Partial Optimal Transport for Co-Registration of Partially-Overlapped Point Clouds

Irvin Chadraa[1], Esteban Tabak[1], Debra F. Laefer[2,3]

[1] Courant Institute of Mathematical Sciences, New York University (NYU) New York, NY
(ic2184@nyu.edu, tabak@cims.nyu.edu)
[2] Dept. of Civil and Urban Engineering, Tandon School of Engineering, NYU, Brooklyn, NY
[3] Center for Urban Science and Progress, Tandon School of Engineering, NYU, Brooklyn, NY
(debra.laefer@nyu.edu)

Keywords: partial optimal transport, data co-registration, subset selection, noise reduction

Abstract

Optimal transport seeks the most efficient way to transform one probability distribution into another, typically under constraints that preserve total mass. However, in many practical applications, such as point cloud and image co-registration, the source and target data distributions may have unequal masses. Herein, this is overcome by relaxing the aforementioned constraints. This proposed "partial" optimal transport framework adaptively selects and matches subsets of both source and target distributions, thereby enabling robust outlier rejection and noise reduction. The method relaxes the strict constraints typically used in linear programming formulations of optimal transport, thus allowing flexibility on both sides of the matching problem. The resulting transport plan is then refined through branch-and-cut and mass bounding procedures that enforce binary mass assignments and further prune undesired points. For scalability, traditional linear programming solvers are replaced by an efficient gradient-based algorithm. This approach is validated on synthetic two-dimensional data and real three-dimensional point cloud data.

1. Introduction

Fully automated co-registration of non-contemporaneous point clouds remains a long-standing challenge in the remote sensing community. The line-of-sight nature of laser-scanning, the multiplicity of platforms from which it can be collected, and the constant upward trajectory of point density collection are all major factors (Vo et al., 2015). The problem is further compounded by the non-static nature of natural and manmade environments: vegetation grows, built infrastructure changes, and entities are transitory (e.g., humans, cars enter and leave scenes at high rates). Such modifications often defy the use of segmentation and object-detection-based co-registration strategies. The size of such data sets are also fundamentally incompatible with iterative point cloud matching. In an effort to overcome these difficulties, this paper proposes a partial optimal transport approach. In this paper, the focus will be to simultaneously address partially overlapping scenes that have different distributions of a single scene.

2. Background

Optimal transport (OT) was first conceived by Gaspard Monge (Monge, 1781) as a method of ideal transport of physical material, such as sand or soil, earning the common moniker "Earth Mover's problem". At its core, OT provides a principled way to compare two probability distributions by computing how much "effort" is needed to transform one into the other. Today, OT is applied to diverse areas of machine learning, including generative models, computer vision, and natural language processing.

The original formulation assumes two distributions: a source probability distribution μ on set X and a target probability distribution ν on set Y. Monge's formulation searches for a continuous map $T : X \to Y$ that pushes source points $x \in X$ to the target points $y \in Y$ such that the externally provided cost $c(x, T(x))$ of moving point x to $T(x)$ is minimized.

This idea is captured mathematically by Equation (1).

$$\min_T \left\{ \int_X c(x, T(x)) \, \mathrm{d}\mu(x) \mid T_{\#}(\mu) = \nu \right\} \tag{1}$$

Here $T_{\#}(\mu) = \nu$ denotes a measure-preserving pushforward. It simply ensures that the transported source mass matches the target exactly, with no mass lost or created in the process. Kantorovich later relaxed Monge's problem to instead find a joint probability distribution $\pi \in X \times Y$ that couples the source and target marginal distributions, μ and ν, respectively (Kantorovich, 1942). In the continuous setting described above, Kantorovich's problem seeks the minimum [Equation (2)]

$$\min_\pi \left\{ \int_{X \times Y} c(x, y) \, \mathrm{d}\pi(x, y) \,\middle|\, \pi \in \Pi(\mu, \nu) \right\} \tag{2}$$

where $\Pi(\mu, \nu)$ is the set of all couplings in which the marginal distributions are exactly μ and ν.

In a discrete setting, Kantorovich's formulation becomes a linear programming problem. The optimal transport map is the real-valued matrix $\Pi = (\pi_{ij}) \in \mathbb{R}^{n \times m}$ that minimizes the total cost C, a sum of pairwise costs $(c_{ij}) \in \mathbb{R}^{n \times m}$ of transport between weighted samples $\{x_i\}_{i=1}^n$ and $\{y_j\}_{j=1}^m$. The Euclidean distance can be used to measure transport cost. Irrespective of the selected measurement metric, each source point x_i has an associated mass $p_i = \mu(x_i)$, each target point y_j has mass $q_j = \nu(y_j)$, and masses in each marginal sum to one. The Kantorovich linear program is written in Equation (3):

$$\begin{aligned} \min_{\Pi \geq 0, \Pi \in \mathbb{R}^{n \times m}} \quad & \sum\nolimits_{i,j} c_{ij} \pi_{ij} \\ \text{subject to} \quad & \sum\nolimits_j \pi_{ij} = p_i \quad \forall i \in \{1, \ldots, n\} \\ & \sum\nolimits_i \pi_{ij} = q_j \quad \forall j \in \{1, \ldots, m\} \end{aligned} \tag{3}$$

That formulation represents the "full" optimal transport problem in a discrete setting. While it can be solved exactly, this comes at significant computational expense. As an alternative, Sinkhorn's algorithm (Cuturi, 2013) can be used to obtain an approximation with less computation.

Essid et al. (2019) proposed a sample-based adaptive OT method for framing full OT as a minimax game between a map $T = \nabla\phi$ and an adversary g to minimize the Kullback-Leibler divergence between the source and target distribution using a variational approach. By composing simple local maps between intermediate distributions, their method avoids density estimation and adaptively learns features from sample data, thereby making it suitable for high-dimensional tasks like point cloud co-registration.

However, classical optimal transport methods assume fully balanced measures with complete mass transfer, which is often restrictive in real-world scenarios. Several extensions have been proposed to address these limitations, particularly in the form of "partial" optimal transport (POT) or unbalanced OT. Caffarelli and McCann (2010) and Figalli (2010) established foundational results on the existence and uniqueness of continuous POT maps and characterize transport regions via free boundaries in Monge–Ampère-type formulations. These works provided important theoretical insights in a continuous space.

More recent efforts focus on computational methods for discrete POT. Riaz et al. (2023) introduced a semi-relaxed formulation with entropic regularization and a proximal solver, aimed at support subset selection on the source dataset. Chapel et al. (2020) reformulated POT as a standard OT problem through dummy mass embedding, enabling exact solvers for tasks such as positive-unlabeled learning. Subsequently, Bai et al. (2023) proposed a version of POT that reduced high-dimensional transport to a set of repeated one-dimensional problems.

Several works have also applied POT using Sinkhorn-based solvers for point cloud registration and domain adaptation. For example, Bešić et al. (2022) employed POT to compute soft correspondences between local descriptors in point clouds as part of a pose estimation pipeline. Similarly, Cattaneo et al. (2022) used Sinkhorn-regulated transport to align intermediate feature distributions across domains for better segmentation transfer. However, entropy-regularized formulations like Sinkhorn's algorithm yield dense, fractional mappings between points that are difficult to interpret in co-registration tasks.

3. Scope and Methodology

3.1 Discrete Partial Kantorovich

The discrete Kantorovich problem, as previously described, can be adapted to address partial optimal transport. In this formulation, it may be preferable to exclude the transportation of certain points if the associated cost becomes comparatively high (as would be happen if there is no matching point in the other dataset). Consequently, the problem is restructured to allow the redistribution of masses, modifying the source masses p_i into new masses $\tilde{p}_i$, and similarly, the target masses q_j into $\tilde{q}_j$. This reconfiguration involves completely removing mass at selected high-cost points and redistributing it among the remaining points. Since the masses are considered probabilities, the constraint that the masses of the source and target each sum to one is preserved.

The corresponding problem shown in Equation (4) reads like Equation (3), but with p_i and q_i replaced by

$$\tilde{p}_i = \begin{cases} 0 \\ \alpha p_i \end{cases} \quad \text{and} \quad \tilde{q}_j = \begin{cases} 0 \\ \beta q_j \end{cases} \tag{4}$$

with scaling parameters $\alpha, \beta \geq 1$ allowing adjustment of the extent of partial transport and the addition of the further constraints shown in Equation (5)

$$\sum_i \tilde{p}_i = \sum_j \tilde{q}_j = 1 \tag{5}$$

To tractably frame the problem into a form solvable by linear programming, the conditions in Equation (4) are relaxed into the inequalities $\tilde{p}_i \leq \alpha p_i$ and $\tilde{q}_j \leq \beta q_j$.

When the scaling parameters α and β equal one, the problem reverts to the original full transport formulation. As α and β grow, the formulation allows the remaining masses at more significant points to grow, while points with the highest transport costs (i.e., the furthest points) are ideally shrunk to a mass of zero and, thus, excluded from the transport plan. Very large values of α and β lead to an "extreme" partial transport, where only the closest points between the source and target sets are matched, effectively converting this formulation into a search for the minimizer of a nearest-neighbor problem. Riaz et al. (2023) also applied a similar scaling parameter in their partial optimal transport formulation, although they only applied this to the source data. In the formulation, herein, it is applied to both the source and the target datasets.

To balance the extent of partial optimal transport, a regularization term λ is introduced into the objective function, which provides the user with control over the degree of partiality in the transport solution. The selection of λ is critical, as it has a direct relationship with the extent of subset selection. Smaller values of λ couple fewer points while larger values revert the mapping to the full transport case. The magnitude of λ depends on the size of the dataset, with one value not necessarily yielding the same behavior for another. This makes the selection criteria inherently dependent on the preference of the user.

With these modifications, partial optimal transport can then be formulated as the following linear programming problem as shown in Equation (6):

$$\begin{aligned} \min_{\Pi \geq 0, \Pi \in \mathbb{R}^{n \times m}} \quad & \sum\nolimits_{i,j} c_{ij}\pi_{ij} + \lambda(\alpha + \beta) \\ \text{subject to} \quad & \alpha \geq 1, \quad \beta \geq 1, \quad \tilde{p}_i \leq \alpha p_i, \quad \tilde{q}_j \leq \beta q_j \\ & \sum\nolimits_j \pi_{ij} = \tilde{p}_i \quad \forall i \in \{1, \ldots, n\} \\ & \sum\nolimits_i \pi_{ij} = \tilde{q}_j \quad \forall j \in \{1, \ldots, m\} \\ & \sum\nolimits_i \tilde{p}_i = 1, \quad \sum\nolimits_j \tilde{q}_j = 1 \end{aligned} \tag{6}$$

Fig. 1 demonstrates the feasibility of this reformulation with uniformly weighted, synthetically generated source points (blue unfilled) and target points (red unfilled) in a two-dimensional coordinate plane. By selecting an arbitrary large value of λ, the full transport map depicted in Fig. 1b is obtained, where each target point accumulates mass from multiple source points, yielding a complex and soft coupling system (shown in green).

Using the partial optimal transport method outlined in Equation (6), only a subset of source (blue filled) and target (red filled) points are coupled in a one-to-one fashion, leaving the rest untouched. This is depicted in Fig. 1c. The subset selection is controlled by arbitrarily setting λ to 0.25. Different values of λ would yield different couplings. The case of extreme partial optimal transport is shown in Fig. 1d with λ set to the very small value of 0.01. In this case, only the closest pair of points in the system is matched, illustrating minimal transport distance.

3.2 Branch-and-Cut

The solution to the above linear programming problem may have points that satisfy strict inequalities $0 < \tilde{p}_i < \alpha p_i$ and $0 < \tilde{q}_j < \beta q_j$ that are allowed by the relaxation, yet remain undesirable. This soft coupling makes no practical sense in Light Detection and Ranging (LiDAR) point matching, as the presence of intermediate mass values indicates that certain points are neither fully included nor excluded in a co-registration setting. This issue is addressed through a finite-depth branch-and-cut procedure described below, which is made practical by the fact that the number of points where the inequality is strict is very small - an empirical observation that can be partially understood through the following theorem.

Theorem 1: For any value of λ, there always exists an optimal solution in which any two points indexed by i_1, i_2 satisfying the strict inequalities

$$0 < \tilde{p}_{i_1} < \alpha p_{i_1} \quad \text{and} \quad 0 < \tilde{p}_{i_2} < \alpha p_{i_2}$$

have disjoint targets, i.e. there is no j such that $\pi_{i_1,j} > 0$ and $\pi_{i_2,j} > 0$.

Proof: Suppose there are two points indexed by i_1 and i_2 that satisfy

$$0 < \tilde{p}_{i_1} < \alpha p_{i_1}, \quad 0 < \tilde{p}_{i_2} < \alpha p_{i_2},$$

and a target point j such that $\pi_{i_1,j} > 0$ and $\pi_{i_2,j} > 0$. If $c_{i_1,j} > c_{i_2,j}$, the solution cannot be optimal, since another solution with

$$\tilde{p}_{i_1} \to \tilde{p}_{i_1} - \delta, \quad \tilde{p}_{i_2} \to \tilde{p}_{i_2} + \delta,$$

$$\pi_{i_1,j} \to \pi_{i_1,j} - \delta, \quad \pi_{i_2,j} \to \pi_{i_2,j} + \delta,$$

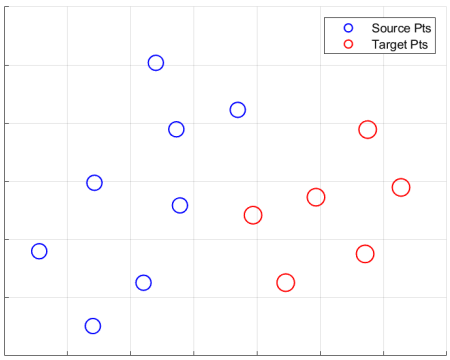
with $\delta > 0$ small enough, satisfies all constraints and has a smaller total cost. Similarly, if $c_{i_1,j} < c_{i_2,j}$, then a solution with lower cost has

$$\tilde{p}_{i_1} \to \tilde{p}_{i_1} + \delta, \quad \tilde{p}_{i_2} \to \tilde{p}_{i_2} - \delta,$$

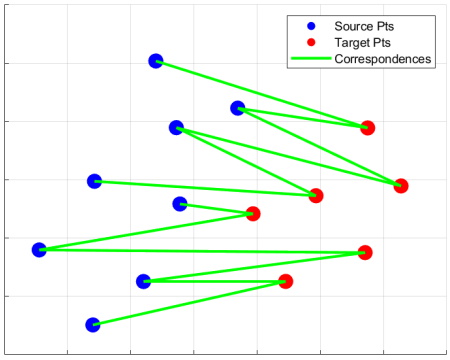
$$\pi_{i_1,j} \to \pi_{i_1,j} + \delta, \quad \pi_{i_2,j} \to \pi_{i_2,j} - \delta.$$

Finally, if $c_{i_1,j} = c_{i_2,j}$, then either of the above transformations preserves the total cost. In this case, either can be applied until one of the corresponding π reaches 0. □

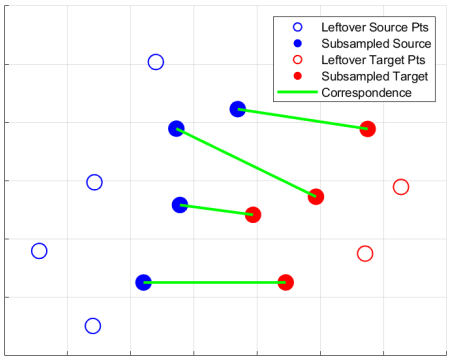
At each branch step, the point with the smallest unresolved mass is identified and two subproblems are generated: one which sets the mass to zero and the other to its maximum value (αp_i or βq_j). Both are solved and the point's status is then fixed based on the lower cost solution. The status of that point is then fixed for all subsequent branches. This process repeats until all mass assignments are resolved to binary values. With a limited number of intermediate cases, this refinement adds only modest computational cost while yielding clean, interpretable couplings. Algorithm 1 outlines this procedure methodically.



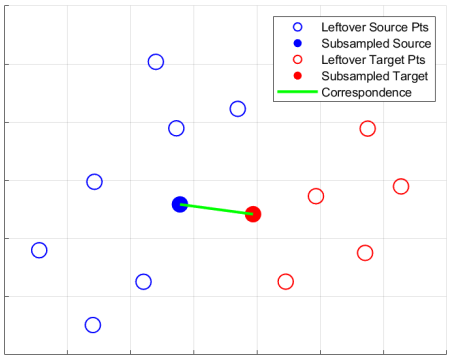
(a) Initial Data: Uniformly weighted source (blue) and target (red) points in two-dimensional space.



(b) Full Optimal Transport ($\lambda = 10$): All points are matched, producing dense, overlapping assignments.

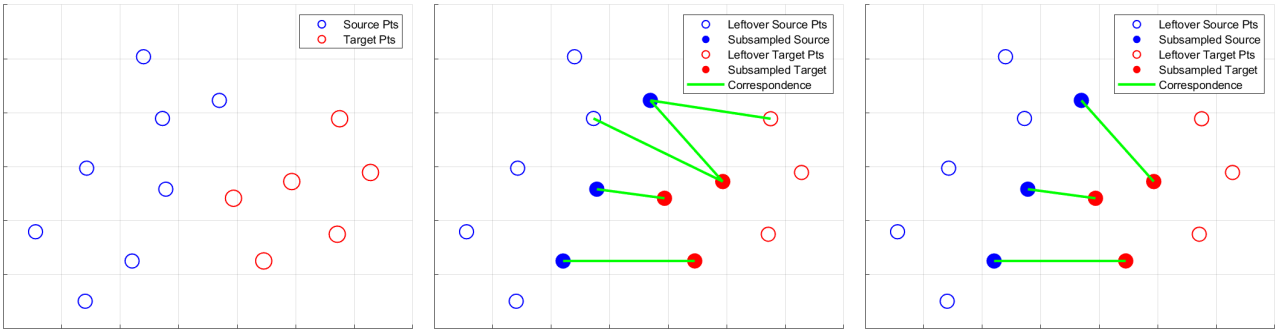


(c) Partial Optimal Transport ($\lambda = 0.25$): A subset of closer points are coupled, leaving others untouched.



(d) Extreme Partial Optimal Transport ($\lambda = 0.01$): Only the closest pair of points is matched.

Figure 1. Illustration of full, partial, and extreme partial optimal transport couplings between source and target sets under different values of λ using our proposed method.



(a) Initial Data (b) Without Branch-and-Cut ($\lambda = 0.06$) (c) With Branch-and-Cut ($\lambda = 0.06$)

Figure 2. In the same setting as Fig. 1, negligible-mass points may remain coupled by the transport map returned by the relaxed linear program form of partial OT depending on the selection of λ, but these are fully pruned with a branch-and-cut procedure.

Algorithm 1 Branch-and-Cut Procedure

Initialize list of active subproblems S with initial relaxed linear program
Set best solution $(\tilde{p}^*, \tilde{q}^*, T^*) \leftarrow$ null, objective $f^* \leftarrow \infty$
while S is not empty **do**
  Select and remove a subproblem from S
  Solve the linear program with fixed constraints
  if the problem is infeasible **then**
    continue
  end if
  Let $(\tilde{p}, \tilde{q}, T)$ be the solution with objective f
  if $f \geq f^*$ **then**
    continue
  end if
  if all entries of $\tilde{p}$ and $\tilde{q}$ are either 0 or full mass **then**
    Update $(\tilde{p}^*, \tilde{q}^*, T^*) \leftarrow (\tilde{p}, \tilde{q}, T)$, $f^* \leftarrow f$
    continue
  end if
  Select a partially assigned index i in $\tilde{p}$ or $\tilde{q}$
  Create two new subproblems by:
  - Fixing $\tilde{p}_i$ or $\tilde{q}_i$ to 0
  - Fixing $\tilde{p}_i$ or $\tilde{q}_i$ to full mass

  Add both subproblems to S
end while
return $(\tilde{p}^*, \tilde{q}^*, T^*)$

Fig. 2 compares results before and after employing the branch-and-cut procedure. Running the normal discrete partial optimal transport procedure with $\lambda = 0.06$ yields a partial mass deletion for one of the source points and one of the target points. Despite these points having a negligible mass, the normal procedure still attempts to couple them. Running the branch-and-cut method with the same regularization parameter entirely removes the negligible point masses (Fig. 2c). All remaining source points have equal mass, as do the remaining target points.

3.3 Implicit Gradient Descent

Traditional linear programming solvers work well for small-scale optimal transport problems, but struggle with scalability. As the size of the data set increases, memory requirements for storing and manipulating the constraint matrix A grow quadratically, which becomes particularly limiting in high-dimensional domains such as point clouds. Meanwhile, algorithms based on entropic smoothing, such as Sinkhorn (Cuturi, 2013), are computationally efficient but tend to produce diffuse, soft matchings that make it difficult to extract sharp, interpretable one-to-one correspondences between points.

To address these challenges, an implicit saddle-point optimization framework first introduced in Essid et al. (2023) is adopted. This method takes a game-theoretic perspective to solve saddle point problems as a "two-player game": one player (the set of primal variables) minimizes an established Lagrangian function, while the other (the set of dual variables) maximizes it. Unlike standard gradient descent, this method takes anticipatory steps that account for the future movement of the other player, resulting in curvature-aware updates that ensure convergence to a true saddle point.

This is begun by deriving the new objective by rewriting the linear programming problem in its canonical form as shown in Equation (7):

$$\min_{z \in \mathbb{R}^n} c^\top z \quad \text{subject to } Az \leq b, z \geq 0 \tag{7}$$

where $z \in \mathbb{R}^{n_z}$ are the primal variables
A is the constraint matrix
b is the constraint vector

As an analog to Equation (6), z contains all the primal variables for which optimization is sought: the entries of the transport map π_{ij}, the relaxed marginals $\tilde{p}$ and $\tilde{q}$, and they scaling parameters α and β.

The Lagrangian form of this problem introduces dual variables $w \in \mathbb{R}^{n_w}$ to enforce inequality constraints as seen in Equation (8):

$$\min_{z \geq 0} \max_{w \geq 0} L(z, w) = c^\top z - w^\top (Az - b) \tag{8}$$

A straightforward primal-dual gradient step looks like

$$\begin{pmatrix} z^{n+1} \\ w^{n+1} \end{pmatrix} = \begin{pmatrix} z^n \\ w^n \end{pmatrix} - \eta J G^n, \tag{9}$$

where η is the learning rate
J is the identity $\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$
G^n is the gradient $\begin{pmatrix} L_x \\ L_y \end{pmatrix}$ at z^n and w^n.

However, updating these variables via simple gradient descent may fail to converge to a saddle point. Instead, an implicit update as shown in Equation (10) is applied that accounts for future positions.

$$\begin{pmatrix} z^{n+1} \\ w^{n+1} \end{pmatrix} = \begin{pmatrix} z^n \\ w^n \end{pmatrix} - \eta J G^{n+1} \tag{10}$$

This update typically does not have a closed-form expression. Instead, the future position G^{n+1} is estimated based on current values of x^n and y^n using a first-order Taylor approximation

$$G^{n+1} \approx G^n + H^n \left[\begin{pmatrix} z^{n+1} \\ w^{n+1} \end{pmatrix} - \begin{pmatrix} z^n \\ w^n \end{pmatrix} \right] \tag{11}$$

with H^n as the Hessian $\begin{pmatrix} L_{zz} & L_{zw} \\ L_{wz} & L_{ww} \end{pmatrix}$ evaluated at z^n and w^n.

The implicit update then yields the following procedure:

$$\begin{pmatrix} z^{n+1} \\ w^{n+1} \end{pmatrix} = \begin{pmatrix} z^n \\ w^n \end{pmatrix} - \eta J G^n - \eta J H^n \left[\begin{pmatrix} z^{n+1} \\ w^{n+1} \end{pmatrix} - \begin{pmatrix} z^n \\ w^n \end{pmatrix} \right] \tag{12}$$

which reduces into the closed form

$$\begin{pmatrix} z^{n+1} \\ w^{n+1} \end{pmatrix} = \begin{pmatrix} z^n \\ w^n \end{pmatrix} - \eta \left(J + \eta H^n \right)^{-1} G^n \tag{13}$$

The variables are then reparameterized

$$Z(z) = z^2, \quad W(w) = w^2 \tag{14}$$

yielding the new objective:

$$\min_{z \geq 0} \max_{w \geq 0} \; L(z, w) = \min_{z} \max_{w} c^\top Z(z) - W(w)^\top (AZ(z) - b) \tag{15}$$

Reparameterizing the primal and dual variables ensures that positivity constraints are satisfied by construction and gives a non-degenerate Hessian suitable for implicit updates. Other mappings like exponential can be used, though squaring offers a straightforward default.

For high-dimensional data, the leading computational cost of this method comes from calculating the Hessian and the inverse $(J + \eta H)^{-1}$, but these can be addressed with relatively low-cost strategies analogous to predictor-corrector schemes in ODEs and quasi-Newton methods. Although this implicit update method incurs a higher cost per-iteration, it offers extensibility to more complex, regularized, or learning-driven transport problems where LP formulations become less practical. Further details and extensive theoretical guarantees of local convergence can be found in Essid et al. (2023), including a case study on the adaptive optimal transport problem introduced in Essid et al. (2019).

3.4 Mass Bounding

In practice, gradient-based optimization for partial optimal transport often results in mass values that approach but never exactly reach a uniform value. This ambiguity makes it the difficult to interpret which points are truly active (selected) or inactive (excluded). To resolve this, a bounding procedure governed by a small tolerance parameter ϵ is introduced, which helps to refine these intermediate values. After solving the relaxed problem, some mass values lie close to the extremes (0 or the maximum uniform weight), but due to numerical noise or early stopping, they rarely become exactly binary.

The bounding step addresses this by snapping any mass less than ϵ to zero, while any mass within ϵ of the maximum uniform weight is set to its full value. This immediately classifies many points as either active or inactive, leaving only a small subset of points in between. For these unresolved points, the same branch-and-cut refinement as before is applied. While setting tolerance ϵ to the halfway point between 0 and the maximum uniform weight is possible, that could yield a suboptimal solution as it circumvents the branch-and-cut procedure. Ultimately, this mass-bounding step helps transition from soft, continuous solutions to hard, interpretable selections, while maintaining sensitivity to the transport cost.

3.5 Adaptive Learning Rate

In saddle-point optimization, particularly in non-convex settings, a fixed learning rate η can lead to unstable updates or convergence to undesirable critical points. To counteract overshooting or stagnation, η is adaptively adjusted using a procedure that extends the classic backtracking line search algorithm (Nocedal and Wright, 1999).

The proposed implementation, enforces two directional (also known as Armijo-Goldstein) conditions as per Equation (16):

$$L(z_{n+1}, w_n) \leq L(z_n, w_n) \leq L(z_n, W_{n+1}) \tag{16}$$

(1) the Lagrangian must decrease when updating the set of primal variables z while keeping the set of dual variables w fixed; and (2) the Lagrangian must increase when updating w, while keeping z fixed. If either of these conditions is violated, the learning rate is iteratively halved until both are satisfied or a maximum number of retries is reached. This ensures stable progress to the optimal saddle point, even in highly nonlinear or ill-conditioned settings.

3.6 Preconditioning

To improve numerical stability and reduce the influence of outliers on the partial OT procedure, a preconditioning step is applied that robustly rescales and centers the set of source points x onto the target y. It uses a robust standardization method based on the median and interquartile range instead of the mean and standard deviation, which are more heavily skewed by extreme values. This reformulation is given by Equation (17):

$$x_{\text{pre}} = \frac{x - Q_2(x)}{Q_3(x) - Q_1(x)} \times (Q_3(y) - Q_1(y)) + Q_2(y) \tag{17}$$

Here, Q_1, Q_2, and Q_3 are the three quartiles (25th, 50th, 75th percentile) of the data. This transformation allows the implicit descent procedure to converge more reliably and efficiently to the local optimum. This step acts as a lightweight but highly effective enhancement to the overall transport pipeline, particularly in noisy, real-world scenarios where assumptions of global alignment may not hold.

4. Results

This proposed method is evaluated using LiDAR point cloud data collected by Laefer and Vo (2019), which includes scans of distinct urban objects including houses, trees, and light poles that were manually extracted to test the proposed approach. To ensure spatial coherence across different experimental setups, additional preprocessing was conducted so that the objects were positioned appropriately nearby, and their coordinate systems were aligned so all structures would lie on approximately the same plane. This setup emulates real-world multi-object scans where irrelevant structures may be present near objects of interest. All experiments were conducted using MATLAB R2023a on a system with 13th Gen Intel Core i7-13700H CPU and 16GB RAM.

4.1 Tree and Light Pole Dataset

Fig. 3 illustrates the point clouds of both the tree and light pole in CloudCompare, as well as the subsampled datasets used in the partial optimal transport experiments in a three-dimensional coordinate plane. From the original point cloud of each object, two subsets were randomly extracted. One was selected as the source and the other as the target. Although the original tree point cloud contained more than 100,000 points, both subsets were artificially limited to only 925 points each to accommodate constrained computing environments with limited processing and memory.

To introduce structural variation and simulate realistic challenges, the source data set was sampled exclusively from the tree and uniformly scaled by a factor of 2 – simulating scale misalignment and testing the robustness of both the proposed preconditioning and transport mechanisms. The target data set includes an additional 150 points sampled from the light pole dataset (originally 18,000 plus points). This additional structure introduces partial mismatch and noise, presenting a challenging case for correspondence estimation and making the co-registration task more representative of real-world scenarios involving occlusion or outlier presence.

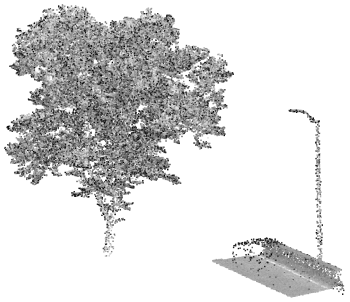
Fig. 3d shows the output of the proposed transport; setting the scaling term to $\lambda = 3$. At this level, only a subset of the tree points are involved in the coupling, while all points associated with the light pole are successfully ignored. Remaining masses are uniform and ambiguous mass values are pruned through the bounding and branch-and-cut steps, preserving interpretable correspondences across domains.

4.2 House and Tree Dataset

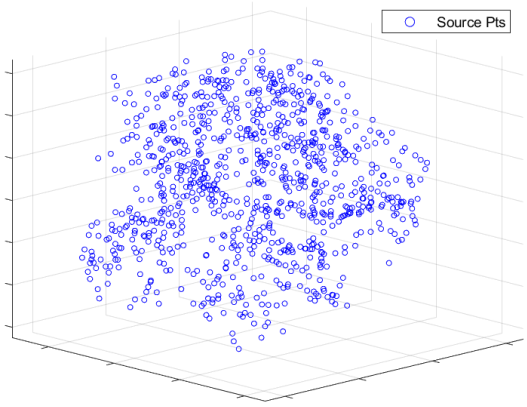
To further validate the method's robustness across different geometric structures, additional experiments comparing the performance of the partial optimal transport approach against several established co-registration methods were conducted using a house and tree point cloud from the previously cited data set. This configuration tested the method's ability to selectively register distinct architectural features while ignoring irrelevant natural structures.

The original house point cloud contains over 250,000 points while the original tree point cloud remains the same as the previous section. After a similar preprocessing stage as the tree and light pole example, the source data randomly sampled 925 points from the house point cloud plus 150 points randomly sampled from the tree point cloud. Both components are drawn at their original scale. The target data samples were only 925 points from the house point cloud and then uniformly scaled by a factor of 2.

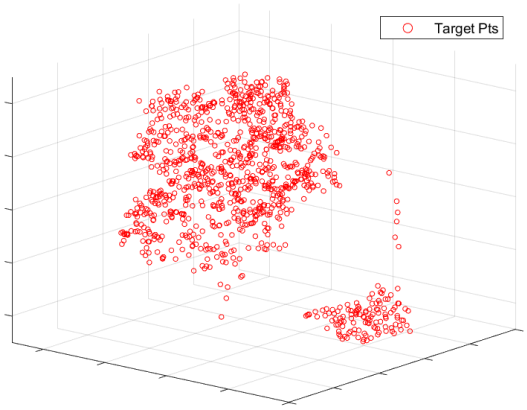
Fig. 4 contains visuals for this point cloud and co-registration results of the Hungarian Algorithm, Robust Point Matching, and partial OT. For fair comparison, all algorithms used identical convergence criteria (tolerance = 1e-6, maximum iterations = 100). Correspondence visuals (shown in green) provide valuable, intuitive understanding of the relative performance of each algorithm which is difficult to convey quantitatively, considering the cost metrics for each method are unique.



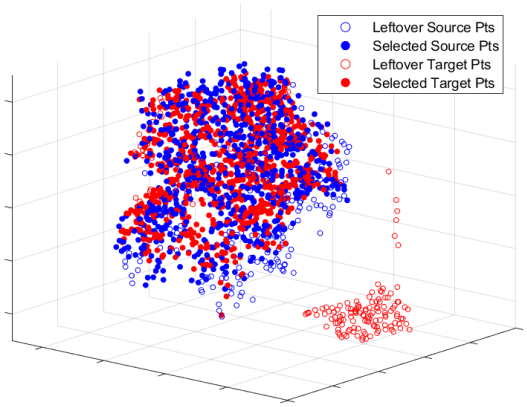
(a) Original tree and light pole point cloud



(b) Initial source: tree only (rescaled to 2x size)

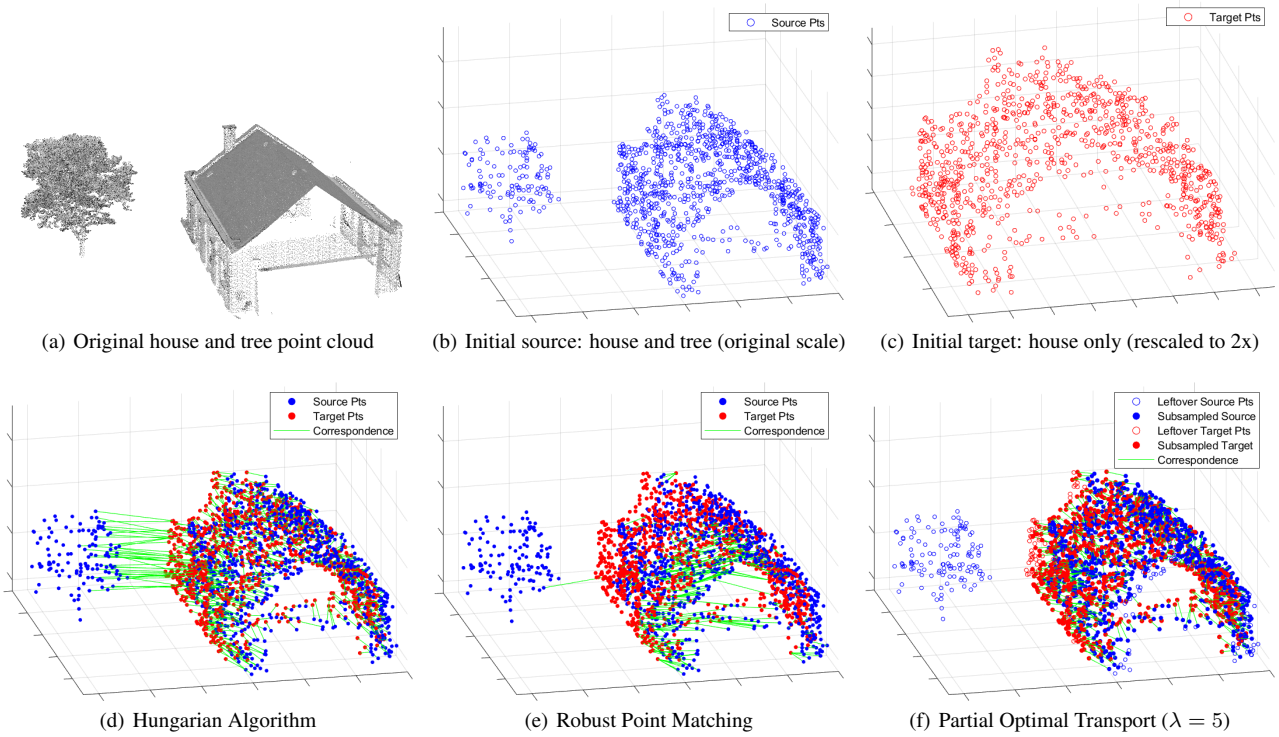


(c) Initial target: tree and light pole (original scale)



(d) Final transport ($\lambda = 3$): only the tree points are coupled

Figure 3. Despite additional noisy structures in the target set, the method correctly couples only the tree structure while discarding the light pole, illustrating robust geometric selectivity.



(a) Original house and tree point cloud (b) Initial source: house and tree (original scale) (c) Initial target: house only (rescaled to 2x)

(d) Hungarian Algorithm (e) Robust Point Matching (f) Partial Optimal Transport ($\lambda = 5$)

Figure 4. Qualitative results of co-registration methods on the house-tree dataset. Partial optimal transport effectively couples the house structure only, demonstrating stronger geometric selectivity compared to traditional co-registration methods.

4.3 Comparison of Co-Registration Methods

The proposed partial OT approach was benchmarked against two previously published techniques. The first is the Hungarian Algorithm, also known as the Munkres algorithm (Kuhn, 1955), which solves the classical assignment problem (a special case of optimal transport) and can be modified to solve unbalanced assignment problems like partial co-registration. Fig. 4d conveys the results of this algorithm, which shows the most indiscriminate correspondence behavior among the three methods, yielding suboptimal assignments that map several house points to the noisy tree structure.

Partial OT was also compared against Robust Point Matching (Yew and Lee, 2020), which aims to establish soft correspondences between point sets and handles outliers through an annealing process. The implementation herein initialized the annealing parameter $\beta = 0.1$, which was increased by a factor of 1.05 per iteration (capped at 50) and combined with a penalty weight $\kappa = 1$ for outliers. Fig. 4e shows results with strong geometric selectivity, as it ignores most tree points, though it includes lengthy correspondences between house structures.

In contrast, partial OT with scaling parameter $\lambda = 5$ effectively eliminates all excess connections to irrelevant geometric features, as evidenced by Fig. 4f, which shows concentrated coupling within the house structure and complete avoidance of any tree points. These correspondence lengths are shorter than those found in Robust Point Matching (as shown in Fig. 4f).

5. Discussion

This paper presents an interpretable framework for discrete partial optimal transport by extending classical OT with a series of algorithmic innovations tailored for point cloud co-registration. The main contributions are as follows.

- A fully relaxed linear programming formulation of partial OT that allows both source and target distributions to vary within bounded mass constraints.
- A regularization term λ to explicitly control the extent of partial transport induced through two learned scaling parameters α and β, unifying the behavior of full and extreme transport plans in a single framework.
- A branch-and-cut refinement that converts soft transport plans into binary support selections, thereby filtering outliers and non-informative points.
- Replacement of memory-intensive linear program solvers with an implicit gradient descent method that is curvature-aware and does not require entropic smoothing.

Despite its strengths, the proposed method is comparatively computationally expensive (273.21 seconds runtime versus 75.47 for the Hungarian Algorithm and 6.96 for the Robust matching). This highlights a trade-off between computational efficiency and correspondence quality. Although partial OT requires significantly more processing time, its strong performance in partial correspondence scenarios and robust handling of geometric outliers make it particularly valuable for applications where registration quality is prioritized over speed.

The performance of partial OT is sensitive to hyperparameter selection, particularly the scaling parameter regularization term λ, the mass bounding coefficient ϵ, and the learning rate η for the gradient descent procedure. Careful tuning is required to avoid the under- or over-pruning of point masses. Sensitivity analysis and a methodical hyperparameter selection process are important future areas to explore.

Finally, the proposed method relies on the preconditioning step to bring the source and target point clouds into approximate alignment. Although the implicit gradient method accounts for mild misalignment, it may struggle in scenarios with significant structural divergence (e.g., introducing a comparatively large structure like a skyscraper in only the source or target dataset but not in other). Notably, traditional full optimal transport and other co-registration methods would also struggle in such regimes due to inherent assumptions of full support matching.

6. Conclusions

This work offers a step toward making optimal transport more adaptable to practical geometric tasks, particularly in settings where full support matching is neither feasible nor desirable. By unifying relaxation, refinement, and curvature-aware optimization, this work advances the use of partial optimal transport to a deployable, scalable tool for co-registration tasks where selective, interpretable matching is critical.

Some interesting directions for future research include:

- Applying this method to voxelized or mesh-based versions of point clouds to reduce memory usage and increase robustness to sampling density variations.
- Incorporating rotation- and scale-invariant cost functions or embedding cost-free transformation to the optimization loop could reduce dependence on manual preconditioning.
- Integrating this method into GPU-accelerated neural pipelines could decrease its computational cost and significantly increase its applicability in the broader remote sensing community.

7. Acknowledgements

The work of Esteban Tabak and Debra Laefer was partially supported by ONR #13456472. The authors thank Sofia Poinelli for valuable discussions on optimal transport during the very early stages of this project. The authors would also like to acknowledge Anni Zheng for assisting in the curation of the processed point cloud data used in the experiments.

References

Bai, Y., Schmitzer, B., Thorpe, M., Kolouri, S., 2023. Sliced optimal partial transport. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13681–13690.

Bešić, B., Gosala, N., Cattaneo, D., Valada, A., 2022. Unsupervised domain adaptation for LiDAR panoptic segmentation. *IEEE Robotics and Automation Letters*, 7(2), 3404–3411. doi.org/10.1109/LRA.2022.3147326.

Caffarelli, L. A., McCann, R. J., 2010. Free boundaries in optimal transport and Monge-Ampere obstacle problems. *Annals of Mathematics*, 673–730. doi.org/10.4007/annals.2010.171.673.

Cattaneo, D., Vaghi, M., Valada, A., 2022. LCDNet: Deep loop closure detection and point cloud registration for LiDAR SLAM. *IEEE Transactions on Robotics*, 38(4), 2074–2093. doi.org/10.1109/TRO.2022.3150683.

Chapel, L., Alaya, M. Z., Gasso, G., 2020. Partial optimal tranport with applications on positive-unlabeled learning. *Advances in Neural Information Processing Systems*, 33, 2903–2913. doi.org/10.48550/arXiv.2002.08276.

Cuturi, M., 2013. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, 26. doi.org/10.48550/arXiv.1306.0895.

Essid, M., Laefer, D. F., Tabak, E. G., 2019. Adaptive optimal transport. *Information and Inference: A Journal of the IMA*, 8(4), 789–816. doi.org/10.1093/imaiai/iaz008.

Essid, M., Tabak, E. G., Trigila, G., 2023. An implicit gradient-descent procedure for minimax problems. *Mathematical Methods of Operations Research*, 97(1), 57–89. doi.org/10.48550/arXiv.1906.00233.

Figalli, A., 2010. The optimal partial transport problem. *Archive for Rational Mechanics and Analysis*, 195(2), 533–560. doi.org/10.1007/s00205-008-0212-7.

Kantorovich, L. V., 1942. On the translocation of masses. *Dokl. Akad. Nauk. USSR (NS)*, 37, 199–201.

Kuhn, H. W., 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2), 83–97. doi.org/10.1002/nav.3800020109.

Laefer, D. F., Vo, A.-V., 2019. 2019 LiDAR Data Collection for Sunset Park, Brooklyn, NY. doi.org/10.17609/4cpx-2h33.

Monge, G., 1781. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, 666–704. doi.org/10.1112/plms/s1-14.1.139.

Nocedal, J., Wright, S. J., 1999. *Numerical Optimization*. Springer.

Riaz, B., Karahan, Y., Brockmeier, A. J., 2023. Partial optimal transport for support subset selection. *Transactions on Machine Learning Research*.

Vo, A.-V., Truong-Hong, L., Laefer, D. F., Bertolotto, M., 2015. Octree-based region growing for point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 88–100. doi.org/10.1016/j.isprsjprs.2015.01.011.

Yew, Z. J., Lee, G. H., 2020. Rpm-net: Robust point matching using learned features. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11824–11833.