Signal Processing 171 (2020) 107474

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Multi-marginal optimal transport using partial information with applications in robust localization and sensor fusion☆

Filip Elvander[a,*], Isabel Haasler[b], Andreas Jakobsson[a], Johan Karlsson[b]

[a] Department of Mathematical Statistics, Lund University, P.O. Box 118, Lund SE-221 00, Sweden
[b] Department of Mathematics, KTH Royal Institute of Technology, Sweden

ARTICLE INFO

Article history:
Received 18 May 2019
Revised 29 December 2019
Accepted 9 January 2020
Available online 11 January 2020

Keywords:
Optimal mass transport
Multi-marginal problems
Entropy regularization
Spectral estimation
Array signal processing
Sensor fusion

ABSTRACT

During recent decades, there has been a substantial development in optimal mass transport theory and methods. In this work, we consider multi-marginal problems wherein only partial information of each marginal is available, a common setup in many inverse problems in, e.g., remote sensing and imaging. By considering an entropy regularized approximation of the original transport problem, we propose an algorithm corresponding to a block-coordinate ascent of the dual problem, where Newton's algorithm is used to solve the sub-problems. In order to make this computationally tractable for large-scale settings, we utilize the tensor structure that arises in practical problems, allowing for computing projections of the multi-marginal transport plan using only matrix-vector operations of relatively small matrices. As illustrating examples, we apply the resulting method to tracking and barycenter problems in spatial spectral estimation. In particular, we show that the optimal mass transport framework allows for fusing information from different time steps, as well as from different sensor arrays, also when the sensor arrays are not jointly calibrated. Furthermore, we show that by incorporating knowledge of underlying dynamics in tracking scenarios, one may arrive at accurate spectral estimates, as well as faithful reconstructions of spectra corresponding to unobserved time points.

1. Introduction

The use of mass, or power, distributions for modeling and describing properties of signals and stochastic processes is ubiquitous in the field of signal processing, being fundamental to areas such as spectral estimation, classification, and localization [55]. Spectral representations, detailing the distribution of power over frequency, serve as compact signal descriptions and constitute the foundation for many applications, such as speech enhancement [29], non-invasive estimation [4], spectroscopy [42], as well as radar and sonar [58]. In such applications, one requires a means of comparing and quantifying distances between spectra. Popular choices include the Kullback–Leibler [35] and Itakura-Saito divergences [27], the Rényi entropy [48], and L_p-norms, where the latter is often implicitly utilized through the use of matrix norms applied to the signal covariance matrix [54,57]. As an alternative to these measures, it has also been proposed to define distances between spectra based on the concept of optimal mass transport [26]. Optimal mass transport (OMT) is concerned with the task of moving one distribution of mass into another, with the description of the movement of mass being referred to as a transport plan. The minimal total cost of rearrangement, i.e., the one associated with the optimal plan, may then be used as a measure of distance between the two distributions [60]. Historically, OMT has been widely used in economics, e.g., for planning and logistics (see [60] for an introduction and an overview of the topic). Recently, there has been a rapid development in the theory and methods for optimal mass transport, and the ideas have attracted considerable attention in various economic and engineering fields, being used, for example, for option pricing [8], color and texture transfer in image processing [20], as well as machine learning [6,50] and other signal processing tasks [34]. These developments have lead to a mature framework for OMT, with computationally efficient algorithms [17,47], and constitute a flexible modeling tool that may be used to address many problems in the areas of signal processing and systems theory.

One of the appeals of using OMT for defining distances between distributions lies in its geometric properties. Specifically, as it models transport between two mass distributions, taking place on their domain of definition, the distance is not only related to the point-

☆ This work was supported in part by eSSENCE grant 2017-4:2, Carl Trygger's foundation grant 16:208, and the Swedish Research Council grants 2015-04148 and 2014-5870.

* Corresponding author.
E-mail addresses: filip.elvander@matstat.lu.se (F. Elvander), haasler@kth.se (I. Haasler), andreas.jakobsson@matstat.lu.se (A. Jakobsson), johan.karlsson@math.kth.se (J. Karlsson).

https://doi.org/10.1016/j.sigpro.2020.107474
0165-1684/

wise differences between the distributions, but also depends on where the mass is located. As a consequence of this, OMT distances directly reflect proximity of mass on the underlying space [26]. This property becomes very attractive when used in array processing and spatial spectral estimation problems, in which the support of the spectrum directly details the location of signal emitting sources [55]. For example, OMT has been shown to produce robust and predictable results when applied to direction of arrival (DoA) and localization problems [22,23]. Also, when computing the distance between two mass distributions, the OMT problem naturally induces a way of interpolating distributions by considering the flow of mass on the underlying space [5,41]. This procedure has been shown to yield physically meaningful results, e.g., when used for modeling heat diffusion on graphs [52], as well as for interpolating and tracking speech/sound signals [30,43] and (structured) covariance matrices [24,44,62].

In this work, we consider multi-marginal OMT problems where the distributions themselves are not directly observed, but manifested through linear measurement equations. We refer to this problem as multi-marginal OMT with partial information. Several problems of interest may be expressed within such a framework, e.g., CT imaging (cf. [9,33]), ensemble estimation [15], image deblurring [36], spectral analysis [30], and radar imaging problems [23]. As motivating examples, we consider problems from radar imaging, being spatial spectral estimation problems, wherein the mass distributions, i.e., power spectra, are observable from covariance matrices corresponding to a set of sensor arrays. For such problems, typically only the co-located sensors, i.e., sensors in an array, can be processed coherently due to calibration errors between the sensor arrays. Herein, we propose an OMT framework for fusing the information provided by the array covariances in a non-coherent manner. For such scenarios, we demonstrate that multi-marginal OMT constitutes a flexible and robust tool for performing information fusion, illustrating the inherent robustness to spectral perturbations, as caused by, e.g., calibration errors, provided by the use of transport models. In addition to this, we show that the proposed OMT framework may be used to fuse information corresponding to different time points, allowing for modeling the time evolution of spatial spectra. In this case, we demonstrate how the framework allows for incorporating prior knowledge of underlying dynamics and the signal generating mechanism, providing a means of forming accurate reconstructions of target trajectories in spatial estimation problems.

In addition to the modeling framework, one of the main contributions of this work is deriving a computationally efficient method for solving multi-marginal OMT problems with partial information. As the original OMT formulation is posed in an infinite-dimensional function space, we formulate discrete approximations of this problem, allowing for practical implementations. Specifically, one of the main contributions of this work is proposing a generalization of the Sinkhorn iterations, which may be applied to OMT problems with partial information. In particular, we consider the corresponding dual problem and propose to solve this with a block-coordinate ascent method, wherein each block update is computed using Newton's method. Furthermore, for certain multi-marginal OMT problems, such as the tracking and barycenter problems, we show that the structure of the cost function may be exploited, allowing for solving high-dimensional problems with a large number of marginals. It should be stressed that even though spatial spectral estimation is here used as a motivating and illustrating example, the proposed framework applies to a large class of multi-marginal OMT problems where only partial information of the marginal distributions is available.

This work is based on several of our previous works on tracking and information fusion. In particular, the original tracking formulation is from [15]. This was further developed for radar problems in [22], with efficient computations using pairwise entropy regularization being presented in [23]. In this paper, by extending this work to the multi-marginal setting, we obtain a more flexible framework, with improved computational efficiency and numerical stability.

This paper is structured as follows. In Section 2, we provide background on the problem of optimal mass transport, multi-marginal optimal transport, and partial information. In Section 3, we present the multi-marginal OMT problem with partial information, and describe the use of this formulation in tracking and information fusion. Section 4 presents computational tools for the multi-dimensional OMT problem, which is then extended to problems with partial information. In Section 5, we derive efficient computations for the algorithms in Section 4, by exploiting structures which arise in certain multi-marginal optimal transport problems. In Section 6, we provide numerical examples, illustrating the properties of the proposed framework, whereas Section 7 concludes upon the work.

2. Background: optimal mass transport

In this section, we provide a brief background on the problem of OMT, together with extensions to settings with underlying dynamics and incomplete information.

2.1. Monge's problem and the Kantorovich relaxation

The problem of OMT, as formulated by Monge, is concerned with finding a mapping between two distributions of mass such that a measure of cost, interpreted as the cost of transportation, is minimized. That is, given two non-negative functions $\Phi_0 \in L_1(\mathcal{X}_0)$ and $\Phi_1 \in L_1(\mathcal{X}_1)$, defined on the spaces $\mathcal{X}_0$ and $\mathcal{X}_1$, respectively, one seeks a function $g : \mathcal{X}_0 \to \mathcal{X}_1$ minimizing the functional [60]

$$\int_{\mathcal{X}_0} c(x_0, g(x_0))\Phi_0(x_0)dx_0, \tag{1}$$

subject to the constraint

$$\int_{x_1 \in \mathcal{U}} \Phi_1(x_1)dx_1 = \int_{g(x_0)\in\mathcal{U}} \Phi_0(x_0)dx_0, \tag{2}$$

for all $\mathcal{U} \subset \mathcal{X}_1$. Here, $c : \mathcal{X}_0 \times \mathcal{X}_1 \to \mathbb{R}_+$ denotes the cost of transporting one unit of mass from points in $\mathcal{X}_0$ to points in $\mathcal{X}_1$, yielding that (1) equals the total cost of transporting the distribution Φ_0 to Φ_1 for a given mapping g. The constraint in (2) enforces the condition that g transports Φ_0 to Φ_1, i.e., the mapping g moves the mass in Φ_0 onto Φ_1 without any loss or addition of mass in the transition. This may be expressed using the shorthand notation $g_\#\Phi_0 = \Phi_1$, where $g_\#\Phi_0$ denotes the so-called push-forward measure, i.e.,

$$\int_{x_1 \in \mathcal{U}} (g_\#\Phi_0)(x_1)dx_1 = \int_{g(x_0)\in\mathcal{U}} \Phi_0(x_0)dx_0, \tag{3}$$

for all $\mathcal{U} \subset \mathcal{X}_1$. The notation thus implies that for any subset $\mathcal{U}$ of $\mathcal{X}_1$, the mass of $g_\#\Phi_0$ on $\mathcal{U}$ is equal to the mass of Φ_0 on the pre-image of $\mathcal{U}$. It directly follows that the total mass of $g_\#\Phi_0$ equals that of Φ_0. Finding such a g that minimizes (1), in general, constitutes a non-convex, non-trivial problem. In contrast, the Kantorovich relaxation of (1) results in a convex problem by replacing the mapping g with an optimal coupling between Φ_0 and Φ_1. This also allows for generalizing the problem to non-negative mass distributions,[1] which we denote $\Phi_0 \in \mathcal{M}_+(\mathcal{X}_0)$ and $\Phi_1 \in \mathcal{M}_+(\mathcal{X}_1)$. The

[1] Such mass distributions are sometimes referred to as measures and may contain, e.g., Dirac delta functions.

set of valid couplings is the set of measures on the product space $\mathcal{X}_0 \times \mathcal{X}_1$ with marginals coinciding with Φ_0 and Φ_1, i.e.,

$$\Omega(\Phi_0, \Phi_1) = \Big\{ M \in \mathcal{M}_+(\mathcal{X}_0 \times \mathcal{X}_1) \mid \int_{\mathcal{X}_1} M(\cdot, x_1) dx_1 = \Phi_0, \int_{\mathcal{X}_0} M(x_0, \cdot) dx_0 = \Phi_1 \Big\}. \quad (4)$$

Here, if $M \in \Omega(\Phi_0, \Phi_1)$, then M is referred to as a transport plan, as the function value $M(x_0, x_1)$ may be interpreted as the amount of mass that is transported between x_0 and x_1. Thus, the marginal constraint (4) replaces the condition $g_{\#}\Phi_0 = \Phi_1$ used in the Monge formulation. It should also be noted that it is assumed that the mass of Φ_0 and Φ_1 are the same, i.e., $\int_{\mathcal{X}_0} \Phi_0(x_0) dx_0 = \int_{\mathcal{X}_1} \Phi_1(x_1) dx_1$, in which case $\Omega(\Phi_0, \Phi_1)$ is always non-empty. The Monge–Kantorovich problem of optimal mass transport may then be stated as

$$\underset{M \in \Omega(\Phi_0, \Phi_1)}{\text{minimize}} \quad \int_{\mathcal{X}_0 \times \mathcal{X}_1} M(x_0, x_1) c(x_0, x_1) dx_0 dx_1, \quad (5)$$

i.e., it seeks the transport plan M minimizing the cost of transportation between Φ_0 and Φ_1, where the point-wise cost of moving a unit mass from $x_0 \in \mathcal{X}_0$ to $x_1 \in \mathcal{X}_1$ is given by $c(x_0, x_1)$. In contrast with (1), the objective function in (5) is linear in M and the optimization problem always has an optimal solution. Further, for $\Phi_0 \in L_1(\mathcal{X}_0)$ and $\Phi_1 \in L_1(\mathcal{X}_1)$, the problems (1) and (5) are equivalent [60].

The resulting minimal objective value of (5) may be used in order to define a notion of distance $S : \mathcal{M}_+(\mathcal{X}_0) \times \mathcal{M}_+(\mathcal{X}_1) \to \mathbb{R}$ according to

$$S(\Phi_0, \Phi_1) = \min_{M \in \Omega(\Phi_0, \Phi_1)} \int_{\mathcal{X}_0 \times \mathcal{X}_1} M(x_0, x_1) c(x_0, x_1) dx_0 dx_1. \quad (6)$$

Recently, the idea of using an OMT-induced concept of proximity has attracted considerable interest in a plethora of modeling applications, such as serving as the learning criterion in machine learning scenarios [3,6], for performing texture and color transfer [20], dictionary learning [50], as well as being used as an evaluation metric [19]. Also, it has been used for inducing metric structure on the space of power spectra [26], where it was shown that choosing c as a metric on $\mathcal{X} = \mathcal{X}_0 = \mathcal{X}_1$ in general results in S being metric on $\mathcal{M}_+(\mathcal{X})$. Similar efforts have been directed towards defining distances for Toeplitz covariance matrices [24], allowing for formulating a spectral estimation framework with inherent robustness [22,23]. One of the main motivations for using OMT-based distances such as (6) is its ability to reflect the geometric properties of the underlying space, $\mathcal{X}$. As a consequence of this, using OMT distances in order to perform interpolation-like tasks, such as, e.g., morphing of images [34], or, recently, for describing the time evolution of a signal on a graph [52], typically yields meaningful and interpretable signal reconstructions. This should be contrasted with the standard Euclidean metric that, when used in the same applications, has been shown to give rise to fade-in fade-out effects, thereby failing to respect underlying assumptions for the signal generating mechanisms, such as, e.g., finite movement speed [24,26,52].

2.2. Selection of cost function

It may be noted that the specific choice of cost function c used in the OMT objective in (5) may greatly influence the transport plan M representing the optimal solution. Preferably, c should be chosen as to reflect properties or limitations of the considered application, e.g., in order to ensure expected smooth trajectories of transported mass that may be dictated by physical considerations. In order to allow for incorporating knowledge of the dynamics of the process generating the signal observations, we here utilize the idea to model the underlying dynamics using a linear system which was proposed in [13]. Specifically, let the underlying space be $\mathcal{X} \subset \mathbb{R}^d$, and consider a mass particle with time-varying location, or state, $x(t) \in \mathcal{X}$. Furthermore, assume that the dynamics of the particle may be well modeled by the (continuous time) linear differential equation[2]

$$\dot{x}(t) = Ax(t) + Bu(t), \quad (7)$$

where $x(t) \in \mathcal{X}$, $A \in \mathbb{R}^{d \times d}$, $B \in \mathbb{R}^{d \times f}$, and where $u(t) \in \mathbb{R}^f$ denotes an input signal, steering the evolution of the state $x(t)$. Following the theory from [13], we will herein consider this type of model in order to define a cost function $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$, describing the cost of transport for a unit mass between two points of $\mathcal{X}$, corresponding to the time instances $t = t_0$ and $t = t_1$. For notational simplicity we here present the case $t_0 = 0$ and $t_1 = 1$. Specifically, we define the cost for transporting mass between $x_0 = x(0)$ to $x_1 = x(1)$ via a quadratic optimal control problem, according to

$$c(x_0, x_1) = \min_{u} \int_0^1 \|u(t)\|_2^2 dt \quad \text{subject to} \quad \dot{x}(t) = Ax(t) + Bu(t), \; x(0) = x_0, \; x(1) = x_1, \quad (8)$$

i.e., the cost is equal to the minimal input signal energy required to steer the state from x_0 at $t = 0$ to x_1 at $t = 1$. For this case, the cost may be expressed in closed form as

$$c(x_0, x_1) = (x_1 - \mathcal{A}(1,0)x_0)^T W(1,0)^{-1} (x_1 - \mathcal{A}(1,0)x_0), \quad (9)$$

where $\mathcal{A}(t, \tau) = expm(A(t - \tau))$ is the state transition matrix of (7), with $expm(\cdot)$ denoting the matrix exponential, and where

$$W(t, s) = \int_s^t \mathcal{A}(t, \tau) B B^T \mathcal{A}(t, \tau)^T d\tau$$

is the controllability Gramian. It may be noted that the problem in (8) admits closed form solutions also for the time-varying case (see, e.g., [13]).

Thus, as the cost of transport is dictated by the system dynamics, the optimal transport plan is the coupling $M \in \mathcal{M}_+(\mathcal{X} \times \mathcal{X})$ that provides the mass association most in accordance with the trajectories implied by (7). Furthermore, having access to a transport plan, M, that minimizes the OMT criterion with the dynamic cost in (8) directly allows for computing intermediate mass distributions interpolating the given marginals using the system dynamics in (7); this is termed a displacement interpolation [41]. Specifically, for a particle $x(t)$ such that $x(0) = x_0$ and $x(1) = x_1$, by solving (7) the optimal trajectory is [13], Eq. (25)

$$\hat{x}(t; x_0, x_1) = \mathcal{A}(t,1) W(1,t) W(1,0)^{-1} \mathcal{A}(1,0) x_0 + W(t,0) \mathcal{A}(1,t)^T W(1,0)^{-1} x_1.$$

Then, one may construct an interpolating mass distribution, or spectrum, Φ_τ^M, corresponding to time $\tau \in [0,1]$ and parametrized by the transport plan M, according to

$$\Phi_\tau^M(x) = \int_{\mathcal{X} \times \mathcal{X}} M(x_0, x_1) \delta\big(x - \hat{x}(\tau; x_0, x_1)\big) dx_0 dx_1,$$

where δ denotes the Dirac delta function. That is, $\Phi_\tau^M(x)$ describes the mass located at x at time τ by considering the mass of all pairs (x_0, x_1) such that their trajectories contain the point (x, τ). Also, this framework directly allows for defining extrapolating spectra, i.e., distributions of mass for $\tau > 1$, e.g., by letting the input signal be identically zero for $\tau > 1$, implying that the unforced system dynamics are used to describe the time evolution of the state,

[2] For notational simplicity, we here limit the exposition to the time-invariant case, although this may be generalized to a dynamical system in a straightforward manner.

and thus the spectrum. It may be noted that the interpolation for Toeplitz covariance matrices presented in [24] may be obtained as a special case of this formulation, using $A = 0$ and $B = I$.

2.3. Partial information

In the OMT formulations in Section 2.1, the input data, i.e., the marginals, used for computing distances and obtaining transport plans, are given. However, in many problems these distributions are only partially known, and are in fact often the quantities of interest to be estimated. In this work, we propose modeling tools that may be applied to several such classes of reconstruction and estimation problems, e.g., inverse problems encountered in imaging, radar, tomography, and signal processing. However, in order to give a concise description, we will herein focus on a few motivating applications in spatial spectral estimation and radar imaging. In such reconstruction problems, one typically seeks to recover a spectrum based on a number of measurements, e.g., estimates of the covariance function of a time series, or estimates of the covariance matrices corresponding to a set of sensor arrays. In this subsection, we describe the basic mathematical model relating the covariance estimates to the power spectra in a localization problem. In the subsequent sections, we will develop theory and formulate OMT problems which may be used to fuse such measurements obtained from different sensors or from different time points in an informed and dependable manner.

2.3.1. Localization and direction of arrival estimation

Consider a scenario in which point-like sources located in a space $\mathcal{X} \subset \mathbb{R}^d$ emit waves impinging on a set of sensors located at $x_k \in \mathbb{R}^d$ for $k = 1, \dots, q$, with $d \leq 3$. Letting the corresponding sensor signal vector be denoted $z \in \mathbb{C}^q$, one may consider the sensor array covariance matrix, often called the spatial covariance matrix,

$$R = \mathbb{E}\left(zz^H\right).$$

Then, the covariance R may be related to a positive measure on $\mathcal{X}$, denoted $\Phi \in \mathcal{M}_+(\mathcal{X})$, according to $R = \Gamma(\Phi)$, where the linear operator $\Gamma : \mathcal{M}(\mathcal{X}) \to \mathbb{M}^q$ is given by [25]

$$\Gamma(\Phi) = \int_{\mathcal{X}} a(x)\Phi(x)a(x)^H dx, \tag{10}$$

where dx denotes the Lebesgue measure on $\mathcal{X}$, where $\mathbb{M}^q$ denotes the space of Hermitian matrices of size $q \times q$, and where $\mathcal{M}(\mathcal{X})$ denotes the set of signed measures on $\mathcal{X}$. Here, $a : \mathcal{X} \to \mathbb{C}^q$ denotes the array steering vector, i.e., the manifold vector, whose specific functional form depends on the sensor locations, the wavelength, as well as on the propagation properties of this space. For example, the array manifold vector corresponding to spherical wave fronts in $\mathcal{X} \subset \mathbb{R}^d$ may be defined as [31]

$$a(x) = \left(\|x_k - x\|_2^{-(d-1)/2} \exp\left(-2\pi \mathrm{i} \frac{\|x_k - x\|_2}{\xi} \right) \right)_{k=1}^{q}, \tag{11}$$

where $\|\cdot\|_2$ denotes the Euclidean norm, i is the imaginary unit, and ξ is the source signal wavelength. Here, Φ is referred to as a spatial spectrum, as it describes the distribution of mass, or power, on $\mathcal{X}$, with the locations of the signal sources on $\mathcal{X}$ corresponding to the support of Φ.

In this work, we will use the problem of localization, seen as a spectral estimation problem, to illustrate how to utilize the proposed multi-marginal OMT framework for fusing information. Specifically, we will consider the problems of sensor fusion and spectral tracking. Sensor fusion corresponds to combining measurements obtained from separate sensor arrays in forming a single estimate of the spatial spectrum, whereas tracking refers to combining measurements collected at consecutive time points in order to reconstruct the time evolution of a spatial spectrum.

2.3.2. Dynamical models with partially observed states

In the context of using dynamical models such as (7), the partial information contained in the measurements may be related to an output space of lower dimension than the state space $\mathcal{X}$. Specifically, the system description in (7) may be extended with an observation equation according to

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Fx(t), \end{aligned} \tag{12}$$

where $F \in \mathbb{R}^{g \times d}$ is a matrix describing the output of the system, implying that covariances are related to spectra detailing mass distributions on a space $\mathcal{Y} \subset \mathbb{R}^g$, potentially different from $\mathcal{X}$, where $y(t) \in \mathcal{Y}$ is the observed state. Using the system in (12), one may thus incorporate prior knowledge of underlying dynamics, expressed through the state equation, even though one only has access to partial measurements of an observed state. For the corresponding spectra, it holds that if $\Phi \in \mathcal{M}_+(\mathcal{X})$, then $F_\# \Phi \in \mathcal{M}_+(\mathcal{Y})$, where $F_\#$ denotes the push-forward operation, as defined in (3). Note that, using the system in (12), covariance measurements corresponding to observable quantities are related to spectra on the space $\mathcal{Y}$, i.e., to elements of $\mathcal{M}_+(\mathcal{Y})$, and only have an indirect link to spectra on the space $\mathcal{X}$. However, the push-forward operation $F_\#$ allows for expressing the connection between covariance matrices and elements of $\mathcal{M}_+(\mathcal{X})$ by modifying the operators Γ to incorporate $F_\#$ according to

$$\Gamma(\Phi) = \int_{\mathcal{Y}} a(y)(F_\#\Phi)(y)a(y)^H dy. \tag{13}$$

It is worth noting that the domain of definition of the linear operator Γ is still $\mathcal{M}(\mathcal{X})$, whereas the domain of definition for the array steering vectors is $\mathcal{Y}$, reflecting that the covariance measurements are related to elements of $\mathcal{M}_+(\mathcal{Y})$. Also, the transport plan M is still an element of $\mathcal{M}_+(\mathcal{X} \times \mathcal{X})$, allowing for modeling dynamical structures on the space generating data, i.e., $\mathcal{X}$, that are not present in the measurement space $\mathcal{Y}$. Thus, in this context, the notion of partial information is twofold. Firstly, mass distributions are not observed directly, but only through linear measurements on a considerably smaller dimension, i.e., covariances in a finite-dimensional space as opposed to the infinite-dimensional space of measures for spectra. Secondly, these spectra are mappings of spectra that are elements of an even larger space. We will in the next section examine how these ideas, together with the concept of multi-marginal transport plans, can be used in order to arrive at compact formulations for spectral estimation.

3. Multi-marginal optimal mass transport

It is worth noting that the formulation in (5) seeks a transport plan connecting two mass distributions, or marginals, Φ_0 and Φ_1, which may be interpreted as a description of how to morph Φ_0 into Φ_1. A natural extension of this is to consider transport between a larger set of marginals, i.e., a set $\{\Phi_\mathcal{T}\}_{t=0}^{\mathcal{T}}$ for $\mathcal{T} \geq 1$ [45]. To this end, let $\mathbf{M} \in \mathcal{M}_+(\boldsymbol{\mathcal{X}})$, where $\boldsymbol{\mathcal{X}} = \mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_\mathcal{T}$, and define the set of projection operators $\mathcal{P}_t : \mathcal{M}(\boldsymbol{\mathcal{X}}) \to \mathcal{M}(\mathcal{X}_t)$, for $t = 0, \dots, \mathcal{T}$, as

$$\mathcal{P}_t(\mathbf{M}) = \int_{\boldsymbol{\mathcal{X}}_{-t}} \mathbf{M}(x_0, \dots, x_t, \dots, x_\mathcal{T}) d\mathbf{x}_{-t}, \tag{14}$$

where $d\mathbf{x}_{-t} = dx_0 \dots dx_{t-1} dx_{t+1} \dots dx_\mathcal{T}$ and

$$\boldsymbol{\mathcal{X}}_{-t} = \mathcal{X}_0 \times \cdots \times \mathcal{X}_{t-1} \times \mathcal{X}_{t+1} \times \cdots \times \mathcal{X}_\mathcal{T}.$$

It may be noted that one may express marginalization also for the standard two-marginal case using the operator $\mathcal{P}_t$. Then, $\mathbf{M} \in \mathcal{M}_+(\boldsymbol{\mathcal{X}})$ is referred to as a multi-marginal transport plan for the set $\{\Phi_t\}_{t=0}^{\mathcal{T}}$ if $\mathcal{P}_t(\mathbf{M}) = \Phi_t$, for $t = 0, \dots, \mathcal{T}$. Note here that $\mathbf{M}$ provides a complete description of the association of mass for the set

$\{\Phi_t\}_{t=0}^{\mathcal{T}}$, i.e., $\mathbf{M}(x_0, x_1, \dots, x_{\mathcal{T}})$ denotes the amount of mass at x_0, corresponding to the marginal Φ_0, that is transported to location x_t, corresponding to marginal Φ_t, for $t = 1, 2, \dots, \mathcal{T}$. Correspondingly, one may define a cost function $\mathcal{C} : \boldsymbol{\mathcal{X}} \to \mathbb{R}_+$, yielding a generalization of the OMT problem as (see also [45])

$$\begin{aligned} \underset{\mathbf{M}\in\mathcal{M}_+(\boldsymbol{\mathcal{X}})}{\text{minimize}} \quad & \int_{\boldsymbol{\mathcal{X}}} \mathbf{M}(\mathbf{x})\mathcal{C}(\mathbf{x})d\mathbf{x} \\ \text{subject to} \quad & \mathcal{P}_t(\mathbf{M}) = \Phi_t, \qquad \text{for } t = 0, \dots, \mathcal{T}, \end{aligned} \tag{15}$$

where $\mathbf{x} = (x_0, \dots, x_{\mathcal{T}})$ and $d\mathbf{x} = dx_0 dx_1 \dots dx_{\mathcal{T}}$. Thus, the minimizer of (15) is the transport plan associated with the minimal cost of transportation for the set $\{\Phi_t\}_{t=0}^{\mathcal{T}}$. This type of formulation has previously found application in, e.g., tomography [1] and fluid dynamics [11], with actual numerical approximations and implementations being considered in [9]. In this work, we consider extending this problem to scenarios in which the marginals $\{\Phi_t\}_{t=0}^{\mathcal{T}}$ are not directly observable. Instead, we assume that one has access to partial information in the form of linear mappings of the marginals. In the context of spatial spectral estimation, this corresponds to the spatial covariance matrix, or estimates thereof, i.e., a sequence $\{R_t\}_{t=0}^{\mathcal{T}}$, where $R_t \in \mathbb{M}^{q_t}$ is related to a marginal $\Phi_t \in \mathcal{M}_+(\mathcal{X}_t)$ through the measurement equation $R_t = \Gamma_t(\Phi_t)$. It should be noted that the operators $\Gamma_t : \mathcal{M}(\mathcal{X}_t) \to \mathbb{M}^{q_t}$ may, in general, be different for different indices t. With this, the problem considered herein may be stated as

$$\begin{aligned} \underset{\substack{\mathbf{M}\in\mathcal{M}_+(\boldsymbol{\mathcal{X}}) \\ \Delta_t\in\mathbb{C}^{q_t\times q_t}}}{\text{minimize}} \quad & \int_{\boldsymbol{\mathcal{X}}} \mathbf{M}(\mathbf{x})\mathcal{C}(\mathbf{x})d\mathbf{x} + \textstyle\sum_{t=0}^{\mathcal{T}} \gamma_t \|\Delta_t\|_{\mathrm{F}}^2 \\ \text{subject to} \quad & \Gamma_t(\mathcal{P}_t(\mathbf{M})) = R_t + \Delta_t \qquad \text{for } t = 0, \dots, \mathcal{T}, \end{aligned} \tag{16}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm and where $\gamma_t > 0$, for $t = 0, 1, \dots, \mathcal{T}$, are user-specified weights. Note that the measurement equations have been augmented by error terms $\Delta_t \in \mathbb{C}^{q_t\times q_t}$ in order to allow for noisy covariance matrix estimates R_t. The parameters γ_t then allow for determining a trade-off between how much consideration should be taken to the transport cost and the measurement error, respectively. For simplicity, we here penalize Δ_t by the squared Frobenius norm. In order to compute couplings between marginals corresponding to index pairs (t, s), one requires the corresponding bi-marginal transport plan. Due to the structure of the multi-marginal transport plan $\mathbf{M}$, this is directly computable by means of the bi-marginal projection operator $\mathcal{P}_{t,s} : \mathcal{M}(\boldsymbol{\mathcal{X}}) \to \mathcal{M}(\mathcal{X}_t \times \mathcal{X}_s)$ defined by

$$\mathcal{P}_{t,s}(\mathbf{M}) = \int_{\boldsymbol{\mathcal{X}}_{-t,-s}} \mathbf{M}(x_0, \dots, x_{\mathcal{T}})d\mathbf{x}_{-t,-s}, \tag{17}$$

where $d\mathbf{x}_{-t,-s} = dx_0 \dots dx_{t-1}dx_{t+1} \dots dx_{s-1}dx_{s+1} \dots dx_{\mathcal{T}}$ and

$$\boldsymbol{\mathcal{X}}_{-t,-s} = \mathcal{X}_0 \times \cdots \times \mathcal{X}_{s-1} \times \mathcal{X}_{s+1} \times \cdots \times \mathcal{X}_{t-1} \times \mathcal{X}_{t+1} \times \cdots \times \mathcal{X}_{\mathcal{T}}.$$

This may then be used, e.g., for interpolation between adjacent marginals for which the cost describes a dynamic model as in Section 2.2. In Section 5, we present computationally efficient methods for computing discrete approximations of the projections in (14) and (17). In fact, computing the marginal projections in (14) will constitute an integral component in the algorithm for solving (16), which will be presented in Section 4.

Next, we will see that a number of spectral estimation problems may be cast in the form (16). As will be clear in the following, the only thing distinguishing these spectral estimation problems is the choice of cost function $\mathcal{C}$. For ease of notation, we will in this description often assume that the domains of the marginals are the same, however, this generalizes straightforwardly to general domains.

3.1. Sensor fusion

Consider a scenario in which waves impinge on a set of $J \in \mathbb{N}$ sensor arrays, with corresponding array manifold vectors $a_j : \mathcal{X} \to \mathbb{C}^{q_j}$ and operators $\Gamma_j : \mathcal{M}(\mathcal{X}) \to \mathbb{M}^{q_j}$, for $j = 1, \dots, J$, where it may be noted that the number of sensors, q_j, may differ among the different arrays. Then, a spatial spectrum $\Phi \in \mathcal{M}_+(\mathcal{X})$ gives rise to a set of covariance matrices R_j, for $j = 1, \dots, J$, where $R_j = \Gamma_j(\Phi)$. Thus, as all sensor array measurements have been generated by the same signal, the problem of estimating the spatial spectrum Φ, and thereby the location of the signal sources, corresponds to the inverse problem of finding Φ such that $R_j = \Gamma_j(\Phi)$, for $j = 1, \dots, J$. However, in practice, calibration errors in the sensor arrays, i.e., errors in the operators Γ_j, may result in there existing no single $\Phi \in \mathcal{M}_+(\mathcal{X})$ that is consistent with all array measurements. For scenarios in which only the covariance matrices for the separate sensor arrays are available, for example if these are estimated locally [32,38], we propose to address this issue using a spectral barycenter formulation. Specifically, we propose to solve

$$\begin{aligned} \underset{\substack{\Phi_j\in\mathcal{M}_+(\mathcal{X}) \\ \Delta_j\in\mathbb{C}^{q_j\times q_j}}}{\text{minimize}} \quad & \textstyle\sum_{j=1}^{J} S(\Phi_0, \Phi_j) + \gamma_j \left\|\Delta_j\right\|_F^2 \\ \text{subject to} \quad & \Gamma_j(\Phi_j) = R_j + \Delta_j, \ \ j = 1, \dots, J, \end{aligned} \tag{18}$$

and estimate the spatial spectrum as the minimizer Φ_0. Here, S is the OMT distance defined in (6). Note here that Φ_j, for $j = 1, \dots, J$, are spectra consistent with the set of observations R_j, whereas Φ_0 is the spectrum closest, in the OMT sense, to this set of spectra, i.e., Φ_0 is the spectral barycenter. It may be noted that the formulation in (18) induces robustness to the estimate Φ_0 as small spatial perturbations caused by, e.g., calibration errors, do not result in large values of the distance measure S. One may formulate the problem in (18) as a multi-marginal problem on the form (16) by specifying the multi-marginal cost function $\mathcal{C}$ as

$$\mathcal{C}(\mathbf{x}) = \sum_{j=1}^{J} c(x_0, x_j), \tag{19}$$

where $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ is the pair-wise cost function in the definition of S. With this, the barycenter problem may be stated as

$$\begin{aligned} \underset{\substack{\mathbf{M}\in\mathcal{M}_+(\boldsymbol{\mathcal{X}}) \\ \Delta_j\in\mathbb{C}^{q_j\times q_j}}}{\text{minimize}} \quad & \int_{\boldsymbol{\mathcal{X}}} \mathbf{M}(\mathbf{x})\Big(\textstyle\sum_{j=1}^{J} c(x_0, x_j)\Big)d\mathbf{x} + \textstyle\sum_{j=1}^{J} \gamma_j \|\Delta_j\|_{\mathrm{F}}^2 \\ \text{subject to} \quad & \Gamma_j\big(\mathcal{P}_j(\mathbf{M})\big) = R_j + \Delta_j \qquad \text{for } j = 1, \dots, J. \end{aligned} \tag{20}$$

To see that (19) indeed induces a barycenter solution when used in (20), note that the cost function $\mathcal{C}$ is structured as to penalize transport between a central marginal, corresponding to the index $j = 0$, and all other marginals, corresponding to indices $j = 1, \dots, J$. As only marginals with indices $j = 1, \dots, J$ correspond to actual measurements, this implies that the marginal $t = 0$, which is not constrained to be related to any measurements, should be close, in the OMT sense, to the entire ensemble of marginals. Note here that the barycenter does not appear as an explicit variable, but is instead given by the projection of $\mathbf{M}$ on the zeroth marginal, i.e., $\Phi_0 = \mathcal{P}_0(\mathbf{M})$. It may be noted that, in general, one may utilize different OMT distances S_j for different arrays, e.g., one may replace the sums in (18) and (19) with a weighted sum to get a weighted barycenter.

3.2. Tracking

One may also consider a scenario in which a single sensor array[3] is used to observe the evolution of a signal over time. Specifically, assume that estimates of the array covariance matrix are obtained at $\mathcal{T} + 1$ time instances corresponding to the nominal times $t = 0, 1, \dots, \mathcal{T}$, giving rise to the sequence $\{R_t\}_{t=0}^{\mathcal{T}}$. Under the assumption that the corresponding spatial spectrum should evolve

[3] We present a formulation for handling multi-array cases in the next section.

smoothly, implying signal sources moving at finite speed, one may estimate the corresponding sequence of spectra, $\{\Phi_t\}_{t=0}^{\mathcal{T}}$, using the tracking formulation

$$\begin{aligned} \underset{\substack{\Phi_t\in\mathcal{M}_+(\mathcal{X})\\ \Delta_t\in\mathbb{C}^{q\times q}}}{\text{minimize}} \quad & \textstyle\sum_{t=1}^{\mathcal{T}} S_t(\Phi_{t-1},\Phi_t)+\gamma_t\|\Delta_t\|_F^2 \\ \text{subject to} \quad & \Gamma(\Phi_t)=R_t+\Delta_t,\ t=0,\dots,\mathcal{T}, \end{aligned} \tag{21}$$

i.e., the sequence $\{\Phi_t\}_{t=0}^{\mathcal{T}}$ minimizes a sequential OMT distance, while interpolating the (de-noised) measurement sequence $\{R_t\}_{t=0}^{\mathcal{T}}$. It may here be noted that the OMT distance, S_t, is indexed by t, as to express that the underlying cost of transport may change depending on the time. Correspondingly, a multi-marginal formulation may be obtained by considering a cost function $\mathcal{C}$ such that

$$\mathcal{C}(\mathbf{x})=\sum_{t=1}^{\mathcal{T}} c_t(x_{t-1},x_t), \tag{22}$$

with the index t expressing time-dependence, yielding

$$\begin{aligned} \underset{\substack{\mathbf{M}\in\mathcal{M}_+(\boldsymbol{\mathcal{X}})\\ \Delta_t\in\mathbb{C}^{q\times q}}}{\text{minimize}} \quad & \textstyle\int_{\boldsymbol{\mathcal{X}}}\mathbf{M}(\mathbf{x})\big(\sum_{t=1}^{\mathcal{T}} c_t(x_{t-1},x_t)\big)d\mathbf{x}+\sum_{t=0}^{\mathcal{T}}\gamma_t\|\Delta_t\|_{\mathrm{F}}^2 \\ \text{subject to} \quad & \Gamma(\mathcal{P}_t(\mathbf{M}))=R_t+\Delta_t \qquad \text{for } t=0,\dots,\mathcal{T}. \end{aligned} \tag{23}$$

Here, the estimated spectral sequence is given by the marginals of $\mathbf{M}$, i.e., $\Phi_t=\mathcal{P}_t(\mathbf{M})$, for $t=0,1,\dots,\mathcal{T}$. It is worth noting that the cost function $\mathcal{C}$ penalizes transport between consecutive marginals, i.e., between pairs corresponding to indices $(t-1,t)$, for $t=1,\dots,\mathcal{T}$. This allows for exploiting assumptions of an underlying dynamic model, expressed in c_t in (22), as described in Section 2.2. Specifically, let the space on which the spatial spectra are defined be denoted $\mathcal{Y}$, and let $\mathcal{X}$ correspond to a larger space, allowing for expressing the system's dynamics. Using the system description in (12), the cost function c_t may be defined according to (8), and the operators Γ may be extended to include the push-forward operation, as detailed in (13). The sequence of spatial spectra may then be obtained as $\Phi_t=F_{\#}\mathcal{P}_t(\mathbf{M})$. We will in Section 6 examine the implications of using a dynamical model as opposed to a static one when performing OMT-based spectral tracking.

Example 1. Consider a DoA estimation scenario in $\mathbb{R}^2$, in which a sequence of array covariance matrices is available, where one in addition to the direction of arrival, $\theta\in[-\pi,\pi)$ aims to model also the velocity of the targets, in order to enforce smooth trajectories. Letting $v(t)\in\mathbb{R}$ denote the angular speed, i.e., $\dot\theta(t)=v(t)$, one may thus select $\mathcal{Y}=[-\pi,\pi)$ and $\mathcal{X}=[-\pi,\pi)\times\mathbb{R}$. In order to impose the assumption of smooth trajectories, one may then enforce finite acceleration by, in the state space representation in (12), using the matrices

$$A=\begin{bmatrix}0&1\\0&0\end{bmatrix},\ B=\begin{bmatrix}0&1\end{bmatrix}^T,\ F=\begin{bmatrix}1&0\end{bmatrix}.$$

It may be noted that the pair (A,B) enforces that the angular position can only be influenced through its angular velocity, i.e., through acceleration, whereas the matrix F expresses the fact that only the angular position is manifested in the array covariance matrix. The corresponding relation between spectra on $\mathcal{X}$ and $\mathcal{Y}$, denoted $\Phi_t^{\mathcal{X}}$ and $\Phi_t^{\mathcal{Y}}$, respectively, may then be expressed as

$$\Phi_t^{\mathcal{Y}}(\theta)=\big(F_{\#}\Phi_t^{\mathcal{X}}\big)(\theta)=\int_{\mathbb{R}}\Phi_t^{\mathcal{X}}(\theta,v)dv,$$

for $\theta\in[-\pi,\pi)$. The cost function c implied by the LQ control signal u is then given by (see, e.g., [15])

$$c(x_0,x_1)=(x_1-expm(A)x_0)^TW(1,0)^{-1}(x_1-expm(A)x_0)$$

with

$$expm(A(t-\tau))=\begin{bmatrix}1&t-\tau\\0&1\end{bmatrix},\ W(1,0)=\begin{bmatrix}1/3&1/2\\1/2&1\end{bmatrix}.$$

3.3. Combining tracking and barycenters

In addition to the problems described in Sections 3.1 and 3.2, one may consider a combination of the two, i.e., the tracking of a spectral barycenter over time. This then allows for addressing issues of resolving spatial ambiguities, as well as promoting smooth spectral trajectories over time as to, e.g., induce robustness to noisy measurements. Therefore, consider J sensor arrays observing a signal during $\mathcal{T}+1$ time instances, giving rise to a set of covariance matrices $R_t^{(j)}$, for $j=1,\dots,J$ and $t=0,\dots,\mathcal{T}$, where the subscript indicates time, and the superscript corresponds to the array index. Thus, we seek to estimate a sequence of spectral barycenters, $\{\Phi_t\}_{t=0}^{\mathcal{T}}$, given covariance matrix estimates $R_t^{(j)}$, for $t=0,1,\dots,\mathcal{T}$ and $j=1,2,\dots,J$. In order to formulate the barycenter tracking problem, one may consider two separate types of transport costs; S_t detailing the OMT distance between consecutive barycenters, and $\tilde S$ describing the distances between barycenters and spectra consistent with the array measurements. For example, the distance S_t may reflect underlying dynamics whereas $\tilde S$ may correspond to a static OMT problem. Correspondingly, one may let the barycenter spectra be elements of $\mathcal{M}_+(\mathcal{X})$, where the space $\mathcal{X}$ may detail the system dynamics, whereas the consistent spectra are elements of $\mathcal{M}_+(\mathcal{Y})$, where $\mathcal{Y}$ is the measurement space.

With this, the barycenter tracking problem may be formulated as

$$\begin{aligned} \underset{\substack{\Phi_t\in\mathcal{M}_+(\mathcal{X}),\ \Phi_t^{(j)}\in\mathcal{M}_+(\mathcal{Y})\\ \Delta_t^{(j)}\in\mathbb{C}^{q_j\times q_j}}}{\text{minimize}} \quad & \mathbf{S}(\boldsymbol{\Phi})+\alpha\tilde{\mathbf{S}}(\boldsymbol{\Phi}) \\ \text{subject to} \quad & \Gamma^{(j)}\big(\Phi_t^{(j)}\big)=R_t^{(j)}+\Delta_t^{(j)}, \end{aligned} \tag{24}$$

for $t=0,\dots,\mathcal{T}$, and $j=1,\dots,J$, where

$$\mathbf{S}(\boldsymbol{\Phi})=\sum_{t=1}^{\mathcal{T}}S_t(\Phi_{t-1},\Phi_t)$$

$$\tilde{\mathbf{S}}(\boldsymbol{\Phi})=\sum_{t=0}^{\mathcal{T}}\sum_{j=1}^{J}\Big(\tilde S\big(\Phi_t^{(j)},F_{\#}\Phi_t\big)+\gamma_t^{(j)}\big\|\Delta_t^{(j)}\big\|_F^2\Big),$$

with $\alpha>0$ being a user-specified parameter allowing for deciding a trade-off between the smoothness of the tracking and the distance between a barycenter Φ_t and its corresponding consistent spectra $\Phi_t^{(j)}$, for $t=0,1,\dots,\mathcal{T}$ and $j=1,2,\dots,J$. It may here be noted the latter distance is stated in terms of the push-forward of the barycenter spectrum, thereby allowing for some states to be measured (e.g., position) whereas others are not (e.g., velocity). In order to arrive at a multi-marginal formulation on the form (16), one may define the cost function $\mathcal{C}$ as

$$\mathcal{C}(\mathbf{x})=\sum_{t=1}^{\mathcal{T}}c_t(x_{t-1,0},x_{t,0})+\alpha\sum_{t=0}^{\mathcal{T}}\sum_{j=1}^{J}\tilde c(x_{t,0},x_{t,j}) \tag{25}$$

where $\mathbf{x}\in\boldsymbol{\mathcal{X}}^J=\boldsymbol{\mathcal{X}}\times\boldsymbol{\mathcal{X}}\times\ldots\times\boldsymbol{\mathcal{X}}$, which is the direct product of $\mathcal{T}+1$ instances of $\boldsymbol{\mathcal{X}}$, each corresponding to a barycenter problem at one time point. That is, this space is a product of the form $\boldsymbol{\mathcal{X}}=\mathcal{X}\times\mathcal{Y}\times\cdots\times\mathcal{Y}$, with one instance of $\mathcal{X}$ and J instances of $\mathcal{Y}$. Here, $\mathbf{x}=(\mathbf{x}_0,\mathbf{x}_1,\dots,\mathbf{x}_{\mathcal{T}})$, where $\mathbf{x}_t=(x_{t,0},x_{t,1},\dots,x_{t,J})$, where the first index indicate the time t, for $t=0,1,\dots,\mathcal{T}$, and the second index indicates the array j, for $j=0,1,\dots,J$, with 0 corresponding to the barycenter. Here, c_t denotes the cost of transport for the tracking, whereas $\tilde c$ defines the cost of transportation between the barycenters and the marginal spectra. With this setup, one may formulate the multi-marginal problem as

$$\begin{aligned} \underset{\substack{\mathbf{M}\in\mathcal{M}_+(\boldsymbol{\mathcal{X}}^J)\\ \Delta_t^{(j)}\in\mathbb{C}^{q_j\times q_j}}}{\text{minimize}} \quad & \textstyle\int_{\boldsymbol{\mathcal{X}}^J}\mathbf{M}(\mathbf{x})\mathcal{C}(\mathbf{x})d\mathbf{x}+\alpha\sum_{t=0}^{\mathcal{T}}\sum_{j=1}^{J}\gamma_t^{(j)}\|\Delta_t^{(j)}\|_{\mathrm{F}}^2 \\ \text{subject to} \quad & \Gamma^{(j)}(\mathcal{P}_{(t,j)}(\mathbf{M}))=R_t^{(j)}+\Delta_t^{(j)}, \end{aligned} \tag{26}$$

for $t = 0, \dots, \mathcal{T}$, and $j = 1, \dots, J$. Here, the operator $\mathcal{P}_{(t,j)}$ projects onto the marginal (t, j) of $\mathbf{M}$ in direct analogy with (14), which for each time t corresponds to a spectrum consistent with the measurements of array j, for $j = 1, 2, \dots, J$. The corresponding sequence of barycenters may then be constructed as $\Phi_t = \mathcal{P}_{(t,0)}(\mathbf{M})$, for $t = 0, 1, \dots, \mathcal{T}$.

Having shown that the multi-marginal OMT problem with partial information in (16) may be used to formulate these spectral estimators, we proceed to present the main results of this work, i.e., how to construct discrete approximations of (16) that allow for computationally efficient solvers. Numerical examples that illustrate the behavior of (16) in spatial spectral estimation scenarios are then provided in Section 6.

4. Solving multi-marginal OMT problems with partial information of marginals

The multi-marginal OMT problem with partial information in (16) is an infinite-dimensional problem, as the multi-marginal transport plan $\mathbf{M}$ is an element of the function space $\mathcal{M}_+(\boldsymbol{\mathcal{X}})$. Fortunately, one may in practical implementations approximate (16) by a discrete counterpart. In this section, we present a computationally efficient solution algorithm for such a problem. Starting by describing a discretization of OMT problems with full information, i.e., problems of the form (15), we derive an algorithm for solving discrete OMT problems with partial information. We show that the critical component of the resulting algorithms is to compute the projections in (14). Section 5 presents several cases for which the projections in (14) can be computed efficiently.

4.1. Entropy regularized OMT and Sinkhorn iterations

In order to discretize the OMT problem in (5), one may construct a grid consisting of n grid points $X = \{\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_n\}$ on the support of the marginals. The discretized marginals are then represented by vectors $\Phi_t \in \mathbb{R}^n$ for $t = 0, \dots, \mathcal{T}$, with $\mathcal{T} \geq 1$. For ease of notation, we will herein assume that all marginals are discretized using the same number of points. However, this, as well as all presented results, generalizes straightforwardly to the case when the number of discretization points differ among the marginals. In the discretized formulation, the cost function c is represented by a $(\mathcal{T}+1)$-dimensional tensor $\mathbf{C} \in \mathbb{R}^{n^{\mathcal{T}+1}}$, where the element $\mathbf{C}_{i_0,i_1,\dots,i_\mathcal{T}} = \mathcal{C}(\mathrm{x}_{i_0}, \mathrm{x}_{i_1}, \dots, \mathrm{x}_{i_\mathcal{T}})$ denotes the transport cost associated with the tuple $(\mathrm{x}_{i_0}, \mathrm{x}_{i_1}, \dots, \mathrm{x}_{i_\mathcal{T}})$. Correspondingly, the transport plan may be represented by a tensor $\mathbf{M} \in \mathbb{R}_+^{n^{\mathcal{T}+1}}$, whose elements $\mathbf{M}_{i_0,i_1,\dots,i_\mathcal{T}}$ denote the amount of mass associated with the tuple $(\mathrm{x}_{i_0}, \mathrm{x}_{i_1}, \dots, \mathrm{x}_{i_\mathcal{T}})$. As a discrete analog to the projection in (14), let $P_t : \mathbb{R}^{n^{\mathcal{T}+1}} \to \mathbb{R}^n$ be the projection on the tth marginal of a tensor $\mathbf{M}$, which is computed by summing over all modes of $\mathbf{M}$ except t, i.e., the jth element of the vector $P_t(\mathbf{M}) \in \mathbb{R}^n$ is given by

$$P_t(\mathbf{M})_j = \sum_{\substack{i_0,\dots,i_{t-1},\\ i_{t+1},\dots,i_\mathcal{T}}} \mathbf{M}_{i_0,\dots,i_{t-1},j,i_{t+1},\dots,i_\mathcal{T}}.$$

Similarly, in analog to the bi-marginal projection in (17), let $P_{t_1,t_2}(\mathbf{M}) \in \mathbb{R}^{n\times n}$ denote the projection of $\mathbf{M}$ on the two joint marginals t_1 and t_2, i.e.,

$$P_{t_1,t_2}(\mathbf{M})_{j,\ell} = \sum_{\substack{i_t:\\ t\in\{0,\dots,\mathcal{T}\}\setminus\{t_1,t_2\}}} \mathbf{M}_{i_0,\dots,i_{t_1-1},j,i_{t_1+1},\dots,i_{t_2-1},\ell,i_{t_2+1},\dots,i_\mathcal{T}}.$$

With this, the discretized and entropy regularized multi-marginal OMT problem is given by

$$\begin{aligned} &\underset{\mathbf{M}\in\mathbb{R}^{n^{\mathcal{T}+1}}}{\text{minimize}} && \langle \mathbf{C}, \mathbf{M}\rangle \\ &\text{subject to} && P_t(\mathbf{M}) = \Phi_t, \quad t = 0, \dots, \mathcal{T}, \end{aligned} \tag{27}$$

where $\langle \mathbf{C}, \mathbf{M}\rangle \triangleq \sum_{i_0,\dots,i_\mathcal{T}} \mathbf{M}_{i_0,\dots,i_\mathcal{T}} \mathbf{C}_{i_0,\dots,i_\mathcal{T}}$. Although (27) is a linear program, finding an optimal transport plan can be computationally cumbersome due to the large number of variables. In particular, the classical bi-marginal OMT problem with a grid of size n yields a discrete problem with n^2 variables. It may be noted that standard linear programming methods require $\mathcal{O}(n^3 \log(n))$ operations to find the optimum for such a problem [46], proving infeasible even for moderate grid sizes.

One way to bypass the expensive computations inherent in finding the exact transport plan is to regularize the OMT problem, yielding an approximation of (27). To this end, it was first suggested in [17] to introduce an entropy term to the objective function. In this setting, the problem in (27) is modified to

$$\begin{aligned} &\underset{\mathbf{M}\in\mathbb{R}^{n^{\mathcal{T}+1}}}{\text{minimize}} && \langle \mathbf{C}, \mathbf{M}\rangle + \epsilon\mathcal{D}(\mathbf{M}) \\ &\text{subject to} && P_t(\mathbf{M}) = \Phi_t, \quad t = 0, \dots, \mathcal{T}, \end{aligned} \tag{28}$$

where

$$\mathcal{D}(\mathbf{M}) \triangleq \sum_{i_0,\dots,i_\mathcal{T}} \left(\mathbf{M}_{i_0,\dots,i_\mathcal{T}} \log\left(\mathbf{M}_{i_0,\dots,i_\mathcal{T}}\right) - \mathbf{M}_{i_0,\dots,i_\mathcal{T}} + 1\right),$$

and $\epsilon > 0$ is a small regularization parameter. The entropy regularized OMT problem is related to the SchrÃ¶dinger bridge problem, which studies the most likely evolution of a particle cloud observed at two different time instances [14,37]. Note that due to the entropy term, the optimal transport plan $\mathbf{M}$ corresponding to problem (28) always has full support, whereas the solution to (27) may be sparse. However, in contrast to (27), the regularized problem (28) is strictly convex and thus always guarantees a unique solution. Furthermore, in [17], it was shown for the bi-marginal setting that, as $\epsilon \to 0$, the solution of (28) converges to the solution of (27) with maximal entropy. More importantly, though, the formulation in (28) allows for deriving an efficient method to find the (approximate) optimal transport plan $\mathbf{M}$, as it implies a low-dimensional representation requiring only $(\mathcal{T}+1)n$ variables instead of the original $n^{\mathcal{T}+1}$ variables.

Specifically, in the original bi-marginal case, where the mass transport plan and cost tensor are given by matrices $M, C \in \mathbb{R}^{n\times n}$, the regularized problem allows for expressing the solution M through diagonal scaling of an $n \times n$ constant matrix. To see this, consider the Lagrange function of (28) with $\mathcal{T} = 1$ and dual variables $\lambda_0, \lambda_1 \in \mathbb{R}^n$, i.e.,

$$\begin{aligned} L(M, \lambda_0, \lambda_1) = \operatorname{tr}\left(C^T M\right) + \epsilon\mathcal{D}(M) \\ + \lambda_0^T(\Phi_0 - M\mathbf{1}) + \lambda_1^T(\Phi_1 - M^T\mathbf{1}). \end{aligned} \tag{29}$$

For fixed dual variables, the minimum of (29) with respect to M is attained when the gradient with respect to the matrix entries m_{ij} vanishes, yielding a solution of the form

$$M = \operatorname{diag}(u_0) K \operatorname{diag}(u_1), \tag{30}$$

where $K = \exp(-C/\epsilon)$, $u_0 = \exp(\lambda_0/\epsilon)$, $u_1 = \exp(\lambda_1/\epsilon)$, and where $\exp(\cdot)$ denotes element-wise application of the exponential function. Then, as the matrix K has strictly positive elements, it follows from Sinkhorn's theorem [53] that there is a unique matrix M on the form (30) with prescribed strictly positive row and column sums Φ_0 and Φ_1. Moreover, the two positive vectors u_0 and u_1 are unique up to multiplication with a scalar and may be found via Sinkhorn iterations, i.e., by iteratively updating u_0 and u_1 as to satisfy the marginal constraints. Specifically, the Sinkhorn iterations are given by [17]

$$\begin{aligned} u_0 &= \Phi_0 ./ (K u_1) \\ u_1 &= \Phi_1 ./ (K^T u_0), \end{aligned} \tag{31}$$

where ./ denotes elementwise division. These iterations converge linearly [39], with the computational bottleneck of the scheme being the two matrix-vector multiplications. The Sinkhorn iterations

Algorithm 1 Sinkhorn method for the multi-marginal optimal mass transport problem [9].

Given: Initial guess $u_t > 0$, for $t = 0, \dots, \mathcal{T}$; starting point t
while Sinkhorn not converged **do**
  $u_t \leftarrow u_t \odot \Phi_t ./ P_t(\mathbf{K} \odot \mathbf{U})$,
  $t \leftarrow t + 1 \pmod{\mathcal{T}+1}$
end while
return $\mathbf{M} = \mathbf{K} \odot \mathbf{U}$

thus provide an efficient technique for finding approximate solutions to OMT problems.

In direct analog to the bi-marginal case, by considering the Lagrangian relaxation of (28), one finds that the transport tensor may be represented by "diagonal" scaling of a constant tensor. Specifically, one may write $\mathbf{M} = \mathbf{K} \odot \mathbf{U}$, where $\odot$ denotes elementwise multiplication, for the two tensors $\mathbf{K}, \mathbf{U} \in \mathbb{R}^{n^{\mathcal{T}+1}}$, given by $\mathbf{K} = \exp(-\mathbf{C}/\epsilon)$ and

$$\mathbf{U} = (u_0 \otimes u_1 \otimes \cdots \otimes u_{\mathcal{T}}), \quad \iff \mathbf{U}_{i_0,\dots,i_{\mathcal{T}}} = \prod_{t=0}^{\mathcal{T}} (u_t)_{i_t},$$

for a set of vectors $u_0, \dots, u_{\mathcal{T}} \in \mathbb{R}^n$. It may be noted that this representation for the mass transfer tensor $\mathbf{M}$ is a direct generalization of the representation in the standard bi-marginal OMT theory, where $\mathbf{M}$, $\mathbf{K}$, and $\mathbf{U} = u_0 \cdot u_1^T$ are matrices, and

$$\mathbf{M} = \mathbf{K} \odot \mathbf{U} = \mathbf{K} \odot \left(u_0 \cdot u_1^T\right) = \operatorname{diag}(u_0)\mathbf{K}\operatorname{diag}(u_1),$$

as in (30). Thus, the problem in (28) may be reduced to finding $\mathcal{T}+1$ scaling vectors in $\mathbb{R}^n$ instead of directly optimizing over the complete mass tensor $\mathbf{M} \in \mathbb{R}^{n^{\mathcal{T}+1}}$. Interestingly, the scaling vectors $u_0, \dots, u_{\mathcal{T}}$ are unique up to a scalar for any $\mathcal{T} \geq 1$ [39]. Furthermore, they may be determined via a Sinkhorn iteration scheme similar to (31), as presented in [9] for multi-marginal OMT problems. For completeness, we state the result here: given an initial set of positive vectors $u_t \in \mathbb{R}^n$, for $t = 0, 1, \dots, \mathcal{T}+1$, the Sinkhorn method (Algorithm 1) is to iteratively update according to

$$u_t \leftarrow u_t \odot \Phi_t ./ P_t(\mathbf{K} \odot \mathbf{U}), \quad t = 0, 1, \dots, \mathcal{T}. \tag{32}$$

It may be noted that also this scheme reduces to the standard Sinkhorn iterations (31) for the case $\mathcal{T} = 1$, i.e., a bi-marginal problem.

Parallels between the Sinkhorn iteration scheme and established algorithms for convex optimization problems have been explored in several works. For instance, it has been shown that Sinkhorn iterations correspond to iterative Bregman projections [9]. In the bi-marginal case they have been derived as a block-coordinate ascent in the dual formulation of (28) [33]. In this work, we utilize the latter result in order to develop new Sinkhorn-type schemes for OMT problems with partial information of the marginals, which the problem with full information in (28) is a special case of.

4.2. Method for partial information of marginals

In this section, we derive an efficient algorithm for solving discrete OMT problems where only partial information of the marginals are available, i.e., formulating discrete approximations of the problem in (16). In this setting, one does not have access to the marginal vectors $\Phi_t \in \mathbb{R}^n$ directly, but instead has measurements of the form $r = G\Phi$, where $r \in \mathbb{R}^m$, with $m < n$, is a vector of available information, and where $G \in \mathbb{R}^{m \times n}$ represents a mapping from the full state information to the partial information. For example, G may represent a discretization of the push-forward operator $F_\#$ in the case of an OMT problem with dynamics detailed by (12), or of the operator Γ, mapping spectra to covariance matrices. Correspondingly, r may be a discretization of the push-forward measure, or the (vectorized) covariance matrix, respectively. Note that all quantities in this section are assumed to be real-valued. As the considered covariance matrices R are, in general, complex, we arrive at the equivalent real-valued problems by constructing the corresponding information vectors r as

$$r = \begin{bmatrix} \mathfrak{Re}(\operatorname{vec}(R))^T & \mathfrak{Im}(\operatorname{vec}(R))^T \end{bmatrix}^T, \tag{33}$$

where $\mathfrak{Re}(\cdot)$ and $\mathfrak{Im}(\cdot)$ denote the real and imaginary parts, respectively. The discretizations of the operators Γ, i.e., G, are then structured as to be consistent with this construction. It should be noted that the solution algorithm presented in this section coincides with the multi-marginal Sinkhorn iterations in (32) for the special case of full information (corresponding to letting G be the identity operator) and noiseless observations of the distribution marginals ($\gamma_t = \infty$).

Consider the multi-marginal OMT problem in (16), where only noisy and partial measurements of the marginals are available. A direct discretization of this problem, with added entropy regularization, may be expressed as

$$\underset{\mathbf{M}, \Delta_t}{\text{minimize}} \quad \langle \mathbf{C}, \mathbf{M} \rangle + \epsilon \mathcal{D}(\mathbf{M}) + \sum_t \gamma_t \|\Delta_t\|_2^2 \tag{34a}$$

$$\text{subject to} \quad G_t P_t(\mathbf{M}) = r_t + \Delta_t \quad \text{for } t = 0, \dots, \mathcal{T}, \tag{34b}$$

where $G_t \in \mathbb{R}^{m_t \times n}$ is a mapping from the full state information to the partial information $r_t \in \mathbb{R}^{m_t}$, as described above. Also, to allow for noisy measurements r_t, the problem is augmented by perturbation vectors $\Delta_t \in \mathbb{R}^{m_t}$ by direct analog to the matrix perturbations of (16), which are penalized by the squared ℓ_2-norm scaled by the penalty parameters $\gamma_t > 0$. It may be noted that due to the presence of the entropy term, $\mathcal{D}$, the problem in (34) is strictly convex. Also, for finite γ_t, a solution always exists due to the introduction of the perturbation vectors Δ_t. Altogether, this implies that for any $\epsilon > 0$ and $\gamma_t < \infty$, the optimization problem (34) has a unique solution. The following proposition characterizes the solution $\mathbf{M}$ in terms of the variables of the Lagrange dual problem, yielding the dual formulation.

Proposition 1. *The optimal solution to (34) may be expressed as* $\mathbf{M} = \mathbf{K} \odot \mathbf{U}$, *for two tensors* $\mathbf{K}, \mathbf{U} \in \mathbb{R}^{n^{\mathcal{T}+1}}$, *given by* $\mathbf{K} = \exp(-\mathbf{C}/\epsilon)$, *and*

$$\mathbf{U} = u_0 \otimes u_1 \otimes \cdots \otimes u_{\mathcal{T}}, \quad \iff \mathbf{U}_{i_0,\dots,i_{\mathcal{T}}} = \prod_{t=0}^{\mathcal{T}} (u_t)_{i_t}, \tag{35}$$

where the vectors u_t *are given by* $u_t = \exp(G_t^T \lambda_t / \epsilon)$ *for* $t = 0, \dots, \mathcal{T}$, *with* $\lambda_t \in \mathbb{R}^{m_t}$ *denoting the Lagrange dual variable corresponding to the constraint* t *in (34b). The optimal perturbation vectors may be expressed as* $\Delta_t = -\frac{1}{2\gamma_t}\lambda_t$, *for* $t = 0, 1, \dots, \mathcal{T}$. *Furthermore, a Lagrange dual of the multi-marginal OMT problem (34) is given by*

$$\underset{\lambda_0,\dots,\lambda_{\mathcal{T}}}{\text{maximize}} \; -\epsilon \, \langle \mathbf{K}, \mathbf{U} \rangle - \sum_{t=0}^{\mathcal{T}} \frac{1}{4\gamma_t} \|\lambda_t\|_2^2 + \sum_{t=0}^{\mathcal{T}} \lambda_t^T r_t. \tag{36}$$

where $\mathbf{U}$ *is given by (35) with* $u_t = \exp(G_t^T \lambda_t / \epsilon)$ *for* $t = 0, \dots, \mathcal{T}$.

Proof. See Appendix 1. □

With the result from Proposition 1, we are now ready to state the method for solving (34). Based on [33], and similarly to [23], we propose to solve the dual (36) by a block coordinate ascent.

Theorem 1. *Given an initial set of vectors $\lambda_0, \dots, \lambda_{\mathcal{T}}$, iterate the following steps repeatedly for $t \in \{0, \dots, \mathcal{T}\}$:*

- *Let*

$$v_t = P_t(\mathbf{K} \odot \mathbf{U})./u_t \tag{37}$$

where $\mathbf{U} = u_0 \otimes u_1 \otimes \cdots \otimes u_{\mathcal{T}}$ for the vectors $u_t = \exp(G_t^T \lambda_t/\epsilon)$.

- *Update the vector λ_t as the solution to*

$$G_t\left(v_t \odot \exp\left(\frac{G_t^T\lambda_t}{\epsilon}\right)\right) + \frac{\lambda_t}{2\gamma_t} - r_t = 0. \tag{38}$$

Then, the vectors $\lambda_0, \dots, \lambda_{\mathcal{T}}$ converge linearly to the unique optimal solution of the dual problem (36). Furthermore, in the limit point of the iteration, the marginals of $\mathbf{M}$ are given directly as $\Phi_t = u_t \odot v_t$, for $t = 0, 1, \dots, \mathcal{T}$.

Proof of Theorem 1. We first show that the iterates in the theorem correspond to a block coordinate ascent in the dual problem (36). A block-coordinate ascent method is to iteratively update one of the dual variables as to maximize the objective function, while keeping the other variables fixed. Here, this is to iteratively update one λ_t, for $t = 0, 1, \dots, \mathcal{T}$, such that it maximizes (36). The maximizing λ_t may be found as the root of the corresponding gradient. Therefore, note that substituting the explicit expression for the elements in $\mathbf{U}$, as detailed in Proposition 1, into the first term of the dual objective (36) yields

$$\sum_{i_0,\dots,i_{\mathcal{T}}} \mathbf{K}_{i_0,\dots,i_{\mathcal{T}}} \mathbf{U}_{i_0,\dots,i_{\mathcal{T}}} = \sum_{i_0,\dots,i_{\mathcal{T}}} \mathbf{K}_{i_0,\dots,i_{\mathcal{T}}} \prod_{t=0}^{\mathcal{T}} \left(\exp(G_t^T\lambda_t/\epsilon)\right)_{i_t}.$$

The gradient of this term with respect to λ_t can be written as

$$G_t(v_t \odot u_t),$$

with $u_t = \exp(G_t^T\lambda_t/\epsilon)$ and the vector v_t defined as

$$\begin{aligned}(v_t)_{i_t} &= \sum_{\substack{i_0,\dots,i_{t-1},\\ i_{t+1},\dots,i_{\mathcal{T}}}} K_{i_0,\dots,i_{t-1},i_t,i_{t+1},\dots,i_{\mathcal{T}}} \prod_{\substack{s=1\\ s\neq t}}^{\mathcal{T}} (u_s)_{i_s} \\ &= \left(\sum_{\substack{i_0,\dots,i_{t-1},\\ i_{t+1},\dots,i_{\mathcal{T}}}} \mathbf{K}_{i_0,\dots,i_{t-1},i_t,i_{t+1},\dots,i_{\mathcal{T}}} \prod_{s=0}^{\mathcal{T}} (u_s)_{i_s}\right)_{i_t} /(u_t)_{i_t} \\ &= (P_t(\mathbf{K}\odot\mathbf{U}))_{i_t}/(u_t)_{i_t}.\end{aligned}$$

Hence, the gradient of the dual objective function (36) with respect to λ_t is

$$-G_t(v_t \odot u_t) - \lambda_t/(2\gamma_t) + r_t.$$

Furthermore, note that the dual objective function (36) is strongly concave. This can be seen by noting that the first term is concave, the second term is strongly concave, and the last term is linear. Since the dual is twice continuously differentiable, the block coordinate ascent method converges (locally) q-linearly to the unique global maximum [10], Theorem 2. □

Remark 1. It is worth noting that, given the dual optimal variables λ_t, for $t = 0, 1, \dots, \mathcal{T}$, obtained as the limit point of the iterations in Theorem 1, the primal optimal variables may be recovered using Proposition 1.

Remark 2. The number of constraints in (34) may be smaller than the number of modes in the mass transfer tensor $\mathbf{M}$. In cases where there are no constraints on the tth projection of $\mathbf{M}$, the corresponding dual variable λ_t is set to zero. The method described in Theorem 1 may then be modified by setting $u_t = \mathbf{1}$ and only iterating over the smaller set of remaining vectors.

Algorithm 2 Sinkhorn-Newton method for the multi-marginal optimal mass transport with partial information of the marginals.

```
Given: Initial guess λ_t, for t = 0, ..., T; starting point t
u_t ← exp(G_t λ_t/ε) for t = 0, ..., T
while Sinkhorn not converged do
    Construct v_t according to (37), i.e.,
    v_t ← P_t(K ⊙ U)./u_t
    while Newton not converged do
        f ← G_t diag(v_t) u_t − r_t + 1/(2γ_t) λ_t
        df ← (1/ε) G_t diag(v_t ⊙ u_t) G_t^T + 1/(2γ_t) I
        Δλ ← −df\f
        λ_t ← λ_t + η Δλ, with η determined by a linesearch
        u_t ← exp(G_t λ_t/ε)
    end while
    t ← t + 1 (mod T + 1)
end while
return M = K ⊙ U
```

Recall that in the standard Sinkhorn iterations (31) and (32), the scaling factors u_t are iteratively updated as to satisfy the marginal constraints. The same property holds for the iterations in Theorem 1. Specifically, if λ_t is a solution to (38), then the implied transport plan in that iteration, i.e., $\mathbf{M} = \mathbf{K} \odot \mathbf{U}$, where $\mathbf{U} = u_0 \otimes u_1 \otimes \cdots \otimes u_{\mathcal{T}}$, with $u_t = \exp(G_t^T\lambda_t/\epsilon)$, satisfies

$$G_t P_t(\mathbf{M}) = r_t + \Delta_t = r_t - \frac{1}{2\gamma_t}\lambda_t.$$

In the case of full information and exact matching in a marginal (i.e., $G_t = I$ and $\gamma_t = \infty$), the update (38) in Theorem 1 reduces to the Sinkhorn iterations (Algorithm 1), i.e.,

$$u_t \leftarrow r_t./v_t = u_t \odot r_t./P_t(\mathbf{K}\odot\mathbf{U}),$$

whereas in the case with only full information (i.e., $G_t = I$ and $0 < \gamma_t < \infty$), the update reduces to solving

$$v_t \odot \exp(\lambda_t/\epsilon) + \lambda_t/(2\gamma_t) - r_t = 0. \tag{39}$$

This equation may be solved element-wise using the Wright omega function (see appendix B in [33]). This allows for computing the proximal operator of the regularized optimal mass transport [33], which is often used in first-order methods for non-smooth optimization. Similar expressions may also be obtained for other penalization terms than the squared ℓ_2-error, which relates to the entropic proximal operator [16]. Herein, we propose to solve (38) using Newton's method. Therefore, note that the corresponding Jacobian is given by

$$\frac{1}{\epsilon} G_t \operatorname{diag}(v_t \odot u_t) G_t^T + \frac{1}{2\gamma_t} I.$$

The full method for solving (34) is summarized in Algorithm 2. It may be noted that the multi-marginal transport plan $\mathbf{M}$ may for some problems be too large to be stored in the memory. However, when used in modeling applications, one is primarily interested in being able to study projections of $\mathbf{M}$, which, due to the structure of the transport plan, do not require $\mathbf{M}$ to be constructed explicitly.

Remark 3. Sufficiently close to the optimal solution, the quadratic approximation of the dual objective underlying the Newton method becomes increasingly accurate. The inner Newton method for solving (37) then converges in the first iteration. In the authors' experience, this is typically achieved within the first few outer Sinkhorn iterations.

The computational bottleneck of Algorithm 2 is the construction of the vectors v_t in (37), requiring the computation of $P_t(\mathbf{K}\odot\mathbf{U})$. Structures in the cost tensor $\mathcal{T}$ may allow for efficiently computing these projections, as we will see in Section 5. With the results

from Section 5, Theorem 1 provides an efficient scheme for implementing the spatial spectral estimators in Sections 3.1–3.3.

5. Efficient computations for structured multi-marginal OMT problems

In many applications, the size of the mass transport tensor $\mathbf{M}$, and thus of the auxiliary tensors $\mathbf{K}$ and $\mathbf{U}$, may be too large to manipulate directly. It is thus crucial to utilize additional structures in the problem whenever this is possible. In this section, some common examples for structures in the cost tensor $\mathbf{C}$ are described, allowing for efficient computation of subproblems in the multi-marginal Sinkhorn algorithm (Algorithm 1). The computational bottleneck of the Sinkhorn scheme, as well as the method for partial information of the marginals (Algorithm 2), is the computation of $P_t(\mathbf{K}\odot\mathbf{U})$, where both the elementwise multiplication $\mathbf{K}\odot\mathbf{U}$ and the application of the projection operator P_t may be expensive. However, in several cases of interest, the choice of cost function, c, induces structure in the tensor $\mathbf{K}$ that may be exploited in order to dramatically increase the efficiency of computing $P_t(\mathbf{K}\odot\mathbf{U})$. Such structures appears in e.g., barycenter problems [47], tracking problems [15,22], and certain hidden Markov chain formulations [28]. Here, we will consider cases relevant for the applications presented in Section 3. In fact, for these cases, the computation of $P_t(\mathbf{K}\odot\mathbf{U})$ may be performed by sequences of matrix-vector multiplications as detailed in the propositions in the following subsections.

5.1. Sequentially decoupling cost function

Consider the multi-marginal OMT problem on the form (15) where the cost function $\mathcal{C}:\mathcal{X}\times\mathcal{X}\to\mathbb{R}_+$ decouples sequentially according to (22), which may be utilized in order to model sequential tracking over time. The following proposition shows how this special case allows for efficient computations of the projections $P_t(\mathbf{K}\odot\mathbf{U})$ and $P_{t_1,t_2}(\mathbf{K}\odot\mathbf{U})$ in the discrete OMT problem.

Proposition 2. *Let the elements of the cost tensor $\mathbf{C}$ in (28) be of the form*

$$\mathbf{C}_{i_0,\dots,i_\mathcal{T}}=\sum_{t=1}^{\mathcal{T}}C_{i_{t-1},i_t},$$

for a cost matrix $C\in\mathbb{R}^{n\times n}$, and let $\mathbf{K}=\exp(-\mathbf{C}/\epsilon)$, and $\mathbf{U}=(u_0\otimes u_1\otimes\cdots\otimes u_\mathcal{T})$. Then, for $0\le t\le\mathcal{T}$,

$$\begin{aligned}P_t(\mathbf{K}\odot\mathbf{U}) = &\left(u_0^TK\mathrm{diag}(u_1)K\dots K\mathrm{diag}(u_{t-1})K\right)^T\odot u_t\\ &\odot(K\mathrm{diag}(u_{t+1})K\dots K\mathrm{diag}(u_{\mathcal{T}-1})Ku_\mathcal{T})\end{aligned}$$

where $K=\exp(-C/\epsilon)$, and, for $0\le t_1<t_2\le\mathcal{T}$,

$$\begin{aligned}P_{t_1,t_2}(\mathbf{K}\odot\mathbf{U}) = &\,\mathrm{diag}\left(u_0^TK\mathrm{diag}(u_1)K\cdots K\mathrm{diag}(u_{t_1-1})K\right)\\ &\odot\mathrm{diag}(u_{t_1})K\mathrm{diag}(u_{t_1+1})K\dots K\mathrm{diag}(u_{t_2})\\ &\odot\mathrm{diag}(K\mathrm{diag}(u_{t_2+1})K\dots K\mathrm{diag}(u_{\mathcal{T}-1})Ku_\mathcal{T}).\end{aligned}\tag{40}$$

Proof. See Appendix 1. □

It is worth noting that the computation of $P_t(\mathbf{K}\odot\mathbf{U})$ requires only $\mathcal{T}$ matrix-vector multiplications of the form Ku_τ, i.e., $\mathcal{O}(\mathcal{T}n^2)$ operations, whereas the complexity for a brute-force approach is $\mathcal{O}(n^{\mathcal{T}+1})$ operations. The same holds for the multiplication of $P_{t_1,t_2}(\mathbf{K}\odot\mathbf{U})$ with a vector. If several projections are computed iteratively, as, e.g., in Algorithm 1 or Algorithm 2, intermediate results of the factors to the left and right of u_t, for $t=0,\dots,\mathcal{T}$, may be stored. Then each iteration requires only one matrix vector multiplication, that is $\mathcal{O}(n^2)$ operations. Furthermore, if the domains of the marginals have cost functions that decouple then the efficiency of the matrix vector computations can be improved even further, as described in the following remark.

Remark 4. In many multidimensional problems, the domain X can be represented as a direct product $X=X_1\times X_2\times\cdots\times X_N$, with $|X_i|=n_i$. If, in addition, the cost decouples in the same way, i.e., for two points $\mathrm{x}_0=(\mathrm{x}_0^{(1)},\mathrm{x}_0^{(2)},\dots,\mathrm{x}_0^{(N)})\in X$ and $\mathrm{x}_1=(\mathrm{x}_1^{(1)},\mathrm{x}_1^{(2)},\dots,\mathrm{x}_1^{(N)})\in X$,

$$c(\mathrm{x}_0,\mathrm{x}_1)=\sum_{i=1}^{N}c_i\left(\mathrm{x}_0^{(i)},\mathrm{x}_1^{(i)}\right)$$

where $c_i:X_i\times X_i\to\mathbb{R}$ is a cost function in the i-th dimension, for $i=1,\dots,N$, then the matrix $K=\exp(-C/\epsilon)\in\mathbb{R}^{n\times n}$, where $n=\prod_{i=1}^{N}n_i$, can be decoupled as a tensor product $K=K_1\otimes K_2\otimes\cdots\otimes K_N$. Multiplications with K can then be done one dimension at a time. Practically, this can be efficiently implemented using specialized tensor toolboxes, e.g., with the command `tmprod` of the Matlab toolbox Tensorlab [59]. Thus, the multiplication Ku requires only $\mathcal{O}(n\sum_i n_i)$ operations instead of $\mathcal{O}(n^2)$ which is required when using standard multiplication. From a memory perspective, this is also important since K does not need to be computed and stored. In particular, this holds for a regular rectangular grid in $\mathbb{R}^N$ with cost function $c(x,y)=\|x-y\|_p^p$, for some $p\ge1$, which is the typical setup for the Wasserstein metric.

5.2. Centrally decoupling cost function

In order to solve OMT barycenter formulations, we may use centrally decoupled cost functions $\mathcal{C}$ according to (19) to model transport between a common barycenter and several marginal distributions. The following proposition describes the projection computations for this case.

Proposition 3. *Let the elements of the cost tensor $\mathbf{C}$ in (28) be of the form*

$$\mathbf{C}_{i_0,\dots,i_J}=\sum_{j=1}^{J}C_{i_0,i_j},$$

for a cost matrix $C\in\mathbb{R}^{n\times n}$, and let $\mathbf{K}=\exp(-\mathbf{C}/\epsilon)$, and $\mathbf{U}=(u_0\otimes u_1\otimes\cdots\otimes u_\mathcal{T})$. Then,

$$P_0(\mathbf{K}\odot\mathbf{U})=u_0\odot\bigodot_{\ell=1}^{J}(Ku_\ell),\tag{41}$$

$$P_j(\mathbf{K}\odot\mathbf{U})=u_j\odot K^T\left(u_0\odot\bigodot_{\ell=1,\ell\neq j}^{J}(Ku_\ell)\right),\tag{42}$$

for $j=1,\dots,J$, where $K=\exp(-C/\epsilon)$. For $j,j_1,j_2=1,\dots,J$, and $j_1\neq j_2$, the pair-wise projections are given by

$$P_{0,j}(\mathbf{K}\odot\mathbf{U})=\mathrm{diag}\left(u_0\odot\bigodot_{\substack{\ell=1\\\ell\neq j}}^{J}(Ku_\ell)\right)K\mathrm{diag}(u_j),\tag{43}$$

$$P_{j_1,j_2}(\mathbf{K}\odot\mathbf{U})=\mathrm{diag}(u_{j_1})K^T\mathrm{diag}\left(p_{j_1,j_2}\right)K\mathrm{diag}(u_{j_2}),\tag{44}$$

where

$$p_{j_1,j_2}=u_0\odot\bigodot_{\substack{\ell=1\\\ell\neq j_1,j_2}}^{J}(Ku_\ell).$$

Proof. See Appendix 1. □

Similarly to the setting of a sequentially decoupling cost, the computation of a projection $P_j(\mathbf{K}\odot\mathbf{U})$, for any $j=0,\dots,J$, in the

case of a centrally decoupling cost, involves J matrix vector multiplications. It thus requires $\mathcal{O}(Jn^2)$ operations, as opposed to $\mathcal{O}(n^{J+1})$ in the brute force approach. For the iterative schemes in Algorithms 1 and 2, storing intermediate results decreases the complexity in each iteration to at most two matrix vector multiplications, i.e., $\mathcal{O}(n^2)$ operations.

5.3. Sequential and central decoupling

Consider a cost function $\mathcal{C}$ being a combination of sequentially and centrally decoupled costs, according to (25), which in Section 3.3 was applied to modeling tracking of barycenters over time. For this case, the following proposition holds.

Proposition 4. *Let the elements of the cost tensor* $\mathbf{C}$ *in (28) be of the form*

$$\mathbf{C}_{(i_{(t,j)}|(t,j)\in\Lambda)} = \sum_{t=1}^{\mathcal{T}} C_{i_{(t-1,0)},i_{(t,0)}} + \sum_{t=1}^{\mathcal{T}}\sum_{j=1}^{J} \tilde{C}_{i_{(t,0)},i_{(t,j)}},$$

for cost matrices $C\in\mathbb{R}^{n\times n}$ *and* $\tilde{C}\in\mathbb{R}^{n\times\tilde{n}}$, *and define* $\Lambda=\{(t,j)\,|\,t\in\{0,1,\dots,\mathcal{T}\}, j\in\{0,1\dots,J\}\}$. *Furthermore, let* $\mathbf{K}=\exp(-\mathbf{C}/\epsilon)$, *and let* $\mathbf{U}=\bigotimes_{(t,j)\in\Lambda} u_{(t,j)}$. *Then, for the central marginals, corresponding to* (t,j) *such that* $j=0$,

$$\begin{aligned} P_{(t,0)}(\mathbf{K}\odot\mathbf{U}) = {} & \left(p_0^T K\mathrm{diag}(p_1)K\dots K\mathrm{diag}(p_{t-1})K\right)^T \odot p_t \\ & \odot(K\mathrm{diag}(p_{t+1})K\dots K\mathrm{diag}(p_{\mathcal{T}-1})Kp_{\mathcal{T}}), \end{aligned} \tag{45}$$

where $p_t = u_{(t,0)}\odot\bigodot_{j=1}^{J}\tilde{K}u_{(t,j)}$ *and* $K=\exp(-C/\epsilon)$, $\tilde{K}=\exp(-\tilde{C}/\epsilon)$. *Furthermore, for the non-central marginals, i.e.,* (t,j) *with* $j=1,2,\dots,J$, *it holds that*

$$\begin{aligned} P_{(t,j)}(\mathbf{K}\odot\mathbf{U}) = {} & u_{(t,j)}\odot\tilde{K}^T\Big((p_0^T K\mathrm{diag}(p_1)K\dots K\mathrm{diag}(p_{t-1})K)^T \\ & \odot\left(p_t./(\tilde{K}u_{(t,j)})\right) \\ & \odot(K\mathrm{diag}(p_{t+1})K\dots K\mathrm{diag}(p_{\mathcal{T}-1})Kp_{\mathcal{T}})\Big). \end{aligned} \tag{46}$$

Proof. See Appendix 1. □

Computing one vector p_t, for any $t=0,\dots,\mathcal{T}$, requires J matrix vector multiplications, i.e., $\mathcal{O}(Jn\tilde{n})$ operations. In order to compute the projection $P_{(t,0)}(\mathbf{K}\odot\mathbf{U})$, a total of $\mathcal{T}+1$ vectors of this form have to be computed, as well as $\mathcal{T}$ aditional matrix vector products of complexity $\mathcal{O}(n^2)$ each. This results in a total of $\mathcal{O}(\mathcal{T}Jn\tilde{n}+\mathcal{T}n^2)$ operations. The complexity is the same for the other projections, $P_{(t,j)}(\mathbf{K}\odot\mathbf{U})$, for $j=1,\dots,J$, and significantly lower than the direct computation of the projections, which requires $\mathcal{O}(n^{(J+1)(\mathcal{T}+1)})$ operations. It is worth noting that in an iterative scheme such as Algorithm 1 or Algorithm 2, when cycling through $j=0,1,\dots,J$, for a given t, one does not have to re-compute p_τ for $\tau\neq t$, i.e., the multiplicative factors to the left and right of $p_t./(Ku_{(t,j)})$ in (46) have to be updated only once for each t. Furthermore, as noted in Section 5.1, these factors may be stored so that updating them for a given t requires only one multiplication with K. Moreover, as noted in Section 5.2, the updates of p_t while cycling through $j=0,\dots,J$ involve at most two multiplications with $\tilde{K}$ in each iteration. Thus, computing the projections $P_{(t,j)}(\mathbf{K}\odot\mathbf{U})$, for all $j=0,\dots,J$ and a fixed t, requires $\mathcal{O}(n^2+Jn\tilde{n})$ operations.

Remark 5. The pairwise projections for index pairs $(t_1,0)$ and $(t_2,0)$, i.e., indices corresponding to the central marginals, may be computed analogously to Proposition 2. That is,

$$\begin{aligned} P_{(t_1,0),(t_2,0)}(\mathbf{K}\odot\mathbf{U}) = {} & \mathrm{diag}\left(p_0^T K\mathrm{diag}(p_1)K\dots K\mathrm{diag}(p_{t_1-1})K\right) \\ & \odot\mathrm{diag}(p_{t_1})K\mathrm{diag}(p_{t_1+1})K\cdots K\mathrm{diag}(p_{t_2}) \\ & \odot\mathrm{diag}(K\mathrm{diag}(p_{t_2+1})K\dots K\mathrm{diag}(p_{\mathcal{T}-1})Kp_{\mathcal{T}}). \end{aligned}$$

In case the index pairs are (t,j_1) and (t,j_2), i.e., when the indices correspond to the same central marginal, the pairwise projections can be computed similarly to Proposition 3. Specifically, the bi-marginal projection for a central marginal and a non-central marginal corresponding to the same time t is given by

$$P_{(t,0),(t,j)}(\mathbf{K}\odot\mathbf{U}) = \mathrm{diag}(u_{(t,0)})\mathrm{diag}\left(\tilde{K}^T\rho_{(t,0),(t,j)}\right)\tilde{K}\mathrm{diag}(u_{(t,j)}),$$

where

$$\begin{aligned} \rho_{(t,0),(t,j)} = {} & \left(p_0^T K\mathrm{diag}(p_1)K\dots K\mathrm{diag}(p_{t-1})K\right)^T \\ & \odot\left(p_t./(u_{(t,0)}\odot\tilde{K}u_{(t,j)})\right) \\ & \odot(K\mathrm{diag}(p_{t+1})K\dots K\mathrm{diag}(p_{\mathcal{T}-1})Kp_{\mathcal{T}}), \end{aligned}$$

whereas the bi-marginal projection for two non-central marginals corresponding to the same time t may be expressed as

$$P_{(t,j_1),(t,j_2)}(\mathbf{K}\odot\mathbf{U}) = \mathrm{diag}(u_{(t,j_1)})\tilde{K}^T\mathrm{diag}(\rho_{(t,j_1),(t,j_2)})\tilde{K}\mathrm{diag}(u_{(t,j_2)}),$$

where

$$\begin{aligned} \rho_{(t,j_1),(t,j_2)} = {} & \left(p_0^T K\mathrm{diag}(p_1)K\dots\mathrm{diag}(p_{t-1})K\right)^T \\ & \odot\left(p_t./\left(\tilde{K}u_{(t,j_1)}\odot\tilde{K}u_{(t,j_2)}\right)\right) \\ & \odot(K\mathrm{diag}(p_{t+1})\dots K\mathrm{diag}(p_{\mathcal{T}-1})Kp_{\mathcal{T}}). \end{aligned}$$

For arbitrary index pairs (t,j), expressions for the pairwise projections may be derived in a similar way.

6. Numerical results

In this section, we demonstrate the behavior of the proposed multi-marginal OMT problem in (16) when applied to the discussed spatial spectral estimation problems. Throughout, the problem in (16) is approximated by the discrete counterpart in (34), and solved using Algorithm 2. For the implementation of the projection operators, P_t, we use the results from the propositions in Section 5. For the example in Section 6.4, we provide a dimensionality analysis of the corresponding OMT problem, illustrating the benefit of the presented computational tools.

6.1. Tracking with static and dynamical models

In order to illustrate the different properties of the static and dynamic OMT formulations, we consider a DoA tracking example with two moving signal sources. Specifically, we consider a ULA consisting of 5 sensors with half-wavelength spacing, measuring the superposition of the source signals, modeled as independent Gaussian processes, together with a spatially white complex Gaussian sensor noise. The signal to noise ratio (SNR), defined as

$$\mathrm{SNR} = 10\log_{10}\left(\sigma^2_{signal}/\sigma^2_{noise}\right),$$

where σ^2_{signal} and σ^2_{noise} are the signal and noise powers, respectively, is 10 dB. The trajectories of the two sources are displayed in the top panel of Fig. 1, with the bottom panel showing the velocities. At six different time instances, evenly spaced in time throughout the trajectories, we collect 25 array signal snapshots, from which the array covariance is estimated using the sample covariance matrix. Using the resulting sequence of covariance matrices, we then attempt to reconstruct the target trajectories using the static and dynamic OMT problems. For the static OMT problem, we use the cost function $c(\theta,\varphi)=|\theta-\varphi|^2$, for $\theta,\varphi\in(-\pi,\pi]$. For the dynamical model, we introduce a latent velocity state, v, such that $\dot{\theta}(t)=v(t)$, and use the state space representation in (12), where

$$A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix},\quad B=\begin{bmatrix}0 & 1\end{bmatrix}^T,\quad F=\begin{bmatrix}1 & 0\end{bmatrix},$$

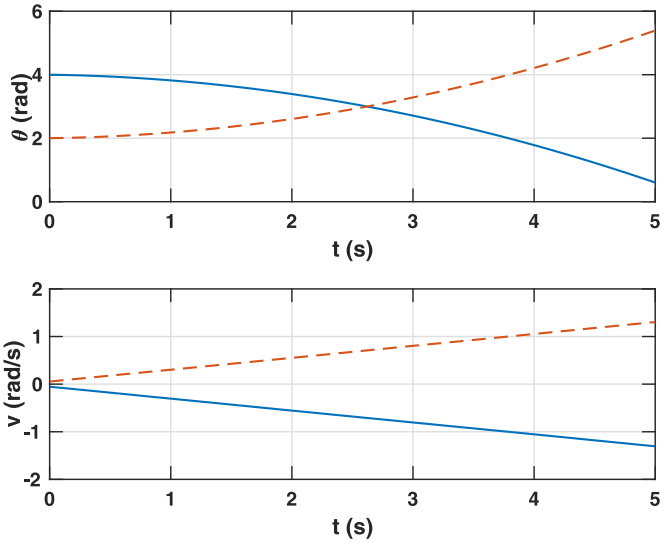

Fig. 1. Ground truth of the DoA tracking example with two moving targets. Top panel: the target DoAs, in radians, as a function of time. Bottom panel: the target velocities, in radians per second, as a function of time.

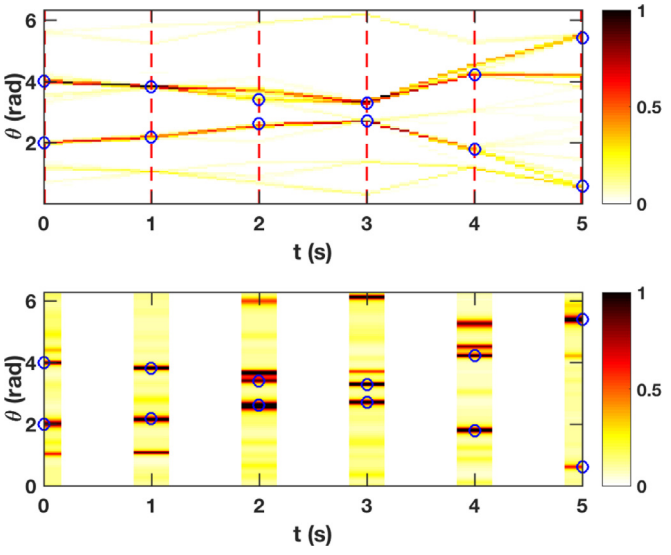

Fig. 2. Top panel: reconstructed DoA spectrum using a static cost function in the OMT formulation, based on six observations of the array covariance matrix. The observation times are marked in dashed lines. The ground truth trajectories at the observation times are marked by rings. Bottom panel: estimates obtained using the Capon estimator applied to the individual covariance matrices.

with the state vector being formed as $x(t) = \begin{bmatrix}\theta(t) & v(t)\end{bmatrix}^T$, i.e., the same model as in Example 1. Note here that the F matrix reflects the fact that only the angle, θ, is directly observable in the array covariance matrix. The choice of B implies that the angle may only be influenced via the velocity, i.e., through acceleration. Thus, the resulting dynamic cost function, as defined in (8), penalizes transport requiring acceleration, whereas the cost function for the static problem in contrast penalizes transport requiring velocity. In the discrete implementations of the methods, we use 100 grid points to represent the angle, θ, and the dynamical model uses 30 grid points for the velocity, v. Also, we use the regularization parameters $\epsilon = 10^{-1}$, and common parameters $\gamma = 30$ and $\gamma = 5$ for all marginals for the static and dynamic OMT models, respectively.

The results for the static and dynamical models are displayed in the top panels of Figs. 2 and 3, respectively. Here, the trajectories in between the observation times, which are indicated by vertical

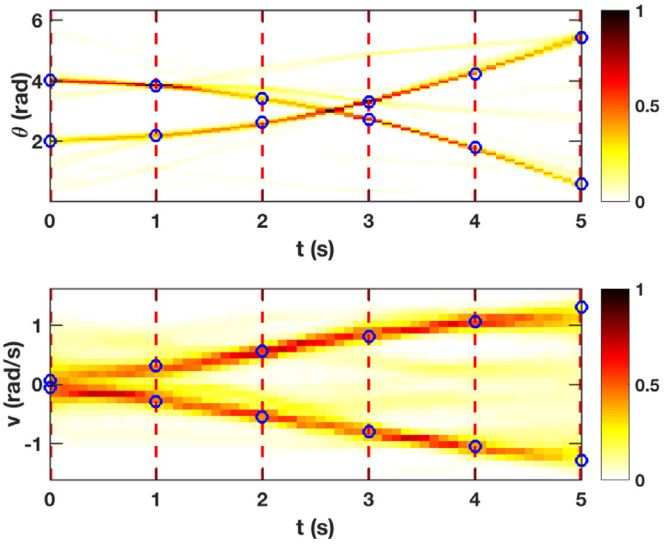

Fig. 3. Reconstructed DoA and velocity spectra using a dynamic cost function in the OMT formulation, based on six observations of the array covariance matrix. The observation times are marked in dashed lines. The ground truth trajectories at the observation times are marked by rings. Top panel: reconstructed DoA spectrum. Bottom panel: reconstructed velocity spectrum.

dashed lines, are reconstructed using the interpolation procedure presented in Section 2.2, as defined by the obtained multi-marginal transport plan. The required bi-marginal transport plans describing the mass transfer between the respective margins, i.e., consecutive observation points, are computed as the bi-marginal projections detailed in Proposition 2. To simplify comparison, the ground truth trajectories from Fig. 1, at the observation times, are superimposed in the plots. Also, the bottom panel of Fig. 2 displays the spectral estimates obtained by applying the Capon estimator [12] to the individual covariance matrix estimates. As may be noted, the Capon estimate contains several spurious peaks, caused by the noisy measurements. In contrast, as seen in Fig. 2, the static model is able to produce reasonable spectral estimates for the observation times. However, the reconstructed trajectories fail to model the crossing of the paths of the targets. This is not unexpected: as the static OMT model penalizes movement, i.e., velocity, the cost of transport is smaller if the targets instead change course (note the trajectory between $t = 2$ and $t = 3$). In contrast, the dynamical formulation, in which movement is expected, is able to produce considerably more accurate estimates, as may be seen in Fig. 3, including the crossing of the paths of the targets. Note also that the dynamical formulation is able to reconstruct the spectrum also for the hidden velocity state.

6.2. Sensor fusion in 3-D - audio example

In order to illustrate the applicability of the multi-marginal OMT formulation in (16), in the form of (20), for performing sensor fusion, we consider a three-dimensional (3-D) localization problem, in which two sensor arrays observe two signal sources. The sources are modeled as localized speech sources, with the source signals being taken from babble noise excerpts from [7]. The arrays consist of ten sensors each, arranged as a ULA for the first array, and as points on a circle, with two additional sensors perpendicular to the plane of the circle, for the second array. The sensor spacing for the ULA is 0.1 meters, and the radius for the circular array is 0.25 meters.

The scenario is shown in Fig. 4. The measured sensor signals are generated using acoustic impulse responses obtained from the randomized image method [18], using the room dimensions

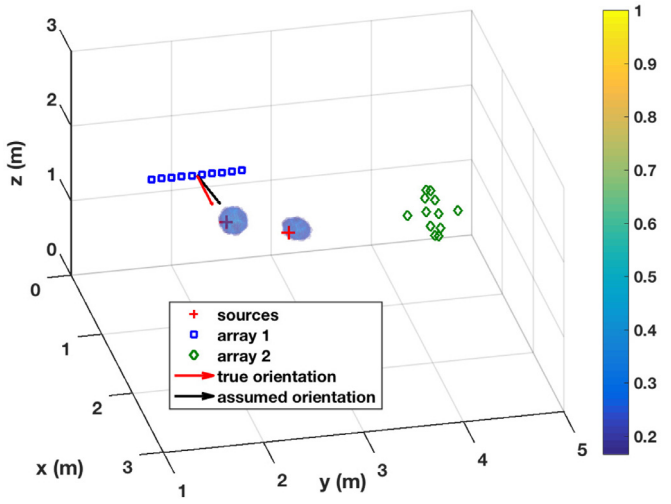

Fig. 4. Scenario for localization in 3D using two sensor arrays and two signal sources. The true and assumed orientations of the first array are indicated by arrows. The three dimensional spectral estimate obtained using the formulation in (20) is superimposed.

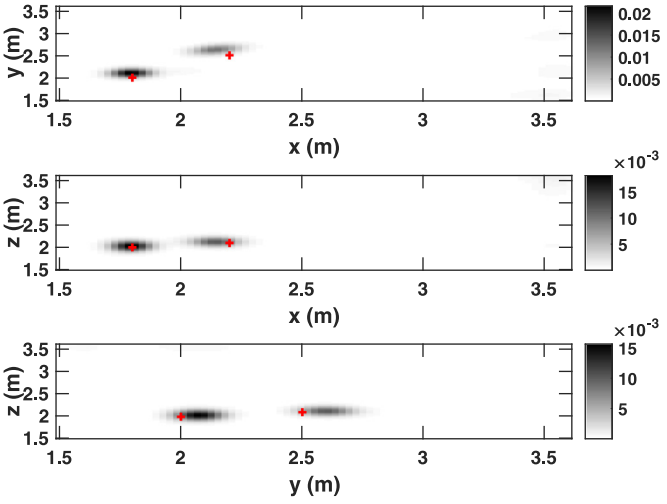

Fig. 5. Two-dimensional projections of the three-dimensional spectral estimate obtained using the formulation in (20) for the estimation scenario in Fig. 4.

$4.3 \times 6.9 \times 2.6$ meters. As the sources are broadband, we consider processing in the short-time Fourier transform (STFT) domain, i.e., equivalent to narrow-band filtering of the signal. Specifically, the signal is sampled at 16 kHz, and the STFT representation consists of 256 frequency bins, where each frame is constructed using 16 ms of the signal, using a Hann window with 50% overlap. We then compute estimates of the covariance matrices for the respective arrays corresponding to the frequency 2437.5 Hz, i.e., for the wavelength 0.1395 m, using the sample covariance estimate constructed from 100 signal snapshots. The proposed barycenter method in (20) is then used to form a joint spatial spectral estimate using the two estimated covariance matrices, with the cost function $c(x_0, x_1) = \|x_0 - x_1\|_2^2$, for $x_0, x_1 \in \mathbb{R}^3$. For the discrete implementation in (34), we use a uniform gridding of the cube $[1.5, 3.5]^3$, using $n = 75$ points in each dimension. The regularization parameters are $\gamma = 1$, common for all marginals, and $\epsilon = 10^{-2}$. Also, in order to illustrate the geometrical properties of the proposed formulation, we assume that the geometry of the ULA is only approximately known. Specifically, the assumed array geometry corresponds to a rotation in the $x-y$ plane of the true array. The obtained spectral estimate is superimposed in Fig. 4. As can be seen, the obtained estimate implies a spatial translation of the actual sources. However, it may be noted that the obtained estimate clearly identifies two targets, i.e., the error in the array orientation only results in a spatial perturbation, but no artifacts in the form of, e.g., spurious sources. This is also illustrated in Fig. 5, showing the projection of the three-dimensional spectrum onto the two dimensional subspaces. As can be seen from the second and third panel, the position in the z-coordinate is unbiased, as the rotation only takes place in the $x-y$ plane.

6.3. Sensor fusion 2-D – simulation study

As illustrated above, the proposed multi-marginal formulation is able to produce easily interpretable spectral estimates when used in 3-D sensor fusion scenarios, despite having erroneous information of the array geometry. Elaborating on this, we here conduct a Monte Carlo simulation study as to investigate the behavior of the spectral barycenter as a function of the array alignment error, and compare to other spectral estimation methods applicable to scenarios in which only the covariance matrices of individual sensor arrays, but not the inter-array covariances, are available. Specifically, consider a 2-D localization scenario, in which two uncorrelated sources impinge on two sensor arrays; one ellipsoidal shaped array consisting of 8 sensors, and one linear array consisting of 7 sensors. The wavefronts are here modeled as being circular. The scenario is shown in Fig. 6. As may be seen, similar to the three-dimensional example, an unknown rotation is introduced to the ellipsoidal array, here varying this rotation between 0 and 10°.

In the Monte Carlo simulation, 100 realizations are generated for each considered rotation angle. In each realization, the locations of the signal sources are randomized uniformly on the square $[-0.5, 0.5] \times [-0.5, 0.5]$. The sources are modeled as uncorrelated circularly symmetric Gaussian white noises with variance 100, and a spatially white noise with variance 1 is added to the sensor measurements. The wavelength of the impinging waves is twice that of the smallest sensor spacing in the linear array. We here consider spectral estimates obtained by the multi-marginal barycenter formulation in (20), as approximated by the discretization in (34). We also consider the formulation in (18) which is discretized and where the entropy regularization is employed separately for each pair-wise transport plan (see [23] for details on the implementation of this problem). For both methods, we utilize the cost function $c(\mathrm{x}_k, \mathrm{x}_\ell) = \|\mathrm{x}_k - \mathrm{x}_\ell\|_2^2$, for grid points $\mathrm{x}_k, \mathrm{x}_\ell \in \mathbb{R}^2$. We grid the square $[-1, 1]^2$ uniformly using $n = 100$ points in each dimension, and the algorithm parameters are $\gamma = 0.01$, common for all marginals, and $\epsilon = 10^{-3}$ and $\epsilon = 5 \cdot 10^{-3}$ for the multi-marginal and pair-wise formulations, respectively. The array covariance matrices are estimated using the sample covariance matrix from 500 signal snapshots. As comparison, we also apply the non-coherent MUSIC and MVDR estimators, as described in [49], as well as the least-squares (LS) estimator from [40], and the non-coherent SPICE estimator from [56] to the estimated covariance matrices in order to obtain estimates of the spatial spectrum. Also included are estimates produced by the robust estimators RARE [51], extended to near-field localization [61], as well as the method introduced in [2], here referred to as AW19. It should here be noted that these robust estimators rely on knowledge of the joint covariance matrix of the sensor arrays, and have therefore, in contrast to the proposed method as well as the other comparison methods, been provided with the inter-array cross-covariances.

Figs. 6–8 provide illustrations of the behavior of the considered methods for the case of 6.7° misalignment, for the multi-marginal formulation, the pair-wise regularized barycenter, and the MUSIC estimator, respectively. Comparing the two barycenter formulations, it may be noted that the estimate provided by

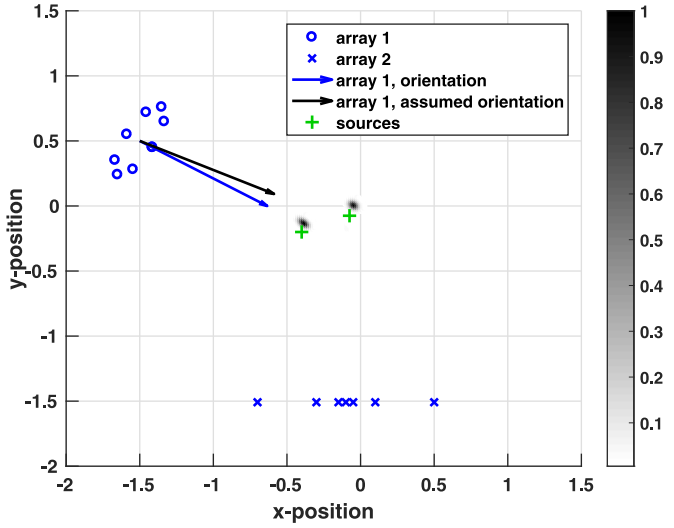

Fig. 6. Spectral estimate as given by the multi-marginal barycenter formulation in (20). The alignment error is 6.7°.

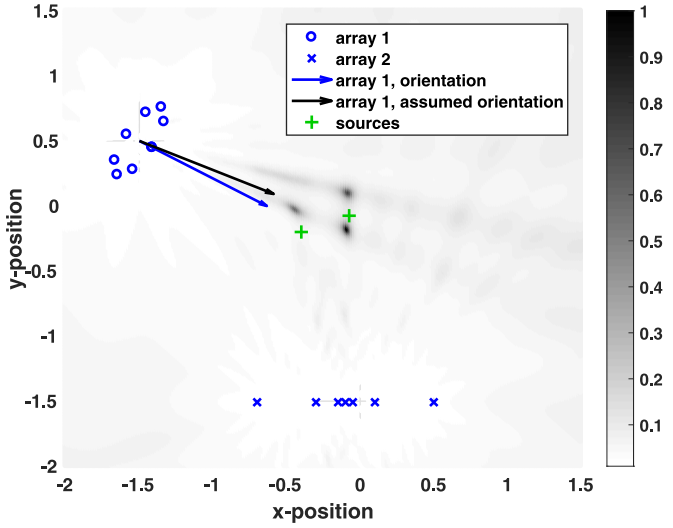

Fig. 8. Pseudo-spectrum as given by non-coherent MUSIC. The alignment error is 6.7°.

the multi-marginal representation is considerably more concentrated, as compared to the formulation with pair-wise regularization. This is due to the optimization problem in (34) being more well-conditioned, allowing for smaller entropy regularization while still retaining numerical stability. It should however be noted that the values of the parameter ϵ does not have exactly the same meaning for the two formulations, i.e., the problems are not equivalent even for identical parameter values. It may also be noted that the spectral estimates obtained using both barycenter formulations only imply a spatial perturbation of the signal sources. In contrast, the pseudo-spectral estimates obtained using the MUSIC estimator contain spurious peaks, in addition to larger deviations from the true source locations.

The results from the full simulation study are shown in Fig. 9, displaying the root mean squared error (RMSE) for the deviation of the spectral peaks to the true source locations. It may here be noted that for the case of no misalignment, the barycenter formulations produce estimates deviating slightly more from the ground truth than the comparison methods. However, as the misalignment increases, the RMSE for the barycenter estimators increase more slowly, indicating a greater robustness. This is not unexpected, as the formulations in (20) and (18) allows for perturbations on the underlying domain, whereas the optimization criteria for the LS and SPICE estimators are related to L_2-distances for the spectra. It may be noted that the robust RARE and AW19 estimators perform similarly to the other comparison methods, arguably due to the quite substantial calibration errors introduced by the array rotations. The sensitivity of RARE to the presence of array rotations and not only pure translations, although for the far-field case, has earlier been noted in [21].

6.4. Tracking of barycenters

As an illustration of the barycenter tracking formulation in Section 3.3, we consider a 2-D localization scenario in which two ULAs, each consisting of 15 sensors, measure signals generated by three moving sources. The scenario is illustrated in Fig. 10. Here, the start location of each source is indicated by an asterisk. As can be seen, the trajectories of the targets all intersect at some point. Note that one of the ULAs has been rotated slightly, as indicated

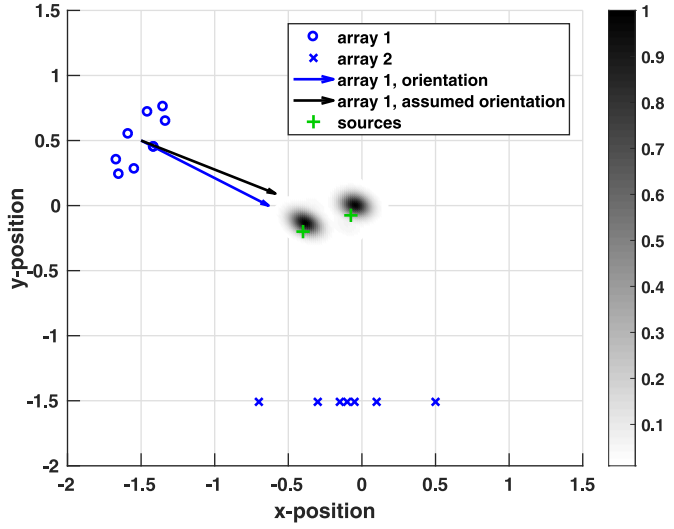

Fig. 7. Spectral estimate as given by the pair-wise barycenter formulation in (18). The alignment error is 6.7°.

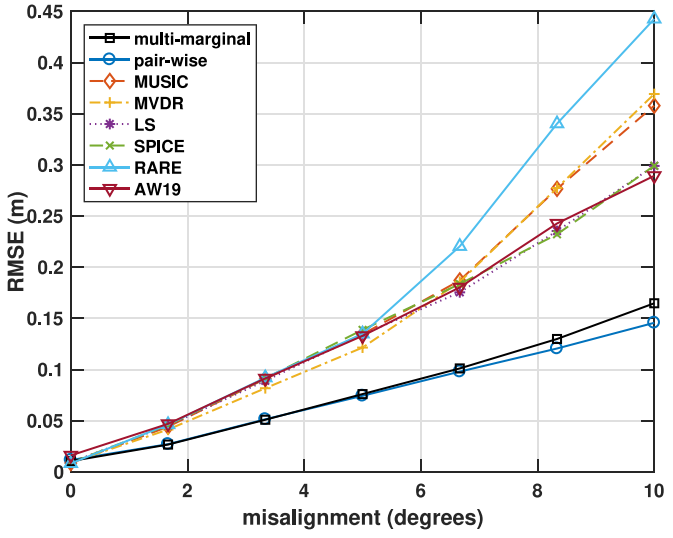

Fig. 9. Error in location of spectral peaks, as function of the misalignment angle.

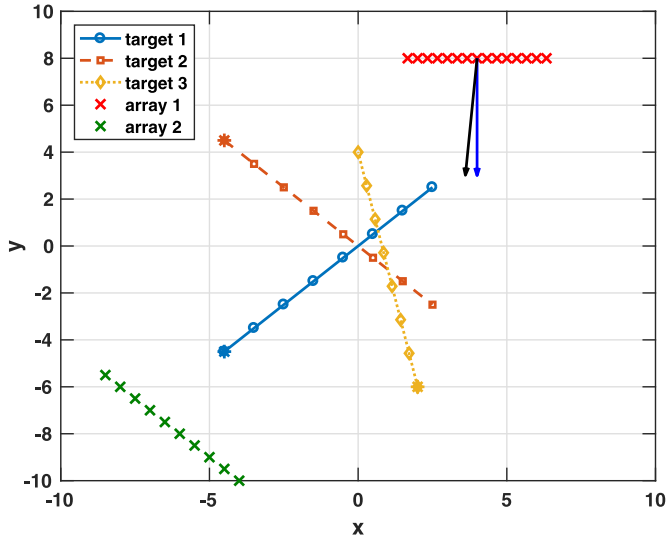

Fig. 10. Scenario for the barycenter tracking problem with three moving sources. The start of the trajectory for each source is indicated by an asterisk (*). Note here that one of the sensor arrays is slightly rotated, as indicated by the orientation arrows.

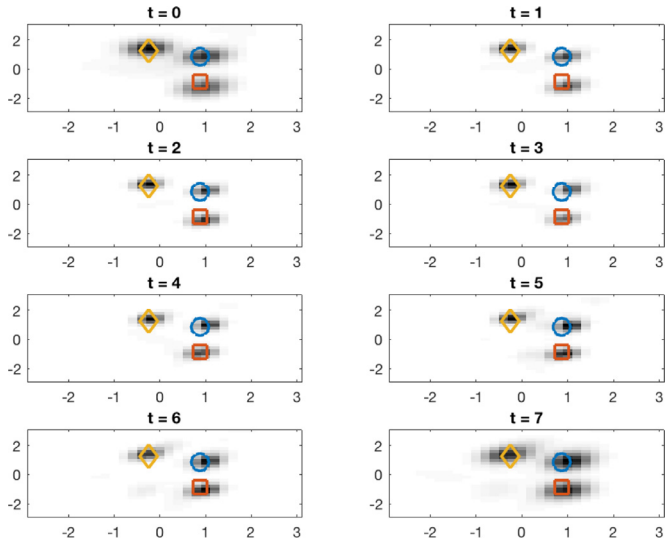

Fig. 12. Estimated velocity spectra corresponding to the time points $t = 0, 1, \ldots, \mathcal{T}$, with $\mathcal{T} = 7$, for the barycenter tracking problem. Note that the ground truth locations have been superimposed, using the same legend as in Fig. 10.

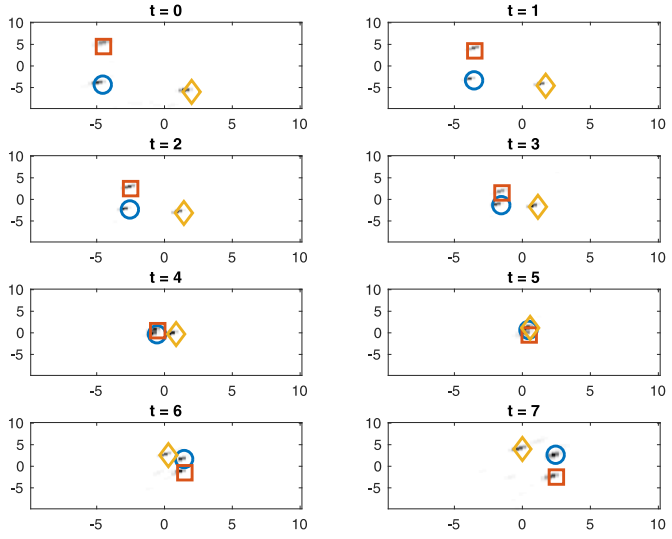

Fig. 11. Estimated spatial spectra corresponding to the time points $t = 0, 1, \ldots, \mathcal{T}$, with $\mathcal{T} = 7$, for the barycenter tracking problem. Note that the ground truth locations have been superimposed, using the same legend as in Fig. 10.

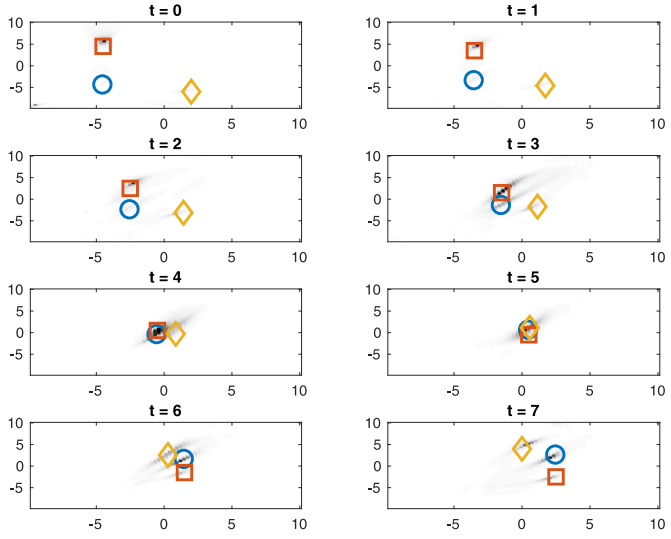

Fig. 13. Estimated spatial spectra corresponding to the time points $t = 0, 1, \ldots, \mathcal{T}$, with $\mathcal{T} = 7$, obtained using the non-coherent MVDR estimator. Note that the ground truth locations have been superimposed, using the same legend as in Fig. 10.

by the orientational arrows, motivating the use of a barycenter formulation in order to form an estimate of the spatial spectrum. At $\mathcal{T}+1$ time points t, $t = 0, 1, \ldots, \mathcal{T}$, with $\mathcal{T} = 7$, we collect 100 signal snapshots which are used to estimate the covariance matrices of the two arrays using the sample covariance matrix. The target signals are modeled as independent Gaussian sources, and spatially white Gaussian noise is added to the sensors, with an SNR of 20 dB. In order to model the tracking part, we use a two-dimensional extension of the state space description utilized in Section 6.1, where each of the two spatial components is endowed with a velocity state. With this description, the barycenter tracking problem considers transport on a four-dimensional space, i.e., over location and velocity for each spatial dimension. As the cost of transport between the barycenter at the observation times and the corresponding observation marginals, we use the Euclidean cost, i.e., $c(\mathbf{x}_0, \mathbf{x}_1) = \|\mathbf{x}_0 - \mathbf{x}_1\|_2^2$ for $\mathbf{x}_0, \mathbf{x}_1 \in \mathbb{R}^2$. In forming the spectral estimates, we use the discretized version of the problem in (26) using the parameters $\epsilon = 0.2$, $\gamma = 0.5$, common for all marginals, and $\alpha = 1$. The spatial domain is gridded uniformly with $n_x = 75$ points in each dimension, and the velocity domain is gridded uniformly with $n_v = 30$ points in each dimension.

The resulting estimated spatial spectra are shown in Fig. 11. As can be seen, the modes of the estimate correspond well to the ground truth, taking into account that the array rotation prevents perfect estimates. Note here that the rotation only causes a slight shift in location of the sources, but no spurious estimates. The corresponding estimated velocity spectra are shown in Fig. 12. As can be seen, the distribution over velocity details three distinct modes that remain fairly constant over time, corresponding well to the ground truth constant velocity. As comparison, Fig. 13 displays the results obtained using the non-coherent MVDR estimator from [49]. Note that this estimator does not take any time dependence into account, and instead forms estimates at each separate time point t using the available pair of array covariance matrices. It may be noted that these estimates display less concentrated estimates, as well as big differences in spectral power.

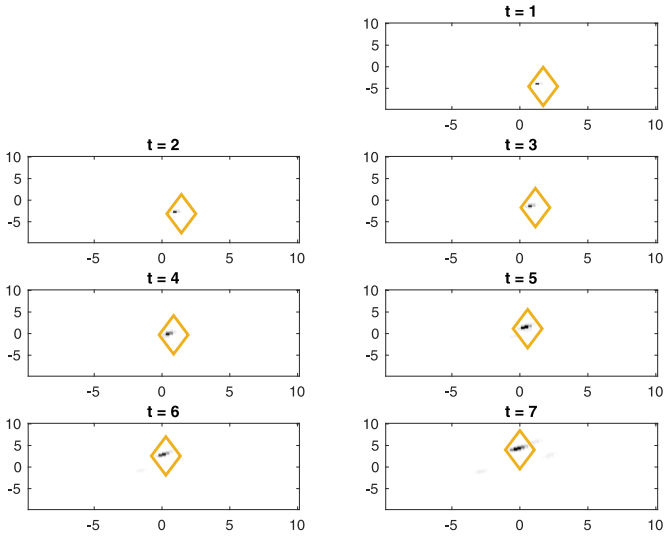

Fig. 14. Tracking of the mass of the peak location corresponding to the third target at time $t = 0$.

Finally, in order to illustrate how mass is transported on the 2-D spatial domain, Fig. 14 tracks the location of the spectral peak corresponding to the third target along the trajectory, identified by computing the bi-marginal projections corresponding to consecutive barycenters, as detailed in Proposition 2 and Remark 5. Note here that the distribution of mass remains fairly concentrated throughout the trajectory.

6.4.1. Dimensionality and complexity analysis

In this section, we provide a brief dimensionality analysis of the OMT problem used to model the barycenter tracking above, as well an analysis of the computational complexity of solving the dual problem in (36) using Algorithm 2. Recall that the state space consists of a location and velocity component for each of the two spatial dimensions. Thus, the size of the discrete state space X is $N = (n_x n_v)^d$, where $n_x = 75$ and $n_v = 30$ are the number of grid points on the spatial and velocity domain, respectively, and $d = 2$ is the number of spatial dimensions. Accordingly, the size of the discrete measurement space Y is $\tilde{N} = n_x^d$. As covariance matrices are estimated for $J = 2$ sensor arrays at $\mathcal{T}+1$ time instances, with $\mathcal{T} = 7$, this implies that the number of elements in the mass transport tensor $\mathbf{M}$ in (34) is

$$N^{(\mathcal{T}+1)}\tilde{N}^{(\mathcal{T}+1)J} \approx 4.3 \cdot 10^{113}.$$

This number is larger than the number of particles in the observable universe; solving the primal OMT problem directly, or even constructing a tensor of this size, is thus infeasible. However, the dual problem in (36) is formulated in terms of the dual vectors $\lambda_{(t,j)} \in \mathbb{R}^m$, for $t = 0, \dots, \mathcal{T}$, and $j = 1, \dots, J$. Here, $m = 225$ denotes the size of covariance vectors in (33), constructed by stacking the real and imaginary components of the corresponding covariance matrices, excluding the redundant parts resulting from the Hermitian structure. Thus, the number of real variables in the dual problem is given by $(\mathcal{T}+1)Jm = 3600$, constituting a dramatic complexity reduction as compared to the primal problem.

To analyze the computational complexity for solving the dual problem, we consider the complexity of performing one iteration sweep of Algorithm 2, i.e., one sweep through all index pairs (t, j), for $t = 0, \dots, \mathcal{T}$, and $j = 1, \dots, J$, as to update the corresponding dual vectors $\lambda_{(t,j)}$, where we run through j in the inner cycle and t in the outer cycle. It may be noted that for each index pair (t, j), this requires computing the vector $v_{(t,j)}$ in (37), followed by solving (38), as detailed in Theorem 1. The computation of $v_{(t,j)}$ is detailed in the projection in (46) in Proposition 4, disregarding the vector $u_{(t,j)}$. As noted in Section 5.3, computing the set of vectors $v_{(t,j)}$, for a fixed t and $j = 1, 2, \dots, J$, requires only one multiplication with $K \in \mathbb{R}^{N\times N}$, and $2J$ multiplications with $\tilde{K} \in \mathbb{R}^{N\times n_x^2}$. By exploiting the structures of K and $\tilde{K}$, as described in Remark 4, i.e., the decoupling in the spatial dimensions, the total complexity for performing the update of the left and right factors and computing the set $\{v_{(t,j)}\}_{j=1}^{J}$, is $\mathcal{O}(d(n_x n_v)^3 + Jdn_x^3 n_v)$. Finding the roots of (38) by Newton's method requires solving a system of linear equations of size $m \times m$ in each inner Newton iteration. However, as indicated in Remark 3, after a few outer Sinkhorn iterations, Newton's method in general converges directly, i.e., it suffices to solve a single system of equations in each outer iteration. Thus, after these initial outer iterations, the complexity for updating all dual variables $\lambda_{(t,j)}$, for $t = 0, 1, \dots, \mathcal{T}$, and $j = 1, 2, \dots, J$, is $\mathcal{O}(\mathcal{T}(d(n_x n_v)^3 + Jdn_x^3 n_v + Jm^3))$.

7. Conclusions

In this work, we have proposed a framework for formulating multi-marginal OMT problems for scenarios in which the underlying mass distributions are only indirectly observable, referred to as partial information of the marginals. Motivated by examples from spatial spectral estimation in array processing, we have shown that the proposed formulations may be used for modeling information fusion, as well as for tracking the evolution of mass distributions over time. By leveraging the geometrical properties of OMT, the proposed formulations have been shown to yield robust spectral estimates, as well as allowing for exploiting prior knowledge of underlying dynamics. Also, we have presented computational tools, leading to computationally efficient solution algorithms for the transport problem. Even though the original primal OMT formulation may be prohibitively large, we have shown that by considering dual formulations, as well as exploiting inherent structures in the problem, one may arrive at tractable solvers even in high-dimensional settings.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix 1. Proofs

In this section, we provide proofs of Propositions 1-4.

Proof of Proposition 1. As to simplify the exposition, let λ and Δ denote the sets $\lambda_0, \dots, \lambda_{\mathcal{T}}$ and $\Delta_0, \dots, \Delta_{\mathcal{T}}$, respectively. The Lagrangian of (34), with dual variable λ, is detailed as

$$L(\mathbf{M}, \Delta, \lambda) = \sum_{i_0,\dots,i_{\mathcal{T}}} \mathbf{C}_{i_0,\dots,i_{\mathcal{T}}} \mathbf{M}_{i_0,\dots,i_{\mathcal{T}}} + \epsilon \mathcal{D}(\mathbf{M}) + \sum_{t=0}^{\mathcal{T}} \gamma_t \|\Delta_t\|_2^2 + \sum_{t=0}^{\mathcal{T}} \lambda_t^T (r_t + \Delta_t - G_t P_t(\mathbf{M})). \quad (47)$$

For fixed λ, the Lagrangian is minimized when the gradients with respect to $\mathbf{M}$ and Δ vanish. For $\mathbf{M}$, this requires

$$\mathbf{C}_{i_0,\dots,i_{\mathcal{T}}} + \epsilon \log \mathbf{M}_{i_0,\dots,i_{\mathcal{T}}} - \sum_{t=0}^{\mathcal{T}} \left(\lambda_t^T G_t\right)_{i_t} = 0$$

$$\Leftrightarrow \mathbf{M}_{i_0,\dots,i_{\mathcal{T}}} = \exp(-\mathbf{C}_{i_0,\dots,i_{\mathcal{T}}}/\epsilon) \prod_{t=0}^{\mathcal{T}} \exp\left(G_t^T \lambda_t/\epsilon\right).$$

Thus, one may express the mass transport tensor as $\mathbf{M} = \mathbf{K} \odot \mathbf{U}$ for two sensors $\mathbf{K}, \mathbf{U} \in \mathbb{R}^{n^{\mathcal{T}+1}}$, defined as $\mathbf{K} = \exp(-\mathbf{C}/\epsilon)$ and

$$\mathbf{U} = u_0 \otimes u_1 \otimes \cdots \otimes u_{\mathcal{T}}, \quad \text{with } u_t = \exp\left(G_t^T \lambda_t/\epsilon\right).$$

This proves the first statement of the proposition. Further, the Lagrangian (47) is minimized with respect to Δ when

$$2\gamma_t \Delta_t + \lambda_t = 0 \quad \Rightarrow \quad \Delta_t = -\frac{1}{2\gamma_t}\lambda_t \quad \text{for } t = 0, 1, \ldots, \mathcal{T}.$$

Thus, the Langrangian (47), when minimized with respect to $\mathbf{M}$ and Δ, becomes

$$\min_{\mathbf{M},\Delta} L(\mathbf{M}, \Delta, \lambda) = -\epsilon \sum_{i_0,\ldots,i_{\mathcal{T}}} \mathbf{M}_{i_0,\ldots,i_{\mathcal{T}}} - \sum_{t=0}^{\mathcal{T}} \left(\frac{1}{4\gamma_t} \|\lambda_t\|_2^2 - \lambda_t^T r_t \right),$$

yielding the dual problem

$$\underset{\lambda_0,\ldots,\lambda_{\mathcal{T}}}{\text{maximize}} \; -\epsilon \sum_{i_0,\ldots,i_{\mathcal{T}}} \mathbf{K}_{i_0,\ldots,i_{\mathcal{T}}} \mathbf{U}_{i_0,\ldots,i_{\mathcal{T}}} - \sum_{t=0}^{\mathcal{T}} \frac{1}{4\gamma_t} \|\lambda_t\|_2^2 + \sum_{t=0}^{\mathcal{T}} \lambda_t^T r_t.$$

□

The proofs of Proposition 2–4, for exploiting tensor structures in order to compute the projections on the marginals, are based on the following two lemmas.

Lemma 1. *Let $\mathbf{U} = u_0 \otimes u_1 \otimes \cdots \otimes u_{\mathcal{T}}$ and suppose that*

$$\langle \mathbf{K}, \mathbf{U} \rangle = w_1^T \mathrm{diag}(u_t) w_2,$$

where w_1 and w_2 are vectors that may depend on u_ℓ, for $\ell \neq t$. Then,

$$P_t(\mathbf{K} \odot \mathbf{U}) = w_1 \odot u_t \odot w_2.$$

Proof. Let e_{i_t} denote the unit vector whose element i_t is equal to one, and with all other elements equal to zero. Then, by replacing u_t with $u_t \odot e_{i_t}$ in the expression for $\langle \mathbf{K}, \mathbf{U} \rangle$, we get the element i_t of the projection on the t:th marginal, i.e.,

$$\begin{aligned}
&\langle \mathbf{K}, u_0 \otimes \cdots \otimes u_{t-1} \otimes (u_t \odot e_{i_t}) \otimes u_{t+1} \otimes \cdots \otimes u_{\mathcal{T}} \rangle \\
&= \sum_{\ell_0,\ell_1\ldots,\ell_{\mathcal{T}}} \mathbf{K}_{\ell_0,\ell_1\ldots,\ell_{\mathcal{T}}} \left(\prod_{s=0}^{t-1} (u_s)_{\ell_s} \right) (u_t \odot e_{i_t})_{\ell_t} \left(\prod_{s=t+1}^{\mathcal{T}} (u_s)_{\ell_s} \right) \\
&= \sum_{\substack{\ell_0,\ldots,\ell_{t-1} \\ \ell_{t+1},\ldots,\ell_{\mathcal{T}}}} \mathbf{K}_{\ell_0,\ell_1\ldots,\ell_{\mathcal{T}}} \left(\prod_{s=0}^{t-1} (u_s)_{\ell_s} \right) (u_t)_{i_t} \left(\prod_{s=t+1}^{\mathcal{T}} (u_s)_{\ell_s} \right) \\
&= (P_t(\mathbf{K} \odot \mathbf{U}))_{i_t}.
\end{aligned}$$

It may be noted that in the second line, a term in the sum is only non-zero if $i_t = \ell_t$, yielding the second equality. By the assumption in the lemma, we get

$$(P_t(\mathbf{K} \odot \mathbf{U}))_{i_t} = w_1^T \mathrm{diag}(u_t \odot e_{i_t}) w_2 = (w_1)_{i_t} (u_t)_{i_t} (w_2)_{i_t},$$

from which the result follows. □

Lemma 2. *Let $\mathbf{U} = u_0 \otimes u_1 \otimes \cdots \otimes u_{\mathcal{T}}$ and suppose that*

$$\langle \mathbf{K}, \mathbf{U} \rangle = w_1^T \mathrm{diag}(u_{t_1}) W_2 \mathrm{diag}(u_{t_2}) w_3,$$

where $w_1, w_3 \in \mathbb{R}^n$ and $W_2 \in \mathbb{R}^{n \times n}$ may depend on u_ℓ, for $\ell \notin \{t_1, t_2\}$. Then,

$$P_{t_1,t_2}(\mathbf{K} \odot \mathbf{U}) = \mathrm{diag}(w_1 \odot u_t) W_2 \mathrm{diag}(u_{t_2} \odot w_3).$$

Proof. Analogously to the proof in Lemma 1, we may express the bi-marginal projections as

$$\begin{aligned}
(P_{t_1,t_2}(\mathbf{K} \odot \mathbf{U}))_{i_{t_1},i_{t_2}} = \langle \mathbf{K}, u_0 \otimes \cdots \otimes u_{t_1-1} \otimes (u_{t_1} \odot e_{i_{t_1}}) \otimes u_{t_1+1} \otimes \cdots \\
\cdots \otimes u_{t_2-1} \otimes (u_{t_2} \odot e_{i_{t_2}}) \otimes u_{t_2+1} \otimes \cdots \otimes u_{\mathcal{T}} \rangle.
\end{aligned}$$

Hence, by the assumption,

$$\begin{aligned}
(P_{t_1,t_2}(\mathbf{K} \odot \mathbf{U}))_{i_{t_1},i_{t_2}} &= w_1^T \mathrm{diag}(u_{t_1} \odot e_{i_{t_1}}) W_2 \mathrm{diag}(u_{t_2} \odot e_{i_{t_2}}) w_3 \\
&= (w_1)_{i_{t_1}} (u_{t_1})_{i_{t_1}} (W_2)_{i_{t_1},i_{t_2}} (u_{t_2})_{i_{t_2}} (w_3)_{i_{t_2}},
\end{aligned}$$

and thus the result follows.

□

Proof of Proposition 2. First, note that due to the assumption of a sequential cost, we have

$$\mathbf{K}_{i_0,\ldots,i_{\mathcal{T}}} = \prod_{t=1}^{\mathcal{T}} K_{i_{t-1},i_t},$$

for the tensor $\mathbf{K} = \exp(-\mathbf{C}/\epsilon)$ and matrix $K = \exp(-C/\epsilon)$. Therefore,

$$\begin{aligned}
\langle \mathbf{K}, \mathbf{U} \rangle &= \sum_{i_0,i_1,\ldots,i_{\mathcal{T}}} \left(\prod_{t=1}^{\mathcal{T}} K_{i_{t-1},i_t} \right) \prod_{t=0}^{\mathcal{T}} (u_t)_{i_t} \\
&= \sum_{i_0,i_1,\ldots,i_{\mathcal{T}}} (u_0)_{i_0} \prod_{s=1}^{\mathcal{T}} (K \mathrm{diag}(u_t))_{i_{t-1},i_t} \\
&= u_0^T K \mathrm{diag}(u_1) K \ldots K \mathrm{diag}(u_{\mathcal{T}-1}) K u_{\mathcal{T}} \\
&= \mathbf{1}^T \mathrm{diag}(u_0) K \mathrm{diag}(u_1) K \ldots K \mathrm{diag}(u_{\mathcal{T}-1}) K \mathrm{diag}(u_{\mathcal{T}}) \mathbf{1}.
\end{aligned}$$

Thus, $\langle \mathbf{K}, \mathbf{U} \rangle$ may be written as in Lemma 1 with

$$\begin{aligned}
w_1 &= \left(\mathbf{1}^T \mathrm{diag}(u_0) K \mathrm{diag}(u_1) K \ldots K \mathrm{diag}(u_{t-1}) K \right)^T, \\
w_2 &= K \mathrm{diag}(u_{t+1}) K \ldots K \mathrm{diag}(u_{\mathcal{T}-1}) K \mathrm{diag}(u_{\mathcal{T}}) \mathbf{1},
\end{aligned}$$

and hence,

$$\begin{aligned}
P_t(\mathbf{K} \odot \mathbf{U}) = &\left(u_0^T K \mathrm{diag}(u_1) K \ldots K \mathrm{diag}(u_{t-1}) K \right)^T \odot u_t \\
&\odot (K \mathrm{diag}(u_{t+1}) K \ldots K \mathrm{diag}(u_{\mathcal{T}-1}) K u_{\mathcal{T}}).
\end{aligned}$$

Moreover, in order to derive an expression for $P_{t_1,t_2}(\mathbf{K} \odot \mathbf{U})$, note that $\langle \mathbf{K}, \mathbf{U} \rangle$ can be written as in Lemma 2 with

$$\begin{aligned}
w_1 &= \left(\mathbf{1}^T \mathrm{diag}(u_0) K \mathrm{diag}(u_1) K \ldots K \mathrm{diag}(u_{t_1-1}) K \right)^T \\
W_2 &= K \mathrm{diag}(u_{t_1+1}) K \ldots K \mathrm{diag}(u_{t_2-1}) K \\
w_3 &= K \mathrm{diag}(u_{t_2+1}) K \ldots K \mathrm{diag}(u_{\mathcal{T}-1}) K \mathrm{diag}(u_{\mathcal{T}}) \mathbf{1},
\end{aligned}$$

and hence the expression (40) follows. □

Proof of Proposition 3. Due to the structure of the cost tensor $\mathbf{C}$, we have

$$\mathbf{K}_{i_0,\ldots,i_J} = \prod_{j=1}^{J} K_{i_0,i_j},$$

for the tensor $\mathbf{K} = \exp(-\mathbf{C}/\epsilon)$ and the matrix $K = \exp(-C/\epsilon)$. Thus, one may write

$$\begin{aligned}
\langle \mathbf{K}, \mathbf{U} \rangle &= \sum_{i_0,i_1,\ldots,i_J} \left(\prod_{\ell=1}^{J} K_{i_0,i_\ell} \right) \prod_{\ell=0}^{J} (u_\ell)_{i_\ell} \\
&= \sum_{i_0} (u_0)_{i_0} \sum_{i_1,\ldots,i_J} \prod_{\ell=1}^{J} K_{i_0,i_\ell} (u_\ell)_{i_\ell} \\
&= \sum_{i_0} (u_0)_{i_0} \prod_{\ell=1}^{J} (K u_\ell)_{i_0} \\
&= u_0^T \left(\bigodot_{\ell=1}^{J} K u_\ell \right).
\end{aligned}$$

It may be noted that this may be expressed as

$$\langle \mathbf{K}, \mathbf{U} \rangle = \mathbf{1}^T \mathrm{diag}(u_0) \left(\bigodot_{\ell=1}^{J} K u_\ell \right),$$

as well as

$$\langle \mathbf{K}, \mathbf{U} \rangle = (Ku_j)^T \left(u_0 \odot \bigodot_{\substack{\ell=1 \\ \ell \neq j}}^{J} Ku_\ell \right) = \mathbf{1}^T \mathrm{diag}(u_j) K^T \left(u_0 \odot \bigodot_{\substack{\ell=1 \\ \ell \neq j}}^{J} Ku_\ell \right).$$

Applying Lemma 1 yields the expressions (41) and (42) for $P_j(\mathbf{K} \odot \mathbf{U})$ for $j = 0, \ldots, J$. Alternatively, one may rewrite $\langle \mathbf{K}, \mathbf{U} \rangle$ as

$$\langle \mathbf{K}, \mathbf{U} \rangle = \mathbf{1}^T \mathrm{diag}(u_0)\, \mathrm{diag}\left(\bigodot_{\substack{\ell=1 \\ \ell \neq j}}^{J} Ku_\ell \right) K \mathrm{diag}(u_j) \mathbf{1},$$

and

$$\langle \mathbf{K}, \mathbf{U} \rangle = \mathbf{1}^T \mathrm{diag}(u_{j_1}) K^T \mathrm{diag}\left(u_0 \odot \bigodot_{\substack{\ell=1 \\ \ell \neq j_1, j_2}}^{J} Ku_\ell \right) K \mathrm{diag}(u_{j_2}) \mathbf{1}.$$

With Lemma 2, this leads to the expressions (43) and (44) for the couplings $P_{j_1,j_2}(\mathbf{K} \odot \mathbf{U})$ for $j_1, j_2 = 0, \ldots, J$. □

Proof of Proposition 4. Recall the definition of the set $\Lambda = \{(t, j) \,|\, t \in \{0, 1, \ldots, \mathcal{T}\}, j \in \{0, 1 \ldots, J\}\}$. The structure of the cost tensor $\mathbf{C}$ then implies that each element of the tensor $\mathbf{K} = \exp(-\mathbf{C}/\epsilon)$ may be expressed as

$$\mathbf{K}_{\left(i_{(t,j)} | (t,j) \in \Lambda\right)} = \left(\prod_{t=1}^{\mathcal{T}} K_{i_{(t-1,0)} i_{(t,0)}} \right) \prod_{t=0}^{\mathcal{T}} \prod_{j=1}^{J} \tilde{K}_{i_{(t,0)} i_{(t,j)}},$$

with the matrices defined as $K = \exp(-C/\epsilon)$ and $\tilde{K} = \exp(-\tilde{C}/\epsilon)$. Furthermore, the elements of the tensor $\mathbf{U}$ are given by

$$\mathbf{U}_{\left(i_{(t,j)} | (t,j) \in \Lambda\right)} = \prod_{t=0}^{\mathcal{T}} \prod_{j=0}^{J} (u_{(t,j)})_{i_{(t,j)}}.$$

Therefore,

$$\begin{aligned} \langle \mathbf{K}, \mathbf{U} \rangle &= \sum_{\substack{i_{(s,\ell)} \\ (s,\ell) \in \Lambda}} \left(\prod_{t=1}^{\mathcal{T}} K_{i_{(t-1,0)} i_{(t,0)}} \right) \left(\prod_{t=0}^{\mathcal{T}} \prod_{j=1}^{J} \tilde{K}_{i_{(t,0)} i_{(t,j)}} \right) \prod_{t=0}^{\mathcal{T}} \prod_{j=0}^{J} (u_{(t,j)})_{i_{(t,j)}} \\ &= \sum_{\substack{i_{(s,0)} \\ s=0,1,\ldots,\mathcal{T}}} \left(\prod_{t=1}^{\mathcal{T}} K_{i_{(t-1,0)} i_{(t,0)}} \right) \prod_{t=0}^{\mathcal{T}} \left(\sum_{\substack{i_{(t,\ell)} \\ \ell=1,2,\ldots,J}} \left(\prod_{j=1}^{J} \tilde{K}_{i_{(t,0)} i_{(t,j)}} \right) \prod_{j=0}^{J} (u_{(t,j)})_{i_{(t,j)}} \right). \end{aligned}$$

Note that the last sum is a projection of the type (41) in Proposition 3, and we denote it as

$$\begin{aligned} \sum_{\substack{i_{(t,\ell)} \\ \ell=1,2,\ldots,J}} \left(\prod_{j=1}^{J} \tilde{K}_{i_{(t,0)} i_{(t,j)}} \right) \prod_{j=0}^{J} (u_{(t,j)})_{i_{(t,j)}} &= \left(u_{(t,0)} \odot \bigodot_{j=1}^{J} \left(\tilde{K} u_{(t,j)} \right) \right)_{i_{(t,0)}} \\ &= (p_t)_{i_{(t,0)}}. \end{aligned}$$

Hence, using the proof of Proposition 2, one may write

$$\begin{aligned} \langle \mathbf{K}, \mathbf{U} \rangle =& p_0^T K \mathrm{diag}(p_1) K \ldots K \mathrm{diag}(p_{\mathcal{T}-1}) K p_{\mathcal{T}} \\ =& p_0^T K \ldots K \mathrm{diag}(p_{t-1}) K \mathrm{diag}(u_{(t,0)}) \mathrm{diag}\left(\bigodot_{j=1}^{J} \left(\tilde{K} u_{(t,j)} \right) \right) \\ & \cdot K \mathrm{diag}(p_{t+1}) K \ldots K p_{\mathcal{T}} \end{aligned}$$

Thus, the projections on the central marginals, corresponding to index $(t, 0)$ for $t = 0, 1, \ldots, \mathcal{T}$, may be computed as in Proposition 2, yielding

$$\begin{aligned} P_{(t,0)}(\mathbf{K} \odot \mathbf{U}) =& \left(p_0^T K \mathrm{diag}(p_1) K \ldots \mathrm{diag}(p_{t-1}) K \right)^T \odot p_t \\ & \odot (K \mathrm{diag}(p_{t+1}) \ldots K \mathrm{diag}(p_{\mathcal{T}-1}) K p_{\mathcal{T}}). \end{aligned}$$

For the other marginals, i.e., for (t, j) such that $j > 0$, one may express $\langle \mathbf{K}, \mathbf{U} \rangle$ similarly as in the proof of (42) of Proposition 3, i.e.,

$$\begin{aligned} \langle \mathbf{K}, \mathbf{U} \rangle =& \mathbf{1}^T \mathrm{diag}(u_{(t,j)}) \tilde{K}^T \left((p_0^T K \mathrm{diag}(p_1) K \ldots \mathrm{diag}(p_{t-1}) K)^T \right. \\ & \left. \odot (p_t ./ (K u_{(t,j)})) \odot (K \mathrm{diag}(p_{t+1}) \ldots K \mathrm{diag}(p_{\mathcal{T}-1}) K p_{\mathcal{T}}) \right), \end{aligned}$$

which, with Lemma 1, yields the expression (46). □

References

[1] I. Abraham, R. Abraham, M. Bergounioux, G. Carlier, Tomographic reconstruction from a few views: a multi-marginal optimal transport approach, Appl. Math. Optim. 75 (1) (2017) 55–73.
[2] A. Adler, M. Wax, Direct localization by partly calibrated arrays: a relaxed maximum likelihood solution, in: 27th European Signal Processing Conference, A Coruna, Spain, 2019.
[3] J. Adler, A. Ringh, O. Öktem, J. Karlsson, Learning to solve inverse problems using Wasserstein loss, 2017, arXiv:1710.10898.
[4] A.N. Amini, E. Ebbini, T.T. Georgiou, Noninvasive estimation of tissue temperature via high-resolution spectral analysis techniques, IEEE Trans. Biol. Eng. 52 (2) (2005) 221–228.
[5] S. Angenent, S. Haker, A. Tannenbaum, Minimizing flows for the Monge–Kantorovich problem, SIAM J. Math. Anal. 35 (1) (2003) 61–97.
[6] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN, 2017. arXiv:1701.07875.
[7] Auditec, Auditory tests (revised), compact disc, auditec, st. louis. st. louis, 1997.
[8] M. Beiglböck, P. Henry-Labordère, F. Penker, Model-independent bounds for option prices - a mass transport approach, Financ. Stoch. 17 (3) (2013) 477–501.
[9] J.D. Benamou, G. Carlier, M. Cuturi, L. Lenna, G. Peyré, Iterative Bregman projections for regularized transportation problems, SIAM J. Sci. Comput. 37 (2) (2015) 1111–1138.
[10] J.C. Bezdek, R.J. Hathaway, Two new convergence results for alternating optimization, in: Computational Intelligence: The Experts Speak, John Wiley and Sons Inc., 2003, pp. 149–164.
[11] Y. Brenier, Generalized solutions and hydrostatic approximation of the Euler equations, Physica D 237 (2008) 1982–1988.
[12] J. Capon, High resolution frequency wave number spectrum analysis, Proc. IEEE 57 (1969) 1408–1418.
[13] Y. Chen, T.T. Georgiou, M. Pavon, Optimal transport over a linear dynamical system, IEEE Trans. Autom. Control 62 (5) (2017) 2137–2152.
[14] Y. Chen, T.T. Georgiou, M. Pavon, On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint, J. Optim. Theory Appl. 169 (2) (2016) 671–691.
[15] Y. Chen, J. Karlsson, State tracking of linear ensembles via optimal mass transport, IEEE Control Syst. Lett. 2 (2) (2018) 260–265.
[16] L. Chizat, G. Peyré, B. Schmitzer, F.X. Vialard, Scaling algorithms for unbalanced optimal transport problems, Math. Comput. 87 (314) (2018) 2563–2609.
[17] M. Cuturi, Sinkhorn distances: lightspeed computation of optimal transport, Proc. Adv. Neural Inf. Process. Syst. (2013) 2292–2300.
[18] E. De Sena, N. Antonello, M. Moonen, T. van Waterschoot, On themodeling of rectangular geometries in room acoustic simulations, IEEE/ACM Trans. Audio Speech Lang. Process. 23 (4) (2015) 774–786.
[19] P. del Aguila Pla, J. Jaldén, Cell Detection by Functional Inverse Diffusion and Non-negative Group Sparsity - Part II: Proximal Optimization and Performance Evaluation, IEEE Trans. Signal Process. 66 (20) (2018) 5422–5437.
[20] A. Dominitz, A. Tannenbaum, Texture mapping via optimal mass transport, IEEE Trans. Vis. Comput. Graph. 16 (3) (2010) 419–433.
[21] S.A. Elkader, A.B. Gershman, K.M. Wong, Rank reduction direction-of-arrival estimators with an improved robustness against subarray orientation errors, IEEE Trans. Signal Process. 54 (5) (2006) 1951–1955.
[22] F. Elvander, I. Haasler, A. Jakobsson, J. Karlsson, Tracking andsensor fusion in direction of arrival estimation using optimal mass transport, in: 26th European Signal Processing Conference, 2018, pp. 1617–1621. Rome, Italy, Sep. 3-7
[23] F. Elvander, I. Haasler, A. Jakobsson, J. Karlsson, Non-coherent sensor fusion via entropy regularized optimal mass transport, in: Proc. 44th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 2019, pp. 4415–4419. Brighton, UK, May 13-17
[24] F. Elvander, A. Jakobsson, J. Karlsson, Interpolation and extrapolation of Toeplitz matrices via optimal mass transport, IEEE Trans. Signal. Process. 66 (20) (2018) 5285–5298.
[25] T.T. Georgiou, Solution of the general moment problem via a one-parameter imbedding, IEEE Trans. Autom. Control 50 (6) (2005) 811–826.
[26] T.T. Georgiou, J. Karlsson, M.S. Takyar, Metrics for power spectra: an axiomatic approach, IEEE Trans. Signal Process. 57 (3) (2009) 859–867.
[27] R.M. Gray, A. Buzo, A.H. Gray Jr, Y. Matsuyama, Distortion measures for speech processing, IEEE Trans. Acoust. Speech Signal Process. 28 (4) (1980) 367–376.
[28] I. Haasler, A. Ringh, Y. Chen, J. Karlsson, Estimating ensemble flows on a hidden Markov chain, in: Proceedings of the 58th IEEE Conference on Decision and Control, 2019.
[29] Y. Hu, P.C. Loizou, A Perceptually Motivated Approach for Speech Enhancement, 11, 2003, pp. 457–465.
[30] X. Jiang, Z.Q. Luo, T.T. Georgiou, Geometric methods for spectral analysis, IEEE Trans. Signal Process. 60 (3) (2012) 1064–1074.
[31] D.H. Johnson, D.E. Dudgeon, Array Signal Processing: Concepts and Techniques, Prentice Hall, Englewood Cliffs, N.J., 1993.

[32] L.M. Kaplan, Global node selection for localization in a distributed sensor network, IEEE Trans. Aerosp. Electron. Syst. 42 (1) (2006) 113–135.
[33] J. Karlsson, A. Ringh, Generalized Sinkhorn iterations for regularizing inverse problems using optimal mass transport, SIAM J. Imaging Sci. 10 (4) (2017) 1935–1962.
[34] S. Kolouri, S.R. Park, M. Thorpe, D. Slepcev, G.K. Rohde, Optimal Mass Transport: signal processing and machine-learning applications, IEEE Signal Process. Mag. 34 (4) (2017) 43–59.
[35] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.
[36] J. Lellmann, D.A. Lorenz, C. Schönlieb, T. Valkonen, Imaging with Kantorovich–Rubinstein Discrepancy, SIAM J. Imaging Sci. 7 (4) (2014) 2833–2859.
[37] C. Léonard, A survey of the Schrödinger problem and some of its connections with optimal transport, Discret. Continuous Dyn. Syst. - A 34 (4) (2014) 1533–1574.
[38] D. Li, K.D. Wong, Y.H. Hu, A.M. Sayeed, Detection, classification, and tracking of targets, IEEE Signal Process. Mag. 19 (2) (2002) 17–29.
[39] J. Lorenz, L. Franklin, On the scaling of multidimensional matrices, Linear Algebra Appl. 114–115 (1989). 717–715
[40] J.-A. Luo, K. Yu, Z. Wang, Y.-H. Hu, Passive source localization from array covariance matrices via joint sparse representations, Neurocomputing 270 (2017) 82–90.
[41] R.J. McCann, A convexity principle for interacting gases, Adv. Math. 128 (1) (1997) 153–179.
[42] S. Mierisová, M. Ala-Korpela, MR spectroscopy quantitation: a review of frequency domain methods, NMR Biomed. 14 (4) (2001) 247–259.
[43] L. Ning, T.T. Georgiou, A. Tannenbaum, On matrix-valued Monge–Kantorovich optimal mass transport, IEEE Trans. Autom. Control 60 (2) (2015) 373–382.
[44] L. Ning, X. Jiang, T.T. Georgiou, On the geometry of covariance matrices, IEEE Signal Process. L 20 (8) (2013) 787–790.
[45] B. Pass, Multi-marginal optimal transport: theory and applications, ESAIM: Math. Model. Numer.Anal. 49 (6) (2015) 1771–1790.
[46] O. Pele, M. Werman, Fast and robust Earth Mover's Distances, in: IEEE 12th Int. Conf. on Comp. Vis., 2009, pp. 460–467. Kyoto, Japan, Sep. 29 - Oct. 2
[47] G. Peyré, M. Cuturi, Computational optimal transport, Found. Trends® Mach.Learn. 11 (5–6) (2019) 355–607.
[48] A. Rényi, On measures of entropy and information, in: Proc. 4th Berkeley Sym. Mathematics of Statistics and Probability, 1961, pp. 547–561.
[49] D.W. Rieken, D.R. Fuhrmann, Generalizing MUSIC MVDR for multiple noncoherent arrays, IEEE Trans. Signal Process. 52 (9) (2004) 2396–2406.
[50] M.A. Schmitz, M. Heitz, N. Bonneel, F. Ngolè, D. Coeurjolly, M. Cuturi, G. Peyré, J.-L. Starck, Wassersteindictionary learning: optimal transport-based unsupervised nonlinear dictionary learning, SIAM J. Imaging Sci. 11 (1) (2018) 643–678.
[51] C.M. See, A.B. Gershman, Direction-of-arrival estimation in partly calibrated subarray-based sensor arrays, IEEE Trans. Signal Process. (2004).
[52] E. Simou, P. Frossard, Graph signal representation with Wasserstein barycenters, in: Proc. 44th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 2019, pp. 5386–5390. Brighton, UK, May 13-17
[53] R. Sinkhorn, Diagonal equivalence to matrices with prescribed row and column sums, Am. Math. Mon. 74 (4) (1967) 402–405.
[54] P. Stoica, P. Babu, J. Li, SPICE : a novel covariance-based sparse estimation method for array processing, IEEE Trans. Signal Process. 59 (2) (2011) 629–638.
[55] P. Stoica, R. Moses, Spectral Analysis of Signals, Prentice Hall, Upper Saddle River, N.J., 2005.
[56] W. Suleiman, P. Parvazi, M. Pesavento, A.M. Zoubir, Non-coherent direction-of-arrival estimation using partly calibrated arrays, IEEE Trans. Signal Process. 66 (21) (2018) 5776–5788.
[57] J. Swärd, S.I. Adalbjörnsson, A. Jakobsson, Generalized sparse covariance-based estimation, Elsevier Signal Process. 143 (2018) 311–319.
[58] H.L.V. Trees, Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise, Krieger Publishing Co., Inc., 1992.
[59] N. Vervliet, O. Debals, L. Sorber, M.V. Barel, L. De Lathauwer, Tensorlab 3.0, mar. 2016. available online.
[60] C. Villani, Optimal Transport: Old and New, Springer Science & Business Media, 2008.
[61] D. Xie, J. Huang, H. Ge, Localization of near-field sources with partly calibrated subarray-based array, in: 5th IEEE Conf. on Industrial Electronics and Applications, 2010, pp. 1758–1761.
[62] K. Yamamoto, Y. Chen, L. Ning, T.T. Georgiou, A. Tannenbaum, Regularization and Interpolation of Positive Matrices, IEEE Trans. Autom. Control 63 (4) (2018) 1208–1212.