Multiview Sensing with Unknown Permutations: An Optimal Transport Approach

Ma, Yanting; Boufounos, Petros T.; Mansour, Hassan; Aeron, Shuchin

TR2021-047   May 11, 2021

Abstract

In several applications, including imaging of deformable objects while in motion, simultaneous localization and mapping, and unlabeled sensing, we encounter the problem of recovering a signal that is measured subject to unknown permutations. In this paper we take a fresh look at this problem through the lens of optimal transport (OT). In particular, we recognize that in most practical applications the unknown permutations are not arbitrary but some are more likely to occur than others. We exploit this by introducing a regularization function that promotes the more likely permutations in the solution. We show that, even though the general problem is not convex, an appropriate relaxation of the resulting regularized problem allows us to exploit the well-developed machinery of OT and develop a tractable algorithm.

MULTIVIEW SENSING WITH UNKNOWN PERMUTATIONS:
AN OPTIMAL TRANSPORT APPROACH

Yanting Ma*, Petros T. Boufounos*, Hassan Mansour*, Shuchin Aeron†

*Mitsubishi Electric Research Laboratories (MERL),{*yma,petrosb,mansour*}*@merl.com*.
†Electrical and Computer Engineering, Tufts University, *shuchin@ece.tufts.edu*.

ABSTRACT

In several applications, including imaging of deformable objects while in motion, simultaneous localization and mapping, and unlabeled sensing, we encounter the problem of recovering a signal that is measured subject to unknown permutations. In this paper we take a fresh look at this problem through the lens of optimal transport (OT). In particular, we recognize that in most practical applications the unknown permutations are not arbitrary but some are more likely to occur than others. We exploit this by introducing a regularization function that promotes the more likely permutations in the solution. We show that, even though the general problem is not convex, an appropriate relaxation of the resulting regularized problem allows us to exploit the well-developed machinery of OT and develop a tractable algorithm.

Index Terms— Multiview Sensing, Unlabeled Sensing, Optimal Transport

1. INTRODUCTION AND MOTIVATION

In this paper, we explore the problem of recovering a signal measured through a linear system while undergoing a partially known permutation. The problem arises in many contexts, including coherent imaging, simultaneous localization and mapping (SLAM) and partially labeled sensing, among others. Our motivation is coherent imaging of moving deformable targets under partially known or partially observable deformations. This has a number of applications, including radar imaging of humans in motion, and in-vivo coherent imaging of moving cells, organisms and live organs, such as beating hearts and breathing lungs.

A common feature in these applications is that, under mild assumptions, the motion can be described as a transformation—typically a permutation of pixels with respect to a reference view—of the reflectivity of the object in front of the measurement system. We assume that the motion is partially known up to a small error, modeled as an unknown permutation from the true position to the assumed known position. In practice, motion is typically estimated using an auxiliary measurement system or a motion model [1–5].

This problem has strong connections to the general problem of unlabeled sensing or shuffled linear regression [6–13], in which the labels or indices of the data are not available during the data acquisition process. In other words, the data has undergone a completely unknown random permutation (shuffling) that also needs to be recovered during reconstruction. The motivating example for this category of problems is typically SLAM [6, 8]. However, the unknown shuffling is rarely completely unknown in practice, even in the case
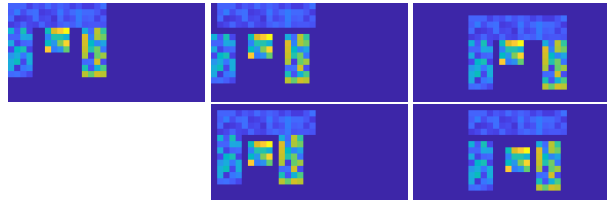
Fig. 1. Illustrative example of our set up. Left: Signal $\mathbf{x}$. Middle and Right: (top) estimates $\mathbf{F}_i\mathbf{x}$ of two different permutations of $\mathbf{x}$, (bottom) actual permutations $\mathbf{x}_i = \mathbf{P}_i\mathbf{F}_i\mathbf{x}$ of $\mathbf{x}$ observed by the acquisition system.

of SLAM. Typically, there is prior knowledge of the permutation, maybe up to small errors, e.g., [14].

Our paper makes the following contributions:

1. A formulation that explicitly regularizes the permutation to be estimated. This incorporates prior knowledge of the permutation, reducing the search space, and, in principle, making the problem easier to solve.

2. A relaxation that allows us to use the well-developed theoretical and algorithmic machinery of optimal transport (OT) [15] in obtaining a solution.

3. A generalization of the unlabeled sensing problem, introducing an optional linear operator measuring the permuted data. This model is more appropriate for coherent imaging under motion, our motivating application.

In the next section we introduce the problem and key assumptions. Section 3 describes prior work on unlabeled sensing, in the context of this paper. In Section 4 we outline our approach to solving the problem, drawing insights from the theory and methods of OT [15–18]. In Section 6 we present results on synthetic data validating our approach. Section 7 discusses our findings and concludes.

2. ACQUISITION MODEL

We assume multiple shots $\mathbf{y}_i$ of a reference reflectivity image $\mathbf{x} \in \mathbb{R}^N$ are acquired through

$$\mathbf{y}_i = \mathbf{A}_i\mathbf{P}_i\mathbf{F}_i\mathbf{x} + \mathbf{n}_i, \ \ i = 1, 2, ...K \tag{1}$$

where,

1. $\mathbf{A}_i$ are known linear measurement operators, modeling the acquisition system,

2. $\mathbf{F}_i$ are known operators that partly predict the deformation of $\mathbf{x}$, into $\mathbf{x}_i$, before it is acquired, and

3. $\mathbf{P}_i$ are *unknown* permutation matrices modeling the uncertainty in our knowledge of $\mathbf{F}_i$.

In the context of our motivating example, $\mathbf{A}_i$ describes the coherent imaging system, such as radar, magnetic resonance, or ultrasound. The operators $\mathbf{F}_i$ describe the object motion, e.g., of the heart or the lung, and are typically estimated using auxiliary information, such as a motion model or an alternative sensing modality; we do not explore how these are estimated in this paper. The unknown permutations $\mathbf{P}_i$ model the correction to the estimation error of $\mathbf{F}_i$, and may be estimated either implicitly or explicitly. Our ultimate goal is to recover $\mathbf{x}$ from the measurements $\mathbf{y}_i$. Estimating the $\mathbf{P}_i$ is incidental, and we are not interested in the quality of this estimation.

Figure 1 provides an illustrative stylized example of our set up. The letter "E" is observed under motion of each of its components. The top left shows the image in the reference pose. The top middle and right show the image under the estimated transformation, i.e., $\mathbf{F}_i\mathbf{x}$, and the bottom middle and right show the actual image $\mathbf{x}_i$ that the acquisition system acquires through $\mathbf{A}_i$. The assumption is that the latter is an unknown limited permutation of the former, i.e., $\mathbf{x}_i = \mathbf{P}_i\mathbf{F}_i\mathbf{x}$, as shown in the figure. The goal is to recover the reference image in the top left of the figure.

To this end we make the following mild assumptions.

(a) The support of $\mathbf{x}$ is known;

(b) Perturbations $\mathbf{P}_i$ that move pixels far from their estimated position in the 2D image domain are less likely.

In the next section we will briefly review some of the related work and contrast our problem formulation. Although related, there are some salient aspects of our set up that makes it unique and that renders itself to novel solution approaches.

3. RELATED WORK

Our model relates to certain signal estimation problems with unknown permutations known as *unlabeled sensing* or *shuffled linear regression* [6–13]. The unlabeled sensing problem can be defined as finding a signal $\mathbf{x}$ and a general permutation $\mathbf{P}$, such that

$$\mathbf{y} = \mathbf{P}\mathbf{B}\mathbf{x} + \mathbf{n}, \qquad (2)$$

where $\mathbf{B}$ is an $M \times N$ matrix with column rank N, and $N < M$.

These problems are instances of well-studied data association and assignment problems [19], [15, Ch. 2]. OT, a key component of our solution, is one approach to handling the 2-D assignment problem. The problem we address is an instance of an N-D (N$>$ 2) assignment problem [20, 21], which is harder to solve in general. Our approach solves smaller 2-D assignment problems, effectively, computing a *barycenter*, an approach that is again inspired by and is related to the OT barycenter problem [22]. Thus we avoid computationally expensive methods that solve the N-D problem [20, 21].

In addition, our formulation has a key difference that necessitates our approach: we target a more general form of the unlabeled sensing problem which includes the (possibly underdetermined) linear measurement operator $\mathbf{A}$. This precludes the aforementioned approaches. Still, if we specify $\mathbf{A}$ as the identity operator, we solve a regularized form of the problem, incorporating additional knowledge that some permutations are more likely to occur.

Specifically, we observe multiple permuted transformations of our signal, wherein permutations for each view are less likely to occur the farther they diverge from the identity operator. Here, we may consider having an equivalent operator to the matrix $\mathbf{B}$ which can be composed by stacking the multi-view transformation matrices $\mathbf{F}_i$

into a single matrix. We also note that, we are only interested in recovering the reflectivity, not necessarily the permutation.

The variation of the unlabeled sensing problem we consider, allows us to develop a computationally efficient algorithm, described in Section 4. Experimentally, our approach performs well under our assumptions, even in the presence of noise, overcoming existing pessimistic computational results for (2).

4. RECOVERY USING PERMUTATION REGULARIZATION

Considering that all permutations are not equally likely in inverting (1), it becomes natural to consider formulations in which the permutation estimation is regularized. In particular, to estimate $\mathbf{P}_i$ and $\mathbf{x}$ from (1), the following is a reasonable formulation:

$$\min_{\mathbf{P}_i \in \mathcal{P}, \mathbf{x}} \sum_{i=1}^{K} \left(\frac{1}{2}\|\mathbf{y}_i - \mathbf{A}_i\mathbf{P}_i\mathbf{F}_i\mathbf{x}\|_2^2 + \beta R(\mathbf{P}_i) \right), \qquad (3)$$

where $\mathcal{P}$ is a set of $N \times N$ permutation matrices and $R(\mathbf{P}_i)$ is a regularization for $\mathbf{P}_i$, which incorporates the prior knowledge that $\mathbf{P}_i$ is more likely to move elements to nearby pixels rather than distant ones, following by Assumption (b).

In particular, $R(\mathbf{P}_i)$ should penalize permutations that move pixels very far from the original position, as measured in the underlying grid of the signal. Using $l[n]$ to denote the true position of coefficient n in the underlying grid $\mathcal{L}$, we assume the cost of permuting coefficient n to position n' to be the squared Euclidean distance in the underlying grid, $\|l[n] - l[n']\|_2^2$. For example, if the underlying grid is 2-dimensional, as in the examples in Section 6, $l[n]$ would be the 2-dimensional position of the nth coefficient of $\mathbf{x}$. Thus, the total regularization cost of permutation matrix $\mathbf{P}_i$ is

$$R(\mathbf{P}_i) := \sum_{n,n'=1}^{N} \|l[n] - l[n']\|_2^2 \, \mathbf{P}_i[n, n']. \qquad (4)$$

This cost promotes permutations with small deviations from the identity, i.e., controlled errors. Of course, when appropriate for the application, other distance metrics, such as the ℓ_1-norm can be used instead of the squared Euclidean distance.

Solving (3) is hard in general, as it requires a combinatorial search. Thus, we consider a relaxation that allows us to use the well-established algorithmic machinery of OT [15–17]. To do so, note that (3) is equivalent to

$$\min_{\mathbf{P}_i \in \mathcal{P}, \mathbf{x}, \mathbf{x}_i} \sum_{i=1}^{K} \left(\frac{1}{2}\|\mathbf{y}_i - \mathbf{A}_i\mathbf{x}_i\|_2^2 + \beta R(\mathbf{P}_i) \right), \qquad (5)$$

$$\text{subject to } \mathbf{x}_i - \mathbf{P}_i\mathbf{F}_i\mathbf{x} = 0, \quad \forall i.$$

Next, we relax the equality constraint to

$$\min_{\mathbf{P}_i \in \mathcal{P}, \mathbf{x}, \mathbf{x}_i} \sum_{i=1}^{K} \left(\frac{1}{2}\|\mathbf{y}_i - \mathbf{A}_i\mathbf{x}_i\|_2^2 + \beta R(\mathbf{P}_i) \right), \qquad (6)$$

$$\text{subject to } \|\mathbf{x}_i - \mathbf{P}_i\mathbf{F}_i\mathbf{x}\|_2^2 \le t, \quad \forall i.$$

Writing the above in Lagrangian form, we have

$$\min_{\mathbf{P}_i \in \mathcal{P}, \mathbf{x}, \mathbf{x}_i} \sum_{i=1}^{K} \left(\frac{1}{2}\|\mathbf{y}_i - \mathbf{A}_i\mathbf{x}_i\|_2^2 + \beta R(\mathbf{P}_i) + \frac{\lambda}{2}\|\mathbf{x}_i - \mathbf{P}_i\mathbf{F}_i\mathbf{x}\|_2^2 \right).$$

$$(7)$$

Algorithm 1 Estimate single view image $\hat{\mathbf{x}}_i = \mathsf{F}_1(\mathbf{A}_i, \mathbf{y}_i, \mathbf{x})$

Compute: , $\mathbf{z}_i = \mathbf{F}_i\mathbf{x}$ and $\mathbf{v}_i = a(\mathbf{z})$.
1: **for** t = 1 to tMax **do**
2:     $\mathbf{u}_i^t = a(\mathbf{x}_i^t)$
3:     $\mathbf{P}_i^* = \underset{\mathbf{P}_i \in \Pi(\mathbf{u}_i^t, \mathbf{v}_i)}{\arg\min} \langle \mathbf{C}(\mathbf{x}_i^t, \mathbf{z}_i), \mathbf{P}_i \rangle$
4:     $\mathbf{x}_i^{t+1} = \mathbf{x}_i^t - \gamma_t \nabla_{\mathbf{x}_i} f(\mathbf{x}, \mathbf{x}_i^t)$
5: **end for**
Output: $\hat{\mathbf{x}}_i = \mathbf{x}_i^{\mathrm{tMax}}$

Algorithm 2 Estimate prototype image $\hat{\mathbf{x}} = \mathsf{F}_2(\mathbf{x}_1, \ldots, \mathbf{x}_K)$

Compute: $\mathbf{u}_i = b(\mathbf{x}_i)$, $i = 1, \ldots, K$.
1: **for** t = 1 to tMax **do**
2:     **for** i=1 to K **do**
3:         $\mathbf{v}_i^t = a(\mathbf{F}_i\mathbf{x}^t)$
4:         $\mathbf{P}_i^* = \underset{\mathbf{P} \in \Pi(\mathbf{u}_i, \mathbf{v}_i^t)}{\arg\min} \langle \mathbf{C}(\mathbf{x}_i, \mathbf{F}_i\mathbf{x}^t), \mathbf{P} \rangle$
5:     **end for**
6:     $\mathbf{x}^{t+1} = \mathbf{x}^t - \gamma_t \sum_{i=1}^K \nabla_{\mathbf{x}} f(\mathbf{x}^t, \mathbf{x}_i)$
7: **end for**
Output: $\hat{\mathbf{x}} = \mathbf{x}^{\mathrm{tMax}}$

Algorithm 3 OT Regularized Multiview Sensing

Input: $\mathbf{A}_i, \mathbf{F}_i, \mathbf{y}_i, i = 1, \ldots, K$.
Initialization for prototype image: $\mathbf{x}^0$.
1: **for** t = 1 to tMax **do**
2:     **for** i = 1 to K **do**
3:         $\mathbf{x}_i^t = \mathsf{F}_1(\mathbf{A}_i, \mathbf{y}_i, \mathbf{x}^{t-1})$
4:     **end for**
5:     $\mathbf{x}^t = \mathsf{F}_2(\mathbf{x}_1^t, \ldots, \mathbf{x}_K^t)$
6: **end for**
Output: $\hat{\mathbf{x}} = \mathbf{x}^{\mathrm{tMax}}$

---

where t and λ are inversely related. Since $\mathbf{P}_i$ is a permutation matrix, the final term in (7) can be expressed as

$$\|\mathbf{x}_i - \mathbf{P}_i\mathbf{F}_i\mathbf{x}\|_2^2 = \sum_{n,n'=1}^N \left(\mathbf{x}_i[n] - (\mathbf{F}_i\mathbf{x})[n']\right)^2 \mathbf{P}[n, n']. \quad (8)$$

The first term in (7) is independent of $\mathbf{P}_i$ and the last two terms are linear in $\mathbf{P}_i$. Therefore, we can incorporate (4) and consolidate them by defining the cost matrix $\mathbf{C}(\mathbf{x}_i, \mathbf{F}_i\mathbf{x})$ as

$$\mathbf{C}(\mathbf{x}_i, \mathbf{F}_i\mathbf{x})[n, n'] := \|l[n] - l[n']\|_2^2 + \frac{\lambda}{2\beta}\left(\mathbf{x}_i[n] - (\mathbf{F}_i\mathbf{x})[n']\right)^2,$$

and rewriting (7) as

$$\min_{\mathbf{x}, \mathbf{x}_i} \sum_{i=1}^K \left(\|\mathbf{y}_i - \mathbf{A}_i\mathbf{x}_i\|_2^2 + \beta \min_{\mathbf{P}_i \in \mathcal{P}} \langle \mathbf{C}(\mathbf{x}_i, \mathbf{F}_i\mathbf{x}), \mathbf{P}_i \rangle \right). \quad (9)$$

The minimization over $\mathbf{P}_i$ in (9) is known as the 2-D assignment problem, which is equivalent to solving a linear program [15]. Specifically, this linear programming relaxation can be interpreted as determining an optimal probabilistic coupling, i.e., a joint distribution $\mathbf{P} \in \Pi(\mathbf{u}, \mathbf{v})$, where $\Pi(\mathbf{u}, \mathbf{v}) = \{\mathbf{P} : 0 \le \mathbf{P}_{i,j} \le 1, \mathbf{P}\mathbf{1} = \mathbf{u}_i, \mathbf{P}^T\mathbf{1} = \mathbf{v}_i\}$, between two probability distributions $\mathbf{u}_i \in [0, 1]^N$ and $\mathbf{v}_i \in [0, 1]^N$ defined on the grid $\mathcal{L}$, where $\mathbf{1}$ is a vector of all ones. Then with $\mathbf{u}_i = \mathbf{v}_i = 1/N$, (9) is equivalent to

$$\min_{\mathbf{x}, \mathbf{x}_i} \sum_{i=1}^K \left(\|\mathbf{y}_i - \mathbf{A}_i\mathbf{x}_i\|_2^2 + \beta \min_{\mathbf{P}_i \in \Pi(\mathbf{u}_i, \mathbf{v}_i)} \langle \mathbf{C}(\mathbf{x}_i, \mathbf{F}_i\mathbf{x}), \mathbf{P}_i \rangle \right) \quad (10)$$

Next, we revisit (10) based on the theory of OT [15, 17] and outline an efficient algorithm to solve the resulting formulation.

5. EFFICIENT RECOVERY USING OPTIMAL TRANSPORT BASED RELAXATION

Our primary goal is to estimate $\mathbf{x}$ and not necessarily the permutations. Keeping in mind the relation between OT and assignment problems [19] [15, Ch. 2], we further make the following relaxations: (a) The marginals $\mathbf{u}, \mathbf{v}$ are assumed to be general, i.e. not necessarily uniform distributions, with possibly different supports; (b) This necessitates relaxing the the constraint on $\mathbf{P}_i$ being permutation matrices to $\mathbf{P}_i$ being *couplings* between marginals $\mathbf{u}, \mathbf{v}$. In this case the estimate obtained from the inner minimization corresponds to what is referred to as a "plan" (i.e. a joint distribution with the given marginals) in the OT literature [15, 17]. In other words we seek a softer coupling instead of a hard assignment. Note that when the supports are equal and the marginals are uniform, then the optimal plan and optimal assignment coincide. Thus, by mapping our problem to an OT one, we can apply efficient algorithms developed in that literature, such as [16, 18], to solve for the optimal plan.

Choice of marginals: Given a signal $\mathbf{x}[n]$, we choose the following marginals, which seem to provide good numerical results. Let $a : \mathbb{R}_+^N \to [0, 1]^N$ be a function that maps reflectivity values to a probability distribution, defined as

$$a(\mathbf{x})[n] := \frac{\mathbb{I}\{\mathbf{x}[n] > T\}}{\sum_{k=1}^N \mathbb{I}\{\mathbf{x}[k] > T\}}, \quad n = 1, \ldots, N, \quad (11)$$

where $\mathbb{I}$ is an indicator function and $T > 0$ is some predefined threshold. Thus, $a(\mathbf{x})$ is a uniform distribution over the grid points with sufficiently large reflectivity values. We then define the marginals as $\mathbf{u} = a(\mathbf{x})$, $\mathbf{v} = a(\mathbf{F}_i\mathbf{x})$ and an optimal transport distance between $a(\mathbf{x}_i)$ and $a(\mathbf{F}_i\mathbf{x})$ as

$$\mathrm{OT}(a(\mathbf{x}_i), a(\mathbf{F}_i\mathbf{x})) = \min_{\mathbf{P}_i \in \Pi(a(\mathbf{x}_i), a(\mathbf{F}_i\mathbf{x}))} \langle \mathbf{C}(\mathbf{x}_i, \mathbf{F}_i\mathbf{x}), \mathbf{P}_i \rangle \quad (12)$$

Given the marginals, we propose to solve for the following relaxed version of (9):

$$\min_{\mathbf{x}, \mathbf{x}_i} \sum_{i=1}^K f(\mathbf{x}, \mathbf{x}_i), \text{ where}$$
$$f(\mathbf{x}, \mathbf{x}_i) = \|\mathbf{y}_i - \mathbf{A}_i\mathbf{x}_i\|_2^2 + \beta\, \mathrm{OT}(a(\mathbf{x}_i), a(\mathbf{F}_i\mathbf{x})). \quad (13)$$

It is worth mentioning that $\mathrm{OT}(a(\mathbf{x}_i), a(\mathbf{F}_i\mathbf{x}))$ may be written as a transportation-L^p distance [23, 24] as $d_{TL^2}^2((\mathbf{x}_i, \mathbf{u}), (\mathbf{F}_i\mathbf{x}, \mathbf{v}))$. The particular case considered here is when the marginals are a function of the signal. While this case may fall in the general set up discussed in [23, 24], it is unclear whether the theoretical results therein still hold when the marginals depend on the signal; we leave the study of theoretical properties of (13) for future work.

In order to use gradient descent methods to solve (13), we need to compute the gradient of $\mathrm{OT}(a(\mathbf{x}_i), a(\mathbf{F}_i\mathbf{x}))$ defined in (12) with respect to $\mathbf{x}_i$ and $\mathbf{x}$. Let $\mathbf{f}$ and $\mathbf{g}$ be Lagrangian multipliers, the Lagrangian form of (12) is

$$L(\mathbf{P}_i, \mathbf{f}, \mathbf{g}, \mathbf{x}_i, \mathbf{x}) = \langle \mathbf{C}(\mathbf{x}_i, \mathbf{F}_i\mathbf{x}), \mathbf{P}_i \rangle + \langle \mathbf{f}, \mathbf{P}_i\mathbf{1} - a(\mathbf{x}_i) \rangle$$
$$+ \langle \mathbf{g}, \mathbf{P}_i^T\mathbf{1} - a(\mathbf{F}_i\mathbf{x}) \rangle.$$
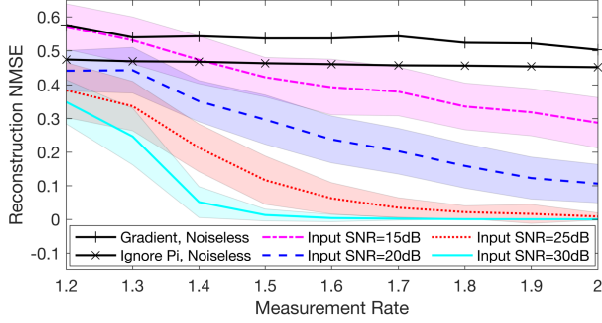
Fig. 2. NMSE as a function of measurement rate at various input SNR, where the number of views is 2. The shaded area represents one standard deviation below and above the mean.
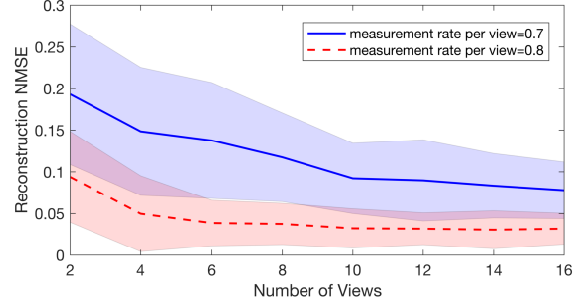

Fig. 3. NMSE as a function of number of views. The measurement rate per view is 0.7 or 0.8 and the input SNR is 20dB.The shaded area represents one standard deviation below and above the mean.

Let $\mathbf{P}_i^*$, $(\mathbf{f}^*, \mathbf{g}^*)$ be the primal and dual optima, respectively. Although they all depend on $\mathbf{x}_i, \mathbf{x}$, the envelope theorem [25] allows us to conveniently compute the gradient:

$$\nabla_{\mathbf{x}_i} \mathrm{OT}(a(\mathbf{x}_i), a(\mathbf{F}_i \mathbf{x})) = \nabla_{\mathbf{x}_i} L(\mathbf{P}_i^*, \mathbf{f}^*, \mathbf{g}^*, \mathbf{x}_i, \mathbf{x})$$
$$\overset{(a)}{=} \nabla_{\mathbf{x}_i} \langle \mathbf{C}(\mathbf{x}_i, \mathbf{F}_i \mathbf{x}), \mathbf{P}_i^* \rangle - \nabla_{\mathbf{x}_i} \langle \mathbf{f}^*, a(\mathbf{x}_i) \rangle$$
$$\overset{(b)}{=} \nabla_{\mathbf{x}_i} \langle \mathbf{C}(\mathbf{x}_i, \mathbf{F}_i \mathbf{x}), \mathbf{P}_i^* \rangle, \text{ a.e.,}$$

where in step (a), $\mathbf{P}_i^*$ and $\mathbf{f}^*$ are considered as constant with respect to $\mathbf{x}_i$ (by the envelope theorem), and step (b) follows by definition of a in (11), whose gradient is zero almost everywhere with respect to Lebesgue measure on $\mathbb{R}^N$. The gradient of $\mathrm{OT}(a(\mathbf{x}_i), a(\mathbf{F}_i \mathbf{x}))$ with respect to $\mathbf{x}$ can be computed in a similar way. Then the gradient of the cost function $f(\mathbf{x}, \mathbf{x}_i)$ can be computed as

$$\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{x}_i) = \lambda \mathbf{F}_i^T \Big(a(\mathbf{F}_i \mathbf{x}) \odot (\mathbf{F}_i \mathbf{x}) - (\mathbf{P}_i^*)^T \mathbf{x}_i \Big),$$
$$\nabla_{\mathbf{x}_i} f(\mathbf{x}, \mathbf{x}_i) = \mathbf{A}_i^T (\mathbf{A}_i \mathbf{x}_i - \mathbf{y}) + \lambda \Big(a(\mathbf{x}_i) \odot \mathbf{x}_i - \mathbf{P}_i^* \mathbf{F}_i \mathbf{x} \Big),$$

where $\odot$ denotes point-wise product.

Our proposed algorithm for estimating $\mathbf{x}$ from (1) solves (13) by alternating between the estimation of $\mathbf{x}_i$ and $\mathbf{x}$. With a fixed $\mathbf{x}$, we estimate $\mathbf{x}_i$ by Algorithm 1, and with a fixed $\mathbf{x}_i$, we estimate $\mathbf{x}$ by Algorithm 2; the full algorithm is summarized in Algorithm 3. The estimation of $\mathbf{P}_i^*$ in Line 3 of Algorithm 1 and Line 4 of Algorithm 2 are solved by IPOT [18].

6. EXPERIMENTS ON SYNTHETIC DATA

To validate our proposed algorithm, we perform two different sets of experiments on simulated data. For the first set, the number of views is fixed to be 2 and we test the algorithm at different measurement rates, defined as the measurement rate per view times the number of views, and input signal to noise ratios (SNR), defined as $\|\mathbf{A}_i \mathbf{x}_i\|_2^2 / \|\mathbf{y}_i - \mathbf{A}_i \mathbf{x}_i\|_2^2$.

In this set of experiments, to demonstrate the effectiveness of the proposed method, we include results for two baseline methods. The first one, labeled Gradient, is a straightforward approach to solving (3) directly, by alternating between estimating $\mathbf{x}$ and $\mathbf{P}_i$ and solving each subproblem using gradient descent. The constraint $\mathbf{P}_i \in \mathcal{P}$ is relaxed to $\mathbf{P}_i \in [0, 1]^{N \times N}$ with an additional regularizer to promote $\mathbf{P}_i$ having the same row and column sums as a permutation: $\|\mathbf{P}_i \mathbf{1} - \mathbf{1}\|_2^2 + \|\mathbf{P}_i^T \mathbf{1} - \mathbf{1}\|_2^2$. The second baseline method, labeled Ignore $\mathbf{P}_i$ in the figure, solves $\mathbf{x}$ from (3) assuming that $\mathbf{P}_i$ is identity.

For the second set of experiments, the input SNR is fixed to be 20dB and the measurement rate per view is fixed to be 0.7 or 0.8. We test the algorithm at different numbers of views. A key issue we explore is that while increasing the number of views increases measurements, it also increases the number of unknown permutations. Given their poor performance of the baselines in the first set, we did not include them in this set.

In all experiments, the reference pose $\mathbf{x}$, the estimated transformation $\mathbf{F}_i \mathbf{x}$, and the actual measured image $\mathbf{x}_i$ are generated in a similar way as the example in Figure 1, except that in the second experiment, the letter "T" instead of "E" is used to simplify data generation. The number of pixels in each image is $N = 512$. The support of the ground truth $\mathbf{x}$ is known to the algorithm and thus so is the support size K. The threshold T in (11) is set to be the Kth largest pixel value in the corresponding vector. The measurement matrix $\mathbf{A}_i$ has i.i.d. Gaussian entries with mean zero and variance $1/N$. Performance is measured by the normalized mean squared error (NMSE), defined as $\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 / \|\mathbf{x}\|_2^2$.

Figure 2 presents the results for the first experiment. The figure shows that the two baseline methods perform poorly even without measurement noise. For our proposed method, the reconstruction performance improves as the measurement rate increases. Moreover, when the input SNR is 25dB or higher, the reconstruction performance improves rather quickly.

Figure 3 presents the results for the second experiment. It shows that an increased number of views improves reconstruction of the reference pose, despite introducing more unknown permutations. Moreover, for a given input SNR, measurement rate, and sparsity of the reference image, performance seems to stop improving after a certain number of views; we defer comprehensive investigation of this effect to future work.

7. DISCUSSION AND CONCLUSIONS

Estimation of signals observed under unknown permutations is a difficult problem in general. Recognizing that in many applications some permutations are more likely than others, we introduce a regularization term which promotes certain permutations over others. By further relaxing the problem, we are able to exploit well-developed techniques in OT to provide tractable algorithms for this problem. A key component in this formulation is the judicious choice of the OT ground cost to incorporate the regularization penalty. While we present a particular choice for this cost, there are several other options, depending on the application, the study of which we defer to later publications.

8. REFERENCES

[1] O. Tuzel, F. Porikli, and P. Meer, "Learning on lie groups for invariant detection and tracking," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.

[2] M. Usman, D. Atkinson, F. Odille, C. Kolbitsch, G. Vaillant, T. Schaeffter, P. G. Batchelor, and C. Prieto, "Motion corrected compressed sensing for free-breathing dynamic cardiac MRI," *Magn Res Med*, vol. 70, pp. 504–516, 2013.

[3] M. Asif, L. Hamilton, M. B. M, and J. Romberg, "Motion-adaptive spatio-temporal regularization for accelerated dynamic MRI," *Magn Res Med*, vol. 70, p. 800812, 2013.

[4] P. Fischer, T. Pohl, A. Maier, and J. Hornegger, "Surrogate-driven estimation of respiratory motion and layers in x-ray fluoroscopy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 282–289.

[5] J. Wulff and M. J. Black, "Efficient sparse-to-dense optical flow estimation using a learned basis and layers," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 120–130.

[6] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, "Unlabeled sensing: Solving a linear system with unordered measurements," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2015, pp. 786–793.

[7] S. Haghighatshoar and G. Caire, "Signal Recovery From Unlabeled Samples," *IEEE Transactions on Signal Processing*, vol. 66, pp. 1242–1257, 2017.

[8] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, "Unlabeled Sensing With Random Linear Measurements," *IEEE Transactions on Information Theory*, vol. 64, pp. 3237–3253, 2018.

[9] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Denoising linear models with permuted data," *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 446–450, 2017.

[10] ——, "Linear Regression With Shuffled Data: Statistical and Computational Limits of Permutation Recovery," *IEEE Transactions on Information Theory*, vol. 64, pp. 3286–3300, 2018.

[11] D. Hsu, K. Shi, and X. Sun, "Linear regression without correspondence," in *Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 1530–1539.

[12] I. Dokmanić, "Permutations unlabeled beyond sampling unknown," *IEEE Signal Processing Letters*, vol. 26, pp. 823–827, 2019.

[13] M. C. Tsakiris and L. Peng, "Homomorphic sensing," in *International Conference on Machine Learning (ICML)*, 2019.

[14] A. Aldroubi and K. Grchenig, "Nonuniform sampling and reconstruction in shift-invariant spaces," *SIAM Review*, vol. 43, no. 4, pp. 585–620, 2001.

[15] G. Peyré and M. Cuturi, "Computational optimal transport," *Foundations and Trends in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[16] M. Cuturi, "Sinkhorn distances: lightspeed computation of optimal transport." in *Advances in Neural Information Processing Systems(NIPS)*, vol. 2, no. 3, 2013, pp. 2292–2300.

[17] F. Santambrogio, "Optimal transport for applied mathematicians," *Birkäuser, NY*, vol. 55, no. 58-63, p. 94, 2015.

[18] Y. Xie, X. Wang, R. Wang, and H. Zha, "A fast proximal point method for computing exact wasserstein distance," in *Uncertainty in Artificial Intelligence*. PMLR, 2020, pp. 433–453.

[19] R. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems*. Society for Industrial and Applied Mathematics, 2012.

[20] A. B. Poore and S. Gadaleta, "Some assignment problems arising from multiple target tracking," *Mathematical and computer modelling*, vol. 43, no. 9-10, pp. 1074–1091, 2006.

[21] P. Emami, P. M. Pardalos, L. Elefteriadou, and S. Ranka, "Machine learning methods for solving assignment problems in multi-target tracking," *arXiv preprint arXiv:1802.06897*, 2018.

[22] M. Cuturi and A. Doucet, "Fast computation of wasserstein barycenters," in *International conference on machine learning*. PMLR, 2014, pp. 685–693.

[23] N. G. Trillos and D. Slepčev, "Continuum limit of total variation on point clouds," *Archive for rational mechanics and analysis*, vol. 220, no. 1, pp. 193–241, 2016.

[24] M. Thorpe, S. Park, S. Kolouri, G. K. Rohde, and D. Slepčev, "A transportation l^p distance for signal analysis," *Journal of mathematical imaging and vision*, vol. 59, no. 2, pp. 187–210, 2017.

[25] S. Afriat, "Theory of maxima and the method of lagrange," *SIAM Journal on Applied Mathematics*, vol. 20, no. 3, pp. 343–357, 1971.