Unlabelled Compressive Sensing under Sparse Permutation and Prior Information

Anonymous authors
Paper under double-blind review

Abstract

In this paper, we study the problem of unlabelled compressed sensing, where the correspondence between the measurement values and the rows of the sensing matrix is lost, the number of measurements is less than the dimension of the regression vector, and the regression vector is sparse in the identity basis. Additionally, motivated by practical situations, we assume that we accurately know a small number of correspondences between the rows of the measurement matrix and the measurement vector. We propose a tractable estimator, based on a modified form of the Lasso, to estimate the regression vector, and we derive theoretical error bounds for the estimate. This is unlike previous approaches to unlabelled compressed sensing, which either do not produce theoretical bounds or which produce bounds for intractable estimators. We show that our algorithm outperforms a hard thresholding pursuit (Htp) approach and an ℓ_1-norm estimator used to solve a similar problem across diverse regimes. We also propose a modified Htp based estimator which has superior properties to the baseline Htp estimator. Lastly, we show an application of unlabelled compressed sensing in image registration, demonstrating the utility of a few known point correspondences.

1 Introduction

Estimation of an unknown vector $\boldsymbol{\beta}^* \in \mathbb{R}^p$ from its (possibly noisy) linear measurements $\boldsymbol{A}\boldsymbol{\beta}^* \in \mathbb{R}^N$, where $\boldsymbol{A} \in \mathbb{R}^{N \times p}$ is the measurement matrix, is a well-studied problem that naturally arises in numerous fields. However, in some scenarios, the correspondences between the rows of the measurement matrix and the measurements get partially or entirely lost due to errors made during the measurement process. Specifically, the permuted and noisy measurements are given as

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{A}\boldsymbol{\beta}^* + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{w} \in \mathbb{R}^N$ is the measurement noise vector and $\boldsymbol{P} \in \mathbb{R}^{N \times N}$ denotes an unknown permutation matrix. Hence, the problem reduces to that of estimating the unknown regression vector from an unknown permutation of its measurements. This problem, known as *unlabelled sensing*, appears in applications such as simultaneous localization and mapping in robotics Thrun & Leonard (2008), record linkage Lahiri & Larsen (2005), image or point cloud registration Pan & Zhang (2005); Li et al. (2021), security and privacy Narayanan & Shmatikov (2008), and simultaneous pose and correspondence determination in computer vision Marques et al. (2009).

Unlabelled sensing is well studied theoretically in the over-determined regime where $N > p$ and typically, under the assumption that the entries of $\mathbf{A}$ are drawn i.i.d. from an arbitrary continuous probability distribution. Within these settings and without measurement noise, Unnikrishnan et al. Unnikrishnan et al. (2018) and Han et al. Han et al. (2018) state that every $\boldsymbol{\beta}^* \in \mathbb{R}^p$ can be uniquely recovered with probability one from its measurements $\boldsymbol{P}\boldsymbol{A}\boldsymbol{\beta}^*$ if $N \geq 2p$. Under some more assumptions on $\boldsymbol{\beta}^*$, the condition $N \geq 2p$ can be relaxed to $N \geq p + 1$ Tsakiris & Peng (2019). The over-determined unlabelled sensing problem is challenging in the presence of noise. In Pananjady et al. (2018); Hsu et al. (2017), authors have derived lower bounds on the signal-to-noise ratio (SNR) for approximate recovery of $\boldsymbol{\beta}^*$. Algorithms to solve the problem for different values of p are suggested in Peng & Tsakiris (2020); Tsakiris & Peng (2019); Candes

& Tao (2005); Slawski & Ben-David (2019). Sresth et al. (2024) extends the approach in Slawski & Ben-David (2019) by assuming that some correspondences between the rows of the measurement matrix and measurement vector are known.

An under-determined unlabelled sensing problem, where $N < p$, can be solved with some additional priors on $\boldsymbol{\beta^*}$. To the best of our knowledge, Peng et al. (2021) is the only work in the context of unlabelled compressive sensing. In Peng et al. (2021), the authors assumed that $\boldsymbol{\beta}^*$ is k-sparse (i.e., it has at most k non-zero elements) and showed that $N \geq 2k$ measurements are sufficient for unique recovery in the absence of noise. Interestingly, the result is analogous to that proved for the over-determined unlabelled sensing problem from Unnikrishnan et al. (2018). On the algorithmic side, a hard-thresholding pursuit-based approach is proposed for unlabelled compressive sensing in Peng et al. (2021). However, no theoretical error bounds on the performance of their *algorithm* are established in Peng et al. (2021), and moreover, their algorithm involves a computationally expensive ℓ_1-norm optimization step to determine the unknown vector $\boldsymbol{\beta^*}$. Also, Peng et al. (2021) employs a subgradient method in their algorithm where the function value may not always decrease across the iterations of the algorithm, as mentioned in their paper.

Possible Prior Information: In several unlabelled sensing applications, such as image alignment, record-linkage, etc., we encounter sparse permutation scenarios, that is, the correspondence between the measurements in $\boldsymbol{y}$ and the rows of $\boldsymbol{A}$ is incorrect for only a small fraction of measurements. Equivalently, we have that $\|\boldsymbol{PA\beta^*} - \boldsymbol{A\beta^*}\|_0 \ll N$. Moreover, in many applications, some of the correspondences between the rows of $\boldsymbol{A}$ and the rows of $\boldsymbol{y}$ are known a priori. For example, in the image alignment problem, many point correspondences in the two images being matched are obtained via feature point matching methods Lowe (2004), with permutations often occurring due to factors such as image self-similarity (i.e., similarity of patches in distant image regions). However, a domain expert can manually mark a few corresponding point pairs in the fixed and moving images, thus yielding some measurements with correspondences known in advance.

Nguyen & Tran (2012) propose a robust form of least absolute shrinkage and selection operator (Lasso), which can estimate the sparse regression vector from its grossly corrupted linear measurements. Treating the permuted measurements as gross corruptions, robust Lasso can be employed to solve unlabelled compressed sensing problems. However, as motivated previously, if there is a prior knowledge of some correspondences between the rows of $\boldsymbol{A}$ and the rows of $\boldsymbol{y}$, robust Lasso is not able to make use of it. A natural question is how the prior knowledge of some correspondences would improve the estimation of $\boldsymbol{\beta^*}$. Such questions are answered in an over-determined setup in Sresth et al. (2024). However, these questions are still open to under-determined, unlabeled sensing problems.

1.1 Our Contribution

Our paper presents the following contributions:

1. We propose two algorithms to solve the unlabelled compressed sensing problem which can make use of known correspondences is available:
   (i) Augmented Robust Lasso (Ar-Lasso), a modification of the standard least absolute shrinkage and selection operator (Lasso).
   (ii) Augmented Hard Thresholding Pursuit (A-Htp) which involves a gradient-descent step and computationally cheaper ℓ_2-norm optimization step, rather than subgradient-method and ℓ_1-norm optimization as required in the earlier approach proposed in Peng et al. (2021).

   Ar-Lasso is suitable for both compressible and perfectly sparse regression vectors, unlike the earlier approach in Peng et al. (2021), which works only for perfectly sparse regression vectors. However, A-Htp can work only in the case of perfectly sparse regression vectors. Note that Ar-Lasso and A-Htp algorithms are applicable even when there are no correspondences known. Moreover, apart from the unlabelled sensing problem, our algorithms are applicable in any regression problem where some measurements are grossly corrupted, and the measurements can be separated into two disjoint sets: (*i*) measurements without any gross corruption and (*ii*) measurements with possible gross corruption.

2. For both Ar-Lasso and A-Htp algorithms, we derive theoretical upper bounds on the estimation error of the unknown vector in terms of the number of measurements N, number of permutations s, dimension of the unknown vector p, sparsity of the unknown vector k, number of known correspondences m and measurement noise variance σ^2. Specifically, for Ar-Lasso, we characterize the notion of a generalized, extended, restricted eigenvalue condition (Gerec), which enables us to prove performance guarantees for the algorithm. Gerec is a generalisation of the extended, restricted eigenvalue condition studied in Nguyen & Tran (2012), to a scenario where some of the correspondences are (possibly) known in advance. We show that the family of Gaussian measurement matrices obeys Gerec with a high probability. We compare our error bounds to those obtained in Nguyen & Tran (2012) and demonstrate that the information of known correspondences allows us to tolerate a larger number of permutations and also results in a lower estimation error.

3. Further, we demonstrate a geometric convergence result for A-Htp under the condition that the sensing matrix obeys some form of restricted isometry property. A-Htp involves joint-estimation of $\boldsymbol{x}^*$, which is a row concatenation of $\boldsymbol{\beta}^*$ and the permutation corruption vector, through an augmented matrix $\left[\begin{array}{c|c} \boldsymbol{A} & \begin{array}{c}\mathbf{0}_{m\times(N-m)} \\ \boldsymbol{I}_{(N-m)\times(N-m)}\end{array}\end{array}\right]$. In order to exploit the distribution of non-zero entries in $\boldsymbol{x}^*$, we introduce a notion of structured-sparsity restricted isometry property. Following this, we demonstrate that $N \geq C(k\log(p/3k) + s\log(n/3s))$ is a sufficient condition for an accurate recovery via A-Htp, which is a more relaxed requirement than $N \geq C((k+s)\log\frac{p+n}{3(k+s)})$ which is what one obtains from a naive analysis without exploiting the specific sparsity structure of $\boldsymbol{x}^*$.

4. Next, we compare our algorithms to the ℓ_1-norm hard-thresholding pursuit approach from Peng et al. (2021) and to another ℓ_1-norm-based estimator motivated from Candes & Tao (2005), across diverse regimes. We demonstrate that our algorithms outperform them across all the regimes examined.

5. Lastly, we demonstrate the impact of utilizing known correspondences in an image registration task. For this task, the problem of unlabeled compressed sensing with sparse permutations and a set of priors is especially relevant in the following manner: (i) In many image registration tasks, a domain expert can mark out a few point correspondences accurately. This provides prior information for the regression problem. (ii) In image registration, the underlying motion vector fields are sparse or compressible in universal dictionaries such as the discrete Fourier transform or discrete cosine transform James et al. (2019). If salient feature point tracking is used for obtaining point correspondences, then the motion vectors at only a small set of points (in the image domain) are observed. This is therefore a compressed sensing or sparse recovery problem. (iii) A fraction of the computed point correspondence pairs suffer from permutation effects due to factors like self-similarity. This is therefore an unlabeled sensing problem where the permutation set is sparse. For more details on this, please refer to Section 1 and Figure 1 of the supplemental material.

2 Problem Formulation and Notations

Consider a set of linear measurements as in Eq. 1 with $N < p$ and the following assumptions:

(C1) $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_N)$ and $\boldsymbol{A}^i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{A}^i \in \mathbb{R}^p$ is the i-th row of the measurement matrix $\boldsymbol{A}$, and $\boldsymbol{I}_N$ stands for the $N \times N$ identity matrix. We denote the smallest eigenvalue of $\boldsymbol{\Sigma}$ by $C_{\min}(\boldsymbol{\Sigma})$, the largest eigenvalue of $\boldsymbol{\Sigma}$ by $C_{\max}(\boldsymbol{\Sigma})$ and the largest entry on the diagonal of $\boldsymbol{\Sigma}$ by $\xi(\boldsymbol{\Sigma})$.

(C2) The regression vector $\boldsymbol{\beta}^*$ is k-sparse where k need not be known in advance. We denote the support of non-zero entries of $\boldsymbol{\beta}^*$ by the set T.

(C3) Any m out of N measurements have accurate correspondences where $m < p$. Without loss of generality, we assume that the correspondences of the top m measurements are accurate, that is, $y_i = \boldsymbol{A}^i \boldsymbol{\beta}^* + w_i$ for $i = 1, 2, \ldots, m$.

(C4) In the remaining $n = N - m < p$ measurements, at most $s \ll n$ measurements have incorrect correspondences. However, we do not know the value of s and which s measurements have incorrect correspondences.

With assumptions (C1)-(C4), we decompose $\boldsymbol{y}$ in Eq. 1 as

$$\boldsymbol{y}_1 = \boldsymbol{A}_1\boldsymbol{\beta}^* + \boldsymbol{w}_1 \text{ and } \boldsymbol{y}_2 = \boldsymbol{P}_2\boldsymbol{A}_2\boldsymbol{\beta}^* + \boldsymbol{w}_2, \tag{2}$$

where $\boldsymbol{y}_1 \in \mathbb{R}^m$ denotes the sub-vector of measurements with known correspondences, $\boldsymbol{y}_2 \in \mathbb{R}^n$ is the sub-vector of remaining measurements (with $m + n = N$) and $\boldsymbol{P}_2 \in \mathbb{R}^{n\times n}$ is an unknown permutation matrix. We denote $\boldsymbol{y} := \begin{bmatrix}\boldsymbol{y}_1^T & \boldsymbol{y}_2^T\end{bmatrix}^T, \boldsymbol{A} := \begin{bmatrix}\boldsymbol{A}_1^T & \boldsymbol{A}_2^T\end{bmatrix}^T$ and $\boldsymbol{w} := \begin{bmatrix}\boldsymbol{w}_1^T & \boldsymbol{w}_2^T\end{bmatrix}^T$. For further analysis, define the permutation corruption vector as

$$\boldsymbol{z}^* = \sqrt{n}\boldsymbol{e}^* := (\boldsymbol{P}_2\boldsymbol{A}_2\boldsymbol{\beta}^* - \boldsymbol{A}_2\boldsymbol{\beta}^*) \in \mathbb{R}^n. \tag{3}$$

The assumption (C4) implies that $\boldsymbol{e}^*$ is s-sparse. We denote the support of non-zero entries of $\boldsymbol{e}^*$ by the set S. Using Eq. 3, we can write Eq. 2 as $\boldsymbol{y}_2 = \boldsymbol{A}_2\boldsymbol{\beta}^* + \sqrt{n}\boldsymbol{e}^* + \boldsymbol{w}_2$. The goal is to estimate $\boldsymbol{\beta}^*$ under assumptions (C1)-(C4). To this end, we provide the formulations for Ar-Lasso and A-Htp algorithms in the next sections.

3 Augmented Robust Lasso

With the introduction of $\boldsymbol{e}^*$, the objective of estimating $\boldsymbol{\beta}^*$ from $\boldsymbol{y}$ under assumptions (C1)-(C4) is posed as a convex optimization problem. Specifically, we propose to minimize an augmented, robust, least-absolute shrinkage and selection operator (Ar-Lasso)-based objective function, given by:

$$L(\boldsymbol{\beta}, \boldsymbol{e}) := \frac{1}{2m}\|\boldsymbol{y}_1 - \boldsymbol{A}_1\boldsymbol{\beta}\|_2^2 + \frac{1}{2n}\|\boldsymbol{y}_2 - \boldsymbol{A}_2\boldsymbol{\beta} - \sqrt{n}\boldsymbol{e}\|_2^2 + \lambda_\beta\|\boldsymbol{\beta}\|_1 + \lambda_e\|\boldsymbol{e}\|_1. \tag{4}$$

Here, the first two terms are data-fit terms, and $\|\boldsymbol{\beta}\|_1$, $\|\boldsymbol{e}\|_1$ are the sparsity-promoting terms in the regression vector and corruption vector, respectively. Next, we formulate the optimization problem as

$$(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{e}}) := \underset{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{e}\in\mathbb{R}^n}{\arg\min}\; L(\boldsymbol{\beta}, \boldsymbol{e}). \tag{5}$$

The question remains, under what conditions Ar-Lasso is able to estimate $\boldsymbol{\beta}^*$ with a low error? In the next subsection, we discuss the generalized, extended, restricted eigenvalue condition based on which we present theoretical guarantees for Ar-Lasso to answer such questions.

3.1 Theoretical Guarantees for Ar-Lasso

First, define set $\mathcal{C} := \{(\boldsymbol{h}, \boldsymbol{f}) \in (\mathbb{R}^p \times \mathbb{R}^n)$ such that $\|\boldsymbol{h}_{\boldsymbol{T}^{\boldsymbol{C}}}\|_1 + \lambda\|\boldsymbol{f}_{\boldsymbol{S}^{\boldsymbol{C}}}\|_1 \leq 3\|\boldsymbol{h}_{\boldsymbol{T}}\|_1 + 3\lambda\|\boldsymbol{f}_{\boldsymbol{S}}\|_1\}$ where $\lambda = \frac{\lambda_e}{\lambda_\beta}$. It can be shown that the error term $(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*, \tilde{\boldsymbol{e}} - \boldsymbol{e}^*)$ belongs to the set $\mathcal{C}$ (see the supplemental material) and hence expectedly, the theoretical guarantees for Ar-Lasso rely on the following two key properties of sensing matrix $\boldsymbol{A}$ over the restricted set $\mathcal{C}$:

(1) **Generalised, extended, restricted eigenvalue condition** (Gerec): We say that the measurement matrix $\boldsymbol{A}$ obeys the generalized, extended, restricted, eigenvalue condition on the restricted set $\mathcal{C}$ if there exists $k_e > 0$ such that $k_e(\|\boldsymbol{h}\|_2^2 + \|\boldsymbol{f}\|_2^2) \leq \frac{1}{N}\|\boldsymbol{A}\boldsymbol{h}\|_2^2 + \frac{n}{N}\|\boldsymbol{f}\|_2^2 \;\forall\; (\boldsymbol{h}, \boldsymbol{f}) \in \mathcal{C}$. Recall that $n \leq N$. The above condition is a generalisation of the extended, restricted eigenvalue condition studied in Nguyen & Tran (2012), to the scenario where some of the correspondences are known, that is when $n \leq N$. To give more context, we use the above condition with $(\boldsymbol{h}, \boldsymbol{f}) := (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*, \tilde{\boldsymbol{e}} - \boldsymbol{e}^*)$ for theoretical analysis of Ar-Lasso. Essentially, $\boldsymbol{f}$ plays the role of error in estimation of the permutation corruption vector $\boldsymbol{z}^* := \boldsymbol{P}_2\boldsymbol{A}_2\boldsymbol{\beta}^* - \boldsymbol{A}_2\boldsymbol{\beta}^*$. Note the difference between Gerec and the extended, restricted eigenvalue condition in Nguyen & Tran (2012). In Nguyen & Tran (2012), the authors assume that each of the N measurements can undergo gross corruption and the aim is to lower-bound the term $(\frac{1}{N}\|\boldsymbol{A}\boldsymbol{h}\|_2^2 + \|\boldsymbol{f}\|_2^2)$ where $\boldsymbol{A} \in \mathbb{R}^{N\times p}, \boldsymbol{h} \in \mathbb{R}^p$ and $\boldsymbol{f} \in \mathbb{R}^N$ (see comment 3(a) after Lemma 1 in Nguyen & Tran (2012)). On the other hand, we have assumed knowledge of m correspondences, and hence, we are only concerned with permutation errors in the remaining $n = N - m$ measurements. Due to this, we have $\boldsymbol{f} \in \mathbb{R}^n$ as opposed to $\boldsymbol{f} \in \mathbb{R}^N$ as in Nguyen & Tran (2012). Alternatively, our $\boldsymbol{f}$ can be thought of as an N-dimensional vector

whose top m entries are zero. When none of the correspondences are known, that is, when $m = 0, n = N$, the generalized, extended, restricted eigenvalue condition reduces to the extended, restricted eigenvalue condition defined in Nguyen & Tran (2012).

(2) **Mutual incoherence condition** (Mic) Nguyen & Tran (2012): We require that there exists $k_m > 0$ such that: $\frac{1}{\sqrt{n}}|\langle \boldsymbol{A}_2\boldsymbol{h}, \boldsymbol{f}\rangle| \leq k_m(\|\boldsymbol{h}\|_2 + \|\boldsymbol{f}\|_2)^2 \ \forall \ (\boldsymbol{h}, \boldsymbol{f}) \in \mathcal{C}$. We refer to $k_m > 0$ as the mutual incoherence constant. The mutual incoherence condition allows decoupling the two error terms $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2$ and $\|\tilde{\boldsymbol{e}} - \boldsymbol{e}^*\|_2$ given that we use $(\boldsymbol{h}, \boldsymbol{f}) := (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*, \tilde{\boldsymbol{e}} - \boldsymbol{e}^*)$ for theoretical analysis of Ar-Lasso.
Next, the question arises: Under what conditions does a measurement matrix obey Gerec and Mic? It can be shown that the family of Gaussian sensing matrices satisfies Gerec with suitable assumptions on (N, m, s, k). To this end, we state the following lemma (proof in supplemental material):

Lemma 1 (**Generalised, extended, restricted eigenvalue condition**). *Consider the Gaussian sensing matrix* $\boldsymbol{A} \in \mathbb{R}^{N\times p}$ *whose rows are i.i.d.* $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. *We have the set* $\mathcal{C} := \{(\boldsymbol{h}, \boldsymbol{f}) \in (\mathbb{R}^p \times \mathbb{R}^n)$ *such that* $\|\boldsymbol{h}_{\boldsymbol{T}^C}\|_1 + \lambda\|\boldsymbol{f}_{\boldsymbol{S}^C}\|_1 \leq 3\|\boldsymbol{h}_{\boldsymbol{T}}\|_1 + 3\lambda\|\boldsymbol{f}_{\boldsymbol{S}}\|_1\}$ *with* $|T| = k$ *and* $|S| = s$ *as defined earlier. Select* $\lambda = \rho\sqrt{\frac{n}{N}\frac{\log n}{\xi(\boldsymbol{\Sigma})\log p}}$, *where* $\rho \in (0, 1)$ *is a constant. If* $s \leq c_1\frac{N}{\rho^2\log n}$ *and* $N \geq c_2\frac{\xi(\boldsymbol{\Sigma})}{C_{min}(\boldsymbol{\Sigma})}k\log p$, *then the following inequality holds with probability at least* $1 - c_3\exp(-c_4N)$: $min\left(\frac{C_{min}(\boldsymbol{\Sigma})}{128}, \frac{n}{8N}\right)(\|\boldsymbol{h}\|_2^2 + \|\boldsymbol{f}\|_2^2) \leq \frac{1}{N}\|\boldsymbol{A}\boldsymbol{h}\|_2^2 + \frac{n}{N}\|\boldsymbol{f}\|_2^2 \ \forall \ (\boldsymbol{h}, \boldsymbol{f}) \in \mathcal{C}$, *where* c_1, c_2, c_3, c_4 *are positive constants.*

We make the following observations: (i) We reiterate that the knowledge of m correspondences requires us to only consider $\boldsymbol{f} \in \mathbb{R}^n$ as opposed to considering $\boldsymbol{f} \in \mathbb{R}^N$. As a result, the sensing matrix $\boldsymbol{A}$ obeys Gerec with a larger tolerance on s, that is, $s \leq c_1\frac{N}{\rho^2\log n} = c_1\frac{N}{\rho^2\log(N-m)}$. As m increases, a larger number of permutations can be tolerated. (ii) In the absence of known correspondences, that is, when $n = N$, the requirement on s becomes $s \leq c_1\frac{N}{\rho^2\log N}$ which is consistent with that obtained in Lemma 1 of Nguyen & Tran (2012). (iii) The choice of λ follows from Theorem 1 in the next section.

Moreover, Nguyen & Tran (2012) prove that the family of Gaussian sensing matrices also satisfies the mutual incoherence condition, again with some suitable assumptions on (n, s, k). For the sake of completeness, we state the following lemma:

Lemma 2 (**Mutual incoherence condition Nguyen & Tran (2012)**). *Consider the Gaussian sensing matrix* $\boldsymbol{A}_2 \in \mathbb{R}^{n\times p}$ *whose rows are i.i.d.* $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. *We have the set* $\mathcal{C} = \{(\boldsymbol{h}, \boldsymbol{f}) \in (\mathbb{R}^p \times \mathbb{R}^n)$ *such that* $\|\boldsymbol{h}_{\boldsymbol{T}^C}\|_1 + \lambda\|\boldsymbol{f}_{\boldsymbol{S}^C}\|_1 \leq 3\|\boldsymbol{h}_{\boldsymbol{T}}\|_1 + 3\lambda\|\boldsymbol{f}_{\boldsymbol{S}}\|_1\}$ *with* $|T| = k$ *and* $|S| = s$ *as defined earlier. Select* $\lambda = \rho\sqrt{\frac{n}{N}\frac{\log n}{\xi(\boldsymbol{\Sigma})\log p}}$, *where* $\rho \in (0, 1)$ *is a constant. Assume that* $s \leq min\left(\frac{N}{n}\frac{\xi(\boldsymbol{\Sigma})}{C_{min}(\boldsymbol{\Sigma})}\frac{k\log p}{\rho^2\log n}, c_5\frac{C_{min}(\boldsymbol{\Sigma})}{C_{max}(\boldsymbol{\Sigma})}n\right)$ *and* $n \geq c_6\xi(\boldsymbol{\Sigma})\frac{C_{max}(\boldsymbol{\Sigma})}{C_{min}^2(\boldsymbol{\Sigma})}k\log p$ *for some sufficiently small positive constant* c_5 *and sufficiently large constant* c_6, *then the following inequality holds with probability greater than* $1 - \exp(-c_7n)$: $\frac{1}{\sqrt{n}}|\langle \boldsymbol{A}_2\boldsymbol{h}, \boldsymbol{f}\rangle| \leq k_m(\|\boldsymbol{h}\|_2 + \|\boldsymbol{f}\|_2)^2 \ \forall \ (\boldsymbol{h}, \boldsymbol{f}) \in \mathcal{C}$, *where* c_7, k_m *are positive constants. We refer to* k_m *as the mutual incoherence constant.*

We pick the above lemma from Nguyen & Tran (2012) with a slight change in the expression for λ to suit our problem. Note that the expression for λ naturally follows from the theoretical analysis of Ar-Lasso (see Theorem 1). Consequently, to obey the mutual incoherence condition, the requirement on s gets relaxed, that is, $s \leq \frac{N}{n}\frac{\xi(\boldsymbol{\Sigma})}{C_{\min}(\boldsymbol{\Sigma})}\frac{k\log p}{\rho^2\log n} = \frac{N}{N-m}\frac{\xi(\boldsymbol{\Sigma})}{C_{\min}(\boldsymbol{\Sigma})}\frac{k\log p}{\rho^2\log n}$. As m increases, more permutations can be tolerated. For a consistency check, setting $m = 0$ gives us the same upper limit on s as obtained in Lemma 2 in Nguyen & Tran (2012).

Now, since we have established that the family of Gaussian sensing matrices obey Gerec and Mic with a high probability, we move on to state the theoretical guarantees for Ar-Lasso. Given the optimization problem Eq. 5, we present an upper bound on the estimation error $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 + \|\tilde{\boldsymbol{e}} - \boldsymbol{e}^*\|_2$ in the following theorem:

Theorem 1. *Consider the optimization problem Eq. 5 and the observation model Eq. 1 under the assumptions (C1)-(C4). Select* $\lambda_\beta = \frac{2}{\rho}\frac{\|\boldsymbol{A^T w}\|_\infty}{N}$ *and* $\lambda_e = \frac{2\sqrt{n}}{N}\|\boldsymbol{w_2}\|_\infty$. *Assume that* $s \leq min\left(c_1\frac{N}{\rho^2 \log n}, \frac{N}{n}\frac{\xi(\boldsymbol{\Sigma})}{C_{min}(\boldsymbol{\Sigma})}\frac{k\log p}{\rho^2\log n}, c_5\frac{C_{min}(\boldsymbol{\Sigma})}{C_{max}(\boldsymbol{\Sigma})}n\right)$ *and that* $N \geq c_2\frac{\xi(\boldsymbol{\Sigma})}{C_{min}(\boldsymbol{\Sigma})}k\log p$, $n \geq c_6\xi(\boldsymbol{\Sigma})\frac{C_{max}(\boldsymbol{\Sigma})}{C^2_{min}(\boldsymbol{\Sigma})}k\log p$, *so that* $\boldsymbol{A}_2$ *satisfies* Mic *with sufficiently small mutual incoherence constant* k_m *such that* $k_m < min\left(\frac{N}{n}\frac{C_{min}(\boldsymbol{\Sigma})}{512}, \frac{1}{32}\right)$, *and* $\boldsymbol{A}$ *satisfies* Gerec *with constant* $k_e > 0$. *Then there exists positive constant* $k_l > 0$ *such that the following inequality holds with probability greater than* $1 - 2/p - 2/n$*:*

$$\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 + \|\tilde{\boldsymbol{e}} - \boldsymbol{e}^*\|_2 \leq 6\sigma k_l^{-2} max\left(\frac{1}{\rho}\sqrt{\frac{\xi(\boldsymbol{\Sigma})k\log p}{N}}, \sqrt{\frac{n}{N}\frac{s\log n}{N}}\right). \tag{6}$$

A few insights on the theorem are in order:

1. In the noiseless scenario ($\sigma = 0$), the error term turns out to be zero. And hence, exact recovery of $(\boldsymbol{\beta^*}, \boldsymbol{e^*})$ is possible with a high probability. In the case of noisy measurements, note that the estimation error scales linearly with the measurement noise standard deviation σ.
2. For fixed p, m and n, the error term scales as $\max(\sqrt{k}, \sqrt{s})$, i.e. sparser $\boldsymbol{\beta^*}$ and $\boldsymbol{e^*}$ yield more accurate reconstructions.
3. As pointed out earlier, the m known correspondences allow $\boldsymbol{A}$ to obey Gerec and Mic with a higher tolerance on s. These are sufficient conditions for error bounds for Ar-Lasso in Theorem 1 to hold. Moreover, note that the term $\sqrt{\frac{n}{N}\frac{s\log n}{N}}$ in Eq. 6 decreases as more correspondences are known. The higher the value of m is for a fixed n or a fixed N, the lower is the estimation error.
4. ρ is a positive constant that controls the sparsity level in the regression vector and sparse error vector. If there are a large number of permutations expected, then a larger value of ρ should be used for specifying λ_β and vice-versa.
5. The choice of λ in Lemma 1 and Lemma 2 follows from the expressions for λ_β and λ_e in Theorem 1. From the Gaussian concentration results, we know that $\|\boldsymbol{A^T w}\|_\infty \leq 2\sigma\sqrt{\xi(\boldsymbol{\Sigma})N\log p}$ with probability at least $1 - 2/p$ and $\|\boldsymbol{w_2}\|_\infty \leq 2\sigma\sqrt{\log n}$ with probability at least $1 - 2/n$. Plugging the expressions in $\lambda = \lambda_\beta/\lambda_e$ gives us $\lambda = \rho\sqrt{\frac{n}{N}\frac{\log n}{\xi(\boldsymbol{\Sigma})\log p}}$. On the other hand, Nguyen & Tran (2012) selects $\lambda = \rho\sqrt{\frac{\log N}{\xi(\boldsymbol{\Sigma})\log p}}$ which is same when we put $m = 0$ in our expression for λ.
6. When no correspondences are known in advance, that is, $m = 0$ and $n = N$, we get the same error bound as obtained in Corollary 1 in Nguyen & Tran (2012).
7. Our approach casts unlabeled sensing with sparse permutations in a structured sparsity framework. This theorem presents performance bounds for it. The particular form of structured sparsity considered here is different from forms such as tree-structured and block-structured sparsity as analyzed in previous works like Baraniuk et al. (2010).

4 Augmented Hard-Thresholding Pursuit

Note that under assumptions (C1)-(C4), $\boldsymbol{y}$ can be re-written as

$$\boldsymbol{y} = \boldsymbol{Hx}^* + \boldsymbol{w}, \text{ where } \boldsymbol{H} := \begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{0}_{m\times n}\\ \boldsymbol{A}_2 & \boldsymbol{I}_{n\times n}\end{bmatrix}$$

is the augmented matrix and $\boldsymbol{x}^* := \begin{bmatrix}\boldsymbol{\beta}^{*T} & \boldsymbol{z}^{*T}\end{bmatrix}^T$. Recalling the notation $\boldsymbol{y} := \begin{bmatrix}\boldsymbol{y}_1^T & \boldsymbol{y}_2^T\end{bmatrix}^T$ and $\boldsymbol{w} := \begin{bmatrix}\boldsymbol{w}_1^T & \boldsymbol{w}_2^T\end{bmatrix}^T$, the problem of estimation of $\boldsymbol{x}^*$ can be posed as

$$\underset{\substack{\boldsymbol{x}=[\,\boldsymbol{\beta}^T\ \boldsymbol{z}^T\,]^T\\ \boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{z}\in\mathbb{R}^n}}{\arg\min} \|\boldsymbol{y} - \boldsymbol{Hx}\|_2 \text{ s.t. } \|\boldsymbol{\beta}\|_0 \leq k, \|\boldsymbol{z}\|_0 \leq s. \tag{7}$$

Algorithm 1 Augmented Hard-Thresholding Pursuit

Input: Measurement vector $\boldsymbol{y}$, augmented matrix $\boldsymbol{H}$, sparsity level k and number of permutations s (both k and s can be estimated via cross-validation – see Sec. 5 under 'Choice of parameters')

Parameter: Learning rate μ

Output: Estimate of $\boldsymbol{\beta}^*$

1: $\boldsymbol{x}^{(0)} = \boldsymbol{0}, t = 0$.

2: **while** not converged **do**

3: $\quad \boldsymbol{l}^{(t+1)} = \boldsymbol{x}^{(t)} - \mu \boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{x}^{(t)} - \boldsymbol{y})$

4: $\quad S^{(t+1)} = \{$indices of k largest entries of $\boldsymbol{l}^{(t+1)}(1 : p)\} \cup \{$indices of s largest entries of $\boldsymbol{l}^{(t+1)}(p+1 : p+n)\}$

5: $\quad \boldsymbol{x}^{(t+1)} = \underset{\substack{\boldsymbol{x} \in \mathbb{R}^{p+n} \text{ s.t.} \\ \text{support}(\boldsymbol{x}) \subseteq S^{(t+1)}}}{\arg\min} \|\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}\|_2$

6: $\quad t = t + 1$

7: **end while**

8: **return** $\boldsymbol{x}^{(t)}(1 : p)$

We solve Eq. 7 using a modified form of the *hard thresholding pursuit* approach Foucart (2011) to suit the specific sparsity structure of our problem. The complete procedure is presented in Alg. 1. We refer to this approach as Augmented Hard Thresholding Pursuit or A-Htp, as it involves sparse recovery of an unknown vector (that is $\boldsymbol{x}^*$ here) using the *augmented* matrix $\boldsymbol{H}$. In Alg. 1, we note that (*i*) Step 2 is a standard gradient descent step where μ is the specified learning rate; (*ii*) Step 3 determines the support of $\boldsymbol{\beta}$ by selecting the indices corresponding to the k largest absolute-value entries in $\boldsymbol{x}(1 : p)$, and the support of $\boldsymbol{e}$ by selecting the indices corresponding to the s largest absolute-value entries in $\boldsymbol{x}(p+1 : p+n)$. The two support sets are combined to obtain the support set of $\boldsymbol{x}$; (*iii*) Step 5 is the debiasing step, and it re-estimates $\boldsymbol{x}$ over the support obtained in the previous step. This step involves computing the Moore-Penrose pseudo inverse of some column-submatrix of $\boldsymbol{H}$, which is computationally cheaper than the ℓ_1-norm optimization step required in Peng et al. (2021). Moreover, Peng et al. (2021) employs the subgradient method in their algorithm where the function value may even increase, whereas our A-htp is a gradient-descent-based method. In the next subsection, we provide a theoretical convergence analysis for A-Htp.

4.1 Convergence Analysis for A-Htp

In the compressed sensing framework, a condition for an accurate recovery of the unknown vector is typically based on the restricted isometry property (RIP) of the measurement matrix Candes & Tao (2005). A-Htp is inherently a compressed sensing algorithm. In this subsection, we state a theoretical guarantee on the recovery of $\boldsymbol{x}^*$ via A-Htp using some form of the restricted isometry constant (RIC) of the augmented measurement matrix $\boldsymbol{H}$. First, we recall the definition of RIC of a matrix. A matrix $\boldsymbol{A} \in \mathbb{R}^{N \times p}$ is said to satisfy the RIP of order t if there exists a constant $\delta_t \in (0, 1)$ such that

$$(1 - \delta_t)\|\boldsymbol{y}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{y}\|_2^2 \leq (1 + \delta_t)\|\boldsymbol{y}\|_2^2, \tag{8}$$

holds for every t-sparse $\boldsymbol{y} \in \mathbb{R}^p$. The smallest constant δ_t for which this holds is called the order-t RIC of the matrix $\boldsymbol{A}$. In other words, the application of $\boldsymbol{A}$ on any t-sparse vector approximately preserves the vector's norm.

Interestingly, in our problem, there is some structure in the distribution of non-zero entries in $\boldsymbol{x}^*$, that is, $\|\boldsymbol{x}^*(1 : p)\|_0 \leq k$ and $\|\boldsymbol{x}^*(p+1 : p+n)\|_0 \leq s$. To exploit this structured sparsity, we characterize a notion of structured-sparsity, restricted isometric property (SS-RIP) in the following definition.

Definition 1. *We say that the matrix $\boldsymbol{H}$ obeys the structured-sparsity, restricted isometric property (SS-RIP) of order $[(p,k),(n,s)]$ if there exists a restricted isometry constant $\delta := \delta^{SS}_{[(p,k),(n,s)]} \in (0,1)$ such that*

$$(1 - \delta)\|\boldsymbol{g}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{g}\|_2^2 \leq (1 + \delta)\|\boldsymbol{g}\|_2^2, \tag{9}$$

holds for every $(k+s)$-sparse $\boldsymbol{g} \in \mathbb{R}^{p+n}$ with $\|\boldsymbol{g}(1 : p)\|_0 \leq k$ and $\|\boldsymbol{g}(p+1 : p+n)\|_0 \leq s$.

With this, we state a geometric convergence result on A-HTP:

Theorem 2. *Consider the Gaussian sensing matrix $\boldsymbol{A}$ whose entries are i.i.d. $\mathcal{N}(0,1/N)$. Suppose that $\delta^{SS}_{[(p,3k),(n,3s)]}$ of the augmented measurement matrix $\boldsymbol{H}$ satisfies $\delta^{SS}_{[(p,3k),(n,3s)]} \leq \frac{1}{\sqrt{3}}$. Then the sequence $\boldsymbol{x}^{(t)}$ defined by A-HTP satisfies the following inequality $\forall\ t \geq 0$:*

$$\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^*\|_2 \leq \gamma^n \|\boldsymbol{x}^{(0)} - \boldsymbol{x}^*\|_2 + \tau \frac{1-\gamma^n}{1-\gamma} \|\boldsymbol{w}\|_2, \tag{10}$$

where $\tau := \dfrac{\sqrt{2\big(1-\delta^{SS}_{[(p,2k),(n,2s)]}\big)} + \sqrt{1+\delta^{SS}_{[(p,k),(n,s)]}}}{1-\delta^{SS}_{[(p,2k),(n,2s)]}}$ *and* $\gamma := \sqrt{\dfrac{2{\delta^{SS}_{[(p,3k),(n,3s)]}}^2}{1-{\delta^{SS}_{[(p,2k),(n,2s)]}}^2}} < 1.$

The proof of the above theorem follows along the lines of the work in Foucart (2011). The following lemma (proved in the supplemental material) outlines under what scenarios the augmented matrix $\boldsymbol{H}$ satisfies the criterion $\delta^{SS}_{[(p,3k),(n,3s)]} \leq \frac{1}{\sqrt{3}}$.

Lemma 3. *Consider the Gaussian sensing matrices $\boldsymbol{A}_1 \in \mathbb{R}^{m\times p}$ and $\boldsymbol{A_2} \in \mathbb{R}^{n\times p}$ with i.i.d. $\mathcal{N}(0,1/N)$ entries. There exist positive constants c_8, c_9 such that the augmented matrix $\boldsymbol{H} := \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{0}_{m\times n} \\ \boldsymbol{A}_2 & \boldsymbol{I}_{n\times n}\end{bmatrix}$ satisfies the structured-sparsity restricted isometry property (SS-RIP) of order $[(p,k),(n,s)]$ provided that $k\log(p/k) + s\log(n/s) \leq c_8 N$, with probability at least $1 - 3\exp(-c_9 N)$. The constants c_8, c_9 depend on the RIC δ. Equivalently, we have the following result:*

$$\mathbb{P}\bigg((1-\delta)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{H}\boldsymbol{x}\|_2^2 \leq (1+\delta)\|\boldsymbol{x}\|_2^2 \textit{ for all } \boldsymbol{x} \textit{ with } \|\boldsymbol{x}(1:p)\|_0 \leq k \textit{ and } \|\boldsymbol{x}(p+1:p+n)\|_0 \leq s\bigg) \geq 1 - 3\exp(-c_9 N). \tag{11}$$

We offer the following remarks:

1. From Lemma 3, we require that $N \geq C(k\log(p/3k) + s\log(n/3s))$ for $\boldsymbol{H}$ to obey SS-RIP of order $[(p,3k),(n,3s)]$. And hence, from Theorem 2, it follows that $N \geq C(k\log(p/3k) + s\log(n/3s))$ is sufficient for an accurate recovery of $\boldsymbol{x}^*$ via A-HTP.

2. Since $n = N - m$, the requirement $N \geq C(k\log(p/3k) + s\log(n/3s))$ relaxes as more correspondences are known.

3. In the scenario $n = 0$, that is, when all the correspondences are known, the problem simplifies to the standard compressed sensing problem. Consequently, we require $N \geq C(k\log(p/3k))$ measurements for an accurate recovery of $\boldsymbol{\beta}^*$, which is consistent with the well-known requirement for compressed sensing via HTP.

Comparison with the standard RIP condition. Note that if we do not exploit the specific sparsity structure in $\boldsymbol{x}^*$, a condition on an accurate recovery of $\boldsymbol{x}^*$ via A-HTP is that the matrix $\boldsymbol{H}$ should satisfy RIP of order $3(k+s)$ Foucart (2011), which requires $N \geq C(k+s)\log(p+n)/(3(k+s))$. On the other hand, by exploiting the sparsity structure in $\boldsymbol{x}^*$, we demonstrate that $N \geq C(k\log(p/3k) + s\log(n/3s))$ is sufficient for an accurate recovery of $\boldsymbol{x}^*$ via our A-HTP algorithm. This is less than $C(k+s)\log(p+n)/(3(k+s))$.

5 Numerical Experiments

In order to assess the impact made by knowledge of known correspondences, we compare AR-LASSO from Eq. 5 and A-HTP from Alg. 1 to the following estimators, none of which use the prior information of known correspondences.:

(*i*) The robust LASSO (R-LASSO) estimator given by $\arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2N}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta} - \sqrt{n}\boldsymbol{e}\|_2^2 + \lambda_\beta\|\boldsymbol{\beta}\|_1 + \lambda_e\|\boldsymbol{e}\|_1$, which is effectively AR-LASSO with $m = 0$.

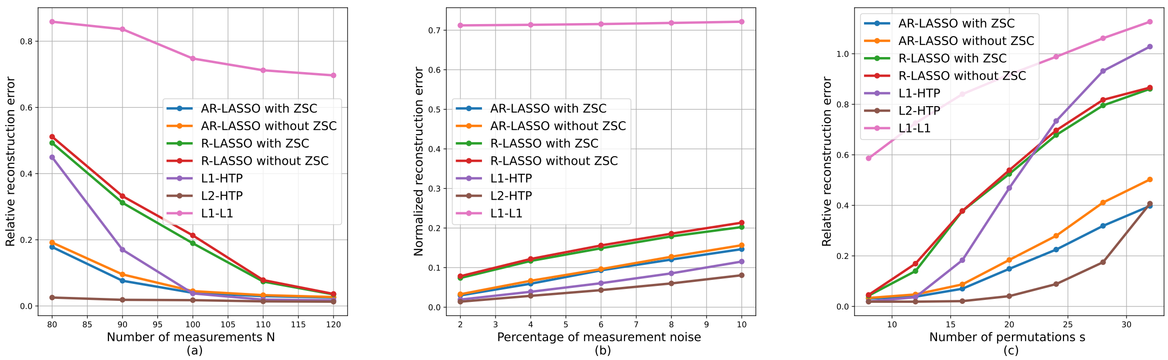

Figure 1: A comparison of the algorithms for unlabelled compressive sensing problem: The error plots as a function of (a) the number of measurements N, (b) the noise level in terms of f_r, and (c) the permutation level, show that using known correspondences as in ℓ_2-Htp and Ar-Lasso, results in a lower reconstruction error.

(*ii*) The ℓ_1-norm hard thresholding pursuit approach in Peng et al. (2021) that minimizes $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta}\|_1$ w.r.t. $\boldsymbol{\beta} \in \mathbb{R}^p$ such that $\|\boldsymbol{\beta}\|_0 \leq k$. We refer to this approach as ℓ_1-Htp.

(*iii*) The $\ell_1 - \ell_1$ estimator motivated from Candes & Tao (2005); Candes et al. (2005) that is posed as $\underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta}\|_1 + \lambda_\beta \|\boldsymbol{\beta}\|_1$.

We refer to our algorithm A-Htp henceforth as ℓ_2-Htp to distinguish it from ℓ_1-Htp. We use CVXPY Diamond & Boyd (2016) to solve all the optimization problems. Also, note that by definition of the concept of permutation, we have a zero-sum constraint (referred to as ZSC henceforth) on $\boldsymbol{e}$ given by $\sum_{i=1}^{n} e_i = 0$. It is worth investigating whether this information helps to improve the estimate of $\boldsymbol{\beta}^*$. To this end, we incorporate this additional hard constraint in the Ar-Lasso and R-Lasso estimators, yielding two additional variants.

Data generation: In all the experiments, the entries of $\boldsymbol{A}$ and the non-zero values of $\boldsymbol{\beta}^*$ are sampled from $\mathcal{N}(0,1)$. $\boldsymbol{P}_2$ is generated by randomly sampling from the family of s-sparse permutation matrices. The entries of $\boldsymbol{w}$ are independently sampled from $\mathcal{N}(0, \sigma^2)$ where $\sigma := f_r \times$ the mean absolute value of the entries of the noiseless measurement vector $\boldsymbol{P}\boldsymbol{A}\boldsymbol{\beta}^*$ with fraction $f_r \in (0,1)$.

Choice of parameters: The regularization parameters λ_β and λ_e in Ar-Lasso, Lasso and $\ell_1 - \ell_1$ algorithms are chosen through cross-validation on a held-out set of 5 measurements. Also note that ℓ_1-Htp and ℓ_2-Htp require the knowledge of (k, s), which are typically unknown in practice but can be chosen via cross-validation. In our experiments, we observe that cross-validation overestimates (k, s) by a factor of 2. Hence, we directly set (k, s) to twice of their true value in the ℓ_1-Htp and ℓ_2-Htp algorithms. We select the learning rate in ℓ_1-Htp and ℓ_2-Htp through cross-validation. The number of iterations in ℓ_1-Htp is set to 200, and that in ℓ_2-Htp is set to 100. We always observed convergence within these iteration counts.

Evaluation metric: The evaluation metric used for all simulations is the relative reconstruction error (RRMSE) $\frac{\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2}{\|\boldsymbol{\beta}^*\|_2}$, where $\tilde{\boldsymbol{\beta}}$ is an estimate of $\boldsymbol{\beta}^*$. We report the RRMSE averaged over 50 randomly chosen instances of $\boldsymbol{P}_2$, with each permutation instance averaged over 50 random instances of measurement noise $\boldsymbol{w}$.

Results: In Fig. 1(a), we have plotted the RRMSE as a function of the number of measurements N for $p = 240$, $k = 14$, $m = 32$, $s = 16$, and 2% noise (i.e. $f_r = 0.02$). We note that the approaches Ar-Lasso and ℓ_2-Htp, which utilize the information of known correspondences, result in lower errors. ℓ_2-Htp results in the least errors, followed by Ar-Lasso. $\ell_1 - \ell_1$ clearly produces much higher errors. Incorporation of ZSC in Ar-Lasso and R-Lasso only marginally improves the estimate of $\boldsymbol{\beta}^*$. As N increases, ℓ_1-Htp starts performing better than Ar-Lasso, indicating diminishing utility of known correspondences at a sufficiently large N.

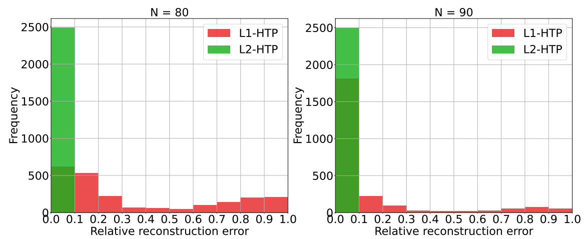

Figure 2: Distribution of the RRMSE values obtained using ℓ_1-HTP and ℓ_2-HTP for different instances of permutation matrix and measurement noise. Utilizing the prior known correspondences results in a much lower standard deviation of the errors.

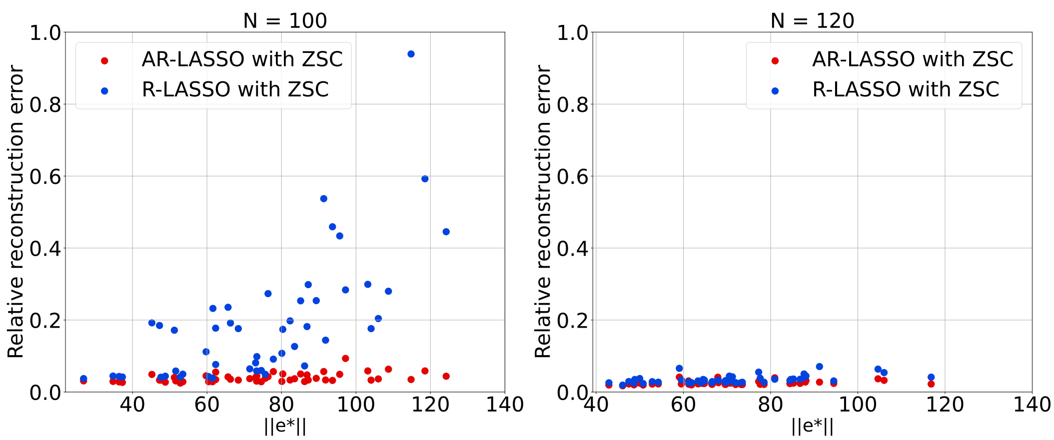

Figure 3: Scatter plots of RRMSE and permutation error magnitude $\|\boldsymbol{e}^*\|_1$ for $N = 100$ (left) and $N = 120$ (right) measurements. For $N = 100$, note the strong correlation of the estimation error with $\|\boldsymbol{e}^*\|_1$ when known correspondences are not utilized as in R-Lasso.

In Fig. 2, histograms of the RRMSE values are displayed for ℓ_1-HTP and ℓ_2-HTP. Note that the estimation error depends on the specific (randomly chosen) instance of $\boldsymbol{P}_2$. We observe that ℓ_1-HTP results in large errors many times. Moreover, we note that by using known correspondences in ℓ_2-HTP, the variance of the RRMSE distribution significantly decreases for a given N. This is particularly prominent given fewer measurements, as seen in the case of $N \in \{80, 90\}$.

In Fig. 3, we display the scatter plot of the errors obtained on the randomly chosen permutation matrices averaged over 50 random noise instances as a function of $\|\boldsymbol{e}^*\|_1$ for Ar-Lasso and R-Lasso. Observe that for $N = 100$, the term $\|\boldsymbol{y}_1 - \boldsymbol{A}_1\boldsymbol{\beta}\|_2^2$ in Ar-Lasso ensures that the errors do not depend much on the permutation error magnitude $\|\boldsymbol{e}^*\|_1$. On the other hand, without the information of known correspondences in R-Lasso, the scatter plot shows a clear positive correlation with $\|\boldsymbol{e}^*\|_1$. This correlation, however, dies down as we get more measurements, as seen in the case $N = 120$.

Next, in Fig. 1(b), we compare the algorithms for different noise levels (i.e., different values of f_r) for $p = 240$, $N = 110$, $k = 14$, $m = 32$, and $s = 16$. We observe that for the Lasso-based approaches and ℓ_2-HTP, the estimation error scales linearly with σ as we expect from the upper bounds in Theorem 1 and Theorem 2.

Lastly, in Fig. 1(c), we show errors for different permutation levels for $p = 240$, $N = 90$, $k = 14$, $m = 32$ and 2% measurement noise. Recall that the error upper bound for Ar-Lasso scales as $\sqrt{s}$. We note that the algorithms that utilize known correspondences are more robust in terms of the number of permutations in the measurements.

Average compute time values, and a practical application of our algorithms for unlabelled compressive sensing with known correspondences to image registration is presented in the supplemental material.

6 Conclusion

We studied the problem of unlabelled compressive sensing with the assumption that the regression vector is sparse and given additional knowledge of a few correspondences. We proposed a tractable Lasso-based estimator and derived theoretical performance bounds for our algorithm. We also presented another estimator based on a modified form of Hard Thresholding Pursuit, with theoretical analysis. We verified the theoretical findings through numerical experiments. We compared our algorithm with a hard thresholding approach and an ℓ_1 norm formulation and demonstrated that our algorithm outperforms them. Additionally, we illustrated that having information about a small number of accurate correspondences reduces the sensitivity of estimation error on the severity of permutation corruption. Lastly, we demonstrated a practical application of our framework in image registration.

References

Richard G Baraniuk, Volkan Cevher, Marco F Duarte, and Chinmay Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010.

E.J. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, 2005. doi: 10.1109/TIT.2005.858979.

Emmanuel Candes, Mark Rudelson, Terence Tao, and Roman Vershynin. Error correction via linear programming. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pp. 668–681. IEEE, 2005.

Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *J. Mach. Learn. Res.*, 17(83):1–5, 2016.

Simon Foucart. Hard thresholding pursuit: An algorithm for compressive sensing. *SIAM J. Numer. Anal.*, 49(6):2543–2563, 2011. doi: 10.1137/100806278.

Deguang Han, Fusheng Lv, and Wenchang Sun. Recovery of signals from unordered partial frame coefficients. *Appl. Comput. Harmon. Anal.*, 44(1):38–58, 2018. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2016.04.002.

Daniel J Hsu, Kevin Shi, and Xiaorui Sun. Linear regression without correspondence. In *Adv. Neural Inf. Process. Syst.*, volume 30. Curran Associates, Inc., 2017.

Jerin Geo James, Pranay Agrawal, and Ajit Rajwade. Restoration of non-rigidly distorted underwater images using a combination of compressive sensing and local polynomial image representations. In *IEEE/CVF Int. Conf. Comput. Vis.*, 2019.

P. Lahiri and Michael D. Larsen. Regression analysis with linked data. *J. Am. Stat. Assoc.*, 100(469): 222–230, 2005. ISSN 01621459.

Feiran Li, Kent Fujiwara, Fumio Okura, and Yasuyuki Matsushita. Generalized shuffled linear regression. In *IEEE/CVF Int. Conf. Comput. Vis.*, pp. 6454–6463, 2021. doi: 10.1109/ICCV48922.2021.00641.

David G Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60:91–110, 2004.

Manuel Marques, Marko Stošić, and João Costeira. Subspace matching: Unique solution to point matching with geometric constraints. In *IEEE/CVF Int. Conf. Comput. Vis.*, pp. 1288–1294, 2009. doi: 10.1109/ICCV.2009.5459318.

Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *IEEE Symp. Secur. Priv.*, pp. 111–125, 2008. doi: 10.1109/SP.2008.33.

Nam H Nguyen and Trac D Tran. Robust LASSO with missing and grossly corrupted observations. *IEEE Trans. Inf. Theory*, 59(4):2036–2058, 2012.

Jun-Jun Pan and Yan-Ning Zhang. A new point matching method based on position similarity. In *Int. Conf. Mach. Learn. Cybern.*, pp. 5154–5158 Vol. 8, 2005. doi: 10.1109/ICMLC.2005.1527852.

Ashwin Pananjady, Martin J. Wainwright, and Thomas A. Courtade. Linear regression with shuffled data: Statistical and computational limits of permutation recovery. *IEEE Trans. Inf. Theory*, 64(5):3286–3300, 2018. doi: 10.1109/TIT.2017.2776217.

Liangzu Peng and Manolis C. Tsakiris. Linear regression without correspondences via concave minimization. *IEEE Signal Process. Lett.*, 27:1580–1584, 2020. doi: 10.1109/LSP.2020.3019693.

Liangzu Peng, Boshi Wang, and Manolis Tsakiris. Homomorphic sensing: Sparsity and noise. In *Int. Conf. Mach. Learn.*, volume 139 of *Proc. Mach. Learn. Res.*, pp. 8464–8475. PMLR, 2021.

Martin Slawski and Emanuel Ben-David. Linear regression with sparsely permuted data. *Electron. J. Stat.*, 13(1):1 – 36, 2019. doi: 10.1214/18-EJS1498.

Garweet Sresth, Ajit Rajwade, and Satish Mulleti. Unlabelled sensing with priors: Algorithm and bounds. In *IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 9551–9555, 2024. doi: 10.1109/ICASSP48485.2024.10445886.

Sebastian Thrun and John J. Leonard. *Simultaneous Localization and Mapping*, pp. 871–889. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-30301-5. doi: 10.1007/978-3-540-30301-5_38.

Manolis Tsakiris and Liangzu Peng. Homomorphic sensing. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Int. Conf. Mach. Learn.*, volume 97, pp. 6335–6344. PMLR, 09–15 Jun 2019.

Jayakrishnan Unnikrishnan, Saeid Haghighatshoar, and Martin Vetterli. Unlabeled sensing with random linear measurements. *IEEE Trans. Inf. Theory*, 64(5):3237–3253, 2018. doi: 10.1109/TIT.2018.2809002.

A Appendix

Please refer to supplemental material for additional experiments and the proofs of lemmas and theorems.