ENTROPY-REGULARIZED OPTIMAL TRANSPORT IN INFORMATION DESIGN

By

Jorge Justiniano, Andreas Kleiner, Benny Moldovanu,

Martin Rumpf and Philipp Strack

December 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2419

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Entropy-Regularized Optimal Transport in Information Design

Jorge Justiniano[1], Andreas Kleiner[2], Benny Moldovanu[2], Martin Rumpf[1], and Philipp Strack[3]

[1]Institute for Numerical Simulation, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany
jorge.justiniano@outlook.de, martin.rumpf@ins.uni-bonn.de
[2]Institute for Microeconomics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany
andreas.kleiner@uni-bonn.de, mold@uni-bonn.de
[3]Department of Economics, Yale University, New Haven CT 06520-8268
philipp.strack@yale.edu

arXiv:2412.09316v1 [math.NA] 12 Dec 2024

December 13, 2024

Abstract

In this paper, we explore a scenario where a sender provides an information policy and a receiver, upon observing a realization of this policy, decides whether to take a particular action, such as making a purchase. The sender's objective is to maximize her utility derived from the receiver's action, and she achieves this by careful selection of the information policy. Building on the work of Kleiner et al., our focus lies specifically on information policies that are associated with power diagram partitions of the underlying domain. To address this problem, we employ entropy-regularized optimal transport, which enables us to develop an efficient algorithm for finding the optimal solution. We present experimental numerical results that highlight the qualitative properties of the optimal configurations, providing valuable insights into their structure. Furthermore, we extend our numerical investigation to derive optimal information policies for monopolists dealing with multiple products, where the sender discloses information about product qualities.

1 Introduction

Mechanism Design is the branch of Game Theory/Economics that analyzes the design of optimal institutions (or so called *game-forms*) governing the interaction of a group composed of self-interested, strategic and privately informed agents. An optimal mechanism needs to solve an information aggregation problem and an incentive problem in order to achieve a desired collective decision (cf. the introductory textbook [5]). Major applications have been made to social choice and voting, to market design e.g., auctions and matching, and to contracting. Numerous Nobel Prizes have been awarded for both theoretical and applied work in this field, e.g. for Myerson's work on optimal auctions [19].

In classical mechanism design analysis, the private information available to the agents about the environment is exogenously given. A more recent branch of inquiry, called *Information Design*, takes the institution that governs the agents' interaction as given, but endogenizes the information structure: in turn, the latter is chosen in order to optimize some given goal.

A large literature within information design has focused on the *Bayesian Persuasion* problem. A particularly important class is that of moment Bayesian persuasion problems which have been studied by Kamenica and Gentzkow [12], Kolotilin [17] and Dworczak and Martini [10]. With few exceptions, the literature on these problems assumes that the setting is unidimensional. In this paper we deal with the multidimensional case and propose an efficient algorithm to compute optimal information designs.

Organization. This paper is organized as follows. In Section 2 the moment Bayesian persuasion model will be introduced. Using recent insights in the model it will be sufficient to optimize over the closure of Lipschitz-exposed points of a subset of measures induced by information policies. A particularly important class of these measures turns out to be represented by Laguerre partitions. Then, in Section 3 the optimization problem of moment Bayesian persuasion is transformed into an optimization over such Laguerre partitions along with a relaxation strategy that guarantees the existence of optimal designs. To compute these optimal

Laguerre partitions we recall in Section 4 the role of Laguerre diagrams in the context of semi-discrete optimal transport, and derive in Section 5 the associated entropy relaxed optimal transport formulation, which allows for a computationally efficient algorithm to optimize over Laguerre partions and thus solve the moment Bayesian persuasion problem numerically. In this context, we demonstrate the convergence of maximizers of the relaxed problems to a maximizer of the non-relaxed problem. To ensure computational reproducibility, the spacial discretization is presented in full detail in Section 6. Then, in Section 7 we show some qualitative features of the algorithm, and finally in Section 8 we use our algorithm to compute optimal information policies for a multi-product monopolist.

2 Information Design and the Moment Bayesian Persuasion Problem

In this section we will rigorously formulate the moment Bayesian persuasion problem. Let $(D, \mathcal{D}, \nu)$ be a probability space, where $D \subseteq \mathbb{R}^d$ is compact and convex, $\mathcal{D}$ is its Borel $\sigma-$algebra, and ν is a probability measure which is assumed to be absolutely continuous with respect to the Lebesgue measure $\mathcal{L}$. For later usage, we assume throughout this paper that the Radon-Nikodym density $\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}} \in L^\infty(\mathcal{L})$. The *state of the world* ω is a realization of a random variable that is distributed according to the *prior* ν. An informed *sender* wants to persuade an uninformed *receiver* to take an action that the sender prefers. The receiver's optimal action depends on her information about the state, and initially she only knows that the state of the world is distributed according to ν. The sender, who knows ν and observes the realization of the state of the world ω, may reveal information to the receiver about the realized state. This revelation may often be strategic if the goals of sender and receiver are different. For example, if the state is represented by a one-dimensional random variable, the sender could reveal to the receiver if the realized state is above or below a certain threshold but provide no additional information — this may be better for the sender than revealing all information.

Specifically, the sender commits to an *information policy* (S, π), where S is a measurable space, $\mathcal{P}(S)$ is the space of probability distributions on S, and $\pi : D \to \mathcal{P}(S)$. If the sender commits to information policy (S, π) and if the realized state of the world is ω, the receiver observes the information policy and the realization of a random variable with values in S that is drawn according to the probability distribution $\pi(\omega)$. The prior ν and the information policy π induce a joint probability distribution γ on $D \times S$ defined by $\gamma(E \times T) := \int_E \pi(\omega)(T)\,\mathrm{d}\nu(\omega)$ for any measurable sets $E \subseteq D$, $T \subseteq S$. After seeing a realization $s \in S$, the receiver updates her beliefs about the state of the world according to Bayes' rule to a *posterior belief* given by $\gamma(\cdot \times \{s\} | D \times \{s\}) \in \mathcal{P}(D)$, which is a regular conditional probability measure of γ. This posterior belief determines the receiver's optimal behaviour and implicitly the sender's payoff. The expected value of the posterior belief is called the *posterior mean*; therefore, each information policy induces a probability distribution over posterior means.

As an example, consider the fully revealing information policy given by $S = (D, \mathcal{D})$ and $\pi(\omega) = \delta_\omega$, where δ_ω denotes the Dirac measure at ω. Under the fully revealing information policy, the random variable observed by the receiver equals the realized state of the world with probability 1 and hence the receiver perfectly learns the state of the world. For an example of a partially informative information policy, let $\{B_1, \dots, B_n\}$ be a partition of D, $S = \left(\{1, \dots, n\}, 2^{\{1,\dots,n\}}\right)$, and $\pi(\omega) = \delta_i$ if $\omega \in B_i$. Under this information policy, the receiver learns in which partition element the realized state lies, and updates her prior belief accordingly.

In the moment Bayesian persuasion model it is assumed that the sender's payoff from inducing a posterior belief depends only on the posterior mean, and is given by an upper semicontinuous function $\Phi : D \to \mathbb{R}$. Note that, in reduced form, this formulation allows for the sender's payoff to depend on an action taken by the receiver. To characterize the distributions of posterior means that can be induced by some information policy we use the following stochastic dominance concept.

Definition 2.1 (Shaked and Shanthikumar [24], Cartier et al. [7] and Phelps [21])**.** For measures ν and ρ, we say that ν *dominates* ρ *in the convex order* (or that ρ is a *fusion* of ν), denoted by $\nu \succeq \rho$, if $\int \psi \,\mathrm{d}\nu \geq \int \psi \,\mathrm{d}\rho$ for all convex functions ψ for which both integrals exist. We write $\nu \succ \rho$ if ν dominates ρ in the convex order and $\nu \neq \rho$. We denote by $F_\nu = \{\rho : \rho \preceq \nu\}$ the set of fusions of a given measure ν.

Any information policy necessarily reveals weakly less information than the state of the world and the generated distribution of posterior means is dominated in the convex order by the prior. Conversely, for any probability measure $\rho \preceq \nu$ there exists an information policy that generates ρ as its probability measure of posterior means (see Blackwell [4] and Strassen [25]).

Therefore, instead of modeling the sender's choice of an information policy we can assume that the sender directly chooses a probability measure that is dominated in the convex order by ν. The sender chooses such a

probability measure to maximize her expected payoff, and therefore solves the problem

$$\max_{\lambda \in F_\nu} \int_D \Phi(y)\, \mathrm{d}\lambda(y). \tag{2.1}$$

This is the information design problem in the focus of this paper.

To deduce qualitative properties of solutions, let us recall some basic concepts of convex analysis. An *extreme* point of a convex set A is a point $y \in A$ that cannot be represented as a convex combination of two other points in A. I.e. $y \in A$ is an extreme point of A if $y = \alpha w + (1-\alpha)z$, for $w, z \in A$ and $\alpha \in [0,1]$ jointly imply that $y = w$ or $y = z$.

The *Krein–Milman Theorem* states that any convex and compact set A in a locally convex space is the closed, convex hull of its extreme points. In particular, such a set has extreme points. The usefulness of extreme points for optimization stems from *Bauer's Maximum Principle*, which states that a convex, upper semicontinuous functional on a non-empty, compact and convex set A of a locally convex space attains its maximum at an extreme point of A. An element y of a convex set A is called *exposed* if there exists a supporting hyperplane H such that $H \cap A = \{y\}$ or, equivalently, there is a continuous, linear functional that attains its unique maximum on A at y. Every exposed point is extreme, but the converse is not true in general.

The set of fusions F_ν appearing in the above maximization problem is convex and compact in the weak* topology of measures. As the objective is linear in the measure λ, a maximum is attained at one of the extreme points of F_ν. It is thus of interest to further explore the structure of extreme and exposed points, and we focus here on those measures that have a finite support.

For any measure ρ on D and a measurable set $B \subseteq D$, we denote by $\rho|_B$ the restriction of ρ to B. As stated in the following theorem a key feature of any extremal measure in F_ν with finite support is that there is a partition of the domain D into convex sets B such that all the original mass restricted to B, $\nu(B)$, remains within B and is fused into a measure $\rho|_B$ whose support is an affinely independent set of points.

Theorem 2.2 (Kleiner et al. [16])**.** *Let $D \subseteq \mathbb{R}^n$ be compact and convex, and let ν be an absolutely continuous probability measure on D. Suppose that ρ is an extreme point of F_ν with finite support. Then there exists a finite partition $\mathcal{P}$ of D into convex sets such that, for each $B \in \mathcal{P}$, $\rho|_B \preceq \nu|_B$ and $\rho|_B$ has affinely independent support.*

The above result generalizes a unidimensional result found in Kleiner et al. [15] and Arieli et al. [2]. In fact, more precise information is available about the geometric structure of the following subset of extreme points.

Definition 2.3. A measure $\rho \in F_\nu$ is a *Lipschitz-exposed* point of F_ν if there exists a Lipschitz-continuous function $\Phi\colon D \to \mathbb{R}$ such that ρ is the unique solution to the problem

$$\max_{\lambda \in F_\nu} \int \Phi(y)\, \mathrm{d}\lambda(y).$$

The following theorem provides a characterization of Lipschitz-exposed points with finite support using Laguerre diagrams, also known as *power diagrams*, see Aurenhammer [3]. Given an n-tupel $\mathbf{X} = (x_1, \ldots, x_n)$ of pairwise distinct sites $x_i \in \mathbb{R}^d$ and a weight vector $\mathbf{g} = (\mathrm{g}_1, \ldots, \mathrm{g}_n) \in \mathbb{R}^n$, the associated Laguerre cells $L_i[\mathbf{X}, \mathbf{g}] \in \mathcal{D}$ are the convex polyhedra

$$L_i[\mathbf{X}, \mathbf{g}] \coloneqq \{y \in D : |y - x_i|^2 - \mathrm{g}_i \le |y - x_j|^2 - \mathrm{g}_j \ \ \forall \ \ 1 \le j \le n\}, \tag{2.2}$$

for $i = 1, \ldots, n$. The ensemble of all Laguerre cells forms the Laguerre diagram.

Theorem 2.4 (Kleiner et al. [16])**.** *Let $D \subseteq \mathbb{R}^n$ be compact and convex, and let ν be a probability measure with full support on D that is absolutely continuous with respect to the Lebesgue measure. Let $\rho \in F_\nu$ have finite support. Then ρ is a Lipschitz-exposed point of F_ν if and only if there exists a Laguerre diagram $\mathcal{P}$ of D such that, for all probability measures λ, if $\lambda|_L \preceq \nu|_L$ for all $L \in \mathcal{P}$ and $\mathrm{supp}\,(\lambda) \subseteq \mathrm{supp}\,(\rho)$ then $\lambda = \rho$.*

In particular, $\rho \in F_\nu$ is a finitely supported Lipschitz-exposed point if there is a Laguerre diagram such that for each Laguerre cell L with non-vanishing measure $\nu(L)$, the relation $\rho|_L \preceq \nu|_L$ holds and if the support of $\rho|_L$ is affinely independent. Indeed, in this case $\lambda|_L \preceq \nu|_L$ implies that $\lambda|_L$ and $\rho|_L$ have the same mean. In turn, this implies that if $\mathrm{supp}\,(\lambda) \subseteq \mathrm{supp}\,(\rho)$ and if the support of ρ is affinely independent then $\rho|_L = \lambda|_L$.

For any compact convex set in a normed vector space, the set of exposed points is dense in the set of extreme points of F_ν (Klee, [14]). Therefore, to optimize a continuous linear objective functional on such a set, it would be sufficient to optimize over the closure of exposed points. In the following, we will focus on the subset of exposed points that are Lipschitz-exposed and where, on each cell L of the corresponding Laguerre diagram, the

mass within the respective cell is fused to a unique mass point. These extreme points have received considerable attention in the Economics literature in one-dimensional settings (see for example Dworczak and Martini [10] or Ivanov [11]). They are simple to use in practice because they can be implemented by only revealing, for a partition of the state space into convex sets, in which partition element the realized state lies. Finally, the relation to Laguerre diagrams makes these exposed points numerically tractable. Our computational method derived below focuses on solving the sender's problem among such exposed points.

3 The optimization task

Motivated by Theorem 2.4 we discuss in this section the optimization problem (2.1) of moment Bayesian persuasion as an optimization over Laguerre partitions along with a relaxation strategy. At first, we study the case where the sender commits to an information policy induced by general partitions. To this end, we consider a probability space $(D, \mathcal{D}, \nu)$ on a compact and convex domain $D \subset \mathbb{R}^d$ for $d \geq 2$, equipped with a probability measure $\nu \in \mathcal{P}(D)$ on the Borel-σ-algebra $\mathcal{D}$, which is assumed to be absolutely continuous with respect to the Lebesgue measure $\mathcal{L}$ with Radon-Nikodym density $\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}} \in L^\infty(\mathcal{L})$. For fixed $n \in \mathbb{N}$, let $(B_i)_{i=1,\dots,n} \subset \mathcal{D}$ be a ν-partition of D, i.e. $\nu(B_i \cap B_j) = 0$ for $i \neq j$, and $\nu\left(\cup_{i=1}^n B_i\right) = 1$. Let $(S, \mathcal{S}) := \left(\{1, \dots, n\}, 2^{\{1,\dots,n\}}\right)$, and assume the sender commits to the information policy $\pi : D \to \mathcal{P}(S)$, with $\pi(\omega) = \delta_i \in \mathcal{P}(S)$ if and only if $\omega \in B_i$. After receiving the signal realization $i \in S$, the receiver updates her belief to the posterior $\gamma(\cdot \times \{i\} \mid D \times \{i\})$. It holds that

$$\gamma(B \times \{i\} \mid D \times \{i\}) = \frac{\int_B \pi(\omega)(\{i\})\,\mathrm{d}\nu(\omega)}{\int_D \pi(\omega)(\{i\})\,\mathrm{d}\nu(\omega)} = \frac{\int_B \mathbb{1}_{B_i}(\omega)\,\mathrm{d}\nu(\omega)}{\int_D \mathbb{1}_{B_i}(\omega)\,\mathrm{d}\nu(\omega)} = \frac{\nu(B \cap B_i)}{\nu(B_i)} = \nu(B|B_i).$$

For any non-ν-null set $B \in \mathcal{D}$, we define the ν-barycenter

$$b[B] := \frac{\int_B y \;\mathrm{d}\nu(y)}{\nu(B)}.$$

The ν-barycenter of B_i coincides then with the mean of the posterior $\nu(\,\cdot\,|B_i)$. To conclude, the posterior mean distribution generated by the information policy π is given by $\sum_{i=1}^n \nu(B_i)\delta_{b[B_i]} \in F_\nu$. In this case, for a continuous function $\Phi : D \to \mathbb{R}$ the functional (2.1) takes the following explicit form:

$$\sum_{\substack{i=1,\dots,n \\ \nu(B_i)>0}} \nu(B_i)\Phi(b[B_i]). \tag{3.1}$$

Now, in the light of Section 2 we focus on partitions described by Laguerre diagrams. In this case, we may optimize the cost functional (3.1) directly on the parameters $(\mathbf{X}, \mathbf{g})$ describing Laguerre cells

$$\mathcal{F}_n[\mathbf{X}, \mathbf{g}] := \sum_{\substack{i=1,\dots,n \\ m_i[\mathbf{X},\mathbf{g}]>0}} m_i[\mathbf{X}, \mathbf{g}]\Phi(b_i[\mathbf{X}, \mathbf{g}]), \tag{3.2}$$

where $m_i[\mathbf{X}, \mathbf{g}] := \nu(L_i[\mathbf{X}, \mathbf{g}])$, and $b_i[\mathbf{X}, \mathbf{g}] := b[L_i[\mathbf{X}, \mathbf{g}]]$ (cf. equation (2.2)), to be maximized over n-tuples of pairwise distinct sites $\mathcal{X}_n := \{\mathbf{X} = (x_1, \dots, x_n) \in \left(\mathbb{R}^d\right)^n : x_i \in \mathbb{R}^d,\ x_i \neq x_j \ \text{ for } \ i \neq j\}$ and weight vectors $\mathbf{g} \in \mathbb{R}^n$. Note that Laguerre cells might be sets of vanishing measure. Since Laguerre cells are by definition a ν-partition of D, the set of associated characteristic functions $\chi_{L_i[\mathbf{X},\mathbf{g}]}$ of the Laguerre cells $L_i[\mathbf{X}, \mathbf{g}]$ for $i = 1, \dots, n$ forms a partition of unity of D, i.e. $\sum_{i_1,\dots n} \chi_{L_i[\mathbf{X},\mathbf{g}]} = 1$ a.e. in D. We define the power diagram associated with an n-tuple of sites $\mathbf{X}$ and a weight vector $\mathbf{g}$ as

$$\boldsymbol{L}[\mathbf{X}, \mathbf{g}] := (\chi_{L_1[\mathbf{X},\mathbf{g}]}, \dots, \chi_{L_n[\mathbf{X},\mathbf{g}]}). \tag{3.3}$$

Along maximizing sequences for the functional $\mathcal{F}_n[\cdot, \cdot]$ it might happen that subsets of Laguerre cells collapse, or that sites as well as weights diverge. Here, the notion of power diagram as partitions of unity helps to deal with these degenerate cases. At first, we obtain the following relative compactness result:

Lemma 3.1. *The set $\mathscr{L} := \boldsymbol{L}(\mathcal{X}_n \times \mathbb{R}^n)$ is relatively compact in $L^1(\nu)^n$ and any limit of a converging sequences in $\mathscr{L}$ is a partition of unity a.e. on D.*

Proof. Let ν be trivially extended onto $\mathbb{R}^d$ with density 0 outside D. Each non-empty Laguerre cell $L_i[\mathbf{X}, \mathbf{g}]$ is convex and its boundary in the interior of D is polygonal and consists of at most $n-1$ planar interfaces. Each

of these interior interfaces has at most a $\mathcal{H}^{d-1}$ measure $\operatorname{diam}(D)^{d-1}$ inside of D. Thus, for $h \in \mathbb{R}$ we observe that

$$\|\chi_{L_i[\mathbf{X},\mathbf{g}]}(\cdot - h) - \chi_{L_i[\mathbf{X},\mathbf{g}]}\|_{L^1(\nu)} \leq (n-1) \left\| \tfrac{\mathrm{d}\nu}{\mathrm{d}x} \right\|_\infty \operatorname{diam}(D)^{d-1} h\,.$$

This, together with the compactness of D, implies by the Fréchet-Kolmogorov theorem the relative compactness of $\mathbf{L}(\mathcal{X}_n \times \mathbb{R}^n)$ in $L^1(\nu)$. Let $(\mathbf{X}^k)_k \subset \mathcal{X}_n$, and $(\mathbf{g}^k)_k \subset \mathbb{R}^n$. Since $\chi_{L_i[\mathbf{X}^k,\mathbf{g}^k]} = 0$ outside D any $L^1(\nu)$-limit of $(\chi_{L_i[\mathbf{X}^k,\mathbf{g}^k]})_{i=1,\dots,n}$ for $k \to \infty$ is of the form $\big(\chi^1, \dots, \chi^n\big) \in L^1(\nu; \{0,1\})^n$ and $\operatorname{supp} \chi_i \subseteq D$. Finally, $\|\sum_{i=1,\dots,n} \chi^i - 1\|_{L^1} = \lim_{k\to\infty} \|\sum_{i=1,\dots,n} \chi_{L_i[\mathbf{X}^k,\mathbf{g}^k]} - 1\|_{L^1} = 0$ and thus $\big(\chi^1, \dots, \chi^n\big)$ is a partition of unity. □

Let us denote by $\mathscr{L}^c$ the closure of $\mathscr{L} = \boldsymbol{L}(\mathcal{X}_n \times \mathbb{R}^n) \in L^1(\nu; \{0,1\})^n$. For $\boldsymbol{L}^c = \big(\chi^1, \dots, \chi^n\big) \in \mathscr{L}^c$ we define the relaxed functional

$$\mathcal{F}_n^c[\boldsymbol{L}^c] := \sum_{\substack{i=1,\dots,n \\ m[\chi^i]>0}} m[\chi^i]\, \Phi(b[\chi^i]) \tag{3.4}$$

where we define with a slight misuse of notation the ν-mass $m[\chi] := \int_D \chi \,\mathrm{d}\nu$ of a characteristic function χ and its ν-barycenter $b[\chi] := m[\chi]^{-1} \int_D x\, \chi \,\mathrm{d}\nu$ for $m[\chi] > 0$. Let us remark that some of the χ^i in $\boldsymbol{L}^c$ (but not all) might have zero mass $m[\chi^i]$. For this relaxed functional we obtain the following existence result of a maximum.

Theorem 3.2. *Assume Φ to be a upper semicontinuous function on D. Then $\mathcal{F}_n^c$ attains its maximum on $\mathscr{L}^c$.*

Proof. At first, we recall that upper semicontinuous functions on compact domains are bounded from above. Let $\bar{\Phi}$ denote the maximum of Φ on D, which exists due to the upper semicontinuity of Φ. Let

$$\big(\boldsymbol{L}_k^c = (\chi_k^1, \dots, \chi_k^n)\big)_{k\in\mathbb{N}} \subset \mathscr{L}^c$$

be a maximizing sequence of (3.4). Due to the compactness of $\mathscr{L}^c$, we obtain that, up to the selection of a subsequence, $(\chi_k^1, \dots, \chi_k^n)$ converges in $L^1(\nu)$ to a limit $\boldsymbol{L}^c = (\chi^1, \dots, \chi^n)$. For all $i \in \{1, \dots, n\}$ with $m[\chi^i] = 0$ it holds that $m[\chi_k^i] \to 0$ for $k \to \infty$ and thus $m[\chi_k^i]\Phi(b[\chi^i]) \leq m[\chi_k^i]\bar{\Phi} \to 0$. For $m[\chi^i] > 0$ the sequence of barycenters $\big(b[\chi_k^i]\big)_{k\in\mathbb{N}}$ converges to $b[\chi^i]$. Taking into account the upper semicontinuity of Φ, this implies that $\limsup_{k\to\infty} \Phi(b[\chi_k^i]) \leq \Phi(b[\chi^i])$. Altogether, we obtain

$$\limsup_{k\to\infty} \mathcal{F}_n^c[\boldsymbol{L}_k^c] = \limsup_{k\to\infty} \sum_{\substack{i=1,\dots,n \\ m[\chi_k^i]>0}} m[\chi_k^i]\, \Phi(b[\chi_k^i]) \leq \sum_{\substack{i=1,\dots,n \\ m[\chi^i]>0}} m[\chi^i]\, \Phi(b[\chi^i]) = \mathcal{F}_n^c[\boldsymbol{L}^c]\,.$$

Hence, the relaxed functional $\mathcal{F}_n^c$ attains its maximum on $\mathscr{L}^c$ at $\boldsymbol{L}^c$. □

Remark 3.3. The relation to the concept of stochastic dominance is as follows. As an upper semicontinuous function, Φ is a ν-measurable function on D. Given a polyhedral partition of D into Laguerre cells $(L_i[\mathbf{X},\mathbf{g}])_{i=1,\dots,n}$, one may collapse the mass of each cell at its barycenter. This induces an atomic probability measure

$$\rho := \sum_{\substack{i=1,\dots,n \\ m_i[\mathbf{X},\mathbf{g}]>0}} m_i[\mathbf{X},\mathbf{g}]\, \delta_{b_i[\mathbf{X},\mathbf{g}]}$$

on D. Recall that a probability measure ν dominates a probability measure ρ in convex order if and only if $\mathbb{E}_{X\sim\nu}[\Phi(X)] \geq \mathbb{E}_{X\sim\rho}[\Phi(X)]$ for all convex functions $\Phi : \mathbb{R}^d \to \mathbb{R}$ such that both expectations exist. By Jensen's inequality, this indeed holds for the initial probability measure ν and the atomic probability measure ρ considered here. Hence, in our ansatz we consider the expected value of the given function Φ with respect to an atomic measure associated with some power diagram $\boldsymbol{L}^c$ and maximize this expected value over all atomic measures induced by power diagrams $\boldsymbol{L}^c$.

To avoid the relaxation, one might consider hard constraints to ensure that pairwise distances between sites and cell masses do not vanish in the limit along a maximizing sequence. Alternatively, a penalty formulation can be used as a more robust and effective alternative. To this end, we define for a penalty parameter $\eta > 0$

$$\mathcal{F}_n^\eta(\mathbf{X},\mathbf{g}) = \mathcal{F}_n(\mathbf{X},\mathbf{g}) - \eta \mathcal{R}_n(\mathbf{X},\mathbf{g}), \tag{3.5}$$

where the penalty term $\mathcal{R}_n$ is given as

$$\mathcal{R}_n(\mathbf{X},\mathbf{g}) := \sum_{\substack{i=1,\dots,n \\ m_i[\mathbf{X},\mathbf{g}]>0}} \int_D |y - x_i|^2 \chi_{L_i[\mathbf{X},\mathbf{g}]}(y)\,\mathrm{d}\nu(y) + \sum_{\substack{1\leq i,j\leq n \\ i\neq j \\ m_i[\mathbf{X},\mathbf{g}],m_j[\mathbf{X},\mathbf{g}]>0}} \frac{m_i[\mathbf{X},\mathbf{g}] m_j[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2}. \tag{3.6}$$

Here, the cell masses as scaling factors and the characteristic functions of cells as weight functions are to be understood as the natural scaling of the corresponding penalty terms. In particular, we observe a stronger penalization of the drift of sites away from the associated cell and the fusion of pairs of sites in case of larger cell masses.

It might happen that in the limit along a maximizing sequence $(\mathbf{X}^N, \mathbf{g}^N)_N \subset \mathcal{X}_l \times \mathbb{R}^l$ of $\mathcal{F}_l^\eta$ for some $l \in \mathbb{N}$ Laguerre cells collapse and the effective number of cells decreases. The following existence theorem takes this into account.

Theorem 3.4. *Let us assume that $\Phi \in C(D)$ with maximum $\bar{\Phi}$ on D. Then, for given number of sites $l \in \mathbb{N}$ there exists an $n \leq l$, such that a maximizer $(\mathbf{X}^*, \mathbf{g}^*)$ of $\mathcal{F}_n^\eta$ exists with $m_i[\mathbf{X}^*, \mathbf{g}^*] > 0$ for $i = 1, \dots, n$ and $\mathcal{F}_n^\eta(\mathbf{X}^*, \mathbf{g}^*) \geq \sup_{(\mathbf{X},\mathbf{g})} \mathcal{F}_l^\eta(\mathbf{X}, \mathbf{g})$.*

Proof. Consider a maximizing sequence for $\mathcal{F}_l^\eta$. Since $m_i[\mathbf{X}^N, \mathbf{g}^N] \in [0,1]$, and $b_i[\mathbf{X}^N, \mathbf{g}^N] \in D$ for $m_i[\mathbf{X}^N, \mathbf{g}^N] > 0$, we may assume that, up to the selection of a subsequence, the sequences of masses and barycenters converge to limits m_i^* and b_i^*, respectively, and that there exists an $n \leq l$ with $m_i^* > 0$ if and only if $i \leq n$. Then, we obtain

$$\begin{aligned}
\sup_{(\mathbf{X},\mathbf{g})} \mathcal{F}_l^\eta(\mathbf{X}, \mathbf{g}) &= \lim_{N\to\infty} \mathcal{F}_l^\eta(\mathbf{X}^N, \mathbf{g}^N) \leq \limsup_{N\to\infty} \mathcal{F}_l(\mathbf{X}^N, \mathbf{g}^N) - \eta \liminf_{N\to\infty} \mathcal{R}_l(\mathbf{X}^N, \mathbf{g}^N) \\
&\leq \limsup_{N\to\infty} \mathcal{F}_n(\mathbf{X}^N, \mathbf{g}^N) + \limsup_{N\to\infty} \sum_{\substack{i=n+1,\dots,l \\ m_i[\mathbf{X}^N,\mathbf{g}^N]>0}} m_i[\mathbf{X}^N, \mathbf{g}^N]\Phi(b_i[\mathbf{X}^N, \mathbf{g}^N]) - \eta \liminf_{N\to\infty} \mathcal{R}_n(\mathbf{X}^N, \mathbf{g}^N) \\
&\quad - \eta \liminf_{N\to\infty} \big(\mathcal{R}_l(\mathbf{X}^N, \mathbf{g}^N) - \mathcal{R}_n(\mathbf{X}^N, \mathbf{g}^N)\big) \\
&= \lim_{N\to\infty} \mathcal{F}_n(\mathbf{X}^N, \mathbf{g}^N) + 0 - \eta \lim_{N\to\infty} \mathcal{R}_n(\mathbf{X}^N, \mathbf{g}^N) - \eta \liminf_{N\to\infty} \big(\mathcal{R}_l(\mathbf{X}^N, \mathbf{g}^N) - \mathcal{R}_n(\mathbf{X}^N, \mathbf{g}^N)\big) \\
&\leq \lim_{N\to\infty} \mathcal{F}_n(\mathbf{X}^N, \mathbf{g}^N) - \eta \lim_{N\to\infty} \mathcal{R}_n(\mathbf{X}^N, \mathbf{g}^N) = \lim_{N\to\infty} \mathcal{F}_n(\mathbf{X}^N, \mathbf{g}^N) - \eta \mathcal{R}_n(\mathbf{X}^N, \mathbf{g}^N) \\
&= \lim_{N\to\infty} \mathcal{F}_n^\eta(\mathbf{X}^N, \mathbf{g}^N).
\end{aligned}$$

Hence, if it exists, a maximizer of $\mathcal{F}_n^\eta$ attains a greater or equal objective value than a maximizer of $\mathcal{F}_l^\eta$.

Now, let $(\mathbf{X}^N, \mathbf{g}^N)_N \subset \mathcal{X}_n \times \mathbb{R}^n$ be a maximizing sequence for $\mathcal{F}_n^\eta$ and in analogy to before assume that $m_i[\mathbf{X}^N, \mathbf{g}^N]$ and $b_i[\mathbf{X}^N, \mathbf{g}^N]$ have limits $m_i^* > 0$, and $b_i^* \in D$ for $i = 1, \dots, n$, that $\mathcal{F}_n^\eta(\mathbf{X}^N, \mathbf{g}^N)$ is monotonically increasing in N and that $m_i[\mathbf{X}^N, \mathbf{g}^N] \geq \frac{1}{2} m_i^*$. Consequently, for $\bar{\Phi}$ being the maximal value of Φ on D the estimate

$$\begin{aligned}
\mathcal{F}_n^\eta(\mathbf{X}^0, \mathbf{g}^0) &\leq \nu(D)\bar{\Phi} - \eta \left(\sum_{i=1,\dots,n} \int_D |y - x_i^N|^2 \chi_{L_i[\mathbf{X}^N,\mathbf{g}^N]} \, \mathrm{d}\nu(y) + \sum_{\substack{1\leq i,j\leq n \\ i\neq j}} \frac{m_i[\mathbf{X}^N, \mathbf{g}^N] m_j[\mathbf{X}^N, \mathbf{g}^N]}{|x_i^N - x_j^N|^2} \right) \\
&\leq \bar{\Phi} - \eta \left(\frac{1}{2} \sum_{i=1,\dots,n} \mathrm{dist}^2(x_i^N, D) m_i^* + \frac{1}{4} \sum_{\substack{1\leq i,j\leq n \\ i\neq j}} \frac{m_i^* m_j^*}{|x_i^N - x_j^N|^2} \right)
\end{aligned} \tag{3.7}$$

is obtained. This implies the following a priori bounds:

$$\mathrm{dist}\,(x_i^N, D) \leq \sqrt{\frac{2(\bar{\Phi} - \mathcal{F}_n^\eta(\mathbf{X}^0, \mathbf{g}^0))}{\eta m_i^*}}, \quad |x_i^N - x_j^N| \geq \sqrt{\frac{\eta m_i^* m_j^*}{4(\bar{\Phi} - \mathcal{F}_n^\eta(\mathbf{X}^0, \mathbf{g}^0))}}$$

for all N. For the uniform bound on $\mathbf{g}$, recall that $\mathbf{g}$ and $\mathbf{g} + \lambda \mathbb{1}_n$ both induce the same Laguerre diagram. Hence, we may assume without loss of generality that $\mathrm{g}_1^N = 0$ for all $N \in \mathbb{N}$. Then, $\lim_{N\to\infty} \mathrm{g}_j^N = \infty$ for $j = 2, \dots, n$ would imply $\lim_{N\to\infty} m_1[\mathbf{X}^N, \mathbf{g}^N] = 0$, which contradicts our choice of n. Similarly, $\lim_{N\to\infty} \mathrm{g}_j = -\infty$ would imply $\lim_{N\to\infty} m_j[\mathbf{X}^N, \mathbf{g}^N] = 0$, which again is a contradiction. Hence, $|\mathbf{g}^N| \leq C$ for some $C > 0$ and all $N \in \mathbb{N}$. Finally, given these a priori bounds, the existence of a maximizer of $\mathcal{F}_n^\eta$ follows directly from the Weierstraß extreme value theorem. □

4 Semi-discrete optimal transport revisited

In the previous section we stated existence results of maximizers of the sender's revenue over the class of Laguerre partitions. To compute these optimal Laguerre partitions we recall in this section the connection of Laguerre diagrams to solutions of semi-discrete optimal transport problems and the associated dual formulation.

In optimal transport theory, one considers optimal couplings $\Pi \in \mathcal{P}(D \times D) \in U(\mu, \nu)$ of two probability measures $\mu, \nu \in \mathcal{P}(D)$. Here, $\mathcal{U}(\mu, \nu)$ is the set $\Pi \in \mathcal{P}(D \times D)$ with $\Pi(A \times D) = \mu(A)$ and $\Pi(D \times A) = \nu(A)$ for all Borel sets $A \subset D$. For the given cost function $(x, y) \mapsto |x - y|^2$ on $D \times D$ measuring the cost of transport from x to y on D, a coupling Π is optimal if it minimizes

$$\mathcal{W}^2[\mu, \nu] := \inf_{\Pi \in \mathcal{U}(\mu,\nu)} \int_{D \times D} |x - y|^2 \, \mathrm{d}\Pi(x, y). \tag{4.1}$$

The functional $\mathcal{W}(\mu, \nu)$ is called the 2-Wasserstein distance and defines a metric between the probability measures μ and ν in $\mathcal{P}(D)$ (cf. [22, Chapter 5]). In semi-discrete optimal transport the measure μ is assumed to be a discrete (empirical) probability measure, i.e.

$$\mu := \sum_{i=1,\dots,n} m_i \delta_{x_i}$$

with $\sum_{i=1,\dots n} m_i = 1$ and $m_i \geq 0$. The minimization of (4.1) is a constrained linear minimization problem and denoted the primal Kantorovich problem. As such, it can naturally be paired with a constrained linear maximization problem as the dual problem, the dual Kantorovich problem [1, Section 6.1]. We obtain

$$\mathcal{W}^2[\mu, \nu] = \sup_{(f,\mathbf{g}) \in \mathcal{R}} \mathbf{m} \cdot \mathbf{g} + \int_D f(y) \, \mathrm{d}\nu(y), \tag{4.2}$$

where $\mathbf{m} := (m_1, \dots, m_n)$, $\mathbf{g} := (\mathrm{g}_1, \dots, \mathrm{g}_n)$, and

$$\mathcal{R} := \{(f, \mathbf{g}) \in C(D) \times \mathbb{R}^n : f(y) + \mathrm{g}_i \leq |x_i - y|^2 \text{ for all } i = 1, \dots, n\}.$$

For given $\mathbf{g} \in \mathbb{R}^n$ we obtain for the optimal f which is consistent with the constrained condition $(f, \mathbf{g}) \in \mathcal{R}$ that $f(y) = \mathbf{g}^C(y)$ for all $y \in D$ with

$$\mathbf{g}^C(y) := \min_{1 \leq i \leq n} |y - x_i|^2 - \mathrm{g}_i. \tag{4.3}$$

Given the C-transform $\mathbf{g}^C : D \to \mathbb{R}^D$ one can reformulate the dual Kantorovich problem (4.2) as the unconstrained convex program

$$\mathcal{W}^2[\mu, \nu] = \max_{\mathbf{g} \in \mathbb{R}^n} \mathcal{D}[\mathbf{g}] \tag{4.4}$$

$$\mathcal{D}[\mathbf{g}] := \int_D \mathbf{g}^C(y) d\nu(y) + \mathbf{m} \cdot \mathbf{g} = \sum_{i=1,\dots,n} \int_{L_i[\mathbf{X},\mathbf{g}]} (|y - x_i|^2 - \mathrm{g}_i) d\nu(y) + \mathbf{m} \cdot \mathbf{g}, \tag{4.5}$$

where $L_i[\mathbf{X}, \mathbf{g}]$ are the Laguerre cells associated with the weight vector $\mathbf{g}$ and the vector of fixed sites $\mathbf{X}$ defined in (2.2). Hence, solving (4.5) for ν and $\mathbf{m}$ consists in finding a Laguerre cell partition of D described via the weight vector $\mathbf{g} \in \mathbb{R}^n$ with cells centered at the given $\mathbf{X}$, and with $\mathbf{m} = (m_i)_{i=0,\dots,n}$ and $m_i = m(L_i[\mathbf{X}, \mathbf{g}])$ for all $i \in \{1, \dots, n\}$. In what follows, we will need to differentiate the functions $\mathcal{F}_n^\eta$ defined in (3.2) with respect to the weights g_j. For the differentiation of $\mathcal{D}(\mathbf{g})$ we obtain

$$\partial_{g_j} \mathcal{D}[\mathbf{g}] = -\int_D \chi_{L_j[\mathbf{X},\mathbf{g}]} \, \mathrm{d}\nu(y) + m_j. \tag{4.6}$$

5 Entropy regularization

Let us recall that our goal is to maximize the function $\mathcal{F}_n^\eta$ via an optimization of the Laguerre cells $L_j[\mathbf{X}, \mathbf{g}]$ described in terms of the sites x_j and the weights g_j for $j = 1, \dots, n$. In general, changing the sites and the weights fosters topological changes in the diagram's topology, leading to a challenging combinatorial optimization problem. To avoid this, we introduce in this section the entropy relaxed optimal transport formulation [20] of the associated semi-discrete optimal transport described in Section 3, which will allow us to derive a computationally efficient algorithm to optimize over Laguerre partitions and thus solve the moment Bayesian persuasion problem numerically. An extensive overview on entropy regularization of optimal transport is also given by Chewi et. al. in [9].

We begin by considering the regularized Wasserstein distance

$$\mathcal{W}_\varepsilon^2[\mu,\nu] := \inf_{\Pi\in\mathcal{U}(\mu,\nu)} \int_{\{x_1,\dots,x_n\}\times D} c(x,y)\,\mathrm{d}\Pi(x,y) + \varepsilon \mathrm{KL}[\Pi|\xi] \tag{5.1}$$

with transport cost $c(x,y) = |x-y|^2$ from x to y and for a regularization parameter $\varepsilon > 0$. We consider a measure $\xi \in \mathcal{P}(\{x_1,\dots,x_n\}\times D)$ with $\operatorname{supp}\mu\otimes\nu \subseteq \operatorname{supp}\xi$ for μ and ν as above, and we define the Kullback-Leibler divergence

$$\mathrm{KL}[\Pi|\xi] := \int_{\{x_1,\dots,x_n\}\times D} \log\left(\frac{\mathrm{d}\Pi}{\mathrm{d}\xi}(x,y)\right)\,\mathrm{d}\Pi(x,y) + (\,\mathrm{d}\xi(x,y) - \,\mathrm{d}\Pi(x,y)) \tag{5.2}$$

between the measures Π and ξ in the case that Π is absolutely continuous with respect to ξ, and otherwise $\mathrm{KL}[\Pi|\xi] := \infty$. The Kullback-Leibler divergence is a concave functional measuring the dissimilarity of the measures Π and ξ and acts here as a regularizing entropy functional. The standard choice for ξ is $\xi = \mu\otimes\nu$. In fact, as long as the above support property holds the functional in (5.1) only changes by an additive constant and hence, the minimizer remains the same. To simplify the optimization algorithm for an entropy regularization of the functional $\mathcal{F}_l^\eta$ defined in (3.5) $\xi = \mathcal{N}\otimes\mathcal{L}$ is a particularly suitable choice where $\mathcal{N} = \sum_{j=1,\dots,n}\delta_{x_j}$ is the counting measure on the support of μ and $\mathcal{L}$ the Lebesque measure. For this choice we proceed as follows.

Associated with the constraint optimization problem (5.1) is the Lagrangian

$$\begin{aligned}\mathscr{L}^\varepsilon(\Pi,f,g) = &\int_{D\times D} c(x,y)\,\mathrm{d}\Pi(x,y) + \varepsilon\int_{D\times D}\left(\log\frac{\mathrm{d}\Pi}{\mathrm{d}\xi} - 1\right)(x,y)\,\mathrm{d}\Pi(x,y)\\ &- \left(\int_{D\times D} f(y)\,\mathrm{d}\Pi(x,y) - \int_D f(y)\,\mathrm{d}\nu(y)\right) - \left(\int_{D\times D} g(x)\,\mathrm{d}\Pi(x,y) - \int_D g(x)\,\mathrm{d}\mu(x)\right).\end{aligned}$$

Now, assume that Π has the density p with respect to ξ, i.e. $\mathrm{d}\Pi(x,y) = p(x,y)\,\mathrm{d}\xi(x,y)$. Then, a necessary condition of a saddle point is that the derivative of $\mathscr{L}^\varepsilon$ vanishes in all directions $q : (x,y)\mapsto q(x,y)$. Hence,

$$\begin{aligned}0 = \left(\frac{\partial\mathscr{L}^\varepsilon}{\partial p}(p)\right)(q) = &\int_{D\times D} c(x,y)q(x,y)\,\mathrm{d}\xi(x,y) + \varepsilon\int_{D\times D}(\log p(x,y) - 1)\,q(x,y)\,\mathrm{d}\xi(x,y)\\ &+ \varepsilon\int_{D\times D} q(x,y)\,\mathrm{d}\xi(x,y) - \int_{D\times D} f(y)q(x,y)\,\mathrm{d}\xi(x,y) - \int_{D\times D} g(x)q(x,y)\,\mathrm{d}\xi(x,y).\end{aligned} \tag{5.3}$$

Hence, we obtain $c(x,y) + \varepsilon\log(p(x,y)) - f(y) - g(x) = 0$ pointwise, or equivalently

$$p(x,y) = \frac{\mathrm{d}\Pi}{\mathrm{d}\xi}(x,y) = \exp\left(\frac{-c(x,y)+g(x)+f(y)}{\varepsilon}\right).$$

and

$$\begin{aligned}\mathscr{L}^\varepsilon(\Pi(f,g),f,g) &= \int_D f(y)\,\mathrm{d}\nu(y) + \int_D g(x)\,\mathrm{d}\mu(x) - \varepsilon\int_{D\times D}\exp\left(\frac{-c(x,y)+g(x)+f(y)}{\varepsilon}\right)\,\mathrm{d}\xi(x,y)\\ &= \int_D f(y)\,\mathrm{d}\nu(y) + \mathbf{m}\cdot\mathbf{g} - \varepsilon\int_D\sum_{j=1,\dots,n}\exp\left(\frac{-c(x_j,y)+g(x_j)+f(y)}{\varepsilon}\right)\,\mathrm{d}\mathcal{L}(y)\end{aligned}$$

Finally, we obtain a dual, entropy regularized, unconstrained formulation for the Wasserstein functional

$$\begin{aligned}\mathcal{W}_\varepsilon^2[\mu,\nu] &:= \sup_{(f\in C(D),\mathbf{g}\in\mathbb{R}^n)}\mathscr{L}^\varepsilon(\Pi(f,g),f,g)\\ &= \sup_{(f\in C(D),\mathbf{g}\in\mathbb{R}^n)}\left(\mathbf{m}\cdot\mathbf{g} + \int_D f(y)\,\mathrm{d}\nu(y) - \varepsilon\int_D\sum_{j=1,\dots,n}\exp\left(\frac{-|y-x_j|^2+f(y)+\mathrm{g}_j}{\varepsilon}\right)\,\mathrm{d}\mathcal{L}(y)\right)\end{aligned} \tag{5.4}$$

Here, the exponent $\exp(\varepsilon^{-1}(f(y)+\mathrm{g}_j - |y-x_j|^2))$ acts as a soft penalty in place of the original hard constraint $f(y)+\mathrm{g}_j \le |y-x_j|^2$. The optimal f for fixed g is characterized by

$$\begin{aligned}0 = \partial_f\mathscr{L}^\varepsilon(\Pi(f,g),f,g)(r) &= \int_D r(y)\,\mathrm{d}\nu(y) - \int_{D\times D}\exp\left(\frac{-c(x,y)+g(x)+f(y)}{\varepsilon}\right)r(y)\,\mathrm{d}\xi(x,y)\\ &= \int_D r(y)\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}}\right)(y)\,\mathrm{d}\mathcal{L}(y) - \int_{D\times D}\exp\left(\frac{-c(x,y)+g(x)+f(y)}{\varepsilon}\right)r(y)\,\mathrm{d}\mathcal{N}(x)\otimes\,\mathrm{d}\mathcal{L}(y)\end{aligned}$$

for all directions r. Hence, we obtain

$$0 = \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}}\right)(y) - \int_D \exp\left(\frac{-c(x,y)+g(x)+f(y)}{\varepsilon}\right)\mathrm{d}\mathcal{N}(x),$$

and finally

$$\mathbf{g}^{C,\varepsilon}(y) \coloneqq \varepsilon \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}}\right)(y) - \varepsilon \log\left(\sum_{j=1,\dots,n} \exp\left(\frac{\mathrm{g}_j - |y-x_j|^2}{\varepsilon}\right)\right) \tag{5.5}$$

defines the optimal f for given $\mathbf{g} \in \mathbb{R}^n$ and all $y \in D$. Given this C-transform, the regularized dual formulation in (5.4) can be rewritten as $\mathcal{W}_\varepsilon^2[\nu,\mu] = \max_{\mathbf{g}\in\mathbb{R}^n} \mathcal{D}^\varepsilon(\mathbf{g})$ with

$$\mathcal{D}^\varepsilon[\mathbf{g}] \coloneqq \mathscr{L}^\varepsilon(\Pi(\mathbf{g}^{C,\varepsilon}, g), \mathbf{g}^{C,\varepsilon}, g) \tag{5.6}$$

$$= \int_D \mathbf{g}^{C,\varepsilon}(y)\,\mathrm{d}\nu(y) + \mathbf{g}\cdot\mathbf{m} - \varepsilon\int_D \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}}\right)(y)\,\mathrm{d}\mathcal{L}(y) = \int_D \mathbf{g}^{C,\varepsilon}(y)\,\mathrm{d}\nu(y) + \mathbf{g}\cdot\mathbf{m} - \varepsilon \tag{5.7}$$

A sufficient condition for a vector $\mathbf{g} \in \mathbb{R}^n$ to maximize (5.6) for given sites $\mathbf{X}$ and masses $\mathbf{m}$ is

$$0 = \partial_{\mathrm{g}_i}\mathcal{D}^\varepsilon[\mathbf{g}] = \partial_{\mathrm{g}_j}\left(\int_D \mathbf{g}^{C,\varepsilon}(y)d\nu(y) + \mathbf{g}\cdot\mathbf{m}\right) = \int_D \partial_{\mathrm{g}_i}\mathbf{g}^{C,\varepsilon}(y)\,\mathrm{d}\nu(y) + m_i \tag{5.8}$$

$$\text{with } \partial_{\mathrm{g}_i}\mathbf{g}^{C,\varepsilon}(y) = -\frac{\exp\left(\frac{\mathrm{g}_i - |y-x_i|^2}{\varepsilon}\right)}{\sum_{j=1,\dots,n}\exp\left(\frac{\mathrm{g}_j - |y-x_j|^2}{\varepsilon}\right)} \tag{5.9}$$

for $i = 1,\dots,n$. The set of functions $\{\chi_i^\epsilon[\mathbf{X},\mathbf{g}]\}_{i=1,\dots,n}$ with $\chi_i^\epsilon[\mathbf{X},\mathbf{g}](y) = -\partial_{\mathrm{g}_i}\mathbf{g}^{C,\varepsilon}(y)$ for $y \in D$ forms a partition of unity on D. Furthermore,

$$\lim_{\varepsilon\to 0}\chi_i^\varepsilon[\mathbf{X},\mathbf{g}] = \chi_{L_i[\mathbf{X},\mathbf{g}]} \tag{5.10}$$

for $i = 1,\dots,n$, where the convergence is in $L^1(\nu)$. For given $[\mathbf{X},\mathbf{g}]$ with $\varepsilon > 0$, $\mathbf{X} = (x_1,\dots,x_n) \in D^n$, $x_j \neq x_i$ for $j,\ i = 1,\dots,n$

$$m_i^\varepsilon[\mathbf{X},\mathbf{g}] \coloneqq \int_D \chi_i^\varepsilon[\mathbf{X},\mathbf{g}](y)\,\mathrm{d}\nu(y), \quad b_i^\varepsilon[\mathbf{X},\mathbf{g}] \coloneqq \int_D \frac{y\,\chi_i^\varepsilon[\mathbf{X},\mathbf{g}](y)\,\mathrm{d}\nu(y)}{m_i^\varepsilon[\mathbf{X},\mathbf{g}]} \tag{5.11}$$

define regularized masses and regularized barycenters, respectively. By the L^1-convergence of $\chi_i^\varepsilon[\mathbf{X},\mathbf{g}]$ to χ_{L_i} and by the compactness of D, one gets

$$\lim_{\varepsilon\to 0} m_i^\varepsilon[\mathbf{X},\mathbf{g}] = m_i[\mathbf{X},\mathbf{g}], \quad \lim_{\varepsilon\to 0} b_i^\varepsilon[\mathbf{X},\mathbf{g}] = b_i[\mathbf{X},\mathbf{g}] \text{ for } m_i[\mathbf{X},\mathbf{g}] > 0. \tag{5.12}$$

Consequently, we obtain the entropy-regularized cost functional

$$\mathcal{F}_n^\varepsilon[\mathbf{X},\mathbf{g}] \coloneqq \sum_{i=1,\dots,n} m_i^\varepsilon[\mathbf{X},\mathbf{g}]\Phi(b_i^\varepsilon[\mathbf{X},\mathbf{g}]) \tag{5.13}$$

as an approximation of the original cost functional $\mathcal{F}_n$ on power diagrams defined in (3.2). Analogously, one finally obtains the entropy-regularized cost functional with penalty parameter $\eta > 0$:

$$\mathcal{F}_n^{\eta,\varepsilon}[\mathbf{X},\mathbf{g}] \coloneqq \mathcal{F}_n^\varepsilon[\mathbf{X},\mathbf{g}] - \eta\mathcal{R}_n^\varepsilon[\mathbf{X},\mathbf{g}], \tag{5.14}$$

where the regularized penalty term is defined as

$$\mathcal{R}_n^\varepsilon[\mathbf{X},\mathbf{g}] \coloneqq \sum_{i=1,\dots,n}\int_D |y-x_i|^2\chi_i^\varepsilon[\mathbf{X},\mathbf{g}](y)\,\mathrm{d}\nu(y) + \sum_{\substack{1\le i,j\le n\\ i\neq j}} \frac{m_i^\varepsilon[\mathbf{X},\mathbf{g}]m_j^\varepsilon[\mathbf{X},\mathbf{g}]}{|x_i-x_j|^2}. \tag{5.15}$$

in analogy to (3.6).

We are now in the position to prove the convergence of maximizers of the entropy regularized functional $\mathcal{F}_n^{\eta,\varepsilon}$ given in (5.14) to a maximizer of the original functional defined in (3.5) for $\varepsilon \to 0$.

Proposition 5.1. *Let $(\varepsilon_N)_N \subset \mathbb{R}^+$, $(\mathbf{X}^N, \mathbf{g}^N)_N \subset \mathcal{X}_n \times \mathbb{R}^n$ so that $\varepsilon_N \to 0$ and $(\mathbf{X}^N, \mathbf{g}^N) \to (\mathbf{X}, \mathbf{g}) \in \mathcal{X}_n \times \mathbb{R}^n$ for $N \to \infty$. Then, we have that*

$$\lim_{N\to\infty} \chi_i^{\varepsilon^N}[\mathbf{X}^N, \mathbf{g}^N] = \chi_{L_i[\mathbf{X},\mathbf{g}]},$$

where the limit is taken in $L^1(\nu)$.

Proof. Let $y \in D$ be in the interior of $L_i(\mathbf{X}, \mathbf{g})$. Then, there is $\delta > 0$, such that $|y - x_i|^2 - \mathrm{g}_i \leq |y - x_j|^2 - \mathrm{g}_j - 3\delta$, for all $j \neq i$. and by the continuity of the function $(\mathbf{X}, \mathbf{g}) \mapsto |y - x_i|^2 - \mathrm{g}_i$ we obtain that $|y - x_i^N|^2 - \mathrm{g}_i^N \leq |y - x_j^N|^2 - \mathrm{g}_j^N - \delta$ for sufficiently large N. This implies that

$$\begin{aligned}\chi_i^{\varepsilon}[\mathbf{X}^N, \mathbf{g}^N](y) =& \frac{\exp\left(\frac{\mathrm{g}_i^N - |y - x_i^N|^2}{\varepsilon^N}\right)}{\sum_{j=1}^n \exp\left(\frac{\mathrm{g}_j^N - |y - x_j^N|^2}{\varepsilon^N}\right)} \geq \frac{\exp\left(\frac{\mathrm{g}_i^N - |y - x_i^N|^2}{\varepsilon^N}\right)}{\exp\left(\frac{\mathrm{g}_i^N - |y - x_i^N|^2}{\varepsilon^N}\right) + (n-1)\exp\left(\frac{\mathrm{g}_i^N - |y - x_i^N|^2}{\varepsilon^N}\right)\exp(-\delta/\varepsilon^N)} \\ =& \frac{1}{1 + (n-1)\exp(-\delta/\varepsilon^N)} \to 1\end{aligned}$$

for $N \to \infty$. Now, let $y \notin \overline{L_i(\mathbf{X}, \mathbf{g})}$. Then, there is $j \neq i$, and $\delta > 0$, such that $|y - x_i|^2 - \mathrm{g}_i \geq |y - x_j|^2 - \mathrm{g}_j + 3\delta$. and $|y - x_i^N|^2 - \mathrm{g}_i^N \geq |y - x_j^N|^2 - \mathrm{g}_j^N + \delta$ for sufficiently large N. From this it follows that

$$\begin{aligned}\chi_i^{\varepsilon}[\mathbf{X}^N, \mathbf{g}^N](y) =& \frac{\exp\left(\frac{\mathrm{g}_i^N - |y - x_i^N|^2}{\varepsilon^N}\right)}{\sum_{j=1}^n \exp\left(\frac{\mathrm{g}_j^N - |y - x_j^N|^2}{\varepsilon^N}\right)} \leq \frac{\exp\left(\frac{\mathrm{g}_i^N - |y - x_i^N|^2}{\varepsilon^N}\right)}{\exp\left(\frac{\mathrm{g}_j^N - |y - x_j^N|^2}{\varepsilon^N}\right)} = \exp\left(\frac{\mathrm{g}_i^N - \mathrm{g}_j^N - |y - x_i^N|^2 + |y - x_j^N|^2}{\varepsilon^N}\right) \\ \leq& \exp\left(-\frac{\delta}{\varepsilon^N}\right) \to 0,\end{aligned}$$

as $N \to \infty$. By the absolute continuity of ν with respect to the Lebesgue measure, we have proven that $\chi_i^{\varepsilon^N}[\mathbf{X}^N, \mathbf{g}^N] \to \chi_{L_i[\mathbf{X},\mathbf{g}]}$, ν-almost everywhere, and by the dominated convergence theorem, we obtain the claim. □

In preparation for the convergence of maximizers, we now state the entropy-regularized "pendant" of Theorem 3.4.

Theorem 5.2. *Let us assume that $\varepsilon > 0$, and $\Phi \in C(D)$ with maximum $\bar{\Phi}$ on D. Then, for given number of sites $l \in \mathbb{N}$ there exists an $n \leq l$, such that a maximizer $(\mathbf{X}^*, \mathbf{g}^*)$ of $\mathcal{F}_n^{\eta,\varepsilon}$ exists with $m_i^{\varepsilon}[\mathbf{X}^*, \mathbf{g}^*] > 0$ for $i = 1, \ldots, n$ and $\mathcal{F}_n^{\eta,\varepsilon}(\mathbf{X}^*, \mathbf{g}^*) \geq \sup_{(\mathbf{X},\mathbf{g})} \mathcal{F}_l^{\eta,\varepsilon}(\mathbf{X}, \mathbf{g})$.*

Proof. The proof is analogous to Theorem 3.4. □

We are now in the position to prove the convergence of maximizers of the entropy regularized functional $\mathcal{F}_n^{\eta,\varepsilon}$ given in (5.14) to a maximizer of the original functional $\mathcal{F}_n^{\eta}$ defined in (3.5) for $\epsilon \to 0$.

Theorem 5.3. *Let $\Phi \in C(D)$, $l \in \mathbb{N}$ and $\eta > 0$. For each $N \in \mathbb{N}$, consider maximizers $(\mathbf{X}^N, \mathbf{g}^N) \in \mathcal{X}_{\beta(N)} \times \mathbb{R}^{\beta(N)}$, where $\beta : \mathbb{N} \to \{1, \ldots, n\}$, so that $\mathcal{F}_{\beta(N)}^{\eta,\varepsilon_N}(\mathbf{X}^N, \mathbf{g}^N) \geq \max_{k=1,\ldots,n}\{\sup_{(\mathbf{X},\mathbf{g})\in\mathcal{X}_k\times\mathbb{R}^k} \mathcal{F}_k^{\eta,\varepsilon_N}(\mathbf{X}, \mathbf{g})\}$. Then, there is $n \in \{1, \ldots, l\}$ such that up to the selection of a subsequence $(\mathbf{X}^N, \mathbf{g}^N)_{N\in\mathbb{N}} \subset \mathcal{X}_n \times \mathbb{R}^n$ converges to a limit $(\mathbf{X}^*, \mathbf{g}^*) \in \mathcal{X}_n \times \mathbb{R}^n$ for which it holds*

$$\mathcal{F}_n^{\eta}(\mathbf{X}^*, \mathbf{g}^*) \geq \max_{k=1,\ldots,l} \left\{ \sup_{(\mathbf{X},\mathbf{g})\in\mathcal{X}_k\times\mathbb{R}^k} \mathcal{F}_k^{\eta}(\mathbf{X}, \mathbf{g}) \right\}.$$

Proof. The proof of the first statement follows the proof of Theorem 3.4 very closely, albeit with a few key differences, which will be highlighted now: Let $n' \in \{1, \ldots, l\}$, such that $|\{N : \beta(N) = n'\}| = \infty$. Choose a subsequence such that up to relabeling of indices $(\mathbf{X}^N, \mathbf{g}^N)_{N\in\mathbb{N}} \subset \mathcal{X}_{n'} \times \mathbb{R}^{n'}$, such that the sequence of masses $(m_1^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N], \ldots, m_{n'}^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N])_{N\in\mathbb{N}}$ converges to some $(m_1^*, \ldots, m_{n'}^*) \in [0, 1]^{n'}$. Moreover, assume without loss of generality that there is $n \in \{1, \ldots, n'\}$ with $m_i^* > 0$ if and only if $i \leq n$, and $m_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N] \geq \frac{1}{2}m_i^*$ for all $N \in \mathbb{N}$, $i \leq n$. It then holds for all $k \in \{1, \ldots, l\}$ and all $(\overline{\mathbf{X}}, \overline{\mathbf{g}}) \in \mathcal{X}_k \times \mathbb{R}^k$ with $m_i[\overline{\mathbf{X}}, \overline{\mathbf{g}}] > 0$ for $i = 1, \ldots, k$:

$$\mathcal{F}_k^{\eta}[\overline{\mathbf{X}}, \overline{\mathbf{g}}] = \lim_{N\to\infty} \mathcal{F}_k^{\eta,\varepsilon_N}[\overline{\mathbf{X}}, \overline{\mathbf{g}}] \leq \limsup_{N\to\infty} \mathcal{F}_{n'}^{\eta,\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N] \leq \limsup_{N\to\infty} \mathcal{F}_{n'}^{\varepsilon_N}(\mathbf{X}^N, \mathbf{g}^N) - \eta \liminf_{N\to\infty} \mathcal{R}_{n'}^{\varepsilon_N}(\mathbf{X}^N, \mathbf{g}^N)$$

$$
\begin{aligned}
&\leq \limsup_{N\to\infty} \mathcal{F}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) + \limsup_{N\to\infty} \sum_{i=n+1,\dots,n'} m_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N]\Phi(b_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N]) - \eta \liminf_{N\to\infty} \mathcal{R}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N)\\
&\quad - \eta \liminf_{N\to\infty} \left(\mathcal{R}_{n'}^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) - \mathcal{R}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N)\right)\\
&= \limsup_{N\to\infty} \mathcal{F}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) + 0 - \eta \liminf_{N\to\infty} \mathcal{R}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) - \eta \liminf_{N\to\infty} \left(\mathcal{R}_{n'}^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) - \mathcal{R}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N)\right)\\
&\leq \limsup_{N\to\infty} \mathcal{F}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) - \eta \liminf_{N\to\infty} \mathcal{R}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) \leq \bar{\Phi} - \eta \liminf_{N\to\infty} \mathcal{R}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N).
\end{aligned}
$$

We obtain $\liminf_{N\to\infty} \mathcal{R}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) \leq \frac{\bar{\Phi}-\mathcal{F}_k^\eta[\overline{\mathbf{X}},\overline{\mathbf{g}}]}{\eta}$. Without loss of generality, we can assume that for all $\delta > 0$, $\mathcal{R}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) < \frac{\bar{\Phi}-\mathcal{F}_k^\eta[\overline{\mathbf{X}},\overline{\mathbf{g}}]+\delta}{\eta}$ holds for all $N \in \mathbb{N}$, $i = 1,\dots,l$. Then, we obtain the bounds

$$
\operatorname{dist}(x_i^N, D) < \sqrt{\frac{2(\bar{\Phi} - \mathcal{F}_k^\eta[\bar{\mathbf{X}},\bar{\mathbf{g}}] + \delta)}{\eta m_i^*}}, \quad |x_i^N - x_j^N| > \sqrt{\frac{\eta m_i^* m_j^*}{4(\bar{\Phi} - \mathcal{F}_k^\eta[\bar{\mathbf{X}},\bar{\mathbf{g}}] + \delta)}},
$$

for all pairwise different $i,j = 1,\dots,n$. Using the same arguments as in the proof of Theorem 3.4, one also finds analogous uniform bounds on $|\mathrm{g}^N|$. Hence, up to a subsequence, $(\mathbf{X}^N,\mathbf{g}^N)_{N\in\mathbb{N}}$ converges to some $(\mathbf{X}^*,\mathbf{g}^*) \in \mathcal{X}_n \times \mathbb{R}^n$. By the above computations, we finally obtain

$$
\mathcal{F}_k^\eta[\overline{\mathbf{X}},\overline{\mathbf{g}}] \leq \lim_{N\to\infty} \mathcal{F}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) - \eta\mathcal{R}_n^{\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) = \lim_{N\to\infty} \mathcal{F}_n^{\eta,\varepsilon_N}(\mathbf{X}^N,\mathbf{g}^N) = \mathcal{F}_n^\eta(\mathbf{X}^*,\mathbf{g}^*).
$$

For the last equality, we used the fact that the masses $m_i^{\varepsilon_N}$ do not vanish in the limit and Proposition 5.1 for the convergence of masses $m_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N] \to m_i[\mathbf{X}^*,\mathbf{g}^*]$. For the convergence of barycenters $b_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N] \to b_i[\mathbf{X}^*,\mathbf{g}^*]$, we additionally use the fact that the function $y \mapsto y$ is bounded in D. Finally, for the convergence of

$$
\int_D |y - x_i^N|^2 \chi_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N](y)\,\mathrm{d}\nu(y) \to \int_D |y - x_i^*|^2 \chi_i[\mathbf{X}^*,\mathbf{g}^*](y)\,\mathrm{d}\nu(y),
$$

we also use the fact that the family of functions $(y \mapsto |y - x_i^N|^2)_{N\in\mathbb{N}}$ is uniformly bounded in $L^\infty(\nu)$ and converges in $L^\infty(\nu)$ to $y \mapsto |y - x_i^*|^2$, and the uniform boundedness principle. □

To numerically implement the maximization of 5.13 or 5.14 via a gradient ascent approach we need to compute the gradient of $\mathcal{F}_n^\varepsilon$. Using $z = x_k$ or $z = \mathrm{g}_k$ for $k = 1,\dots,n$ we obtain

$$
\begin{aligned}
\partial_z \mathcal{F}_n^\varepsilon[\mathbf{X},\mathbf{g}] &= \sum_{j=1,\dots,n} \partial_z m_j^\varepsilon[\mathbf{X},\mathbf{g}]\Phi(b_j^\varepsilon[\mathbf{X},\mathbf{g}]) + m_j^\varepsilon[\mathbf{X},\mathbf{g}]\partial_z\Phi(b_j^\varepsilon[\mathbf{X},\mathbf{g}]),\\
\partial_{x_k} \mathcal{R}_n^\varepsilon[\mathbf{X},\mathbf{g}] &= \int_D 2(x_k - y)\chi_k^\varepsilon[\mathbf{X},\mathbf{g}](y)\,\mathrm{d}\nu(y) + \sum_{i=1,\dots,n} \int_D |y - x_i|^2 \partial_{x_k}\chi_i^\varepsilon[\mathbf{X},\mathbf{g}](y)\,\mathrm{d}\nu(y)\\
&\quad + \sum_{\substack{1\leq i,j\leq n\\ i\neq j}} \Bigg(\frac{\partial_{x_k} m_i^\varepsilon[\mathbf{X},\mathbf{g}] m_j^\varepsilon[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2} + \frac{m_i^\varepsilon[\mathbf{X},\mathbf{g}]\partial_{x_k} m_j^\varepsilon[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2}\\
&\quad - 2\frac{m_i^\varepsilon[\mathbf{X},\mathbf{g}] m_j^\varepsilon[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^4}((x_i - x_j)\delta_{ik} + (x_j - x_i)\delta_{jk})\Bigg),\\
\partial_{\mathrm{g}_k} \mathcal{R}_n^\varepsilon[\mathbf{X},\mathbf{g}] &= \sum_{i=1,\dots,n} \int_D |y - x_i|^2 \partial_{\mathrm{g}_k}\chi_i^\varepsilon[\mathbf{X},\mathbf{g}](y)\,\mathrm{d}\nu(y)\\
&\quad + \sum_{\substack{1\leq i,j\leq n\\ i\neq j}} \left(\frac{\partial_{\mathrm{g}_k} m_i^\varepsilon[\mathbf{X},\mathbf{g}] m_j^\varepsilon[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2} + \frac{m_i^\varepsilon[\mathbf{X},\mathbf{g}]\partial_{\mathrm{g}_k} m_j^\varepsilon[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2} \right),\\
\partial_z\Phi(b_j^\varepsilon[\mathbf{X},\mathbf{g}]) &= \nabla\Phi(b_j^\varepsilon[\mathbf{X},\mathbf{g}]) \cdot \partial_z b_j^\varepsilon[\mathbf{X},\mathbf{g}],\\
\partial_z m_j^\varepsilon[\mathbf{X},\mathbf{g}] &= \int_D \partial_z\chi_j^\varepsilon[\mathbf{X},\mathbf{g}](y)\,\mathrm{d}\nu(y),\\
\partial_z b_j^\varepsilon[\mathbf{X},\mathbf{g}] &= \frac{1}{m_j^\varepsilon[\mathbf{X},\mathbf{g}]} \int_D y\partial_z\chi_j^\varepsilon[\mathbf{X},\mathbf{g}](y)\,\mathrm{d}\nu(y) - \frac{b_j^\varepsilon[\mathbf{X},\mathbf{g}]\partial_z m_j^\varepsilon[\mathbf{X},\mathbf{g}]}{m_j^\varepsilon[\mathbf{X},\mathbf{g}]^2},\\
\partial_{x_k}\chi_j^\varepsilon[\mathbf{X},\mathbf{g}](y) &= \tfrac{2x_j}{\varepsilon}\left(\chi_j^\varepsilon[\mathbf{X},\mathbf{g}](y) - \delta_{kj}\right)\chi_k^\varepsilon[\mathbf{X},\mathbf{g}](y),\\
\partial_{\mathrm{g}_k}\chi_j^\varepsilon[\mathbf{X},\mathbf{g}](y) &= -\tfrac{1}{\varepsilon}\left(\chi_j^\varepsilon[\mathbf{X},\mathbf{g}](y) - \delta_{kj}\right)\chi_k^\varepsilon[\mathbf{X},\mathbf{g}](y).
\end{aligned}
$$

Hence, since the gradient of $\mathcal{F}^\varepsilon$ has the form of an integral with respect to ν, a stochastic gradient ascent is a very suitable method to numerically obtain an optimizer of this problem: since evaluating the integral of gradients becomes very expensive, one can save computational costs at every iteration by sampling the measure ν, and by approximating the integral by the sum of the integrand evaluated at the samples' values, cf. [6]. This method has the advantage of not having to discretize the measure ν in space, which might lead to further inaccuracies. Alternatively, to implement the maximization of 5.13 numerically, we have to further discretize the continuous given measure ν in space.

6 Spatial discretization

As the next step, we have to discretize our optimization problem in space. To this end, we restrict to the domain $D = [0,1]^d$ and consider a dyadic mesh with grid size $h = 2^{-N}$ for $N \in \mathbb{N}$. The domain is subdivided into cells $D_h^\alpha = \times_{1,\dots,d}[\alpha_j h, (\alpha_j+1)h]$ where α is a multi-index in $(\mathcal{I}^h)^d$ for the index set $\mathcal{I}^h = \{0,\dots,2^N-1\}$. We consider a discrete measure $\nu^h = \sum_{\alpha\in\mathcal{I}^d} \nu_\alpha \delta_{y_h^\alpha}$ with $\nu_\alpha = \nu(D_h^\alpha)$, where $y_h^\alpha = \big((\alpha_j+\frac{1}{2})h\big)_{j=1,\dots,d}$ are the cell centers. This discretization ansatz gives rise to discrete counterparts of the continuous, regularized characteristic functions $\chi_i^\varepsilon[\mathbf{X},\mathbf{g}]$, masses $m_i^\varepsilon[\mathbf{X},\mathbf{g}]$, and barycenters $b_i^\varepsilon[\mathbf{X},\mathbf{g}]$:

$$\chi_{i,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}] := \chi_i^\varepsilon[\mathbf{X},\mathbf{g}](y_h^\alpha) = \frac{\exp\left(\frac{\mathrm{g}_i - |y_h^\alpha - x_i|^2}{\varepsilon}\right)}{\sum_{j=1,\dots,n}\exp\left(\frac{\mathrm{g}_j - |y_h^\alpha - x_j|^2}{\varepsilon}\right)}, \tag{6.1}$$

$$m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}] := \sum_{\alpha\in\mathcal{I}^d} \chi_{i,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\nu^\alpha, \quad b_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}] := \sum_{\alpha\in\mathcal{I}^d} \frac{y_h^\alpha\, \chi_{i,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}{m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}\nu^\alpha. \tag{6.2}$$

Based on this discretization, we obtain a discrete functional

$$\mathcal{F}_n^{\varepsilon,h}[\mathbf{X},\mathbf{g}] := \sum_{i=1,\dots,n} m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\Phi(b_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]), \tag{6.3}$$

and the usual penalty-enhanced discrete functional

$$\mathcal{F}_n^{\eta,\varepsilon,h}[\mathbf{X},\mathbf{g}] := \sum_{i=1,\dots,n} m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\Phi(b_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]) - \eta\mathcal{R}_n^{\varepsilon,h}[\mathbf{X},\mathbf{g}], \tag{6.4}$$

where the discrete penalty term is defined by

$$\mathcal{R}_n^{\varepsilon,h}[\mathbf{X},\mathbf{g}] := \sum_{i=1,\dots,n}\sum_{\alpha\in\mathcal{I}^d} |y_h^\alpha - x_i|^2 \chi_{i,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\nu^\alpha + \sum_{\substack{1\le i,j\le n\\ i\neq j}} \frac{m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\, m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2}. \tag{6.5}$$

Now, in analogy to Proposition 5.1 we demonstrate the consistency of entropy regularized and discrete characteristic functions and the continuous counterpart. To this end, let us define the ν-almost surely piecewise continuous function $\chi_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}] \in L^1(\nu)$ as follows:

$$\chi_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}](y) = \chi_{i,\alpha}^{\varepsilon,h} \text{ for } \nu-\text{almost all } y \text{ in a cell interior } \overset{\circ}{D}{}_h^\alpha .$$

Proposition 6.1. *For $(\varepsilon_N)_N \subset \mathbb{R}^+$, $(\mathbf{X}^N,\mathbf{g}^N)_N \subset \mathcal{X}_n \times \mathbb{R}^n$ with $\varepsilon_N \to 0$, $(\mathbf{X}^N,\mathbf{g}^N) \to (\mathbf{X},\mathbf{g}) \in \mathcal{X}_n \times \mathbb{R}^n$ for $N\to\infty$ and for grid sizes $h_N = 2^{-N}$ we obtain that $\lim_{N\to\infty}\chi_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N] = \chi_{L_i[\mathbf{X},\mathbf{g}]}$, in $L^1(\nu)$.*

Proof. By the triangle inequality, we have that

$$\left\|\chi_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N] - \chi_{L_i[\mathbf{X},\mathbf{g}]}\right\|_{L^1(\nu)} \le \left\|\chi_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N] - \chi_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N]\right\|_{L^1(\nu)} + \left\|\chi_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N] - \chi_{L_i[\mathbf{X},\mathbf{g}]}\right\|_{L^1(\nu)}.$$

In Proposition 5.1, we already showed that the second term converges to zero. To verify that the first term vanishes as well in the limit for $N\to\infty$, we estimate the norm of the gradient of $\chi_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N]$ given by

$$\nabla_y\chi_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N](y) = \chi_i^{\varepsilon_N}[\mathbf{X}^N,\mathbf{g}^N](y)\left(-2\frac{y-x_i^N}{\varepsilon_N} + \frac{\sum_{j=1}^n 2(y-x_j^N)\exp\left(\frac{\mathrm{g}_j^N - |y-x_j^N|^2}{\varepsilon_N}\right)}{\varepsilon_N\sum_{j=1}^n\exp\left(\frac{\mathrm{g}_j^N - |y-x_j^N|^2}{\varepsilon_N}\right)}\right).$$

Let $y \in D$ be in the interior of $L_i(\mathbf{X}, \mathbf{g})$. Then, as in the proof of Proposition 5.1, there is a $\delta > 0$, such that

$$|y - x_i^N|^2 - \mathrm{g}_i^N \leq |y - x_j^N|^2 - \mathrm{g}_j^N - \delta$$

for for $j \neq i$ and for sufficiently large N. Then, one obtains

$$\begin{aligned}
|\nabla_y \chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y)| &= \frac{2\chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y)}{\varepsilon_N} \left| (y - x_i^N)(\chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y) - 1) + \frac{\sum_{j \neq i}(y - x_j^N) \exp\left(\frac{\mathrm{g}_j^N - |y - x_j^N|^2}{\varepsilon_N}\right)}{\sum_{j=1}^n \exp\left(\frac{\mathrm{g}_j^N - |y - x_j^N|^2}{\varepsilon_N}\right)} \right| \\
&\leq \frac{2}{\varepsilon_N} \left(|y - x_i^N|(1 - \chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y)) + \sum_{j \neq i} |y - x_j^N| \exp\left(\frac{-\delta}{\varepsilon_N}\right) \right) \\
&\leq \frac{2}{\varepsilon_N} \left(|y - x_i^N| \left(\frac{(n-1)\exp(-\delta/\varepsilon_N)}{1 + (n-1)\exp(-\delta/\varepsilon_N)} \right) + \sum_{j \neq i} |y - x_j^N| \exp\left(\frac{-\delta}{\varepsilon_N}\right) \right) \\
&\leq \frac{2}{\varepsilon_N} \exp(-\delta/\varepsilon_N) \left(|y - x_i^N|(n-1) + \sum_{j \neq i} |y - x_j^N| \right) \longrightarrow 0,
\end{aligned}$$

for $N \to \infty$. Next, let $y \notin \overline{L_i(\mathbf{X}, \mathbf{g})}$. Then, it holds

$$\begin{aligned}
|\nabla_y \chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y)| &= \frac{2\chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y)}{\varepsilon_N} \left| -(y - x_i^N) + \frac{\sum_{j \neq i}(y - x_j^N) \exp\left(\frac{\mathrm{g}_j^N - |y - x_j^N|^2}{\varepsilon_N}\right)}{\sum_{j=1}^n \exp\left(\frac{\mathrm{g}_j^N - |y - x_j^N|^2}{\varepsilon_N}\right)} \right| \\
&= \frac{2\chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y)}{\varepsilon_N} \left| -(y - x_i^N) + \tilde{x} \right| \leq \frac{2}{\varepsilon_N} \exp(-\delta/\varepsilon_N) \left| x_i^N - y + \tilde{x} \right| \to 0,
\end{aligned}$$

where $\tilde{x} = \sum_{j \neq i}(y - x_j^N)\chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y)$ is an element in the convex hull of $(y - x_j^N)_{j=1,\ldots,n}$. and thus due to the convergence of $(\mathbf{X}^N)_N$ uniformly bounded. This implies that $|x_i^N - y + \tilde{x}|$ is uniformly bounded.

For every $y \in D$ in the interior of $L_i(\mathbf{X}, \mathbf{g})$ and sufficiently large N we observe that y is in the cell $D_{h_N}^{\alpha^N}$ for some multi-indices α^N and $D_{h_{N_0}}^{\alpha^{N_0}}$ is completely contained in the interior of $L_i(\mathbf{X}, \mathbf{g})$. There is a $\delta > 0$ such that $|y - x_i^N|^2 - \mathrm{g}_i^N \leq |y - x_j^N|^2 - g_j^N - \delta$ holds for $j \neq i$ and all $y \in D_{h_N}^{\alpha^N}$. Hence, the restriction of $\chi_i^{\varepsilon_M}[\mathbf{X}^M, \mathbf{g}^M]$ onto this grid cell for $M \geq N$ is a family of uniformly Lipschitz functions, and in particular equicontinuous. It follows that

$$\begin{aligned}
\left| \chi_i^{\varepsilon_N, h_N}[\mathbf{X}^N, \mathbf{g}^N](y) - \chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y) \right| &= \left| \chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y_{h_N}^{\alpha^N}) - \chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N](y) \right| \\
&\leq \left(\sup_{N \geq N_0} \mathrm{Lip}(\chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N]) \right) |y - y_{h_N}^{\alpha^N}| \to 0
\end{aligned}$$

The analogous statement holds for $y \notin \overline{L_i(\mathbf{X}, \mathbf{g})}$ as well. Thus, $\chi_i^{\varepsilon_N, h_N}[\mathbf{X}^N, \mathbf{g}^N]$ converges to $\chi_i^{\varepsilon_N}[\mathbf{X}^N, \mathbf{g}^N]$ ν-almost everywhere and, by the dominated convergence theorem, one obtains the desired convergence in $L^1(\nu)$. □

We obtain the entropy-regularized, fully discrete analogue of Theorem 3.4.

Theorem 6.2. *Let us assume that $\varepsilon, h > 0$, and $\Phi \in C(D)$ with maximum $\bar{\Phi}$ on D. Then, for given number of sites $l \in \mathbb{N}$ there exists an $n \leq l$, such that a maximizer $(\mathbf{X}^*, \mathbf{g}^*)$ of $\mathcal{F}_n^{\eta,\varepsilon,h}$ exists with $m_i^{\varepsilon,h}[\mathbf{X}^*, \mathbf{g}^*] > 0$ for $i = 1, \ldots, n$ and $\mathcal{F}_n^{\eta,\varepsilon,h}(\mathbf{X}^*, \mathbf{g}^*) \geq \sup_{(\mathbf{X},\mathbf{g})} \mathcal{F}_l^{\eta,\varepsilon,h}(\mathbf{X}, \mathbf{g})$.*

Proof. The proof is completely analogous to the proof of Theorem 3.4. □

Next, we investigate the convergence of maximizers of the fully discrete, entropy regularized functional $\mathcal{F}_n^{\eta,\varepsilon,h}$ given in (6.4) to a maximizer of the original functional $\mathcal{F}_n^{\eta}$ defined in (3.5) for both $\varepsilon \to 0$ and $h \to 0$.

Theorem 6.3. *Let $\Phi \in C(D)$, $l \in \mathbb{N}$ and $\eta > 0$. For each $N \in \mathbb{N}$, consider maximizers $(\mathbf{X}^N, \mathbf{g}^N) \in \mathcal{X}_{\beta(N)} \times \mathbb{R}^{\beta(N)}$, where $\beta : \mathbb{N} \to \{1, \ldots, n\}$, so that $\mathcal{F}_{\beta(N)}^{\eta,\varepsilon_N,h_N}(\mathbf{X}^N, \mathbf{g}^N) \geq \max_{k=1,\ldots,n}\{\sup_{(\mathbf{X},\mathbf{g}) \in \mathcal{X}_k \times \mathbb{R}^k} \mathcal{F}_k^{\eta,\varepsilon_N,h_N}(\mathbf{X}, \mathbf{g})\}$.*

Then, there is $n \in \{1,\dots,l\}$ such that up to the selection of a subsequence $(\mathbf{X}^N,\mathbf{g}^N)_{N\in\mathbb{N}} \subset \mathcal{X}_n \times \mathbb{R}^n$ converges to a limit $(\mathbf{X}^*,\mathbf{g}^*) \in \mathcal{X}_n \times \mathbb{R}^n$. Furthermore, it holds

$$\mathcal{F}_n^\eta(\mathbf{X}^*,\mathbf{g}^*) \geq \max_{k=1,\dots,l} \left\{ \sup_{(\mathbf{X},\mathbf{g})\in\mathcal{X}_k\times\mathbb{R}^k} \mathcal{F}_k^\eta(\mathbf{X},\mathbf{g}) \right\}.$$

Proof. The proof of the first statement is along the same lines as the proof of Theorem 3.4. For the last statement, notice that

$$\begin{aligned}
b_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N] &= \sum_{\alpha\in\mathcal{I}^d} \frac{y_h^\alpha\, \chi_{i,\alpha}^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N]}{m_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N]}\nu^\alpha = \int_D \frac{g^N(y)\,\chi_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N]}{m_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N](y)}\,\mathrm{d}\nu(y)\\
\mathcal{R}_n^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N] &= \sum_{i=1,\dots,n}\sum_{\alpha\in\mathcal{I}^d} |y_h^\alpha - x_i^N|^2 \chi_{i,\alpha}^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N]\nu^\alpha + \sum_{\substack{1\leq i,j\leq n\\ i\neq j}} \frac{m_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N]\; m_j^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N]}{|x_i^N - x_j^N|^2}\\
&= \sum_{i=1,\dots,n}\int_D f^N(y)\chi_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N](y)\,\mathrm{d}\nu(y) + \sum_{\substack{1\leq i,j\leq n\\ i\neq j}} \frac{m_i^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N]\; m_j^{\varepsilon_N,h_N}[\mathbf{X}^N,\mathbf{g}^N]}{|x_i^N - x_j^N|^2},
\end{aligned}$$

where $f^N(y) := y_h^\alpha$ for $y \in \mathring{D}_h^\alpha$, and $g^N(y) := |y_h^\alpha - x_i^N|^2$ for $y \in \mathring{D}_h^\alpha$ are the piecewise constant approximations of the functions $y \mapsto y$ and $y \mapsto |y - x_i^N|^2$ consistent with our grid cells. The rest of the proof is then completely analogous to the proof of the respective statement in Theorem 5.3. □

To numerically implement the maximization of 5.13 via a gradient ascent approach we need to compute the gradient of $\mathcal{F}_n^\varepsilon$. We specifically obtain the following derivatives for spatially discrete quantities using $z = x_k$ or $z = \mathrm{g}_k$ for $k = 1,\dots,n$

$$\begin{aligned}
\partial_z \mathcal{F}_n^{\varepsilon,h}[\mathbf{X},\mathbf{g}] &= \sum_{j=1,\dots,n} \partial_z m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\Phi(b_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]) + m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\partial_z\Phi(b_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]),\\
\partial_{x_k}\mathcal{R}_n^{\varepsilon,h}[\mathbf{X},\mathbf{g}] &= 2\sum_{\alpha\in\mathcal{I}^d}(x_k - y_h^\alpha)\chi_{k,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\nu^\alpha + \sum_{i=1,\dots,n}\sum_{\alpha\in\mathcal{I}^d}|y_h^\alpha - x_i|^2\partial_{x_k}\chi_{i,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\nu^\alpha\\
&\quad + \sum_{\substack{1\leq i,j\leq n\\ i\neq j}}\left(\frac{\partial_{x_k} m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2} + \frac{m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\partial_{x_k} m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2}\right.\\
&\quad \left. - 2\frac{m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^4}((x_i - x_j)\delta_{ik} + (x_j - x_i)\delta_{jk})\right),\\
\partial_{\mathrm{g}_k}\mathcal{R}_n^{\varepsilon,h}[\mathbf{X},\mathbf{g}] &= \sum_{i=1,\dots,n}\sum_{\alpha\in\mathcal{I}^d}|y_h^\alpha - x_i|^2\partial_{\mathrm{g}_k}\chi_{i,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\nu^\alpha\\
&\quad + \sum_{\substack{1\leq i,j\leq n\\ i\neq j}}\left(\frac{\partial_{\mathrm{g}_k} m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2} + \frac{m_i^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\partial_{\mathrm{g}_k} m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}{|x_i - x_j|^2}\right),\\
\partial_z\Phi(b_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]) &= \nabla\Phi(b_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}])\cdot\partial_z b_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}],\\
\partial_z m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}] &= \sum_{\alpha\in\mathcal{I}^d}\partial_z\chi_{j,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\nu^\alpha,\\
\partial_z b_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}] &= \frac{1}{m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}\sum_{\alpha\in\mathcal{I}^d} y_h^\alpha\partial_z\chi_{j,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\nu^\alpha - \frac{b_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]\partial_z m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]}{m_j^{\varepsilon,h}[\mathbf{X},\mathbf{g}]^2},\\
\partial_{x_k}\chi_{j,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}] &= \tfrac{2x_j}{\varepsilon}\left(\chi_{j,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}] - \delta_{kj}\right)\chi_{j,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}],\\
\partial_{\mathrm{g}_k}\chi_{j,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}] &= -\tfrac{1}{\varepsilon}\left(\chi_{j,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}] - \delta_{kj}\right)\chi_{j,\alpha}^{\varepsilon,h}[\mathbf{X},\mathbf{g}].
\end{aligned}$$

7 Numerical experiments

Now, we will apply the above derived method to compute the optimal partition of the unit cube $[0,1]^2$ for given n, a function Φ and a probability density ν. In the presented numerical results, we assume ν to be the Lebesgue

measure on the unit cube. Moreover, for every numerical experiment in this and the next section, we made use of the sparse multi-scale algorithm by Schmitzer [23] and the implementation in [8] to efficiently compute the Sinkhorn iterations. To optimize the parameters of the power diagrams, we used the *Adam* optimizer (cf. [13]) as a stochastic gradient ascent method. Unless otherwise explicitly stated, we use $\eta = 0$, $N = 256$ (i.e. grid size $h = 1/256$), and initialize our algorithm with $n = 12$ sites/weights for each of the experiments in this and the next section. On each plot in this section, the barycenters (triangles) and sites (circles) are plotted, for visualisation purposes the latter ones only if inside the unit cube. On the left of each plot, the respective values of the component of the g vector is also plotted.

In Figure 1 we show the dependence of the optimal numerical solution on the entropy parameter ε. The interfaces between the Laguerre cells become fuzzier with increasing ε, whereas the structure of the optimal solutions does not change much.

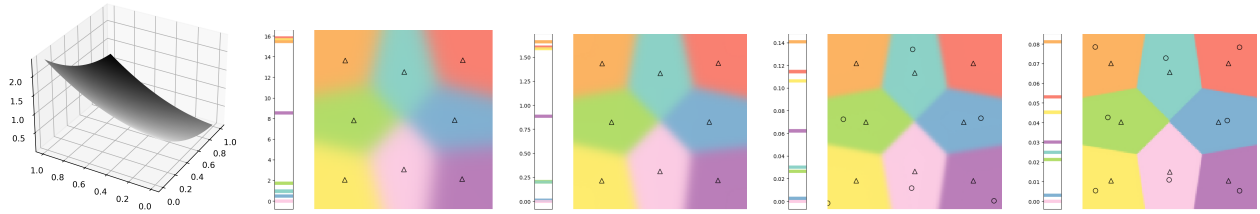

Figure 1: Convergence of the optimal power diagram for different entropy parameters $\varepsilon = 25, 5, 1, 0.2$. (four right-most plots), where the blur parameter values are given in units of $1/N$, for $N = 128$. The function Φ is plotted on the left-most panel.

Next, we plot in Figure 2 the optimal configurations for different values of η for a function Φ (left), which has global/local maxima of equal value at the points $(\frac{1}{2}, \frac{1}{4}), (\frac{3}{4}, \frac{3}{4})$, and $(\frac{1}{4}, \frac{3}{4})$. We know that the solution for $\eta = 0$ is not contained in the space of power diagrams, but we observe for $\eta \to 0$ the convergence of the solutions. Indeed, for large values of η a single cell has positive mass and no other cells contribute to the cost. For vanishing η, two additional cells appear as optimal configurations and the three cells meet in a triple-point converging to $(\frac{1}{2}, \frac{1}{2})$ with one angle converging to π.

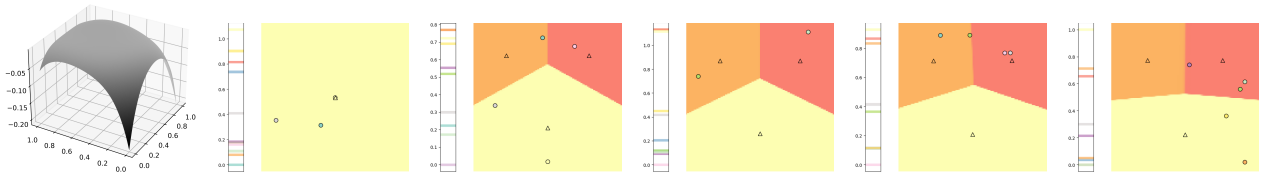

Figure 2: Convergence of the optimal power diagram for a function Φ (column (i)) with global maxima at $(0.25, 0.75), (0.75, 0.75)$ and $(0.5, 0.25)$ for regularization parameter $\eta = 1\text{e}{-1}, 1\text{e}{-2}, 1\text{e}{-3}, 1\text{e}{-4}, 1\text{e}{-5}$ (columns (ii)-(vi)).

Finally, we check how our algorithm deals with fusing/pushing cells away, when the optimal solution requires a smaller number of sites/weights than what the algorithm was initialized with. To this end, in Figure 3 we consider a concave function Φ, that has the trivial partition as solution, and plot some iterations of our algorithm that show the proper recovery of this solution, i.e. cells disappear by either pushing the respective sites away (orange, green cells), or by increasing the difference between the g-values of the purple cell and those of the respective cells (beige, blue).

8 Application to information design

Finally, we consider the application of our algorithm to a particular problem in information design: the monopolist's problem. In this problem, a seller (sender) can disclose information about the qualities of the products she sells. There are two objects available individually at fixed prices p_1, p_2, and the bundle of both objects can be bought at fixed price $p_3 = p_1 + p_2 + \Delta$ for some bundling "surcharge" $\Delta > 0$ or "discount" $\Delta < 0$. A consumer (receiver) has valuations $\mathbf{v} = (v_1, v_2)$ per unit of quality, distributed according to the Lebesgue measure on the unit square $[0, 1]^2$. The quality of the objects $\mathbf{q} = (q_1, q_2)$ is distributed on $[\underline{q}, \overline{q}]^2$ according to

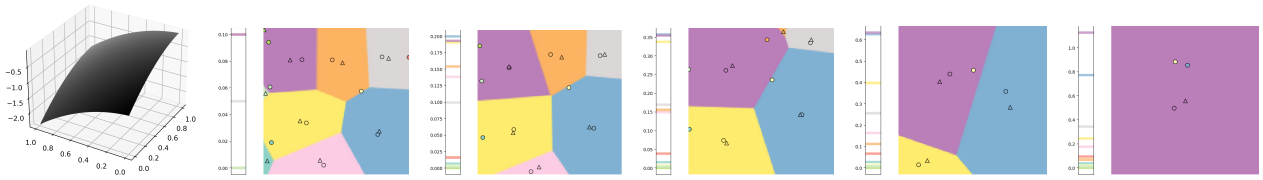

Figure 3: Convergence of the optimal power diagram for a concave function Φ (left). One sees how the algorithm automatically pushes cells outside of the relevant unit square to enact the trivial solution. Plotted are the computed power diagrams after iterations $it = 1, 2, 4, 8, 16$ (five right-most plots).

a measure μ that is absolutely continuous with respect to the Lebesgue measure. Throughout this section, we assume that μ is the (re-scaled) Lebesgue measure; the general case works analogously. The realized qualities are known to the seller but not to the buyer, and the realized valuations are known to the buyer but not to the seller.

We first consider a buyer with unit demand and then consider a buyer who demands more than one object and has additive valuations. A buyer has *unit demand* if she values at most one object. If the seller provides a signal about the qualities such that the buyer believes the expected qualities of the objects are (q_1, q_2), then a buyer with unit demand buys only good $i = 1, 2$ if

$$q_i v_i - p_i > \max\{0,\ q_{-i} v_{-i} - p_{-i}\}$$

and buys nothing if

$$\max_i \{q_i v_i - p_i\} < 0.$$

Here, we ignore ties, which have probability zero of occurring. If a buyer with *additive valuations* buys only object i, her payoff is $q_i v_i - p_i$ but if she buys both objects her payoff is $q_1 v_1 + q_2 v_2 - p_3$. Therefore, a buyer with additive valuations buys only object i if

$$q_i v_i - p_i > \max\{0,\ q_{-i} v_{-i} - p_{-i},\ q_1 v_1 + q_2 v_2 - p_3\},$$

buys both objects if

$$q_1 v_1 + q_2 v_2 - p_3 > \max\{0,\ q_1 v_1 - p_1,\ q_2 v_2 - p_2\},$$

and buys nothing otherwise.

For fixed expected qualities (q_1, q_2), let $C_i(q_1, q_2)$ be the probability assigned to the set of valuations for which the buyer only buys object $i = 1, 2$, let $C_3(q_1, q_2)$ denote the probability assigned to the set of valuations for which the consumer buys both objects, and let $C_0(q_1, q_2)$ be the probability assigned to the set of valuations where the consumer buys nothing. Recall that these depend on whether we consider a consumer with unit demand or with additive valuations, and are computed from the uniform distribution of valuations: these are the areas of the respective convex polygons defined by the above inequalities. The seller's revenue is then given by

$$R(q_1, q_2) = \sum_{i=1}^{3} p_i C_i(q_1, q_2).$$

Recalling the characterization of exposed points, the seller chooses a Laguerre diagram $\pi = \{D^1, D^2, \ldots, D^n\}$ of $[\underline{q}, \overline{q}]^2$ with respective barycenters $\mathbf{q}^j = (q_1^j, q_2^j),\ j = 1, 2, \ldots, n$. In other words, for each realization of qualities the seller reveals to the consumer only to which cell the qualities belong. The consumer then updates her belief about the expected qualities to the barycenter $\mathbf{q}^j$ of the cell D^j which contains the true qualities.

The designer's expected revenue is then given by

$$\sum_{j=1}^{n} \mu(D^j) R(q_1^j, q_2^j)$$

and the designer chooses a Laguerre diagram to maximize this expected revenue.

Unit demand. We first consider a buyer with unit demand. Figure 4 shows that the seller discloses only a coarse signal about the qualities of the products. Each cell corresponds to one expected quality pair of the products which we indicate by a triangle in the figure. There are at most 4 cells in the optimal signal: These cells can be roughly interpreted as corresponding to different quality pairs: (1) the orange cell corresponds to both products being of low expected quality, (2) the purple cell corresponds to product 1 being of low quality and product 2 being of high quality, (3) the red cell corresponds to product 1 being of high quality and product 2 being of low quality, and (4) the green cell corresponds to both products being of high quality. In the orange cells, the expected qualities are so low that the buyer never buys either of the products, independent of their valuations. In the red and purple cells, the expected quality of the lower-quality object is so low, that the buyer will never buy the lower-quality object; the buyer will either buy the higher-quality object or buy nothing at all. As the price of product 2 increases, in the optimal information policy, a signal indicating a high quality of product 2 becomes more informative in that it indicates a higher expected quality. This offsets the higher price and still induces some buyer types to purchase the more expensive product. If the price of product 2 becomes too high, no types will buy product 2 and the optimal Laguerre diagram has only two cells, as the right-most panel illustrates.

Under the optimal Laguerre diagram, the seller does not provide full information to the buyer even though that would raise efficiency. By revealing only imprecise information, the buyer's information rents are reduced and the seller's revenue raised. To evaluate the benefit of choosing an optimal information policy induced by a Laguerre diagram, we consider additional, non-optimal information policies, as benchmarks. Table 1 specifies the revenue generated under various information policies. It shows in row (i) the different values of the price p_2, in row (ii) the revenue R_{opt} induced by the optimal Laguerre diagram partition computed with our algorithm, in row (iii) the revenue $R(1,1)$ for an information policy where no information is given to the buyer, and in row (iv) the revenue for an additional benchmark information policy based on a partition generated with Lloyd's algorithm. Here, for the same number of cells as in the computed optimal Laguerre diagram, we generate with Lloyd's algorithm a Laguerre diagram partition which imposes the barycenter of each cell to coincide with the respective site, cf. [18]. The respective Lloyd diagrams used for benchmarking are exemplified in Figure 7. Row (v) of Table 1 shows the revenue $\mathbb{E}(R)$ in case of the full information policy, and in row (vi) the percentage increase in revenue of the optimal power diagram policy compared to the full information policy is displayed. As one can see, optimal information design creates significant value to the seller: in this example it increases profits relative to full information in excess of 10%.

Figure 5 shows the optimal Laguerre diagrams for prices $p_1 = 1$ and $p_2 = 1.25$ and various values of the lower bound on the quality. As the lower bound increases, the optimal signals become less likely to contain significant information about the quality of object 2. The corresponding revenue and the revenue under alternative benchmarks is shown in Table 2.

p_2	1	1.25	1.5	1.75	2
$\mathbf{R_{opt}}$	**0.3153**	**0.2648**	**0.2175**	**0.1839**	**0.1716**
$R(1,1)$	0.0000	0.0000	0.0000	0.0000	0.0000
R_{Lloyd}	0.3056	0.2543	0.1667	0.1667	0.1667
$\mathbb{E}(R)$	0.2833	0.2407	0.1977	0.1657	0.1534
$pp(\%)$	11.30	10.01	10.01	10.98	11.86

Table 1: Monopolist's problem with unit demand, price $p_1 = 1$, quality boundaries $\underline{q} = 0$ and $\overline{q} = 2$ for different prices p_2 (first row) under selected information policies.

$\underline{q}$	0.25	0.5	0.75	1.	1.25
$\mathbf{R_{opt}}$	**0.2999**	**0.3457**	**0.4074**	**0.4853**	**0.5687**
$R(1,1)$	0.1111	0.2000	0.3564	0.4757	0.5687
R_{Lloyd}	0.2838	0.3105	0.3604	0.4583	0.5687
$\mathbb{E}(R)$	0.2728	0.3146	0.3714	0.4528	0.5508
$pp(\%)$	11.30	10.01	10.01	10.98	11.86

Table 2: Monopolist problem with unit demand, prices $p_1 = 1$, $p_2 = 1.25$, upper quality bound $\overline{q} = 2$ for different lower quality bounds $\underline{q} = 0.25, 0.5, 0.75, 1., 1.25$ (first row) under selected information policies.

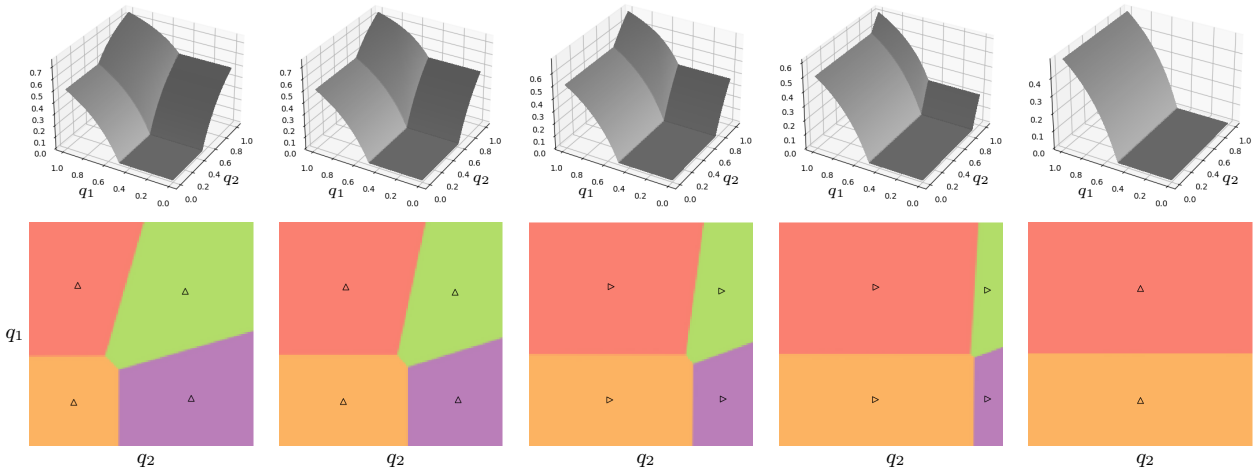

Figure 4: Optimal configurations for the monopolist's problem with unit demand, with prices $p_1 = 1$ and $p_2 = 1, 1.25, 1.5, 1.75, 2$ (second row, from left to right) and quality boundaries $\underline{q} = 0$ and $\overline{q} = 2$. The respective revenue function R for each case is plotted in the first row.

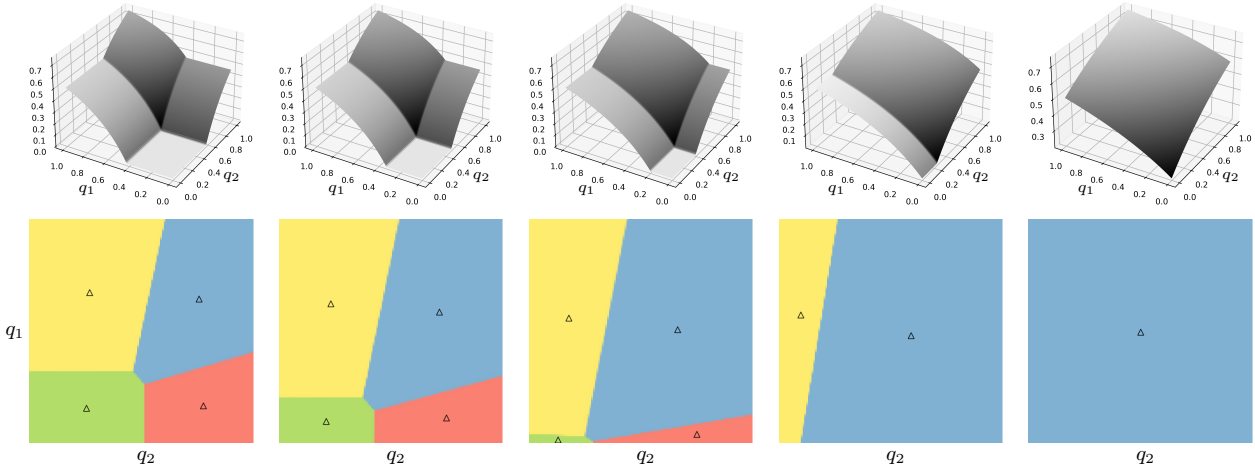

Figure 5: Optimal configurations for the monopolist's problem with unit demand, prices $p_1 = 1$ and $p_2 = 1.25$, upper quality bound $\overline{q} = 2$ and, from left to right, lower quality bounds $\underline{q} = 0.25, 0.5, 0.75, 1, 1.25$ (second row). The respective revenue function R (defined above) is plotted in the first row.

Additive valuations. Figure 6 illustrates optimal information policies induced by Laguerre diagrams if the buyer has additive valuations. The optimal Laguerre diagrams become richer, with up to seven cells. Moreover, there is significant variation in the shape of the diagrams as the bundling surplus/discount varies. As Tables 3 and 4 show, there is again a significant benefit to only partially revealing information compared to fully revealing the qualities.

Δ	-1	-0.875	-0.75	-0.625	-0.5	-0.375	-0.25	-0.125	0
$\mathbf{R_{opt}}$	**0.6577**	**0.5914**	**0.5215**	**0.4602**	**0.4196**	**0.3944**	**0.3754**	**0.3582**	**0.3432**
$R(1,1)$	0.5000	0.4307	0.3516	0.2686	0.1875	0.1143	0.0547	0.0146	0.0000
R_{Lloyd}	0.6528	0.5868	0.4824	0.3843	0.3510	0.3433	0.3533	0.3436	0.3333
$\mathbb{E}(R)$	0.6417	0.5196	0.4578	0.4138	0.3803	0.3544	0.3343	0.3188	0.3069
$pp(\%)$	2.49	13.82	13.91	11.21	10.33	11.29	12.29	12.36	11.83

Table 3: Monopolist's problem with additive valuations for different bundling discounts Δ (first row), quality boundaries $\underline{q} = 0$ and $\overline{q} = 2$ and prices $p_1 = p_2 = 1$, under selected information policies.

Δ	0.125	0.25	0.375	0.5	0.625	0.75
$\mathbf{R_{opt}}$	**0.3303**	**0.3205**	**0.3154**	**0.3153**	**0.3153**	**0.3153**
$R(1,1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R_{Lloyd}	0.3231	0.3142	0.3079	0.3056	0.3056	0.3056
$\mathbb{E}(R)$	0.2979	0.2916	0.2875	0.2851	0.2839	0.2834
$pp(\%)$	10.88	9.91	9.70	10.59	11.06	11.26

Table 4: Monopolist's problem with additive valuations for different bundling surcharges Δ (first row), quality boundaries $\underline{q} = 0$ and $\overline{q} = 2$ and prices $p_1 = p_2 = 1$ under selected information policies.

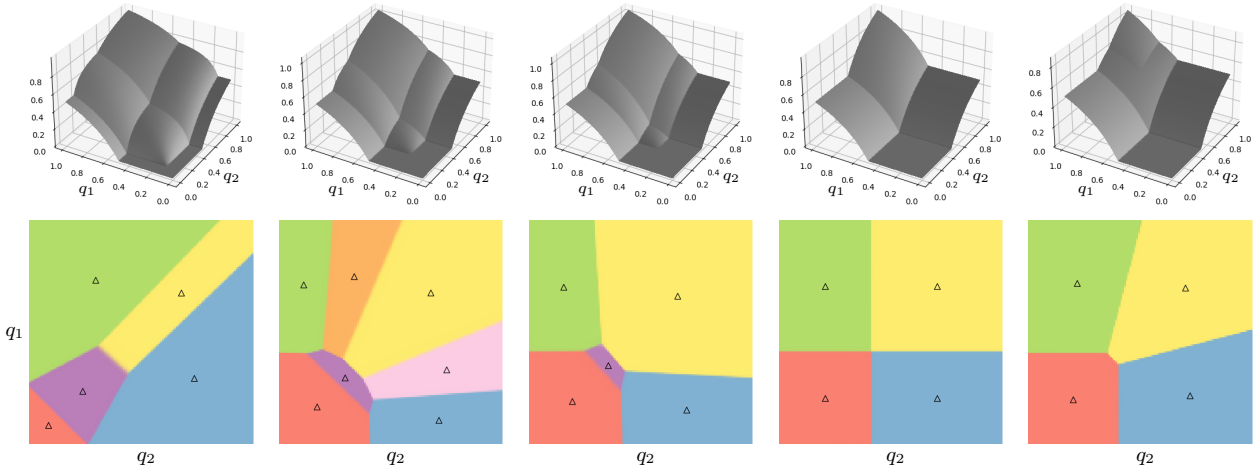

Figure 6: Optimal configurations for the monopolist's problem with additive valuations, with prices $p_1 = p_2 = 1$, quality bounds $\underline{q} = 0$ and $\overline{q} = 2$ and, from left to right, bundling surcharges/discounts $\Delta = -0.75, -0.5, -0.375, 0, 0.375$ (second row). The respective revenue function R (defined above) is plotted in the first row.

Acknowledgement

This work was supported by the German Research Foundation (DFG) via Germany's Excellence Strategy project 390685813 – Hausdorff Center for Mathematics.

References

[1] L. Ambrosio and N. Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.

[2] I. Arieli, Y. Babichenko, R. Smorodinsky, and T. Yamashita. Optimal persuasion via bi-pooling. *Theoretical Economics*, 18(1):15–36, 2023.

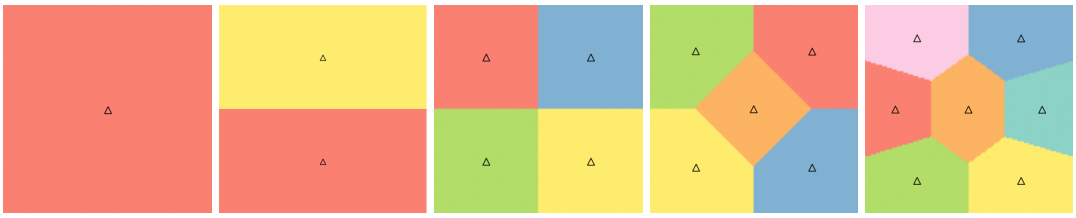

Figure 7: Plot of the Lloyd's centroidal power diagrams for $n = 1, 2, 4, 5, 7$ cells from left to right. The sites coincide with the barycenters of the cells.

[3] F. Aurenhammer. Power diagrams: properties, algorithms and applications. *SIAM Journal on Computing*, 16(1):78–96, 1987.

[4] D. Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

[5] T. Börgers. *An introduction to the theory of mechanism design.* Oxford University Press, USA, 2015.

[6] L. Bottou and O. Bousquet. The tradeoffs of large scale learning. *Advances in neural information processing systems*, 20, 2007.

[7] P. Cartier, J. Fell, and P.-A. Meyer. Comparaison, des mesures portées par un ensemble convexe compact. *Bulletin de la Société Mathématique de France*, 92:435–445, 1964.

[8] B. Charlier, J. Feydy, J. A. Glaunès, F.-D. Collin, and G. Durif. Kernel operations on the gpu, with autodiff, without memory overflows. *Journal of Machine Learning Research*, 22(74):1–6, 2021.

[9] S. Chewi, J. Niles-Weed, and P. Rigollet. Statistical optimal transport. *arXiv preprint arXiv:2407.18163*, 2024.

[10] P. Dworczak and G. Martini. The simple economics of optimal persuasion. *Journal of Political Economy*, 127(5):1993–2048, 2019.

[11] M. Ivanov. Optimal monotone signals in bayesian persuasion mechanisms. *Economic Theory*, 72(3):955–1000, 2021.

[12] E. Kamenica and M. Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

[13] D. P. Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[14] V. L. Klee Jr. Extremal structure of convex sets. ii. *Mathematische Zeitschrift*, 69(1):90–104, 1958.

[15] A. Kleiner, B. Moldovanu, and P. Strack. Extreme points and majorization: Economic applications. *Econometrica*, 89(4):1557–1593, 2021.

[16] A. Kleiner, B. Moldovanu, P. Strack, and M. Whitmeyer. Extreme points of multidimensional fusions. *discussion paper*, 2024.

[17] A. Kolotilin. Optimal information disclosure: A linear programming approach. *Theoretical Economics*, 13(2):607–635, 2018.

[18] S. Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[19] R. B. Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

[20] G. Peyré, M. Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[21] R. R. Phelps. *Lectures on Choquet's theorem.* Springer, 2001.

[22] F. Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

[23] B. Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM Journal on Scientific Computing*, 41(3):A1443–A1481, 2019.

[24] M. Shaked and J. G. Shanthikumar. *Stochastic orders.* Springer, 2007.

[25] V. Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.