

CHALMERS
UNIVERSITY OF TECHNOLOGY

Fast Computation of the TGOSPA Metric for Multiple Target Tracking Via Unbalanced Optimal Transport

Downloaded from: https://research.chalmers.se, 2026-11-02 21:06 UTC

Citation for the original published paper (version of record):
Wernholm, V., Warnsater, A., Ringh, A. (2025). Fast Computation of the TGOSPA Metric for Multiple Target Tracking Via Unbalanced Optimal Transport. IEEE Control Systems Letters, 9: 498-503.
http://dx.doi.org/10.1109/LCSYS.2025.3573880

N.B. When citing this work, cite the original published paper.

© 2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, or reuse of any copyrighted component of this work in other works.

(article starts on next page)

Fast Computation of the TGOSPA Metric for Multiple Target Tracking via Unbalanced Optimal Transport

Viktor Nevelius Wernholm, Alfred Wärnsäter, and Axel Ringh

***Abstract*—In multiple target tracking, it is important to be able to evaluate the performance of different tracking algorithms. The trajectory generalized optimal sub-pattern assignment metric (TGOSPA) is a recently proposed metric for such evaluations. The TGOSPA metric is computed as the solution to an optimization problem, but for large tracking scenarios, solving this problem becomes computationally demanding. In this letter, we present an approximation algorithm for evaluating the TGOSPA metric, based on casting the TGOSPA problem as an unbalanced multimarginal optimal transport problem. Following recent advances in computational optimal transport, we introduce an entropy regularization and derive an iterative scheme for solving the Lagrangian dual of the regularized problem. Numerical results suggest that our proposed algorithm is more computationally efficient than the alternative of computing the exact metric using a linear programming solver, while still providing an adequate approximation of the metric.**

***Index Terms*—Estimation, numerical algorithms, optimization, optimization algorithms.**

I. Introduction

MULTIPLE target tracking (MTT) deals with the task of estimating targets that appear, disappear, and move within a scene, given data from noisy measurements. A wide range of algorithms that solves this task has been developed, see, e.g., [4], [5], [10]. To objectively evaluate the performance of different MTT algorithms in test scenarios, where the ground truth trajectories of the objects are known, one needs a distance function that quantifies the error between the estimated target trajectories and the ground truth trajectories. A recently proposed metric for evaluating MTT algorithms is the trajectory generalized optimal sub-pattern assignment (TGOSPA) metric [11], [12], [16]. This is an extension of the GOSPA metric [22]. The latter is a metric between snapshots of ground truth and identified targets in a single time frame, and it penalizes localization errors for properly detected targets, missed targets in the ground truth, and falsely detected targets that do not exist in the ground truth. The TGOSPA metric generalizes this to a metric for tracks over multiple time frames by also including a penalty for so-called track switching, where the identities of ground truth targets erroneously get swapped.

Both the GOSPA metric and the TGOSPA metric are formulated as optimization problems. While the GOSPA metric involves solving an assignment problem, which can be done efficiently using, e.g., the Hungarian algorithm (see [20, Ch. 11]), computation of the TGOSPA metric requires solving a set of coupled, consecutive assignment problems. Computing the TGOSPA metric is therefore in general only tractable when the tracking scenario contains a small number of targets [11, Sec. III.C]. Therefore, [11], [12] also suggests a linear programming (LP) relaxation of the TGOSPA metric, and this relaxation is in fact also a metric. Nevertheless, for large tracking scenarios, solving the corresponding LP still requires significant computational resources.

In this letter, which is based on the master's thesis [19], we derive an efficient method for approximately solving the LP-relaxed TGOSPA problem. This is done by first casting the problem as an unbalanced multimarginal optimal transport problem and deriving the Sinkhorn iterations [2], [3], [7], [14], [15], [21], [23]. However, in the multimarginal setting, the Sinkhorn iterations only partially alleviate the computational cost, as computing the projections needed in the iterations in general scales exponentially with the number of marginals. This can be overcome if, e.g., the cost function has a graph structure [2], [3], [14], [15], [23], but that is not the case for the problem considered here. Nevertheless, the cost function in the problem is a decomposable structured tensor, and using the structure we develop a computationally efficient way to evaluate the projections needed. Finally, we prove that the

Received 12 March 2025; revised 9 May 2025; accepted 20 May 2025. Date of publication 26 May 2025; date of current version 16 June 2025. This work was supported in part by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, Sweden; in part by the Swedish Research Council (VR) under Grant 2020-03454; and in part by the KTH Digital Futures. Recommended by Senior Editor C. Briat. *(Viktor Nevelius Wernholm and Alfred Wärnsäter contributed equally to this work.) (Corresponding author: Axel Ringh.)*

Viktor Nevelius Wernholm is with Saab Surveillance, Saab AB, 412 89 Gothenburg, Sweden (e-mail: viktor.neveliuswernholm@saabgroup.com).

Alfred Wärnsäter is with the Department of Mathematics, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden (e-mail: alfwar@kth.se).

Axel Ringh is with the Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96 Gothenburg, Sweden (e-mail: axelri@chalmers.se).

Digital Object Identifier 10.1109/LCSYS.2025.3573880

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

algorithm converges linearly, and demonstrate its performance on a number of examples.

II. Background

In this section, we introduce the TGOSPA metric and the basics of entropy regularized multimarginal optimal transport. The section is also used to set up notation. To this end, we use $^\top$ to denote the transpose of a vector or a matrix, $\mathbb{R}_+$ to denote non-negative real numbers, and the operation exp on a vector, matrix, or tensor means elementwise exponential.

A. TGOSPA

A trajectory X on time steps $\{1,\dots,T\} \subset \mathbb{N}$ is defined as the sequence $(\boldsymbol{x}^1,\dots,\boldsymbol{x}^T)$ of sets $\boldsymbol{x}^t$. If a trajectory is alive and in a state $x^t \in \mathbb{R}^N$ at time step t, then $\boldsymbol{x}^t = \{x^t\}$. Otherwise, $\boldsymbol{x}^t = \varnothing$. Let $\mathcal{T}$ denote the set of all possible such trajectories. The TGOSPA metric is a metric between sets of trajectories, i.e., a function that maps $\mathcal{T}\times\mathcal{T} \to \mathbb{R}_+$, and it can be formulated as an integer linear program. To this end, let $\mathbf{X}, \mathbf{Y} \in \mathcal{T}$, and denote with $\boldsymbol{x}_i^t$ and $\boldsymbol{y}_j^t$ the (possibly empty) state of ground truth i and estimate j at time step t. For a set of ground truths $\mathbf{X}$ consisting of m trajectories, and a set of estimates $\mathbf{Y}$ consisting of n trajectories, we can represent their associations in time step t as an $(m+1)\times(n+1)$ binary matrix W^t. For such a matrix, $W^t_{i,j} = 1$ means that ground truth i and estimate j are assigned to each other at time step t, and $W^t_{i,j} = 0$ means that they are not. Every such matrix thus satisfies

$$W^t_{(i,j)} \in \{0,1\}, \quad i = 1,\dots,m, \quad j = 1,\dots,n, \tag{1a}$$

$$\sum_{i=1}^{m+1} W^t_{(i,j)} = 1, \quad j = 1,\dots,n, \tag{1b}$$

$$\sum_{j=1}^{n+1} W^t_{(i,j)} = 1, \quad i = 1,\dots,m. \tag{1c}$$

Here, (1b) implies that each ground truth trajectory is either assigned to exactly one estimated trajectory or unassigned (in the latter case, $W^t_{(m+1,j)} = 1$), and (1c) implies the corresponding property among the estimated trajectories.

Let $\mathcal{W}_{(m,n)}$ be the set of assignment matrices described by (1). The TGOSPA metric is the power $1/p$ of

$$\begin{aligned}\underset{\substack{W^t\in\mathcal{W}_{(m,n)},\\ t=1,\dots,T}}{\text{minimize}} \quad & \sum_{t=1}^{T}\sum_{i=1}^{m+1}\sum_{j=1}^{n+1} D^t_{(i,j)} W^t_{(i,j)} \\ & + \frac{\gamma^p}{2}\sum_{t=1}^{T-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|W^{t+1}_{(i,j)} - W^t_{(i,j)}\right|,\end{aligned} \tag{2}$$

see [11, Lem. 1]. Here, $1 \le p < \infty$ determines to what extent outliers are penalized, $\gamma > 0$ determines how much we penalize track-switches, and

$$\left(D^t_{(i,j)}\right)^{1/p} = \begin{cases} \min\left(||x_i^t - y_j^t||, c\right), & \text{if } \boldsymbol{x}_i = \{x_i\}, \boldsymbol{y}_j = \{y_j\}, \\ 0, & \text{if } \boldsymbol{x} = \varnothing, \boldsymbol{y} = \varnothing, \\ c/2^{1/p} & \text{else,} \end{cases}$$

where $c > 0$ is a cut-off parameter, and where $\boldsymbol{x}^t_{m+1} = \varnothing$ and $\boldsymbol{y}^t_{n+1} = \varnothing$ for all t.

Remark 1: For $t = 1,\dots,T$, the problem

$$\underset{W^t\in\mathcal{W}_{(m,n)}}{\text{minimize}} \quad \sum_{i=1}^{m+1}\sum_{j=1}^{n+1} D^t_{(i,j)} W^t_{(i,j)}$$

in (2) is an unbalanced optimal transport problem [2], [13].

Problem (2) is an integer linear program, and to reduce the computational cost, [11] also proposes the relaxed TGOSPA problem obtained by relaxing the binary constraints (1a) to $W^t_{(i,j)} \ge 0$.[1] It can be shown that this relaxed TGOSPA problem also defines a metric on $\mathcal{T}$, see [11, Sec. IV].

B. Multimarginal Optimal Transport

Optimal transport deals with problems of how mass can be moved between an initial distribution and a target distribution as efficiently as possible. Such problems can be extended to multimarginal optimal transport problems, where an optimal transport plan between several distributions is sought.

Let $\hat{M} \in \mathbb{R}_+^{N_1\times\dots\times N_T}$ denote the transport tensor, $\hat{C} \in \mathbb{R}_+^{N_1\times\dots\times N_T}$ the cost tensor, and $\hat{\mu}_t \in \mathbb{R}_+^{N_t}$, for $t = 1,\dots,T$, the marginals of a multimarginal optimal transport problem. Additionally, for $t = 1,\dots,T$, define the projections $\hat{P}_t(\hat{M})_{i_t} = \sum_{i_1,\dots,i_{t-1},i_{t+1},\dots,i_T} \hat{M}_{i_1,\dots,i_T}$. The most common type of multimarginal optimal transport problems is then

$$\begin{aligned}\underset{\hat{M}\in\mathbb{R}_+^{N_1\times\dots\times N_T}}{\text{minimize}} \quad & \langle \hat{C}, \hat{M}\rangle \\ \text{subject to} \quad & \hat{P}_t\left(\hat{M}\right) = \hat{\mu}_t, \quad t = 1,\dots,T,\end{aligned} \tag{3}$$

where $\langle A, B\rangle \coloneqq \sum_{i_1,\dots,i_L} A_{i_1,\dots,i_L} B_{i_1,\dots,i_L}$, for tensors A and B with L indices.

Even though (3) is an LP, it is, in general, difficult to solve directly for larger instances since the number of variables increases exponentially with T. This makes manipulating and storing the tensors $\hat{C}$ and $\hat{M}$ computationally infeasible.

A first step towards addressing this is to compute approximate solutions using entropy regularization [3], [7], [14], [15], [21], [23]. To this end, for a tensor A with L indices, let the entropy of A be defined by

$$E(A) = \sum_{i_1,\dots,i_L}\left(A_{i_1,\dots,i_L}\log(A_{i_1,\dots,i_L}) - A_{i_1,\dots,i_L} + 1\right).$$

For some regularization parameter $\varepsilon > 0$, adding the term $\varepsilon E(\hat{M})$ to the objective function of (3) yields the regularized problem

$$\begin{aligned}\underset{\hat{M}\in\mathbb{R}^{N^T}}{\text{minimize}} \quad & \langle \hat{C}, \hat{M}\rangle + \varepsilon E\left(\hat{M}\right) \\ \text{subject to} \quad & \hat{P}_t\left(\hat{M}\right) = \hat{\mu}_t, \quad t = 1,\dots,T.\end{aligned} \tag{4}$$

A Sinkhorn algorithm for solving this problem is obtained by doing block coordinate ascent in the Lagrangian dual problem. However, for such an algorithm to be efficient, one must also find an efficient way to compute the projections needed in the algorithm; cf. [3], [14], [15], [23].

[1] Note that $W^t_{(i,j)} \le 1$ is implicitly enforced by (1b) and (1c). Also note that the LP relaxation is known to not be tight, see [19, Prop. 4].

III. A Novel Algorithm for Approximation of the Relaxed TGOSPA Metric

In this section, we derive a Sinkhorn algorithm for approximating the optimal value of the LP-relaxed version of the TGOSPA metric. This is done in three main steps. First, we reformulate the LP-relaxed version of (2) as a multimarginal optimal transport problem over a high-order tensor. Second, following the steps in the literature, we derive a block coordinate ascent algorithm in the dual to the entropy regularized version of this problem. Third, we show that the structure of the cost function can be used to efficiently compute the projections needed in the algorithm. Using this, we then formulate the full algorithm, and also show that it converges linearly.

A. TGOSPA as Multimarginal Optimal Transport

Let $\mathcal{M}_+$ denote the set of all non-negative tensors with $2T$ indices such that odd indices have dimension $m+1$, and even indices have dimension $n+1$. We index such tensors using the notation $M_{(i_1,j_1),\dots,(i_T,j_T)}$ for $i_1,\dots,i_T \in \{1,\dots,m+1\}$ and $j_1,\dots,j_T \in \{1,\dots,n+1\}$, and the element $M_{(i_1,j_1),\dots,(i_T,j_T)}$ should be interpreted as the amount of mass transported along the trajectory $(i_1,j_1),\dots,(i_T,j_T)$. Analogously to the projections defined in Section II-B, let

$$P_t(M)_{(i_t,j_t)} = \sum_{(i_1,j_1),\dots,(i_T,j_T)\setminus(i_t,j_t)} M_{(i_1,j_1),\dots,(i_T,j_T)},$$

for $t=1,\dots,T$, and let

$$P_{t,t+1}(M)_{(i_t,j_t),(i_{t+1},j_{t+1})} = \sum_{\substack{(i_1,j_1),\dots,(i_T,j_T)\setminus\\(i_t,j_t),(i_{t+1},j_{t+1})}} M_{(i_1,j_1),\dots,(i_T,j_T)},$$

for $t=1,\dots,T-1$. Here, $P_t(M)_{(i_t,j_t)}$ should be interpreted as the amount of mass at position (i_t,j_t) at time step t, and $P_{t,t+1}(M)_{(i_t,j_t),(i_{t+1},j_{t+1})}$ should be interpreted as the mass transported between position (i_t,j_t) and position (i_{t+1},j_{t+1}) from time step t to $t+1$.

Next, note that the assignment matrices W^t can be naturally identified with the projections $P_t(M)$ for $t=1,\dots,T$. Using this idea, we prove the following.

Theorem 1: Let $\bar{\mu} \in \mathbb{R}^{m+1}$ and $\tilde{\mu} \in \mathbb{R}^{n+1}$ be defined by $\bar{\mu} = (\mathbf{1}_m^\top, n)^\top$ and $\tilde{\mu} = (\mathbf{1}_n^\top, m)^\top$, where $\mathbf{1}_\ell \in \mathbb{R}^\ell$ is a vector with all elements equal to one. Further, let

$$F_{(i,j),(k,l)} = \begin{cases} \frac{\gamma^p}{2}\big(1-\delta_{ik}\delta_{jl}\big)\big(\big(1-\delta_{k,m+1}\big)\big(1-\delta_{l,n+1}\big) \\ \quad +\big(1-\delta_{i,m+1}\big)\big(1-\delta_{j,n+1}\big)\big), & \text{if } i=k, \\ \infty, \quad \text{otherwise}, \end{cases}$$

where δ denotes the Kronecker delta, i.e., $\delta_{i,j}=1$ if $i=j$ and 0 otherwise. Then the multimarginal optimal transport problem

$$\begin{aligned} \underset{M\in\mathcal{M}_+}{\text{minimize}} \quad & \langle C, M\rangle, \\ \text{subject to} \quad & P_t(M)\mathbf{1}_{n+1} = \bar{\mu},\ t=1,\dots,T, \\ & P_t(M)^\top \mathbf{1}_{m+1} = \tilde{\mu},\ t=1,\dots,T, \end{aligned} \tag{5a}$$

where

$$C_{(i_1,j_1),\dots,(i_T,j_T)} = \sum_{t=1}^{T} D^t_{(i_t,j_t)} + \sum_{t=1}^{T-1} F_{(i_t,j_t),(i_{t+1},j_{t+1})}, \tag{5b}$$

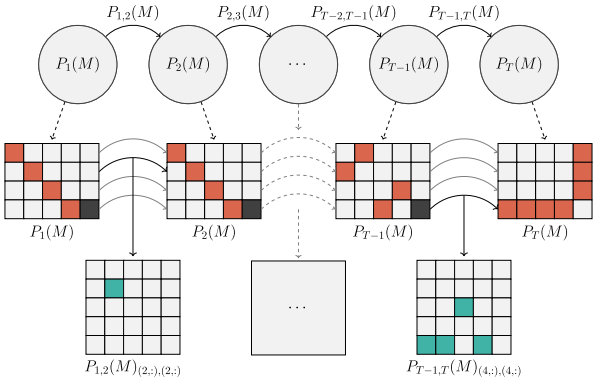


Fig. 1. Illustration of the problem structure. The top row shows the high level structure, the middle row the assignment matrices, and the bottom row the flows between rows in the assignment matrices.

and the LP-relaxed version of (2) has the same optimal value.

Proof: To show that the optimal values coincide, we first claim that for any feasible solution M of (5), there exists a feasible solution to the LP-relaxed version (2) with a lower or equal objective value. To see this, let $W^t = P_t(M)$ for $t=1,\dots,T$. From the constraints in (5), it follows that $\{W^t\}_t$ is a feasible solution to the LP relaxation of (2). Next, a direct calculation gives that

$$\begin{aligned} \langle F, P_{t,t+1}(M)\rangle &\geq \frac{\gamma^p}{2}\sum_{i,j=1}^{m,n}\Big[W^t_{(i,j)} + W^{t+1}_{(i,j)} \\ &\quad -2\min\Big(W^t_{(i,j)}, W^{t+1}_{(i,j)}\Big)\Big] = \frac{\gamma^p}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\Big|W^{t+1}_{(i,j)} - W^t_{(i,j)}\Big|, \end{aligned} \tag{6}$$

where the inequality is due to $P_{t,t+1}(M)_{(i,j),(i,j)} \leq W^s_{ij}$ for $s=t,t+1$, and where the last equality follows from the identity $a+b-2\min(a,b) = |a-b|$ for $a,b\in\mathbb{R}$. Using (6), we get $\langle C, M\rangle = \sum_{t=1}^{T}\langle D^t, P_t(M)\rangle + \sum_{t=1}^{T-1}\langle F, P_{t,t+1}(M)\rangle \geq \sum_{t=1}^{T}\sum_{i=1}^{m+1}\sum_{j=1}^{n+1} D^t_{(i,j)}W^t_{(i,j)} + \frac{\gamma^p}{2}\sum_{t=1}^{T-1}\sum_{i=1}^{m}\sum_{j=1}^{n}|W^{t+1}_{(i,j)} - W^t_{(i,j)}|$, which proves the claim. Conversely, let $\{W^t\}_t$ be a feasible solution to the LP-relaxed version of (2). From (6), we see that a sufficient condition for equality is that M is feasible to (5), $P_t(M) = W^t$ for $t=1,\dots,T$, and that M is such that $P_{t,t+1}(M)_{(i,j),(i,j)} = \min(W^t_{(i,j)}, W^{t+1}_{(i,j)})$ for $t=1,\dots,T-1$, $i=1,\dots,m$ and $j=1,\dots,n$. Such M exists as it simply corresponds to a transport plan where as much mass as possible is unmoved. ■

In Figure 1, the structure of problem (5) is shown. The first part of the cost tensor (5b) corresponds to costs on the assignment matrices directly (middle row), while the second part corresponds to transports between them (bottom row).

Remark 2: The marginals $\bar{\mu}$ and $\tilde{\mu}$ have equal mass, which is a prerequisite for utilizing the optimal transport framework. Introducing the last elements in $\bar{\mu}$ and $\tilde{\mu}$ is one way to handle unbalanced transport problems [2], [13].

Remark 3: There are multiple ways in which the LP-relaxed version of (2) can be formulated as an optimal transport problem. For another formulation, see [19, Sec. 4.1].

B. Block Coordinate Ascent in the Lagrangian Dual

To derive a Sinkhorn algorithm for (5), we first add an entropy regularization to the objective function, which yields

$$\begin{aligned} \underset{M\in\mathcal{M}_+}{\text{minimize}} \quad & \langle C, M\rangle + \varepsilon E(M), && \text{(7a)}\\ \text{subject to} \quad & P_t(M)\mathbf{1}_{n+1} = \bar{\mu},\ t = 1,\dots,T, && \text{(7b)}\\ & P_t(M)^\top \mathbf{1}_{m+1} = \tilde{\mu},\ t = 1,\dots,T. && \text{(7c)} \end{aligned}$$

For small values of ε, the optimal value of (7) is close to the optimal value of (5) (cf. [21, Sec. 4.5]). Next, following similar steps as in the literature (see, e.g., [3], [7], [14], [15], [21], [23]), we Lagrangian relax the constraints (7b) and (7c) with the dual variables $\bar{\lambda}^t \in \mathbb{R}^{m+1}$ and $\tilde{\lambda}^t \in \mathbb{R}^{n+1}$, respectively. Taking the derivative of the Lagrangian with respect to M and setting it equal to zero gives

$$\begin{aligned} M^{(\bar{\lambda},\tilde{\lambda})}_{(i_1,j_1),\dots,(i_T,j_T)} = \exp\left(-\frac{C_{(i_1,j_1),\dots,(i_T,j_T)}}{\varepsilon}\right) \\ \cdot \prod_{t=1}^{T} \exp\left(\frac{\bar{\lambda}^t_{i_t}}{\varepsilon}\right) \prod_{t=1}^{T} \exp\left(\frac{\tilde{\lambda}^t_{j_t}}{\varepsilon}\right), \end{aligned} \quad (8)$$

which, in turn, gives the dual problem

$$\underset{\substack{\bar{\lambda}^t\in\mathbb{R}^{m+1},\ \tilde{\lambda}^t\in\mathbb{R}^{n+1}\\ t=1,\dots,T}}{\text{maximize}} \quad \phi\left(\bar{\lambda},\tilde{\lambda}\right), \quad (9)$$

where $\phi(\bar{\lambda},\tilde{\lambda}) = \tilde{\phi}(\bar{\lambda},\tilde{\lambda}) + \varepsilon(m+1)^T(n+1)^T$, and where

$$\begin{aligned} \tilde{\phi}\left(\bar{\lambda},\tilde{\lambda}\right) = -\varepsilon \sum_{(i_1,j_1),\dots,(i_T,j_T)} \exp\left(-\frac{C_{(i_1,j_1),\dots,(i_T,j_T)}}{\varepsilon}\right) \\ \cdot \prod_{t=1}^{T} \exp\left(\frac{\bar{\lambda}^t_{i_t}}{\varepsilon}\right) \prod_{t=1}^{T} \exp\left(\frac{\tilde{\lambda}^t_{j_t}}{\varepsilon}\right) \\ + \sum_{t=1}^{T}\sum_{i=1}^{m+1} \bar{\mu}_i\bar{\lambda}^t_i + \sum_{t=1}^{T}\sum_{j=1}^{n+1} \tilde{\mu}_j\tilde{\lambda}^t_j. \end{aligned}$$

We have the following result regarding the primal-dual problems (7) and (9).

Proposition 1: The optimization problems (7) and (9) are both convex, both have optimal solutions, and they attain the same optimal value. Moreover, the optimal solutions to the two problems are related via (8).

Proof: Due to space limitations, only a proof outline is given. The first step is to observe that there exists a feasible solution to (7) with the property that $M_{(i_1,j_1),\dots,(i_T,j_T)} > 0$ for all indices such that $C_{(i_1,j_1),\dots,(i_T,j_T)} < \infty$. The remaining results them follows by appropriate use of strong duality; see, e.g., [6, Ch. 5]. ∎

Based on the proposition above, an algorithm to solve (7) is now constructed as a coordinate ascent algorithm to solve (9). To simplify the presentation, we introduce the transformed dual variables $\bar{u}^t = \exp(\bar{\lambda}^t/\varepsilon)$ and $\tilde{u}^t = \exp(\tilde{\lambda}^t/\varepsilon)$, and the tensor $K = \exp(-C/\varepsilon)$. Next, let $M^{(\bar{\lambda},\tilde{\lambda})}$ be a tensor of the form (8), and let $\bar{w}^\tau_{i_\tau} = (P_\tau(M^{(\bar{\lambda},\tilde{\lambda})})\mathbf{1}_{n+1})_{i_\tau}/\bar{u}^\tau_{i_\tau}$. We observe that

$$\bar{w}^\tau_{i_\tau} = \sum_{j_t} \sum_{(i_1,j_1),\dots,(i_T,j_T)\setminus(i_t,j_t)} K_{(i_1,j_1),\dots,(i_T,j_T)} \prod_{\substack{t=1\\ t\neq\tau}}^{T} \bar{u}^t_{i_t} \prod_{t=1}^{T} \tilde{u}^t_{j_t}$$

is independent of $\bar{u}^\tau_{i_\tau}$. Now, since problem (9) is convex, in order to maximize $\phi(\bar{\lambda},\tilde{\lambda})$ with respect to $\lambda^\tau_{i_\tau}$, we take the derivative of the function and set it to zero. This gives

$$0 = \frac{\partial\phi}{\partial\lambda^\tau_{i_\tau}} = \bar{\mu}_{i_\tau} - \bar{u}^\tau_{i_\tau}\bar{w}^\tau_{i_\tau},$$

from which we obtain the update rule

$$\bar{u}^\tau \leftarrow \bar{\mu} \oslash P_\tau\left(M^{(\bar{\lambda},\tilde{\lambda})}\right)\mathbf{1}_{n+1} \odot \bar{u}^\tau, \quad \tau = 1,\dots,T, \quad (10)$$

where $\oslash$ and $\odot$ denotes elementwise division and multiplication, respectively. Analogously, we obtain the update rule

$$\tilde{u}^\tau \leftarrow \tilde{\mu} \oslash P_\tau\left(M^{(\bar{\lambda},\tilde{\lambda})}\right)^\top \mathbf{1}_{m+1} \odot \tilde{u}^\tau, \quad \tau = 1,\dots,T. \quad (11)$$

C. Efficient Computation of Projections of Transport Tensors and Convergence of the Sinkhorn Algorithm

Iteratively performing the updates (10) and (11) is coordinate-wise ascent in the dual problem (9), i.e., it is a Sinkhorn algorithm for solving (7). However, the Sinkhorn iterations only partially alleviate the computational cost in the multimarginal setting. More specifically, (10) and (11) require the computation of $P_t(M^{(\bar{\lambda},\tilde{\lambda})})$. The latter involves nested sums which, if evaluated naively, requires $\mathcal{O}(m^T n^T)$ operations to evaluate. To derive a method for efficiently computing $P_t(M^{(\bar{\lambda},\tilde{\lambda})})$, let $k^t_{(i_t,j_t)} = \exp(-D^t_{(i_t,j_t)}/\varepsilon)$ and $\hat{k}_{(i_t,j_t),(i_{t+1},j_{t+1})} = \exp(-F_{(i_t,j_t),(i_{t+1},j_{t+1})}/\varepsilon)$, and note that by the structure of the cost in (5b), the tensor K can be factorized as

$$K_{(i_1,j_1),\dots,(i_T,j_T)} = \prod_{t=1}^{T} k^t_{(i_t,j_t)} \prod_{t=1}^{T-1} \hat{k}_{(i_t,j_t),(i_{t+1},j_{t+1})}.$$

Using this structure, we have the following result.

Theorem 2: $P_\tau(M^{(\bar{\lambda},\tilde{\lambda})})_{(i_\tau,j_\tau)}$ can be written as

$$P_\tau\left(M^{(\bar{\lambda},\tilde{\lambda})}\right)_{(i_\tau,j_\tau)} = \overrightarrow{\Phi}^\tau_{(i_\tau,j_\tau)} k^\tau_{(i_\tau,j_\tau)} \bar{u}^\tau_{i_\tau} \tilde{u}^\tau_{j_\tau} \overleftarrow{\Phi}^\tau_{(i_\tau,j_\tau)}, \quad (12)$$

where $\overrightarrow{\Phi}^\tau_{(i_\tau,j_\tau)}$ and $\overleftarrow{\Phi}^\tau_{(i_\tau,j_\tau)}$ are defined recursively as

$$\begin{aligned} \overrightarrow{\Phi}^\tau_{(i_\tau,j_\tau)} = \sum_{(i_{\tau-1},j_{\tau-1})} \overrightarrow{\Phi}^{\tau-1}_{(i_{\tau-1},j_{\tau-1})} \\ \cdot k^{\tau-1}_{(i_{\tau-1},j_{\tau-1})} \hat{k}_{(i_{\tau-1},j_{\tau-1}),(i_\tau,j_\tau)} \bar{u}^{\tau-1}_{i_{\tau-1}} \tilde{u}^{\tau-1}_{j_{\tau-1}}, \end{aligned} \quad (13)$$

with $\overrightarrow{\Phi}^1_{(i_1,j_1)} = 1$, and

$$\begin{aligned} \overleftarrow{\Phi}^\tau_{(i_\tau,j_\tau)} = \sum_{(i_{\tau+1},j_{\tau+1})} \overleftarrow{\Phi}^{\tau+1}_{(i_{\tau+1},j_{\tau+1})} \\ \cdot k^{\tau+1}_{(i_{\tau+1},j_{\tau+1})} \hat{k}_{(i_\tau,j_\tau),(i_{\tau+1},j_{\tau+1})} \bar{u}^{\tau+1}_{i_{\tau+1}} \tilde{u}^{\tau+1}_{j_{\tau+1}}. \end{aligned} \quad (14)$$

with $\overleftarrow{\Phi}^T_{(i_T,j_T)} = 1$.

Proof: With the definitions in (13) and (14), it can be readily shown that (12) holds. Details are omitted. ∎

The structure in (12)–(14) means that we have a way of computing the projections in $\mathcal{O}(m^2n^2)$ operations instead of $\mathcal{O}(m^T n^T)$ operations,[2] greatly improving the computational

[2]This can further be reduced to $\mathcal{O}(\max(m,n)\min(m,n)^2)$ by utilizing that infinite elements in F give zero elements in $\hat{k}$.

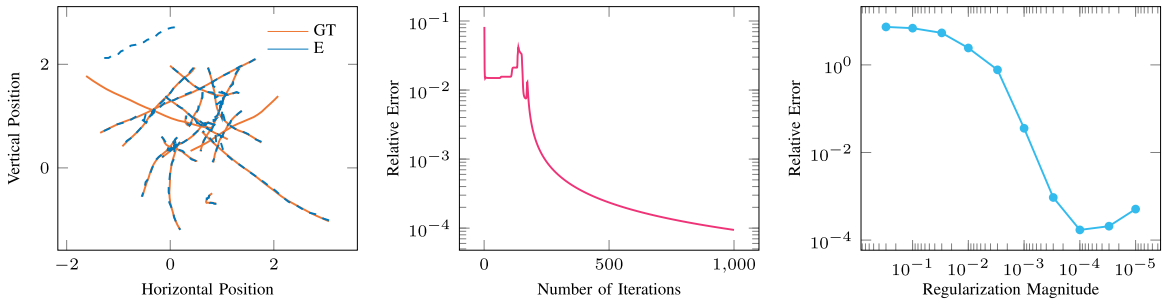


Fig. 2. **Left:** Simulated data. GT is the simulated ground truth data, and E is the simulated estimated trajectories. **Middle:** Convergence plot of Algorithm 1 on the data in Fig. 2 (left). **Right:** Relative error depending on the parameter η.

Algorithm 1 Block Coordinate Ascent Method for (9)

Initialize $\bar{u}$ and $\tilde{u}$ to one everywhere
while Not converged **do**
  $\overleftarrow{\Phi}^T \leftarrow 1$
  for $t = T-1, \dots, 1$ **do**
    Compute $\overleftarrow{\Phi}^t$ from $\overleftarrow{\Phi}^{t+1}$ using (14)
  end for
  $\overrightarrow{\Phi}^1 \leftarrow 1$
  for $t = 1, \dots, T-1$ **do**
    $\bar{u}^t \leftarrow \bar{\mu} \oslash P_t(M^{(\bar{\lambda},\tilde{\lambda})})\mathbf{1}_{n+1} \odot \bar{u}^t$, using (12)
    $\tilde{u}^t \leftarrow \tilde{\mu} \oslash P_t(M^{(\bar{\lambda},\tilde{\lambda})})^\top \mathbf{1}_{m+1} \odot \tilde{u}^t$, using (12)
    Compute $\overrightarrow{\Phi}^{t+1}$ from $\overrightarrow{\Phi}^t$ using (13)
  end for
  $\bar{u}^T \leftarrow \bar{\mu} \oslash P_T(M^{(\bar{\lambda},\tilde{\lambda})})\mathbf{1}_{n+1} \odot \bar{u}^T$
  $\tilde{u}^T \leftarrow \tilde{\mu} \oslash P_T(M^{(\bar{\lambda},\tilde{\lambda})})^\top \mathbf{1}_{m+1} \odot \tilde{u}^T$
end while

speed of the updates (10) and (11). Finally, the results from Section III-B and Theorem 2 yields Algorithm 1. By using [18], we get that the algorithm is convergent in the following sense (cf. [15, Prop. 1]).

Theorem 3: Algorithm 1 converges linearly to an optimal solution of (9).

Remark 4: From the discussion above, we note that one full iteration in Algorithm 1 requires $\mathcal{O}(Tm^2n^2)$ operations. While Theorem 3 shows that Algorithm 1 converges linearly, it neither gives an explicit rate in terms of the parameters of the problem, nor a bound on the number of iterations needed in order to get close to the optimal solution. Such bounds are known to exist for some Sinkhorn methods, see, e.g., [1], [8], [17], [21, Sec. 4.2]. Deriving such bounds for this problem would be interesting, but it is left for future work.

IV. Numerical Results

To demonstrate the efficiency of Algorithm 1, we test it on simulated data. We compare the algorithm with four off-the-shelf LP solvers, namely the well-known commercial solvers Gurobi, CPLEX and MOSEK, as well as the open source solver CLP [9]. For all simulations we use $c = 0.25$, $p = 1$, $\gamma = 1$ in the definition of the TGOSPA metric. Moreover, the value of ε needs to be selected in relation to the magnitude of the rest of the objective function (7a). For this reason, we take $\varepsilon = \eta \cdot T \cdot \max(\max(D), \max(F))$,[3] where η becomes the scaled regularization parameter. For stability reasons, we use the log-sum-exp rewriting and do all computations in the log-domain (see, e.g., [21, Sec. 4.4], [24]).

Data is generated using the algorithm for structured data described in [19, Sec. 5.1.1]. Figure 2 (left) shows an example of such data, simulated with parameters $m_t = 14$, $m_f = 2$, $n_f = 1$, $T = 20$, $r = 1$, $q = 0.9$, $c_s = 0.25$, $N_{\text{ts}} = 20$, $N_{\max} = 10^5$, and $\sigma = 0.01$. In the below results, we use the relative error $|f_{\text{opt}} - f|/f_{\text{opt}}$ as a measure of accuracy. Here, f_{opt} is the optimal value of the LP relaxation of (2), and $f = \langle C, M_{\text{iter}} \rangle$ where M_{iter} is the approximate transport plan either at the current iterate of Algorithm 1 or after termination of the algorithm. Which of the two interpretations of f that is used is clear from context. The value f_{opt} is computed using an LP solver.

Our method also allows for parallelization, and can thus be accelerated by running it on a GPU. The accelerated version is also included in the comparisons, and it should be noted that LP solvers can in general not be accelerated in the same way. The experiments were run on a AMD Ryzen 5 5600X 3701 MHz CPU and a NVIDIA GeForce GTX 1060 GPU. Our method was implemented using PyTorch, and the LP solvers were interfaced using the PuLP package in Python.

A. Illustrating Convergence and Effect of Regularization

To illustrate the behavior of the algorithm, we first apply it to the simulated data shown in Figure 2 (left). For $\eta = 10^{-5}$, the relative error as a function of the number of iterations of Algorithm 1 is shown in Figure 2 (middle). We note that most of the improvement happens in a transient phase, suggesting that early stopping should be considered when computational resources are limited.

Figure 2 (right) shows the effect of η on the result of Algorithm 1. In the simulations, we terminate the algorithm when the relative step size $||\boldsymbol{u} - \boldsymbol{u}_{\text{prev}}||/||\boldsymbol{u}_{\text{prev}}||$ is less than $1.5 \cdot 10^{-6}$. Here, $\boldsymbol{u}$ and $\boldsymbol{u}_{\text{prev}}$ are vectors constructed by concatenating all transformed dual variables from two consecutive iterations. As expected, small values of η leads to good results, and decreasing η generally leads to better approximations. The latter is true except for very small η. We believe that the this is due to that the convergence speed is expected to decrease with ε, i.e., with η (cf. Remark 4), but that we still use the same breaking criteria in all cases.

[3]The maximum over F is taken over all finite values of the tensor.

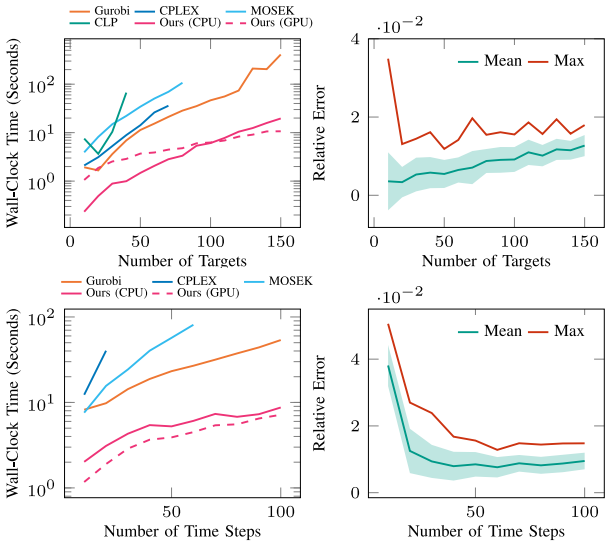


Fig. 3. Average runtimes and errors for Algorithm 1 on varying data sizes. The top row shows results for fixed number of time steps T and varying number of targets m. The bottom row shows results for varying number of time steps T and fixed number of targets m. The shaded regions in the right column is one standard deviation from the mean.

B. Computational Efficiency

In order to evaluate the computational efficiency of Algorithm 1, we vary the size of the input data and compare the wall-clock running time to those of the LP solvers. Note that the LP solvers compute the exact value of the LP-relaxed TGOSPA metric, while our algorithm only provides an approximation. For this reason, we also record the relative error compared to the exact solution. Using the data generation procedures with parameters $m_t = m - m_f$, $m_f = 0.1m$ (rounded to the nearest integer), $n_f = m_f$, $r = 1$, $q = 0.9$, $c_s = 0.25$, $N_{\text{ts}} = m$, $N_{\max} = 100m$, and $\sigma = 0.01$, we generate random scenarios for varying number of trajectories and time steps. For the case of varying number of trajectories, we fix $T = 100$, and for varying number of time steps, we fix $m = 100$. We use $\eta = 10^{-4}$, and terminate the algorithm when the relative step size is less than 10^{-3}.

For each set of simulation parameters we generate 25 scenarios. Averaged results over the scenarios are shown in Figure 3. Among the LP solvers, Gurobi was the fastest, and therefore it is the only one that was run for the largest scenarios. For example, CLP was not able to solve even the smallest scenario in the bottom row of Figure 3 in under two minutes. We see that for larger tracking scenarios, our algorithm computes solutions, with a mean relative error of around 1%, significantly faster than the LP solvers.

V. Conclusion

In this letter, we present an algorithm for fast approximation of the LP-relaxed TGOSPA metric. The algorithm is derived by reformulating the TGOSPA problem as an unbalanced multimarginal optimal transport problem, and leveraging ideas from the literature on entropy regularized optimal transport. Numerical results show that the method provides adequate approximations of the metric while significantly reducing computational costs. An interesting future direction for research would be to explore the possibility of utilizing the differentiability of Algorithm 1 (in the sense of reverse mode of automatic differentiation) to compute gradients with respect to the elements of the cost tensor, which would open a path toward data-driven MTT algorithms.

References

[1] J. M. Altschuler and E. Boix-Adsera, "Polynomial-time algorithms for multimarginal optimal transport problems with structure," *Mathe. Program.*, vol. 199, no. 1, pp. 1107–1178, 2023.
[2] F. Beier, J. von Lindheim, S. Neumayer, and G. Steidl, "Unbalanced multi-marginal optimal transport," *J. Math. Imag. Vis.*, vol. 65, no. 3, pp. 394–413, 2023.
[3] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, "Iterative Bregman projections for regularized transportation problems," *SIAM J. Sci. Comput.*, vol. 37, no. 2, pp. A1111–A1138, 2015.
[4] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 5–18, Jan. 2004.
[5] S. S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems*. Boston, MA, USA: Artech House, 1999.
[6] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[7] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transportation distances," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Assoc., Inc., 2013.
[8] J. Fan, I. Haasler, J. Karlsson, and Y. Chen, "On the complexity of the optimal transport problem with graph-structured cost," in *Proc. Int. Conf. Artif. Intell. Stat.*, 2022, pp. 9147–9165.
[9] J. Forrest et al., "Coin-or/Clp: Release/1.17.9." Oct. 2023. [Online]. Available: https://github.com/coin-or/Clp
[10] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association," *IEEE J. Ocean. Eng.*, vol. 8, no. 3, pp. 173–184, Jul. 1983.
[11] Á. F. García-Fernández, A. S. Rahmathullah, and L. Svensson, "A metric on the space of finite sets of trajectories for evaluation of multi-target tracking algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 3917–3928, 2020.
[12] Á. F. García-Fernández, A. S. Rahmathullah, and L. Svensson, "A time-weighted metric for sets of trajectories to assess multi-object tracking algorithms," in *Proc. IEEE 24th Int. Conf. Inf. Fusion (FUSION)*, 2021, pp. 1–8.
[13] T. T. Georgiou, J. Karlsson, and M. S. Takyar, "Metrics for power spectra: An axiomatic approach," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 859–867, Mar. 2009.
[14] I. Haasler, A. Ringh, Y. Chen, and J. Karlsson, "Multimarginal optimal transport with a tree-structured cost and the Schrödinger bridge problem," *SIAM J. Control Optim.*, vol. 59, no. 4, pp. 2428–2453, 2021.
[15] I. Haasler, A. Ringh, Y. Chen, and J. Karlsson, "Scalable computation of dynamic flow problems via multimarginal graph-structured optimal transport," *Math. Oper. Res.*, vol. 49, no. 2, pp. 986–1011, 2024.
[16] J. Krejcí, O. Kost, O. Straka, Y. Xia, L. Svensson, and Á. F. García-Fernández, "TGOSPA metric parameters selection and evaluation for visual multi-object tracking," 2024, *arXiv:2412.08321*.
[17] T. Lin, N. Ho, M. Cuturi, and M. I. Jordan, "On the complexity of approximating multimarginal optimal transport," *J. Mach. Learn. Res.*, vol. 23, no. 65, pp. 1–43, 2022.
[18] Z. Q. Luo and P. Tseng, "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Optim. Theory Appl.*, vol. 72, no. 1, pp. 7–35, Jan. 1992.
[19] V. N. Wernholm and A. Wärnsäter, "Efficient evaluation of target tracking using entropic optimal transport," M.S. thesis, Dept. Math. Sci., Chalmers Univ. Technol., Gothenburg, Sweden, 2024.
[20] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1982.
[21] G. Peyré and M. Cuturi, "Computational optimal transport: With applications to data science," *Found. Trends® Mach. Learn.*, vol. 11, nos. 5–6, pp. 355–607, 2019.
[22] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, "Generalized optimal sub-pattern assignment metric," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, 2017, pp. 1–8.
[23] A. Ringh, I. Haasler, Y. Chen, and J. Karlsson, "Graph-structured tensor optimization for nonlinear density control and mean field games," *SIAM J. Control Optim.*, vol. 62, no. 4, pp. 2176–2202, 2024.
[24] B. Schmitzer, "Stabilized sparse scaling algorithms for entropy regularized transport problems," *SIAM J. Sci. Comput.*, vol. 41, no. 3, pp. A1443–A1481, 2019.