Uncertainty Quantification for Multi-Scan Registration

XIANGRU HUANG, The University of Texas at Austin, USA
ZHENXIAO LIANG, The University of Texas at Austin, USA
QIXING HUANG, The University of Texas at Austin, USA

A fundamental problem in scan-based 3D reconstruction is to align the depth scans under different camera poses into the same coordinate system. While there are abundant algorithms on aligning depth scans, few methods have focused on assessing the quality of a solution. This quality checking problem is vital, as we need to determine whether the current scans are sufficient or not and where to install additional scans to improve the reconstruction. On the other hand, this problem is fundamentally challenging because the underlying ground-truth is generally unavailable, and it is challenging to predict alignment errors such as global drifts manually. In this paper, we introduce a local uncertainty framework for geometric alignment algorithms. Our approach enjoys several appealing properties, such as it does not require re-sampling the input, no need for the underlying ground-truth, informative, and high computational efficiency. We apply this framework to two multi-scan alignment formulations, one minimizes geometric distances between pairs of scans, and another simultaneously aligns the input scans with a deforming model. The output of our approach can be seamlessly integrated with view selection, enabling uncertainty-aware view planning. Experimental results and user studies justify the effectiveness of our approach on both synthetic and real datasets.

CCS Concepts:
• Computing methodologies → Point-based models; Shape analysis;

Additional Key Words and Phrases: uncertainty quantification, multi-scan registration, approximation error, view planning

ACM Reference Format:

1 INTRODUCTION

Reconstructing 3D models from depth scans is a fundamental problem in geometry processing and beyond. A standard 3D reconstruction pipeline typically combines a geometry registration phase and a geometry reconstruction phase (c.f. [Huber 2002]). The goal of geometry registration is to bring depth scans obtained from different camera poses into the same coordinate system. The accuracy of this phase largely determines the quality of the reconstructed 3D model. Existing multi-scan registration techniques have predominantly
focused on developing suitable objective terms and novel optimization methods for computing a single registration result (c.f. [Tam et al. 2013]). In contrast, they place minimal efforts on studying an equally important problem of how to quantify the uncertainties of scan registration. Successful solutions to this task allow us to answer critical questions such as whether the current result is accurate, and if not, where to instill additional inputs to improve the quality of scan registration. A prominent example is how to quantify the uncertainties of global pose error, which is very hard to assess manually (particularly for models with non-planar structures).

In this paper, we introduce a local uncertainty quantification (or UQ) framework for multi-scan registration. The key idea is to treat multi-scan registration as solving an optimization problem. This optimization problem implicitly defines a map from the input scans to the output optimized poses. By assuming a noise model for generating the depth scans, this map then enables us to quantify local uncertainties of the optimized scans via propagating uncertainties associated with the input scans. We introduce new uncertainty measures derived from an approximation of the output covariance matrix, i.e., through the inverse Hessian evaluated at the current reconstruction. These uncertainty measures (e.g., diagonal blocks and leading eigenvectors of the approximated covariance matrix) enjoy multiple appealing properties, such as computational efficiency, quantified approximation errors, and information-rich. Compared to prior UQ approaches [Bosse and Zlot 2008; Manoj et al. 2015; Szeliski 1990; Szeliski and Kang 1997; Triggs et al. 1999; Wahba 1983; Zhang 1998] that utilize the inverse Hessian matrix to approximate the output covariance matrix, we provide a first analysis of the approximation errors, from which we develop informative and robust uncertainty measures. Such measures are critical for multi-scan registration, where it is infeasible and unnecessary to visualize the entire predicted covariance matrix.

We have applied this framework under two multi-scan registration settings, i.e., 1) geometric alignments between pairs of scans (or joint pairwise registration (JPR)) [Benjemaa and Schmitt 1998; Besl and McKay 1992; Brown and Rusinkiewicz 2007; Chen and Medioni 1992b; Krishnan et al. 2005; Pottmann et al. 2006; Raghuramu 2015], and 2) aligning the input scans to a deforming model (or simultaneous registration and reconstruction (SRAR)) [Huang et al. 2007a; Huang and Anguelov 2010; Jin et al. 1995; Liu et al. 2006; Tubic et al. 2003]. Under both settings, we develop geometric representations that lead to small discretization errors of our method on discrete objects (e.g., point clouds).

We have evaluated our approach on synthetic and real datasets. On synthetic datasets, experimental results justify that the predicted uncertainty measures match the simulated uncertainty measures obtained from repeated sampling the input depth-scans. On real datasets with ground-truth poses, the predicted covariance matrix is consistent with the pose errors. User studies demonstrate the effectiveness of our UQ approaches for enhancing user experiences in interactive scanning, and our approach accurately reveals the registration errors and significantly outperforms alternative methods.

As an application of our UQ approaches, we consider model-based view planning, which optimizes the camera poses on a similar 3D model to maximize the quality of the resulting 3D reconstruction [Scott et al. 2003]. We introduce a novel model-based view planner, where the objectives are to minimize the length of the scanning trajectory and the trace of the approximated covariance matrix. We present a practical greedy approach to solve the induced optimization problem efficiently. Compared to other model-based view planning approaches, our approach offers direct control of the reconstruction error and also generalizes better when the model and testing instances are different.

In summary, we present the following contributions:

- We define uncertainty measures for multi-scan registration that are computationally efficient, robust, and information-rich.
- We apply the proposed approach under two formulations, namely, joint pairwise registration and aligning scans to a deforming template model.
- We develop efficient approaches to visualize the quantified uncertainties for interactive scanning.
- We demonstrate the usefulness of our approach in interactive quality checking and model-based view planning.

2 RELATED WORKS

It is beyond the scope of this paper to provide a thorough overview of the literature on geometry registration. We refer to [Berger et al. 2014; Tam et al. 2013; Zollhöfer et al. 2018] for standard surveys on these topics. In the following, we discuss works on multi-scan registration and uncertainty quantification, which are most relevant to the context of this paper.

2.1 Multi-Scan Registration

Multi-scan registration seeks to bring depth scans stored in the individual camera coordinate systems into the same coordinate system. Depending on our assumptions about the input depth scans, existing approaches fall into the category of global methods [Arrigoni et al. 2016a,b; Govindu and Pooja 2014; Huang and Guibas 2010, 2013; Huang et al. 2017, 2019; Torsello et al. 2011; Wang and Singer 2013; Zhang et al. 2019], where we do not have any prior knowledge about the poses of the input scans, and the category of local methods [Benjemaa and Schmitt 1998; Brown and Rusinkiewicz 2007; Chen and Medioni 1992b; Huang et al. 2007a; Huang and Anguelov 2010; Krishnan et al. 2005; Pottmann et al. 2006; Raghuramu 2015], where we assume that we have some initial guesses about the underlying ground-truth poses. This paper studies uncertainty quantification for local methods. A standard strategy among local methods is to minimize the cumulative distances among the overlapping regions [Benjemaa and Schmitt 1998; Brown and Rusinkiewicz 2007; Chen and Medioni 1992b; Krishnan et al. 2005; Pottmann et al. 2006; Raghuramu 2015]. Although these methods work well on dense and clean depth scans, where the matches among overlapping regions are accurate, it becomes less reliable for sparse and noisy point clouds. In this case, depending on the structure of the overlapping areas, the alignment errors can accumulate.

Another category of methods unifies registration and reconstruction. The key idea is to optimize scan poses by minimizing the distances between the input scans and a deforming surface model. The representation of this dynamic model includes point clouds [Henry et al. 2012; Keller et al. 2013; Rusinkiewicz et al. 2002; Whelan et al.
Regarding multi-view reconstruction, several works [Kanatani and Morris 2006; Lhuillier and Perriollat 2006; Polie et al. 2018] studied uncertainty quantification under a feature-based structure-from-motion setting. Yet these approaches do not model the strong correlations among the location errors of image feature points, which depend on the camera poses and the underlying shape. In contrast, we follow a correspondence free approach, enabling us to model independent samples.

As the performance of algebraic methods depends on the discretization error, a technical focus of geometric uncertainty analysis is to develop suitable representations and error metrics that are insensitive to such errors. For example, in the application of pairwise geometric registration, the point-2-plane distance metric has proven to be much more accurate than the point-2-point distance metric (c.f. [Landry et al. 2019]). [Pauly et al. 2004] introduced a point-based surface representation to quantify the uncertainties of implicit surfaces derived from a point cloud. [Pothkow et al. 2011] developed a grid-based representation for quantifying the uncertainties of converting an iso-surface into a triangular mesh. In contrast, we focus on modeling multi-scan registration under both the setting of minimizing pairwise distances between overlapping scans and the context of aligning input scans to a deforming template.

2.3 Geometrically Stable Sampling

A relevant task in geometry registration is to analyze the condition number of the Hessian matrix of the objective function. For example, [Gelfand et al. 2003] analyzed the condition number of the Hessian matrix for registering a pair of scans under the point-2-plane distance metric [Chen and Medioni 1992a]. They also introduced a greedy approach that selects a subset of the sample points so that the induced condition number is minimized. The fundamental difference between our UQ approach and the approach of optimizing the condition number is that we focus on the uncertainties of the optimal solution. In contrast, the method of minimizing the condition number can help the convergence to the optimal solution (c.f. [Nocedal and Wright 2006]).

2.4 Model-based View Planning

Besides maximizing the registration and reconstruction quality from a fixed set of scans, optimizing the set of input scans also dictates the reconstruction quality. Model-based view planning seeks to pre-compute a set of camera poses using a similar model, e.g., a coarse reconstruction or a model of the same category, and then uses the planned trajectory for scanning. The planned trajectory can be enforced exactly, e.g., using a camera rig, or showing it to the user as a suggestion for interactive scanning. Model-based view planning has been studied extensively in the literature. We refer to [Scott et al. 2005] for a literature review and to [Fan et al. 2016; Hepp et al. 2018] for recent advances.

In model-based view planning, designing a suitable 3D reconstruction quality score is crucial for both the reconstruction quality and the generalization behavior between training and testing models. This paper shows that when using the trace-norm of the pose covariance matrix, we obtain better reconstruction quality and generalization behavior than prior techniques.

In contrast to developing new multi-scan registration techniques, this paper focuses on the new topic of quantifying the uncertainties of multi-scan registration techniques.

2.2 Uncertainty Analysis in Vision and Graphics

Uncertainty analysis is a long-standing topic in computer graphics and vision. In terms of methodology, present works fall into the category of simulation-based approaches and the group of algebraic methods. Simulation-based methods [Bengtsson and Baerveldt 2003; Iversen et al. 2017; Landry et al. 2019] apply sampling to generate inputs and then calculate statistics (e.g., the covariance matrix) among the induced outputs. These approaches can provide accurate estimations. However, they are computationally expensive, as the same algorithm has to run many times (typically in the order of thousands) on different inputs (c.f. [Landry et al. 2019]).

Algebraic methods [Bengtsson and Baerveldt 2003; Biber and Straßer 2003; Bosse and Zlot 2008; Censi 2007; Manoj et al. 2015; Nieto et al. 2005; Szeliski 1990; Szeliski and Kang 1997; Triggs et al. 1999; Wahba 1983; Zhang 1998] build on a noise model of the input and compute an approximation of the output covariance matrix via the inverse Hessian of the objective function. These approaches are efficient, as they only need to run each algorithm one time. However, they are less accurate than simulation-based methods due to approximation errors of the Hessian matrix and discretization errors (c.f. [Landry et al. 2019]). Our approach, which falls into this category, introduces innovative and rigorous means to address both issues. Specifically, we analyze the approximation errors of the inverse Hessian, from which we derive informative measures of the covariance matrix that are insensitive to these approximations. Besides, we carefully model the objective functions to minimize the effects of the discretization error.

In terms of applications, most existing works have focused on various pairwise alignment problems, e.g., structure-from-motion [Szeliski and Kang 1997; Triggs et al. 1999; Zhang 1998] and ICP registration [Bengtsson and Baerveldt 2003; Biber and Straßer 2003; Bosse and Zlot 2008; Censi 2007; Manoj et al. 2015; Nieto et al. 2005]. In contrast to these works, we focus on multi-scan registration. In this setting, the dimension of the covariance matrix is significantly higher than that of the pairwise scenario, and there are critical questions on what uncertainty measures are both informative and robust, how to compute them efficiently, and how to visualize them for interactive reconstruction.
We will discuss how to address this issue in Section 4.2.

We consider an unconstrained optimization problem, where the Jacobi matrix in the linear map, from which we develop informative uncertainty measures. This strategy allows us to perform uncertainty quantification from a single set of input scans. The sub-figures show norms of the diagonal blocks of the covariance matrices.

3 OVERVIEW

In this section, we provide a roadmap of the technical sections (Section 4 to Section 7) of this paper. As multi-scan registration algorithms typically solve optimization problems to compute the aligned scan poses, we describe in Section 4 a general framework for computing informative and robust uncertainty measures among the outputs of solving unconstrained optimization problems. In Section 5 and Section 6, we apply the proposed UQ framework to two popular formulations for multi-scan registration and reconstruction, namely, minimizing the distances between pairs of overlapping scans (Section 5) and minimizing the distances between the input scans and the deforming template (Section 6). In each setting, we present algorithms for uncertainty quantification and visualization schemes to interpret the quantified uncertainties. Finally, Section 7 presents a novel uncertainty-driven model-based view planner.

4 UNCERTAINTY QUANTIFICATION FOR UNCONSTRAINED OPTIMIZATION

In this section, we describe how to compute informative and robust uncertainty measures of the outputs that are defined by solving optimization problems. Our approach roots in an estimation of the output covariance matrix and an analysis of the approximation error. In the following, we first describe how to compute the approximated covariance matrix in Section 4.1. We then introduce robust and informative uncertainty measures in Section 4.2.

4.1 Implicitly Defined Optimal Solutions

We consider an unconstrained optimization problem, where the objective function admits the form \( f(x, y, w, p^{x}) \). Here \( p^{x} \) denotes the underlying ground-truth; \( y \) encodes the input noise model; \( w \) collects the hyper-parameters of \( f \); \( x \) parameterizes the output variables. In the context of this paper (See Figure 2), \( p^{y} \) denotes the underlying ground-truth scan points (i.e., when there is no noise) in a world coordinate system. Note that \( p^{y} \) is generally unavailable. We will discuss how to address this issue in Section 4.2. \( y \) models the uncertainties when generating the input scans. Each sampling of \( y \) gives one set of input scans for geometry registration. \( x \) collects pose variables (joint pairwise registration and simultaneous registration and reconstruction) and parameters of the underlying 3D model (simultaneous registration and reconstruction). \( w \) encodes the scan indicators, i.e., whether each scan participates in multi-scan registration or not. Note that we treat \( w \) as a constant vector until Section 7, where we study scan selection. However, we keep \( w \) in our notations for the completeness of our approach.

Without losing generality, we place the additional assumption that \( f(x, y, w, p^{x}) \geq 0 \) and \( f(x, y, w, p^{x}) = 0 \) if and only if \( x = 0 \) and \( y = 0 \). In other words, \( x \) encodes deviations from the underlying ground-truth pose, \( y \) characterizes the noise-level, and the underlying ground-truth is recovered when there is no input noise. Under this setup, we formulate the map from \( y \) to \( x \) as solving the following optimization problem:

\[
\begin{align*}
    x^{*}(y, w, p^{x}) &= \arg\min_{x} f(x, y, w, p^{x}).
\end{align*}
\]

Given a probability distribution over \( y \), the map defined in (1) induces a probability distribution over \( x \). Quantifying this distribution to visualize the uncertainties of multi-scan registration introduces several ground challenges. First, \( x^{*}(y, w, p^{x}) \) is a non-linear map of \( y \). Second, the underlying ground-truth \( p^{x} \) is unknown. Finally, for high-dimensional problems such as multi-scan registration, we need concise and informative measures. Our approach is motivated from a line of works [Bengtsson and Baerveldt 2003; Bosse and Zlot 2008; Censi 2007; Manoj et al. 2015; Szeliski and Kang 1997; Wahba 1983; Zhang 1998] that approximates implicitly defined non-linear maps by their linear counterparts. The results lead to approximations of the underlying covariance matrix. A key novelty of this paper is that we analyze the approximation error to obtain accurate, informative, and concise uncertainty measures that are suitable for high-dimensional problems such as multi-scan registration. Another subtle difference is that we consider two approximated covariance matrices, one for computation, and another for the application of model-based view planning (See Section 7). In the following, we describe the approximated covariance matrices. In Section 4.2, we introduce our uncertainty quantification algorithm.

Specifically, we first approximate the non-linear map defined in (1) by its linear approximation, i.e.,

\[
\begin{align*}
    x^{*}(y, w, p^{x}) &\approx \frac{\partial x^{*}}{\partial y}(0, w, p^{x}) \cdot y.
\end{align*}
\]

ACM Trans. Graph., Vol. 39, No. 4, Article 130. Publication date: July 2020.
We can compute the partial derivative \( \frac{\partial x^*}{\partial y}(0, w, p^{GT}) \) via the implicit function theorem [Krantz and Parks 2013]:

\[
\frac{\partial x^*}{\partial y}(0, w, p^{GT}) := -\frac{\partial^2 f}{\partial x^2}(0, 0, w, p^{GT})^{-1}\left(\frac{\partial^2 f}{\partial x^2}(0, 0, w, p^{GT})\right) (3)
\]

To address the second challenge that \( p^{GT} \) is unavailable, we simply replace \( p^{GT} \) by a reconstruction \( p^* \) derived from one set of input scans. This approach leads to the following approximation

\[
x^*(y, w, p^*) = \frac{\partial x^*}{\partial y}(0, w, p^*) \cdot y. (4)
\]

The approximations described in (2) and (4) offer two approximations of the output covariance matrix \( V[x^*] \):

\[
\overline{C} := \frac{\partial x^*}{\partial y}(0, w, p^*) \cdot V[y] \cdot \frac{\partial x^*}{\partial y}(0, w, p^*)^T, (5)
\]

\[
C := \frac{\partial x^*}{\partial y}(0, w, p^*) \cdot V[y] \cdot \frac{\partial x^*}{\partial y}(0, w, p^*)^T. (6)
\]

where \( V[y] \) denotes the input covariance matrix. Our UQ approach will leverage the second covariance matrix \( C \), which only requires a single output [Szeliski and Kang 1997; Wahba 1983; Zhang 1998] (e.g., a single run of multi-scan registration). In contrast, we will use the first covariance matrix \( \overline{C} \) to develop a model-based view planner (e.g., where the underlying ground-truth is known).

When the approximations in (2) and (4) are precise and the elements of \( y \) are independently random variables with zero mean, then \( x^* \) is approximately a Gaussian distribution with zero mean (due to the central limit theorem [Van Der Vaart 1998]). Since the covariance matrix fully characterizes a Gaussian distribution with zero mean, both \( \overline{C} \) and \( C \) are informative in this setting.

For multi-scan registration, both (2) and (4) are inexact. Therefore, it is critical to analyze the approximation errors. In the following, we use one such analysis to derive measures of \( C \) that are informative and insensitive to the approximation errors. Note that such uncertainty measures are also preferred for multi-scan registration, as it is infeasible to visualize any entire covariance matrix.

4.2 Informative Uncertainty Measurements

To understand the approximation errors of \( C \) and \( \overline{C} \), we study the following model for generating the input noise. Specifically, we assume that the elements of \( y \) are independent, and each \( y_i \) satisfies

\[
\text{Expectation} : \quad E[y_i] = 0 \quad (7)
\]

\[
\text{Variance} : \quad V[y_i] = \sigma^2 \quad (8)
\]

\[
\text{Bounded} : \quad |y_i| \leq \delta \sigma \quad (9)
\]

for some \( \delta > 1 \). Note that this model does not require the distribution of each \( y_i \) to be a Gaussian. The following proposition states a bound on the difference between \( C \) (\( \overline{C} \)) and the output covariance matrix \( V[x^*] \) with respect to the matrix spectral norm.

**Proposition 1.** (Informal) Under the noise model described above and mild conditions on the derivatives of \( f \), we have

\[
\frac{1}{\sigma^2} \max \{||C - V[x^*]|, ||\overline{C} - V[x^*]|\} \leq \sigma C(\kappa(f_{xx}), \delta) \quad (10)
\]

with high probability. Here \( || \cdot || \) denotes the spectral norm; \( f_{xx} = \frac{\partial^2 f}{\partial x^2}; C \) depends on \( \delta; \kappa(f_{xx}) := \frac{\lambda_{\text{max}}(f_{xx})}{\lambda_{\text{min}}(f_{xx})} \) denotes the ratio between the maximum and the minimum eigenvalues of \( f_{xx} \), i.e., its the condition number. The smaller \( \delta \) and \( \kappa(f_{xx}) \), the bigger \( C(\kappa(f_{xx}), \delta) \).

The formal statement of Prop. 1 and the proof are derived from the Taylor expansions of \( x^*(y) \) and \( f \). We defer the details to the supplemental material.

**Remark.** Prop. 1 states that the approximation error is dictated by the condition number of \( f_{xx} \). In the context of multi-scan registration, this condition number connects to the sparsity pattern of \( f_{xx} \), which depends on how the input scans overlap with each other. As we will see in Section 8, an implication is that when the overlapping relations among the input scans are strong, the approximation error of \( C \) is small. On the other hand, for thin objects or when input scans form strongly overlapping components, the approximation error of \( C \) can be large.

Motivated from Prop 1, we introduce two groups of measures to access the uncertainties of \( x^* \):

- **Diagonal blocks.** The first group consists of diagonal blocks of \( C \), each of which corresponds to one group of variables, e.g., 6 × 6 blocks for scan poses. These measures are accurate, as the spectral norms of the deviations in diagonal blocks are upper bounded by those of the entire matrices. They are also easy to compute. On the other hand, they do not capture the correlations among uncertainties of multiple scans.

- **Leading eigenvectors.** The second group consists of leading eigenvectors of \( C \), which capture structural correlations among groups of variables, e.g., clusters of scans that strongly overlap with each other.

The motivation comes from the fact that leading eigenvectors are also stable under perturbations. Specifically, Given two symmetric matrices \( A, E \in \mathbb{R}^{n \times n} \), a variant of the Davis-Kahan theorem [Davis and Kahan 1970] states that the distance between the \( k \)-th eigenvector \( u_k \) of \( A \) and the eigenspace \( U_k \in \mathbb{R}^{n \times k} \) formed by the top \( k' > k \) eigenvectors of \( A + E \) is bounded as \( ||u_k - (U_k \cdot U_k^T) u_k||^2 \leq \|E\|/\|\lambda_k - \lambda_{k'} - \|E\|\| \) where \( \lambda_k \) is the \( k \)-th eigenvalue of \( A \). In other words, the top eigenvectors of \( V[x^*] \) are well approximated by their projections in the leading eigenspace of \( C \).

5 JOINT PAIR-WISE REGISTRATION

In this section, we apply the framework described in Section 4 to joint pairwise registration (or JPR), i.e., minimizing distances between all pairs of overlapping scans. We begin by describing the formulation of JPR in Section 5.1. We then discuss the uncertainty quantification setup in Section 5.2. Finally, we discuss how to visualize the quantified uncertainties in Section 5.3.

5.1 Formulation of JPR

This formulation considers the setting of aligning \( N \) scans \( S = \{S_1, \cdots, S_N\} \). Each scan \( S_i \) is a set of \( p_{i1}, \cdots, p_{iN_i} \) consisting of a set of surfels. \( p_{ij} \) and \( n_{ij} \) denote the position and the normal of \( p_{ij} \), respectively. Let \( T_i = (R_i, t_i) \) be the rigid pose associated
5.2 Uncertainty Quantification

We first describe a noise model for generating the input scans. We then introduce our UQ algorithm (See Section 5.2.2).

### 5.2.1 Noise model

Figure 3(Right) illustrates the uncertainty quantification setup for joint pair-wise registration. We model the input scans as perturbations from a latent surface model. Specifically, consider a latent surface model represented as a set of surfels \( \{s_k = (q_k, n_k), 1 \leq k \leq K \} \), where \( q_k \) and \( n_k \) denote the position and the normal of \( s_k \), respectively. Let \( k_{ij} \in \{1, \cdots, K\} \) be the index of the corresponding surfel for scan point \( p_{ij} \). With \( p_{ij}^g \) we denote the ground-truth location for \( p_{ij} \), which is close to \( q_{k_{ij}} \) and lies on the plane specified by \( (q_{k_{ij}}, n_{k_{ij}}) \), i.e., \( (p_{ij}^g - q_{k_{ij}})^T n_{k_{ij}} = 0 \). We model \( p_{ij} \) and \( n_{ij} \) so that they perturb \( p_{ij}^g \) and \( n_{k_{ij}} \), respectively:

\[
\begin{align*}
    p_{ij} &= p_{ij}^g + y_{ij} n_{k_{ij}}, \\
    n_{ij} &= n_{k_{ij}} + z_{ij}
\end{align*}
\]  

where \( y_{ij} \in \mathbb{R} \) and \( z_{ij} \in \mathbb{R}^3 \) encode the noises in positions and normals, respectively. In our UQ approach, we assume \( \{y_{ij}\} \cup \{z_{ij}\} \) are independent with each other, and they satisfy

\[
E[y_{ij}] = 0, V[y_{ij}] = \sigma^2, \quad E[z_{ij}] = 0, V[z_{ij}] = \Sigma_n.
\]

where \( \Sigma_n \in \mathbb{R}^{3 \times 3} \) is a pre-defined covariance matrix.

In our experiments, we set \( \sigma \) and \( \Sigma_n \) as the variance of point-2-plane distance and the variance of the normal differences, respectively. We evaluate these variances among all correspondences in \( \{C_{ij}\} \). We them divide both variances by two to predict the variances from the latent surface model.

For uncertainty quantification, we place one additional assumption. Define the induced correspondences between \( S_i \) and \( S_i' \) as

\[
C_{ii}' := \{(p_{ij}, p_{ij}')|1 \leq j \leq N_i, 1 \leq j \leq N_i', k_{ij} = k_{ij}'\}.
\]

We assume the correspondence set \( C_{ii}' \) employed in joint pairwise registration is contained in \( C_{ii}^g \), i.e., \( C_{ii}' \subset C_{ii}^g \). In other words, if \( (p_{ij}, p_{ij}') \in C_{ii}' \) then \( k_{ij} = k_{ij}' \). As analyzed in [Pottmann and Hofer 2002; Pottmann et al. 2004], this is a feasible assumption as the squared point-2-plane distance is a quadratic approximation of the squared distance function.
Uncertainty Quantification for Multi-Scan Registration

Fig. 4. Visualizing pose uncertainties of reconstructing a wrench model under joint pairwise registration (or JPR). We show the input model, the aligned scans, the diagonal blocks, and three leading eigenvectors. The top row shows the predicted results, and the bottom row shows the simulated results. We can see that the predicted uncertainties are consistent with the intuition that this model possesses planar faces that can glide when performing registration. Such planar face structures are revealed in the leading eigenvectors. Also, our approach reveals high-frequency signals of the simulated covariance matrix.

Let \( C = \bigcup_{(i, i') \in E} C_{i i'} \) collect all point-wise correspondences. Introduce

\[
C_m := \sum_{(p_{ij}, p'_{ij}) \in C} (h_{ij} - h'_{ij})(h_{ij} - h'_{ij})^T \\
A_{dis} := \sigma \cdot \sum_{(p_{ij}, p'_{ij}) \in C} (h_{ij} - h'_{ij})n_{ij}^T((e_{ij} - e'_{ij})^T \otimes I)
\]

where \( e_{ij} \in \mathbb{R}^{\Sigma N_i} \) is the canonical basis with the \( ij \)-th element being one and \( e_{ij} \otimes I \in \mathbb{R}^{[\Sigma N_i] \times 3} \). Then we have

\[
C_{\text{JPR}} = C_m^{-1} A_{\text{dis}} A_{\text{dis}}^T C_m^{-1}.
\]

**Proof.** Please refer to the supplemental material. \(\square\)

**5.3 Uncertainty Visualization**

As described in Section 4, we interpret \( C_{\text{JPR}} \) using its diagonal blocks and leading eigenvectors. This section discusses how to visualize these quantities. Let \( C_{\text{JPR}}^{(i)} \in \mathbb{R}^{6 \times 6}, 1 \leq i \leq N \) denote the diagonal block that corresponds to scan \( S_i \); we convert it into an uncertainty value for scan \( S_i \). To this end, we first compute its leading eigenvalue \( \lambda_{\text{max}}^{(i)} \in \mathbb{R} \) and leading eigenvector \( u_{\text{max}}^{(i)} \in \mathbb{R}^6 \), respectively. Decompose \( \sqrt{\lambda_{\text{max}}^{(i)}} u_{\text{max}}^{(i)} = (\tilde{c}_i; c_i) \), where \( \tilde{c}_i \) and \( c_i \) correspond the translation and rotation components, respectively. We then define the uncertainty value for \( S_i \) as

\[
\sigma_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \| \tilde{c}_i + c_i \times p_{ij} \|
\]

which is the average displacement among \( S_i \) incurred by \( (\tilde{c}_i; c_i) \). As shown in Figure 4, we reconstruct a mesh from the aligned scans and visualize \( \sigma_i, 1 \leq i \leq N \) as a scalar field on the reconstructed mesh. This is done by performing a range query on the aligned scans from each mesh vertex and then computing the mean among
The input to SRAR consists of \( \text{c.f.}\ [\text{Gelfand and Guibas} \ 2004] \). We can see that the diagonal blocks \( AC \)M Trans. Graph., Vol. 39, No. 4, Article 130. Publication date: July 2020.

Section 5 to represent the latent surface as a collection of surfels aligning each scan with the latent surface. We adapt the setup in the latent surface model and how to define the distance metric for potential, SRAR does not require point normals as input.

6.1 Formulation of SRAR

In this section, we apply the proposed framework to the setting of simultaneous registration and reconstruction (or SRAR) [Huang et al. 2007a; Huang and Anguelov 2010]. A standard reconstruction pipeline usually consists of two sequential steps, where the first step performs joint pairwise registration, and where the second step reconstructs a 3D model from the aligned scans. In contrast, SRAR unifies registration and reconstruction by solving a single optimization problem. The key idea is to align the input scans to a latent surface model, which is also jointly optimized with the scan poses. In the following, we first describe the formulation of SRAR in Section 6.1. We then introduce the uncertainty quantification formulation for SRAR in Section 6.2.

Fig. 4 shows visualizations of a wrench model that possesses planar structures. Intuitively, scans of these planar faces can drift (c.f. [Gelfand and Guibas 2004]). We can see that the diagonal blocks and the leading eigenvectors nicely reveal such structural patterns.

6 SIMULTANEOUS REGISTRATION AND RECONSTRUCTION

In this section, we apply the proposed framework to the setting of simultaneous registration and reconstruction (or SRAR) [Huang et al. 2007a; Huang and Anguelov 2010]. A standard reconstruction pipeline usually consists of two sequential steps, where the first step performs joint pairwise registration, and where the second step reconstructs a 3D model from the aligned scans. In contrast, SRAR unifies registration and reconstruction by solving a single optimization problem. The key idea is to align the input scans to a latent surface model, which is also jointly optimized with the scan poses. In the following, we first describe the formulation of SRAR in Section 6.1. We then introduce the uncertainty quantification formulation for SRAR in Section 6.2.

6.1 Formulation of SRAR

The input to SRAR consists of \( N \) depth scans \( S_i = \{ p_{ij}, 1 \leq j \leq N_j \}, 1 \leq i \leq N \). Note that unlike joint pairwise registration (or JPR) which relies on the point normals to formulate the registration potential, SRAR does not require point normals as input.

The key difference between SRAR and JPR lies in how to represent the latent surface model and how to define the distance metric for aligning each scan with the latent surface. We adapt the setup in Section 5 to represent the latent surface as a collection of surfels \( s_k = (q_k, n_k), 1 \leq k \leq K \). As we will use the point-2-plane distance metric [Chen and Medioni 1992a] to align the input scans with the latent surface, each surfel possesses three independent variables, i.e., the corresponding plane equation \( (d_k, n_k) \), where \( d_k \) denotes the distance from the origin to \( s_k \).

Let \( k_{ij} \in \{1, \cdots, K\} \) denote the corresponding plane of point \( p_{ij} \) (i.e., the one with closest surfel position \( q_{k_{ij}} \)). SRAR amounts to optimize the pose \( (R_i, t_i) \) of each scan \( S_i \) and the plane equations \( \{(d_k, n_k)\} \) by solving the following optimization problem:

\[
\begin{align*}
\arg\min_{\{R_i, t_i, \{(d_k, n_k)\}\}} & \sum_{i=1}^{N} \sum_{j=1}^{N_i} \left( (R_i p_{ij} + t_i)^T n_{k_{ij}} - d_{k_{ij}} \right)^2 \\
\text{subject to } & \quad R_i = I_3, t_i = 0.
\end{align*}
\]

In other words, (19) utilizes the point-2-plane distance metric to align the data points with the latent surface model. As discussed in [Huang and Anguelov 2010], (19) admits a simple and effective minimization strategy, which alternates between optimizing \( (R_i, t_i) \) and \( \{(d_k, n_k)\} \).

**Remark 3.** In the same spirit as ICP registration [Besl and McKay 1992], we also update the associations between the scan points and the latent planes by recursively finding the closest surfel positions. Such operations are particularly useful when the input scans are not aligned. To achieve fast convergence, we can also use a coarse latent surface and gradually refine it when scans become more aligned. Fig. 5(Left) shows a typical optimization procedure for SRAR on the same dataset shown in Fig. 3(Left). However, when performing uncertainty quantification, we assume these associations are fixed.

6.2 Uncertainty Quantification Formulation

We proceed to apply the proposed framework described in Section 4 to the objective function of SRAR.
6.2.1 Noise model. We employ a similar noise model as the one used in JPG (See Figure 5(Right)). Specifically, consider a set of ground-truth planes \((d_{k,i}^{gt}, n_{k,i}^{gt})\). Let \(p_{k,i}^{gt}\) be the ground-truth location for scan point \(p_{ij}\). With \(k_{ij}\) we denote the index of the corresponding plane of \(p_{ij}\):

\[
\begin{align*}
  n_{k_{ij}}^{gt} & = \sum_{k=1}^{K} d_{k_{ij}}^{gt} n_{k_{ij}}^{gt} = 0. \\
  \end{align*}
\]

The input positions \(\{p_{ij}\}\) are generated by following (13). Unlike JPR, we compute \(\sigma\) as the variance of the point-2-plane distance between the optimized scan points and their corresponding planes.

6.2.2 Formulation. In the same spirit as JPG, we assume that for each sampled input described above, SRAR converges to the underlying point-plane associations \(\{k_{ij}\}\). In practice, we found that the discretization error of the proposed SRAR formulation is small. Intuitively, point-2-plane distances are insensitive to wrong point-plane associations when the planes are compatible with each other locally.

To apply (4), the remaining task is to convert (19) into an unconstrained optimization problem by using the following local parameterizations for rotations and scan poses:

\[
\begin{align*}
  R_i &= \exp(c_i \times) \quad 1 \leq i \leq N \\
  n_i &= \frac{n_i^{gt} + x_{k_{ij}}^{gt} n_{k_{ij}}^{gt}}{\|n_i^{gt} + x_{k_{ij}}^{gt} n_{k_{ij}}^{gt}\|} \\
  d_i &= d_i^{gt} + x_{k_{ij}}^{gt} 1 \leq k \leq K
\end{align*}
\]

where \((n_i^{gt}, t_i^{gt}, t_{k_{ij}}^{gt})\) forms a local coordinate system.

Applying (3) to the objective function (19) of SRAR, we arrive at the following proposition for computing the approximated covariance matrix.

**PROPOSITION 3.** Introduce sparse coefficient vectors \(g_{ij} \in \mathbb{R}^{6(N-1)+3K}\), \(1 \leq i \leq N, 1 \leq j \leq N_i\), whose non-zero elements are given by

\[
\begin{align*}
  V_{(6(i-1))(6(i-1)-6)}(g_{ij}) &= n_{k_{ij}}^{gt} \quad 2 \leq i \leq N, 1 \leq j \leq N_i \\
  V_{(6(i-1)+6)(6(i-1)+6)}(g_{ij}) &= n_{k_{ij}}^{gt} \times n_{k_{ij}}^{gt} \quad 2 \leq i \leq N, 1 \leq j \leq N_i \\
  V_{(6(i-1)+3k_{ij})(6(i-1)+3k_{ij})}(g_{ij}) &= t_{k_{ij},1}^{gt} \times p_{k_{ij}}^{gt} \quad 1 \leq i \leq N, 1 \leq j \leq N_i \\
  V_{(6(i-1)+3k_{ij}+1)(6(i-1)+3k_{ij}+1)}(g_{ij}) &= t_{k_{ij},2}^{gt} \times p_{k_{ij}}^{gt} \quad 1 \leq i \leq N, 1 \leq j \leq N_i \\
  V_{(6(i-1)+3k_{ij}+2)(6(i-1)+3k_{ij}+2)}(g_{ij}) &= -1, \quad 1 \leq i \leq N, 1 \leq j \leq N_i
\end{align*}
\]

Denote

\[
C_{XX} := \sum_{i=1}^{N} \sum_{j=1}^{N_i} g_{ij} g_{ij}^T
\]

Then the uncertainty quantification approach described above outputs the following approximated covariance matrix:

\[
C_{SRAR} = \sigma^2 C_{XX}^{-1}
\]

**PROOF.** Please refer to the supplemental material. \qed

6.2.3 Uncertainty Visualization. Uncertainty visualization employs the same visualization schemes described in Section 5.3. As shown in Figure 6(Left), our predicted covariance matrices are consistent with the simulated covariance matrices. Another observation is that the predicted diagonal blocks of SRAR have smaller norms than those of JPR (See Figure 6(Right)). An interpretation is that...
JPR is sensitive to the accuracy of point normals. In contrast, SRAR does not require point normals.

7 UNCERTAINTY-DRIVEN VIEW-PLANNING

In this section, we introduce a model-based view-planning approach. Generally speaking, a model-based view-planner takes a 3D model as input and outputs a set of camera poses. These camera poses typically optimize a two-term objective function (c.f. [Scott et al. 2003]), where the first term characterizes a coverage score of the camera poses and where the second term measures the length of the shortest path that connects the camera poses. Our model-based view planner replaces the coverage score by an uncertainty score derived from the approach described in Section 6. This formulation allows us to plan views to minimize the reconstruction error of the resulting scans directly. In the following, we first describe the uncertainty score in Section 7.1. We then describe our model-based view planner in Section 7.2.

7.1 An Uncertainty Score

This section defines an uncertainty score among scans \( S \), \( 1 \leq i \leq n \) of a model \( M \), where \( S_i \) are generated from pre-defined camera poses \( C_i \) using the noisy model described in Section 6. To model the process of selecting a subset of scans for reconstruction, we consider a reweighted total objective function under SRAR:

\[
\begin{align*}
  f_{\text{weighted}} & := \sum_{i=1}^{n} w_i \sum_{j=1}^{n} ((R_t p_{ij} + t_i)^T n_{k_{ij}} - d_{k_{ij}})^2 \\
  & \text{(23)}
\end{align*}
\]

where \( w_i \) is an indicator that specifies whether \( S_i \) is selected or not. We seek to set \( w_i \) as either 0 or 1.

Applying (5) to (23), we obtain the following approximated covariance matrix among both the scan poses and the reconstruction:

\[
C_{\text{SRAR}}(w, p^f) = \left( \sum_{i=1}^{n} w_i G_i G_i^T \right)^{-1} \left( \sum_{i=1}^{n} w_i^2 G_i G_i^T \right) \left( \sum_{i=1}^{n} w_i G_i G_i^T \right)^{-1}
\]

where \( G_i = (g_{ij}) \) collects vectors \( g_{ij} \) associated with \( S_i \) in its columns. Note that since the ground-truth model \( M \) is present, we define \( C_{\text{SRAR}}(w, p^f) \) with respect to the ground-truth scan points, i.e., each \( S_i \) stays on \( M \).

The following theorem states an important property of \( C_{\text{SRAR}}(w, p^f) \), which sheds lights on the proposed uncertainty score that we will introduce.

**Theorem 1.** Suppose

\[
\sum_{i=1}^{n} w_i G_i G_i^T > 0, \quad \sum_{i=1}^{n} G_i G_i^T > 0.
\]

Then

\[
C_{\text{SRAR}}(w, p^f) \geq \left( \sum_{i=1}^{n} G_i G_i^T \right)^{-1}.
\]

In particular, equality holds if \( w_1 = w_2 = \cdots = w_n \).

**Proof:** Please refer to the supplemental material. \( \square \)

**Remark 4.** Theorem 1 states that adding more scans always reduces variance, i.e., when scan weights of active scans are identical. This property motivates us to use \( C(w, p^f) \) to define the quality score.

We proceed to define the quality score as a suitable function of the covariance matrix.

**Corollary 1.** Consider a broad range of matrix functions \( \mathcal{H} \) that include polynomials of eigenvalues of principal sub-matrices with non-negative weights, e.g., trace or determinant of a principal sub-matrix. Then for each \( h \in \mathcal{H} \), we have

\[
h(C_{\text{SRAR}}(w, p^f)) \geq h\left( \sum_{i=1}^{n} G_i G_i^T \right)^{-1}.
\]

Following Corollary 1, we define the uncertainty score for view-planning as the trace of the sub-matrix \( C_{\text{pose}}(w, p^f) \in \mathbb{R}^{(N-1)\times(N-1)} \) of \( C_{\text{SRAR}}(w, p^f) \) that corresponds to scan poses:

\[
f_{\text{uncertainty}}(w) := \text{Trace}(C_{\text{pose}}(w, p^f)).
\]

7.2 A Greedy Uncertainty-Driven View Planner

Besides \( f_{\text{uncertainty}}(w) \), which evaluates the reconstruction quality from a subset of scans specified by \( w \), another essential goal in...
model-based view planning is to minimize the length of the path that connects the selected views. While this quantity, i.e., a solution to the travel salesman problem (or TSP), admits integer programming relaxations. We found that it is challenging to solve the induced global optimization problem. Fundamental challenges are 1) the linear programming relaxations are loose, and 2) rounding fractional solutions into integer solutions effectively. To address these issues, we utilize a greedy approach in this paper.

Specifically, let $f_{\text{length}}(w)$ denote the length of the trajectory among selected cameras specified by $w$. We use a greedy approach to determine the best binary indicator vector $w$ that minimizes the following cumulative objective function:

$$
\min_{w \in \{0,1\}^n} \text{f}_{\text{uncertainty}}(w) + \lambda f_{\text{length}}(w)
$$

where $\lambda$ is a tradeoff-parameter that balances the reconstruction error and the length of the TSP path. Starting from $w_i = 1, 1 \leq i \leq n$, we calculate the value of the objective function in (28) when removing each scan. We then remove the scan that leads to the smallest objective value after removing that scan. This procedure is iterated when we cannot reduce the objective function further. In our implementation, we use the Christofides’s algorithm [Christofides 1976] for computing $f_{\text{length}}(w)$.

In practice, we found that this simple strategy leads to high-quality solutions. The total running time for a model with 100K latent planes and 500 camera poses is between 110 seconds to 210 seconds on a desktop with a dual-core 3.4GHz CPU and 128G main memory. As shown in Figure 7, increasing the value of $\lambda$ leads to smoother paths but with increased reconstruction errors.

8 EXPERIMENTAL EVALUATION

This section presents experimental evaluations of the proposed UQ and view-planning methods. In Section 8.1 and Section 8.2, we describe experimental evaluations of our UQ approaches on synthetic examples and real examples, respectively. The goal of evaluating on synthetic examples is to assess the approximation errors of the proposed UQ techniques empirically. In contrast, experimental evaluations on real examples study whether the quantified uncertainties match the pose and reconstruction errors compared to the underlying ground-truth. In Section 8.3, we present a user study to gauge the effectiveness of the proposed UQ approaches on model checking. In Section 8.4, we present a user study on the model-based view-planning approach described in Section 7.

8.1 Experimental Evaluations on Synthetic Datasets

In this section, we first describe the experimental setup. We then analyze the experimental results on approximation errors of the proposed UQ approaches. We also quantitatively compare the accuracies of the two multi-scan registration strategies discussed in this paper, i.e., SRAR and JPR.

8.1.1 Experimental Setup. We have collected a synthetic dataset that consists of four categories of models (two models per category), namely, Plant, Engine, Airplane, and Bridge (See Figure 8). These models cover a diverse range of model scales. 3D reconstructions of these models have direct scientific and industrial impacts. For example, reconstructing plants over time helps us to understand the internal mechanisms of biological forms (c.f. [Li et al. 2013]). In this scenario, quantified uncertainties of the reconstructions are critical for analyzing growing patterns. The Engine, Airplane, and Bridge categories collect diverse man-made objects, where uncertainty quantification can significantly enhance all phases of interactive scanings, such as view planning and quality checking. The resulting accurate reconstructions offer various downstream applications, e.g., damage detection for engines, airplanes, and bridges.

The camera locations of the synthetic scans are uniformly among the viewing sphere centered at the origin of each model. We set the viewing sphere’s radius as a random number between two times and four times the diameter of each model. Moreover, the upright direction of the camera always passes through the upright orientation of each model. In total, we place 50 scans for Plant and Engine models, 200 scans for Airplane models, and 500 scans for Bridge models (the number of scans increases as the model size increases).

When generating the input scans, we choose three noise ratios, i.e., $\epsilon \in \left\{\frac{1}{1000}, \frac{1}{500}, \frac{1}{200}\right\}$, where $L$ is the diameter of that model in the Euclidean distance. At each level, we place 10000 sets of depth scans independently. Each scan point is independently perturbed from one set to another, their overlapping patterns are fixed.

When generating the input to JPR and SRAR, we randomly perturb the pose of each input scan by a random rotation specified by $\exp(-c\epsilon x)$ (where $x$ is a random vector in $[-0.02, 0.02]^3$) and a random translation in $[-4\epsilon, 4\epsilon]^3$. Since the ground-truth correspondences are not given in these experiments, they also evaluate the discretization errors of the proposed approaches.

8.1.2 Analysis of Approximation Errors. We first analyze the approximation errors of our UQ approaches by comparing the diagonal blocks and the leading eigenvectors of the simulated covariance matrix with those predicted by our approach. For each predicted diagonal block $A_{\text{pred}}$ and that obtained from the simulated co-variance matrix $A_{\text{simul}}$, we report the relative error as $\frac{|A_{\text{pred}} - A_{\text{simul}}|}{|A_{\text{simul}}|}$. We then plot the cumulative distribution of relative errors for each approach per model.

The evaluation protocol for the leading eigenspace is similar. For each eigenvector $v$ of a simulated covariance matrix $C$, we first compute $v_{\text{perp}} = \arg \min_{v_{\perp} \in \mathcal{Q}_C} \|v_{\perp} + v\|^2$, where $\mathcal{Q}_C$ is the leading eigenspace of the corresponding predicted covariance matrix. Then for all the leading eigenvectors $v$, we collect statistics on the relative errors for all the corresponding sub-vectors of $v$ and $v_{\text{perp}}$, i.e., $\|v_{\text{perp}} - v_i\|/\|v_i\|$. Intuitively, these relative errors characterize the coverage of the leading eigenspace of the predicted covariance matrix $C$ on that of the simulated covariance matrix $\hat{C}$. They are zero if and only if these two eigenspaces coincide. The same as diagonal blocks, we plot the cumulative distribution of relative errors for all eigenvectors of the simulated covariance matrix.

The sixth and the seventh columns of Figure 8 show cumulative distributions of the approximation errors of the predicted diagonal blocks and the predicted eigenspace. Note that we conducted the experiments under both JPR and SRAR while varying the noise levels
Fig. 8. Each block shows the experimental results on one synthetic model. (From left to right): A sample set of the input scans, quantified uncertainties derived from the diagonal blocks of the predicted covariance matrix under JPG, corresponding quantified uncertainties of the simulated covariance matrix under JPG, quantified uncertainties derived from the diagonal blocks of the predicted covariance matrix under SRAR, corresponding quantified uncertainties of the simulated covariance matrix under SRAR, approximation errors of our approach under JPR, and approximation errors of our method under SRAR.
of the input scans. Overall, our UQ approaches offer accurate predictions of the underlying uncertainty measures. For JPR, the mean relative errors range from 0.108 to 0.171 for diagonal blocks and from 0.158 to 0.216 for the leading eigenspace, respectively. For SRAR, the mean relative errors range from 0.117 to 0.182 for diagonal blocks and from 0.149 to 0.228 for the leading eigenspace, respectively. These quantitative results are consistent with the qualitative results (from the second to the fifth columns of Figure 8), which demonstrate that the predicted results offer meaningful approximations of the corresponding simulated ones.

We can also observe that the relative errors increase when the noise level increases. Moreover, the relative errors increase when the shapes become thin and long. For example, the approximation errors on Airplane and Bridge are more significant than those on Plant and Engine. Intuition is that uncertainties on Airplane and Bridge are more prominent, and the linear map approximation is less accurate. However, even at the highest noise level, the relative errors of the diagonal blocks and the leading eigenspace are below 0.250 under both settings. Such results justify that uncertainty measures derived from diagonal blocks and leading eigenvectors of the predicted covariance matrix are robust, informative, and efficient.

8.1.3 Multi-Scan Registration Comparison. Another interesting question is: how do the two registration methodologies compare against each other in terms of the registration error? As compared in the third and the fifth columns of Figure 8, the approach of simultaneous registration and reconstruction (or SRAR) yields better registration results than joint pairwise registration (or JPR). The simulated uncertainties of SRAR are consistently smaller than those of JPR. On Airplane and Bridge, the performance of SRAR is considerably better than JPR. Again, the advantage of SRAR is that it does not utilize point normals, which may be noisy.

8.2 Experimental Evaluations on Real Datasets

In this section, we evaluate the proposed UQ approaches on real datasets. Similar to Section 8.1, we first describe the experimental setup. We then analyze the experimental results of UQ and compare the two registration approaches.

8.2.1 Experimental Setup. We have collected a real benchmark dataset that consists of four categories of models, namely, Teeth, MechPart, ScanNet [Dai et al. 2017a], and Waymo [Sun et al. 2019] (See Figure 9). Similar to the synthetic dataset, 3D reconstructions of these models enjoy rich medical and industrial applications. Applications of 3D teeth reconstructions include surgery planning for implanting and orthodontics [Yau et al. 2014]. Both applications benefit from certified 3D reconstructions as input. 3D reconstructions of mechanical parts are used for model inspection, in which quantified uncertainties of 3D reconstructions enable us to evaluate the manufacturing error. ScanNet and Waymo are used to assess the performance of our approach to large-scale scenes. Such reconstructions are essential for applications of virtual reality and 3D mapping.

Each model in this dataset is associated with a set of input scans and ground-truth poses and reconstructions. The scans of Teeth and MechPart come from an Artec Space Spider scanner. The number of scans per model ranges from 300 to 500. For both Teeth and MechPart, the underlying ground-truth models come from an Artec Micro high precision industrial scanner. We obtain the ground-truth pose of each input scan by aligning it with the underlying ground-truth model. For ScanNet and Waymo, we use the scans provided by the original datasets. Similar to the synthetic dataset, the overlapping graph among the input scans of each model exhibits great diversity, e.g., from sparse graphs to dense graphs.

For Teeth and MechPart, we use the ground-truth poses and reconstructions described above for experimental evaluation. Since scenes of ScanNet do not possess ground-truth reconstructions, we utilize planar areas of each scene and evaluate deviations of the data points in each planar area from the corresponding fitted plane. For models in the Waymo dataset, we directly assess the pose error using the provided ground-truth poses.

8.2.2 Analysis of UQ Results. The setting of real scans differ from that of synthetic scans in two ways: 1) we do not know the noise model for generating the input scans, and 2) we do not have an exact underlying ground-truth. Therefore, we employ a different evaluation protocol for quantitative evaluation. Specifically, for each reconstruction, we compare it against the approximated underlying ground truth (i.e., obtained using the high-precision industrial scanner) to pose errors and shape errors. We then evaluate whether the quantified uncertainties are consistent with these errors, as detailed below.

We first consider the cumulative distribution of the relative difference between the ground-truth pose error described above and the predicted variance vector associated with each scan pose. The statistics collect the results of 10000 experiments, i.e., we pick 200 subsets of scans per model, where each subset randomly picks 90% of the input scans 50 times. Note that given two vectors $a, b$, their relative difference is given by $\frac{\|a - b\|}{\max(\|a\|, \|b\|)}$. As shown in the third column of Figure 9, most normalized differences fall between 0.042 to 0.225. In other words, the predicted pose variance per scan nicely indicates the underlying ground-truth pose error per scan.

We proceed to analyze whether our UQ approaches capture the structural patterns among the ground-truth pose errors. To this end, we compute the L2 projection of the ground-truth pose error vector onto the leading eigenspace computed using our method. We then calculate the relative projection error between each pose error vector and its projection. As shown in the fourth column of Figure 9, most projection errors fall between 0.052 to 0.238. In other words, the predicted leading eigenspace provides a meaningful approximation of the simulated eigenspace.

8.2.3 Multi-Scan Registration Comparison. As shown in Figure 9, the pose error of SRAR is usually smaller than that of JPR. For example, the mean rotation/translation errors when using JPR are $0.29^\circ/59\text{micron}$, $0.32^\circ/87\text{micron}$, and $0.51^\circ/32\text{mm}$ on Teeth, MechPart, and Waymo, respectively. Likewise, the corresponding errors when using SRAR are $0.19^\circ/27\text{micron}$, $0.27^\circ/39\text{micron}$, and $0.31^\circ/13\text{mm}$, respectively. The relative error reductions range from 12.1% to 52.3% on these datasets.
Fig. 9. Each block shows the experimental results on one real model. (From left to right): A sample of the input scans, quantified uncertainties derived from the diagonal blocks of the predicted covariance matrix under JPG, the ground-truth reconstruction error under JPG, quantified uncertainties under SRAR, the ground-truth reconstruction error under SRAR, approximation errors of the pose components of the covariance matrices (JPR and SRAR), and approximation errors of the shape component of the covariance matrix (SRAR).
We proceed to evaluate the effects of the proposed UQ approaches applying to the optimized latent surface. For SRAR, the reconstruction applies to the aligned data points. For JPR, we employed Poisson Surface Reconstruction [Kazhdan et al. 2006] to obtain the reconstructions. For JPR, the reconstruction applies to the aligned data points. For SRAR, the reconstruction applies to the optimized latent surface.

The reconstructions conducted in this study are generated by running SRAR and JPR on the synthetic dataset. We sample pairs of 3D reconstructions, compute mean reconstruction errors, and sample reconstruction pairs. In total, we collected 400 reconstruction pairs for the user study (200 using JPR and the other 200 using SRAR). Half of them form an Easy subset where the mean error of one reconstruction is smaller than half of another reconstruction. The remaining pairs form the Hard subset.

8.3 User Study on UQ for Model Checking

We proceed to evaluate the effects of the proposed UQ approaches for model checking, i.e., whether they help users to understand the quality of a 3D reconstruction in interactive scanning.

8.3.1 Experimental Setup. For quantitative evaluation, we have designed a user interface that consists of three sub-windows (See Figure 10). The left sub-window shows the ground-truth model, which mimics images of the underlying physical model. The middle and right sub-windows show two reconstructions of the underlying model, where the mean error of one reconstruction is smaller than that of the other one. We employed Poisson Surface Reconstruction [Kazhdan et al. 2006] to obtain the reconstructions. For JPR, the reconstruction applies to the aligned data points. For SRAR, the reconstruction applies to the optimized latent surface.

Each pair of reconstructions is displayed in four modes:

- **M-I: Raw reconstruction.** The first mode merely displays raw reconstructions without additional information. The user judges which one is better via visual comparisons with the underlying ground-truth.

- **M-II: Local alignment error.** The second mode shows the mean distances between the aligned point clouds and the reconstructed model. We collect statistics on the reconstructed model using the approach described in Section 5.3.

- **M-III: Results of our approach.** The third mode visualizes diagonal blocks (the default option) and leading eigenvectors of the predicted covariance matrix using the method described in Section 5.3.

- **M-IV: Ground-truth reconstruction error.** The last mode shows the ground-truth reconstruction error by measuring the closest distance between each point of the reconstruction and the underlying ground-truth reconstruction.

8.3.2 Analysis of Results. Figure 10 shows the results of a user study among 20 AMT users. We can see that when only displaying the rough reconstructions, picking the reconstruction with the smaller error is close to a random guess. This behavior is consistent among both the Easy subset and the Hard subset. In other words, it is critical to provide users with additional information to make accurate judgments. The local registration error is useful to some extent, yielding 67.5%/32.5% (±4.9%), 53.8%/46.2% (±5.5%), and 59.4%/40.6% (±5.2%) success/failure rates on Easy, Hard, and Overall, respectively. In other words, the local registration error may indicate the reconstruction error when one reconstruction is significantly better than the other reconstruction. On the other hand, it is ineffective for instances in the Hard subset.

In contrast, our approach yields far better results. The success/failure rates on settings of Easy, Hard, and Overall are 85.0%/15.0% (±3.2%), 72.6%/27.4% (±3.7%), and 78.8%/21.2% (±3.5%), respectively. The performance matches the results of using ground-truth reconstruction errors, i.e., 88.9%/11.1% (±2.8%), 78.8%/21.2% (±3.5%), and 83.9%/16.1% (±3.2%), respectively. Note that using the ground-truth reconstruction errors do not offer perfect results, as salient error regions of some reconstruction pairs are different, making it difficult to judge which one has a smaller mean error.

<table>
<thead>
<tr>
<th>Question</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective indication of global drifts?</td>
<td>94%</td>
</tr>
<tr>
<td>Effective indication of reconstruction errors?</td>
<td>88%</td>
</tr>
<tr>
<td>Eigenvectors indicate uncertainty structures?</td>
<td>85%</td>
</tr>
</tbody>
</table>

Table 1. A questionnaire for assessing our UQ approach.

We have conducted an additional questionnaire to gauge the effectiveness of our visualization schemes (See Table 1). For example, 94% of the users indicated that our UQ approach is useful for revealing...
global drifts. Moreover, 88% of the users reported that the output of our approach reflects magnitudes and patterns of reconstruction errors. Finally, 85% of the users commented that visualizations of eigenvectors indicate structures in the quantified uncertainties. All percentages are above 85%, and they justify the effectiveness of our approach.

8.4 Uncertainty-Driven View Planning

In this section, we present an experimental analysis of the proposed uncertainty-driven view planning approach under the framework of SRAR.

8.4.1 Experimental Setup. We perform experimental evaluations under the setting of category-specific view planning. Specifically, consider the task of scanning a physical object whose category label is known. We use a similar 3D model that belongs to the same category to compute the scanning trajectory. We then display the planned path to a user for interactive scanning. Note that the planned trajectory only suggests how to scan a new object. We find that users do not precisely follow the intended trajectory.

We have performed an experimental study on a Wrench category that consists of one reference model and three testing objects. Note that the differences between the testing model and the reference model vary, and such differences help assess the generalization ability of our approach. Interactive scanning employed an Artec Space Spider scanner. To generate the planned trajectory, we placed 1000 camera poses around the reference model. The distance from each camera origin to the reference model is 2cm, the viewing direction of each camera targets at the closest point. We run the algorithm described in Section 7 to generate the planned trajectory.

8.4.2 Analysis of View-Planning. For a baseline comparison, we pick a state-of-the-art model-based view planning approach [Scott 2009], which applies a set cover method to optimize a visibility-driven score among the selected views. As shown in Figure 11, our UQ-driven view planner delivers results that are better than the baseline approach. The mean/max reconstruction errors (in micros) on the three testing objects are 10.4/56.2, 15.2/68.1, and 18.3/83.2, respectively. In contrast, the mean/max reconstruction errors (in micros) of the baseline approach are 24.1/132.1, 41.2/145.1, and 49.4/161.7, respectively. Visually, our approach offers much evenly distributed reconstruction errors than the baseline approach, which exhibits certain global drifts. These results show the advantage of modeling uncertainties of geometry reconstruction for view planning.

For both approaches, the reconstruction errors increase when the differences between the testing model and the reference model escalate. However, our method offers a smaller variance among the three testing models than the baseline approach. These statistics show that the uncertainty score, which considers how scans overlap, is more stable than the visibility score for characterizing the reconstruction quality.

9 CONCLUSIONS AND FUTURE WORK

In this paper, we have studied the problem of uncertainty quantification for multi-scan registration. Our approach examines the approximation errors of utilizing the inverse of the Hessian matrix of the objective function to estimate the covariance matrix among the output derived from minimizing this objective function. This analysis results in robust and informative uncertainty measures. We have applied this approach under two settings of multi-scan registration, namely, minimizing pairwise distances between overlapping scans and aligning the input scans with a deforming template. Under both settings, we showed how to formulate objective functions that possess small discretization errors. Our uncertainty quantification methods also allowed us to compare these two formulations, both empirically and theoretically. Experimental results
on a variety of synthetic and real datasets and user studies on real datasets demonstrate the usefulness of our approaches. Finally, we showed the effectiveness of our uncertainty measures for the task of model-based view planning.

There are ample opportunities for future research. For example, it would be interesting to explore how to extend the linear approximation scheme employed in this paper to non-linear schemes, e.g., by introducing high-order components of the Taylor expansion. Besides, we would also like to incorporate non-rigid scan registration (that can address scan warping) into our formulation. Finally, it would be interesting to investigate uncertainty quantification for other geometry processing tasks such as shape editing, shape modeling, and reverse engineering.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable comments. Qixing Huang would like to acknowledge support from NSF DMS-1700234, NSF CIF-1729496, NSF IIS-1618648, a gift from Snap Research and a GPU donation from Nvidia Inc.

REFERENCES


Uncertainty Quantification for Multi-Scan Registration • 130:17
A PROOFS OF PROPOSITIONS AND THEOREMS IN SECTION 4

In this section, we present a formal statement of the main analysis result of the paper. We begin with introducing the notations in Section A.1. We then discuss the main result, key lemmas, and a proof of the main result in Section A.2. In Section A.3, we show a bound for the error Taylor’s approximation of the implicit function induced from some energy function that fits some regularity conditions. Finally in Section A.4, we discuss the error due to the reconstruction. Except in Section A.4 we always use the ground truth reconstruction parameter, i.e. \( p^{gt} \), as in (5).

A.1 Notations

**Definition 2.** Given a vector \( y \in \mathbb{R}^n \), we define the \( k \)-order product tensor induced from \( y \) as \( \otimes^k y \in \mathbb{R}^{n \times \cdots \times n} \).

**Definition 3.** Given two \( k \)-order tensors \( A, B \in \mathbb{R}^{n_1 \times \cdots \times n_k} \), we define their inner products as

\[
\langle A, B \rangle = \sum_{i_1=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} a_{i_1 \cdots i_k} b_{j_1 \cdots j_k},
\]

where \( a_{i_1 \cdots i_k} \) and \( b_{j_1 \cdots j_k} \) denote the corresponding elements of \( A \) and \( B \).

**Definition 4.** Suppose the implicit function \( x(y) \) is given by binary function \( f \) such that \( f(x(y), y) = 0 \). For simplicity we assume \( f(0, 0) = 0 \) and now we are interested about the local behavior near \((0, 0)\). The \( k \)-th order approximation for \( x(y) \) is referred to as

\[
S_n(y) := \sum_{k=0}^{n} \frac{1}{k!} \langle x^{(k)}(0), \otimes^k y \rangle
\]

where \( x^{(k)} \) is the \( k \)-th order derivative of \( x \) with respect to \( y \). The \( k \)-th order remainder is naturally defined as

\[
R_n(y) := x(y) - S_n(y) = \sum_{k=n+1}^{\infty} \frac{1}{k!} \langle x^{(k)}(0), \otimes^k y \rangle.
\]

For convenience we use abbreviation \( f_{x^k y^s} \) to denote the partial derivative \( \frac{\partial^{k+s}}{\partial x^k \partial y^s} f(0, 0) \) which is a \((b+s)\)-order tensor. Sometimes we treat \( f_{x^k y^s} \) as a matrix of shape

\[
\text{size}(x)^b \times \text{size}(y)^s
\]

where \( \text{size}(\cdot) \) takes the number of components of operand.

**Lemma 2.**

\[
\sum_{k=0}^{n} \frac{1}{k!} \langle \otimes^k R_n(y), f_{x^k} (S_n(y), y) \rangle = 0
\]

\[
\sum_{p \geq 2} \frac{1}{p!} \left\langle \otimes^p R_n(y), \sum_{q \geq 2} \frac{1}{q!} \langle \otimes^q S_n(y), f_{x^{p+q} y^s} \rangle \otimes (\otimes^s y) \right\rangle = 0
\]

**Proof.** By the Taylor’s theorem for binary function \( f(x, y) \) we have that

\[
\sum_{k,s \geq 0} \frac{1}{k! s!} f_{x^k y^s} (\otimes^k x(y)) \otimes (\otimes^s y) = 0
\]

Substituting \( x(y) \) by \( R_n(y) + S_n(y) \), (30) turns into

\[
0 = \sum_{k,s \geq 0} \frac{1}{k! s!} f_{x^k y^s} (\otimes^k x(y)) \otimes (\otimes^s y)
\]

\[
= \sum_{p \geq 2} \frac{1}{p!} \left\langle \otimes^p R_n(y), \sum_{q \geq 2} \frac{1}{q!} \langle \otimes^q S_n(y), f_{x^{p+q} y^s} \rangle \otimes (\otimes^s y) \right\rangle
\]

(31)

\[
= \sum_{p \geq 2} \frac{1}{p!} \langle \otimes^p R_n(y), f_{x^{p+q} y^s} \rangle \cdot (\otimes^q S_n(y), y)
\]

(32)

The equation (32) was obtained by shrinking the Taylor expansion of \( f_{x^p(x, y)} \) at point \( (S_n(y), y) \). In general the binomial theorem cannot apply to tensor product because of the lack of commutativity. As a matter of fact in (31) we utilized the symmetric property of the tensor \( f_{x^{p+q} y^s} \).

\[\square\]

Lemma 2 allows us to estimate \( |R_n(y)| \) only depending on partial derivatives of binary function \( f \) and low-order Taylor’s expansion of \( x(y) \).

**Corollary 2 (Scalar Version).** Suppose

\[
\sum_{k \geq 2} \frac{1}{k!} \left\langle \otimes^k R_n(y), f_{x^k} (S_n(y), y) \right\rangle \leq c R_n(y) \cdot f_x (S_n(y), y)
\]

holds for some constant \( c < 1 \). Then

\[
|R_n(y)| \leq \frac{1}{1-c} \left| f_x (S_n(y), y) \right|^{-1} |f(S_n(y), y)|.
\]

**Remark.** It is worth noting that the condition (33) is not that strong as it seems at first look. In the neighborhood of \( y_0 = 0 \), \( |R_n(y)| \) has the order of magnitude \( O(|y|^{n+1}) \), which means \( |f_{x^{n+1}}| \) just needs to grow in the rate slower than \( O \left( \frac{k!}{|y|^{(n+1)}} \right) \). It can be shown that

\[
f_x (S_n(y), y) = -\left( \frac{1}{n+1!} \right) (x^{(n+1)} \otimes \nabla^{n+1} y) + o(y)^{n+1},
\]

hence the dominating term in \( |R_n(y)| \) would be

\[
\frac{1}{(n+1)!} |x^{(n+1)}|^2 |\nabla^{n+1} y|^{n+1},
\]

compliant with the Taylor’s expansion of \( x(y) \).

A.2 Main Result and Key Lemmas

In this subsection we consider implicit function \( x(y) \) induced by energy function \( T(x, y) \):

\[
x(y) = \arg \min T(x, y).
\]

Assume \( x(0) = 0 \) as before. As we focus on the local behavior of \( x \) near \((0, 0)\) with some appropriate smooth conditions, this definition is equivalent to

\[
T_x(x, y) = 0
\]

where \( T_x = \frac{\partial T}{\partial x} \).
A subtle difference between $T_x$ and $f$ introduced in A.1 is that $T_x$ is a vector instead of scalar. But Lemma 2 still holds true by adding one more dimension from vector-valued $T_x$ to tensors $f_T^{x,y}$ and $f_{x,y}$ through proof. Similar to Corollary 2, we have the following lemma when $T(x, y)$ is known:

**LEMMA 3 (VECTOR VERSION).** Denote by $\lambda_{\text{max}}$ and $\lambda_{\text{min}}$ the largest and smallest eigenvalues of matrix $T_{xx}(S_n(y), y)$ respectively, and suppose

$$\sum_{k \geq 2} \frac{1}{k^2} \left( \otimes^k R_n(y) \right) \cdot T_{x+k}(S_n(y), y) \leq c |R_n(y)|^2 \cdot \lambda_{\text{max}}$$  \hspace{1cm} (34)

holds for some constant $c < 1$. Then

$$|R_n(y)| \leq \frac{2|T_x(S_n, y)|}{\lambda_{\text{min}} + \sqrt{\lambda_{\text{min}}^2 - 4c \lambda_{\text{max}} |T_x(S_n, y)|}}$$

whenever $|T_x(S_n, y)| \leq \frac{\lambda_{\text{min}}^2}{4c \lambda_{\text{max}}}$. Here $\| \cdot \|$ represents the spectral norm of matrices.

**PROOF.** By Lemma 2 and the triangle inequality we have

$$|T_{xx}(S_n(y), y) \cdot R_n(y)| \leq |T_x(S_n, y)| + \sum_{k \geq 2} \frac{1}{k^2} \left( \otimes^k R_n(y) \right) \cdot T_{x+k}(S_n(y), y) \leq |T_x(S_n, y)| + \frac{\lambda_{\text{max}} |R_n(y)|^2}{\lambda_{\text{min}}}$$

By definition of eigenvalues,

$$|T_{xx}(S_n(y), y) \cdot R_n(y)| \geq \lambda_{\text{min}} |R_n(y)|.$$  

Thus

$$\lambda_{\text{min}} |R_n(y)| \leq |T_x(S_n, y)| + c \cdot \lambda_{\text{max}} |R_n(y)|^2$$

which is a quadratic inequality about $|R_n(y)|$ thus has two branches. Since $R_n(y) = O(y^{n+1})$ in the neighborhood of $y_0 = 0$, by continuity we could assume $|R_n(y)|$ falls into the smaller branch. Solving this inequality reveals that

$$|R(y)| \leq \frac{2|T_x(S_n, y)|}{\lambda_{\text{min}} + \sqrt{\lambda_{\text{min}}^2 - 4c \lambda_{\text{max}} |T_x(S_n, y)|}}$$

which completes our proof. \hspace{1cm} \Box

We next bound the difference between $\overline{C}$ and $\text{Var}[x(y)]$ using tools introduced above. Recall that

$$\overline{C} := \text{Var}(x \cdot y) = \text{Var}[S_1(y)]$$

where $S_n$ was introduced in A.1. In the following we would discuss the more general situation where

$$\overline{C}_n := \text{Var}[S_n(y)]$$

in which we use $S_n(y)$ to approximate $x(y)$ rather than the linear approximation $S_1(y)$. In this way

$$\text{Var}[x(y)] - \overline{C}_n = \text{Var}[x(y)] - \text{Var}[S_n(y)] = \text{Var}[R_n(y)] + \text{Cov}[R_n(y), S_n(y)] + \text{Cov}[S_n(y), R_n(y)]$$

where $\text{Cov}(X, Y) := \mathbb{E}[X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^T]$ is the covariance matrix. By Cauchy’s inequality we have

$$\|\text{Var}[x(y)] - \overline{C}_n\| \leq 2\sqrt{\mathbb{E}[|R_n(y)|^2]} \|\text{Var}[S_n(y)]\| + 2\sqrt{\mathbb{E}[|S_n(y)|^2]}$$

(35)

Lemma 3 together with inequality (35) provides a general framework to bound the error of our approximated variance. In general $S_n(y)$ can be calculated explicitly and $T_x(S_n(y), y)$ has the order of magnitude $O(|y|^{n+1})$ in the neighborhood of $y_0 = 0$. Although $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$ are functions with respect to $y$, we could expect they are stable within the area of our interest since the implicit function $x(y)$ would be ill-conditioned if $\lambda_{\text{min}}(T_{xx})$ tends to zero.

**LEMMA 4.**

$$T_x(S_n(y), y) = -\frac{1}{(n+1)!} \langle \otimes^{n+1} y, x^{(n+1)}(0) \rangle + o(|y|^{n+2})$$

**PROOF.** By definition of $S_n(y)$ we have

$$T_x(S_n(y) + R_n(y), y) = 0$$

(36)

The Taylor’s expansion of $T_x$ has the form like Lemma 2.

It is worth noting that if $R_n(y)$ appeared in some product then this product would have no terms of degree less or equal to $n$. Hence the all terms of degree at most $n$ on left side of Lemma 2 is contributed by $T_x(S_n(y), y)$, which is of course zero from the right side, i.e., constant zero.

As for the coefficient of $\otimes^{n+1} y$, it suffices to compare the coefficient corresponding to degree $n + 1$ on both side of Lemma 2. \hspace{1cm} \Box

As far as the linear approximation $S_1(y)$ is concerned, (35) can be simplified further.

**THEOREM 5.** Suppose the random variable $\mathbb{E}[y] = 0$. The function $T(x, y)$ satisfies the condition in (34) and

$$|T_x(S_1(y), y)| \leq \|T_{xx}(S_1, y)\|/(4c\kappa(y))$$

where

$$\kappa(y) := \lambda_{\text{max}}(T_{xx}(S_1, y))/\lambda_{\text{min}}(T_{xx}(S_1, y)).$$

Then we have

$$\text{Var}[x(y)] - \overline{C}_n \leq 4\mathbb{E}\left[\frac{|T_x(S_1, y)|^2 |\kappa(y)|^2}{\|T_{xx}(S_1, y)\|^2} + 2|\kappa(0)| \mathbb{E}\left[\frac{2|T_x(S_1, y)| \cdot |\kappa(y)| \cdot |y|}{\|T_{xx}(S_1, y)\|}\right]\right]$$

**PROOF.** Just note that

$$\text{Cov}[R_1(y), S_1(y)] = \mathbb{E}[R_1(y)S_1(y)^T]$$

and

$$\|\text{Var}[R_1(y)]\| \leq \mathbb{E}[|R_1(y)|^2].$$

Then directly using Lemma 3 completes the proof. \hspace{1cm} \Box
A.3 An Explicit Bound for Functions with Regularity

Finally we show that a family of functions with regularity conditions satisfy the condition (34).

**Theorem 6.** Suppose there are positive constants $M, \alpha, \beta, \zeta, \kappa$ such that:

- $\|T_{x^1, y^1}(0, 0)\| \leq M\alpha^b \beta^s$

Here $T_{x^1, y^1}$ is regarded as a matrix whose column size is the dimension of variable $x$ and $\| \cdot \|$ takes the spectral norm of matrices.

- $\lambda_{\min}(T_{xx}(0, 0)) > \frac{M\alpha}{\kappa}$, which implies that $\frac{\lambda_{\max}(T_{xx}(0, 0))}{\lambda_{\min}(T_{xx}(0, 0))} \leq \kappa$.

- $|R_n(y)| \leq \frac{\zeta}{\alpha}$

and

$$|T_x(S_n(y), y)| \leq \frac{(1/\kappa - \tau e^\tau)^2}{2M\alpha^2 e^{\kappa \tau}},$$

where $\tau(y) = a|S_n(y)| + \beta|y|$. Then we have

$$|R_n(y)| \leq \frac{2|T_x(S_n(y), y)|}{\|T_{xx}(0, 0)\| \|1/\kappa - \tau(y)e^{\tau(y)}\|}$$

**Remark.** We make some explanation for the last two conditions above. (37) means $|R_n(y)|$ is constrained inside the area what we consider into the neighborhood of $(x_0, y_0) = (0, 0)$. This condition is necessary because the higher-order terms would play a critical role and the polynomial approximation fails when $y$ go beyond the area of constant radius.

The (38) is in fact rather weak. We have shown that

$$|T_x(S_n(y), y)| = -\frac{1}{(n + 1)!}T^{(n+1)}(0, 0)|F_x \cdot |y|^{n+1} + o(|y|^{n+1})$$

and the right side of (38) has a lower bound when constrained in the neighborhood of $(0, 0)$ where the growth of $r$ would be nearly linear.

**Proof.** By Taylor’s expansion, for $b \geq 1$ we have

$$\|T_{x^1, \cdots, x^b, y^1, \cdots, y^b}(0, 0)\| \leq \sum_{p, q \geq 0} \frac{1}{p!q!} |T_{x^{p+1} y^q}(0, 0)| \cdot |R_n(y)|^p \cdot |y|^q$$

$$\leq M\alpha^b \sum_{p, q \geq 0} \frac{1}{p!q!} \alpha^p \beta^q |R_n(y)|^p \cdot |y|^q$$

$$= M\alpha^b \sum_{k \geq 1} \frac{1}{k!} |a| |R_n(y)|^k + \beta |y|^k$$

$$= M\alpha^b e^{\tau(y)}. \quad (40)$$

For $T_{xx}$ we have

$$\|T_{xx}(S_n(y), y) - T_{xx}(0, 0)\| \leq \sum_{p+q \geq 1} \frac{1}{p!q!} |T_{x^{p+1} y^q}(0, 0)| \cdot |R_n(y)|^p \cdot |y|^q$$

$$\leq M\alpha \sum_{k \geq 1} \frac{1}{k!} \alpha^k |R_n(y)|^k \cdot |y|^q$$

$$= M\alpha \sum_{k \geq 1} \frac{1}{k!} \tau(y)^k$$

$$\leq M\alpha \tau(y)e^{\tau(y)} < \kappa \tau(y)e^{\tau(y)} \lambda_{\min}(T_{xx}(0, 0)).$$

Thus

$$\lambda_{\min}(T_{xx}(S_n(y), y)) \geq \lambda_{\min}(T_{xx}(0, 0))(1 - \kappa \tau(y)e^{\tau(y)})$$

$$\geq \|T_{xx}(0, 0)\| (1/\kappa - \tau(y)e^{\tau(y)}) \quad (1/\kappa - \tau(y)e^{\tau(y)})$$

(41)

Dividing left side of the equation

$$\sum_{b \geq 2} \frac{1}{b!} T_{x^{b+1}}(S_n(y), y) \cdot \phi^b R_n(y) = 0$$

into three parts we obtain that

$$T_{xx}(S_n(y), y) \cdot R_n(y)$$

$$= -T_x(S_n(y), y) \cdot R_n(y) - \sum_{b \geq 2} \frac{1}{b!} T_{x^{b+1}}(S_n(y), y) \cdot \phi^b R_n(y). \quad (42)$$

Inequality (41) implies that

$$\|T_{xx}(S_n(y), y) \cdot R_n(y)\| \geq \|T_{xx}(0, 0)\| (1/\kappa - \tau(y)e^{\tau(y)}) |R_n(y)|.$$  \quad (43)

Inequality (40) implies that

$$\left| \sum_{b \geq 2} \frac{1}{b!} T_{x^{b+1}}(S_n(y), y) \cdot \phi^b R_n(y) \right|$$

$$\leq M \sum_{b \geq 2} \frac{1}{b!} \alpha^b e^{\tau(y)} |R_n(y)|^b$$

$$\leq \frac{1}{2} M\alpha^2 \exp \{ \tau(y) + a|R_n(y)| \}. \quad (44)$$

Combining (42), (43) and (44) we obtain the inequality

$$\|T_{xx}(0, 0)\| (1/\kappa - \tau(y)e^{\tau(y)}) |R_n(y)|$$

$$\leq |T_x(S_n(y), y))| + \frac{1}{2} M\alpha^2 \exp \{ \tau(y) + a|R_n(y)| \} |R_n(y)|^2$$

$$\leq |T_x(S_n(y), y))| + \frac{1}{2} M\alpha^2 \exp \{ \tau(y) + \zeta \} |R_n(y)|^2. \quad (45)$$

The inequality (45) is a quadratic inequality about variable $|R_n(y)|$ and the solution has two branches. Since now we are interested in the neighborhood of $y_0 = 0$ and $R_n(y) = O(|y|^{n+1})$, we assume the solution falls into the branch corresponding to the smaller solutions. Solving this quadratic inequality and utilizing (38), the conclusion (39) immediately follows. \qed
A.4 The Error from Reconstruction Bias

In the last section we mainly consider the difference between \( \text{Var}[x(y)] \) and \( \tilde{C} \) where the reconstruction parameter \( p \) in (4) is assumed to be the ground truth \( p^{GT} \), but such an assumption is impractical in general. Instead we use \( p^* \) that is derived from one set of scans. In other words, \( p^* \) is a function of \( y \). In particular \( p^*(0) = p^{GT} \). In this section we are about to estimate the error of (6) compared with (5) where the ground truth reconstruction \( p^{GT} \) is replaced by \( p^* \).

The first order approximation of \( p^* \) is

\[
p^*(y) \approx p^{GT} + p'(0) \cdot y.
\]

(46)

In our noise model we assumed \( y \) has zero mean and covariance matrix \( \sigma^2 I \), which means the components of random variable \( y \) are pairwise independent. In the following argument we would use the first order approximation to analyze the error caused by the reconstruction bias. This reduction is reasonable since the reconstruction function \( p \) can be regarded locally linear about the input noise \( y \).

Thus by Markov’s inequality,

\[
\Pr[|p^*(y) - p^{GT}| \geq y\sigma] \leq \frac{\text{diag}(p^{GT}T^p(0))}{y^2} = \frac{|p'(0)|^2}{y^2},
\]

where diag means the sum of diagonal entries.

For some fixed \( y \), we consider solution \( x \) and \( x' \) for \( p^{GT} \) and \( p^* \) respectively:

\[
f(x, y, p^{GT}) = 0
\]

(47)

\[
f(x', y, p^*) = 0
\]

(48)

The Tyalor’s expansion for \( f \) with respect to variables \( x \) and \( p \) would be

\[
\sum_{b, l \geq 0} \frac{1}{b!l!} (\frac{\partial^b}{\partial x^b} f(x, y, p^{GT}) \cdot (\frac{\partial^l}{\partial p^l} (p^* - p^{GT}))) = 0.
\]

(49)

The first order approximation of \( x' \) would be

\[
x' \approx x - T_{xx}^{-1}T_{xp}(p^* - p^{GT}).
\]

Thus

\[
\text{Var}[x'] \approx \text{Var}[x] + T_{xx}^{-1}T_{xp}(p^* - p^{GT})(p^* - p^{GT})T_{px}T_{xx}^{-1}.
\]

(50)

Formula (50) is rather crude since we have not taken \( y \) into consideration. The following theorem gives a more accurate and formal estimation for the discussion above.

**Theorem 7.** Function \( x(y, p^*) \) is determined by equation

\[
x(y, p^*) = \arg\max_x T(x, y, p^*)
\]

(51)

where \( p^* \) is the predicted parameter. Suppose \( x(y^{GT}, p^{GT}) = x^{GT} \) and \( y \) follows the noise model that \( \text{Var}[y] = \sigma^2 I \) as the main text. For brevity we assume that

\[
x^{GT} = y^{GT} = p^{GT} = 0.
\]

(52)

Let \( S_1(y, p^*) \) be the first order approximation of function \( x(y, p^*) \) with respect to variable \( y \), i.e.,

\[
S_1(y, p^*) = \frac{\partial x}{\partial y}(0, p^*) \cdot y.
\]

Here we ignored the constant term \( x(0, p^*) \) because \( x(0, p^{GT}) = 0 \). The following inequality provides a bound for bias between variances of \( S_1(y, p^*) \) and \( S_1(y, p^{GT}) \).

\[
-Z(p^*) \leq \text{Var}[S_1(y, p^*)] - \text{Var}[S_1(y, p^{GT})] \leq Z(p^*)
\]

(52)

in which \( Z(p) \) is a matrix such that

\[
||Z(p^*)||
\]

\[
\leq 2\sigma^2 \max \left\{ ||T_{xx}^{-1}||^2 \cdot ||T_{xy}|| \left( ||T_{xx}^{-1}|| \cdot ||T_{xy}|| \cdot ||T_{xp}|| + ||T_{xy}|| \right) \right\}
\]

\[
\cdot |p^* - p^{GT}| \cdot \text{dim}(y)
\]

(53)

where max takes arguments along the segment joining \( (0, 0, p^*) \) and \( (0, 0, p^{GT}) \) for partial derivatives \( T_{xx}, T_{xy}, T_{xp} \) and \( T_{xy} \).

**Remark.** If the partial derivatives of \( T \) are in the neighborhood of \( p^{GT} \) and \( |p^* - p^{GT}| < \delta \), we can simplify (53) into

\[
||Z(p^*)|| \leq \sigma^2 \cdot |p^* - p^{GT}| \cdot C(p^{GT}, \delta)
\]

(54)

where \( C(p^{GT}, \delta) \) depends on the ground truth reconstruction \( p^{GT} \) and the reconstruction error \( \delta \). When \( y \) is normally distributed, \( p^* \) can be also regarded as approximately normally distributed concentrated in \( p^{GT} \) in local sense. In this way the error matrix \( Z \) is typically smaller than \( \text{Var}[S_1(y, p^{GT})] \) by several orders of magnitude.

**Proof.** Using the mean value theorem of vector-valued functions we have

\[
|S_1(y, p^*) - S_1(y, p^{GT})| \leq \frac{\partial S_1}{\partial p} \cdot |p^* - p^{GT}|
\]

\[
\leq \left( ||T_{xx}^{-1}||^2 \cdot ||T_{xy}|| \cdot ||T_{xp}|| + ||T_{xy}|| \right) \cdot |y| \cdot |p^* - p^{GT}|.
\]

(55)

where \( p_i \) is some point on segment joining \( p^* \) and \( p^{GT} \), and \( T_{xx}, T_{xy}, T_{xp}, T_{xy} \) are all taken at point \( (x^{GT}, y^{GT}, p_i) \). Next we consider the variance. First we have inequality

\[
\text{Var}[S_1(y, p^*)] \leq \mathbb{E}[S_1(y, p^*) S_1^T(y, p^*)].
\]

Utilizing the face that \( \mathbb{E}[S_1(y, p^{GT})] = 0 \), we have

\[
\text{Var}[S_1(y, p^*)] \leq \mathbb{E}[S_1(y, p^*) S_1^T(y, p^*) - S_1(y, p^{GT}) S_1^T(y, p^{GT})].
\]

Similarly, again applying mean value theorem we have

\[
|S_1(y, p^*) S_1^T(y, p^*) - S_1(y, p^{GT}) S_1^T(y, p^{GT})| \leq \frac{\partial S_1}{\partial p} \cdot |p^* - p^{GT}|
\]

\[
\leq 2 ||T_{xx}^{-1}||^2 \cdot ||T_{xy}|| \left( ||T_{xx}^{-1}|| \cdot ||T_{xy}|| \cdot ||T_{xp}|| + ||T_{xy}|| \right) \cdot |y|^2 \cdot |p^* - p^{GT}|
\]

where \( p_i \) is some point on segment joining \( p^* \) and \( p^{GT} \), and \( T_{xx}, T_{xy}, T_{xp}, T_{xy} \) are all taken at point \( (x^{GT}, y^{GT}, p_i) \). Since \( \mathbb{E}[|y|^2] = \sigma^2 \cdot \text{dim}(y) \) we complete our proof. \( \square \)
B PROOFS OF PROPOSITIONS AND THEOREMS IN SECTION 5

B.1 Proof of Prop. 2

Since we only care about the local behavior near the ground truth, we use the $p_{ij}^{gt}, n_{ij}^{gt}$ to replace the optimized poses $p_{ij}^*, n_{ij}^*$. Also for convenience we assume the output $(e_i, t_i)$ is $(0, 0)$ corresponding to $p_{ij}^{gt}, n_{ij}^{gt}$ since otherwise we can just consider the shift from the ground truth.

$\mathbf{x}^*(y)$ is determined by

$$
\begin{align*}
\mathbf{x}^*(y) & := \arg \min_{\{R, t\}} \frac{\partial^2}{\partial y^2} \sum_{(i', r') \in C} d^2(S_i, T_i, S_{i'}, T_{i'}) \\
& = \arg \min_{x} \{R, t\} \frac{\partial^2}{\partial y^2} \sum_{(i', r') \in C_{ir}} \sum_{\alpha_{ijr}^{\prime}} (y_{ij}^{\prime})^2,
\end{align*}
$$

in which

$$
\begin{align*}
d^2(S_i, T_i, S_{i'}, T_{i'}) & := \sum_{(p_{ij}, p_{i'j'}) \in C} \frac{\partial^2}{\partial y^2} \sum_{(i', r') \in C_{ir}} \sum_{\alpha_{ijr}^{\prime}} (y_{ij}^{\prime})^2,
\end{align*}
$$

$$
\alpha_{ijr}^{\prime}(x, y) = (R_ip_{ij} + t_i - R_{i'r'}p_{i'j'} - t_{i'})^T (R_{i'r'}n_{i'j'}),
$$

and

$$
\begin{align*}
f(x, y) & = \sum_{(p_{ij}, p_{i'j'}) \in C} \frac{\partial^2}{\partial y^2} \sum_{(i', r') \in C_{ir}} \sum_{\alpha_{ijr}^{\prime}} (y_{ij}^{\prime})^2.
\end{align*}
$$

Introducing notation $\alpha_{ijr}^{\prime}$ is to emphasize the dependence upon variables $x$ and $y$. It is worth noting that $\alpha_{ijr}^{\prime}(x^{gt}, y^{gt}) = 0$ for all $(i, i', r') \in E$.

By definition

$$
C := \text{Var}[x'(y^{gt}) \cdot y] = \mathbb{E}[x'(y^{gt}) \cdot y[y^{gt}]^T = \sigma^2 x'(y^{gt}) x'(y^{gt})^T,
$$

where $x'(y^{gt})$ is the first order derivative of $x$ at the point of ground truth. It is known that

$$
\begin{align*}
x'(y^{gt}) & = \left(\frac{\partial^2}{\partial x^2} f(x^{gt}, y^{gt})\right)^{-1} \frac{\partial^2}{\partial x^2} f(x^{gt}, y^{gt}).
\end{align*}
$$

Here we clarify the format of the variable $x$. Recall that $y$ collects input $p_{ij}$ and $n_{ij}$, and $x$ consists of pose estimation $R_{ij} = \exp(e_i \times)$ and $t_i$. Note $x$ is a $6(N-1)$-dimensional vector since $R_{ij}$ is always $I$ and $t_i$ is 0. We put $t_i$ in $(6i - 1)$th to $(6i - 9)$th components of $x$, $e_i$ in $(6i - 8)$th to $(6i - 6)$th components of $x$.

Define $h_{ij} \in \mathbb{R}^{6(N-1)}$ whose non-zero elements are

$$
\begin{align*}
V_{6i-11:6i-9}(h_{ij}) & := n_{ij}^{gt} \\
V_{6i-6:6i-6}(h_{ij}) & := p_{ij}^{gt} \times n_{ij}^{gt}
\end{align*}
$$

Using such a definition we have

$$
\begin{align*}
\frac{\partial^2}{\partial x^2} f(x^{gt}, y^{gt}) & = 2 \sum_{(p_{ij}, p_{i'j'}) \in C} \frac{\partial^2}{\partial x^2} f(x^{gt}, y^{gt}) \cdot \left(\frac{\partial^2}{\partial x^2} f(x^{gt}, y^{gt})\right)^T \\
& = 2 \sum_{(p_{ij}, p_{i'j'}) \in C} (h_{ij} - h_{i'j'})(h_{ij} - h_{i'j'})^T.
\end{align*}
$$

where $P_{ij} \in \mathcal{P}_3, |y|$ are column permutations so that all $P_{ij}$ have disjoint non-zero columns. Let $e_{ij} \in \mathbb{R}^{6(N-1)}$ be the canonical basis with the $i, j$-th element being one. Since

$$
\begin{align*}
(P_{ij,h_i} - P_{i'j',h_{i'}})(P_{ij} - P_{i'j'})^T & = [\|((t_1, j_1) = (i_2, j_2)) + \|((i'_1, j'_1) = (i'_2, j'_2)) \\
- \|((i'_1, j'_1) = (i_2, j_2)) - \|((i_1, j_1) = (i'_2, j'_2))] I \\
& = (e_{ij} - e_{i'j'})(C_{ij} - e_{i'j'} I).
\end{align*}
$$

Finally we obtain that

$$
\begin{align*}
C := \text{Var}[x'(y^{gt}) \cdot y] & = C_{m1} A_{dis} A_{dis} C_m^{-1}.
\end{align*}
$$

C PROOFS OF PROPOSITIONS AND THEOREMS IN SECTION 6

In this section, we provide proofs of the propositions and theorems in Section 6.

C.1 Proof of Prop. 3

The energy function is

$$
f = \sum_{i=1}^{N} \sum_{j=1}^{N} h_{ij}^2
$$

where $h_{ij} = (R_{ij}p_{ij} + t_i)^T n_{ij} - d_{kij}$.

Similar to Prop. 2, we have

$$
\begin{align*}
\text{Var}[x^*(y)] & = \left(\frac{\partial^2 f}{\partial x^2} \right)^{-1} \frac{\partial^2 f}{\partial y^2} \cdot \text{Var}[y] \cdot \left(\frac{\partial^2 f}{\partial y^2}\right)^{-1} \\
& = \sigma^2 \left(\frac{\partial^2 f}{\partial x^2} \right)^{-1} \frac{\partial^2 f}{\partial y^2} \cdot \frac{\partial^2 f}{\partial y^2} \cdot \left(\frac{\partial^2 f}{\partial x^2}\right)^{-1},
\end{align*}
$$

in which all partial derivatives were taken at the point of ground truth $(x^{gt}, y^{gt})$.

Note that $h_{ij}(x^{gt}, y^{gt}) = 0$ for ground truth $(x^{gt}, y^{gt})$. Taking advantage of this property we have

$$
\begin{align*}
\frac{\partial^2 h_{ij}^2}{\partial x^2} & = \frac{\partial h_{ij}^2}{\partial x} \cdot \frac{\partial h_{ij}}{\partial x} \cdot \frac{\partial h_{ij}}{\partial x} \left(\frac{\partial h_{ij}}{\partial x}\right)^T \\
& = 2 \frac{\partial h_{ij}}{\partial x} \cdot \frac{\partial h_{ij}}{\partial x} \left(\frac{\partial h_{ij}}{\partial x}\right)^T
\end{align*}
$$

ACM Trans. Graph., Vol. 39, No. 4, Article 130. Publication date: July 2020.
\[
\frac{\partial^2 h_{ij}}{\partial x \partial y} = 2 \left( h_{ij} \frac{\partial^2 h_{ij}}{\partial x^2} + \frac{\partial h_{ij}}{\partial x} \frac{\partial h_{ij}}{\partial y} \right) = 2 \frac{\partial h_{ij}}{\partial x} \frac{\partial h_{ij}}{\partial y} \quad (58)
\]

Observe that the only non-zero components in \( \frac{\partial h_{ij}}{\partial y} \) is corresponding to parameter \( p_{ij} \), which leads to

\[
\left( \frac{\partial h_{ij}}{\partial y} \right)^T \frac{\partial h_{ij}}{\partial y} = \begin{cases} p_{ij}^T p_{ij} = 1, & \text{if } (i, j) = (i', j') \\ 0, & \text{otherwise} \end{cases}
\]

Thus

\[
\frac{\partial^2 f}{\partial x \partial y} = 4 \sum_{i=1}^N \sum_{j=1}^N h_{ij} \left( \frac{\partial h_{ij}}{\partial x} \right)^T = \frac{\partial^2 f}{\partial x^2},
\]

and furthermore,

\[
\operatorname{Var}[x^*(y)] = \sigma^2 \left( \sum_{i=1}^N \sum_{j=1}^N \frac{\partial h_{ij}}{\partial x} \left( \frac{\partial h_{ij}}{\partial y} \right)^T \right)^{-1}.
\]

Simple calculus figures out that

\[
\frac{\partial h_{ij}}{\partial x} = g_{ij}
\]

where \( g_{ij} \) is just as defined in Prop. 3. Hence the proof completes. \( \square \)

C.2 Efficient Inverse

One challenge of (22) in the main paper is that for large-scale problems, e.g., hundreds of scans with hundreds of thousands of latent planes, it is infeasible to compute the inverse of \( C_{XX} \) directly. In the following, we describe how to leverage the structure in \( C_{XX} \) to efficiently compute the pose component \( C_{SRAR, \text{Pose}} \) of \( C_{SRAR} \).

The following proposition characterizes the structure of \( C_{XX} \).

**Proposition 4.** \( C_{XX} \) can be expressed as

\[
C_{XX} = \begin{pmatrix} A & B \\ B^T & \Sigma \end{pmatrix}
\]

where \( A \in \mathbb{R}^{N \times (N-1)} \) is a \((N-1) \times (N-1)\) block matrix, where the \( (i, i') \)-th block is non-zero if and only if the corresponding planes of the data points of \( S_{i+1} \) and \( S_{i-1} \) overlap. \( B \in \mathbb{R}^{(N-1) \times K} \) is a \((N-1) \times K\) block matrix, where the \( (i, k) \)-th block is non-zero if and only if the \( k \)-th plane is the corresponding plane of a data point in \( S_{i+1} \), \( \Sigma \in \mathbb{R}^{K \times K} \) is a \( K \times K \) block diagonal matrix.

The proof is straightforward and we omit it for brevity.

The \( 2 \times 2 \) block structure of \( C_{XX} \) offers a simple way to express the its inverse:

**Proposition 5.** The inverse of \( C_{XX} \) is given by

\[
\begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & \Sigma^{-1} & 0 & 0 \\
0 & 0 & I & 0 \\
0 & 0 & 0 & \Sigma^{-1} \\
\end{pmatrix}
\begin{pmatrix}
I & 0 & 0 & 0 \\
0 & I & 0 & 0 \\
0 & 0 & I & 0 \\
0 & 0 & 0 & I \\
\end{pmatrix}
\]

**Proof.** See Appendix C.3. \( \square \)

Since \( \Sigma \) is a block diagonal matrix, we can use (58) to efficiently encode \( C_{XX} \).

C.3 Proof of Proposition 5

Denote

\[
\begin{pmatrix} A & B \\ B^T & \Sigma \end{pmatrix}^{-1} = \begin{pmatrix} X & Y \\ Y^T & Z \end{pmatrix}.
\]

It follows that

\[
\begin{align*}
AX + BY &= I \\
AY + BZ &= 0 \\
B^T Y + \Sigma Z &= I
\end{align*}
\]

(62) leads to

\[
Z = \Sigma^{-1} (I - B^T Y)
\]

Substitute (63) into (61), we obtain

\[
(A - B\Sigma^{-1} B^T) Y = -B\Sigma^{-1} \Rightarrow Y = -(A - B\Sigma^{-1} B^T)^{-1} B\Sigma^{-1}.
\]

Substitute (64) into (62), we obtain

\[
Z = \Sigma^{-1} (I + B^T (A - B\Sigma^{-1} B^T)^{-1} B\Sigma^{-1}) = \Sigma^{-1} + (B\Sigma^{-1})^T (A - B\Sigma^{-1} B^T)^{-1} (B\Sigma^{-1}).
\]

Substitute (64) into (60), we obtain

\[
X = A^{-1} (I + B\Sigma^{-1} B^T (A - B\Sigma^{-1} B^T)^{-1} B\Sigma^{-1}) = A^{-1} + A^{-1} B\Sigma^{-1} B^T (I - A^{-1} B\Sigma^{-1} B^T)^{-1} A^{-1}
\]

Substituting (64)(65), we obtain

\[
X = A^{-1} (I - A^{-1} B\Sigma^{-1} B^T)^{-1} A^{-1} = (A - B\Sigma^{-1} B^T)^{-1}.
\]

Combining (64),(65), and (66), we end the proof. \( \square \)

D PROOFS OF PROPOSITIONS AND THEOREMS IN SECTION 7

D.1 Proof of Theorem 1

First of all, we have

\[
\sum_{i=1}^n \left( \begin{array}{c} w_i G_i \\ \hat{G}_i \end{array} \right) \cdot \left( \begin{array}{c} w_i \hat{G}_i^T \\ G_i \end{array} \right) \geq 0.
\]

Denote the left side of (67) as \( \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \). Then we have

\[
\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \geq 0, \quad A > 0.
\]

It follows that

\[
0 \leq \begin{pmatrix} I & 0 \\ -B^T A^{-1} & I \end{pmatrix} \cdot \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \cdot \begin{pmatrix} I & 0 \\ 0 & -A^{-1} B \end{pmatrix}
\]

\[
= \begin{pmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{pmatrix}
\]

This means

\[
C - B^T A^{-1} B \geq 0,
\]

which ends the proof.