No. 600 March 2019

Quadratically regularized optimal transport

D. A. Lorenz, P. Manns, C. Meyer

ISSN: 2190-1767

Noname manuscript No.
(will be inserted by the editor)

Quadratically regularized optimal transport

Dirk A. Lorenz · Paul Manns · Christian Meyer

the date of receipt and acceptance should be inserted later

Abstract We investigate the problem of optimal transport in the so-called Kantorovich form, i.e. given two Radon measures on two compact sets, we seek an optimal transport plan which is another Radon measure on the product of the sets that has these two measures as marginals and minimizes a certain cost function.

We consider quadratic regularization of the problem, which forces the optimal transport plan to be a square integrable function rather than a Radon measure. We derive the dual problem and show strong duality and existence of primal and dual solutions to the regularized problem. Then we derive two algorithms to solve the dual problem of the regularized problem: A Gauss-Seidel method and a semismooth quasi-Newton method and investigate both methods numerically. Our experiments show that the methods perform well even for small regularization parameters. Quadratic regularization is of interest since the resulting optimal transport plans are sparse, i.e. they have a small support (which is not the case for the often used entropic regularization where the optimal transport plan always has full measure).

1 Introduction

In this paper we will investigate a regularized version of the optimal transport problem. Optimal transport dates back to the work of Monge in 1781 but the problem formulation we use here is the one of Kantorovich [14]. Let us fix some notation and formulate the problem: Let $\Omega_1 \subset \mathbb{R}^{d_1}$, $\Omega_2 \subset \mathbb{R}^{d_2}$ be two compact domains, denote $\Omega = \Omega_1 \times \Omega_2$, and assume we are given two positive regular Radon measures $\mu_{1/2}$ on $\Omega_{1/2}$, respectively. Further we assume that a cost function $c : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is given that models the cost of transporting a unit of mass from $x_1 \in \Omega_1$ to $x_2 \in \Omega_2$. The optimal transport problem asks to find a transport plan π, which is a Radon measure on Ω, such that it has minimal overall transport cost $\int_\Omega c(x_1, x_2)\,\mathrm{d}\pi(x_1, x_2)$ among all measures π which have μ_1 and μ_2 as first and second marginals, respectively, i.e. for all Borel sets $A \in \Omega_1$ it holds that

Address(es) of author(s) should be given

$\pi(A \times \Omega_2) = \mu_1(A)$ and for all Borel sets $B \in \Omega_2$ it holds that $\pi(\Omega_1 \times B) = \mu_2(B)$. This problem has been studied extensively and we refer to the books [17, 18, 22, 23, 20]. One particular result is, that an optimal plan π^* exists and that the support of optimal plans is contained in the so-called c-superdifferential of a c-concave function [1, Theorem 1.13]. For many cost functions c, this means that optimal transport plans are supported on small sets and that they are in fact singular with respect to the Lebesgue measure on Ω. This makes the numerical treatment of optimal transport problems difficult and one can employ regularization to obtain approximately optimal plans π that are functions on Ω. The regularization method that has got the most attention recently is regularization with the negative entropy of π and we refer to [15, 9, 4]. Entropic regularization has gotten popular in machine learning applications due to the fact that it allows for the very simple Sinkhorn algorithm (in the discrete case), see [8, 12] and also [16] for a recent and thorough review of the computational aspects of optimal transport.

Regularization different from entropic regularization has been much less studied. We are only aware of works in the discrete case, e.g. [3, 10]. In this work we will investigate the case where we regularize the problem in $L^2(\Omega)$. The paper is organized as follows: In Section 2 we state the problem and analyze existence and duality. It will turn out that existence of solutions of the dual problem will be quite tricky to show, but we will show that dual solutions exist in respective L^2 spaces and that a straightforward optimality system characterizes primal-dual optimality. In Section 3 we derive two different algorithms for the discrete version of the quadratically regularized optimal transport problem, and in Section 4 we comment on a simple discretization scheme and report numerical examples.

Notation. We will abbreviate $x_+ = \max(x, 0)$ (and will apply this also to functions and to measures where $_+$ will mean the positive part from the Hahn-Jordan decomposition). By $C(\Omega)$ we denote that space of continuous functions on Ω (and we will always work on compact sets) equipped with the supremum norm $\|\cdot\|_\infty$ and by $\mathfrak{M}(\Omega)$ we denote the space of Radon measures on a compact domain and we use the norm $\|\mu\|_{\mathfrak{M}} = \sup\{\int f \,\mathrm{d}\mu \mid f \in C(\Omega),\ |f| \leq 1\}$. The Lebesgue measure will be λ (and we also use $\lambda_{1/2}$ to specify the Lebesgue measure on $\Omega_{1/2}$). For convenience, we use $|\Omega|$ for the Lebesgue measure of the set Ω. Furthermore, for a Radon measure $w \in \mathfrak{M}$, we denote the absolutely and singular part arising from the Lebesgue decomposition with respect to the Lebesgue measure by w_{ac} and w_s, i.e. they satisfy $w_{ac} \ll \lambda$ and $w_s \perp \lambda$.

2 Quadratic regularization in the continuous case

For the quadratically regularized optimal transport problem we seek a transport plan $\pi \in L^2(\Omega_1 \times \Omega_2)$ which for a given cost function $c \in L^2(\Omega_1 \times \Omega_2)$, a regularization parameter $\gamma > 0$, and given functions $\mu_{1/2} \in L^2(\Omega_{1/2})$ solves

$$\min_{\pi} \ \langle c, \pi\rangle_{L^2} + \tfrac{\gamma}{2}\|\pi\|_{L^2}^2 \quad \text{subject to} \quad \begin{aligned} &\int_{\Omega_2} \pi(x_1, x_2)\,\mathrm{d}\lambda_2 = \mu_1(x_1), \\ &\int_{\Omega_1} \pi(x_1, x_2)\,\mathrm{d}\lambda_1 = \mu_2(x_2), \\ &\pi(x_1, x_2) \geq 0 \end{aligned} \tag{1}$$

where the constraints are understood pointwise almost everywhere.

2.1 Solutions of the primal problem

It is straight forward to show, that optimal transport plans exist:

Lemma 2.1 *Problem* (1) *has an optimal solution if and only if* $\mu_1 \in L^2(\Omega_1)$, $\mu_2 \in L^2(\Omega_2)$, $\mu_1, \mu_2 \geq 0$ *almost everywhere, and* $\int_{\Omega_1} \mu_1(x_1)\,\mathrm{d}\lambda_1 = \int_{\Omega_2} \mu_2(x_2)\,\mathrm{d}\lambda_2$.

Proof Assume that there is an optimal solution $\pi^* \in L^2(\Omega_1 \times \Omega_2)$. By Jensen's inequality we get

$$\begin{aligned}\int_{\Omega_1} \mu_1^2(x_1)\,\mathrm{d}\lambda_1 &= \int_{\Omega_1} \left(\int_{\Omega_2} \pi^*(x_1,x_2)\,\mathrm{d}\lambda_2\right)^2 \mathrm{d}\lambda_1 \\ &\leq |\Omega_2| \iint_{\Omega_1\times\Omega_2} \pi^*(x_1,x_2)^2\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2 < \infty\end{aligned}$$

which shows $\mu_1 \in L^2(\Omega_1)$. The argument for μ_2 is similar. Non-negativity of $\mu_{1/2}$ follows from non-negativity of π^*. Finally, by Fubini's theorem

$$\begin{aligned}\int_{\Omega_1} \mu_1(x_1)\,\mathrm{d}\lambda_1 &= \iint_{\Omega_1\times\Omega_2} \pi^*(x_1,x_2)\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2 \\ &= \int_{\Omega_2} \mu_2(x_2)\,\mathrm{d}\lambda_2\end{aligned}$$

Conversely, if $\mu_1 \in L^2(\Omega_1)$ and $\mu_2 \in L^2(\Omega_2)$ and $\mu_{1/2} \geq 0$ we set $C := \int_{\Omega_1} \mu_1(x_1)\,\mathrm{d}\lambda_1 = \int_{\Omega_2} \mu_2(x_2)\,\mathrm{d}\lambda_2$. Then $\pi(x_1,x_2) = \frac{1}{C}\mu_1(x_1)\mu_2(x_2)$ is feasible for (1) and since the objective is continuous, coercive, and strongly convex a (unique) minimizer exists. □

2.2 Dual problem and existence of dual solutions

In the following section, we apply the classical Lagrange duality to the linear-quadratic program (1). To this end, let us define the Lagrangian associated with (1). In order to shorten the notation, we set

$$\mu := \gamma\,\mu_1 \otimes \mu_2.$$

Furthermore, we define

$$P_1 : L^2(\Omega) \ni \pi \mapsto \int_{\Omega_2} \pi\,\mathrm{d}\lambda_2 \in L^2(\Omega_1),\; P_2 : L^2(\Omega) \ni \pi \mapsto \int_{\Omega_1} \pi\,\mathrm{d}\lambda_1 \in L^2(\Omega_2), \quad (2)$$

and denote the the primal objective by

$$E_\gamma : L^2(\Omega) \to \mathbb{R}, \quad E_\gamma(\pi) := \int_\Omega c\,\pi\,\mathrm{d}\lambda + \frac{\gamma}{2}\,\|\pi\|^2_{L^2(\Omega)}. \quad (3)$$

Then, the Lagrangian associated with (1) is given by

$$\begin{aligned}
&\mathcal{L} : L^2(\Omega) \times L^2(\Omega_1) \times L^2(\Omega_2) \times L^2(\Omega) \to \mathbb{R},\\
&\mathcal{L}(\pi, \alpha_1, \alpha_2, \varrho) := E_\gamma(\pi) - \langle \varrho, \pi \rangle_{L^2(\Omega)}\\
&\qquad\qquad + \langle \alpha_1, P_1\pi - \mu_1 \rangle_{L^2(\Omega_1)} + \langle \alpha_2, P_2\pi - \mu_2 \rangle_{L^2(\Omega_2)}.
\end{aligned}$$

Moreover, we abbreviate the feasible set of the dual problem by

$$\mathcal{F}_\mathrm{D} := \{(\alpha_1, \alpha_2, \varrho) \in L^2(\Omega_1) \times L^2(\Omega_2) \times L^2(\Omega) : \varrho \geq 0 \ \text{ a.e. in } \Omega\}. \tag{4}$$

Then, by standard arguments, the primal problem in (1) is equivalent to

$$\inf(P) := \inf_{\pi \in L^2(\Omega)} \sup_{(\alpha_1, \alpha_2, \varrho) \in \mathcal{F}_\mathrm{D}} \mathcal{L}(\pi, \alpha_1, \alpha_2, \varrho), \tag{PP}$$

while its (Lagrangian) dual is given by

$$\sup(D) := \sup_{(\alpha_1, \alpha_2, \varrho) \in \mathcal{F}_\mathrm{D}} \inf_{\pi \in L^2(\Omega)} \mathcal{L}(\pi, \alpha_1, \alpha_2, \varrho), \tag{DP}$$

where we used the usual abbreviations for the optimal values. In order to verify the existence of Lagrange multipliers, we need to show that

- there is no duality gap, i.e., $\inf(P) = \sup(D)$,

and that

- the dual problem (DP) admits a solution.

We start with the latter question. For this purpose, we first reformulate the dual problem. Since $\mathcal{L}$ is quadratic w.r.t. π, the inner inf-problem is solved by

$$\pi = \frac{1}{\gamma}(\rho + \alpha_1 \oplus \alpha_2 - c), \tag{5}$$

where the mapping $\oplus : L^2(\Omega_1) \times L^2(\Omega_2) \to L^2(\Omega)$ is defined via

$$(v_1 \oplus v_2)(x_1, x_2) := v_1(x_1) + v_2(x_2) \tag{6}$$

for almost all $(x_1, x_2) \in \Omega$ and all $v_i \in L^2(\Omega_i)$, $i = 1, 2$.

Remark 2.2 The map $\oplus$ is related to the adjoints of the projections $P_{1/2}$ from (2) by $\alpha_1 \oplus \alpha_2 = P_1^*\alpha_1 + P_2^*\alpha_2$.

Inserting (5) into (DP) yields

$$\begin{aligned}
\sup_{\alpha_1 \in L^2(\Omega_1), \alpha_2 \in L^2(\Omega_2)} \sup_{\rho \geq 0} \Big(&- \frac{1}{2\gamma} \int_\Omega (\rho + \alpha_1 \oplus \alpha_2 - c)^2 \,\mathrm{d}\lambda\\
&+ \int_{\Omega_1} \mu_1 \alpha_1 \,\mathrm{d}\lambda_1 + \int_{\Omega_2} \mu_2 \alpha_2 \,\mathrm{d}\lambda_2 \Big)
\end{aligned} \tag{7}$$

Again, the inner optimization problem is quadratic w.r.t. ρ so that its solution is given by

$$\rho = -(\alpha_1 \oplus \alpha_2 - c)_-. \tag{8}$$

Inserted in (7), this results in the following dual problem

$$\left.\begin{aligned} \min \quad & \Phi(\alpha_1,\alpha_2) := \tfrac{1}{2}\|(\alpha_1 \oplus \alpha_2 - c)_+\|^2_{L^2(\Omega)} \\ & \qquad -\gamma \int_{\Omega_1} \mu_1\, \alpha_1 \,\mathrm{d}\lambda_1 - \gamma \int_{\Omega_2} \mu_2\, \alpha_2 \,\mathrm{d}\lambda_2 \\ \text{s.t.} \quad & \alpha_i \in L^2(\Omega_i),\ i = 1,2. \end{aligned}\right\} \tag{D}$$

To prove existence of solutions for this problem, we need to require the following

Assumption 1 *The domains Ω_1 and Ω_2 are compact. Moreover, the cost function c is continuous and fulfills $c \geq \underline{c} > -\infty$. Furthermore, the marginals μ_1 and μ_2 satisfy $\mu_i \in C(\Omega_i)$ and $\mu_i \geq \delta > 0$, $i = 1,2$. In addition we assume that $\int_{\Omega_1} \mu_2 \,\mathrm{d}\lambda_1 = \int_{\Omega_1} \mu_2 \,\mathrm{d}\lambda_1 = 1$.*

Remark 2.3 The last assumption on the normalization of the marginals is just to ease the subsequent analysis and can be relaxed by $\int_{\Omega_1} \mu_2 \,\mathrm{d}\lambda_1 = \int_{\Omega_1} \mu_2 \,\mathrm{d}\lambda_1$, which is needed anyway to ensure the existence of a solution to the primal problem, see Lemma 2.1.

Remark 2.4 Note that there is an obvious source of non-uniqueness for the dual problem (D): We can add a constant to α_1 and subtract it from α_2 and this does not change the dual objective, i.e for any constant C it holds that $\Phi(\alpha_1 + C, \alpha_2 - C) = \Phi(\alpha_1, \alpha_2)$. This non-uniqueness will not cause trouble in the proofs and when convenient, we remove it, e.g. by demanding that $\int_{\Omega_2} \alpha_2 d\lambda_2 = 0$.

Remark 2.5 We emphasize that, for some of the following results, not all hypotheses in Assumption 1 are necessary. For instance, a duality gap can be excluded with less restrictive assumptions. However, the existence of solutions to (D) requires the complete Assumption 1 and, in order to ease the presentation, we require it as a standing assumption for the whole section.

Observe that the objective Φ in (D) is also well defined for functions in $\alpha_i \in L^1(\Omega_i)$ with $(\alpha_1 \oplus \alpha_2 - c)_+ \in L^2(\Omega)$. This gives rise to the following auxiliary dual problem:

$$\left.\begin{aligned} \min \quad & \Phi(\alpha_1,\alpha_2) \\ \text{s.t.} \quad & \alpha_i \in L^1(\Omega_i),\ i = 1,2, \quad (\alpha_1 \oplus \alpha_2 - c)_+ \in L^2(\Omega). \end{aligned}\right\} \tag{D'}$$

Our strategy to prove existence of solutions to (D) is now as follows:

1. First, we show that (D') admits a solution $(\alpha_1^*, \alpha_2^*) \in L^1(\Omega_1) \times L^1(\Omega_2)$, see Proposition 2.10.
2. Then, we prove that α_1^* and α_2^* possess higher regularity, namely that they are functions in $L^2(\Omega_i)$, $i = 1,2$, cf. Theorem 2.11.
3. Thus, (α_1^*, α_2^*) is feasible for (D) and, since the feasible set of (D') contains the one of (D), while the objective of (D') restricted to L^2-functions coincides with the objective in (D), this finally gives that (α_1^*, α_2^*) is indeed optimal for (D).

The reason to consider (D') is essentially that the objective Φ is not coercive in $L^2(\Omega)$, but only in $L^1(\Omega)$ (at least w.r.t. the negative part of α_i). Therefore, we have to deal with weakly* converging sequences in the space of Radon measures within the proof of existence of solutions. For this purpose, we need to extend the objective to a suitable set. To that end, let us define

$$G : L^2(\Omega) \ni w \mapsto \int_\Omega \tfrac{1}{2} w_+^2 - w\mu \,\mathrm{d}\lambda \in \mathbb{R}. \tag{9}$$

Note that, thanks to $\int_{\Omega_1} \mu_2 \,\mathrm{d}\lambda_1 = \int_{\Omega_1} \mu_2 \,\mathrm{d}\lambda_1 = 1$, it holds

$$\Phi(\alpha_1, \alpha_2) = G(\alpha_1 \oplus \alpha_2 - c) - \int_\Omega c\,\mu \,\mathrm{d}\lambda \quad \forall\, \alpha_i \in L^2(\Omega_i),\, i = 1, 2. \tag{10}$$

Of course, G is also well defined as a functional on the feasible set of (D') and we will denote this functional by the same symbol to ease notation. In order to extend G to the space of Radon measures, consider for a given measure $w \in \mathfrak{M}(\Omega)$, the Hahn-Jordan decomposition $w = w_+ + w_-$ and assume that $w_+ \in L^2(\Omega)$. Then, we set $G(w) = \int_\Omega \frac{1}{2} w_+^2 \,\mathrm{d}\lambda - \int_\Omega \mu \,\mathrm{d}w$. With a slight abuse of notation, we denote this mapping by G, too. Note in this context that, if the singular part of w (w.r.t. the Lebesgue measure) vanishes, then also $w_+ \in L^1(\Omega)$ and $w_+(x) = \max\{0, w(x)\}$ λ-a.e. in Ω so that both functionals coincide on $L^2(\Omega)$, which justifies this notation. Furthermore, we also generalize the map $\oplus$ to the measure space by setting

$$\alpha_1 \oplus \alpha_2 \coloneqq \alpha_1 \otimes \lambda_2 + \lambda_1 \otimes \alpha_2, \quad \alpha_i \in \mathfrak{M}(\Omega_i),\, i = 1, 2.$$

Again, it is easily seen that, for $\alpha_i \in L^2(\Omega_i)$, $i = 1, 2$, this definition boils down to the one in (6). Also Remark 2.2 applies in that we can express $\alpha_1 \oplus \alpha_2$ in terms of the adjoints of P_1 and P_2 from (2) when defined appropriately.

The next lemma is rather obvious and covers the coercivity of G in $L^1(\Omega)$ as indicated above.

Lemma 2.6 *Let Assumption 1 hold and suppose that a sequence $\{w^n\} \subset L^2(\Omega)$ fulfills*

$$G(w^n) \le C < \infty \quad \forall\, n \in \mathbb{N}.$$

Then, the sequences $\{w_+^n\}$ and $\{w_-^n\}$ are bounded in $L^2(\Omega)$ and $L^1(\Omega)$, respectively.

Proof We rewrite G as $G(w) = \int_\Omega \frac{1}{2} w_+^2 - w_+\mu \,\mathrm{d}\lambda + \int_\Omega w_-\mu \,\mathrm{d}\lambda$. The positivity of μ then implies

$$\|w_+^n\|_{L^2(\Omega)}^2 = G(w^n) + \int_\Omega w_+^n \mu \,\mathrm{d}\lambda - \int_\Omega w_-^n \mu \,\mathrm{d}\lambda \le C + \|\mu\|_{L^2(\Omega)} \|w_+^n\|_{L^2(\Omega)},$$

which gives the first assertion. To see the second one, we use $\mu \ge \delta$ to estimate

$$\begin{aligned} C \ge G(w^n) &= \int_\Omega \tfrac{1}{2} (w_+^n - \mu)^2 \,\mathrm{d}\lambda - \int_\Omega \mu_2/2 \,\mathrm{d}\lambda + \int_\Omega w_-^n \mu \,\mathrm{d}\lambda \\ &\ge - \int_\Omega \mu_2/2 \,\mathrm{d}\lambda + \delta \|w_-^n\|_{L^1(\Omega)}, \end{aligned}$$

which finishes the proof. □

The next lemma provides a lower semicontinuity result for G w.r.t. weak* convergence in $\mathfrak{M}(\Omega)$. Note that, here, we need the extension of G as introduced above.

Lemma 2.7 *Let Assumption 1 be fulfilled and a sequence $\{w_n\} \subset L^2(\Omega)$ be given such that $w^n \rightharpoonup^* w^*$ in $\mathfrak{M}(\Omega)$ and $G(w^n) \le C < \infty$ for all $n \in \mathbb{N}$. Then there holds $w_+^* \in L^2(\Omega)$ and*

$$G(w^*) \le \liminf_{n\to\infty} G(w^n).$$

Proof Define the function

$$f : \mathbb{R} \ni \xi \mapsto \xi_+^2 \in \mathbb{R},$$

which is clearly convex and continuous. Therefore, in light of the weak* convergence of w^n, [11, Theorem 5.19] yields

$$\liminf_{n\to\infty} \int_\Omega f(w^n)\,\mathrm{d}\lambda \ge \int_\Omega f(w_{ac}^*)\,\mathrm{d}\lambda + \int_\Omega f^\infty\left(\frac{\mathrm{d}w^*}{\mathrm{d}|w_s^*|}\right)\mathrm{d}|w_s^*| \tag{11}$$

where $\frac{\mathrm{d}w^*}{\mathrm{d}|w_s^*|}$ denotes the Radon-Nikodym derivative of w^* w.r.t the total variation measure of w_s^* and f^∞ is the recession function of f, see (12) below. For $x \in \operatorname{supp}(w_s^*)$ we get by the definition of the Radon-Nikodym derivative that

$$\frac{\mathrm{d}w^*}{\mathrm{d}|w_s^*|}(x) = \lim_{r\searrow 0} \frac{w^*(B_r(x))}{|w_s^*|(B_r(x))} = \lim_{r\searrow 0} \frac{w_{ac}^*(B_r(x)) + w_s^*(B_r(x))}{|w_s^*|(B_r(x))} = \frac{\mathrm{d}w_s^*}{\mathrm{d}|w_s^*|}(x)$$

since $\lim_{r\searrow 0} w_{ac}^*(B_r(x)) = 0$.

Regarding the recession function f^∞, we obtain

$$\begin{aligned} f^\infty(z) &= \sup_w \{f(w+z) - f(w)\} &(12)\\ &= \sup_w \{(w+z)_+^2 - w_+^2\} \\ &\le \sup_w \{w_+^2 + 2z_+w_+ + z_+^2 - w_+^2\} &(13)\end{aligned}$$

by definition. If $z > 0$, we obtain $f^\infty(z) = \infty$ because the inequality (13) becomes an equality for $w > 0$ and the sup can be made arbitrarily large. If $z \le 0$, we obtain $f^\infty(z) \le 0$ from (13) and this is attained with the choice $w = 0$. This shows

$$f^\infty(z) = \begin{cases} \infty: & z > 0 \\ 0: & z \le 0 \end{cases}. \tag{14}$$

Now, let us return to (11). Due to $f \ge 0$, the assumed boundedness of $\{G(w^n)\}$, the weak* convergence of $\{w^n\}$, and the continuity of μ, this inequality leads to

$$\begin{aligned}\int_\Omega f^\infty\left(\frac{\mathrm{d}w^*}{\mathrm{d}|w_s^*|}\right)\mathrm{d}|w_s^*| &\le \liminf_{n\to\infty}\Big(G(w^n) + \int_\Omega w^n\,\mu\,\mathrm{d}\lambda\Big)\\ &\le \limsup_{n\to\infty} G(w^n) + \lim_{n\to\infty}\int_\Omega w^n\,\mu\,\mathrm{d}\lambda \le C + \int_\Omega \mu\,\mathrm{d}w^* < \infty.\end{aligned}$$

In view of (14), this can only be true if $\frac{\mathrm{d}w_s^*}{\mathrm{d}|w_s^*|} = \frac{\mathrm{d}w^*}{\mathrm{d}|w_s^*|} \leq 0$ $|w_s^*|$-a.e. on $\mathrm{supp}(|w_s^*|)$. Since $w_s^* \ll |w_s^*|$, we have that $w_s^* = \frac{\mathrm{d}w_s^*}{\mathrm{d}|w_s^*|}|w_s^*|$ and thus $w_s^* \leq 0$, in other words $(w_s^*)_+ = 0$. Since for any Lebesgue null set A we have

$$(w_+^*)_s(A) = w_+^*(A) = \sup_{B\subset A} (w_s^*(B) + w_{ac}^*(B)) = (w_s^*)_+(A),$$

which shows that $(w_+^*)_s = 0$ so that we already have $w_+^* \in L^1(\Omega)$. Moreover, for every Borel set $A \subset \Omega$, we find

$$w_+^*(A) \leq \sup_{B\subset A} (w_s^*(B)) + \sup_{B\subset A} (w_{ac}^*(B)) = (w_s^*)_+(A) + (w_{ac}^*)_+(A) = (w_{ac}^*)_+(A)$$

and, since w_+^* is absolutely continuous w.r.t. the Lebesgue measure, this also holds for the associated density functions so that $f(w_{ac}^*) = (w_{ac}^*)_+^2 \geq (w_+^*)^2$ λ-a.e. in Ω. Together with $f^\infty(\frac{\mathrm{d}w^*}{\mathrm{d}|w_s^*|}) = 0$ $|w_s^*|$-a.e. on $\mathrm{supp}(|w_s^*|)$ as seen above and $w^n \rightharpoonup^* w^*$, this in combination with (11) yields

$$\begin{aligned}
G(w^*) &= \int_\Omega \tfrac{1}{2}\,(w_+^*)^2\,\mathrm{d}\lambda - \int_\Omega \mu\,\mathrm{d}w^* \\
&\leq \tfrac{1}{2}\int_\Omega f(w_{ac}^*)\,\mathrm{d}\lambda + \tfrac{1}{2}\int_\Omega f^\infty\left(\frac{\mathrm{d}w^*}{\mathrm{d}|w_s^*|}\right)\mathrm{d}|w_s^*| - \int_\Omega \mu\,\mathrm{d}w^* \\
&\leq \liminf_{n\to\infty} \tfrac{1}{2}\int_\Omega f(w^n)\mathrm{d}\lambda - \lim_{n\to\infty}\int_\Omega w^n\,\mu\,\mathrm{d}\lambda = \liminf_{n\to\infty} G(w^n) < \infty.
\end{aligned}$$

This shows $w_+^* \in L^2(\Omega)$ and $G(w^*) \leq \liminf_{n\to\infty} G(w^n)$ as claimed. □

Before we are in the position to prove existence for (D'), we need two additional results on the $\oplus$-operator in the space of Radon measures.

Lemma 2.8 *If $\alpha_i \in \mathfrak{M}(\Omega_i)$, $i = 1, 2$ and $\int_{\Omega_2} d\alpha_2 = 0$, then it holds that*

$$\|\alpha_1\|_{\mathfrak{M}} \leq \tfrac{1}{|\Omega_2|}\|\alpha_1 \oplus \alpha_2\|_{\mathfrak{M}} \quad \text{and} \quad \|\alpha_2\|_{\mathfrak{M}} \leq \tfrac{2}{|\Omega_1|}\|\alpha_1 \oplus \alpha_2\|_{\mathfrak{M}}$$

Proof We estimate

$$\begin{aligned}
\|\alpha_1 \oplus \alpha_2\|_{\mathfrak{M}} &= \sup_{\|\phi\|_\infty \leq 1} \iint_{\Omega_1\times\Omega_2} \phi(x_1, x_2)\,\mathrm{d}(\alpha_1(x_1) + \alpha_2(x_2)) \\
&\geq \sup_{\substack{\|\phi_1\|_\infty\leq 1\\ \|\phi_2\|_\infty \leq 1}} \iint_{\Omega_1\times\Omega_2} \phi_1(x_1)\phi_2(x_2)\,\mathrm{d}(\alpha_1(x_1) + \alpha_2(x_2)) \\
&= \sup_{\substack{\|\phi_1\|_\infty\leq 1\\ \|\phi_2\|_\infty \leq 1}} \Bigg[\iint_{\Omega_1\times\Omega_2} \phi_1(x_1)\phi_2(x_2)\,\mathrm{d}\alpha_1(x_1)\mathrm{d}\lambda_2 \\
&\qquad + \iint_{\Omega_1\times\Omega_2} \phi_1(x_1)\phi_2(x_2)\,\mathrm{d}\lambda_1\mathrm{d}\alpha_2(x_2)\Bigg].
\end{aligned} \tag{15}$$

We estimate the right hand side from below by taking $\phi_2 \equiv 1$ and get, since $\int_{\Omega_2} \mathrm{d}\alpha_2(x_2) = 0$, that

$$\begin{aligned}\|\alpha_1 \oplus \alpha_2\|_{\mathfrak{M}} &\geq \sup_{\|\phi_1\|_\infty \leq 1} \int_{\Omega_1} \phi_1(x_1) d\alpha_1(x_1)|\Omega_2| + \int_{\Omega_2} d\alpha_2(x_2) \int_{\Omega_1} \phi_1(x_1)\mathrm{d}\lambda_1 \\ &= |\Omega_2|\|\alpha_1\|_{\mathfrak{M}}.\end{aligned}$$

Now we start again at (15) and estimate from below by taking $\phi_1 \equiv 1$ to get

$$\begin{aligned}\|\alpha_1 \oplus \alpha_2\|_{\mathfrak{M}} &\geq \sup_{\|\phi_2\|_\infty \leq 1} \int_{\Omega_1} \mathrm{d}\alpha_1(x_1) \int_{\Omega_2} \phi(x_2)\,\mathrm{d}\lambda_2 + \int_{\Omega_2} \phi_2(x_2)\,\mathrm{d}\alpha(x_2)|\Omega_1| \\ &\geq -|\Omega_2| \int_{\Omega_1} d\alpha_1(x_1) + |\Omega_1|\|\alpha_2\|_{\mathfrak{M}}\end{aligned}$$

which implies

$$|\Omega_1|\|\alpha_2\|_{\mathfrak{M}} \leq \|\alpha_1 \oplus \alpha_2\|_{\mathfrak{M}} + |\Omega_2| \int_{\Omega_1} d\alpha_1(x_1) \leq \|\alpha_1 \oplus \alpha_2\|_{\mathfrak{M}} + |\Omega_2|\|\alpha_1\|_{\mathfrak{M}}$$

which completes the proof. □

Lemma 2.9 *Let $c \in C(\Omega)$ and $\alpha_i \in \mathfrak{M}(\Omega_i)$ for $i \in \{1,2\}$ with Lebesgue decompositions, $\alpha_i = f_i + \eta_i$ satisfying $f_i \ll \lambda$ and $\eta_i \perp \lambda$ for $i \in \{1,2\}$. Then,*

$$(\alpha_1 \oplus \alpha_2 - c)_+ = (f_1 \oplus f_2 - c + (\eta_1)_+ \oplus (\eta_2)_+)_+. \tag{16}$$

Proof The measures f_i,η_i exist by Lebesgue's decomposition theorem, see Theorem 1.155 in [11]. We combine these decompositions with $\alpha_1 \oplus \alpha_2 = \alpha_1 \otimes \lambda + \lambda \otimes \alpha_2$ to arrive at Lebesgue's decomposition of $\alpha_1 \oplus \alpha_2$ with respect to $\lambda \otimes \lambda$, namely

$$\alpha_1 \oplus \alpha_2 - c = f_1 \oplus f_2 - c + \eta_1 \oplus \eta_2 \tag{17}$$

$$f_1 \oplus f_2 - c \ll \lambda \otimes \lambda \tag{18}$$

$$\eta_1 \oplus \eta_2 \perp \lambda \otimes \lambda \tag{19}$$

(which holds true because $c \in C(\Omega) \hookrightarrow \mathfrak{M}(\Omega)$). Now, we consider the Hahn-Jordan decomposition of η_1,

$$\begin{aligned}\eta_1 &= (\eta_1)_+ - (\eta_1)_- \\ (\eta_1)_+ &\perp (\eta_1)_-,\end{aligned} \tag{20}$$

and obtain from (17) that

$$\begin{aligned}\alpha_1 \oplus \alpha_2 - c &= (f_1 + \eta_1) \oplus (f_2 + \eta_2) - c \\ &= f_1 \oplus f_2 + \eta_1 \oplus \eta_2 - c \\ &= f_1 \oplus f_2 + \big((\eta_1)_+ - (\eta_1)_-\big) \oplus \eta_2 - c \\ &= f_1 \oplus f_2 + (\eta_1)_+ \otimes \lambda - (\eta_1)_- \otimes \lambda + \lambda \otimes \eta_2 - c \\ &= f_1 \oplus f_2 - c + (\eta_1)_+ \oplus \eta_2 - (\eta_1)_- \otimes \lambda.\end{aligned}$$

Furthermore,

$$(\eta_1)_- \otimes \lambda \perp f_1 \oplus f_2 - c + (\eta_1)_+ \oplus \eta_2$$

where the singularity with respect to $f_1 \oplus f_2 - c$ is due to (18) and (19) and the singularity with respect to $(\eta_1)_+ \oplus \eta_2$ is due to (20). Thus,

$$\begin{aligned}(\alpha_1 \oplus \alpha_2 - c)_- &= (f_1 \oplus f_2 - c + (\eta_1)_+ \oplus \eta_2)_- + (-(\eta_1)_- \otimes \lambda)_- \\ &= (f_1 \oplus f_2 - c + (\eta_1)_+ \oplus \eta_2)_- + (\eta_1)_- \otimes \lambda.\end{aligned}$$

as $(\eta_1)_- \otimes \lambda$ is a positive measure. Consequently,

$$(\alpha_1 \oplus \alpha_2 - c)_+ = (f_1 \oplus f_2 - c + (\eta_1)_+ \oplus \eta_2)_+.$$

Repeating this argument with the Hahn-Jordan decomposition of η_2 yields the claim. □

Now we are ready to prove the existence result for (D'):

Proposition 2.10 *Under Assumption 1 the minimization problem* (D') *admits a solution* $(\alpha_1^*, \alpha_2^*) \in L^1(\Omega_1) \times L^1(\Omega_2)$.

Proof We proceed via the classical direct method of the calculus of variations. For this purpose, let $\{(\alpha_1^n, \alpha_2^n)\} \subset L^1(\Omega_1) \times L^1(\Omega_2)$ with $(\alpha_1^n \oplus \alpha_2^n - c)_+ \in L^2(\Omega)$ be a minimizing sequence for (D'), where we shift α_1 and α_2 by adding and subtracting constants such that we obtain $\int_{\Omega_2} \alpha_2 \, \mathrm{d}\lambda_2 = 0$. Note that, due to its additive structure, this does not change the objective Φ in (D'), cf. Remark 2.4.

Next, let us define $w^n := \alpha_1^n \oplus \alpha_2^n - c$. Then, thanks to (10) and Lemma 2.6, the sequence $\{w^n\}$ is bounded in $L^1(\Omega)$. Hence, there is a weakly* converging subsequence, which we denote by the same symbol w.l.o.g., i.e., $w^n \rightharpoonup^* \tilde{w}$ in $\mathfrak{M}(\Omega)$. Now, Lemma 2.7 applies giving that

$$\tilde{w}_+ \in L^2(\Omega), \tag{21}$$

$$G(\tilde{w}) \le \liminf_{n\to\infty} G(w_n). \tag{22}$$

Since $\{w^n\}$ is bounded in $\mathfrak{M}(\Omega)$, the same holds for $\{\alpha_1^n \oplus \alpha_2^n\}$ and, as α_2^n is normalized, Lemma 2.8 gives that $\{\alpha_i^n\}$ is bounded in $\mathfrak{M}(\Omega_i)$, $i = 1, 2$. Therefore, we can select a further (sub-)subsequence, still denoted by the same symbol to ease notation, such that

$$\alpha_i^n \rightharpoonup^* \tilde{\alpha}_i \quad \text{in } \mathfrak{M}(\Omega_i), \quad i = 1, 2.$$

Since the mapping $\mathfrak{M}(\Omega_1) \times \mathfrak{M}(\Omega_2) \ni (\alpha_1, \alpha_2) \mapsto \alpha_1 \oplus \alpha_2 \in \mathfrak{M}(\Omega)$ is the adjoint of the projection mapping $C(\Omega) \ni \varphi \mapsto \big(\int_{\Omega_2} \varphi \, \mathrm{d}\lambda_2, \int_{\Omega_1} \varphi \, \mathrm{d}\lambda_1\big) \in C(\Omega_1) \times C(\Omega_2)$, see Remark 2.2, it is weakly* continuous so that

$$\tilde{w} = \tilde{\alpha}_1 \oplus \tilde{\alpha}_2 - c. \tag{23}$$

Next, we investigate the singular parts of $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$. We start with the positive part and employ Lebesgue's decomposition of $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$:

$$\tilde{\alpha}_i = \alpha_i^* + \tilde{\eta}_i, \quad \alpha_i^* \ll \lambda_i, \quad \tilde{\eta}_i \perp \lambda_i, \quad i = 1, 2.$$

In the following we will see that the regular parts $\alpha_i^* \in L^1(\Omega_i)$, $i = 1, 2$, are exactly the solution of (D'). For this purpose, we first show that the positive parts of $\tilde{\eta}_1$ and $\tilde{\eta}_2$ vanish. We have $\alpha_1^* \oplus \alpha_2^* - c \ll \lambda$, $\tilde{\eta}_1 \oplus \tilde{\eta}_2 \perp \lambda$, and, by uniqueness of

Lebesgue's decomposition, $\tilde{w}_s = \tilde{\eta}_1 \oplus \tilde{\eta}_2$. But from (21), we know that $(\tilde{w}_s)_+ = 0$. Combining this fact with Lemma 2.9, applied to the case $f_1 = 0$, $f_2 = 0$, and $c = 0$, we obtain

$$(\tilde{\eta}_1 \oplus \tilde{\eta}_2)_+ = (\tilde{\eta}_1)_+ \oplus (\tilde{\eta}_2)_+.$$

and consequently, $(\tilde{\eta}_i)_+ = 0$ for $i = 1, 2$ by positivity. Therefore, $(\tilde{\alpha}_i)_+$ are L^1-functions rather than measures. Moreover, by applying once again Lemma 2.9, we deduce from $(\tilde{\eta}_i)_+ = 0$ that

$$(\alpha_1^* \oplus \alpha_2^* - c)_+ = (\tilde{\alpha}_1 \oplus \tilde{\alpha}_2 - c)_+ = \tilde{w}_+ \in L^2(\Omega). \tag{24}$$

This shows the feasibility of (α_1^*, α_2^*) for (D').

To show its optimality, consider the objective in (D'). Due to (24), we obtain

$$\int_\Omega (\tilde{\alpha}_1 \oplus \tilde{\alpha}_2 - c)_+^2 \, \mathrm{d}\lambda = \int_\Omega (\alpha_1^* \oplus \alpha_2^* - c)_+^2 \, \mathrm{d}\lambda. \tag{25}$$

Regarding the other summands of the objective Φ from (D'), we get

$$\int_{\Omega_i} \mu_i \, \mathrm{d}\tilde{\alpha}_i = \int_{\Omega_i} \mu_i \, \alpha_i^* \, \mathrm{d}\lambda_i - \int_{\Omega_i} \mu_i \, \mathrm{d}(\tilde{\eta}_i)_- \leq \int_{\Omega_i} \mu_i \, \alpha_i^* \, \mathrm{d}\lambda_i \tag{26}$$

Together with (25) and (22), this implies for the objective in (D')

$$\begin{aligned} \Phi(\alpha_1^*, \alpha_2^*) &= \tfrac{1}{2} \int_\Omega (\alpha_1^* \oplus \alpha_2^* - c)_+^2 \, \mathrm{d}\lambda - \gamma \int_{\Omega_1} \mu_1 \, \alpha_1^* \, \mathrm{d}\lambda_1 - \gamma \int_{\Omega_2} \mu_2 \, \alpha_2^* \, \mathrm{d}\lambda_2 \\ &\leq G(\tilde{\alpha}_1 \oplus \tilde{\alpha}_2 - c) - \int_\Omega c \, \mu \, \mathrm{d}\lambda \\ &= G(\tilde{w}) - \int_\Omega c \, \mu \, \mathrm{d}\lambda \\ &\leq \liminf_{n \to \infty} \Phi(\alpha_1^n, \alpha_2^n), \end{aligned} \tag{27}$$

which demonstrates the optimality of (α_1^*, α_2^*). □

In the following, we assume that $\int_{\Omega_2} \alpha_2^* \, \mathrm{d}\lambda_2 = 0$. If this is not the case, then we can again shift α_1^* and α_2^* without changing the value of Φ, cf. Remark 2.4.

Theorem 2.11 *Let Assumption 1 hold. Then every optimal dual solution* (α_1^*, α_2^*) *from Proposition 2.10 satisfies* $\alpha_i^* \in L^2(\Omega_i)$*,* $i = 1, 2$*, and is therefore also a solution of the original dual problem* (D).

Proof We again consider the positive and the negative part separately and start with $(\alpha_1^*)_-$. Let $\varphi \in C_c^\infty(\Omega_1)$ and $t > 0$ be fixed, but arbitrary. Then, thanks to

$$0 \leq ((\alpha_1^* + t\,\varphi) \oplus \alpha_2^* - c)_+ \leq (\alpha_1^* \oplus \alpha_2^* - c)_+ + t\,\varphi_+,$$

Proposition 2.10 implies that $((\alpha_1^* + t\varphi) \oplus \alpha_2^* - c)_+ \in L^2(\Omega)$ so that $(\alpha_1^* + t\varphi, \alpha_2^*)$ is feasible for (D'). Therefore, the optimality of (α_1^*, α_2^*) for (D') yields

$$\tfrac{1}{2} \int_\Omega \frac{1}{t} \Big(((\alpha_1^* + t\,\varphi) \oplus \alpha_2^* - c)_+^2 - (\alpha_1^* \oplus \alpha_2^* - c)_+^2 \Big) \, \mathrm{d}\lambda - \gamma \int_{\Omega_1} \mu_1 \, \varphi \, \mathrm{d}\lambda_1 \geq 0 \quad \forall\, t > 0.$$

Owing to the continuous differentiability of $\mathbb{R} \ni r \mapsto r_+^2 \in \mathbb{R}$, the first integrand converges to $2(\alpha_1^* \oplus \alpha_2^* - c)_+ \varphi$ λ-a.e. in Ω. Moreover, the Lipschitz continuity of the max-function gives that

$$\frac{1}{t}\Big(((\alpha_1^* + t\,\varphi) \oplus \alpha_2^* - c)_+^2 - (\alpha_1^* \oplus \alpha_2^* - c)_+^2\Big) \leq |\varphi|^2 + 2\,|\varphi|\,(\alpha_1^* \oplus \alpha_2^* - c)_+^2 \quad \text{a.e. in } \Omega.$$

Hence, due to Lebesgue's dominated convergence theorem, we are allowed to pass to the limit $t \searrow 0$ and obtain in this way

$$\int_{\Omega_1} \Big(\int_{\Omega_2} (\alpha_1^* \oplus \alpha_2^* - c)_+ \,\mathrm{d}\lambda_2 - \gamma\mu_1\Big)\varphi \,\mathrm{d}\lambda_1 \geq 0.$$

Since $\varphi \in C_c^\infty(\Omega)$ was arbitrary, the fundamental lemma of the calculus of variations thus gives

$$\int_{\Omega_2} (\alpha_1^* \oplus \alpha_2^* - c)_+ \,\mathrm{d}\lambda_2 = \gamma\mu_1 \quad \lambda_1\text{-a.e. in } \Omega_1. \tag{28}$$

Next, define the following sequence of functions in $L^1(\Omega_2)$:

$$f_n(x_2) := (-n + \alpha_2^*(x_2) - \underline{c})_+, \quad n \in \mathbb{N},$$

where $\underline{c}$ is the lower bound for c from Assumption 1. Then we have $f_n \geq 0$ λ_2-a.e. Ω_2 and $f_n \searrow 0$ λ_2-a.e. in Ω_2 so that the monotone convergence theorem gives

$$\int_{\Omega_2} (-n + \alpha_2^*(x_2) - \underline{c})_+ \,\mathrm{d}\lambda_2 = \int_{\Omega_2} f_n(x_2) \,\mathrm{d}\lambda_2 \to 0 \quad \text{as } n \to \infty.$$

Thus there exists $N \in \mathbb{N}$ such that

$$\int_{\Omega_2} (-N + \alpha_2^*(x_2) - \underline{c})_+ \,\mathrm{d}\lambda_2 < \gamma\,\delta, \tag{29}$$

where $\delta > 0$ is the threshold for μ_1 from Assumption 1. Now assume that $\alpha_1^* \leq -N$ λ_1-a.e. on a set of $E \subset \Omega_1$ of positive Lebesgue measure. Then

$$\int_{\Omega_2} (\alpha_1^* \oplus \alpha_2^* - c)_+ \,\mathrm{d}\lambda_2 \leq \int_{\Omega_2} (-N \oplus \alpha_2^* - \underline{c})_+ \,\mathrm{d}\lambda_2 < \gamma\,\delta \leq \gamma\,\mu_1 \quad \lambda_1\text{-a.e. in } E,$$

which contradicts (28). Therefore, $\alpha_1^* > -N$ λ_1-a.e. in Ω_1, which even implies that $(\alpha_1^*)_- \in L^\infty(\Omega_1)$. Concerning $(\alpha_2^*)_-$, one can argue in exactly the same way to conclude that $(\alpha_2^*)_- \in L^\infty(\Omega_2)$, too.

For the positive parts we find

$$\begin{aligned}
&|\Omega_2|\,\|\alpha_1^*\|_{L^2(\Omega_1)}^2 + |\Omega_1|\,\|\alpha_2^*\|_{L^2(\Omega_2)}^2 \\
&\quad = \int_\Omega |\alpha_1^* \oplus \alpha_2^*|^2 \,\mathrm{d}\lambda \qquad\qquad \Big(\text{since } \int_{\Omega_2} \alpha_2^* \,\mathrm{d}\lambda_2 = 0\Big) \\
&\quad = \int_\Omega (\alpha_1^* \oplus \alpha_2^*)_+^2 + (\alpha_1^* \oplus \alpha_2^*)_-^2 \,\mathrm{d}\lambda \\
&\quad \leq 2\int_\Omega (\alpha_1^* \oplus \alpha_2^* - c)_+^2 + c_+^2 + (\alpha_1^*)_-^2 + (\alpha_2^*)_-^2 \,\mathrm{d}\lambda < \infty,
\end{aligned}$$

where we used (24) and the boundedness of the negative parts proven above. Note that the constant shift, potentially needed to ensure $\int_{\Omega_2} \alpha_2^* \, d\lambda_2 = 0$ has no effect on the equation in (24) due to the additive structure of $\oplus$.

We have thus shown that (α_1^*, α_2^*) is feasible for (D). Since (α_1^*, α_2^*) solves (D'), whose objective is the same as in (D), while its feasible set is larger, this implies that we have found a solution to (D). □

Now that we have established the existence of solutions to the dual problem, we turn to the duality gap. We follow standard arguments, see e.g. [21], and show the following by means of the strict separation theorem:

Proposition 2.12 *Under Assumption 1 the unique minimizer π^* of* (1) *satisfies*

$$E_\gamma(\pi^*) \le \sup(D), \tag{30}$$

where E_γ is the primal objective from (3) *and* $\sup(D)$ *again denotes the optimal value of the dual problem.*

Proof We define the set

$$\begin{aligned} K := \{(r, v_1, v_2, w) \in \mathbb{R} \times L^2(\Omega_1) \times L^2(\Omega_2) \times L^2(\Omega) \,| \\ \exists\, \pi \in L^2(\Omega) : E_\gamma(\pi) \le r,\, -\pi \le w,\, P_i\, \pi - \mu_i = v_i,\, i = 1, 2\}, \end{aligned}$$

where P_i is as defined in (2). This set is convex (due to linearity of P_i, $i = 1, 2$, and convexity of E_γ) and non-empty, since $(E_\gamma(\pi^*), 0, 0, 0) \in K$. Moreover, it is closed. To see this, consider an arbitrary sequence $\{(r_n, v_{1,n}, v_{2,n}, w_n)\}_{n\in\mathbb{N}} \subset K$ with

$$(r_n, v_{1,n}, v_{2,n}, w_n) \to (r, v_1, v_2, w) \quad \text{in } \mathbb{R} \times L^2(\Omega_1) \times L^2(\Omega_2) \times L^2(\Omega).$$

Then, for every $n \in \mathbb{N}$, there exists $\pi_n \in L^2(\Omega)$ such that the conditions in the definition of K are fulfilled, in particular $E_\gamma(\pi_n) \le r_n$. Thus, $\{\pi_n\}$ is bounded in $L^2(\Omega)$ and, consequently, there is a weakly convergent subsequence, w.l.o.g. the whole sequence itself, i.e., $\pi_n \rightharpoonup \pi$ in $L^2(\Omega)$. Due to convexity and continuity, E_γ is weakly semicontinuous, which, together with the linearity of P_i, $i = 1, 2$, implies that $(r, v_1, v_2, w) \in K$ giving in turn the closedness of K.

Now, since K is non-empty, closed, and convex, we can apply the strict separation theorem. For this purpose, note that $(E_\gamma(\pi^*) - 1/n, 0, 0, 0) \notin K$, $n \in \mathbb{N}$, by optimality of π^*. Therefore, the strict separation theorem implies the existence of a separating hyperplane, i.e., there exists

$$(\sigma_n, \alpha_{1,n}, \alpha_{2,n}, \varrho_n) \in \mathbb{R}\times L^2(\Omega_1)\times L^2(\Omega_2)\times L^2(\Omega), \quad (\sigma_n, \alpha_{1,n}, \alpha_{2,n}, \varrho_n) \ne (0, 0, 0, 0)$$

such that, in view of the construction of K and noting that $(E_\gamma(\pi) + \delta, P_1\pi - \mu_1, P_2\pi - \mu_2, -\pi + z) \in K$ for all $\pi \in L^2(\Omega)$, $\delta > 0$, and $z \in L^2(\Omega)$, $z \ge 0$,

$$\begin{aligned} \sigma_n\Big(E_\gamma(\pi^*) - \frac{1}{n}\Big) < \sigma_n(E_\gamma(\pi) + \delta) + \langle \varrho_n, -\pi + z\rangle_{L^2(\Omega_1\times\Omega_2)} \\ + \langle \alpha_{1,n}, P_1\pi - \mu_1\rangle_{L^2(\Omega_1)} + \langle \alpha_{2,n}, P_2\pi - \mu_2\rangle_{L^2(\Omega_2)} \\ \forall\, (\delta, \pi, z) \in \mathbb{R} \times L^2(\Omega)^2 : \delta \ge 0,\, z \ge 0. \end{aligned} \tag{31}$$

If we choose $\pi = \pi^*$, $\delta = 1$, and $z = \pi^* \ge 0$, the feasibility of π^* for (1) implies

$$0 \le \sigma_n(1 + 1/n) \quad \Longrightarrow \quad \sigma_n \ge 0.$$

By choosing $\pi = \pi^*$, $\delta = 0$, and $z = \pi^* + \zeta$ with $\zeta \geq 0$, one obtains

$$(\varrho_n, \zeta)_{L^2(\Omega)} \geq -\sigma_n/n \quad \forall \zeta \in L^2(\Omega) : \zeta \geq 0.$$

To show the sign condition on ϱ_n, assume the contrary, i.e., $\varrho_n < 0$ a.e. in $E \subset \Omega$ with $|E| > 0$. Then we choose $\zeta = m\,\chi_E$, $m \in \mathbb{N}$, such that

$$-\sigma_n/n \leq m \int_E \varrho_n \,\mathrm{d}\lambda \to -\infty \quad \text{for } m \to \infty \quad \Longrightarrow \quad \varrho_n \geq 0 \text{ a.e. in } \Omega.$$

Next, we show that $\sigma_n > 0$. To this end, assume $\sigma_n = 0$ so that (31) becomes

$$0 < \langle \varrho_n, -\pi + z \rangle_{L^2(\Omega)} + \langle \alpha_{1,n}, P_1\pi - \mu_1 \rangle_{L^2(\Omega_1)} + \langle \alpha_{2,n}, P_2\pi - \mu_2 \rangle_{L^2(\Omega_2)} \\ \forall\, (\pi, z) \in L^2(\Omega)^2 : z \geq 0. \tag{32}$$

Now, define

$$\hat{\pi}(x_1, x_2) := \mu_1(x_1)\, \mu_2(x_2).$$

Then, according to our hypotheses on μ_i in Assumption 1,

$$\hat{\pi} \geq 0 \quad \text{a.e. in } \Omega, \quad P_i\hat{\pi} = \mu_i,\ i = 1, 2.$$

Thus, if we choose $z = 0$ and $\pi = \hat{\pi}$ in (32), then we obtain the desired contradiction:

$$0 < -\int_\Omega \underbrace{\varrho_n(x_1, x_2)}_{\geq 0}\, \underbrace{\hat{\pi}(x_1, x_2)}_{\geq 0}\, \mathrm{d}\lambda \leq 0.$$

Thus, the assumption $\sigma_n = 0$ cannot hold, and we may assume w.l.o.g. that $\sigma_n = 1$ (otherwise argue with $\tilde{\varrho}_n := \varrho_n/\sigma_n$ and $\tilde{\alpha}_{1,n}$, $\tilde{\alpha}_{2,n}$ analogously).

Therefore, (31) becomes with $\delta = 0$ and $z = 0$

$$E_\gamma(\pi^*) - \frac{1}{n} < E_\gamma(\pi) - \langle \varrho_n, \pi \rangle_{L^2(\Omega)} \\ + \langle \alpha_{1,n}, P_1\pi - \mu_1 \rangle_{L^2(\Omega_1)} + \langle \alpha_{2,n}, P_2\pi - \mu_2 \rangle_{L^2(\Omega_2)} \tag{33}$$

for all $\pi \in L^2(\Omega)$ and thus

$$E_\gamma(\pi^*) - \frac{1}{n} \leq \sup_{(\alpha_1, \alpha_2, \varrho) \in \mathcal{F}_\mathrm{D}} \inf_{\pi \in L^2(\Omega)} \mathcal{L}(\pi, \alpha_1, \alpha_2, \varrho),$$

where $\mathcal{F}_\mathrm{D}$ is the feasible set of the dual problem from (4). Since $n \in \mathbb{N}$ was arbitrary, this gives the result. ◻

Corollary 2.13 *Provided that Assumption 1 holds, there is no duality gap, i.e.,*

$$\inf(P) = \sup(D). \tag{34}$$

Proof Based on Proposition 2.12, the arguments are standard. Since weak duality always holds and π^* solves the primal problem, we have

$$\sup(D) \leq \inf(P) = E_\gamma(\pi^*),$$

which, together with (30), yields the assertion. ◻

Now, we are finally in the position to state the main result of this section concerning first-order necessary and sufficient optimality conditions for problem (1) involving Lagrange multipliers:

Theorem 2.14 *A function $\pi^* \in L^2(\Omega)$ is a solution of (1) if and only if there exist Lagrange multipliers $\alpha_1^* \in L^2(\Omega_1)$ and $\alpha_2^* \in L^2(\Omega_2)$, which solve the dual problem, so that*

$$\pi^* - \tfrac{1}{\gamma}\left(\alpha_1^* \oplus \alpha_2^* - c\right)_+ = 0 \qquad \lambda\text{-a.e. in } \Omega, \tag{35a}$$

$$\int_{\Omega_2} \left(\alpha_1^* \oplus \alpha_2^* - c\right)_+ \mathrm{d}\lambda_2 = \gamma\mu_1 \quad \lambda_1\text{-a.e. in } \Omega_1, \tag{35b}$$

$$\int_{\Omega_1} \left(\alpha_1^* \oplus \alpha_2^* - c\right)_+ \mathrm{d}\lambda_1 = \gamma\mu_2 \quad \lambda_2\text{-a.e. in } \Omega_2. \tag{35c}$$

Proof Based on the previous results, the arguments are standard. First of all, we know from (8) that, if (α_1^*, α_2^*) solves (D), then $(\alpha_1^*, \alpha_2^*, \rho^*)$ with $\rho^* = -(\alpha_1^* \oplus \alpha_2^* - c)_- \in L^2(\Omega)$ is a solution of (7) and (DP), respectively. Now, assume that π^* solves (1) and equivalently (PP). Since we know that dual problem admits a solution and there is no duality gap, $(\pi^*, \alpha_1^*, \alpha_2^*, \rho^*)$ forms a saddle point of the Lagrangian and thus satisfies the Karush-Kuhn-Tucker conditions, i.e.,

$$\partial_\pi \mathcal{L}(\pi^*, \alpha_1^*, \alpha_2^*, \rho^*) = 0, \tag{36a}$$

$$\rho \geq 0, \quad \rho\,\pi = 0, \quad \pi \geq 0, \tag{36b}$$

$$\int_{\Omega_2} \pi \,\mathrm{d}\lambda_2 = \mu_1, \qquad \int_{\Omega_1} \pi \,\mathrm{d}\lambda_1 = \mu_2. \tag{36c}$$

Resolving the complementarity system (36b) for ρ by means of the max-function and inserting this into the gradient equation in (36a) gives (35).

On the other hand, if $(\pi^*, \alpha_1^*, \alpha_2^*)$ satisfies (35), then it is easily seen that $(\pi^*, \alpha_1^*, \alpha_2^*, \rho^*)$ with $\rho^* = -(\alpha_1^* \oplus \alpha_2^* - c)_-$ fulfills the KKT-system (36). Thus, since the problem is convex, it is a saddle point of the Lagrangian and hence, π^* is a global minimizer of the primal problem (1). □

The significance of Theorem 2.14 lies in the fact that we can characterize optimality of π by just two equalities in $L^2(\Omega_1)$ and $L^2(\Omega_2)$, respectively, namely (35b) and (35c). Thus, we effectively reduce the size of the problem from searching one function on $\Omega = \Omega_1 \times \Omega_2$ to searching two functions, one on Ω_1 and one on Ω_2 (similarly as for entropic regularization, cf. [4]). This will be exploited numerically in Section 3.

2.3 Regularization of the dual problem

As seen before, the dual problem in (D) is not uniquely solvable. One source of non-uniqueness is of course the kernel of the map $(\alpha_1, \alpha_2) \mapsto \alpha_1 \oplus \alpha_2$. This kernel is one-dimensional and is spanned by the function $(1, -1)$, which could be easily taken into account in an algorithmic framework. However, there is another source of non-uniqueness due to the max-operator that cuts of the negative part. For instance, if $\Omega_1 = \Omega_2 = [0,1]$, $\mu_1 = \mu_2 \equiv 1$, and

$$c(x,y) := \begin{cases} C, & \text{if } \frac{1}{2} \le x \le 1,\ \frac{1}{2} \le y \le 1, \\ 0, & \text{else}, \end{cases} \qquad \text{with} \quad C > 4,$$

then a straight forward calculation shows that, for every $\delta \in [0, \frac{C-4}{2}]$, the tuple

$$\alpha_1^*(x) = \begin{cases} 1+\delta, & \text{if } x \in [0,\frac{1}{2}), \\ -1-\delta, & \text{if } x \in [\frac{1}{2}, 1], \end{cases} \qquad \alpha_2^*(y) = \begin{cases} 3+\delta, & \text{if } y \in [0,\frac{1}{2}), \\ 1-\delta, & \text{if } y \in [\frac{1}{2}, 1], \end{cases}$$

solves the optimality system (35b)–(35c). This shows that the potential structure of non-uniqueness might become fairly intricate. Therefore, we investigate the following regularization of the dual problem:

$$\left.\begin{array}{ll} \min & \Phi_\varepsilon(\alpha_1,\alpha_2) := \Phi(\alpha_1,\alpha_2) + \frac{\varepsilon}{2}\big(\|\alpha_1\|^2_{L^2(\Omega_1)} + \|\alpha_2\|^2_{L^2(\Omega_2)}\big) \\ \text{s.t.} & \alpha_i \in L^2(\Omega_i),\ i = 1,2, \end{array}\right\} \qquad (\mathrm{D}_\varepsilon)$$

with a regularization parameter $\varepsilon > 0$. It is clear that the additional quadratic terms in the regularized objective Φ_ε yield that the latter is strictly convex and coercive in $L^2(\Omega_1) \times L^2(\Omega_2)$. Therefore, for every $\varepsilon > 0$, (D_ε) admits a unique solution.

Proposition 2.15 *Let $\{\varepsilon_n\} \subset \mathbb{R}^+$ be a sequence converging to zero and denote the solutions of* (D_ε) *with $\varepsilon = \varepsilon_n$ by $(\alpha_1^n, \alpha_2^n) \in L^2(\Omega_1) \times L^2(\Omega_2)$. Then the sequence $\{(\alpha_1^n, \alpha_2^n)\}$ admits a weak accumulation point. Furthermore, every weak accumulation point is automatically a strong one and a solution of the original dual problem* (D).

Proof Let $(\alpha_1^*, \alpha_2^*) \in L^2(\Omega_1) \times L^2(\Omega_2)$ denote an arbitrary globally optimal solution of (D) (whose existence is guaranteed by Theorem 2.11). Then the optimality of (α_1^*, α_2^*) for (D) and of (α_1^n, α_2^n) for (D_ε) (with $\varepsilon = \varepsilon_n$) gives

$$\begin{aligned} &\Phi(\alpha_1^*, \alpha_2^*) + \tfrac{\varepsilon_n}{2}\big(\|\alpha_1^n\|^2_{L^2(\Omega_1)} + \|\alpha_2^n\|^2_{L^2(\Omega_2)}\big) \\ &\quad \le \Phi(\alpha_1^n, \alpha_2^n) + \tfrac{\varepsilon_n}{2}\big(\|\alpha_1^n\|^2_{L^2(\Omega_1)} + \|\alpha_2^n\|^2_{L^2(\Omega_2)}\big) = \Phi_{\varepsilon_n}(\alpha_1^n, \alpha_2^n) \le \Phi_{\varepsilon_n}(\alpha_1^*, \alpha_2^*). \end{aligned}$$

Taking the precise structure of Φ_ε into account, this implies

$$\|\alpha_1^n\|^2_{L^2(\Omega_1)} + \|\alpha_2^n\|^2_{L^2(\Omega_2)} \le \|\alpha_1^*\|^2_{L^2(\Omega_1)} + \|\alpha_2^*\|^2_{L^2(\Omega_2)} \qquad (37)$$

and thus, the boundedness of $\{(\alpha_1^n, \alpha_2^n)\}$ in $L^2(\Omega_1) \times L^2(\Omega_2)$. This in turn gives the existence of a weak accumulation point as claimed.

Now assume that $(\tilde\alpha_1, \tilde\alpha_2)$ is such a weak accumulation point, i.e., there is a subsequence, to simplify the notation denoted by the same symbol, such that

$$(\alpha_1^n, \alpha_2^n) \rightharpoonup (\tilde\alpha_1, \tilde\alpha_2) \quad \text{in } L^2(\Omega_1) \times L^2(\Omega_2). \qquad (38)$$

Using again the optimality of (α_1^*, α_2^*) and (α_1^n, α_2^n) for their respective optimization problems, we obtain

$$\Phi(\alpha_1^*, \alpha_2^*) \le \Phi(\alpha_1^n, \alpha_2^n) \le \Phi_{\varepsilon_n}(\alpha_1^n, \alpha_2^n) \le \Phi_{\varepsilon_n}(\alpha_1^*, \alpha_2^*) \to \Phi(\alpha_1^*, \alpha_2^*) \qquad (39)$$

as $n \to \infty$. On the other hand, Φ is convex and continuous and therefore lower semicontinuous w.r.t. weak convergence in $L^2(\Omega_1) \times L^2(\Omega_2)$ so that (38) and (39) lead to

$$\Phi(\tilde{\alpha}_1, \tilde{\alpha}_2) \leq \liminf_{n\to\infty} \Phi(\alpha_1^n, \alpha_2^n) = \lim_{n\to\infty} \Phi(\alpha_1^n, \alpha_2^n) = \Phi(\alpha_1^*, \alpha_2^*),$$

which gives in turn the optimality of the weak limit.

As (α_1^*, α_2^*) was chosen arbitrarily among the optimal solutions, the estimate (37) has to hold for the choice $(\alpha_1^*, \alpha_2^*) = (\tilde{\alpha}_1, \tilde{\alpha}_2)$, i.e.

$$\|\alpha_1^n\|^2_{L^2(\Omega_1)} + \|\alpha_2^n\|^2_{L^2(\Omega_2)} \leq \|\tilde{\alpha}_1\|^2_{L^2(\Omega_1)} + \|\tilde{\alpha}_2\|^2_{L^2(\Omega_2)}$$

for all $n \in \mathbb{N}$. Thus, we have

$$\liminf_{n\to\infty} \|\alpha_1^n\|^2_{L^2(\Omega_1)} + \|\alpha_2^n\|^2_{L^2(\Omega_2)} \leq \|\tilde{\alpha}_1\|^2_{L^2(\Omega_1)} + \|\tilde{\alpha}_2\|^2_{L^2(\Omega_2)},$$

but if $(\alpha_1^n, \alpha_2^n) \nrightarrow (\tilde{\alpha}_1, \tilde{\alpha}_2)$, we would have

$$\|\tilde{\alpha}_1\|^2_{L^2(\Omega_1)} + \|\tilde{\alpha}_2\|^2_{L^2(\Omega_2)} < \liminf_{n\to\infty} \|\alpha_1^n\|^2_{L^2(\Omega_1)} + \|\alpha_2^n\|^2_{L^2(\Omega_2)}$$

by $(\alpha_1^n, \alpha_2^n) \rightharpoonup (\tilde{\alpha}_1, \tilde{\alpha}_2)$ and consequently, we have $(\alpha_1^n, \alpha_2^n) \to (\tilde{\alpha}_1, \tilde{\alpha}_2)$ in $L^2(\Omega_1) \times L^2(\Omega_2)$. □

Theorem 2.16 *Let $\{\varepsilon_n\} \subset \mathbb{R}^+$ be a sequence converging to zero and denote the solutions of* (D_ε) *with $\varepsilon = \varepsilon_n$ again by $(\alpha_1^n, \alpha_2^n) \in L^2(\Omega_1) \times L^2(\Omega_2)$. Moreover, define*

$$\pi_n := \tfrac{1}{\gamma}(\alpha_1^n \oplus \alpha_2^n - c)_+. \tag{40}$$

Then π_n converges strongly in $L^2(\Omega)$ to the unique solution of (1).

Proof From (37), we know that $\{(\alpha_1^n, \alpha_2^n)\}$ is bounded in $L^2(\Omega_1) \times L^2(\Omega_2)$. Consequently, the same holds for $\{\pi_n\}$ as a sequence in $L^2(\Omega)$ by its definition in (40) and the Lipschitz continuity of the max-function. Hence, there is a weakly converging subsequence, which we denote by the same symbol to ease notation, i.e.,

$$\pi_n \rightharpoonup \tilde{\pi} \quad \text{in } L^2(\Omega). \tag{41}$$

In the following, we show that $\tilde{\pi}$ is the optimal solution of (1). We start with its feasibility. Since the set $\{\pi \in L^2(\Omega) : \pi(x_1, x_2) \geq 0 \text{ a.e. in } \Omega\}$ is clearly convex and closed thus weakly closed, $\tilde{\pi}$ satisfies the inequality constraint in (1). The equality constraints can be derived from the necessary optimality conditions associated with (D_ε) as follows: The first-order necessary and sufficient optimality conditions for (D_ε) read

$$\int_{\Omega_2} (\alpha_1^n \oplus \alpha_2^n - c)_+ \,\mathrm{d}\lambda_2 + \varepsilon\, \alpha_1^n = \gamma\, \mu_1 \quad \lambda_1\text{-a.e. in } \Omega_1 \tag{42}$$

$$\int_{\Omega_1} (\alpha_1^n \oplus \alpha_2^n - c)_+ \,\mathrm{d}\lambda_1 + \varepsilon\, \alpha_2^n = \gamma\, \mu_2 \quad \lambda_2\text{-a.e. in } \Omega_2. \tag{43}$$

Testing the first equation with an arbitrary $\varphi_1 \in C_c^\infty(\Omega_1)$, inserting the definition of π_n, and integrating over Ω_1 yields

$$\int_{\Omega_1}\int_{\Omega_2} \pi_n \,\mathrm{d}\lambda_2\, \varphi_1 \,\mathrm{d}\lambda_1 = \int_{\Omega_1} \mu_1\, \varphi_1 \,\mathrm{d}\lambda_1 + \frac{\varepsilon_n}{\gamma} \int_{\Omega_1} \alpha_1^n\, \varphi_1 \,\mathrm{d}\lambda_1$$

Thanks to the weak convergence in (41), the boundedness of $\{\alpha_1^n\}$ in $L^2(\Omega_1)$, and $\varepsilon_n \searrow 0$, we obtain in the limit

$$\int_{\Omega_1}\int_{\Omega_2} \tilde{\pi}\,\mathrm{d}\lambda_2\,\varphi_1\,\mathrm{d}\lambda_1 = \int_{\Omega_1} \mu_1\,\varphi_1\,\mathrm{d}\lambda_1,$$

and, since $\varphi_1 \in C_c^\infty(\Omega_1)$ was arbitrary, $\tilde{\pi}$ satisfies the first equality constraint in (1). The second equality constraint can be verified completely analogously so that the weak limit $\tilde{\pi}$ is indeed feasible for (1).

To show its optimality, let us consider the regularized objective Φ_ε. By the definition of π_n in (40) and the optimality conditions in (42) and (43), tested with α_1^n and α_2^n, respectively, we find

$$\begin{aligned}
\Phi_{\varepsilon_n}(\alpha_1^n,\alpha_2^n) &= \tfrac{1}{2}\|(\alpha_1^n \oplus \alpha_2^n - c)_+\|_{L^2(\Omega)}^2 - \gamma\int_{\Omega_1}\mu_1\alpha_1^n\,\mathrm{d}\lambda_1 - \gamma\int_{\Omega_2}\mu_2\alpha_2^n\,\mathrm{d}\lambda_2 \\
&\qquad + \tfrac{\varepsilon_n}{2}\|\alpha_1^n\|_{L^2(\Omega_1)}^2 + \tfrac{\varepsilon_n}{2}\|\alpha_2^n\|_{L^2(\Omega_2)}^2 \\
&= \tfrac{\gamma^2}{2}\|\pi_n\|_{L^2(\Omega)}^2 - \gamma\int_{\Omega}\pi_n(\alpha_1^n \oplus \alpha_2^n)\,\mathrm{d}\lambda - \tfrac{\varepsilon_n}{2}\|\alpha_1^n\|_{L^2(\Omega_1)}^2 - \tfrac{\varepsilon_n}{2}\|\alpha_2^n\|_{L^2(\Omega_2)}^2 \\
&= -\tfrac{\gamma^2}{2}\|\pi_n\|_{L^2(\Omega)}^2 - \gamma\int_{\Omega}c\,\pi_n\,\mathrm{d}\lambda - \tfrac{\varepsilon_n}{2}\|\alpha_1^n\|_{L^2(\Omega_1)}^2 - \tfrac{\varepsilon_n}{2}\|\alpha_2^n\|_{L^2(\Omega_2)}^2 \\
&= -\gamma E_\gamma(\pi_n) - \tfrac{\varepsilon_n}{2}\|\alpha_1^n\|_{L^2(\Omega_1)}^2 - \tfrac{\varepsilon_n}{2}\|\alpha_2^n\|_{L^2(\Omega_2)}^2,
\end{aligned}$$

where E_γ is the primal objective from (3). Completely analogously, one shows for the original objective Φ by means of the optimality system (35) that

$$\Phi(\alpha_1^*,\alpha_2^*) = -\gamma\,E_\gamma(\pi^*),$$

where $\pi^* \in L^2(\Omega)$ is the unique solution of (1) and $(\alpha_1^*,\alpha_2^*) \in L^2(\Omega_1)\times L^2(\Omega_2)$ solves the dual problem (D). Now, putting everything so far together, we obtain from the convergence of the regularized dual objective in (39) and the boundedness of $\{\alpha_i^n\}$, $i=1,2$, that

$$\begin{aligned}
\lim_{n\to\infty} E_\gamma(\pi_n) &= \lim_{n\to\infty}\Big(-\tfrac{1}{\gamma}\,\Phi_{\varepsilon_n}(\alpha_1^n,\alpha_2^n) - \tfrac{\varepsilon_n}{2\gamma}\|\alpha_1^n\|_{L^2(\Omega_1)}^2 - \tfrac{\varepsilon_n}{2\gamma}\|\alpha_2^n\|_{L^2(\Omega_2)}^2\Big) \\
&= -\tfrac{1}{\gamma}\,\Phi(\alpha_1^*,\alpha_2^*) = E_\gamma(\pi^*).
\end{aligned}$$

On the other hand, E_γ is convex and continuous, thus weakly lower semicontinuous, and therefore

$$E_\gamma(\tilde{\pi}) \le \liminf_{n\to\infty} E_\gamma(\pi_n) = E_\gamma(\pi^*).$$

Since $\tilde{\pi}$ is feasible, as seen above, this gives the optimality of $\tilde{\pi}$. As the objective of (1) is strictly convex, the solution of (1) is unique so that $\tilde{\pi} = \pi^*$. Thus, the weak limit is unique and a well known argument by contradiction therefore implies the weak convergence of the whole sequence $\{\pi_n\}$ to π^*.

To show strong convergence, assume the contrary, i.e., there is a subsequence $\{\pi_{n_k}\}_{k\in\mathbb{N}}$ and $\delta > 0$ so that $\|\pi_{n_k} - \pi^*\|_{L^2(\Omega)} \ge \delta$ for all $k \in \mathbb{N}$. According to Proposition 2.15, the associated sequence $\{(\alpha_1^{n_k},\alpha_2^{n_k})\}$ admits a weak accumulation point, which is also a strong one and a solution of the dual problem, i.e., there is a subsequence $\{(\alpha_1^{n_{k_\ell}},\alpha_2^{n_{k_\ell}})\}_{\ell\in\mathbb{N}}$ such that, by the uniqueness of π^*,

$$\pi_{n_{k_\ell}} = \tfrac{1}{\gamma}(\alpha_1^{n_{k_\ell}} \oplus \alpha_2^{n_{k_\ell}} - c)_+ \to \tfrac{1}{\gamma}(\alpha_1^* \oplus \alpha_2^* - c)_+ = \pi^* \quad \text{as } \ell\to\infty,$$

which gives the desired contradiction. □

2.4 The discrete dual problem

We show a simple discretization of the quadratically regularized optimal transport problem (1) by piecewise constant approximation in Appendix A. To keep the notation concise, we state the corresponding discrete optimal transport problem and illustrate the duality already here. This will be the basis of our algorithms we derive in Section 3. A discrete version of the continuous problem (1) is the finite-dimensional problem

$$\min_{\pi\in\mathbb{R}^{M\times N}} \langle \pi, c\rangle + \tfrac{\gamma}{2}\|\pi\|_F^2 \quad \text{s.t.} \quad \pi^T\mathbf{1}_M = \mu^+,\ \pi\mathbf{1}_N = \mu^-,\ \pi \geq 0 \tag{44}$$

where $\mathbf{1}_N \in \mathbb{R}^N$ denotes the vector of all ones, $\mu^+ \in \mathbb{R}^N$, $\mu^- \in \mathbb{R}^M$ denote the discretized marginals with $\sum_{j=1}^N \mu_j^+ = \sum_{i=1}^M \mu_j^-$, and $c \in \mathbb{R}^{M\times N}$ denotes the discretized cost. For the discrete form of the optimality system (35) we slightly changed notation from $\mu_{1/2}$ to $\mu^\pm$ and we also replace the Lagrange multipliers α_1 and α_2 by α and β and get

$$\pi = \tfrac{1}{\gamma}\left(\alpha \oplus \beta - c\right)_+ \tag{45a}$$

$$\sum_{i=1}^M \left(\alpha_i + \beta_j - c_{ij}\right)_+ = \gamma\mu_j^+,\ j = 1,\ldots,N \tag{45b}$$

$$\sum_{j=1}^N \left(\alpha_i + \beta_j - c_{ij}\right)_+ = \gamma\mu_i^-,\ i = 1,\ldots,M \tag{45c}$$

where $\alpha \in \mathbb{R}^M$, $\beta \in \mathbb{R}^N$ and $(\alpha \oplus \beta)_{i,j} = \alpha_i + \beta_j$ is the “outer sum”. The discrete counterpart of Φ from (D) is

$$\Phi(\alpha,\beta) = \tfrac{1}{2}\|(\alpha\oplus\beta - c)_+\|_F^2 - \gamma\langle\mu^-,\alpha\rangle - \gamma\langle\mu^+,\beta\rangle$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

We write the optimality condition (45b)-(45c) as a non-smooth equation $F(\alpha,\beta) = 0$ in $\mathbb{R}^{M+N}$ with

$$F(\alpha,\beta) = \begin{pmatrix} F_1(\alpha,\beta) \\ F_2(\alpha,\beta)\end{pmatrix} = \begin{pmatrix} (\sum_{j=1}^N \left(\alpha_i+\beta_j - c_{ij}\right)_+ - \gamma\mu_i^-)_{i=1,\ldots,M} \\ (\sum_{i=1}^M \left(\alpha_i+\beta_j - c_{ij}\right)_+ - \gamma\mu_j^+)_{j=1,\ldots,N}\end{pmatrix} \tag{46}$$

(note that $F_1 = \partial_\alpha\Phi$ and $F_2 = \partial_\beta\Phi$).

Since F is the composition of Lipschitz continuous and semismooth functions, we have the following result (for the chain rule for semismooth functions, see e.g. [13, Thm. 2.10]):

Lemma 2.17 *The function F (and thus, the gradient of Φ) is (globally) Lipschitz continuous and semismooth.*

3 Algorithms

The optimality system (45b), (45c) for the smooth and convex problem (D) can be solved by different methods. In [3] the authors propose to use a generic L-BFGS solver and also derive an alternating minimization scheme, which is similar to the non-linear Gauss-Seidel method in the next section, but differs slightly in the numerical realization and [19] also uses an off-the-shelf solver. Here we propose methods that exploit the special structure of the optimality system: A non-linear Gauss-Seidel method and a semismooth Newton method.

3.1 Non-linear Gauss-Seidel

The method in this section is similar to the one described in the Appendix of [3], but we describe it here for the sake of completeness. A close look at the optimality system

$$\sum_{j=1}^{N}(\alpha_i+\beta_j-c_{ij})_+ = \gamma\mu_i^-,\quad i=1,\dots,M. \tag{47a}$$

$$\sum_{i=1}^{M}(\alpha_i+\beta_j-c_{ij})_+ = \gamma\mu_j^+,\quad j=1,\dots,N \tag{47b}$$

shows that we can solve all M equations in (47a) for the α_i in parallel (for fixed β) since the ith equation depends on α_i only. Similarly, all N equations in (47b) can be solved for the β_j if α is fixed. Hence, we can perform a non-linear Gauss-Seidel method for these non-smooth equations (also known as alternating minimization, nonlinear SOR or coordinate descent method for Φ [6, 24]), i.e. alternatingly solving the equations (47a) for α (for fixed β) and then the equations (47b) for β (for fixed α). The whole method is stated in Algorithm 1. Since Φ is convex with Lipschitz continuous gradient (cf. Lemma 2.17) the convergence of the algorithm follows from results in [2].

Algorithm 1 Non-linear Gauss-Seidel for quadratically regularized optimal transport

```
Initialize: β^0 ∈ ℝ^N, set k = 0
repeat
    Set α^{k+1} to be the solution of (47a) with β = β^k.
    Set β^{k+1} to be the solution of (47b) with α = α^{k+1}.
    k ← k + 1
until some stopping criterion
```

Each equation for an α_i or β_j is just a single scalar equation for a scalar quantity and the structure of the equation is of the following form: For a given vector $y\in\mathbb{R}^n$ and right hand side $b\in\mathbb{R}$, solve

$$f(x) := \sum_{j=1}^{n}(x-y_j)_+ = b. \tag{48}$$

Of course, one can solve this problem by bisection, but here are two other, more efficient methods to solve equations of the type (48):

Direct search. If we denote by $y_{[j]}$ the j-th smallest entry of y (i.e. we sort y in an ascending way), we get that

$$f(x) = \sum_{j=1}^{n} (x - y_{[j]})_+$$
$$= \begin{cases} 0, & x \le y_{[1]} \\ kx - \sum_{j=1}^{k} y_{[j]}, & y_{[k]} \le x \le y_{[k+1]},\ k = 1, \dots, n-1 \\ nx - \sum_{j=1}^{n} y_{[j]}, & x \ge y_{[n]}. \end{cases}$$

To obtain the solution of (48) we evaluate f at the break points $y_{[j]}$ until we find the interval $[y_{[k]}, y_{[k+1]}[$ in which the solution lies (by finding k such that $f(y_{[k]}) \le b < f(y_{[k+1]})$), and then setting

$$x = \frac{b + \sum_{j=1}^{k} y_{[j]}}{k}.$$

The complexity of the method is dominated by the sorting of the vector y, its complexity is $\mathcal{O}(n \log(n))$.

Semismooth Newton. Although f is non-smooth, we may perform Newton's method here. The function f is piecewise linear and on each interval $]y_{[j]}, y_{[j+1]}[$ is has the slope j (a simple situation with $n = 3$ is shown in Figure 1). At the break points we may define $f'(y_{[j]}) = j$ and then we iterate

$$x^{k+1} = x^k - \tfrac{f(x^k)}{f'(x^k)}.$$

If we start with $x^0 \ge y^{[n]} = \max_k y_k$, the method will produce a monotonically decreasing sequence which converges in at most n steps. Actually, we can initialize the method with any x^0 that is strictly larger than $y_{[1]} = \min_k y_k$. Note that we do not need to sort the values of y_k to calculate the derivative since we have $f'(x) = \#\{i \ : \ x \ge y_i\}$. In practice, the method usually needs much less iterations than n.

3.2 Semismooth Newton

As seen in Lemma 2.17, the mapping F is semismooth and hence, we may use a semismooth Newton method [5, 7].

A simple calculation proves the following lemma.

Lemma 3.1 *A Newton derivative of F from (46) at (α, β) is given by*

$$G = \begin{pmatrix} \operatorname{diag}(\sigma \mathbf{1}_N) & \sigma \\ \sigma^T & \operatorname{diag}(\sigma^T \mathbf{1}_M) \end{pmatrix} \in \mathbb{R}^{(M+N)\times(M+N)}$$

where $\sigma \in \mathbb{R}^{M\times N}$ is given by

$$\sigma_{ij} = \begin{cases} 1 & \alpha_i + \beta_j - c_{ij} \ge 0 \\ 0 & \textit{otherwise.} \end{cases}$$

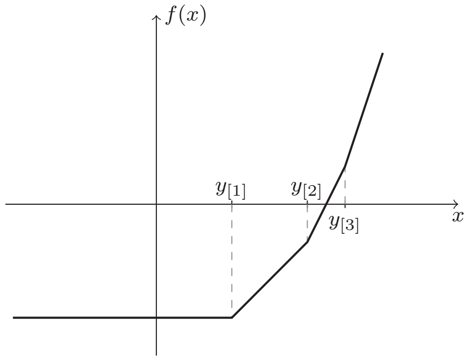


Fig. 1 Illustration of the non-smooth function f from (48).

A step of the semismooth Newton method for the solution of $F(\alpha,\beta)=0$ would consist of setting

$$\begin{pmatrix}\alpha^{k+1}\\ \beta^{k+1}\end{pmatrix}=\begin{pmatrix}\alpha^{k}\\ \beta^{k}\end{pmatrix}-\begin{pmatrix}\delta\alpha^{k}\\ \delta\beta^{k}\end{pmatrix}\quad\text{where}\quad F(\alpha^k,\beta^k)=G\begin{pmatrix}\delta\alpha^{k}\\ \delta\beta^{k}\end{pmatrix}.$$

However, the next lemma shows, that G has a non-trivial kernel.

Lemma 3.2 *Let G be the Newton derivative of F at (α,β) defined in Lemma 3.1. Then the following holds true:*

1. *$G\in\mathbb{R}^{(M+N)\times(M+N)}$ is symmetric,*
2. *G is positive semi-definite,*
3. *$(a,b)\in\operatorname{kern}(G)$ if and only if $\sigma_{ij}(a_i+b_j)=0$ for all $1\le i\le M$, $1\le j\le N$.*

Proof Symmetry of G is clear by construction. To see that G is positive semi-definite we calculate

$$\begin{aligned}(a,b)^\top G(a,b)&=\sum_{j=1}^N\sum_{i=1}^M\sigma_{ij}a_i^2+\sum_{j=1}^N\sum_{i=1}^M\sigma_{ij}b_j^2+2\sum_{j=1}^N\sum_{i=1}^M\sigma_{ij}a_ib_j\\&=\sum_{j=1}^N\sum_{i=1}^M\sigma_{ij}(a_i+b_j)^2\ge 0.\end{aligned}$$

Due to the non-negativity of σ, this also shows the last point. □

The third point of the lemma shows that the kernel of G may have a high dimension, depending on the matrix σ. Hence we resort to a quasi Newton method where we regularize the Newton step arising from the dual problem from Section 2.2 by setting

$$\begin{pmatrix}\alpha^{k+1}\\ \beta^{k+1}\end{pmatrix}=\begin{pmatrix}\alpha^{k}\\ \beta^{k}\end{pmatrix}-\begin{pmatrix}\delta\alpha^{k}\\ \delta\beta^{k}\end{pmatrix}\quad\text{where}\quad F(\alpha^k,\beta^k)=(G+\varepsilon I)\begin{pmatrix}\delta\alpha^{k}\\ \delta\beta^{k}\end{pmatrix}$$

with a small $\varepsilon>0$. By [5], the method still converges, but only a local linear rate is guaranteed. We note that we have not applied the semismooth Newton method to the regularized dual problem from Section 2.3. This would also be possible, but lead not only to the regularized Newton matrix from above but we would also have to adapt the objective F in the computation of the update.

Let us make a few remarks on the the regularized Newton step and its numerical treatment.

- The matrix σ (and hence the Newton matrix G) is usually very sparse. The closer α and β are to the optimal ones, the closer $(\alpha_i + \beta_j - c_{ij})_+$ is to the optimal regularized transport plan π and for small γ this usually very sparse.
- Since G is positive semi-definite, the regularized step could be done by the method of conjugate gradients. However, any linear solver that can exploit the sparsity of G can be used.

As usual, the regularized semismooth Newton method may not converge globally. A simple globalization technique is an Armijo linesearch in the Newton direction. The full method is described in Algorithm 2.

Algorithm 2 Globalized and regularized semismooth Newton method quadratically regularized optimal transport

Initialize: $\alpha^0 \in \mathbb{R}^M$, $\beta^0 \in \mathbb{R}^N$, set $k = 0$, choose regularization parameter $\varepsilon > 0$, Armijo parameters $\theta, \kappa \in]0, 1[$, and a tolerance $\tau > 0$

repeat

Calculate

$$P_{ij} = \alpha_i^k + \beta_j^k - c_{ij}, \quad \sigma_{ij} = \begin{cases} 1 & P_{ij} \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad \text{and} \quad \pi_{ij} = \max(P_{ij}, 0)/\gamma.$$

Calculate $\delta\alpha$ and $\delta\beta$ by solving

$$\left(\begin{pmatrix} \operatorname{diag}(\sigma \mathbf{1}_N) & \sigma \\ \sigma^T & \operatorname{diag}(\sigma^T \mathbf{1}_M) \end{pmatrix} + \varepsilon I\right) \begin{pmatrix} \delta\alpha \\ \delta\beta \end{pmatrix} = -\gamma \begin{pmatrix} \pi \mathbf{1}_N - \mu^- \\ \pi^T \mathbf{1}_M - \mu^+ \end{pmatrix}$$

Set $t = 1$ and compute the directional derivative

$$d = D_{(\delta\alpha, \delta\beta)} \Phi(\alpha^k, \beta^k) = \gamma \sum_{ij} \pi_{ij} (\delta\alpha_i + \delta\beta_j) - \gamma(\langle \delta\alpha, \mu^- \rangle + \langle \delta\beta, \mu^+ \rangle).$$

while $\Phi(\alpha^k + t\delta\alpha, \beta^k + t\delta\beta) \geq \Phi(\alpha^k, \beta^k) + t\theta d$ **do**

$t \leftarrow \kappa t$

end while

Set $\alpha^{k+1} = \alpha^k - t\delta\alpha$, $\beta^{k+1} = \beta^k - t\delta\beta$

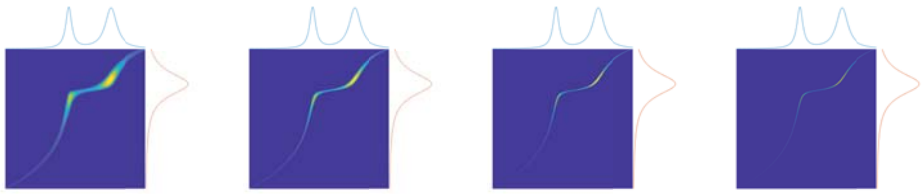
$k \leftarrow k + 1$

until $\|\pi \mathbf{1}_N - \mu^-\|_\infty, \|\pi^T \mathbf{1}_M - \mu^+\|_\infty \leq \tau$

4 Numerical examples

4.1 Illustration of $\gamma \to 0$

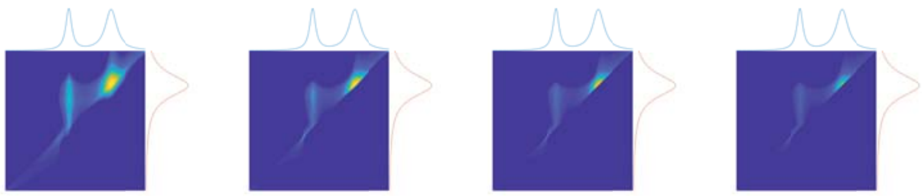
In our first numerical example we illustrate the how the solutions π^* of the regularized problem converge for vanishing regularization parameter $\gamma \to 0$. We generate some marginals, fix a transport cost and compute solutions of the discretized transport problems (44) for a sequence $\gamma_n \to 0$ and illustrate the optimal transport plans (and the related regularized transport costs). Our marginals are non-negative functions sampled at equidistant points x_i, y_i in the interval $[0,1]$ and we used $M = N = 400$ and the cost $c_{ij} = (x_i - y_j)^2$ is the squared distance between the sampling points. The results are shown in Figure 2. One observes that the optimal transport plans converge to a measure that is singular and is supported on the graph of a monotonically increasing function, exactly as the fundamental theorem of optimal transport [1] predicts.



$\gamma = 10$ $\gamma = 1$ $\gamma = 0.1$ $\gamma = 0.01$

Fig. 2 Visualization of transport plans of the quadratically regularized optimal transport problem with $M = N = 400$ and quadratic transport cost $c_{ij} = (x_i - y_j)^2$.

We repeat the same experiment where the cost is the (non-squared) distance $c_{ij} = |x_i - y_j|$. Here we had to choose larger regularization parameters as it turned out that values similar to Figure 2 would lead to almost undistinguishable results. The results are shown in Figure 3. Note the different structure of the transport plan (which is again in agreement with the predicted results from the fundamental theorem of optimal transport). In Figure 4 we show the results for the concave but increasing cost $c_{ij} = \sqrt{|x_i - y_j|}$ and again observe the expected effect that a concave transport cost encouraged that as much mass as possible stays in place (as can be seen by the concentration of mass along the diagonal of the transport plan).



$\gamma = 1{,}000$ $\gamma = 100$ $\gamma = 10$ $\gamma = 1$

Fig. 3 Visualization of transport plans of the quadratically regularized optimal transport problem with $M = N = 400$ and metric transport cost $c_{ij} = |x_i - y_j|$.

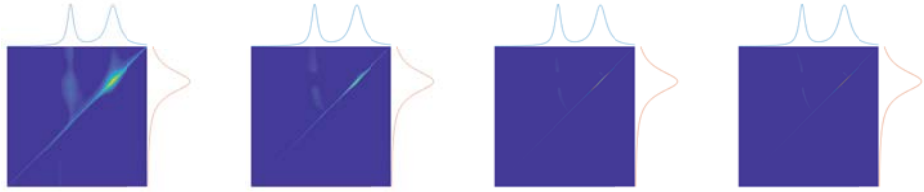


Fig. 4 Visualization of transport plans of the quadratically regularized optimal transport problem with $M = N = 400$ and concave increasing transport cost $c_{ij} = \sqrt{|x_i - y_j|}$.

4.2 Mesh independence and comparsion of SSN and NLGS

While we did not analyze our algorithms in the continuous case, we made an experiment to see how the methods converge when we change the mesh size of the discretization. To that end, we did a simple piecewise constant approximation of the marginals, the cost and the transport plan as described in Appendix A. We also took care to adapt the termination criteria so that we terminate the algorithms when the continuous counterpart of the termination criteria is satisfied (again, see Appendix A for details).

We used marginals $\mu^{\pm} : [0,1] \to [0,\infty[$ of the form

$$\mu^{+}(x) = r\tfrac{1}{1+m(x-a)^2},, \qquad \mu^{-}(x) = s\Big(\tfrac{1}{1+m_1(x-a_1)^2} + \tfrac{1}{1+m_2(x-a_2)^2}\Big)$$

with varying $m, m_1, m_2 > 0$, $0 < a, a_1, a_2 < 1$ and appropriate normalization factors r, s and quadratic cost $c(x,y) = (x-y)^2$ and discretized each instance of the problem with $M = N$ varying from 10 to $1,000$. We solved the problem for each size for regularization parameter $\gamma = 0.001$ with the semismooth Newton method from Algorithm 2 (with parameters $\epsilon = 10^{-6}$ and Armijo parameters $\kappa = 0.5$ and $\theta = 0.1$) up to tolerance 10^{-3} and report the number of iterations needed in Figure 5. As can be observed, the number of iterations is comparable for each instance of the problem. Moreover, it seems that the number of iterations does not grow with finer discretization (however, the number of iterations seems to oscillate unpredictable for coarse discretization). The would hint at mesh independence of the method and one could hope to prove this is future research. We performed a similar experiment for the nonlinear Gauss-Seidl method from Algorithm 1 (with larger regularization parameter $\gamma = 0.05$ and only up to $M = N = 500$ and show the results in Figure 6. We see an overall increase of the number of iterations but only very slightly (with several instances where the number of iterations does not increasing with finer discretization).

5 Conclusion

We analyzed the quadratically regularized optimal transport problem in Kantorovich form. While it is straight forward to derive the dual problem, our proof of existence of dual optima is quite intricate. We note that we are not aware of any

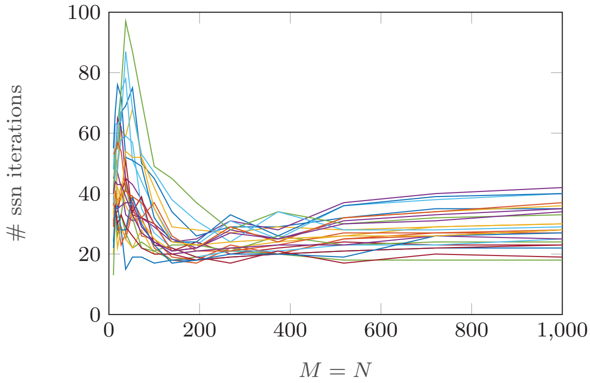


Fig. 5 Number of iteration for the semismooth Newton method to achieve a desired accuracy. Each graph corresponds to one instance of the problem.

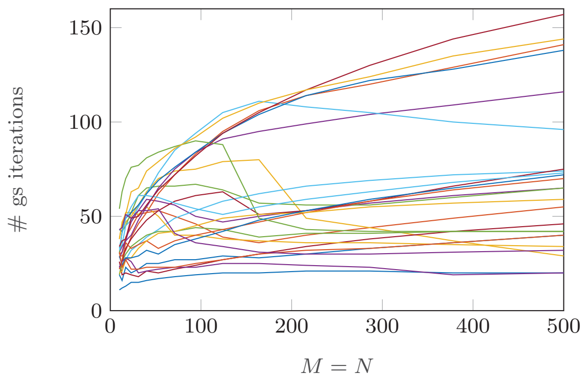


Fig. 6 Number of iteration for the nonlinear Gauss-Seidel semismooth method to achieve a desired accuracy. Each graph corresponds to one instance of the problem.

proof of existence of the dual of other regularized transport problems in the continuous case. We derived two algorithms to solve the dual problems, both of which converge by standard results. It turns out that the semismooth quasi-Newton methods converges fast in all cases and that it behaves stable with respect to the regularization parameter in our numerical experiments. We even observe mesh independence of the method in the experiments. One drawback of the semismooth Newton method is (compared with, e.g., the Sinkhorn iteration [8]), is that we need to assemble the Newton matrix in each step. While this matrix is usually very sparse, one still need to check MN cases, which may be too large for large scale problems. We did not investigate, how special structure of the cost function c may help to reduce the cost to assemble the sparse matrix σ.

A Discretization with piecewise-constant ansatz functions

For sake of brevity, we just consider an equidistant discretization of $[0,1]$ into N intervals using piecewise constant ansatz functions, i.e.

$$\pi(x,y) \coloneqq \sum_{i,j=0}^{N-1} \pi_{ij}\chi_{(\frac{i}{N},\frac{i+1}{N})\times(\frac{j}{N},\frac{j+1}{N})}(x,y),$$

for coefficients π_{ij} and assume analogous definitions for the quantities c, μ^+, μ^-, α and β. They have to coincide on average over the intervals. Again, we study this for π and obtain that the identity

$$\begin{aligned}\int_{\frac{i}{N}}^{\frac{i+1}{N}}\int_{\frac{j}{N}}^{\frac{j+1}{N}} \pi(x,y)\,\mathrm{d}y\mathrm{d}x &= \int_{\frac{i}{N}}^{\frac{i+1}{N}}\int_{\frac{j}{N}}^{\frac{j+1}{N}} \sum_{i,j=0}^{N-1} \pi_{ij}\chi_{(\frac{i}{N},\frac{i+1}{N})\times(\frac{j}{N},\frac{j+1}{N})}(x,y)\,\mathrm{d}y\mathrm{d}x \\ &= \frac{1}{N^2}\pi_{ij}\end{aligned}$$

holds. Again, analogous identities hold for the quantities c, μ^+, μ^-, α and β. The ones with one-dimensional domain are scaled by $\frac{1}{N}$ instead of $\frac{1}{N^2}$.

Now, we consider the discrete Algorithm 2, which operates on discrete quantities and establish a consistent mapping of the quantities from the discretization to the ones of the solver. We denote its input quantities by $\bar{c}_{ij}$, $\bar{\mu}_i^-$, $\bar{\mu}_i^+$ and its output quantities by $\bar{\alpha}_i$, $\bar{\beta}_j$, $\bar{pi}_{ij}$, and $\bar{E}$. It solves for

$$\sum_{j=0}^{N-1} \bar{\pi}_{ij} = \gamma\bar{\mu}_i^+,$$

which we desire to correspond to

$$\int_{\frac{i}{N}}^{\frac{i+1}{N}}\int_0^1 \pi(x,y)\,\mathrm{d}y\mathrm{d}x = \int_{\frac{i}{N}}^{\frac{i+1}{N}} \mu^+(x)\,\mathrm{d}x.$$

We plug in the ansatz functions and obtain the identity

$$\frac{1}{N^2}\sum_{j=0}^{N-1} \pi_{ij} = \frac{1}{N}\mu_i^+.$$

We set $\bar{\pi}_{ij} \coloneqq \gamma\pi_{ij}$ and obtain

$$\sum_{j=0}^{N-1} \bar{\pi}_{ij} = \sum_{j=0}^{N-1} \pi_{ij} = N\mu_i^-.$$

Thus, the choice $\bar{\mu}_i^- \coloneqq N\mu_i^-$ gives a consistent conversion. Similarly, we obtain $\bar{\mu}_j^+ \coloneqq N\mu_j^+$. We proceed with the objective. Plugging in the ansatz functions into the continuous objective gives

$$E = \frac{\gamma}{2}\frac{1}{N^2}\sum_{i,j=0}^{N-1} \pi_{ij}^2 - \frac{1}{N}\left(\sum_{i=0}^{N-1} \alpha_i\mu_i^- + \sum_{j=0}^{N-1} \beta_j\mu_j^+\right).$$

The solver computes

$$\bar{E} = \frac{1}{2}\sum_{i,j=0}^{N-1} \bar{\pi}_{ij}^2 - \gamma\left(\sum_{i=0}^{N-1} \bar{\alpha}_i\bar{\mu}_i^- + \sum_{j=0}^{N-1} \bar{\beta}_j\bar{\mu}_j^+\right),$$

Plugging in $N\mu_i^- = \bar{\mu}_i^-$, $N\mu_j^+ = \bar{\mu}_j^+$ and $\gamma\pi_{ij} = \bar{\pi}_{ij}$ gives

$$\bar{E} = \gamma^2 \frac{1}{2} \sum_{i,j=0}^{N-1} \pi_{ij}^2 - \gamma N \left(\sum_{i=0}^{N-1} \bar{\alpha}_i \mu_i^- + \sum_{j=0}^{N-1} \bar{\beta}_j \mu_j^+ \right).$$

Thus, the consistent identity $E = \frac{1}{\gamma N^2} \bar{E}$ follows if we choose $\bar{\alpha}_i \coloneqq \alpha_i$ and $\bar{\beta}_i \coloneqq \beta_i$. The solver computes $\bar{\alpha}_i$ as the solution of

$$\sum_{j=0}^{N-1} (\bar{\alpha}_i + \bar{\beta}_j - \bar{c}_{ij})_+ = \gamma \bar{\mu}_i^-,$$

whereas the discretization of the corresponding continuous equation reads

$$\frac{1}{N} \sum_{j=0}^{N-1} (\alpha_i + \beta_j - c_{ij})_+ = \gamma \mu_i^-$$

in terms of the coefficients. Plugging in the choices $\alpha_i = \bar{\alpha}_i$, $\beta_j = \bar{\beta}_j$, $c_{ij} = \bar{c}_{ij}$ and $N\mu_i^- = \bar{\mu}_i^-$ yields equivalence of the latter equation to

$$\frac{1}{N} \sum_{j=0}^{N-1} (\bar{\alpha}_i + \bar{\beta}_j - \bar{c}_{ij})_+ = \gamma \frac{1}{N} \bar{\mu}_j^-,$$

which is equivalent to the equation that is solved by Algorithm 2. The argument for $\bar{\mu}_j^+$ is carried out analogously.

Regarding termination, the solver checks the criteria

$$\left| \frac{1}{\gamma} \sum_{j=0}^{M-1} \bar{\pi}_{ij} - \bar{\mu}_i^- \right| < \tau \quad \text{and} \quad \left| \frac{1}{\gamma} \sum_{i=0}^{M-1} \bar{\pi}_{ij} - \bar{\mu}_j^+ \right| < \tau.$$

We only consider the first and plug the identity $\gamma\pi_{ij} = \bar{\pi}_{ij}$ into it, which gives equivalence to

$$\left| \sum_{j=0}^{M-1} \pi_{ij} - N\mu_i^- \right| < \tau.$$

This in turn is equivalent to

$$N^2 \left| \sum_{j=0}^{M-1} \int_{\frac{i}{N}}^{\frac{i+1}{N}} \int_{\frac{j}{N}}^{\frac{j+1}{N}} \pi(x,y)\,\mathrm{d}y\mathrm{d}x - \int_{\frac{i}{N}}^{\frac{i+1}{N}} \mu^-(x)\,\mathrm{d}x \right| < \tau,$$

$$\left| \int_{\frac{i}{N}}^{\frac{i+1}{N}} \int_0^1 \pi(x,y)\,\mathrm{d}y - \mu^-(x)\,\mathrm{d}x \right| < \frac{\tau}{N^2}.$$

Moreover, the ansatz functions for π and μ^- are constant on $\left(\frac{i}{N}, \frac{i+1}{N}\right)$, which induces equivalence to

$$\int_{\frac{i}{N}}^{\frac{i+1}{N}} \left| \int_0^1 \pi(x,y)\,\mathrm{d}y - \mu^-(x) \right| \mathrm{d}x < \frac{\tau}{N^2}.$$

This implies that if the solver terminates, we have

$$\left\| \int_0^1 \pi(\cdot,y)\,\mathrm{d}y - \mu^-(\cdot) \right\|_{L^1((0,1))} < \frac{\tau}{N}.$$

We summarize the choices for the consistent mapping of quantities arising from the discretization to quantities the solver operates on in Table 1. Finally, we make a note on the calculation of the coefficients c_{ij} for the cost function $c(x,y) \coloneqq (x-y)^2$:

$$c_{ij} = N^2 \int_{\frac{i}{N}}^{\frac{i+1}{N}} \int_{\frac{j}{N}}^{\frac{j+1}{N}} (x-y)^2\,\mathrm{d}y\mathrm{d}x = \ldots = \frac{1}{N^2}\left((i-j)^2 + \frac{1}{6} \right).$$

Table 1 Mapping discretization quantities to solver quantities.

Coefficient	Solver Quantity	Conversion
π_{ij}	$\bar{\pi}_{ij}$	$\bar{\pi}_{ij} = \gamma\pi_{ij}$
c_{ij}	$\bar{c}_{ij}$	$\bar{c}_{ij} = c_{ij}$
μ_i^-	$\bar{\mu}_i^-$	$\bar{\mu}_i^- = N\mu_i^-$
μ_j^+	$\bar{\mu}_j^+$	$\bar{\mu}_j^+ = N\mu_j^+$
α_i	$\bar{\alpha}_i$	$\bar{\alpha}_i = \alpha_i$
β_j	$\bar{\beta}_j$	$\bar{\beta}_j = \beta_j$
J	$\bar{J}$	$\bar{J} = JN^2\gamma$
τ	$\bar{\tau}$	$\bar{\tau} = \tau N$

References

1. Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.
2. Dimitri P. Bertsekas. *Nonlinear programming*. Athena Scientific Optimization and Computation Series. Athena Scientific, Belmont, MA, third edition, 2016.
3. Mathieu Blondel, Vivien Seguy, and Antoine Rolet. Smooth and sparse optimal transport. In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 880–889, Playa Blanca, Lanzarote, Canary Islands, 09–11 Apr 2018. PMLR.
4. Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, 2017.
5. Xiaojun Chen. Superlinear convergence of smoothing quasi-newton methods for nonsmooth equations. *Journal of Computational and Applied Mathematics*, 80(1):105 – 126, 1997.
6. Xiaojun Chen. On convergence of SOR methods for nonsmooth equations. *Numer. Linear Algebra Appl.*, 9(1):81–92, 2002.
7. Xiaojun Chen, Zuhair Nashed, and Liqun Qi. Smoothing methods and semismooth methods for nondifferentiable operator equations. *SIAM J. Numer. Anal.*, 38(4):1200–1216, 2000.
8. Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.
9. Marco Cuturi and Gabriel Peyré. A smoothed dual approach for variational Wasserstein problems. *SIAM J. Imaging Sci.*, 9(1):320–343, 2016.
10. Montacer Essid and Justin Solomon. Quadratically regularized optimal transport on graphs. *SIAM Journal on Scientific Computing*, 40(4):A1961–A1986, 2018.
11. Irene Fonseca and Giovanni Leoni. *Modern methods in the calculus of variations: L^p spaces*. Springer Monographs in Mathematics. Springer, New York, 2007.
12. Aude Genevay, Gabriel Peyre, and Marco Cuturi. Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pages 1608–1617, 2018.
13. Michael Hinze, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. *Optimization with PDE constraints*, volume 23. Springer Science & Business Media, 2008.
14. Leonid V. Kantorovič. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.
15. Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. Optimal transport with proximal splitting. *SIAM J. Imaging Sci.*, 7(1):212–238, 2014.
16. Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
17. Svetlozar T. Rachev and Ludger Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.
18. Svetlozar T. Rachev and Ludger Rüschendorf. *Mass transportation problems. Vol. II.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Applications.

19. Lucas Roberts, Leo Razoumov, Lin Su, and Yuyang Wang. Gini-regularized optimal transport with an application to spatio-temporal forecasting. arXiv preprint arXiv:1712.02512, 2017.
20. Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.
21. Fredi Tröltzsch. Regular Lagrange multipliers for control problems with mixed pointwise control-state constraints. *SIAM Journal on Optimization*, 15:616–634, 2005.
22. Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
23. Cédric Villani. *Optimal transport. Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.
24. Stephen J. Wright. Coordinate descent algorithms. *Math. Program.*, 151(1, Ser. B):3–34, 2015.