algorithm.

Contribution. Assume that the feasible set of (1.1) (resp. (1.2)) is compact and Slater's condition holds for (1.1) (resp. (1.2)). Then we prove that for fixed $\varepsilon > 0$:

(i) Solving \mathbf{P}_ε reduces to solving the unconstrained convex optimization problem

$$(1.7) \quad \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \mathbf{b}^T \boldsymbol{\lambda} - \frac{\varepsilon}{e} \sum_{i=1}^d \exp((\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{c})_i / \varepsilon) \quad (=:\max_{\boldsymbol{\lambda}} G_\varepsilon(\boldsymbol{\lambda})),$$

where G_ε is concave, differentiable, and attains its maximum at some $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ because $-G_\varepsilon$ is coercive.

(ii) Solving \mathbf{Q}_ε reduces to solving the unconstrained convex optimization problem

$$(1.8) \quad \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \mathbf{b}^T \boldsymbol{\lambda} - \varepsilon \operatorname{Tr}(\exp(\mathcal{A}^* \boldsymbol{\lambda} - \mathbf{C}) / \varepsilon) - \mathbf{I} \quad (=:\max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \hat{G}_\varepsilon(\boldsymbol{\lambda})),$$

where \hat{G}_ε is concave, differentiable, and attains its maximum at some $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, also because \hat{G}_ε is coercive.

(iii) We also provide asymptotics as $\varepsilon \downarrow 0$. Namely, $\tau_\varepsilon \rightarrow \tau$ as $\varepsilon \downarrow 0$, and therefore with $\varepsilon > 0$ fixed, sufficiently small, one obtains a close approximation of τ by solving the single unconstrained convex optimization (1.7). Similarly, $\rho_\varepsilon \rightarrow \rho$ as $\varepsilon \downarrow 0$, and therefore with $\varepsilon > 0$ fixed, sufficiently small, one obtains a close approximation of ρ by solving the single unconstrained convex optimization (1.8).

In addition and interestingly, consider a sequence $\boldsymbol{x}^*(\varepsilon_k)_{k \in \mathbb{N}}$ of optimal solutions of $\mathbf{P}_{\varepsilon_k}$, with $\varepsilon_k \downarrow 0$ as $k \rightarrow \infty$. Then the sequence converges to the unique optimal solution \boldsymbol{x}^* of \mathbf{P} which minimizes the Boltzmann-Shannon entropy among the set of all optimal solutions of \mathbf{P} . The same result also holds for a sequence $\mathbf{X}^*(\varepsilon_k)_{k \in \mathbb{N}}$ of optimal solutions of $\mathbf{Q}_{\varepsilon_k}$.

(iv) For illustration purposes and also as a proof of concept, we provide a (limited) set of numerical experiments on LPs and SDPs of non-trivial size. For SDPs note that even though (1.8) is an unconstrained problem, evaluation of $\exp((\mathcal{A}^* \boldsymbol{\lambda} - \mathbf{C}) / \varepsilon)$ for a given $\boldsymbol{\lambda} \in \mathbb{R}^m$ requires a singular value decomposition $\mathbf{U}^T \mathbf{L} \mathbf{U}$ of $\mathcal{A}^* \boldsymbol{\lambda} - \mathbf{C}$ to exponentiate the diagonal elements of \mathbf{L} , and therefore as expected, maximizing \hat{G}_ε in (1.8) is significantly more difficult² than maximizing G_ε in (1.7).

Remarks and discussion.

- Comparing (1.3) with (1.5): To solve (1.5) one needs to maintain the semidefinite constraint $\boldsymbol{x} > 0$, that is, (1.5) is *not* an unconstrained minimization problem. In contrast, (1.7) is the Lagrangian relaxation-scheme applied to \mathbf{P}_ε and thanks to

²Notice that in the log-barrier method one has to evaluate $\ln(\det(\mathbf{X}))$, a non-trivial task, let alone that \mathbf{X} should be maintained to be positive definite.

the particular form of the Shannon entropy, the unconstrained inner minimization

$$\min_{\mathbf{x} \in \mathbb{R}^d} \mathbf{c}^T \mathbf{x} + \boldsymbol{\lambda}^T (\mathbf{b} - \mathbf{A} \mathbf{x}) + \varepsilon f(\mathbf{x}) \quad (= G_\varepsilon(\boldsymbol{\lambda}) - \mathbf{b}^T \boldsymbol{\lambda})$$

(where $f(\mathbf{x}) := \sum_{i=1}^d x_i \ln(x_i)$ if $\mathbf{x} \geq 0$ and $+\infty$ otherwise) is obtained in closed form with a strictly positive minimizer $\mathbf{x}(\varepsilon)$, which yields the concave differentiable function G_ε in explicit form (1.7).

The Shannon-entropy has been used in ε -regularization of (static) optimal transport (OT), and the characterization of optimal solutions, has resulted in the Sinkhorn algorithm widely used (e.g., in Machine Learning applications of OT). For more details on OT and the Sinkhorn algorithm the interested reader is referred to e.g. [4, 11] and references therein. However to the best of our knowledge we are not aware of the use of Boltzmann-Shannon entropy to help solve LPs.

- Comparing (1.4) with (1.6): To solve (1.6) with interior point methods, one needs to maintain the semidefinite constraint $\mathbf{X} \succ 0$, that is, (1.6) is *not* an unconstrained minimization problem. In contrast, (1.8) is the Lagrangian relaxation-scheme applied to \mathbf{Q}_ε and thanks to the particular form of the Shannon entropy, the unconstrained inner minimization

$$\min_{\mathbf{X} \in \mathcal{S}^n} \text{Tr}(\mathbf{C} \cdot \mathbf{X}) + \boldsymbol{\lambda} (\mathbf{b} - \mathcal{A}(\mathbf{X})) + \varepsilon g(\mathbf{X}) \quad (= \hat{G}_\varepsilon(\boldsymbol{\lambda}) - \mathbf{b}^T \boldsymbol{\lambda})$$

where $g(\mathbf{X}) := \text{Tr}(\mathbf{X} \cdot \ln(\mathbf{X}))$ if $\mathbf{X} \succ 0$, 0 if $\mathbf{X} \succeq 0$ and \mathbf{X} is singular, and $+\infty$ otherwise) is obtained in closed form with a strictly positive definite minimizer $\mathbf{X}(\varepsilon)$, and yields the function concave differentiable function \hat{G}_ε in explicit form (1.8).

In fact the entropy function $\mathbf{X} \mapsto \text{Tr}(\mathbf{X} \cdot \ln(\mathbf{X}))$ is rather called the Von Neumann entropy, and is used in some Quantum Optimal Transport problems formulated as (complex) SDPs with Hermitian matrices of very large dimension. So while inspired from this approach in Quantum Optimal Transport, our formulation \mathbf{Q}_ε for general SDPs (1.2) turns out to be in a much more favorable situation, at least for reasonable dimensions.

2. MAIN RESULT

Recall that $\lim_{u \downarrow 0} u \ln(u) = 0$ and so by continuity at 0, consider the convex function

$$\mathbf{x} \mapsto f(\mathbf{x}) := \begin{cases} \sum_{i=1}^d x_i \log(x_i) & \text{if } \mathbf{x} \geq 0 \\ +\infty & \text{otherwise,} \end{cases},$$

(with the convention $0 \ln(0) = 0$) which is convex and continuous on its domain $\text{Dom}(f) (= \{\mathbf{x} : \mathbf{x} \geq 0\})$, and differentiable on $\text{int}(\text{Dom}(f)) = \{\mathbf{x} > 0\}$.

2.1. Entropy-regularization of LPs. Consider the linear program (LP) (1.1) and its ε -regularization (1.3) where $\mathbf{c} \in \mathbb{R}^d$, $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{A} \in \mathbb{R}^{m \times d}$. For fixed $\varepsilon > 0$ define the Lagrangian

$$(\mathbf{x}, \boldsymbol{\lambda}) \mapsto \mathcal{L}_\varepsilon(\mathbf{x}, \boldsymbol{\lambda}) := \mathbf{c}^T \mathbf{x} + \boldsymbol{\lambda}^T (\mathbf{b} - \mathbf{A} \mathbf{x}) + \varepsilon f(\mathbf{x}), \quad (\mathbf{x}, \boldsymbol{\lambda}) \in \mathbb{R}^d \times \mathbb{R}^m,$$

associated with \mathbf{P}_ε , and let

$$\boldsymbol{\lambda} \mapsto G_\varepsilon(\boldsymbol{\lambda}) := \inf_{\mathbf{x} \in \mathbb{R}^d} \mathcal{L}_\varepsilon(\boldsymbol{\lambda}, \mathbf{x}), \quad \boldsymbol{\lambda} \in \mathbb{R}^d.$$

Lemma 2.1. *Assume that there exists $\lambda_0 \in \mathbb{R}^m$ such that $\mathbf{A}^T \lambda_0 > 0$ (so that $\Omega := \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}; \mathbf{x} \geq 0\}$ is compact). Then for every fixed $\varepsilon > 0$, G_ε is concave, differentiable and satisfies*

$$(2.1) \quad G_\varepsilon(\boldsymbol{\lambda}) = \mathcal{L}_\varepsilon(\boldsymbol{\lambda}, \mathbf{x}^*(\boldsymbol{\lambda})) = \mathbf{b}^T \boldsymbol{\lambda} - \varepsilon \sum_{i=1}^d \mathbf{x}_i^*(\boldsymbol{\lambda}), \quad \forall \boldsymbol{\lambda} \in \mathbb{R}^m$$

$$(2.2) \quad \text{with } \mathbf{x}_i^*(\boldsymbol{\lambda}) := \exp((\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{c})_i / \varepsilon - 1) > 0, \quad i = 1, \dots, d;$$

$$(2.3) \quad \nabla G_\varepsilon(\boldsymbol{\lambda}) = \mathbf{b} - \mathbf{A}\mathbf{x}^*(\boldsymbol{\lambda}), \quad \forall \boldsymbol{\lambda} \in \mathbb{R}^m.$$

Proof. The function $\mathbf{x} \mapsto \mathcal{L}_\varepsilon(\boldsymbol{\lambda}, \mathbf{x})$ is strictly convex and differentiable on $\text{int}(\text{Dom}(f))$. Let $\mathbf{x}^*(\boldsymbol{\lambda})$ be as in (2.2) and so $\mathbf{x}^*(\boldsymbol{\lambda}) \in \text{int}(\text{Dom}(f))$. Observe that $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0$, which implies that $\mathbf{x}^*(\boldsymbol{\lambda})$ is the unique minimizer of $\mathbf{x} \mapsto \mathcal{L}_\varepsilon(\boldsymbol{\lambda}, \mathbf{x})$, and yields (2.1)-(2.2). To get (2.3), substitute (2.2) in the gradient $\nabla_{\mathbf{x}} G_\varepsilon(\mathbf{x}^*(\boldsymbol{\lambda}))$. Finally, concavity of G_ε follows substituting (2.2) in $\mathcal{L}_\varepsilon(\mathbf{x}^*(\boldsymbol{\lambda}), \boldsymbol{\lambda})$. \square

Coercivity of G_ε . Let $\mathbb{S}^{m-1} \subset \mathbb{R}^m$ denote the Euclidean unit sphere, and assume that \mathbf{A} is full row rank.

Lemma 2.2. *Assume that there exists $0 < \mathbf{x}_0 \in \Omega$ and let G_ε be as in (1.7). Then $-G_\varepsilon$ is coercive, that is, $G_\varepsilon \rightarrow -\infty$ as $\|\boldsymbol{\lambda}\| \rightarrow +\infty$.*

Proof. We first prove that if a sequence $(\boldsymbol{\lambda}_k)_{k \in \mathbb{N}}$ is such that $\|\mathbf{b}^T \boldsymbol{\lambda}_k\| \rightarrow \infty$ as $k \rightarrow \infty$, then there exists i such that $|(\mathbf{A}^T \boldsymbol{\lambda}_k)_i| \rightarrow +\infty$ as $k \rightarrow \infty$. Suppose not, that is, $\sup_k \sup_i |(\mathbf{A}^T \boldsymbol{\lambda}_k)_i| < a$ for some scalar $a > 0$. Then $\boldsymbol{\lambda}_k / \|\boldsymbol{\lambda}_k\| \in \mathbb{S}^{m-1}$, and by compactness, there exists a subsequence relabelled $(\boldsymbol{\lambda}_k)_{k \in \mathbb{N}}$ for convenience, such that $\boldsymbol{\lambda}_k / \|\boldsymbol{\lambda}_k\| \rightarrow \boldsymbol{\theta} \in \mathbb{S}^{m-1}$. Hence for every $i = 1, \dots, d$, $|(\mathbf{A}^T \boldsymbol{\lambda}_k / \|\boldsymbol{\lambda}_k\|)_i| < a / \|\boldsymbol{\lambda}_k\| \rightarrow 0$ as $k \rightarrow \infty$. Therefore $\mathbf{A}^T \boldsymbol{\theta} = 0$ which in turns implies $\boldsymbol{\theta} = 0$ (as \mathbf{A} is full row rank), in contradiction with $\|\boldsymbol{\theta}\| = 1$.

Next, write $\mathbf{b} = \mathbf{A}\mathbf{x}_0$ (with $\mathbf{x}_0 > 0$), so that $\mathbf{b}^T \boldsymbol{\lambda} = (\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{c})^T \mathbf{x}_0 + \mathbf{c}^T \mathbf{x}_0$. Then for a sequence $(\boldsymbol{\lambda}_k)_{k \in \mathbb{N}}$ with $\|\boldsymbol{\lambda}_k\| \rightarrow \infty$,

$$\begin{aligned} G_\varepsilon(\boldsymbol{\lambda}_k) &= \mathbf{b}^T \boldsymbol{\lambda}_k - \varepsilon \sum_{i=1}^d \mathbf{x}_i(\boldsymbol{\lambda}_k) \\ &= \mathbf{c}^T \mathbf{x}_0 + (\mathbf{A}^T \boldsymbol{\lambda}_k - \mathbf{c})^T \mathbf{x}_0 - \varepsilon \|\mathbf{x}(\boldsymbol{\lambda}_k)\|_1 \\ &= \mathbf{c}^T \mathbf{x}_0 + \sum_{i=1}^d (\mathbf{A}^T \boldsymbol{\lambda}_k - \mathbf{c})_i \left[(\mathbf{x}_0)_i - \varepsilon \frac{\exp((\mathbf{A}^T \boldsymbol{\lambda}_k - \mathbf{c})_i / \varepsilon)}{e^{(\mathbf{A}^T \boldsymbol{\lambda}_k - \mathbf{c})_i}} \right]. \end{aligned}$$

So let $(\|\boldsymbol{\lambda}_k\|) \rightarrow \infty$, and let $I := \{i : |(\mathbf{A}^T \boldsymbol{\lambda}_k)_i| \rightarrow +\infty\}$ so that $I \neq \emptyset$. If $i \in I$ and $(\mathbf{A}^T \boldsymbol{\lambda}_k - \mathbf{c})_i \rightarrow -\infty$ then the item i of the bracket is nonnegative (as $(\mathbf{x}_0)_i > 0$) and therefore the item i of the sum goes to $-\infty$. Similarly, if $(\mathbf{A}^T \boldsymbol{\lambda}_k - \mathbf{c})_i \rightarrow +\infty$ then in the item i the bracket goes to $-\infty$ and therefore the item i of the sum goes to $-\infty$. Therefore the contribution in sum of every item $i \in I$ gives $-\infty$, which yields the desired result. \square

Corollary 2.3. *For every $\varepsilon > 0$, the regularization \mathbf{P}_ε of \mathbf{P} has a unique optimal solution $0 < \mathbf{x}^*(\varepsilon) \in \Omega$, and under the assumption of Lemma 2.2,*

$$(2.4) \quad \tau_\varepsilon = \min_{\mathbf{x} \geq 0} \mathbf{c}^T \mathbf{x} + \varepsilon f(\mathbf{x}) = \mathbf{c}^T \mathbf{x}^*(\varepsilon) + \varepsilon \sum_{i=1}^d \mathbf{x}_i^*(\varepsilon) \ln(\mathbf{x}_i^*(\varepsilon))$$

$$(2.5) \quad = \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} G_\varepsilon(\boldsymbol{\lambda}) = \mathbf{b}^T \boldsymbol{\lambda}^* - \varepsilon \sum_{i=1}^d \mathbf{x}_i^*(\varepsilon).$$

Proof. \mathbf{P}_ε is equivalent to $\min_{\mathbf{x} \in \Omega} \mathbf{c}^T \mathbf{x} + \varepsilon f(\mathbf{x})$, and the infimum of a continuous function on a compact set is attained. Moreover as $\mathbf{x} \mapsto \mathbf{c}^T \mathbf{x} + \varepsilon f(\mathbf{x})$ is strictly convex, a minimizer $\mathbf{x}^*(\varepsilon) \in \Omega$ is unique.

Next, recall that $G_\varepsilon(\boldsymbol{\lambda}) \leq \tau_\varepsilon$ for all $\boldsymbol{\lambda} \in \mathbb{R}^m$, G_ε is continuous, differentiable and by Lemma 2.2, $-G_\varepsilon$ is coercive. Therefore its maximum is attained at $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ where $\nabla G_\varepsilon(\boldsymbol{\lambda}^*) = 0 = \mathbf{b} - \mathbf{A} \mathbf{x}^*(\boldsymbol{\lambda}^*)$, and so $\mathbf{x}^*(\boldsymbol{\lambda}^*) \in \Omega$. Moreover a minimizer $\boldsymbol{\lambda}^*$ is unique because in view of (2.1), and (2.2), G_ε is strictly concave. Finally, as

$$G_\varepsilon(\boldsymbol{\lambda}^*) = \mathbf{L}_\varepsilon(\boldsymbol{\lambda}^*, \mathbf{x}^*(\boldsymbol{\lambda}^*)) = \mathbf{c}^T \mathbf{x}^*(\boldsymbol{\lambda}^*) + \varepsilon f(\mathbf{x}^*(\boldsymbol{\lambda}^*)) \leq \tau_\varepsilon,$$

and $\mathbf{x}^*(\boldsymbol{\lambda}^*) \in \Omega$, we conclude that $\mathbf{x}^*(\boldsymbol{\lambda}^*)$ is a minimizer of \mathbf{P}_ε and by uniqueness $\mathbf{x}^*(\boldsymbol{\lambda}^*) = \mathbf{x}^*(\varepsilon)$. \square

As a result, one may obtain τ_ε by solving an unconstrained optimization problem (2.4), i.e., by maximizing the concave (and explicit) differentiable function G_ε over \mathbb{R}^d . Therefore with $\varepsilon > 0$ small and fixed, one obtains an approximation of τ , the optimal value of the LP (1.1). Moreover, at the unique maximizer $\boldsymbol{\lambda}^*$ of (2.4), one extracts the unique minimizer $\mathbf{x}^*(\varepsilon) = \mathbf{x}^*(\boldsymbol{\lambda}^*)$ of \mathbf{P}_ε directly by the formula (2.2).

Finally, observe that in the canonical form (1.1), one has $m \leq d$ and even sometimes $m \ll d$, which is an additional nice feature of (2.4).

2.2. Comparison with the log-barrier. Let $\mu > 0$ be fixed, and consider

$$\mathbf{P}_\mu : \quad \tau_\mu := \inf_{\mathbf{x}} \{ \mathbf{c}^T \mathbf{x} - \mu \sum_{i=1}^d \ln(\mathbf{x}_i) : \mathbf{A} \mathbf{x} = \mathbf{b} \}.$$

At a minimizer $\mathbf{x}(\mu)$, the necessary KKT-optimality conditions yield

$$\mathbf{A} \mathbf{x}(\mu) = \mathbf{b}; \quad (\mathbf{c} - \mathbf{A}^T \boldsymbol{\lambda}(\mu))_i = \mu / \mathbf{x}_i(\mu) \quad ;, \quad i = 1, \dots, d,$$

for some multiplier $\boldsymbol{\lambda}(\mu)$. That is:

$$\mathbf{x}_i(\mu) = \frac{\mu}{\mathbf{c}_i - \mathbf{A}^T \boldsymbol{\lambda}(\mu)}, \quad i = 1, \dots, d,$$

from which nonnegativity is not automatic. It requires that $\mathbf{A}^T \boldsymbol{\lambda}(\mu) < \mathbf{c}$, and so the analogue H_μ of G_ε reads:

$$H_\mu(\boldsymbol{\lambda}) := \mathbf{b}^T \boldsymbol{\lambda} + \sum_{i=1}^d \frac{\mathbf{c}_i \mu}{(\mathbf{c} - \mathbf{A}^T \boldsymbol{\lambda})_i} + \mu \ln((\mathbf{c} - \mathbf{A}^T \boldsymbol{\lambda})_i), \quad \mathbf{A}^T \boldsymbol{\lambda} < \mathbf{c},$$

defined only for $\mathbf{A}^T \boldsymbol{\lambda} < \mathbf{c}$, whereas G_ε is defined for every $\boldsymbol{\lambda} \in \mathbb{R}^m$. This is an important difference if one wants to maximize H_μ to retrieve the optimal value of \mathbf{P}_μ via Lagrangian relaxation-scheme³. Finally one remarks that if $f(\mathbf{x})$ is *not* a

³However, in interior point algorithms one usually directly minimizes $\mathbf{c}^T \mathbf{x} - \mu \sum_{i=1}^d \ln(\mathbf{x}_i)$ over the subspace $\mathbf{A} \mathbf{x} = \mathbf{b}$, and then one iterates with a lower value of the parameter μ , etc.

barrier for the positive orthant, its gradient $\mathbf{x} \mapsto \sum_{i=1}^d \ln(\mathbf{x}_i) + 1$ acts as a barrier in gradient methods.

2.3. Semidefinite Programming via von-Neumann entropy. Consider the semidefinite program (SDP)

$$(2.6) \quad \mathbf{P} : \quad \rho = \min \{ \text{Tr}(\mathbf{C} \cdot \mathbf{X}) : \mathcal{A}(\mathbf{X}) = \mathbf{b}; \mathbf{X} \succeq 0 \}$$

where $\mathbf{C} \in \mathcal{S}^n$ and $\mathcal{A} : \mathcal{S}^n \mapsto \mathbb{R}^m$ is a linear mapping.

We assume that the spectrahedron $\Omega := \{ \mathcal{A}(\mathbf{X}) = \mathbf{b}; \mathbf{X} \succeq 0 \}$ is compact, and we consider the following regularization \mathbf{P}_ε of \mathbf{P} , where $\varepsilon > 0$ is fixed, arbitrary.

$$(2.7) \quad \mathbf{Q}_\varepsilon : \quad \rho_\varepsilon = \min_{\mathbf{X}} \{ \text{Tr}(\mathbf{C} \cdot \mathbf{X}) + \varepsilon \text{Tr}(\mathbf{X} \cdot \ln(\mathbf{X})) : \mathcal{A}(\mathbf{X}) = \mathbf{b}; \mathbf{X} \succ 0 \}.$$

\mathbf{P}_ε is a convex problem as the von-Neumann entropy

$$\mathbf{X} \mapsto \text{Tr}(\mathbf{X} \cdot \ln(\mathbf{X})), \quad \forall \mathbf{X} \in \mathcal{S}_{++}^n \quad (:= \{ \mathbf{X} \in \mathcal{S}^n : \mathbf{X} \succ 0 \})$$

is a convex function. Next define

$$(2.8) \quad \mathbf{X} \mapsto g(\mathbf{X}) := \begin{cases} \text{Tr}(\mathbf{X} \cdot \ln(\mathbf{X})) & \text{if } \mathbf{X} \succ 0 \\ 0 & \text{if } \mathbf{X} \succeq 0 \text{ and } \mathbf{X} \text{ is singular} \\ +\infty & \text{otherwise.} \end{cases}$$

So g is continuous on $\mathcal{S}_+^n = \{ \mathbf{X} : \mathbf{X} \succeq 0 \}$ by using the convention $u \ln(u) = 0$ if $u = 0$.

Proposition 2.4. *The convex optimization problem has an optimal solution.*

Proof. It follows from the fact that Ω is compact and in (2.7) one may rather consider the cost function $\mathbf{X} \mapsto \text{Tr}(\mathbf{C} \cdot \mathbf{X}) + \varepsilon g(\mathbf{X})$ which is continuous. \square

Let $\mathcal{A}^* : \mathbb{R}^m \rightarrow \mathcal{S}^n$ be the adjoint linear, injective mapping associated with \mathcal{A} , and introduce the Lagrangian $G_\varepsilon : \mathbb{R}^m \rightarrow \mathbb{R}$:

$$(2.9) \quad G_\varepsilon(\boldsymbol{\lambda}) := \mathbf{b}^T \boldsymbol{\lambda} + \inf_{\mathbf{X}} \{ \text{Tr}(\mathbf{X} \cdot (\mathbf{C} - \mathcal{A}^* \boldsymbol{\lambda})) + \varepsilon \text{Tr}(\mathbf{X} \cdot \ln(\mathbf{X})) \}.$$

Observe that $G_\varepsilon(\boldsymbol{\lambda}) \leq \rho_\varepsilon$, for every $\boldsymbol{\lambda} \in \mathbb{R}^m$.

Lemma 2.5. *For every fixed $\varepsilon > 0$, the function G_ε is strictly concave and satisfies*

$$(2.10) \quad \begin{aligned} G_\varepsilon(\boldsymbol{\lambda}) &= \mathbf{b}^T \boldsymbol{\lambda} - \varepsilon \text{Tr} \mathbf{X}(\boldsymbol{\lambda}) \\ \text{with } \mathbf{X}(\boldsymbol{\lambda}) &= \exp((\mathcal{A}^* \boldsymbol{\lambda} - \mathbf{C})/\varepsilon) - \mathbf{I}. \end{aligned}$$

Proof. The function

$$\mathbf{X} \mapsto \text{Tr}((\mathbf{C} - \mathcal{A}^* \boldsymbol{\lambda}) \cdot \mathbf{X}) + \varepsilon \text{Tr}(\mathbf{X} \cdot \log(\mathbf{X}))$$

is convex. Its gradient, which reads

$$\mathbf{X} \mapsto \mathbf{C} - \mathcal{A}^* \boldsymbol{\lambda} + \varepsilon (\ln(\mathbf{X}) + \mathbf{I}), \quad \forall \mathbf{X} \succ 0,$$

vanishes at $\mathbf{X}(\boldsymbol{\lambda}) := \exp((\mathcal{A}^* \boldsymbol{\lambda} - \mathbf{C})/\varepsilon) - \mathbf{I} \succ 0$. Next, evaluating G_ε at $\mathbf{X}(\boldsymbol{\lambda})$ yields (2.10), from which strict concavity follows. \square

Coercivity of G_ε . We first recall the following technical result.

Proposition 2.6. *Let $\mathbf{A}, \mathbf{B} \in \mathcal{S}^n$ be (n, n) real symmetric matrices, with $\mathbf{B} \succeq 0$, and let (λ_j^A) and (λ_j^B) be their respective eigenvalues arranged in increasing order $\lambda_1^A \leq \dots \leq \lambda_n^A$, and $\lambda_1^B \leq \dots \leq \lambda_n^B$. Then:*

$$(2.11) \quad \sum_{j=1}^n \lambda_j^A \lambda_{n-j+1}^B \leq \text{Tr}(\mathbf{A} \cdot \mathbf{B}) \leq \sum_{j=1}^n \lambda_j^A \lambda_j^B.$$

Proof. If $\mathbf{A} \succeq 0$ the result is well-known, otherwise, write $\mathbf{A} = \mathbf{A} + a\mathbf{I} - a\mathbf{I}$ for $a > 0$ sufficiently large so that $\lambda_j^{\mathbf{A}+a\mathbf{I}} = \lambda_j^A + a \geq 0$ for all j . \square

Lemma 2.7. *Let Ω be compact and assume that:*

- Slater's condition holds, i.e., there exists $0 \prec \mathbf{X}_0 \in \Omega$, and
- $\mathcal{A}^* \boldsymbol{\lambda} = 0 \Rightarrow \boldsymbol{\lambda} = 0$.

Then $G_\varepsilon(\boldsymbol{\lambda}) \rightarrow -\infty$ whenever $\|\boldsymbol{\lambda}\| \rightarrow +\infty$,

Proof. We first prove that $\|\mathcal{A}^* \boldsymbol{\lambda}\| \rightarrow \infty$ whenever $\|\boldsymbol{\lambda}\| \rightarrow \infty$. So let $(\boldsymbol{\lambda}_k)_{k \in \mathbb{N}}$ be a sequence such that $\|\boldsymbol{\lambda}_k\| \rightarrow \infty$ as $k \rightarrow \infty$. Then there is a subsequence still denoted $(\boldsymbol{\lambda}_k)_{k \in \mathbb{N}}$ for convenience, such that $\boldsymbol{\lambda}_k / \|\boldsymbol{\lambda}_k\| \rightarrow \boldsymbol{\lambda}^* \in \mathbb{S}^{m-1}$. Hence suppose that $\sup_k \|\mathcal{A}^* \boldsymbol{\lambda}_k\| < a$ for some $a > 0$. Then $\mathcal{A}^* \boldsymbol{\lambda}_k / \|\boldsymbol{\lambda}_k\| \rightarrow \mathcal{A}^* \boldsymbol{\lambda}^* = 0$, which in turn implies $\boldsymbol{\lambda}^* = 0$, in contradiction with $\boldsymbol{\lambda}^* \in \mathbb{S}^{m-1}$. Next write

$$\mathbf{b}^T \boldsymbol{\lambda} = \text{Tr}(\mathcal{A}^* \boldsymbol{\lambda} - \mathbf{C}) \cdot \mathbf{X}_0 + \text{Tr}(\mathbf{C} \cdot \mathbf{X}_0),$$

so that

$$\begin{aligned} G_\varepsilon(\boldsymbol{\lambda}_k) &= \mathbf{b}^T \boldsymbol{\lambda}_k - \varepsilon \text{Tr}(\mathbf{X}(\boldsymbol{\lambda}_k)) \\ &= \text{Tr}(\mathbf{C} \cdot \mathbf{X}_0) + \text{Tr}((\mathcal{A}^* \boldsymbol{\lambda}_k - \mathbf{C}) \cdot \mathbf{X}_0) - \varepsilon \text{Tr}(\mathbf{X}(\boldsymbol{\lambda}_k)) \\ &= \text{Tr}(\mathbf{C} \cdot \mathbf{X}_0) + \text{Tr}((\mathcal{A}^* \boldsymbol{\lambda}_k - \mathbf{C}) \cdot \mathbf{X}_0) - \varepsilon \text{Tr}(\exp(\mathcal{A}^* \boldsymbol{\lambda}_k - \mathbf{C}) / \varepsilon) - \mathbf{I} \end{aligned}$$

Let $(\sigma_1(\boldsymbol{\lambda}_k), \dots, \sigma_m(\boldsymbol{\lambda}_k))$ be the eigenvalues of the real symmetric matrix $\mathcal{A}^* \boldsymbol{\lambda}_k - \mathbf{C}$, and let (u_1, \dots, u_m) that of $\mathbf{X}_0 \succ 0$, arranged in increasing order. By Proposition 2.6:

$$\begin{aligned} \text{Tr}((\mathcal{A}^* \boldsymbol{\lambda}_k - \mathbf{C}) \cdot \mathbf{X}_0) &\leq \sum_{i=1}^n u_i \sigma_i(\boldsymbol{\lambda}_k) \\ \text{Tr}(\exp((\mathcal{A}^* \boldsymbol{\lambda}_k - \mathbf{C}) / \varepsilon) - \mathbf{I}) &\geq \sum_{i=1}^n \exp(\sigma_i(\boldsymbol{\lambda}_k) / \varepsilon - 1), \end{aligned}$$

and therefore

$$G_\varepsilon(\boldsymbol{\lambda}_k) \leq \sum_{i=1}^n u_i \sigma_i(\boldsymbol{\lambda}_k) - \varepsilon \exp(\sigma_i(\boldsymbol{\lambda}_k) / \varepsilon - 1).$$

Since $\|\mathcal{A}^* \boldsymbol{\lambda}_k\| \rightarrow \infty$ as $k \rightarrow \infty$, one has $|\sigma_i(\boldsymbol{\lambda}_k)| \rightarrow \infty$ for at least one index i , and so let $I^+ := \{i : \sigma_i(\boldsymbol{\lambda}_k) \rightarrow +\infty\}$, and $I^- := \{i : \sigma_i(\boldsymbol{\lambda}_k) \rightarrow -\infty\}$. As $k \rightarrow \infty$:

$$\begin{aligned} u_i \sigma_i(\boldsymbol{\lambda}_k) - \varepsilon \exp(\sigma_i(\boldsymbol{\lambda}_k) / \varepsilon - 1) &\rightarrow -\infty \text{ if } i \in I^+ \\ \underbrace{u_i}_{>0} \sigma_i(\boldsymbol{\lambda}_k) - \varepsilon \exp(\sigma_i(\boldsymbol{\lambda}_k) / \varepsilon - 1) &\rightarrow -\infty \text{ if } i \in I^-. \end{aligned}$$

This shows that $G_\varepsilon \rightarrow -\infty$ as $k \rightarrow \infty$, and so $-G_\varepsilon$ is coercive. \square

Corollary 2.8. *The regularization \mathbf{P}_ε of \mathbf{P} has a unique optimal solution $0 < \mathbf{X}^*(\varepsilon) \in \Omega$, and under the assumption of Lemma 2.7,*

$$(2.12) \quad \tau_\varepsilon = \text{Tr}(\mathbf{C} \cdot \mathbf{X}^*(\varepsilon)) + \varepsilon \text{Tr}(\mathbf{X}^*(\varepsilon) \cdot \log(\mathbf{X}^*(\varepsilon)))$$

$$(2.13) \quad = \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} G_\varepsilon(\boldsymbol{\lambda}) = \mathbf{b}^T \boldsymbol{\lambda}^* - \varepsilon \text{Tr}(\mathbf{X}^*(\varepsilon)).$$

The proof follows similar arguments as the linear programming case.

3. ASYMPTOTIC ANALYSIS AND COMPUTATIONAL EXPERIMENTS

In this section, we analyze the asymptotic behavior of the solution to the entropy-regularized linear program as $\varepsilon \downarrow 0$, and provide some computational experiments that illustrate the approach.

3.1. Asymptotic Behavior of the Regularized Solution. We consider the asymptotic behavior of the solution $\mathbf{x}^*(\varepsilon)$ to the entropy-regularized linear program as the regularization parameter $\varepsilon \downarrow 0$.

Lemma 3.1. *For every $\varepsilon > 0$, let $\mathbf{x}^*(\varepsilon)$ be the unique minimizer of (1.3). Then $\mathbf{x}^*(\varepsilon)$ converges to the unique minimizer \mathbf{x}^* of the LP (1.1) which minimizes the Shannon entropy among all minimizers of (1.1).*

Proof. Consider a sequence $\varepsilon_k \downarrow 0$, and its associated sequence of unique minimizers $\mathbf{x}^*(\varepsilon_k)$ of $\mathbf{P}_{\varepsilon_k}$ as defined in (1.3), and let \mathbf{x} be any minimizer of \mathbf{P} as defined in (1.1). Since Ω is compact, $\mathbf{x}^*(\varepsilon_k) \rightarrow \mathbf{x}^*$ (up to a subsequence). Then we have

$$0 \leq \mathbf{c}^T \mathbf{x}^*(\varepsilon_k) - \mathbf{c}^T \mathbf{x} \leq \varepsilon_k \left(\sum_{i=1}^d \mathbf{x}_i \ln \mathbf{x}_i - \sum_{i=1}^d \mathbf{x}_i^*(\varepsilon_k) \ln \mathbf{x}_i^*(\varepsilon_k) \right)$$

where the first inequality follows from the fact that \mathbf{x} is the minimizer of (1.1) and the second inequality follows from the fact that $\mathbf{x}^*(\varepsilon)$ is the minimizer of (1.3). As $\varepsilon_k \downarrow 0$, we obtain $\mathbf{c}^T \mathbf{x}^* = \mathbf{c}^T \mathbf{x}$. Moreover, dividing ε_k on the right-hand-side inequality, one obtains $\sum_{i=1}^d \mathbf{x}_i \ln \mathbf{x}_i \geq \sum_{i=1}^d \mathbf{x}_i^*(\varepsilon_k) \ln \mathbf{x}_i^*(\varepsilon_k)$. Thus taking the limit as $k \rightarrow \infty$, one obtains $\sum_{i=1}^d \mathbf{x}_i \ln \mathbf{x}_i \geq \sum_{i=1}^d \mathbf{x}_i^* \ln \mathbf{x}_i^*$. That is, \mathbf{x}^* minimizes the Shannon entropy among all minimizers of (1.1). Since entropy is strictly convex, the whole sequence $\mathbf{x}^*(\varepsilon_k)$ converges to the one that minimizes Shannon entropy (and so is unique). \square

Remark 3.2. *For semidefinite programs, a similar result holds for the minimizers $\mathbf{X}(\boldsymbol{\lambda})$ of problem \mathbf{P}_ε , as defined in (2.7), with a similar proof.*

3.2. Computational Experiments. We next provide some numerical results to illustrate the method for solving entropy-regularization of linear programs. To showcase the proposed method, we have considered two different type of examples.

Large-Scale Linear Program. Our first example consists of a non trivial LP with 10000 variables and 50 constraints. The matrix \mathbf{A} and the vector \mathbf{b} were generated randomly while ensuring feasibility by setting $\mathbf{b} = \mathbf{A}\mathbf{x}^*$ for a randomly generated positive vector \mathbf{x}^* (so that Slater's condition holds). The cost vector \mathbf{c} was also generated randomly. To obtain $\boldsymbol{\lambda}^*$, the dual function $G_\varepsilon(\boldsymbol{\lambda})$ was maximized using MATLAB's `fminunc` function based on a quasi-Newton method.

With $\varepsilon := 0.01$, we observed that solving 20 random instances of the regularized LP problem required only an average of 15 iterations to converge, with the norm

of the gradient of \mathbf{G}_ε at optimality being on the order of $1e - 4$; recall that the gradient of \mathbf{G}_ε is the feasibility constraint in (1.1).

Discrete Optimal Transport Problem. As already mentioned, for optimal transport problems (OT) (a very specific type of LPs), the celebrated Sinkhorn algorithm (also based on a Shannon-entropy regularization) is widely used; see e.g. [11]. So we have also compared our algorithm with Sinkhorn algorithm on a toy OT example with cost matrix

$$C = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix},$$

and with source and target distributions (marginals) :

$$(3.1) \quad p = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \quad \text{and} \quad q = \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}.$$

A regularization parameter of $\varepsilon = 0.01$ was applied. The dual problem was solved using again `fminunc`, to yield the optimal dual solution and the corresponding regularized cost. The optimal value $G_\varepsilon(\boldsymbol{\lambda}^*)$ of the regularized LP for the OT problem was **1.7906** with $\|\nabla G_\varepsilon(\boldsymbol{\lambda}^*)\| \approx 1e - 6$, while the optimal value obtained via the Sinkhorn algorithm⁴ was **1.7906**.

Remark 3.3. *The Sinkhorn algorithm is an iterative matrix scaling algorithm used to find a bistochastic matrix by iteratively normalizing its rows and columns. A key property of the Sinkhorn algorithm is the factorization of the exponential map, which characterizes the primal solution as*

$$\exp(\phi + \psi - C).$$

where ϕ, ψ are lagrange multipliers corresponding to the marginal constraints in the Sinkhorn algorithm. An example of such marginals is provided in (3.1). This factorization enables efficient updates and guarantees geometric convergence at a rate of $\mathcal{O}(e^{-k/\varepsilon})$, where k is the iteration count and $\varepsilon > 0$ is the entropic regularization parameter.

However, such a factorization is not possible in (2.3), because the matrix $\mathbf{A}^T \boldsymbol{\lambda}$ contains mixed terms that prevent the exponential matrix from being factorized. This lack of separability is an obstruction to envision Sinkhorn-like algorithms, as iterative matrix scaling techniques rely on independent row and column updates, which are not possible in this setting.

Semidefinite Programming Example. We also implemented our method for solving problem \mathbf{Q}_ε defined in (2.7) by maximizing the dual function G_ε given in (2.10). We solved a problem with $\mathcal{A} : \mathcal{S}^{100} \rightarrow \mathbb{R}^{20}$, where the individual matrices were randomly generated symmetric matrices. The problem was solved by again using MATLAB's `fminunc` function. In 20 random experiments, we observed that the problem converged on average in 15 iterations.

We wish to emphasize that the numerical results provided in this section are meant to showcase the proposed method. While we have successfully implemented the method for medium to relatively large LPs (**with matrices \mathbf{A} of dimension around (50×10000)**), for semidefinite programming problems we have encountered numerical issues with large dimensions and very small regularization parameter ε ,

⁴We refer to the MATLAB implementation by Marco Cuturi, available at <https://github.com/marcocuturi/SinkhornTransport>

very likely due to our use of basic Matlab’s matrix exponential routines which are prone to numerical instabilities. Further investigation beyond the scope of the present paper, is required to (i) develop robust and faster algorithms to implement the proposed method, and (ii) possibly provide rates of convergence for fixed ε . In addition, a detailed comparison of the proposed method with existing methods is a topic further investigation.

Remark 3.4. *Table 1 below draws a comparison between the proposed method and the Sinkhorn algorithm, similar to how the Simplex algorithm is compared to the Hungarian algorithm for solving linear programming and optimal transport problems, respectively.*

	<i>LP</i>	<i>OT</i>
<i>Classical Algorithms</i>	<i>Simplex</i>	<i>Hungarian</i>
<i>Entropy-regularization</i>	<i>Proposed method</i>	<i>Sinkhorn</i>

TABLE 1. (LP) and optimal transport (OT) problems, with their entropy-regularization.

ACKNOWLEDGEMENTS

This work was done when the two authors were invited at the IPAM institute in UCLA during the long program *Non Commutative Optimal Transport* (Spring 2025). It was inspired by the use of Shannon entropy in (static) OT and benefited from various discussions with participants in that program. The authors gratefully acknowledge financial support from IPAM.

REFERENCES

- [1] J.-D. Benamou, G. Carlier, L. Nenna. A numerical method to solve optimal transport problems with Coulomb cost, Chapter in *Splitting Methods in Communication, Imaging, Science, and Engineering*, Roland Glowinski, Stanley J. Osher, and Wotao Yin Eds., Springer Cham, 2016, pp. 577-601.
- [2] F. Beier, J. von Lindheim, S. Neumayer, G. Steidl. Unbalanced multimarginal optimal transport, *J. Math. Imaging and Vision* 65, pp. pp. 394–413, 2023.
- [3] S. Bubeck, R. Eldan. The entropic barrier: Exponential families, log-concave geometry and self-concordance, *Math. Oper. Res.* 44(1), pp. 264–276, 2018.
- [4] Y Chen, T.T. Georgiou, M. Pavon. Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrodinger bridge, *Siam Review* 63 (2), pp. 249-313, 2021.
- [5] S. Cole, M. Eckstein, S. Friedland, K. Zyczkowski. On quantum optimal transport, [arXiv:2105.06922](https://arxiv.org/abs/2105.06922), 2022
- [6] M. Cuturi, G. Peyré. *Computational Optimal Transport: With Applications to Data Science (Foundations and Trends(r) in Machine Learning)*, Now Publishers, 2019.
- [7] A. Gerolin, N. Monina. *Non-commutative Optimal Transport for semi-definite positive matrices*, [arXiv:2309.04846](https://arxiv.org/abs/2309.04846), 2023.
- [8] A. Gerolin, A. Kausamo, T. Rajala. Multi-marginal entropy-transport with repulsive cost, [arXiv:1907.07900](https://arxiv.org/abs/1907.07900), 2019.
- [9] C. Villani. *Optimal Transport: Old and New*, Springer Berlin Heidelberg, 2008.
- [10] Yiling Xie, Yiling Luo, Xiaoming Huo. Solving a Special Type of Optimal Transport Problem by a Modified Hungarian Algorithm, [arXiv:2210.16645](https://arxiv.org/abs/2210.16645), 2023

- [11] M. Cuturi. *Sinkhorn Distances: Lightspeed Computation of Optimal Transport*, *Advances in Neural Information Processing Systems*, vol. 26, 2013

LAAS-CNRS, 7 AVENUE DU COLONEL ROCHE, 31077 TOULOUSE CÉDEX 4, FRANCE
Email address: `spchhatoi@laas.fr`

LAAS-CNRS AND TOULOUSE SCHOOL OF ECONOMICS (TSE), LAAS, 7 AVENUE DU COLONEL
ROCHE, 31077 TOULOUSE CÉDEX 4, FRANCE
Email address: `lasserre@laas.fr`