ON THE OBSERVABILITY IN VISUAL SLAM NETWORKS

Qier An and Yuan Shen

Department of Electronic Engineering, Tsinghua University
Beijing National Research Center for Information Science and Technology, Beijing, China
Email: aqe18@mails.tsinghua.edu.cn, shenyuan_ee@tsinghua.edu.cn

ABSTRACT

Observability is an essential aspect for the performance of a visual Simultaneous Localization and Mapping (SLAM) network. This paper presents a statistical perspective to evaluate the observability of visual SLAM networks and its dependence on network structure. In particular, we first give the general form for the Fisher information matrix (FIM) of each visual observation and the impact of 3D point translation and camera motion on observations. Then the observability of visual SLAM networks is investigated based on the nullspace of the FIM and its relation with network structure to derive the lost rank and its upper and lower bounds. We also propose the ill-conditioned score to evaluate the degradation of visual SLAM performance under any network structure. Numerical results validate our conclusions.

***Index Terms*—** Observability, Visual SLAM, FIM, Network structure

1. INTRODUCTION

Autonomous navigation has received wide concern recently with the rapid development of autonomous driving [1], unmanned aerial vehicle (UAV) [2, 3] and robotics [4]. Although Global Positioning System (GPS) provides global real-time absolute localization service and prevails in smart systems, it may fail in some heavily-occluded environments, such as city canyon and forest [5, 6]. Meanwhile, many navigation methods require a prior map of the environment [7], which is unavailable in some tasks like space exploration. To realize autonomous navigation in unknown environments, the visual SLAM technique simultaneously builds the 3D map of the scene and locates the agent in the map through images captured by onboard cameras, without relying on any infrastructure. Visual SLAM frameworks can be categorized as filter-based [8] and optimization-based [9, 10] methods, and can also be classified as direct [11, 12], semi-direct [13] and indirect [8–10] methods. Despite the prosperity in applications, previous works mainly focus on the implementation accuracy and efficiency, and the lack of theoretic investigation about visual SLAM impedes further improvement of performance.

For a visual SLAM network, the performance can be evaluated from two aspects: accuracy and observability. The accuracy metric gives an overall evaluation for 3D map reconstruction and agent localization, since the errors of multiple 3D points and cameras are added together to obtain the accuracy value. To quantify the accuracy of camera localization, previous works calculate the mean square error (MSE) between the estimated trajectory and the ground-truth, and adopt the iterative closest point (ICP) method for 3D map evaluation [14]. A FIM-based metric has also been proposed which eliminates the reliance on the ground-truth model [15]. The observability metric provides a detailed examination for the unobserved directions in the state space of visual SLAM problem, which explains the reason for the inconsistency of solutions under certain conditions [16]. For linear time-invariant (LTI) and linear time-varying (LTV) systems, the observability is usually investigated based on the observability grammian or observability matrix [17–19], which has been applied in filter-based 3D [7, 16, 20] and range/bearing-based 2D visual SLAM algorithms [21]. Some other works also adopt the Fisher information to inspect the stochastic observability of visual SLAM [19, 22]. The major limitations of previous works about the observability of visual SLAM are twofold: the investigation only applies to some simple networks with a limited number of 3D points and cameras, and meanwhile only the number of unobserved dimensions are given, without detailed results about the unobserved directions and its relation with the network deployment.

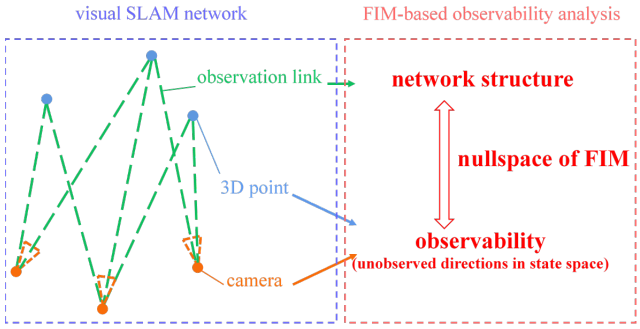


Fig. 1: Observability analysis of visual SLAM networks and its dependence on network structure based on the nullspace of FIM.

In this paper, we establish a statistical framework to investigate the observability of visual SLAM networks and demonstrate the dependence of unobserved directions in state space on network structure (see Fig. 1). In particular, we first formulate the visual SLAM into a state estimation problem to derive the FIM for each visual observation, and analyze the impact of 3D point and camera motion on observations. Then we investigate the observability of visual SLAM networks based on the nullspace of the FIM to derive the lost rank and the relation between unobserved directions and network structure, and also give the upper and the lower bounds for the lost rank. Besides, we also propose the ill-conditioned score to quantitatively evaluate the degradation of visual SLAM performance. Numerical results validate the effectiveness of our results.

This research was supported in part by the National Natural Science Foundation of China under Grant 61871256, National Key R&D Program of China 2020YFC1511803, and Tsinghua University Initiative Scientific Research Program.

2. SYSTEM MODEL

In this section, we formulate the visual SLAM into a state estimation problem and introduce the FIM to investigate the observability of visual SLAM from the perspective of statistical information.

In computer vision, the 3D map of the scene is usually represented by the point cloud model. Assume there are N cameras and M 3D points in the visual SLAM network, with index sets $\mathcal{I}_N = \{1, 2, \ldots, N\}$ and $\mathcal{I}_M = \{1, 2, \ldots, M\}$ respectively. If 3D point $\boldsymbol{x}_j$ is observed by the i-th camera, the imaging process can be described by the pin-hole camera model [23] as[1]

$$\mathbf{x}_i^{(j)} = \frac{1}{z_i^{(j)}} \boldsymbol{K}_i \boldsymbol{R}_i (\boldsymbol{x}_j - \boldsymbol{c}_i) + \mathbf{n}_i^{(j)}, \quad i \in \mathcal{I}_N, j \in \mathcal{I}_M \tag{1}$$

where $\mathbf{x}_i^{(j)}$ represents the projection pixel of $\boldsymbol{x}_j$ in the i-th camera, $\boldsymbol{K}_i, \boldsymbol{R}_i$ and $\boldsymbol{c}_i$ denote the i-th camera's intrinsic parameter matrix, rotation matrix and position, respectively, $z_i^{(j)}$ is the last coordinate of $\boldsymbol{R}_i(\boldsymbol{x}_j - \boldsymbol{c}_i)$, and $\mathbf{n}_i^{(j)} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathrm{n}i}^{(j)2} \boldsymbol{I})$ denotes the imaging noise.

In visual SLAM, the 3D point position and the camera pose are estimated simultaneously. To simplify analysis, we assume the camera orientation is derived in advance, and thus the state vector is formulated as $\boldsymbol{\theta} = (\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}, \ldots, \boldsymbol{x}_M^{\mathrm{T}}, \boldsymbol{c}_1^{\mathrm{T}}, \boldsymbol{c}_2^{\mathrm{T}}, \ldots, \boldsymbol{c}_N^{\mathrm{T}})^{\mathrm{T}}$. The observation vector consists of all the pixels, i.e., $\mathbf{x} = (\mathbf{x}^{(1)\mathrm{T}}, \mathbf{x}^{(2)\mathrm{T}}, \ldots, \mathbf{x}^{(M)\mathrm{T}})^{\mathrm{T}}$ with $\mathbf{x}^{(j)} = (\mathbf{x}_1^{(j)\mathrm{T}}, \mathbf{x}_2^{(j)\mathrm{T}}, \ldots, \mathbf{x}_N^{(j)\mathrm{T}})^{\mathrm{T}}$, which follows the likelihood function as [24]

$$p(\mathbf{x};\boldsymbol{\theta}) \propto \exp\Big(\sum_{\substack{i \in \mathcal{I}_N \\ j \in \mathcal{I}_M}} -\frac{1}{2\sigma_{\mathrm{n}i}^{(j)2}} \big|\big| \mathbf{x}_i^{(j)} - \frac{1}{z_i^{(j)}} \boldsymbol{K}_i \boldsymbol{R}_i (\boldsymbol{x}_j - \boldsymbol{c}_i) \big|\big|^2 \Big). \tag{2}$$

To evaluate the information contained in each observation, we adopt the FIM in statistical analysis which is computed as [25]

$$\boldsymbol{J}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}} \left\{ -\triangle_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \ln p(\mathbf{x}; \boldsymbol{\theta}) \right\}, \tag{3}$$

where $[\triangle_{\boldsymbol{u}}^{\boldsymbol{v}}]_{i,j} = \partial^2 / \partial [\boldsymbol{u}]_i \partial [\boldsymbol{v}]_j$. According to (3), the FIM measures the amount of information for each dimension of the state provided by the observation sample. For our visual SLAM problem, the nullspace of the network's FIM corresponds to the unobserved directions of the state vector. Therefore in the following sections, we focus on the investigation of the network's FIM to derive the observability in visual SLAM networks.

3. THEORETICAL ANALYSIS

The FIM of the visual SLAM network can be derived according to (2) and (3) as

$$\boldsymbol{J}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^{\mathrm{T}} & \boldsymbol{C} \end{bmatrix}, \tag{4}$$

$$\boldsymbol{A} = \sum_{i \in \mathcal{I}_N} \mathrm{diag}\{\boldsymbol{J}_{i1}, \boldsymbol{J}_{i2}, \ldots, \boldsymbol{J}_{iM}\}, \tag{5}$$

$$\boldsymbol{B} = \begin{bmatrix} -\boldsymbol{J}_{11} & -\boldsymbol{J}_{21} & \ldots & -\boldsymbol{J}_{N1} \\ -\boldsymbol{J}_{12} & -\boldsymbol{J}_{22} & \ldots & -\boldsymbol{J}_{N2} \\ \vdots & \vdots & \ldots & \vdots \\ -\boldsymbol{J}_{1M} & -\boldsymbol{J}_{2M} & \ldots & -\boldsymbol{J}_{NM} \end{bmatrix}, \tag{6}$$

$$\boldsymbol{C} = \sum_{j \in \mathcal{I}_M} \mathrm{diag}\{\boldsymbol{J}_{1j}, \boldsymbol{J}_{2j}, \ldots, \boldsymbol{J}_{Nj}\}, \tag{7}$$

where $\boldsymbol{J}_{ij}$ denotes the FIM generated by observation $\mathbf{x}_i^{(j)}$. If 3D point $\boldsymbol{x}_j$ is not observed by camera $\boldsymbol{c}_i$, then $\boldsymbol{J}_{ij} = \mathbf{0}$ in (5)-(7).

3.1. FIM for each visual observation

We first give the general form for $\boldsymbol{J}_{ij}$ in (5)-(7). For simplicity, superscripts and subscripts are omitted in this section.

Proposition 1 *The FIM generated by the observation of 3D point $\boldsymbol{x}$ in camera $\boldsymbol{c}$ follows [15]*

$$\boldsymbol{J} = \frac{1}{\sigma_{\mathrm{n}}^2 z^2} \Big(\boldsymbol{v}_1 \boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{v}_2 \boldsymbol{v}_2^{\mathrm{T}} \Big), \tag{8}$$

where $\boldsymbol{v}_1 = \boldsymbol{r}_1 - (x/z)\boldsymbol{r}_3, \boldsymbol{v}_2 = \boldsymbol{r}_2 - (y/z)\boldsymbol{r}_3$ with $(x, y, z)^{\mathrm{T}} = \boldsymbol{R}(\boldsymbol{x} - \boldsymbol{c})$ and $\boldsymbol{r}_k^{\mathrm{T}}$ denoting the k-th row of $\boldsymbol{R}$.

Remark 1 *According to Proposition 1, the FIM of each visual observation is a 3×3 matrix of rank 2, with the eigenvector corresponding to eigenvalue 0 falling on viewing direction $(\boldsymbol{x} - \boldsymbol{c})$, which means that there is no information gain along $(\boldsymbol{x} - \boldsymbol{c})$. We also derive the impact of 3D point and camera motion on the FIM as follows*

- *3D point translation: If 3D point $\boldsymbol{x}$ moves along the viewing direction to $\boldsymbol{x}' = \boldsymbol{x} + a(\boldsymbol{x} - \boldsymbol{c})$, the FIM generated by $\boldsymbol{x}'$ is derived as $\boldsymbol{J}' = \boldsymbol{J}/(1 + a)^2$.*
- *Camera translation: If camera $\boldsymbol{c}$ moves along the viewing direction to $\boldsymbol{c}' = \boldsymbol{c} + b(\boldsymbol{x} - \boldsymbol{c})$, the FIM generated by $\boldsymbol{c}'$ is derived as $\boldsymbol{J}' = \boldsymbol{J}/(1 - b)^2$.*
- *Camera rotation: If the camera orientation rotates by an angle with the position unchanged, the eigenvector corresponding to eigenvalue 0 remains the same, and only the information gains in the plane perpendicular to $(\boldsymbol{x} - \boldsymbol{c})$ change.*

To give geometric interpretations, we resort to the information ellipsoid [15]. According to Proposition 1, the information ellipsoid of $\boldsymbol{J}$ is a 2D ellipse in the eigenspace, with the normal direction equivalent to $(\boldsymbol{x} - \boldsymbol{c})$. When 3D point $\boldsymbol{x}$ or camera $\boldsymbol{c}$ moves along $(\boldsymbol{x} - \boldsymbol{c})$, only the size of the information ellipsoid changes. When the camera orientation rotates, the new information ellipsoid can be derived by rotating the 2D ellipse around $(\boldsymbol{x} - \boldsymbol{c})$ in the plane.

3.2. Observability of visual SLAM network

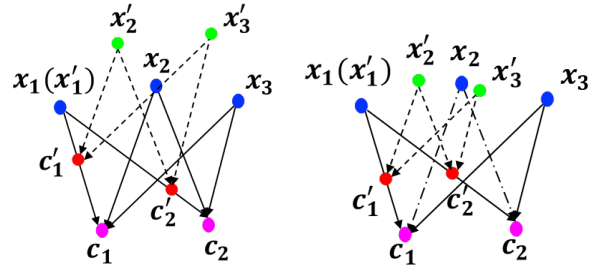
3.2.1. Nullspace and lost rank of FIM

The visual SLAM network topology is determined by the observation links between cameras and 3D points. Assume the N cameras and the M 3D points compose K connected components. Each connected component consists of four types of structures as[2]

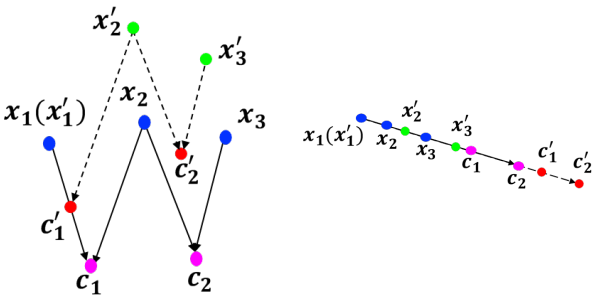
(1) Coplanar butterfly structure: As Fig. 2(a) shows, all the observation links are on the same plane and constitute a butterfly structure. Fix 3D point $\boldsymbol{x}_1$ for example, and move $\boldsymbol{c}_1, \boldsymbol{c}_2$ along their viewing directions to derive $\boldsymbol{c}_1' = \boldsymbol{c}_1 + a\boldsymbol{v}_{11}$ and $\boldsymbol{c}_2' = \boldsymbol{c}_2 + b\boldsymbol{v}_{21}$, respectively. The other 3D points $\boldsymbol{x}_j'$ and cameras $\boldsymbol{c}_i'$ in the structure are derived by drawing lines parallel to corresponding viewing directions. Then we can prove that $\boldsymbol{J}(\boldsymbol{\theta})(\boldsymbol{\theta}' - \boldsymbol{\theta}) = \mathbf{0}$. Since there are two variables $a, b \in \mathbb{R}$ in $\boldsymbol{\theta}'$, each coplanar butterfly structure corresponds to two dimensions of the nullspace of $\boldsymbol{J}(\boldsymbol{\theta})$.

[1] In (1), we omit the last coordinate of the homogeneous form $(\mathbf{x}_i^{(j)\mathrm{T}}, 1)^{\mathrm{T}}$ for the simplicity of expression.

[2] $\boldsymbol{v}_{ij} = \boldsymbol{x}_j - \boldsymbol{c}_i$ and $\boldsymbol{\theta}' = (\boldsymbol{x}_1'^{\mathrm{T}}, \boldsymbol{x}_2'^{\mathrm{T}}, \ldots, \boldsymbol{x}_M'^{\mathrm{T}}, \boldsymbol{c}_1'^{\mathrm{T}}, \boldsymbol{c}_2'^{\mathrm{T}}, \ldots, \boldsymbol{c}_N'^{\mathrm{T}})^{\mathrm{T}}$.



(a) Coplanar butterfly structure (b) Non-coplanar butterfly structure



(c) Chain structure (d) Degenerate butterfly structure

Fig. 2: Examples for four types of structures in connected component.

(2) Non-coplanar butterfly structure: As Fig. 2(b) shows, all the observation links constitute a butterfly structure, while the 3D points and cameras are distributed on different planes, e.g., plane $\boldsymbol{x}_1\boldsymbol{c}_1\boldsymbol{c}_2$ and plane $\boldsymbol{x}_2\boldsymbol{c}_1\boldsymbol{c}_2$. Compared with the coplanar butterfly structure, the non-coplanar property requires $a = b$ in $\boldsymbol{\theta}'$. Thus each non-coplanar butterfly structure corresponds to one dimension of the nullspace of $\boldsymbol{J}(\boldsymbol{\theta})$.

(3) Chain structure: As Fig. 2(c) shows, all the observation links form a chain structure of length l (number of observation links). Fix 3D point $\boldsymbol{x}_1$, and move $\boldsymbol{c}_1$ along to $\boldsymbol{c}_1' = \boldsymbol{c}_1 + a\boldsymbol{v}_{11}$. The other 3D points $\boldsymbol{x}_j'$ and cameras $\boldsymbol{c}_i'$ in the structure move along their parallel viewing directions, and each generates one degree of freedom. Thus each chain structure corresponds to l dimensions of the nullspace of $\boldsymbol{J}(\boldsymbol{\theta})$.

(4) Degenerate butterfly structure: As Fig. 2(d) shows, all the observation links form a butterfly structure, while all the 3D points and cameras are distributed on the same line. It can be proved that the degenerate butterfly structure is equivalent to a chain structure with the same number of 3D points and cameras.

Based on the above analysis, the lost rank of the FIM can be derived as Proposition 2.

Proposition 2 *Assume there are K connected components in the visual SLAM network, and the k-th connected component includes B_{k1} coplanar butterfly structures, B_{k2} non-coplanar butterfly structures, and L_k chain or degenerate butterfly structures with the p-th one of length $l_p^{(k)}$. Then the lost rank of $\boldsymbol{J}(\boldsymbol{\theta})$ is derived as*

$$L(\boldsymbol{J}(\boldsymbol{\theta})) = 3K + \sum_{k=1}^{K}(\sum_{p=1}^{L_k} l_p^{(k)} + 2B_{k1} + B_{k2}), \tag{9}$$

and thus rank$(\boldsymbol{J}(\boldsymbol{\theta})) = 3(N+M) - L(\boldsymbol{J}(\boldsymbol{\theta}))$.

Remark 2 *The unobserved directions of the visual SLAM network can be derived from the nullspace of FIM $\boldsymbol{J}(\boldsymbol{\theta})$. In (9), the lost rank, i.e., the dimension of the nullspace of $\boldsymbol{J}(\boldsymbol{\theta})$, is composed of two terms: the first term $3K$ corresponds to the translation of connected components, and the second term corresponds to the deformation of network topology. These results reflect the dependence of observability on the structure of visual SLAM network.*

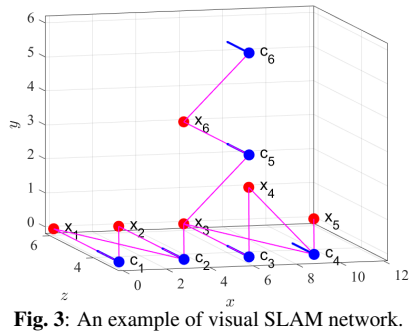


Fig. 3: An example of visual SLAM network.

3.2.2. *Upper and lower bounds for lost rank*

When the whole network is connected, i.e., $K = 1$, and consists of only a non-coplanar butterfly structure, the lost rank of $\boldsymbol{J}(\boldsymbol{\theta})$ reaches the lower bound 4. When each connected component only includes the chain and the degenerate butterfly structures, the lost rank generated by each component reaches its upper bound. Furthermore, when each component only includes a 3D point and a camera, the lost rank of $\boldsymbol{J}(\boldsymbol{\theta})$ reaches the upper bound.

Proposition 3 *The upper and the lower bounds for the lost rank of $\boldsymbol{J}(\boldsymbol{\theta})$ are as follows*

$$4 \leq L(\boldsymbol{J}(\boldsymbol{\theta})) \leq N + M + 2K \leq 2(N+M). \tag{10}$$

3.2.3. *Metric for observability: ill-conditioned score*

Based on the above analysis, we define the ill-conditioned score as a metric for the observability of visual SLAM networks.

Definition 1 (Ill-conditioned Score) *For a visual SLAM network with N cameras and M 3D points, the* ill-conditioned score *is defined as*

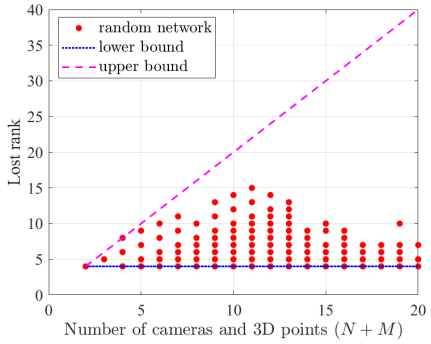
$$F = \frac{L(\boldsymbol{J}(\boldsymbol{\theta}))}{3(N+M)}, \tag{11}$$

where $L(\boldsymbol{J}(\boldsymbol{\theta}))$ is given in (9).

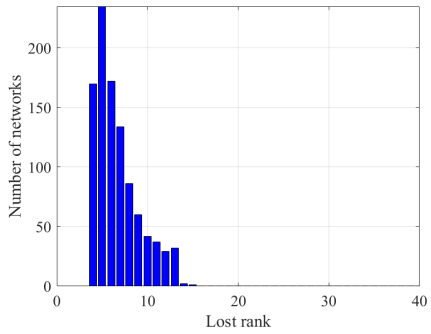
The ill-conditioned score F evaluates the observability degradation of a visual SLAM network from the perspective of information, with the value depending on network structure. A larger F means a higher proportion of unobserved dimensions. Besides, to improve the observability of a visual SLAM network, we should try to improve the connectivity of the network and reduce the number of the chain, coplanar and degenerate butterfly structures.

4. NUMERICAL RESULTS

In this section, we first compare the lost rank of an example network given by the closed-form expression in (9) and the simulation result to validate Proposition 2. Then we give the lost rank and the ill-conditioned score of 1000 random networks with different number of cameras and 3D points to verify our analysis about observability.

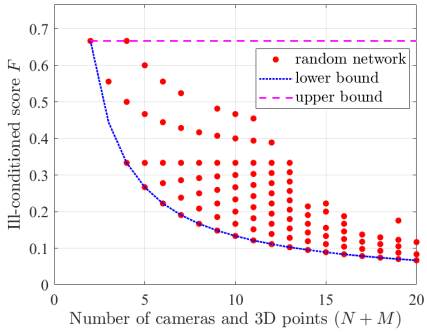


(a) Lost rank

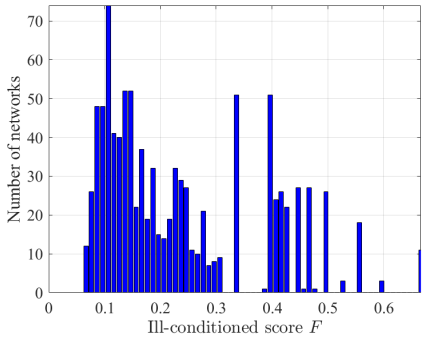


(b) Distribution of lost rank

Fig. 4: Lost rank for 1000 random networks. (a) Scatter figure for the lost rank of random networks with different number of cameras and 3D points, and the lower and upper bounds. (b) Distribution for the lost rank of random networks.



(a) Ill-conditioned score



(b) Distribution of ill-conditioned score

Fig. 5: Ill-conditioned score for 1000 random networks. (a) Scatter figure for the ill-conditioned score of random networks with different number of cameras and 3D points, and the lower and upper bounds. (b) Distribution for the ill-conditioned score of random networks.

To validate Proposition 2, we give an example network in Fig. 3 with $N=6$ cameras and $M=6$ 3D points, and the network consists of $B_1=1$ coplanar butterfly structure, $B_2=1$ non-coplanar butterfly structure, $L=3$ chain structures with lengths $l_1=1, l_2=1, l_3=3$ respectively. According to (9), the closed-form result is derived as $L(\boldsymbol{J}(\boldsymbol{\theta}))=11$, which is consistent with our simulation result. Moreover, the following simulation results of random networks also validate our closed-form result in Proposition 2.

To validate Proposition 3 and further interpret the ill-conditioned score, we generate 1000 random networks with different number of cameras and 3D points, in which the 3D points (cameras) are distributed randomly in the cube with coordinate range $[-3,3]\times[-3,3]\times[6,12]$ ($[-3,3]\times[-3,3]\times[-3,3]$), and the observation links are randomly generated between 3D points and cameras. The lost rank of each random network is derived based on the eigendecomposition of FIM through numerical simulation.

Fig. 4 shows the scatter figure and the distribution of the lost rank of the 1000 random networks. As Fig. 4(a) shows, the lost rank of the 1000 random networks are bounded by the upper and the lower bounds given in Proposition 3. Meanwhile, the lost rank of most random networks is located in the region close to the lower bound, which coincides with the distribution in Fig. 4(b). These results mean that extreme structures are rare in actual networks, and thus the observability of most visual SLAM networks will not degrade to the upper bound.

Fig. 5 gives the scatter figure and the distribution of the ill-conditioned score of the 1000 random networks. In Fig. 5(a), the ill-conditioned score of random networks are distributed in the region within the upper and the lower bounds. Besides, when the number of cameras and 3D points increases, the distribution for the ill-conditioned score of random networks approaches the lower bound. This indicates that the ill-conditioned degree is ameliorated since the probability with extreme network structures reduces with more cameras and 3D points. Meanwhile, the distribution of ill-conditioned score in Fig. 5(b) also indicates that the ill-conditioned degree of most visual SLAM networks is at a low level and the observability does not suffer severe degradation.

5. CONCLUSION

In this work we investigated the observability of visual SLAM networks from a perspective of statistical information and demonstrated the dependence of unobserved directions in state space on network structure. Specifically, we first formulated the visual SLAM into a state estimation problem to give the FIM for each visual observation, and also analyzed the impact of 3D point translation and camera motion on observations. Then we investigated the dependence of the observability of visual SLAM on network structure based on the nullspace of the network's FIM, and derived closed-form expressions for the lost rank with its upper and lower bounds. Besides, a metric for observability, the ill-conditioned score, was also proposed to quantitatively evaluate the degradation of visual SLAM performance, which may serve as a guideline for future algorithm design. Numerical results verified our conclusions.

6. REFERENCES

[1] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2012, pp. 3354–3361.

[2] P. Tokekar, J. Vander Hook, D. Mulla, and V. Isler, "Sensor planning for a symbiotic UAV and UGV system for precision agriculture," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1498–1511, 2016.

[3] L. Paull, G. Huang, M. Seto, and J. J. Leonard, "Communication-constrained multi-AUV cooperative SLAM," in *Proc. IEEE Int. Conf. Robot Autom.*, 2015, pp. 509–516.

[4] K. Yousif, A. Bab-Hadiashar, and R. Hoseinnezhad, "An overview to visual odometry and visual SLAM: Applications to mobile robotics," *Intell. Ind. Syst.*, vol. 1, no. 4, pp. 289–311, 2015.

[5] J. Liao, X. Li, X. Wang, S. Li, and H. Wang, "Enhancing navigation performance through visual-inertial odometry in GNSS-degraded environment," *GPS Solutions*, vol. 25, no. 2, pp. 1–18, 2021.

[6] Z. Gong and et al., "Graph-based adaptive fusion of GNSS and VIO under intermittent gnss-degraded environment," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–16, 2021.

[7] M. Bryson and S. Sukkarieh, "Observability analysis and active control for airborne SLAM," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 1, pp. 261–280, 2008.

[8] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, 2007.

[9] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. IEEE ACM Int. Symp. Mixed Augmented Reality*, vol. 3, Nov. 2007, pp. 225–234.

[10] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, 2017.

[11] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: large-scale direct monocular SLAM," in *Proc. Eur. Conf. Comput. Vision*, 2014, pp. 834–849.

[12] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 611–625, 2018.

[13] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: fast semi-direct monocular visual odometry," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2014, pp. 15–22.

[14] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, vol. 1, 2006, pp. 519–528.

[15] Q. An, Y. Wang, and Y. Shen, "Sensor deployment for visual 3D perception: a perspective of information gains," *IEEE Sensors J.*, vol. 21, no. 6, pp. 8464–8478, 2021.

[16] A. Nemra and N. Aouf, "Robust airborne 3D visual simultaneous localization and mapping with observability and consistency analysis," *J. Intell. Robot. Syst.*, vol. 55, no. 4, pp. 345–376, 2009.

[17] Z. Chen, K. Jiang, and J. Hung, "Local observability matrix and its application to observability analyses," in *Proc. Annu. Conf. IEEE Ind. Electron. Soc.*, 1990, pp. 100–103 vol.1.

[18] J. J. Morales and Z. M. Kassas, "Stochastic observability and uncertainty characterization in simultaneous receiver and transmitter localization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 2, pp. 1021–1031, 2019.

[19] Z. M. Kassas and T. E. Humphreys, "Observability analysis of collaborative opportunistic navigation with pseudorange measurements," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 260–273, 2014.

[20] F. Aghili, "Integrating IMU and landmark sensors for 3D SLAM and the observability analysis," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2010, pp. 2025–2032.

[21] K. W. Lee, W. S. Wijesoma, and J. I. Guzman, "On the observability and observability analysis of SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2006, pp. 3569–3574.

[22] Z. Wang and G. Dissanayake, "Observability analysis of SLAM using fisher information matrix," in *Proc. IEEE Conf. Control Autom. Robot.*, 2008, pp. 1242–1247.

[23] R. Hartley and A. Zisserman, Multiple view geometry in computer vision. Cambridge Univ. Press, 2003.

[24] Q. An and Y. Shen, "Camera configuration design in cooperative active visual 3d reconstruction: A statistical approach," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2020, pp. 1–5.

[25] K. M. Steven, Fundamentals of statistical signal processing: Estimation theory. New Jersey, NJ. USA: Prentice Hall, 1993.