

Generalized Shuffled Linear Regression

Feiran Li¹ Kent Fujiwara² Fumio Okura¹ Yasuyuki Matsushita¹

¹Osaka University

² LINE Corporation

Abstract

We consider the shuffled linear regression problem where the correspondences between covariates and responses are unknown. While the existing formulation assumes an ideal underlying bijection in which all pieces of data should match, such an assumption barely holds in real-world applications due to either missing data or outliers. Therefore, in this work, we generalize the formulation of shuffled linear regression to a broader range of conditions where only part of the data should correspond. Moreover, we present a remarkably simple yet effective optimization algorithm with guaranteed global convergence. Distinct tasks validate the effectiveness of the proposed method.¹

1. Introduction

Linear regression has always been a powerful tool for parameter estimation problems in science and engineering. In the conventional form, it is designed to estimate the parameters of a linear system given pairs of matched data. Its objective function is written as²

$$\min_{\mathbf{x} \in \mathcal{X}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{n \times d}$ ($n \gg d$) and $\mathbf{b} \in \mathbb{R}^n$ denote the collections of covariates and responses, respectively, and \mathbf{x} is the regression variable belonging to set \mathcal{X} .

There also exist cases where the correspondences between \mathbf{A} and \mathbf{b} are unknown. For example, in computer vision tasks such as point cloud registration, shape matching, perspective- n -points, and image registration, the regression variable \mathbf{x} stands for the rigid transformation, spectral mapping, camera projection, and transformation matrices, respectively. In such cases, correspondences among points, vertices, or pixels are often unknown. Similar situations also appear in other fields, such as data de-anonymization [25] in data science, artifacts dating [12] in archaeology, and sampling with jitter [28] in signal processing.

All these issues lead to a recently raised variant of linear regression, namely, the *shuffled linear regression* (SLR) problem. It aims at simultaneously recovering both correspondences between data and the regression variable \mathbf{x} [28, 45]. Formally, its objective function is written as

$$\min_{\mathbf{P} \in \Pi^*, \mathbf{x} \in \mathcal{X}} \|\mathbf{P}\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2, \quad (2)$$

where \mathbf{P} is an n -dimensional square permutation matrix from the discrete and finite set Π^* satisfying

$$\mathbf{P}\mathbf{1} = \mathbf{P}^T\mathbf{1} = \mathbf{1}, \quad \mathbf{P}_{ij} \in \{0, 1\}, \quad (3)$$

where $\mathbf{1}$ is the all-one vector, and \mathbf{P}_{ij} denotes the ij^{th} entry of \mathbf{P} . Intuitively, constraints (3) bring some plausible properties on the relations between each covariate \mathbf{A}_i and response \mathbf{b}_j :

- $\mathbf{P}_{ij} \in \{0, 1\}$ ensures \mathbf{A}_i and \mathbf{b}_j are either absolutely matched or completely independent from each other.
- Together with the binary property, $\mathbf{P}\mathbf{1} = \mathbf{P}^T\mathbf{1} = \mathbf{1}$ further guarantees bijectivity, *i.e.*, each \mathbf{A}_i and \mathbf{b}_j possess exactly one match.

This vanilla formulation of SLR demands exactly the same cardinalities of $\{\mathbf{A}_i\}$ and $\{\mathbf{b}_j\}$, and all of them to be inliers. However, such requirements are seldom satisfied in practice due to missing data and outliers. Therefore, it is desired to extend SLR to work well for real-world applications.

In this work, we generalize SLR to a broader range of cases where the cardinalities of covariates $\{\mathbf{A}_i\}$ and responses $\{\mathbf{b}_j\}$ can be different, and only part of them should match. We denote this generalized setup of SLR as *Generalized Shuffled Linear Regression* (GSLR). Our contributions are summarized as follows:

- We present GSLR, a generalized formulation of SLR, making it applicable to practical situations, where only parts of the data should correspond.
- We propose a remarkably simple yet effective algorithm with a detailed theoretical analysis for optimization.
- We employ distinct examples to demonstrate how GSLR can benefit computer vision tasks in achieving state-of-the-art accuracies.

¹Source code can be found at <https://github.com/SILI1994/Generalized-Shuffled-Linear-Regression>.

²Without loss of generality, we assume the Euclidean norm as a metric.

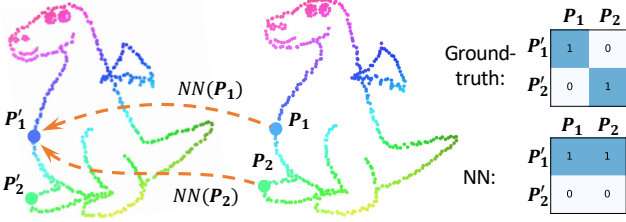


Figure 1: NN-based corresponding strategy may violate the one-to-one correspondences by either assigning a single point to multiple queries or leaving points unmatched.

2. Related work

In this section, we review the prior-arts related to (G)SLR and discuss about relevant techniques and problems.

Shuffled linear regression Some pioneering works of SLR focus on solvability analysis. For example, Unnikrishnan *et al.* [37] prove that for a given set of covariates $\mathbf{A} \in \mathbb{R}^{n \times d}$, $n \geq 2d$ is necessary for obtaining a unique solution of the regression variable \mathbf{x} in noiseless case. Pananjady *et al.* [28] further provide conditions for unique recovery under noisy data. There also exists a work [19] that provides theoretical guidance on the design of global algorithms with polynomial time complexity. However, they are still impracticable due to the $\mathcal{O}(n^d)$ time complexity.

Other literature focuses on practically solving the SLR problem. For example, Abid *et al.* [1, 2] establish a Monte-Carlo expectation-maximization framework by treating correspondences as the hidden parameter. Zhang *et al.* [45] reformulate objective (2) by eliminating \mathbf{x} with maximum likelihood criterion, which in fact leads to a quadratic assignment problem (QAP).

There also exist works that concentrate on simplified versions of objective (2). For example, Ashwin *et al.* [27] estimate $\mathbf{P}^T \mathbf{A} \mathbf{x}$ as a whole in the manner of the denoising problem. Slawski and Ben-David [34] assume only a small amount of data is disordered and propose to model the sparsely-shuffled property via $L1$ -norm. Based on the same assumption, they later suggest rejecting the mismatched ones as outliers via robust estimation [35].

In contrast to these approaches that assume an ideal underlying bijection, our GSLR formulation aims at general cases where only part of the data should correspond.

Similar problems in applications Although it is only recently that SLR was formally formulated [28], the problem itself has already been approximately handled in various applications. For example, the rigid point cloud registration problem, which aims at recovering the spatial transformation \mathbf{T} with unknown point-wise correspondences, can perfectly fit in the (G)SLR formulation. However, existing approaches,

such as iterative closet point (ICP) [3] and its variants, are infeasible algorithms for the SLR problem. In detail, as shown in Fig. 1, their nearest neighbors (NN)-based matching strategy may either link a point to multiple queries or leave it unmatched, which consequently violates the one-to-one matching constraint of permutation matrices. As we will show later in Sec. 5, it is this violation that prevents them from achieving stably accurate results. For a solution, some literature proposes to strengthen the bijectivity. For example, both Jing *et al.* [30] and Szymon [32] propose to symmetrize the registration process; and Gold *et al.* [14] convert the assignment matrix to a doubly stochastic one to balance the weights of correspondences. However, contrary to our GSLR formulation, these attempts still cannot guarantee strict one-to-one correspondences. There also exist works [38, 39] that cast the matching problem into a linear assignment problem (LAP) by maximizing kernel densities in product spaces, whereas their generalizability is limited for assuming the correspondences between the covariates $\{\mathbf{A}_i\}$ and responses $\{\mathbf{b}_j\}$ to be partial-to-all instead of partial-to-partial.

Robust estimation Robust estimation aims at alleviating the effects of outliers on regression problems. A popular tool to realize so are the M-estimators [36], which propose to either down-weight or completely reject suspected outliers. The ability of a robust estimator is typically assessed by two criteria, namely, breakdown point and efficiency [48]. Specifically, the breakdown point, theoretically upper-bounded by 0.5, demonstrates the proportion of outliers an estimator can tolerate before providing an incorrect result. The efficiency, calculated as the ratio of the theoretically minimal variance provided by the Cramer-Rao bound w.r.t. the actual one of the estimator, stands for the quality of an estimator.

Relation to QAP QAP (e.g., second-order graph matching [46]) is another popular problem subject to permutational constraints. However, different from QAP, (G)SLR additionally requires to estimate the regression variable \mathbf{x} . Thus, converting (G)SLR to QAP is hardly possible unless the optimal \mathbf{x} can be explicitly expressed by \mathbf{A} , \mathbf{b} , and \mathbf{P} [45] (e.g., when $\mathbf{x} \in \mathbb{R}^d$, we can eliminate \mathbf{x} from (G)SLR by $\mathbf{x} = (\mathbf{A}^T \mathbf{P}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}^T \mathbf{b}$). However, in general cases where $\mathbf{x} \in \mathcal{X} \neq \mathbb{R}^d$, such an elimination can only present coarse results as mentioned in [22].

3. Formulation and optimization of GSLR

Following the formulation of SLR shown in objective (2), the GSLR problem can be similarly written as

$$\min_{\mathbf{P} \in \Pi, \mathbf{x} \in \mathcal{X}} \|\mathbf{P} \mathbf{A} \mathbf{x} - \mathbf{b}\|_2^2, \quad (4)$$

where $\mathbf{A} \in \mathbb{R}^{m \times d}$, $\mathbf{b} \in \mathbb{R}^n$ ($m \neq n$ in general), and Π is the set of generalized permutation matrices.

3.1. Properties of Π

Given objective (4), we first explore the properties that a generalized permutation matrix $\mathbf{P} \in \Pi$ should satisfy. Specifically, apart from the binary and one-to-one matching properties inherited from the ordinary counterpart Π^* , we also desire it to be capable of characterizing outliers, as

- $\mathbf{P}_{ij} = 1$ if \mathbf{A}_i and \mathbf{b}_j should match otherwise 0.
- $\mathbf{P}_{ij} = 0$ if either \mathbf{A}_i or \mathbf{b}_j is an outlier.
- $\sum_i \mathbf{P}_{ij}, \sum_j \mathbf{P}_{ij} \in \{0, 1\}$ to guarantee that each \mathbf{A}_i and \mathbf{b}_j can have one correspondence at most.

These requirements can be summarized as

$$\mathbf{P}\mathbf{1} \leq \mathbf{1}, \mathbf{P}^T\mathbf{1} \leq \mathbf{1}, \mathbf{P}_{ij} \in \{0, 1\}. \quad (5)$$

However, conditions (5) cannot be directly used for optimization. If done so, the global minimum of objective (4) would be anchored to 0 with the associated trivial solution $\mathbf{P} = \mathbf{0}$. Therefore, it is necessary to introduce some other constraints to ensure the existence of matches.

A simple yet effective solution is to manually designate the number of matches [40, 41]. Specifically, we introduce an integer $k \in \left[\frac{\min(m, n)}{2}, \min(m, n)\right]$ and force $\sum \mathbf{P}_{ij} = k$, indicating that there must exactly exist k inlier matches between the sets of covariates and responses. Consequently, the trivial solution can be avoided, and the constraints over the set of generalized permutation matrices Π is

$$\mathbf{P}\mathbf{1} \leq \mathbf{1}, \mathbf{P}^T\mathbf{1} \leq \mathbf{1}, \mathbf{P}_{ij} \in \{0, 1\}, \sum \mathbf{P}_{ij} = k. \quad (6)$$

For notational simplicity, we hereafter keep using Π to denote the set of all the generalized permutation matrices under all possible k . Similar to the ordinary counterpart Π^* , constraints (6) imply that Π is discrete and combinatorial, and degenerates to Π^* when $k = m = n$. Now, we can obtain the formal formulation of GSLR by joining (4) and (6):

$$\begin{aligned} & \min_{\mathbf{P}, \mathbf{x} \in \mathcal{X}} \|\mathbf{P}\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \\ & \text{s.t. } \mathbf{P}\mathbf{1} \leq \mathbf{1}, \mathbf{P}^T\mathbf{1} \leq \mathbf{1}, \mathbf{P}_{ij} \in \{0, 1\}, \sum \mathbf{P}_{ij} = k. \end{aligned} \quad (7)$$

3.2. A bi-global optimization algorithm

In this section, we consider the optimization method for objective (7). Its difficulties are threefold. First, the discrete property of the permutation \mathbf{P} hinders optimizers designed for smooth functions. Second, its high non-convexity requires careful initialization. Last, the number of matches k is unknown. As a solution, we propose a simple and effective algorithm, which can effectively maintain the discreteness and alleviate the initialization problem. For clarification, we hereby temporarily assume that k is known and put its estimation in Sec. 3.3.

The key observation of our proposal is that slight reformulation of objective (7) can result in a bi-globally optimizable form. Specifically, per the definitions of linear regression, we can facilely rewrite it into the summation form of

$$\min_{\mathbf{P} \in \Pi, \mathbf{x} \in \mathcal{X}} \sum \mathbf{P}_{ji} \|\mathbf{A}_i\mathbf{x} - \mathbf{b}_j\|_2^2. \quad (8)$$

If we apply alternative optimization to objective (8), by treating the correspondences \mathbf{P}_{ji} as known, updating the regression variable \mathbf{x} leads to a weighted linear least square problem, which should be globally solvable in most applications. On the other hand, if \mathbf{x} is fixed, the objective function for updating \mathbf{P} is in the form of

$$\min_{\mathbf{P} \in \Pi} \mathbf{D}_{ij} \mathbf{P}_{ji}, \quad (9)$$

where \mathbf{D} is the cost matrix formulated by $\|\mathbf{A}_i\mathbf{x} - \mathbf{b}_j\|_2^2$.

When the number of matches $k = \min(m, n)$, problem (9) is a standard LAP, which can be globally optimized via efficient approaches such as the Hungarian algorithm [20]. For an LAP with arbitrary k , it forms a k -cardinality linear assignment problem (k -LAP), which can be transformed into a standard one via paddling the cost matrix [40]. Therefore, we can conclude that the GSLR problem is facilely solvable via alternative optimization.

3.3. k as an M-estimator

It is obvious that the number of matches k characterizes the ability of GSLR in rejecting outliers. In detail, it acts similarly to the trimming operator in robust estimation, which sets the weights of a specific proportion of suspected outliers to zero. In fact, such a technique has already been broadly used in various similar tasks. For example, the Trim-ICP algorithm [8] for point cloud registration suggests removing a portion of paired points with large residuals; and the partial graph matching algorithm [41] assumes only half of the nodes as inliers, which leads to $k = \frac{1}{2} \min(m, n)$.

However, we hesitate to directly use the trimming operator for its well-studied limitation on simultaneously maintaining a high breakdown point and high efficiency. Specifically, although a breakdown point of 0.5 can be facilely achieved by ignoring 50% of the data, the associated statistical efficiency would only be 7% [11]. Moreover, it is also preferable for some applications [26, 31] to preserve as many matches as possible while maintaining high robustness.

In this work, we employ another M-estimator to determine k , namely, the Huber-skip estimator [16]. Apart from inheriting the complete-rejection (*i.e.*, set weights as binary) and easy-to-calculate properties of the trimming operator, it also achieves much higher efficiency while being capable of maintaining a breakdown point of 0.5. Specifically, given some data (r_1, r_2, \dots, r_n) , it rejects r_i that violates

$$\left| \frac{r_i - \text{median}(r)}{\text{MAD}(r)} \right| \leq \text{threshold}, \quad (10)$$

Algorithm 1: A bi-global algorithm for GSLR

Input: $\mathbf{A} \in \mathbb{R}^{m \times d}$, $\mathbf{b} \in \mathbb{R}^n$ **Initialization:** $\mathbf{x} = \mathbf{x}_0 \in \mathcal{X}$ **while** not converged **do**

- Calculate $\check{\mathbf{P}}_t$ by solving the standard LAP with $k_t = \min(m, n)$ and \mathbf{x}_{t-1} .
- Compute residuals under $\check{\mathbf{P}}_t$ and \mathbf{x}_{t-1} and apply the Huber-skip estimator to get k_t .
- Set $k_t = \min(k_t, k_{t-1})$
- Calculate $\hat{\mathbf{P}}_t$ by solving the k -LAP with the updated k_t and \mathbf{x}_{t-1} .
- Get $\mathbf{x}_t \in \mathcal{X}$ by solving the weighted linear regression problem under $\hat{\mathbf{P}}_t$.

Output: $\mathbf{x} \in \mathcal{X}$, $\mathbf{P} \in \Pi$

where MAD is the median absolute deviation, and the threshold is set to 3.5 throughout the paper as suggested by [16]. More details are provided in the supplementary material.

To embed the Huber-skip estimator into our aforementioned optimization scheme, we directly follow the formulation of least trimmed squares (LTS) [18] and compute k as the number of pairs whose residuals satisfy Eq. (10). However, such a naive formulation would break the monotonicity of the alternative optimization scheme and make its convergence non-guaranteed. As a solution, in the t^{th} iteration, we set $\hat{k}_t = \min(k_{t-1}, k_t)$ to ensure a convergence. The entire algorithm is summarized in Algorithm 1.

4. Algorithm analysis

In this section, we analyze the time and space complexities of Algorithm 1. Moreover, we also prove that it monotonically decreases the objective function and converges globally to a critical point.

4.1. Complexity

We analyze the space and time complexities of Algorithm 1 under the assumption that $\#\{\mathbf{A}_1, \dots, \mathbf{A}_m\} \geq \#\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ (i.e., $m \geq n$) without loss of generality. The main space complexity comes from the cost matrix \mathbf{D} , which is $\mathcal{O}(mn)$ theoretically and $\mathcal{O}(mnd)$ in our implementation to avoid loops. As for time complexity in each iteration, the Hungarian algorithm for solving the k -LAP is $\mathcal{O}(n(m+n-k)^2)$, and the Huber-skipping procedures is with $\mathcal{O}(n \log n)$ time complexity.

4.2. Monotonicity

We now show that the total energy $E(\mathbf{P}, \mathbf{x})$ of objective (8) monotonically decreases with Algorithm 1. Specifically,

since the linear least square problem on the regression variable \mathbf{x} is globally optimizable, it is easy to see that

$$E(\mathbf{P}_t, \mathbf{x}_{t+1}) \leq E(\mathbf{P}_t, \mathbf{x}_t), \quad (11)$$

where the equality holds if and only if \mathbf{x}_t is already a global optimum. Furthermore, we define a temporary permutation matrix \mathbf{P}_{tmp} by randomly removing $k_{t+1} - k_t$ non-zero assignments from \mathbf{P}_t . Since the cost matrix $\mathbf{D}_{ij} = (\mathbf{A}_i \mathbf{x} - \mathbf{b}_j)^2$ is non-negative, we can obtain that

$$E(\mathbf{P}_{\text{tmp}}, \mathbf{x}_{t+1}) \leq E(\mathbf{P}_t, \mathbf{x}_{t+1}). \quad (12)$$

Since \mathbf{P}_{tmp} satisfies the constraints of the k -LAP formulated by \mathbf{x}_{t+1} and k_{t+1} , it is also within the solution space. On the other hand, as \mathbf{P}_{t+1} is a global optimum of this objective function, we can further derive that

$$E(\mathbf{P}_{t+1}, \mathbf{x}_{t+1}) \leq E(\mathbf{P}_{\text{tmp}}, \mathbf{x}_{t+1}), \quad (13)$$

where the equality holds if and only if \mathbf{P}_{tmp} is already within the set of global minima. Moreover, we can obtain the following relationship by summarizing Eqs. (11)-(13):

$$E(\mathbf{P}_{t+1}, \mathbf{x}_{t+1}) \leq E(\mathbf{P}_{\text{tmp}}, \mathbf{x}_{t+1}) \leq E(\mathbf{P}_t, \mathbf{x}_t), \quad (14)$$

where the equality holds if and only if $(\mathbf{P}_t, \mathbf{x}_t)$ is a critical point of the objective function. From Eq. (14), we can conclude that Algorithm 1 has a monotonic property.

4.3. Convergence

Our proposal can globally converge to an critical point owing to the monotonicity. To be precise, we first make two assumptions about the domain of the regression variable \mathbf{x} :

- The set \mathcal{X} for $\mathbf{x} \in \mathcal{X}$ is compact.
- $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is continuous on \mathcal{X} .

These assumptions are rather weak and can be satisfied in most practical applications formulated in the GSLR manner. For example, the orthogonal groups raised in geometric computer vision and geometric processing tasks can satisfy both conditions. For the general real coordinate space \mathbb{R}^n , despite not being compact by itself, we can reasonably assume that \mathbf{x} lies in a closed and bounded subset of it, since infinities are barely meaningful solutions in practical applications.

Similar to the proof of the global convergence of the expectation-maximization algorithm [42] and the generalized alternative optimization [15], the global convergence property of GSLR is a direct instantiation of the convergence theorem of Zangwill [44]:

Theorem 1 (Zangwill's global convergence theorem). *Let \mathcal{H} be a set-valued map on \mathcal{S} that, given $\mathbf{s}_0 \in \mathcal{S}$, generates a sequence $\{\mathbf{s}_k\}_k^\infty$ via iterating $\mathbf{s}_{k+1} \in \mathcal{H}(\mathbf{s}_k)$. Also, let a solution set $\Gamma \subset \mathcal{S}$ be given. Suppose that*

- The sequence $\{s_k\}_k^\infty \subset \mathcal{S}'$ for $\mathcal{S}' \subset \mathcal{S}$ is compact.
- There exists a continuous function Ψ on \mathcal{S} that
 - if $s \notin \Gamma$: $\Psi(s') < \Psi(s)$ for all $s' \in \mathcal{H}(s)$.
 - if $s \in \Gamma$: $\Psi(s') \leq \Psi(s)$ for all $s' \in \mathcal{H}(s)$.
- The set-valued map \mathcal{H} is closed at all s for $s \in \mathcal{S} \setminus \Gamma$.

Then, the limit of any convergent subsequence of $\{s_k\}_k^\infty$ is in the solution set Γ . Furthermore, $\lim_{k \rightarrow \infty} \Psi(s_k) = \Psi(s^*)$ for all limit points s^* .

An iterative optimization algorithm can be described as a triplet $\{\Psi, \mathcal{H}, \Gamma\}$ in the terminology of Theorem 1. For Algorithm 1, we can straightforwardly set the objective function of GSLR as Ψ , and $\mathcal{H} : \Pi \times \mathcal{X} \rightarrow \Pi \times \mathcal{X}$ to one iteration of it. Also, we can formulate the solution set with all the critical points that satisfy

$$\Gamma = \{(\mathbf{P}, \mathbf{x}) \in \Pi \times \mathcal{X} : \mathbf{P} \in \underset{\mathbf{P}'}{\operatorname{argmin}} \Psi(\mathbf{P}', \mathbf{x}) \text{ and } \mathbf{x} \in \underset{\mathbf{x}'}{\operatorname{argmin}} \Psi(\mathbf{P}, \mathbf{x}')\}. \quad (15)$$

We show in the following that all the three requests in Theorem 1 are satisfied by this GSLR triplet.

For the compactness of the feasible region $\Pi \times \mathcal{X}$, we only need to verify the compactness of Π and \mathcal{X} respectively, since unions of compact sets are still compact. Specifically, \mathcal{X} is compact under our assumption; and for Π , its finite and discrete properties imply its compactness.

For the continuity and monotonicity of the objective function $\Psi = \|\mathbf{P}\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$, since it is already demonstrated to be monotonic in Sec. 4.2, we hereby only focus on the continuity. For proof, since compositions of continuous functions are still continuous, we hereby focus on the continuity of $f(\mathbf{P}, \mathbf{x}) = \mathbf{P}\mathbf{A}\mathbf{x}$. Given the fact that such a function is a bilinear form w.r.t. (\mathbf{P}, \mathbf{x}) , we only need to show that it is separately continuous w.r.t. both \mathbf{P} and \mathbf{x} . For proof, per the assumption mentioned above, we have $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ continuous w.r.t. $\mathbf{x} \in \mathcal{X}$. Moreover, since the set of all possible generalized permutations Π is finitely discrete, we have $f(\mathbf{P}) = \mathbf{P}\mathbf{A}$ continuous w.r.t. $\mathbf{P} \in \Pi$. Therefore, we obtain that $f(\mathbf{P}, \mathbf{x}) = \mathbf{P}\mathbf{A}\mathbf{x}$ is continuous and hence so is Ψ .

Definition 1 (Closedness of set-valued map). A set-valued map $\mathcal{H} : \mathcal{U} \rightarrow \mathcal{V}$ is said to be closed on a point $u \in \mathcal{U}$ provided that the following two characters

- $u_k \rightarrow u$, where $u_k \in \mathcal{U}$,
- $v_k \rightarrow v$, where $v, v_k \in \mathcal{V}$,
- $v_k \in \mathcal{H}(u_k)$,

imply $v \in \mathcal{H}(u)$. Based on this, \mathcal{H} is called closed if it is closed at any $u \in \mathcal{U}$.

We now prove that the map $(\mathbf{P}_{t+1}, \mathbf{x}_{t+1}) \in \mathcal{H}(\mathbf{P}_t, \mathbf{x}_t)$ is closed on $\Pi \times \mathcal{X}$. Specifically, for any $(\mathbf{P}_t, \mathbf{x}_t)$, we can always obtain some $\delta_1 > 0$ and $\delta_2 > 0$ that satisfy

$$\begin{aligned} \forall \mathbf{P}', \mathbf{P}' \rightarrow \mathbf{P} &:= \{\mathbf{P}' : \mathbf{P}' \in \mathcal{N}_{\delta_1}(\mathbf{P}_t)\} = \{\mathbf{P}_t\}, \\ \forall \mathbf{x}', \mathbf{x}' \rightarrow \mathbf{x} &:= |\mathbf{x}', \mathbf{x}_t|_{\mathcal{X}} < \delta_2, \end{aligned} \quad (16)$$

where $\mathcal{N}_{\delta_1}(\mathbf{P}_t)$ is the neighborhood of \mathbf{P}_t . Therefore, we can derive the differences of mapped \mathbf{x} as

$$\begin{aligned} \mathcal{H}(\mathbf{P}', \mathbf{x}') \setminus \mathcal{H}(\mathbf{P}_t, \mathbf{x}_t) &= \mathcal{H}(\mathbf{P}_t, \mathbf{x}_t) \setminus \mathcal{H}(\mathbf{P}', \mathbf{x}') \\ &= \mathcal{H}(\mathbf{P}_t, \mathbf{x}') \setminus \mathcal{H}(\mathbf{P}_t, \mathbf{x}_t) \\ &= \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} \Psi(\mathbf{P}_t, \mathbf{x}) \setminus \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} \Psi(\mathbf{P}_t, \mathbf{x}) = \emptyset. \end{aligned} \quad (17)$$

Since Eq. (17) implies that the cost matrix for the k -LAP generated by the regression variable $\mathbf{x} \in \mathcal{H}(\mathbf{P}, \mathbf{x})$ remains unchanged when the input of \mathcal{H} switches from $(\mathbf{P}_t, \mathbf{x}_t)$ to $(\mathbf{P}', \mathbf{x}')$, so will the mapped \mathbf{P} . Therefore, we have $\mathcal{H}(\mathbf{P}_t, \mathbf{x}_t)$ closed by further joining Eqs. (16) and (17).

As all the requirements of the Zangwill's convergence theorem are satisfied as presented above, we can now conclude that our algorithm is globally convergent to a critical point as defined by the solution set Γ .

5. Experiments and applications

We conduct experiments to demonstrate the differences between SLR and GSLR in terms of robust estimation, as well as to explore how GSLR performs on suitable applications. We limit the discussion within the computer vision field, although GSLR itself is a much more general problem.

5.1. Task 1: Image registration

We first carry out experiments on image registration to demonstrate the effectiveness of GSLR over SLR in terms of robust estimation. Given a pair of images portraying the same plane, or a non-planar scene subject to $\mathbf{0}$ translation in terms of the camera pose, homography describes the relation of $\mathbf{p}' = \alpha \mathbf{H}\mathbf{p}$, where \mathbf{p}' and \mathbf{p} are matched pixels on the two images in homogeneous coordinates, α is a scaling factor, and \mathbf{H} is the homography matrix. Furthermore, if images are captured with long focal lengths, \mathbf{H} can be assumed as affine transformations [17]. Since such an assumption implies that the scaling $\alpha = 1$, we can formulate the problem as

$$\min_{\mathbf{H} \mathbf{P} \in \Pi} \|\mathbf{P}\mathbf{H}\mathbf{A} - \mathbf{B}\|_F^2, \quad (18)$$

where \mathbf{A} and \mathbf{B} stand for the collections of pixels from the two images in homogeneous coordinates. This problem is regarded as a GSLR problem rather than an SLR one due to the inherent existence of outliers.

In the implementation, we use the SIFT feature [21] to retrieve the sets of pixel locations \mathbf{A} and \mathbf{B} . The initial

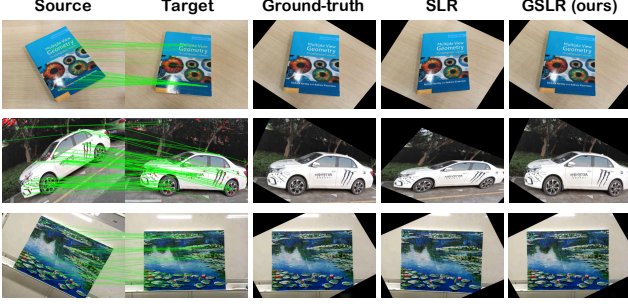


Figure 2: Results of image registration experiments. GSLR achieves visually the same results w.r.t. the ground-truths whereas SLR presents obvious distortions on all the trials. On the source and target pairs, outliers detected by GSLR are marked as red and inliers as green.

transformation \mathbf{H} of all the trials is set to the identity matrix. For optimizing SLR, we fix the number of inliers k of Algorithm 1 to the number of feature points. We use the mature descriptor and RANSAC-based method [17] to approximate the ground-truths. Fig. 2 shows the results. As shown, GSLR can effectively estimate the transformation matrices and provide visually the same results as the descriptor-based approach (denoted as Ground-truth) even without using the descriptors, where SLR results are obviously affected by outliers on all the trials.

5.2. Task 2: Point cloud registration

Point cloud registration works toward estimating the optimal rigid transformation \mathbf{T} between two point clouds. Its objective function can be written as

$$\min_{\mathbf{T} \in \text{SE}(3), \mathbf{P} \in \Pi} \|\mathbf{P}\mathbf{T}\mathbf{A} - \mathbf{B}\|_F^2, \quad (19)$$

where \mathbf{A} and \mathbf{B} represent the two point clouds in homogeneous coordinates. Usually, they would consist of numerous outliers that should never be used for estimating \mathbf{T} .

Experimental setup, metrics and peer methods We conduct experiments on both synthetic and real-world datasets. All pairs of point clouds are registered back and forth due to the unknown alignment order. For space limitation, we only report the transformational errors with the metric $\text{Trans. err} = \|\mathbf{T}_{\text{gt}}\mathbf{T}_{\text{esti}}^{-1} - \mathbf{I}\|_F$ in the main paper and put the respective rotational and translational counterparts in the supplementary material. For competing methods, we use ICP [3], Trim-ICP [8], SVR [7], CPD [24], FGR [47], and FilterReg [13] since they represent different strategies on handling outliers. Specifically, ICP works as the basic baseline, Trim-ICP shares similar outlier-rejection techniques with GSLR but employs the NN-based matching strategy, SVR treats both point clouds as mixture distributions and minimizes the robust $L2$ -divergence between them; FGR uses a

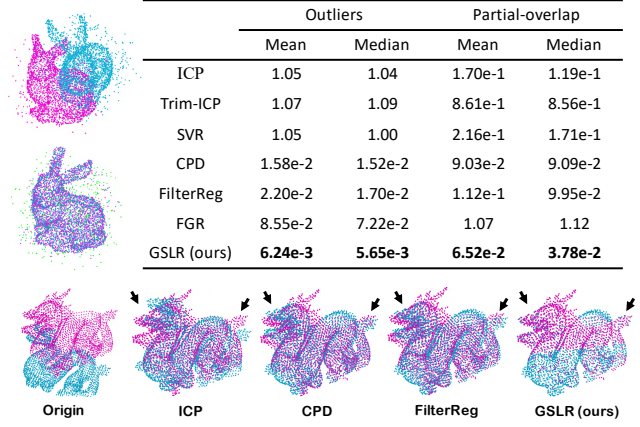


Figure 3: Results on the synthetic data. Upper left: Robustness of GSLR under the outlier-contaminated setup. Identified outliers are marked as green. Upper right: Summary of the mean and median transformational errors. Bottom: An example of the partially overlapped data. Trim-ICP, SVR, and FGR present similar visualizations and thus omitted.

scaled Geman-McClure estimator to remove outliers from matchings obtained via FPFH features [33], and CPD and FilterReg cast the registration problem into a distribution-fitting one and use a uniform distribution to cope with outliers. Other details are given in the supplementary material.

Registration with outliers and partial overlaps We use the synthetic bunny, dragon, and armadillo point clouds [9] for tests. For the outlier-contaminated cases, we first down-sample the point clouds to approximately 2500 points and corrupt them with Gaussian noises. Then outliers are added with ratios sampled from (0.25, 1), representing that 20% to 50% points are outliers. We repeat this procedure 10 times to obtain 30 trials in total. For the partial-overlapping setup, we down-size the point clouds to around 3000 points each, and randomly crop some parts from them. We retrieve 4 cropped point clouds from each shape and conduct 36 trials in total. For all the tests, we set the ground-truth rotation to 30° in Euler angles along each axis together with a random translation within $[0, 1]$. As shown in Fig. 3, SVR fails to provide accurate results as dense clusters of outliers are often similarly weighted as inliers. Moreover, compared to CPD and FilterReg, which can only down-weight the outliers, GSLR presents more accurate results by completely rejecting them.

Why ICP-style algorithms suffer from outliers Considering the fact that Trim-ICP and GSLR are equipped with similar robust estimation techniques (*i.e.*, binarily weighting the pairs), it is worthy of analyzing the substandard performances of Trim-ICP presented in Fig. 3.

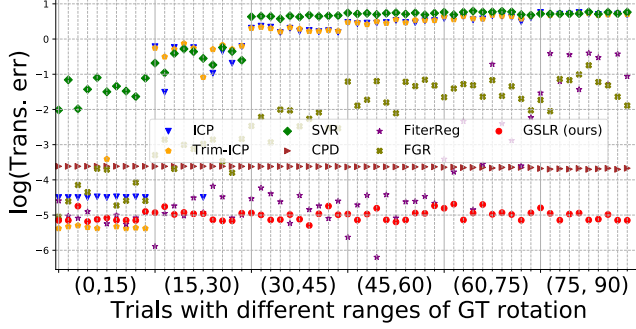


Figure 4: Sensitivity of different approaches w.r.t. initializations on outlier-contaminated point clouds.

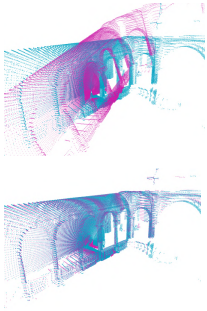


Figure 5: Registration results on the real-world dataset. Left: An example achieved by our GSLR formulation. Right: Mean and median transformational errors of all the trials.

It is the matching policy to blame. In detail, we observe that the NN-based matching strategy of ICP and its variants makes them extremely sensitive to initializations. For demonstration, we conduct a test with different initializations on the aforementioned bunny point clouds with 1/3 points as outliers. The ground-truth rotations in Euler angles are split into 6 ranges, and the registration is repeated 10 times within each of them. Fig. 4 records the results. As shown, ICP and Trim-ICP can also present reasonable results given plausible initializations. In such a condition, Trim-ICP can even provide more accurate estimations compared to the others owing to its large trimming ratio. However, they would fail drastically when the initializations are coarsened. On the contrary, GSLR can remarkably alleviate this problem via the one-to-one correspondences, demonstrated by the stable performances even under large initialization errors.

Real-world registration We also examine the performances of GSLR on real-world point clouds. For experimental setup, we establish 7 groups consisting of 5 consecutive frames from the ETH Hauptgebäude dataset [29], and conduct pairwise registration within each group, resulting in 140 trials altogether. This dataset is known challenging

because of repeated patterns and a minimum overlap ratio of 63.9%. To let ICP-style algorithms perform reasonably, we roughly pre-align the point clouds by fixing the ground-truth rotations to 15° in Euler angles. All point clouds are down-sampled to around 3500 points. Figure 5 reports the results. As shown, our method exhibits the highest accuracy compared to other state-of-the-art methods. On the other hand, since the local features are extremely unreliable when repeated patterns exist, FGR fails to present reasonable results.

Other results and experiments We have also conducted other experiments to analyze statistical efficiencies, elapsed time, as well as sensitivities w.r.t. different scales of noises and ratios of outliers. The results are provided in the supplementary material for interested readers.

5.3. Task 3: Isometric shape matching

In contrast to point cloud registration that concentrates on the regression variable, isometric shape matching looks for vertex-wise correspondences between meshes. Following functional map [26], its objective function can be written as

$$\min_{\mathbf{C} \in \mathbf{O}(s), \mathbf{P} \in \Pi} \|\mathbf{PCF} - \mathbf{G}\|_F^2, \quad (20)$$

where $\mathbf{O}(\cdot)$ denotes the orthogonal group, and \mathbf{F} and \mathbf{G} stand for the vertex-wise Laplacian eigenfunctions of the two meshes. In general, they include outliers since the mappings can seldom be ideally isometric. Therefore, robust estimation is still required although the underlying ground-truth is a full-rank matching [30].

Experimental setup, metrics and peer methods We carry out experiments on commonly used datasets. Specifically, we conduct both intra-class and inter-class matching on the FAUST dataset [5], and intra-class matching on the TOSCA one [6]. Inter-class tests on TOSCA are out of scope due to their high non-isometry. For each task, we randomly select 30 pairs of shapes and report the mean values as the final results. For the vertices that are labeled as outliers by GSLR, we simply map them to their nearest neighbors with the estimated \mathbf{C} , although more complicated methods (e.g., searching for correspondences within the connected components) can be employed to improve smoothness. For evaluation, we employ the commonly used correspondences-geodesic metric to quantify the performances. Moreover, we also follow the definition of coverage presented in [30] to measure the bijectivity, which is defined as the ratio of vertices that possess at least one correspondence to their total amount. We select ICP [26], BCICP [30] and ZoomOut [23] as the competing methods. For setup, we use 50-dimensional eigenfunctions of the Laplace-Beltrami operator to formulate the sets of functions \mathbf{F} and \mathbf{G} , and initialize all the

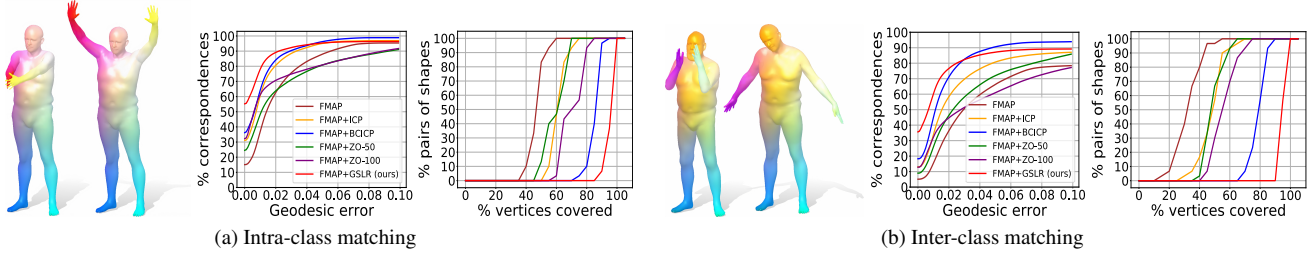


Figure 6: Matching results on the FAUST dataset. Within each sub-figure, left shows an example achieved by GSLR, middle is the geodesic errors, and right shows the coverage rate. For the labels, FMAP is short for functional map, and FMAP+ZO-50 and FMAP+ZO-100 respectively stand for the ZoomOut algorithm with 50 and 100-dimensional eigenfunctions.

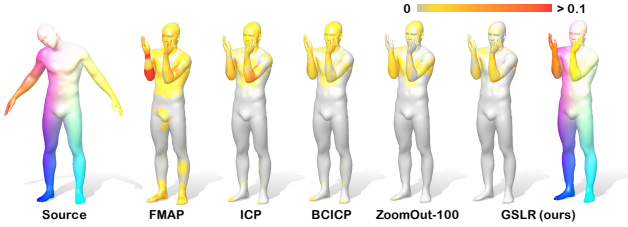


Figure 7: An example of geodesic error maps on the FAUST dataset. Comparing to the others, more correspondences estimated by GSLR coincide with the ground-truth.

algorithms with the wave kernel signature [4] and functional map under the direct operator [30]. More details are given in the supplementary material.

Results Figure 6 shows the overall results on the FAUST dataset, and Fig. 7 depicts the estimation errors. Compared to the others, GSLR can provide more accurate results by estimating obviously higher percentages of matches that coincide with the ground-truth. Moreover, it can also effectively maintain the one-to-one correspondences by covering more than 80% vertices on all the pairs. For the TOSCA dataset, we decimate the meshes to approximately 3400 vertices each and obtain the error curves via NN-composition. Fig. 8 records the results. As presented, the GSLR formulation can still champion both the matching accuracy and the bijectivity. We emphasize that ZoomOut-100 doubles the dimensions of input eigenfunctions although it achieves similar results as GSLR in terms of accuracy. An example of coverages is given in Fig. 9.

6. Discussion and conclusion

In this work, we propose the GSLR formulation together with an optimization algorithm to generalize SLR to broader-ranging conditions where only part of the data should match. Distinct applications demonstrate its effectiveness.

The main limitation of our current algorithm lies in time efficiency. *I.e.*, while the k -LAP is tractable for medium-

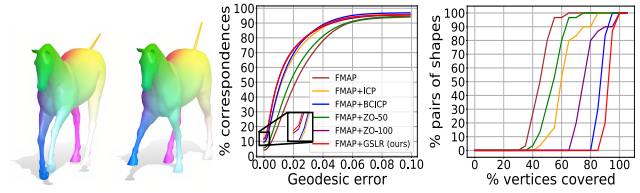


Figure 8: Matching results on the TOSCA dataset. Left: An example matching achieved by GSLR. Right: Geodesic errors and coverages.

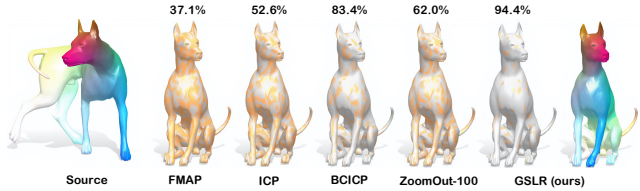


Figure 9: An example of coverages on the TOSCA dataset. Vertices violating bijection are marked as orange. Upper percentages indicate the coverage rates.

sized problems (*e.g.*, the sizes in our experiments vary from $3k$ to $8k$), the elapsed time would increase significantly when dealing with giant ones. Fortunately, this problem can be effectively solved with off-the-shelf CUDA-based Hungarian algorithms [10, 43], which are hundreds of times faster.

For future work, GSLR requires more theoretical analysis. For example, we are interested in studying whether the conditions of the unique solution proposed in [28] for SLR can generalize to GSLR, as well as the convergence rate of Algorithm 1. For applications, we wish to work on variants of GSLR for other tasks, such as partial shape matching and registration with point-to-plane metrics.

Acknowledgment

This work was supported by NII CRIS collaborative research program operated by NII CRIS and LINE Corporation.

References

- [1] Abubakar Abid, Ada Poon, and James Zou. Linear regression with shuffled labels. *arXiv preprint arXiv:1705.01342*, 2017. [2](#)
- [2] Abubakar Abid and James Zou. A stochastic expectation-maximization approach to shuffled linear regression. In *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, 2018. [2](#)
- [3] K Somani Arun, Thomas S Huang, and Steven D Blostein. Least-squares fitting of two 3-d point sets. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 9(5):698–700, 1987. [2](#), [6](#)
- [4] Mathieu Aubry, Ulrich Schlickewei, and Daniel Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *International conference on computer vision workshops (ICCV workshops)*, 2011. [8](#)
- [5] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. [7](#)
- [6] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. *Numerical geometry of non-rigid shapes*. 2008. [7](#)
- [7] Dylan Campbell and Lars Petersson. An adaptive data representation for robust point-set registration and merging. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015. [6](#)
- [8] Dmitry Chetverikov, Dmitry Svirko, Dmitry Stepanov, and Pavel Krsek. The trimmed iterative closest point algorithm. *Object recognition supported by user interaction for service robots*, 3:545–548, 2002. [3](#), [6](#)
- [9] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996. [6](#)
- [10] Ketan Date and Rakesh Nagi. Gpu-accelerated hungarian algorithms for the linear assignment problem. *Parallel Computing*, 57:52–72, 2016. [8](#)
- [11] Jurgen A Doornik. Robust estimation using least trimmed squares. *Institute for Economic Modelling, Oxford Martin School, and Economics Department, University of Oxford, UK*, 2011. [3](#)
- [12] Fajwel Fogel, Rodolphe Jenatton, Francis Bach, and Alexandre d’Aspremont. Convex relaxations for permutation problems. In *Proceedings of Conference on Neural Information Processing Systems (NIPS)*, 2013. [1](#)
- [13] Wei Gao and Russ Tedrake. Filterreg: Robust and efficient probabilistic point-set registration using gaussian filter and twist parameterization. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. [6](#)
- [14] Steven Gold, Anand Rangarajan, Chien-Ping Lu, Suguna Pappu, and Eric Mjolsness. New algorithms for 2d and 3d point matching: pose estimation and correspondence. *Pattern Recognition*, 31(8):1019–1031, 1998. [2](#)
- [15] Asela Gunawardana and William Byrne. Convergence theorems for generalized alternating minimization procedures. *Journal of Machine Learning Research*, 6:2049–2073, 2005. [4](#)
- [16] Frank R Hampel. The breakdown points of the mean combined with some rejection rules. *Technometrics*, 27(2):95–107, 1985. [3](#), [4](#)
- [17] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. 2003. [5](#), [6](#)
- [18] Douglas M Hawkins. The feasible solution algorithm for least trimmed squares regression. *Computational statistics & data analysis*, 17(2):185–196, 1994. [4](#)
- [19] Daniel J Hsu, Kevin Shi, and Xiaorui Sun. Linear regression without correspondence. In *Proceedings of Conference on Neural Information Processing Systems (NIPS)*, 2017. [2](#)
- [20] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955. [3](#)
- [21] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of International Conference on Computer Vision (ICCV)*, 1999. [5](#)
- [22] Joao Maciel and Joao Paulo Costeira. A global solution to sparse correspondence problems. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 25(2):187–199, 2003. [2](#)
- [23] Simone Melzi, Jing Ren, Emanuele Rodolà, Abhishek Sharma, Peter Wonka, and Maks Ovsjanikov. Zoomout: Spectral upsampling for efficient shape correspondence. *Transactions on Graphics (TOG)*, 38(6), 2019. [7](#)
- [24] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(12):2262–2275, 2010. [6](#)
- [25] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *Proceedings of Symposium on Security and Privacy*, 2008. [1](#)
- [26] Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: a flexible representation of maps between shapes. *Transactions on Graphics (TOG)*, 31(4):1–11, 2012. [3](#), [7](#)
- [27] Ashwin Pananjady, Martin J Wainwright, and Thomas A Courtade. Denoising linear models with permuted data. In *Proceedings of International Symposium on Information Theory (ISIT)*, 2017. [2](#)
- [28] Ashwin Pananjady, Martin J Wainwright, and Thomas A Courtade. Linear regression with shuffled data: Statistical and computational limits of permutation recovery. *Transactions on Information Theory*, 64(5):3286–3300, 2017. [1](#), [2](#), [8](#)
- [29] François Pomerleau, M. Liu, Francis Colas, and Roland Siegwart. Challenging data sets for point cloud registration algorithms. *International Journal of Robotics Research*, 31(14):1705–1711, Dec. 2012. [7](#)
- [30] Jing Ren, Adrien Poulenard, Peter Wonka, and Maks Ovsjanikov. Continuous and orientation-preserving correspondences via functional maps. *Transactions on Graphics (TOG)*, 37(6):1–16, 2018. [2](#), [7](#), [8](#)
- [31] Emanuele Rodolà, Luca Cosmo, Michael M Bronstein, Andrea Torsello, and Daniel Cremers. Partial functional correspondence. In *Computer Graphics Forum*, 2017. [3](#)
- [32] Szymon Rusinkiewicz. A symmetric objective function for icp. *Transactions on Graphics (TOG)*, 38(4):1–7, 2019. [2](#)

- [33] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *Proceedings of International Conference on Robotics and Automation (ICRA)*, 2009. 6
- [34] Martin Slawski, Emanuel Ben-David, et al. Linear regression with sparsely permuted data. *Electronic Journal of Statistics*, 13(1):1–36, 2019. 2
- [35] Martin Slawski, Emanuel Ben-David, and Ping Li. A two-stage approach to multivariate linear regression with sparsely mismatched data. *Journal of Machine Learning Research*, 2019. 2
- [36] Robert G Staudte and Simon J Sheather. *Robust estimation and testing*, volume 918. 2011. 2
- [37] Jayakrishnan Unnikrishnan, Saeid Haghighatshoar, and Martin Vetterli. Unlabeled sensing with random linear measurements. *Transactions on Information Theory*, 64(5):3237–3253, 2018. 2
- [38] Matthias Vestner, Zorah Löhner, Amit Boyarski, Or Litany, Ron Slossberg, Tal Remez, Emanuele Rodola, Alex Bronstein, Michael Bronstein, Ron Kimmel, et al. Efficient deformable shape correspondence via kernel matching. In *Proceedings of International Conference on 3D Vision (3DV)*, 2017. 2
- [39] Matthias Vestner, Roei Litman, Emanuele Rodolà, Alex Bronstein, and Daniel Cremers. Product manifold filter: Non-rigid shape correspondence via kernel density estimation in the product space. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2
- [40] Anton Volgenant. Solving the k-cardinality assignment problem by transformation. *European Journal of Operational Research*, 157(2):322–331, 2004. 3
- [41] Fudong Wang, Nan Xue, Jin-Gang Yu, and Gui-Song Xia. Zero-assignment constraint for graph matching with outliers. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3
- [42] CF Jeff Wu. On the convergence properties of the em algorithm. *The Annals of statistics*, pages 95–103, 1983. 4
- [43] Satyendra Singh Yadav, Paulo Alexandre Crisóstomo Lopes, Aleksandar Ilic, and Sarat Kumar Patra. Hungarian algorithm for subcarrier assignment problem using gpu and cuda. *International Journal of Communication Systems*, 32(4):e3884, 2019. 8
- [44] Willard I Zangwill. *Nonlinear programming: a unified approach*, volume 52. 1969. 4
- [45] Hang Zhang, Martin Slawski, and Ping Li. Permutation recovery from multiple measurement vectors in unlabeled sensing. In *Proceedings of International Symposium on Information Theory (ISIT)*, 2019. 1, 2
- [46] Ruonan Zhang and Wenmin Wang. Second-and high-order graph matching for correspondence problems. *Transactions on Circuits and Systems for Video Technology*, 28(10):2978–2992, 2017. 2
- [47] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast global registration. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2016. 6
- [48] Abdelhak M Zoubir, Visa Koivunen, Yacine Chakhchoukh, and Michael Muma. Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. *Signal Processing Magazine*, 29(4):61–80, 2012. 2