Robot kinematic structure classification from time series of visual data

Alberto Dalla Libera¹, Matteo Terzi², Rossi Alessandro¹, Gian Antonio Susto¹ and Ruggero Carli¹

Abstract—In this paper we present a novel algorithm to solve the robot kinematic structure identification problem. Given a time series of data, typically obtained processing a set of visual observations, the proposed approach identifies the ordered sequence of links associated to the kinematic chain, the joint type interconnecting each couple of consecutive links, and the input signal influencing the relative motion. Compared to the state of the art, the proposed algorithm has reduced computational costs, and is able to identify also the joints’ type sequence.

I. INTRODUCTION

In recent years, there has been an increasing interest toward Modular Robotics and reconfigurable and adaptable robots have started to be designed [1], [2], [3]. In particular, reconfigurability and modularity can be exploited to build robots with greater adaptability to several different environments, as well as robots able to accomplish different tasks, obtaining important cost and time reductions.

On the other hand, Modular Robotics introduces new challenging issues. One of these is the need to handle robots with variable kinematic structures, where this variability might result in a partially or complete lack of knowledge about the kinematic structure. It is worth stressing that the forward and inverse kinematics are fundamental in robotics applications; remarkable examples are motion planning [4], robot modeling and control [5].

When dealing with standard robots, the prior knowledge about the robot geometry is extremely precise, since most of the times CAD models are available and direct measurements of the robot parameters are possible. Kinematic models relate joint input signals and robot configurations. Typically they are computed as a sequence of relative transformations between reference frames properly assigned [5]. Unavoidable inaccuracies in the geometrical parameters, wear and systematic errors in the measurements, make sometimes calibration procedures necessary [6].

In the Modular Robotics context, where the uncertainty about the robot geometry might be particularly high, the development of algorithms able to estimate the kinematic structure starting from a time series of visual observations is crucial. The kinematic structure identification problem is defined at different levels, depending on the sensors used and the amount of prior information available. When there is no prior knowledge, firstly it is necessary to identify the rigid bodies composing the robot, and extract information about their poses. Secondly, starting from this piece of information, the robot kinematic is learned identifying the ordered sequence of links, the type of the joint connecting any pair of consecutive links and the corresponding input signal.

The first level is strictly related to the kind of sensors used in the data collection. For example, in [7] point cloud data have been considered. More precisely, the authors have proposed to identify the links by clustering points based on their relative distances, and by assuming that each cluster corresponds to one link. The task is much more complicated when observations come from a standard 2D camera. Indeed in this case the clustering phase should be preceded by the identification of features which are constant over the observations.

In this paper, we consider the setup introduced in [8], [9] and [10], where a distinct fiducial marker is attached to each link. This assumption, that simplifies considerably the data acquisition, is particularly reasonable in the Modular Robotics context.

The focus of this paper is on the second level of the kinematic structure reconstruction. Specifically, starting from a time series containing the marker poses and the joint signals, we propose an algorithm able to reconstruct the ordered markers sequence associated to the robot kinematic chain, together with the joint types connecting subsequent links and the corresponding joint input signals.

Similar problems have been treated in [8] and [10]. In [8], the authors have restricted the scope of their work to the case of only revolute joints, and the learning of the kinematic parameters is obtained by a gradient-based minimization procedure. Moreover, numerical results highlight how the convergence is guaranteed only when each link has a marker attached to. In [10] the authors have proposed the use of Gaussian Processes (GP) methodologies. However, the focus is limited to learn only the sequence of markers, and no joint type identification has been considered. Moreover, even if the authors have proposed a strategy to simplify the standard GP procedure, the algorithm they developed might be quite expensive from a computational point of view, in particular when dealing with manipulator with a relevant number of degrees of freedom.

The kinematic structure identification algorithm we introduce is based on checking the feasibility of three systems of equations, which are obtained starting from elementary kinematic relations between pairs of subsequent links and using information extracted from time series of visual data. More precisely, given a couple of markers attached to subse-
quent links and the corresponding joint input signal, a linear system of equations holds true if the three elements define a prismatic transformation, instead a linear and a non linear systems are satisfied if the transformation is revolute. In general, it is possible to exhibit sets of observations for which systems of equations hold true though the pair of markers and the joint signal considered are not in relation among them. However, by extensive Monte Carlo simulations we show that this false-positive fact is very unlikely to appear. To ensure that the feasibility of the introduced systems of equations is a necessary and sufficient condition to verify if two markers are attached to consecutive links, we need to apply our strategy with data obtained from fully informative sets of observations; in the paper we exhibit a class of trajectories from which it is possible to properly select observation sets which are fully informative.

Compared to the state of the art, the proposed approach is less expensive as regards the computational costs, since it is based on the solution of linear and non linear systems of equations with low dimensionality. Moreover, differently from [10], the proposed algorithm fully reconstructs the kinematic structure, including the joint type sequence.

The paper is organized as follows. In the next Section we describe the considered setup, and in Section [II] we derive the three systems of equations used in the kinematic identification. The fully informative trajectories are reported in Section [III] while the proposed algorithm and the Monte Carlo simulations are in Section [IV].

II. SETUP DESCRIPTION AND PROBLEM FORMULATION

In this section we formally describe the setup and problem considered in this work.

The framework is the same one adopted in [10], and it consists in a camera and a robotic arm, composed by \( n \) links and \( n - 1 \) joints, forming an open kinematic chain. As far as the camera placement is concerned, we adopted an eye-to-hand configuration, namely the camera is observing the robot and its pose is fixed with respect to the world reference frame (RF). For a pictorial description of the overall setup see Figure 1.

We assume that each robot link \( L_j \), \( j = 1, \ldots, n \), has a fiducial marker \( M_i \), \( i = 1, \ldots, n \), attached to it, see for instance, [11], [12] and [13]. The \( j \) subscript denotes the link position in the kinematic chain, in particular \( L_1 \) and \( L_n \) correspond, respectively, to the base link and the last link.

The marker-link relations are unknowns and, for this reason we use different subscripts to enumerate markers and links. More formally, referring to the notation above introduced, given the \( i \)-th marker \( M_i \), the link \( L_j \) at which \( M_i \) is attached to is unknown.

The relative poses between the world RF and each marker are obtained processing each single frame coming from the camera [11], [12], [13]. Specifically, let \( p_{M_i}^c(t) \) be the pose of marker \( M_i \) at time \( t \), denoted with the respect to the camera RF (hereafter indicated by the superscript \( c \)); then

\[
\begin{bmatrix}
I_{M_i}^c(t) \\
o_{M_i}^c(t)
\end{bmatrix}^T
\]

where \( I_{M_i}^c \) denotes the position vector, and it is composed by the three Cartesian coordinates \( x_{M_i}, y_{M_i} \) and \( z_{M_i} \), while, as far as the orientation \( o_{M_i}^c \) is concerned, the yaw-pitch-roll convention is adopted, where \( \gamma_{M_i}(t), \beta_{M_i}(t) \) and \( \alpha_{M_i}(t) \) are respectively the yaw, pitch and roll angles.

For computational reasons, it is convenient to express the relative orientation between \( M_i \) and the camera using the rotation matrix \( R_{M_i}^c \), that is related to the yaw, pitch and roll angles by the standard expression \( R_{M_i}^c = R_z(\gamma_{M_i})R_y(\beta_{M_i})R_x(\alpha_{M_i}) \). Similar definitions hold for \( p_{L_j}^c \), which denotes the pose of the \( j \)-th link.

Then, by processing the frame taken by the camera at time instant \( t \), it is possible to reconstruct the set of poses \( D_p(t) = \{ p_{M_1}^c(t), \ldots, p_{M_n}^c(t) \} \).

In addition, we assume that the vector of the joint configurations \( q(t) = [q_1(t), \ldots, q_{n-1}(t)] \) is available at time \( t \); specifically, \( q_k(t) \) parametrizes the relative displacement between the two links connected by the \( k \)-th joint. However, we assume that also the relations joint-links are unknown, namely, for the \( k \)-th joint we do not know which pair of links \( L_{j_1} \) and \( L_{j_2} \) is connected by joint \( k \). For this reason, similarly to what done when considering the markers, we use different subscripts to denote joints and links.

The main goal of this paper is that of identifying the robot kinematic structure starting from a time series of measurements \( X = \{ (D_p(t_1), q(t_1)), \ldots, (D_p(t_T), q(t_T)) \} \), composed by the joints configurations and the marker poses.

From now on, to keep the notation compact, we point out explicitly the dependencies on time only when it is necessary.

The identification of the robot kinematic structure can be decomposed into three subtasks:

- Identifying \( S_M = \{ M_1, \ldots, M_n \} \), i.e., the sequence of markers associated to the kinematic chain \( L_1, \ldots, L_n \);
- Identifying \( S_Q = \{ Q_k \}, \ldots, Q_{k_{n-1}} \) \), i.e., the sequence of joint types connecting consecutive links along the kinematic chain, starting from the pair \( (L_1, L_2) \), up to the pair \( (L_{n-1}, L_n) \); more precisely \( Q_k \) is a binary variable assuming value 0 (resp. value 1) when the joint connecting \( L_j \) and \( L_{j+1} \) is prismatic (resp. revolute);
- Identifying \( S_q = \{ q_k, \ldots, q_{k_{n-1}} \} \), i.e., the sequence of joint signals that parametrize the corresponding transformations in \( S_Q \).

III. RELATIONS BETWEEN COUPLE OF SUBSEQUENT MARKERS

In this section we provide useful expressions that describe the relative motion between markers which are attached to consecutive links; in particular we will distinguish the case where the joint connecting the links is prismatic from the case where it is revolute.

More formally, let \( M_{i_1} \) and \( M_{i_2} \) be the two markers, and let \( L_{j_1} \) and \( L_{j_2} \) be the links they are attached to. Supposing \( j_1 = j_2 - 1 \) (similar considerations hold for \( j_1 = j_2 + 1 \)), we will assign a RF to each marker and to each link and, based on these RFs, we will provide a mathematical description of the relative motion between \( M_{i_1} \) and \( M_{i_2} \). In
particular, given a set of observations $X$, we will identify three systems of equations, that will be exploited in the next sections, to discriminate if the links associated to $M_l$ and $M_k$ are directly connected or not. Additionally, if $L_{j_1}$ and $L_{j_2}$ are connected, the systems will uniquely identify the joint type connecting the two links and also the joint variable $\theta_{j}$ describing the relative displacement.

A. Reference frames conventions

The definition of a RF for each link and for each maker is required to provide a mathematical description of the transformations occurring along the kinematic chain. As far as the links are concerned, we adopt the Denavit-Hartenberg (DH) convention; for details we refer the interested reader to [5], chapter 2.8.2. Once the RFs of the links have been assigned, the expression of $R_{L_{j-1}L_{j}}$, i.e., the relative orientation between the consecutive links $L_{j-1}$ and $L_{j}$, is given by

$$R_{L_{j-1}L_{j}} = R_{x}(\theta_{j})R_{z}(\alpha_{j}),$$

being $R_{x}$ and $R_{z}$ the elementary rotation matrices around the $z$-axis and the $x$-axis, respectively, and $\alpha_{j}$ a constant parameter (see [5]). In case the joint connecting $L_{j-1}$ and $L_{j}$ is prismatic, then $\theta_{j}$ is constant and equal to $\theta_{0}^{j}$, while, if the joint is revolute and controlled by $q_{k}$, it holds $\theta_{j} = \theta_{0}^{j} + q_{k}$. The relation between the relative positions of $L_{j-1}$ and $L_{j}$ is described by $t_{L_{j-1}L_{j}}^{L_{j}}$, i.e., the expression of the origin of $L_{j}$ with respect to $L_{j-1}$, and it is given by

$$t_{L_{j-1}L_{j}}^{L_{j}} = \begin{bmatrix} 0 \\ 0 \\ d_{j} \end{bmatrix} + R_{z}(\theta_{j}) \begin{bmatrix} a_{j} \\ 0 \\ 0 \end{bmatrix},$$

where $R_{z}(\theta_{j})$ is defined as before and $a_{j}$ is a constant parameter of the kinematic (see [5]). If the joint connecting $L_{j-1}$ and $L_{j}$ is revolute then $d_{j}$ is constant and equal to $d_{0}^{j}$, while, if it is prismatic and parametrized by $q_{k}$, then it holds $d_{j} = d_{0}^{j} + q_{k}$.

Additionally, we need to define position and orientation of the reference frame of each marker with the respect to the reference frame of the link they are attached to. For example, suppose that $M_{l}$ is attached to $L_{j_1}$, then position and orientation of $M_{l}$ w.r.t. $L_{j_1}$ are described, respectively, by $t_{M_{l}L_{j_1}}^{L_{j_1}}$ and $R_{M_{l}L_{j_1}}^{L_{j_1}}$. For later use it is convenient to introduce also $R_{M_{l}L_{j_1}}^{L_{j_1}} = (R_{M_{l}L_{j_1}}^{L_{j_1}})^T$ and $I_{M_{l}L_{j_1}}^{L_{j_1}}$. Similar definitions hold for $M_{k}$ and $L_{j_2}$.

Since the marker-link transformations are fixed, $t_{M_{l}M_{k}}^{L_{j_1}}$ and $R_{M_{l}M_{k}}^{L_{j_1}}$ are independent from the joint values $q$ and constant over the time.

It is worth stressing that $t_{M_{l}M_{k}}^{L_{j_1}}$ and $R_{M_{l}M_{k}}^{L_{j_1}}$ are unknown, and we did not introduce any limitation on the way the markers are attached to links. From a practical point of view, this fact is very interesting, since it allows to adopt the proposed algorithm even in setups different from the one we described in Section II. Indeed, it might happen that the markers are not available and the use of ad-hoc computer-vision algorithms is required to get information about the robot displacements [14], [15], [16]. In this context the markers placement is not controllable, but it still holds that $t_{M_{l}M_{k}}^{L_{j_1}}$ and $R_{M_{l}M_{k}}^{L_{j_1}}$ are constant.

Let $t_{M_{l}M_{k}}^{L_{j_1}}$ and $R_{M_{l}M_{k}}^{L_{j_1}}$ be, respectively, the Cartesian coordinates of the origin of $M_{l}$ w.r.t. $M_{k}$, and the relative orientation between $M_{l}$ and $M_{k}$. Assuming that $L_{j_1}$ and $L_{j_2}$ are subsequent, i.e., $j_2 = j_1 + 1$, $t_{M_{l}M_{k}}^{L_{j_1}L_{j_2}}$ is given by

$$t_{M_{l}M_{k}}^{L_{j_1}L_{j_2}} = t_{M_{l}M_{k}}^{L_{j_1}} + R_{L_{j_1}L_{j_2}}^{L_{j_1}} t_{M_{l}L_{j_2}}^{L_{j_2}} + R_{M_{l}M_{k}}^{L_{j_1}} ( - t_{L_{j_2}}^{M_{l}} ).$$

Moreover, exploiting standard properties of rotation matrices, the following equation holds

$$R_{M_{l}M_{k}}^{L_{j_1}L_{j_2}} = R_{L_{j_1}L_{j_2}}^{L_{j_1}} R_{M_{l}L_{j_2}}^{L_{j_2}} (\theta_{j_2}) R_{L_{j_1}L_{j_2}}^{L_{j_1}}.$$

It is worth remarking that, in the described setup, we have knowledge of $t_{M_{l}M_{k}}^{L_{j_1}}$, since $t_{M_{l}M_{k}}^{L_{j_1}} = ( t_{M_{l}M_{k}}^{L_{j_1}} )^T ( t_{M_{l}M_{k}}^{L_{j_1}} - t_{M_{l}L_{j_1}}^{L_{j_1}} )$, where $t_{M_{l}M_{k}}^{L_{j_1}}$, $t_{M_{l}L_{j_1}}^{L_{j_1}}$, and $t_{M_{l}M_{k}}^{L_{j_1}}$ are obtained processing the information coming from the camera. Moreover, also $R_{M_{l}M_{k}}^{L_{j_1}} = ( R_{M_{l}M_{k}}^{L_{j_1}} )^T R_{M_{l}M_{k}}^{L_{j_1}}$ is known, since by definition it is a function of the camera observations $R_{M_{l}M_{k}}^{L_{j_1}}$.

In the remaining part of this Section we further investigate the above relations, distinguishing the case where the joint connecting two successive links is prismatic from the case where the joint is revolute.

B. Prismatic joint

Assume that the joint connecting $L_{j_1}$ and $L_{j_2}$ is prismatic. Since in this case the angle $\theta_{j}$ is constant, it follows that the relative orientation between $L_{j_1}$ and $L_{j_2}$ is not affected by variations of the joint variable, that is, the matrix $R_{M_{l}M_{k}}^{L_{j_1}L_{j_2}}$ is also constant over the time. In addition, assume that $q_{l}$ is the joint variable associated to the prismatic joint connecting $L_{j_1}$ and $L_{j_2}$. By substituting the expression of $t_{M_{l}M_{k}}^{L_{j_1}L_{j_2}}$ given in

![Fig. 1. Symbolic representation of the setup.](Image)
where the first three terms of the last equation are constant and they can be compacted in the vector $l_{i_1,i_2}^t$, while the last term depends on the joint coordinate $q_k$.

Observe that the last equation defines a system of equations which are linear w.r.t. $l_{i_1,i_2}^t$ and the third column of $R_{L_{i_1},i_1}^M$ that we denote hereafter as $Z (R_{L_{i_1},i_1}^M)$. Specifically we can write

$$l_{M_2}^{M_1} = \begin{bmatrix} I_3 & q_k I_3 \\ Z & R_{L_{i_1},i_1}^M \end{bmatrix} = A(q_k) b_{i_1,i_2},$$

where $A(q_k) = \begin{bmatrix} I_3 & q_k I_3 \end{bmatrix}$ and $b_{i_1,i_2} = \begin{bmatrix} l_{i_1,i_2}^t \end{bmatrix}^T Z (R_{L_{i_1},i_1}^M)^T$ subject to the constraint

$$Z (R_{L_{i_1},i_1}^M)^T Z (R_{L_{j_1},j_1}^M) = 1. \tag{7}$$

Note that Equation $\tag{6}$ holds when we have only one observation. In general, if the set of observations $X$ has cardinality $T$, Equation $\tag{6}$ is replaced by the following linear system of $3T$ equations

$$\begin{bmatrix} l_{M_2}^{M_1}(t_1) \\ \vdots \\ l_{M_2}^{M_1}(t_T) \end{bmatrix} = \begin{bmatrix} A(q_k(t_1)) \\ \vdots \\ A(q_k(t_T)) \end{bmatrix} b_{i_1,i_2} = A(X) b_{i_1,i_2}. \tag{8}$$

We have the following Proposition.

**Proposition 1:** Consider two markers $M_{i_1}$ and $M_{i_2}$, attached to consecutive links connected through a revolute joint. Let $q_k$ be the joint signal influencing the relative motion between the two links. Then, given a set of observations $X$, the rotation matrix $R_{M_2}^{M_1}$ is constant and the linear system of equations in $\tag{8}$ has solution satisfying the constraint in $\tag{7}$.

**C. Revolute joint**

Now assume that the joint connecting $L_{j_1}$ and $L_{j_2}$ is revolute. Expressing $l_{L_{j_2}}^{L_{j_1}}$ as in $\tag{3}$, we have that equation

$$r_{M_2}^{M_1} = l_{L_{j_1}}^{M_1} + R_{L_{j_1}}^M (\begin{bmatrix} 0 \\ 0 \\ q_k \end{bmatrix} + R_{(\theta_j)}^0 \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}) + R_{L_{j_2}}^M (-l_{L_{j_2}}^{M_2}) = l_{L_{j_1}}^{M_1} + R_{L_{j_1}}^M (\begin{bmatrix} 0 \\ 0 \\ q_k \end{bmatrix} + R_{(\theta_j)}^0 \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}) + R_{L_{j_2}}^M (-l_{L_{j_2}}^{M_2}),$$

which is linear w.r.t. the vector of variables

$$\bar{b}_{i_1,i_2} = \begin{bmatrix} l_{M_2}^{M_1} + R_{L_{j_1}}^M \begin{bmatrix} a_{j_2} \\ 0 \\ 0 \end{bmatrix} \\ R_{L_{j_2}}^M \begin{bmatrix} a_{j_2} \\ 0 \\ 0 \end{bmatrix} - l_{L_{j_2}}^{M_2} \end{bmatrix} \tag{9}$$

since Equation $\tag{9}$ can be rearranged as

$$l_{M_2}^{M_1} (X) = \begin{bmatrix} l_{M_2}^{M_1}(t_1) \\ \vdots \\ l_{M_2}^{M_1}(t_T) \end{bmatrix} = \begin{bmatrix} A(D_{X}) b_{i_1,i_2} \end{bmatrix} = \begin{bmatrix} A(D_{X}) \bar{b}_{i_1,i_2} \end{bmatrix}. \tag{10}$$

We have the following Proposition.

**Proposition 2:** Consider two markers $M_{i_1}$ and $M_{i_2}$, attached to consecutive links connected through a revolute joint. Then, given a set of observations $X$, the linear system of equation $\tag{10}$ has solution.

It is worth observing that the dependance of $\tag{10}$ on the revolute joint signal is not explicit since it is incorporated into the evolution of the matrix $R_{M_2}^{M_1}$. To directly consider the effects of varying the joint signal on the relative motion between $M_{i_1}$ and $M_{i_2}$, we analyze their relative orientation. Differently than in the prismatic case, $R_{M_2}^{M_1}$ is not constrained to be constant. Let $q_k$ be the action signal of the revolute joint between $L_{j_1}$ and $L_{j_2}$.

Rewriting $R_{L_{j_2}}^L$ according to $\tag{11}$, we have that Equation $\tag{4}$ can be rewritten as

$$R_{M_2}^{M_1} = R_{L_{j_1}}^L q_k + R_{x}(\theta_j) R_{x}(\alpha_j) R_{j_2}^L,$$

where $R_{M_2}^{L_{j_2}} = R_{x}(\theta_j) R_{x}(\alpha_j) R_{j_2}^L$. Analyzing $\tag{11}$ element-wise, we can identify a system of nine equations where the unknowns are the elements of $R_{L_{j_1}}^L$. 

\( R_{M_i}^{L_j} \), while the output is given by the elements of \( R_{M_j}^{M_i} \).

Now, let \( Vec(\cdot) \) be the operator that maps a \( N \times N' \)-dimensional matrix \( M \) into the \( N^2 \)-dimensional column vector obtained by stacking the columns of the matrix \( M \) on top of one another, then we have that \( Vec \left( R_{M_i}^{M_j} \right) = Vec \left( R_{L_j}^{M_i} R_2(q_k) R_{L_j}^{M_i} \right) \).

In addition observe that the unknown variables must satisfy the following orthogonality constraints

\[
Vec \left( R_{L_j}^{M_i} \left( R_{L_j}^{M_i} \right)^T \right) = Vec \left( I_3 \right) \tag{11}
\]

\[
Vec \left( R_{M_i}^{L_j} \left( R_{M_i}^{L_j} \right)^T \right) = Vec \left( I_3 \right). \tag{12}
\]

It is worth stressing that the aforementioned equations are non linear w.r.t the variables \( Vec \left( R_{L_j}^{M_i} \right) \) and \( Vec \left( R_{M_i}^{L_j} \right) \).

If, instead of considering a single observation we consider a set of observations \( X \), the equations describing the relative orientations between \( M_{i_1} \) and \( M_{i_2} \) are

\[
\begin{bmatrix}
Vec \left( R_{M_{i_1}}^{L_j} (t_1) \right) \\
\vdots \\
Vec \left( R_{M_{i_2}}^{L_j} (t_T) \right)
\end{bmatrix}
= \begin{bmatrix}
Vec \left( R_{L_j}^{M_{i_1}} R_2 (q_k(t_1)) R_{M_{i_2}}^{L_j} \right) \\
\vdots \\
Vec \left( R_{L_j}^{M_{i_1}} R_2 (q_k(t_T)) R_{M_{i_2}}^{L_j} \right)
\end{bmatrix}. \tag{13}
\]

We have the following proposition.

**Proposition 3:** Consider two markers \( M_{i_1} \) and \( M_{i_2} \), attached to consecutive links connected through a revolute joint. Let \( q_k \) be the actuation signal of the joint. Then, given a set of observations \( X \), the system of non-linear equations in \((13)\) admits a solution satisfying the non-linear orthogonality constraints in \((12)\).

### IV. FULLY INFORMATIVE TRAJECTORIES

In the previous Section we have stated three propositions defining conditions that are verified when \( M_{i_1} \) and \( M_{i_2} \) are attached to consecutive links. In general the reverse relations are not true, since it is possible to exhibit sets of observations \( X \) such that the conditions of the previous Propositions are satisfied though the markers are not attached to subsequent links; this fact is strictly related to the sequence of joint configurations which have generated the taken observations. Unfortunately, due to space constraints, we do not include in this paper examples of such false positive observation sets.

Instead, in this section we introduce a class of trajectories from which it is possible to properly select an observation set \( X \) for which the conditions defined in Propositions 1 for the prismatic joint and in Propositions 2 and 3 for the revolute joint, are not only necessary but also sufficient to verify if two markers are attached to consecutive links. An observation set with this property is said to be fully informative. A class of fully informative observation sets can be formally defined as follows.

**Definition 1:** Consider a collection of \( n-1 \) trajectories, where each trajectory is obtained moving only one joint, and keeping all the others stuck. For \( k = 1, \ldots, n-1 \), without loss of generality, assume the \( k \)-th trajectory is obtained varying the joint signal \( q_k \) and let \( t_{1,k} \) and \( t_{2,k} \) be two time instants such that \( mod(q_k(t_{1,k})) \neq mod(q_k(t_{2,k})) \) and \( q_w(t_{1,k}) = q_w(t_{2,k}), w \neq k \), where \( mod(\cdot) \) is the \( 2\pi \) module operator. Then let us define the observation set \( X \) as the collection of the pairs

\[
(D_p(t_{1,k}), q(t_{1,k})), (D_p(t_{2,k}), q(t_{2,k}))
\]

for \( k = 1, \ldots, n-1 \).

Observe that \( X \) has \( 2n-2 \) observations, which for typical robot (i.e., \( n = 3, 4, 5, 6 \)) represents a limited number of observations. It is possible to show that \( X \) is a fully informative set and in particular we have the following results.

**Proposition 4:** Let \( M_{i_1} \) and \( M_{i_2} \) be two markers satisfying equations in Proposition 1 for a set of observations \( X \) defined in Definition 1 and joint signal \( q_k \). Then the corresponding links \( L_{i_1} \) and \( L_{i_2} \) are subsequent in the kinematic chain and the joint between them is prismatic with input signal \( q_k \).

**Proposition 5:** Let \( M_{i_1} \) and \( M_{i_2} \) be two markers satisfying equations in Proposition 2 and Proposition 3 for a set of observations \( X \) defined in Definition 1 and joint signal \( q_k \). Then the corresponding links \( L_{i_1} \) and \( L_{i_2} \) are subsequent in the kinematic chain and the joint between them is revolute with input signal \( q_k \).

The proofs of the above Propositions are reported in Section VII.

A couple of remarks are now in order.

**Remark 1:** As said there are examples of observation sets \( X \) which are not fully informative and that might lead to false positive situations. However in the numerical section we will show, by Monte Carlo simulations, that selecting these false positive observation sets from generic trajectories seems to be a very unlikely event.

**Remark 2:** It is worth noting that the condition \( mod(q_k(t_{1,k})) \neq mod(q_k(t_{2,k})) \) might not be verified with general input trajectories. A straightforward example happens when the actuation signal is periodic, with period \( T_q \) and \( t_2 = t_1 + T_q \). To avoid this situations we assume that the obtained trajectories are post processed, simply removing the redundant values.

### V. PROPOSED APPROACH

In this Section we describe the algorithm we propose to deal with the robot kinematic structure identification problem. Our strategy consists in iterating over all the possible triplets, composed by a pair of markers and a joint signal, the procedure described by the flow chart in Figure 2.

Specifically, given a pair of markers and a joint signal, the algorithm first evaluates if the markers are connected through a prismatic joint and, in case this test is negative, it secondly evaluates if they are connected through a revolute joint.

The first test consists in checking the feasibility of the linear equations defined in Proposition 1. The second test, instead, is composed by two steps; the first step verifies if the linear equations of Proposition 2 admit solution, and, in such case, also the second step is performed which consists
in solving the system of non-linear equations of Proposition 3. Observe that, in this way, the last step is performed only when it is necessary, thus, minimizing its executions. This fact is particularly relevant from the computational point of view, since the non linear test is the most expensive.

A. Empirical results for general trajectories

In this section we investigate the effectiveness of the proposed approach for general trajectories, by running Monte Carlo simulations. In particular, simulating different robot kinematics (with \( n = 6 \)), we obtained a dataset composed of 128 time-series, each one accounting for 50 observations. Among the different simulations we let vary several parameters, like the joint type order \( S_{Q_j} \), the DH parameters and the markers positioning, namely \( R_{M_i}^{L_j} \) and \( L^j_{M_i} \). As far as the input trajectories are concerned, we simulate a sinusoid for each joint signal, with amplitude and frequency randomly selected in each simulation.

For each time-series we have considered all the possible triplets, i.e. all pairs of markers and joint signal, for a total number of 2400 triplets, and we computed the systems of equations defined by Proposition 1, 2 and 3 to verify if the systems have solution.

Results are reported in Figure 3 in the form of confusion matrix. As usual the elements along the diagonal quantify the well classified triplets, while the elements outside the diagonal the ones which are misclassified. For example, when considering Proposition 1 the (1, 1) entry quantifies the cases in which the system defined by Proposition 1 holds true and the elements of the considered triplets identify a prismatic transformation, while the (2, 2) entry the cases in which the system of equations is not verified and the relation between the elements of the triplet is not prismatic. The (1, 2) entry instead, represent the cases in which the system of equations holds even if the triple of elements are not connected, while the (2, 1) element represent the opposite situation.

Results confirm that equations defined by Proposition 2 and 3 identify only a necessary condition when considered alone, since a significant number of false positives occurred. Indeed it happens that the relative systems of equations have solutions even if the two markers are not connected. On the other hand, considering them together as done in Proposition 3 the number of false positives goes to zero, allowing a perfect classification. Basically, empirical evidence highlights how the probability that the dataset collected contains only observations that leads to a false positives is close to zero.

VI. CONCLUSION

In this paper we introduced a novel and computational efficient algorithm able to learn the robot kinematic structure from visual observations. We prove the effectiveness of our approach in several simulated environments, testing different type of kinematics. As future work we plan to extend the proposed approach to the case of noisy measurements, and test the algorithm also in a real robot. In particular we expect to replace the two linear systems with two linear least squared problems, and the non linear system with a non linear least squared problem, as well as defining a threshold in the error, to discriminate among the cases in which the links are connected or not.

VII. APPENDIX

To prove Proposition 4 and 5 we need to introduce the relative transformations between two links \( L_{j_1} \) and \( L_{j_2} \) not subsequent in the kinematic chain. \( R_{L_{j_1}^{L_{j_2}}}^{L_{j_1}} \), the relative orientation between \( L_{j_1} \) and \( L_{j_2} \) is obtained iterating equation
along the kinematic chain, i.e. considering all transformations induced by the joints between \( L_{j_1} \) and \( L_{j_2} \):\[
R_{L_{j_2}}^{L_{j_1}} = \prod_{m=j_{1}+1}^{m=j_2} (R_{\theta_m}(\theta_m)R_x(\alpha_m)),
\]
where \( \theta_m \) is a function of \( q_{km} \) if the joint is revolute.

As regards the position the following general equation holds\[(15)\]
\[
I_{L_{j_2}}^{L_{j_1}} = \sum_{m=j_{1}+1}^{m=j_2} \left( R_{L_{j_2}}^{L_{j_1}} L_{m}^{m-1} \right).
\]

A. Proof of Proposition 4

Proposition 4 states that, when considering the excitation trajectories in Definition 1 and the relative minimal set of observations \( \tilde{X} \), \( M_{i_1}, M_{i_2} \) and \( q_k \) verify equations of Definition 1, then \( L_{j_1} \) and \( L_{j_2} \) are subsequent in the kinematic chain, and the joint between them is prismatic with \( q_k \) as actuation signal.

We prove this proposition by contradiction, showing that, if the system of equations defined by Proposition 1 holds, then triple \( M_{i_1}, M_{i_2} \) and \( q_k \) describe a prismatic transformation.

First of all we exclude the case in which there is at least one revolute joint between \( L_{j_1} \) and \( L_{j_2} \). Without loss of generality let \( j_3 \) be the index of the link coming after the revolute joint. The relative orientation between the markers is \( R_{M_{i_2}}^{M_{i_1}} = R_{L_{j_2}}^{L_{j_1}} R_{L_{j_3}}^{L_{j_2}} R_{M_{i_2}}^{L_{j_2}} \), that exploiting (14), becomes
\[
R_{L_{j_2}}^{L_{j_1}} = R_{M_{i_1}}^{M_{i_2}} \prod_{m=j_{1}+1}^{m=j_2} (R_{\theta_m}(\theta_m)R_x(\alpha_m)) R_{\theta_j_3}(\theta_j_3) R_x(\alpha_j_3)
\]
\[
\prod_{m=j_{1}+1}^{m=j_2} (R_{\theta_m}(\theta_m)R_x(\alpha_m)) R_{L_{j_2}}^{L_{j_1}}
\]
\[
= R_{M_{i_1}}^{M_{i_2}} R_{\theta_j_3}(\theta_j_3) R_x(q_{j_3}) R_{M_{i_2}}^{L_{j_1}} = R_{M_{i_1}}^{M_{i_2}} R_x(q_{j_3}) R_{M_{i_2}}^{L_{j_3}}.
\]

If Equations of Proposition 1 are verified, the relative orientation between \( M_{i_1} \) and \( M_{i_2} \) is constant, and consequently, considering two different time instants \( t_1 \) and \( t_2 \), \( R_{M_{i_2}}^{M_{i_1}}(t_1) = R_{M_{i_2}}^{M_{i_1}}(t_2) \), that implies
\[
R_{L_{j_3}}^{L_{j_1}}(t_1) R_x(q_{j_3}(t_1)) R_{L_{j_2}}^{L_{j_1}}(t_1) = R_{t_1}^{M_{i_1}}(t_2) R_x(q_{j_3}(t_2)) R_{L_{j_2}}^{L_{j_1}}(t_2).
\]

However the last condition is not verified when the input trajectories are in \( \tilde{X} \). Indeed, when considering two input locations belonging to the subset of trajectories with \( \text{mod}(q_{j_3}(t_1)) \neq \text{mod}(q_{j_3}(t_2)) \) and \( q_w(t_1) = q_w(t_2) \) if \( w \neq k_{j_3} \), \( R_{L_{j_3}}^{L_{j_1}}(t_1) = R_{L_{j_3}}^{L_{j_1}}(t_2) \) and \( R_{M_{i_2}}^{L_{j_3}}(t_1) = R_{M_{i_2}}^{L_{j_3}}(t_2) \), since they are function of the \( q_w \) signals, that assume the same values in \( t_1 \) and \( t_2 \). This implies \( R_x(q_{j_3}(t_1)) = R_x(q_{j_3}(t_2)) \), and for the uniqueness if the Euler angles \( \text{mod}(q_{j_3}(t_1)) = \text{mod}(q_{j_3}(t_2)) \), that is against Definition 1.

The last observation proves that, when the joint signals are chosen in accordance with Definition 1, \( R_{M_{i_2}}^{L_{j_3}} \) cannot be constant over \( \tilde{X} \) if there is at least one revolute joint between the links. Consequently, to conclude the proof, we consider a configuration in which there are one or more prismatic joints between \( L_{j_1} \) and \( L_{j_2} \). In these configurations the \( M_{i_2} \) RF origin in the \( M_{i_1} \) RF is
\[
\begin{align*}
I_{M_{i_2}}^{M_{i_1}} &= I_{L_{j_3}}^{M_{i_1}} + I_{L_{j_2}}^{L_{j_3}} + M_{i_2} \left(-M_{L_{j_2}} \right) \\
&= \sum_{m=j_{1}+1}^{m=j_2} \left( \left( R_{M_{i_2}}^{M_{i_1}} \left[ \begin{array}{cc} 0 & R_x(\theta_m) \end{array} \right] + R_x(\theta_m) \left[ \begin{array}{cc} d_m & 0 \\
0 & 0 \end{array} \right] \right) \right)
\end{align*}
\]
\[
\begin{align*}
I_{L_{j_2}}^{M_{i_1}} &= R_{M_{i_1}}^{M_{i_2}} \left[-I_{L_{j_2}}^{M_{i_2}} + R_{M_{i_2}}^{M_{i_1}} \left[-I_{L_{j_2}}^{M_{i_2}} \right] \right] \end{align*}
\]

where the terms before the last sum are constant and equal to \( I_{t_1,t_2} \), while the ones in the last sum depends on the joint signals. As in 5, the last equation defines a linear system with coefficient matrix \( A^* = \left[ I_3 \ q_{j_1} \ I_3 \ldots \ q_{j_2} \ I_3 \right] \) and vector of variables \( b_{t_1,t_2}^* = \left( (l_{t_1,t_2})^T \ Z(R_{L_{j_1}}^{M_{i_1}})^T \ldots \ Z(R_{L_{j_2-1}}^{M_{i_1}})^T \right)^T \), where, as in 5, the \( Z(R_{M_{i_1}}^{M_{i_1+1}}) \) elements have unitary norm. Assuming equations of Proposition 1 are verified for the set of observations \( X \), it holds
\[
I_{M_{i_2}}^{M_{i_1}}(X) = A(X)b_{t_1,t_2}^* = A^*(X)b_{t_1,t_2}^*.
\]

Anyway, if the inputs are selected in accordance with Definition 1 \( \text{rank}(A(X)) = 6 \) and \( \text{rank}(A^*(X)) = 3 + 3(j_2 - j_1) \), and, given the constraints on the \( b_{t_1,t_2} \) and \( b_{t_1,t_2}^* \) elements, \( \text{rank}(A(X)b_{t_1,t_2}) \leq \text{rank}(A^*(X)b_{t_1,t_2}^*) \). Moreover to obtain \( \text{rank}(A(X)b_{t_1,t_2}) = \text{rank}(A^*(X)b_{t_1,t_2}^*) \) we need to consider only one prismatic joint and, since \( \text{span}(\left\{q_k(t_1) \ldots q_k(t_{m-1})\right\}) = \text{span}(\left\{q_k(t_1) \ldots q_k(t_{m-1})\right\}) \) if and only of \( k_i = k_j \), the equivalence in 19 holds if and only if the joint input signal is \( q_k \), proving Proposition 4.

B. Proof of Proposition 5

As done for Proposition 4 we prove the proposition by contradiction. Assume that conditions defined by Proposition 2 and Proposition 3 are satisfied, and the inputs are selected accordingly to Definition 1. Moreover suppose that \( L_{j_3} \) and \( L_{j_2} \) are not consecutive. Let \( N_{rev} \) be the number of revolute joints present in the chain between the two links, and \( k_{j_3} \ldots k_{j_3+N_{rev}} \) their indexes. The relative orientation between \( M_{i_2} \) and \( M_{i_1} \) is \( R_{M_{i_2}}^{M_{i_1}} = R_{L_{j_1}}^{L_{j_3}} R_{L_{j_2}}^{L_{j_1}} R_{M_{i_2}}^{L_{j_2}} \), with \( R_{L_{j_1}}^{L_{j_3}} \) as in 14. As done in 16, the relative orientation between the markers can be rewritten highlighting the contributions of a particular revolute joint, for example the \( j_3 \) joint, obtaining \( R_{M_{i_2}}^{M_{i_1}} = R_{L_{j_3}}^{L_{j_1}} R_{L_{j_1}}^{L_{j_3}} R_{M_{i_2}}^{L_{j_2}} \). At the same time, since conditions defined by Proposition 3 are assumed to be true, it holds \( R_{M_{i_2}}^{L_{j_2}} = R_{L_{j_1}}^{L_{j_3}} R_{L_{j_1}}^{L_{j_3}} R_{M_{i_2}}^{L_{j_2}} \), with \( R_{L_{j_1}}^{L_{j_3}} \) and \( R_{M_{i_2}}^{L_{j_2}} \) corresponding to the solution of the non linear
system. Then it follows

\[ R_{L_{j1}}^{M_{j2}} R_{e}^{j_2} R_{M_{j1}}^{L_{j2}} = R_{L_{j1}}^{M_{j2}} R_{e}^{j_2} R_{M_{j1}}^{L_{j2}}, \]

and exploiting the standard properties of the rotation matrices and the Euler angles, it holds \( R_{e}^{j_2} = R_{L_{j1}}^{M_{j2}} R_{e}^{j_2} R_{M_{j1}}^{L_{j2}} \). Moreover, \( R_{L_{j1}}^{M_{j2}} \) and \( R_{M_{j1}}^{L_{j2}} \) can be expressed with respect to their Euler angles, for example adopting the \( zxz \) convention, obtaining

\[ R_{e}^{j_2}(q_{k_3}) = R_{e}^{j_2} R_{e}^{j_1} R_{L_{j1}}^{M_{j2}} R_{e}^{j_2} R_{M_{j1}}^{L_{j2}}(q_{k_3}) \]

When neglecting the degenerate configurations, for the Euler angles uniqueness the last equation is true if and only if \( \beta_1 \) and \( \beta_2 \) are null, and \( q_k - \alpha_1 - \gamma_2 = q_{k_3} + \alpha_2 \), namely \( q_k - q_{k_3} = \alpha_1 + \gamma_2 + \gamma_1 + \alpha_2 \). Now consider two time instants \( t_1 \) and \( t_2 \) in \( \mathbb{X} \) satisfying \( \mod(q_{j_2}(t_1)) \neq \mod(q_{j_2}(t_2)) \) and \( q_w(t_1) = q_w(t_2) \) if \( w \neq k_{j_3} \). It holds

\[ R_{L_{j3}}^{M_{j2}}(t_1) = R_{L_{j3}}^{M_{j2}}(t_2) \quad \text{and} \quad R_{M_{j2}}^{L_{j3}}(t_1) = R_{M_{j2}}^{L_{j3}}(t_2), \]

implying also that \( \alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2 \) and \( \gamma_2 \) are constant in \( t_1 \) and \( t_2 \). Then

\[ q_k(t_1) - q_k(t_2) = q_{k_3}(t_1) - q_{k_3}(t_2). \]

Under the excitation assumptions in Definition 1 the last equation holds if and only if \( k = k_{j_3} \). Similar considerations can be done considering the other revolute joints and opportune time instants, resulting in a set of conditions that are unfeasible when \( N_{REV} > 1 \). This observations prove two important facts which are verified if the system of equations defined by Proposition 3 holds for \( X \). The first fact is that only one revolute transformation is allowed, and the second is that the input of the revolute joint must be \( q_k \).

To conclude the analysis we need to exclude the case in which there is a revolute joint with input signal \( q_k \) and one or more prismatic joints between \( L_{j_1} \) and \( L_{j_2} \). This case is excluded by equations of Proposition 2. Let \( j_3 \) be the revolute joint index. Consider the \( I_{M_{j2}}^{L_{j3}} \) expression in (3), and expand \( I_{M_{j2}}^{L_{j3}} \) in order to highlight the prismatic transformations occurring before or after the revolute joint \( I_{M_{j2}}^{L_{j3}} \). Then it follows

\[ I_{M_{j1}}^{L_{j3}} = I_{M_{j1}}^{L_{j3}} + R_{L_{j3}}^{M_{j2}}(I_{L_{j2}}^{M_{j3}}) + \sum_{m=1}^{j_2} R_{L_{j3m}}^{M_{j2}} I_{L_{j3m}}^{L_{j2}} \]

where the \( R_{L_{j3m}}^{M_{j2}} \) and \( I_{L_{j3m}}^{L_{j2}} \) matrices in the sums are constant since they refer to prismatic transformations, while the \( I_{L_{j3m}}^{L_{j2}} \) and \( R_{L_{j3m}}^{M_{j2}} \) expressions are defined in (2) with \( d_{j_3} \) dependent on the joint signals. As in equation (18), the contributions due to the first sum and the first term define a linear system with coefficient matrix \( A^* = [I_3, q_{k_{j_3}}, I_3 \ldots q_{k_{j_3}}, I_3] \), and vector of variables \( b_{1,3}^{*} = \begin{bmatrix} (I_{L_{j3}}^{M_{j3}})^T Z(R_{L_{j3}}^{M_{j3}})^T \ldots Z(R_{L_{j3m-1}}^{M_{j3}})^T \end{bmatrix}^T \), with the \( Z(R_{L_{j3m-1}}^{M_{j3}}) \) elements constrained to have unitary norm.

On the other hand, when Proposition 2 holds, \( I_{M_{j2}}^{L_{j3}} \in \text{span} \{ \hat{A}(X) \} \), and considering the constraints on the \( b_{1,3}^{*} \) norm and that \( \text{span} \{ A^*(X) \} \notin \text{span} \{ \hat{A}(X) \} \), it follows that it is impossible that there are prismatic transformations between \( L_{j_1} \) and the link after the revolute joint, i.e. \( j_3 = j_1 + 1 \). To conclude the proof we need to exclude the possibility that there are prismatic transformations between the revolute joint and \( L_{j_2} \). Observe that the second sum contributions in (20) are proportional to \( R_{M_{j2}}^{L_{j3}} \) with time-variant coefficient, due to the joint signals variability. Anyway this is in contradiction with the conditions in Proposition 2 and the proof is concluded. Indeed, since \( \hat{A}(X) \) is a full rank matrix, if (10) holds, its solution is unique, but this is not in accordance with the coefficient variability in (20).

**REFERENCES**

[1] M. Yim, D. G. Duff, and K. D. Roufas, “Polybot: a modular recon¬figurable robot,” in Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065), vol. 1, April 2000, pp. 514–520 vol.1.

[2] M. Yim, W. Shen, B. Salemi, D. Rus, M. Moll, H. Lipson, E. Klavins, and G. S. Chirikjian, “Modular self-reconfigurable robot systems [grand challenges of robotics],” IEEE Robotics Automation Magazine, vol. 14, no. 1, pp. 43–52, March 2007.

[3] L. Brodbeck and F. Iida, “Enhanced robotic body extension with mod¬ular units,” in 2012 IEEE International Conference on Intelligent Robots and Systems, Oct 2012, pp. 1428–1433.

[4] J.-C. Latombe, Robot Motion Planning. Norwell, MA, USA: Kluwer Academic Publishers, 1991.

[5] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, Robotics, Mod¬elling, Planning and Control, 2009.

[6] G. Gao, W. Wang, K. Lin, and Z. Chen, “Kinematic calibration for articulated arm coordinate measuring machines base on particle swarm optimization,” in 2009 Second International Conference on Intelligent Computation Technology and Automation, vol. 1, Oct 2009, pp. 189–192.

[7] T. Zhou and B. E. Shi, “Simultaneous learning of the structure and kinematic model of an articulated body from point clouds,” in 2016 International Joint Conference on Neural Networks (IJCNN), July 2016, pp. 5248–5255.

[8] M. Hersch, E. L. Sauser, and A. Billard, “Online learning of the body schema.” vol. 5, pp. 161–181, 06 2008.

[9] T. Rhr, J. Sturm, D. Pangeric, M. Beetz, and D. Cremers, “A generalized framework for opening doors and drawers in kitchen environments,” in 2012 IEEE International Conference on Robotics and Automation, May 2012, pp. 3852–3858.

[10] J. Sturm, C. Plagemann, and W. Burgard, “Unsupervised body scheme learning through self-perception,” in 2008 IEEE International Confer¬ence on Robotics and Automation, May 2008, pp. 3328–3333.

[11] M. Fiola, “Designing highly reliable fiducial markers,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 7, pp. 1317–1324, July 2010.

[12] E. Olson, “Apriltag: A robust and flexible visual fiducial system,” in 2011 IEEE International Conference on Robotics and Automation, May 2011, pp. 3400–3407.

[13] J. Wang and E. Olson, “Apriltag 2: Efficient and robust fiducial detection,” in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Oct 2016, pp. 4193–4198.

[14] S. Kim and K. Yoon, “Point density-invariant 3d object detection and pose estimation,” in 2017 IEEE International Conference on Image Processing (ICIP), Sept 2017, pp. 2647–2651.

[15] W. Yun, J. Lee, J. Lee, and J. Kim, “Object recognition and pose estimation for modular manipulation system: Overview and initial results,” in 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), June 2017, pp. 198–201.

[16] H. Kim, J. Y. Lee, J. H. Kim, J. B. Kim, and W. Y. Han, “Object recognition and pose estimation using klt,” in 2012 12th International Conference on Control, Automation and Systems, Oct 2012, pp. 214–217.