CHALMERS
UNIVERSITY OF TECHNOLOGY

Global Optimality for Point Set Registration Using Semidefinite Programming

Citation for the original published paper (version of record):
Iglesias, J., Olsson, C., Kahl, F. (2020). Global Optimality for Point Set Registration Using Semidefinite Programming. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020: 8284-8292.
http://dx.doi.org/10.1109/CVPR42600.2020.00831

(article starts on next page)

Global Optimality for Point Set Registration Using Semidefinite Programming

José Pedro Iglesias[1], Carl Olsson[1,2], and Fredrik Kahl[1]

[1]Chalmers University of Technology, Sweden
[2]Lund University, Sweden

Abstract

In this paper we present a study of global optimality conditions for Point Set Registration (PSR) with missing data. PSR is the problem of aligning multiple point clouds with an unknown target point cloud. Since non-linear rotation constraints are present the problem is inherently non-convex and typically relaxed by computing the Lagrange dual, which is a Semidefinite Program (SDP).

In this work we show that given a local minimizer the dual variables of the SDP can be computed in closed form. This opens up the possibility of verifying the optimally, using the SDP formulation without explicitly solving it. In addition it allows us to study under what conditions the relaxation is tight, through spectral analysis. We show that if the errors in the (unknown) optimal solution are bounded the SDP formulation will be able to recover it.

1. Introduction

Point Set Registration (PSR) is a problem with several applications in the area of computer vision [4, 20, 14, 21, 29, 5, 2, 11, 24], which consists of estimating transformations that align two or more sets of 3D points. The instance of PSR with only two views related by a rigid transformation is a particular case which has been been intensively studied and that can be solved in closed form [3]. But as the number of view increases, and incomplete observations are taken in consideration, the problem becomes more complex due to the constraints on the rotations.

In this work we focus on the registration of multiple point clouds to a global coordinate system. The point clouds correspond to observations of subsets of n points from m local coordinate systems. These observations, besides incomplete, can be noisy and are related through unknown rigid transformations. Assuming that sparse correspondences between 3D points are known, the aim of such problem is not only to estimate the unknown transformations from the m local coordinate systems to a global coordinate system, but also the global coordinates of the original set of n points. Solutions or approximated solutions to this problem are not exactly new, as shown in [7], whereby we aim at study global optimality conditions when solving PSR using Semidefinite Programming (SDP) [26].

Global optimization over the manifold of rotations have previously been studied by a number of papers. A classical result is that of Horn *et al*. [13] which showed that registration of two point sets under a Euclidean transformation can be solved globally optimally using singular value decomposition. In [18] a similar problem where the point-to-point correspondences were replaced by point-to-plane correspondences was considered. It was shown how to apply Lagrangian duality to achieve orthogonality or the transformation matrix. The relaxation was tightened in [6], which introduced orientation constraints on the columns of the rotation matrix and added redundant rotation constraints to strengthen the lower bound provided by the dual problem. Yang *et al*. [28] also applies Lagrangian duality to obtain a quaternion-based certifiably optimal solution to the Wahba problem.

A problem that has received lots of attention in recent years due to it usefulness in structure from motion and SLAM applications is that of rotation averaging [9, 1, 30, 19, 8] One of the first attempts at solving this problem was presented in [10] where the unit norm constraint was dropped to achieve a linear solution strategy. Fredriksson *et al*. [9] showed how to incorporate the norm constraint using duality. Martinec and Pajdla [16] used the matrix parameterization but dropped the orthogonality and determinant constraints and extracted an approximate solution from the eigenvalues corresponding to the three smallest eigenvalues of the system matrix. Eriksson *et al*. [8] recently showed that despite the non-convexity of the problem, applying Lagrangian duality allows to enforce orthonormality and determinant constraints exactly under very generous noise levels. In [12] convexity properties of the single rotation averaging problem are studied. As far as we know these results

do not generalize to the case of multiple rotations.

From an application point of view the work that is the closest to ours is that of Chaudhury et al. [7]. This work introduced a spectral and a SDP relaxation that turns global registration of point clouds into a convex problem, which can be solved in polynomial time using a proposed algorithm. The authors also present sufficient conditions on the problem for exact recovery using the proposed algorithm, as well as a detailed stability analysis.

Motivated by [8], and even though it starts from a similar formulation, this work differs from [7] by proposing a certificate and studying the conditions in terms of the estimated point cloud, global pose and missing data that assures global optimality of a candidate primal solution, using Lagrangian duality and spectral analysis. The main contributions of the paper are:

- application of Lagrangian duality to the Point Registration Problem, that given a candidate solution to the primal problem allow us to obtain the corresponding dual variable in closed form. This allows verifying the global optimality of a local minimizer without solving the SDP (Lemma 1), which leads to a significant speedup;
- derivation of certificates/bounds (Theorems 1 and 2) on reconstruction errors that, if fulfilled, are sufficient to guarantee global optimality of a candidate primal solution;
- analysis and evaluation of the certificates as function of missing data and spatial distribution of the estimated 3D scene, using synthetic and real data.

1.1. Notation

Throughout the paper we use lower-case letters to represent scalars and vectors, and upper-case letters for matrices. An element of a matrix A at position (i, j) is referred as $\{A\}_{i,j}$, while $A_{i,j} \in \mathbb{R}^{n,m}$ corresponds to a block of the matrix A with dimensions $n \times m$. We use $\|A\|$ as the operator 2-norm of a matrix A, and $A^\dagger$ as its Moore–Penrose pseudoinverse. We denote $\lambda_i(A)$ as the i:th eigenvalue of a matrix $A \in \mathbb{R}^{n\times n}$, such that $\lambda_1(A) \leq \lambda_2(A) \leq \ldots \leq \lambda_n(A)$.

2. Problem Statement

The Point Set Registration (PSR) problem consists of estimating m transformations, i.e. pairs of rotations and translations, that align m observed point sets w.r.t. a global coordinate system, along with an "average" point set. The m observed point sets and the "average" point set are usually also called *source* point sets and *target* point set, respectively. The problem assumes known correspondences between point sets, and it can have missing data in the sense that a source set might lack observations of some points in the target set. The goal is to minimize the distance between the transformed source sets to the target set, and the corresponding optimization problem is usually formulated as

$$\begin{aligned} \min_{y,R,t} \quad & \sum_{i,j} w_{i,j} \|y_i - (R_j x_{i,j} + t_j)\|^2 \\ \text{subject to} \quad & R_j \in \mathcal{SO}(3),\ j = 1, \ldots, m \end{aligned}, \tag{1}$$

where R_j and t_j are the rotation and translation from frame j to the global frame, y_i the i:th point of the target set, $x_{i,j}$ the observation of the i:th point of the j:th source set, and $w_{i,j} = \{0, 1\}$ determines if point i is observed in the j:th source set. We assume that the target set has n points and there are m source sets, so $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Each term of (1) can be written in more compact form using trace notation as follows

$$\begin{aligned} &\|y_i - (R_j x_{i,j} + t_j)\|^2 = \\ &= \text{tr}\left(\begin{bmatrix} y_i & t_j & R_j \end{bmatrix} \begin{bmatrix} 1 & -1 & -x_{i,j}^T \\ -1 & 1 & x_{i,j}^T \\ x_{i,j} & x_{i,j} & -x_{i,j}x_{i,j}^T \end{bmatrix} \begin{bmatrix} y_i^T \\ t_j^T \\ R_j^T \end{bmatrix} \right), \end{aligned} \tag{2}$$

and if we assume that each source set has zero mean, i.e. $\sum_i w_{i,j} x_{i,j} = 0$, by summing over i and j the total cost in (1) becomes

$$\text{tr}\left(\begin{bmatrix} Y & T & R \end{bmatrix} \begin{bmatrix} A & -W & -X^T \\ -W^T & B & 0 \\ -X & 0 & C \end{bmatrix} \begin{bmatrix} Y^T \\ T^T \\ R^T \end{bmatrix} \right) \tag{3}$$

where $Y = [y_1 \ \ldots \ y_n]$, $T = [t_1 \ \ldots \ t_m]$, $R = [R_1 \ \ldots \ R_m]$, $\{W\}_{i,j} = w_{i,j}$, A and B are diagonal matrices with $\{A\}_{i,i} = \sum_j w_{i,j}$ and $\{B\}_{j,j} = \sum_i w_{i,j}$, C is a block diagonal matrix with 3×3 blocks $C_{j,j} = \sum_i w_{i,j} x_{i,j} x_{i,j}^T$, and $X = \left[X_1^T, \ldots, X_m^T\right]^T$ with $X_j = \begin{bmatrix} w_{1,j}x_{1,j} & \ldots & w_{n,j}x_{n,j} \end{bmatrix}$. The cost function defined in (3) depends on R, Y and T. If we fix R the remaining variables are unconstrained with a least squares objective and we can therefore compute the optimal Y and T in terms of R, which eliminates Y and T. Another way to write (3) is

$$\text{tr}\left(YAY^T - 2YWT^T - 2YX^TR^T + TBT^T + RCR^T\right), \tag{4}$$

where the term RCR^T can be dropped since RR^T is constant along its 3×3 blocks diagonal. From the stationary conditions for the gradient of (4) with respect to the T, we get that $-2YW + 2TB = 0 \Leftrightarrow T = YWB^{-1}$ and by replacing T in (4)

$$\text{tr}\left(YLY^T - 2YX^TR^T\right). \tag{5}$$

The matrix $L = A - WB^{-1}W^T$ is the Laplacian of a weighted graph, with n nodes, each corresponding to a 3D

point from Y. A similar type of graph was studied by Mieghem et al. [25]. Graph Laplacians have some well known properties: 1) they are positive semidefinite; 2) for non-degenerative cases, they have a unique zero eigenvalue, which corresponds to the eigenvector $\mathbb{1}_n$, that is a vector of all ones. For simplicity of notation we let $P = L^{\dagger}$. By considering the eigenvalue decompositions of L and P it is easy to see that the same is true for P, and that non-zero eigenvalues of P can be obtained from the eigenvalues of L as $\lambda_k(P) = 1/\lambda_{n-k+2}(L),\ k = 2, \ldots, n$. Alternatively, one can write $L = \sum_k \Theta_k$, where $\Theta_k = \text{diag}(W_k) - \frac{1}{n_k} W_k^T W_k$ and W_k is row k of W. Note that for source point sets with zero mean, we have that $t_k = \frac{\sum_i w_{i,k} y_k}{\sum_i w_{i,k}} = \frac{1}{n_k} Y W_k^T$, where $n_k = \sum_i w_{i,k}$. This means that T_k can be obtained from Y as $T_k = \frac{1}{n_k} Y W_k^T W_k$, and consequently $Y_k - T_k = Y\left(\text{diag}(W_k) - \frac{1}{n_k} W_k^T W_k\right) = Y\Theta_k$. Each Θ_k is an orthogonal projection, which means $\Theta_k^2 = \Theta_k = \Theta_k^T$, and it can be interpreted as an operator that selects the set of points of Y that are observed in the point set k and makes it zero-meaned by subtracting its average. For a complete source point set, W_k is a all-ones vector and we have $\Theta = I_n - \frac{1}{n}\mathbb{1}_n$, and as so $\Theta P = P\Theta = P$.

To eliminate Y from the expression (5) we again use the stationary point conditions, now using the gradient with respect to Y, which gives

$$2YL - 2RX = 0 \Leftrightarrow Y = RXP + Q, \tag{6}$$

where Q is an arbitrary matrix that fulfills $QL = 0$. The matrix Q, who's rows are in the nullspace of L, must have constant rows. Since the columns of X^T have zero mean it is clear that $QX^T = 0$ and therefore substituting (6) into (5) we can write our cost function as $-\text{tr}\left(RXPX^TR^T\right)$. Since Q has no influence on the cost function, we can choose $Q = 0$ and as so $Y = RXP$. We define our primal problem (P) as minimizing the cost function over the rotation matrices group, i.e.,

$$\begin{aligned} (P) \quad & \min_{R} \quad -\text{tr}\left(RXPX^TR^T\right) \\ & \text{subject to} \quad R_j \in \mathcal{SO}(3),\ j = 1, \ldots, m. \end{aligned} \tag{7}$$

In the following we let $M = -XPX^T$.

3. Duality and Optimality Conditions

This section will focus on sufficient conditions, in terms of the dual variables, for which there is no duality gap between the primal problem (7) and its dual problem. Then, in the next section, we will analyze these conditions in terms of the input data, noise level and missing data.

Eriksson et al. [8] have studied these conditions for the dual problem applied to the rotation averaging problem. The reasoning and main conclusions, given the similar structure of the primal problem, can also be directly applying for the semidefine program to the PSR problem. In this section we give a brief review of this analysis. For additional details we refer to [8].

3.1. Relaxation of the Primal Problem

For a solution R^* to be a minimum of the primal problem (P) it has to fulfill KKT conditions. If we relax the primal by removing the determinant constraint ($\det(R_j) = 1,\ j = 1, \ldots, m$), the Lagrangian of the problem is

$$L(R, \Lambda) = \text{tr}\left(R(\Lambda + M)R^T\right) - \text{tr}\left(\Lambda\right), \tag{8}$$

where $\Lambda = \text{diag}(\Lambda_1, \ \ldots, \ \Lambda_m)$ and Λ_j are 3×3 symmetric matrices. From the gradient of (8) we get that the KKT conditions are

$$\begin{aligned} &(\Lambda^* + M)\, {R^*}^T = 0, \\ &R^* \in \mathcal{SO}(3)^m. \end{aligned} \tag{9}$$

The first equation of (9) allow us to get a closed form solution for Λ^* as a function of R^*, where each block Λ_i^* is given by

$$\Lambda_i^* = \sum_j X_i P X_j^T {R_j^*}^T R_i^*,\ i = 1, \ldots, m. \tag{10}$$

From (10), and by defining $M_\Lambda = \Lambda^* + M$, we get that

$$(M_\Lambda)_{i,j} = \begin{cases} \sum_{k \neq i} X_i P X_k^T {R_k^*}^T R_i^*, & i = j \\ -X_i P X_j^T, & \text{otherwise} \end{cases}. \tag{11}$$

The result from (11) will be used in later sections.

3.2. Sufficient Global Optimality Conditions

In this section we will define the dual problem of (P) and state conditions in terms of M_Λ which guarantee that there is no duality gap between the primal and the dual problems. From (8) we can write the dual (D) problem of (P) as

$$\max_{M_\Lambda \succeq 0} \quad \min_{R} \quad L(R, \Lambda). \tag{12}$$

Note that we get

$$\min_{R} \quad L(R, \Lambda) = \begin{cases} -\text{tr}(\Lambda), & M_\Lambda \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}, \tag{13}$$

which gives the dual problem

$$(D) \quad \max_{M_\Lambda \succeq 0} \text{tr}(\Lambda). \tag{14}$$

The following lemma relates the KKT conditions to the dual problem and forms the basis of our approach:

Lemma 1 *If R^* is a KKT point with corresponding Lagrange multiplier Λ^* then R^* is globally optimal and there is no duality gap between (P) and (D) if $M_\Lambda \succeq 0$.*

If the dual variable Λ^* can be computed by other means than solving the SDP then $M_\Lambda^* \succeq 0$ can be used as a certificate of optimality. Hence this gives us a way of verifying optimality of a local solution for the PSR problem.

4. Spectral Analysis of M_Λ

In this section we investigate under what conditions we can ensure that $M_\Lambda \succeq 0$ for a KKT point R^*. Our goal is to understand why these problems often give tight relaxations even for relatively high noise levels (see Section 6).

4.1. Decoupling into Noise-free and Noisy Data

We will start by defining D_R as a block diagonal matrix with $(D_R)_{i,i} = R_i^*$. Since D_R is an orthogonal matrix, we have that if $D_R M_\Lambda D_R^T$ is positive semidefinite, then M_Λ is positive semidefinite as well. From (11), we get that each 3×3 block of $D_R M_\Lambda D_R^T$ is given by

$$\left(D_R M_\Lambda D_R^T\right)_{i,j} = \begin{cases} \sum_{k \neq i} R_i^* X_i P X_k^T R_k^{*T}, & i = j \\ -R_i^* X_i P X_j^T R_j^{*T}, & \text{otherwise} \end{cases} \tag{15}$$

From now on we will again drop the $(.)^*$ notation for simplicity. In the absence of noise, $R_k X_k = Y\Theta_k$, we get that the block elements outside of the diagonal of $D_R M_\Lambda D_R^T$ become $-Y\Theta_i P \Theta_j Y^T$ and the diagonal blocks become $\sum_{k \neq i} Y\Theta_i P \Theta_j Y^T$. This results in a positive semidefinite matrix $D_{Y\Theta} \Phi D_{Y\Theta}^T$, where

$$D_{Y\Theta} = \begin{bmatrix} Y\Theta_1 & 0^{3\times n} & \dots \\ 0^{3\times n} & Y\Theta_2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{16}$$

and each $n \times n$ block of Φ is given by

$$\Phi_{i,j} = \begin{cases} I - P, & i = j \\ -P, & \text{otherwise} \end{cases}. \tag{17}$$

To the difference between $D_R M_\Lambda D_R^T$ and $D_{Y\Theta} \Phi D_{Y\Theta}^T$ we will call Δ, which consist of a matrix that captures the noise dependence of $D_R M_\Lambda D_R^T$. As so, we have $D_R M_\Lambda D_R^T = D_{Y\Theta} \Phi D_{Y\Theta}^T + \Delta$.

4.2. Spectral Analysis of the Decoupled Data

The matrix $D_R M_\Lambda D_R^T$ is positive semidefinite if and only if the eigenvalues of $D_R M_\Lambda D_R^T$ are greater or equal to zero. Using Weyl's eigenvalue inequality and our noise-free and noisy decomposition of $D_R M_\Lambda D_R^T$, we can write that

$$\lambda_1\left(D_R M_\Lambda D_R^T\right) \geq \lambda_1(D_{Y\Theta} \Phi D_{Y\Theta}^T) - |\lambda_{\max}(\Delta)| \geq 0. \tag{18}$$

Note that, according to (15), $\sum_j \left(D_R M_\Lambda D_R^T\right)_{i,j} = 0$, meaning that any matrix $\mu N = \mu[I \quad I \quad \dots]^T$, with arbitrary μ belongs to the null space of $D_R M_\Lambda D_R^T$. For a large enough μ, we have that $\lambda_1\left(D_R M_\Lambda D_R^T + \mu N N^T\right) = \lambda_4\left(D_R M_\Lambda D_R^T\right)$. The exact same reasoning can be applied to $D_{Y\Theta} \Phi D_{Y\Theta}^T$, resulting in $\lambda_1\left(D_{Y\Theta} \Phi D_{Y\Theta}^T + \mu N N^T\right) = \lambda_4\left(D_{Y\Theta} \Phi D_{Y\Theta}^T\right)$. And so, by adding $\mu N N^T$ to both sides we get

$$\lambda_4(D_R M_\Lambda D_R^T) \geq \lambda_4(D_{Y\Theta} \Phi D_{Y\Theta}^T) - |\lambda_{\max}(\Delta)| \geq 0. \tag{19}$$

We will use the result from (19) as a lower bound to $\lambda_4(D_R M_\Lambda D_R^T)$, meaning that if we get $\lambda_4(D_{Y\Theta} \Phi D_{Y\Theta}^T) \geq |\lambda_{\max}(\Delta)|$ then M_Λ is positive semidefinite and the KKT point is the global minimum of (P).

Now, as stated in [8], we have that $|\lambda_{\max}(\Delta)| \leq \sum_j \|\Delta_{i,j}\|$, where $\Delta_{i,j} = Y\Theta_i P \Theta_j Y^T - R_i X_i P X_j^T R_j^T$ for $i \neq j$, and $\Delta_{i,i} = -\sum_{i \neq j} \Delta_{i,j}$. By defining

$$\eta := \max_{i \neq j} \quad m \|Y\Theta_i P \Theta_j Y^T - R_i X_i P X_j^T R_j^T\| \tag{20}$$

we can conclude that $|\lambda_{\max}(\Delta)| \leq 2\sum_{i \neq j} \|\Delta_{i,j}\| \leq 2\frac{m-1}{m}\eta$. As so, from (19) we obtain a sufficient condition for global optimality.

Theorem 1 *Given a candidate solution R^* that fulfills (9), there is no duality gap between (P) and (D) if*

$$\epsilon := \frac{\eta}{\lambda_4(D_{Y\Theta} \Phi D_{Y\Theta}^T)} \leq \frac{m}{2(m-1)}. \tag{21}$$

No missing data case For the particular case of no missing data, we have that $\lambda_4(D_{Y\Theta} \Phi D_{Y\Theta}^T)$ can be simplified. Note that, in this case, $D_Y \Phi D_Y^T = (mI_m - \mathbb{1}_m) \otimes \frac{1}{m} YY^T$. The eigenvalue property of the Kronecker product tells us that the eigenvalues of $D_Y \Phi D_Y^T$ are given by $\lambda_i(mI_m - \mathbb{1}_m)\lambda_j(\frac{1}{m}YY^T)$ with $i = 1, \dots, m$ and $j = 1, 2, 3$. The matrix $mI_m - \mathbb{1}_m$ has m eigenvalues, where $\lambda_1(mI_m - \mathbb{1}_m) = 0$ and $\lambda_k(mI_m - \mathbb{1}_m) = m,\ k = 2, \dots, m$. This means that the smallest non-zero eigenvalue of $D_Y \Phi D_Y^T$ is given by $\lambda_4(D_Y \Phi D_Y^T) = \lambda_1(YY^T)$. The upper bound of $|\lambda_{\max}(\Delta)|$ can also be simplified since now we have

$$\eta_0 := \max_{i \neq j} \quad \|YY^T - R_i X_i X_j^T R_j^T\| \tag{22}$$

and consequently a sufficient condition for global optimality for no missing data is obtained.

Corollary 1.1 (No Missing Data) *If all points are visible in all views, i.e., $w_{i,j} = 1,\ \forall i, j$, then, given a candidate solution R^* that fulfills (9), there is no duality gap between (P) and (D) if*

$$\epsilon_0 := \frac{\eta_0}{\lambda_1(YY^T)} \leq \frac{m}{2(m-1)}. \tag{23}$$

4.3. Interpretation and Evaluation

The results Theorem 1 and specially Corollary 1.1 allow us to obtain some interpretation about the (sufficient) conditions for global optimality. For the no missing data case, one can see that global optimality is guaranteed when the ratio between the biggest eigenvalue of the noise dependent η_0 and the smallest eigenvalue is smaller than $\frac{m}{2(m-1)}$. The η_0 encapsulates the noise in the problem by quantifying the difference between the covariance matrix of Y and the cross-correlation matrix of each pair of point sets, over $i \neq j$. Note that in the noise-free case we have $\eta_0 = 0$. A similar interpretation can be extended to the general case with missing data, where the covariance and correlation matrices in (20), as well as the clean data in $\lambda_4(D_{Y\Theta}\Phi D_{Y\Theta}^T)$, are weighted based on the observed data through P.

To evaluate the duality gap and the results from Theorem 1 and Corollary 1.1, we will generate several (100) instances of PSR problems for different noise levels, percentage of missing data, and spatial distribution of the target point cloud. Each instance has $n = 250$ points and $m = 10$ rotations and translations, generated from a Gaussian distribution. The noise in the source point sets is also generated from a Gaussian distribution with standard deviation from 0 to 4 times the standard deviation of the instance's point cloud. For the missing data, we randomly remove $p\%$ of the n points, with $p = \{0, 25, 50\}$. Since ϵ depends on $\lambda_4(D_{Y\Theta}\Phi D_{Y\Theta}^T)$, we also change the instance's point cloud covariance, such that $\frac{\lambda_1(YY^T)}{\lambda_3(YY^T)} = q$, with $q = \{0.1, 0.5, 1\}$. Note that if $q = 0$, the point cloud is contained in a 2D-plane. Regarding the duality gap of the SDP relaxation, we solve the SDP problem similarly to what Eriksson et al. [8] do, and then look at the rank of the obtained solution. If the rank is superior to 3 then the relaxation is non-tight.

For these two experiments, the average rank of ϵ and of the SDP solution over the PSR problem instances are shown in Figure 1 and 2, respectively, and we have sufficient conditions for global optimality when $\epsilon < \frac{m}{2(m-1)} = 0.56$. Since ϵ can only be evaluated for KKT points, when the SDP gives a solution with rank higher than 3, we project it to a rank 3 solution and then perform local optimization (see Section 6.1 for more details about this).

One of the main things to note is the high amount of noise that the SDP relaxation allows before there is duality gap. We can also see that the tightness of the relaxation is highly affected by changes in p and q. In particular, from our experiments it is clear that the bound in Theorem 1 gets looser as $\lambda_1(YY^T) \to 0$, and consequently $\lambda_4(D_{Y\Theta}\Phi D_{Y\Theta}^T) \to 0$, resulting in a reduced usefulness for real problems when the estimated point cloud is close to planar. To improve this, one has to find an alternative decoupling in which the noise-free component does not depend on the point cloud spatial distribution.

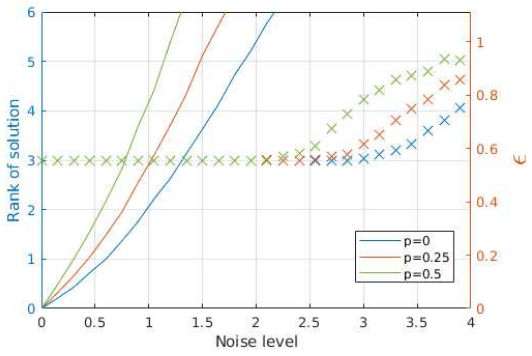


Figure 1: Evaluation of the result from Theorem 1 for 0%, 25% and 50% of missing data and under different noise levels. The plotted data points correspond to the average over the generated problem instances. For this experiment, $\frac{\lambda_1(YY^T)}{\lambda_3(YY^T)} = q$ is fixed as 1.

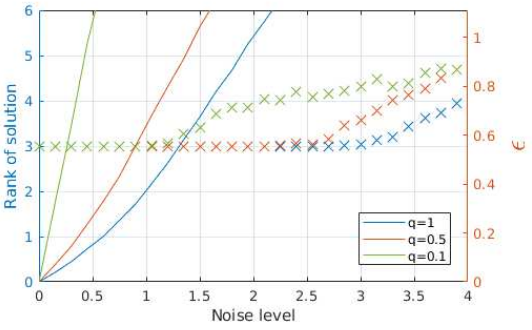


Figure 2: Evaluation of the result from Theorem 1 for $q = \{0.1, 0.5, 1\}$ and under different noise levels. The plotted data points correspond to the average over the generated problem instances. For this experiment, we set $p = 0$ (no missing data).

5. An Alternative Spectral Analysis

In order to find a better noise-free/noisy data decomposition we go back to the matrix $D_R M_\Lambda D_R^T$ defined in (15). This matrix can also be written as $D_R M_\Lambda D_R^T = D_1 - D_2 P D_2^T$ where $D_1 = \text{blkdiag}(R_1 X_1 Y^T, \ldots, R_m X_m Y^T)$ and $D_2 = D_R X$. This means that $D_R M_\Lambda D_R^T$ is the Schur complement $Q \setminus L$ of a matrix

$$Q = \begin{bmatrix} D_1 & D_2 \\ D_2^T & L \end{bmatrix}. \tag{24}$$

From the Schur complement properties, it follows that if $L \succeq 0$ (which is true since it is the Laplacian of a weighted graph as stated in Section 2), then $D_R M_\Lambda D_R^T \succeq 0$ iff $Q \succeq 0$. The matrix Q has other Schur complement, $Q \setminus D1 = L - D_2^T D_1^{-1} D_2$ and similarly, we have that if $D1 \succeq 0$, then $Q \succeq 0$ iff $L - D_2^T D_1^{-1} D_2 \succeq 0$. This means that finding

the sufficient conditions for $D_R M_\Lambda D_R^T \succeq 0$ is equivalent to doing the same for $D_1 \succeq 0$ and $Q \setminus D_1 \succeq 0$. Since D_1 is a block diagonal matrix, $D_1 \succeq 0$ iff $R_k X_k Y^T \succeq 0$, for all $k = 1, \ldots, m$.

The matrix $Q \setminus D_1$ has rank $n-4$, and its null-space its spanned by a all-ones vector $\mathbb{1}$ and Y. Note also that $\mathbb{1}$ and Y are orthogonal since $Y^T \mathbb{1} = 0$.

5.1. Decoupling into Noise-free and Noisy Data

The procedure for decoupling $Q \setminus D_1$ into noise-free and noisy matrices is similar to the one followed in Section 4.1.

In a noise-free situation, we have that $Q \setminus D_1$ becomes $L - \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i$ with $\tilde{C}_i = Y \Theta_i Y^T$. Note that this matrix shares the same rank and null space as $Q \setminus D_1$.

The noisy matrix, here denominated as Δ', can be obtained by subtracting $L - \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i$ from $Q \setminus D_1$, which results in

$$\Delta' = \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i - X_i^T R_i^T C_i^{-1} R_i X_i \tag{25}$$

where $C_i^{-1} = (R_i X_i Y^T)^{-1}$.

5.2. Spectral Analysis of the Decoupled Data

Just as in Section 4.2, we apply the Weyl's eigenvalue inequality to our decoupled data, and use the 4-dimensional null space of $Q \setminus D_1$(and $L - \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i$) to consider the smallest non-zero eigenvalues, resulting in the inequality

$$\lambda_1(Q \setminus D_1) \geq \lambda_5(L - \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i) - \eta' \geq 0, \tag{26}$$

where we define

$$\begin{aligned} \eta' &:= |\lambda_{\max}(\Delta')| = \\ &= \| \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i - X_i^T R_i^T C_i^{-1} R_i X_i \|. \end{aligned} \tag{27}$$

From the eigenvalue inequality (26) we are able to draw a new sufficient condition for global optimality.

Theorem 2 *Given a candidate solution R^* that fulfills (9), if $R_k X_k Y^T \succeq 0$ for all $k = 1, \ldots, m$, then there is no duality gap between (P) and (D) if*

$$\epsilon' := \frac{\eta'}{\lambda_5(L - \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i)} \leq 1. \tag{28}$$

No missing data case In the absence of missing data, we have $L = mI - \mathbb{1}_n$, $Y\Theta_i = Y$ and $\tilde{C}_i = \tilde{C} = YY^T$. In this case we are able to obtain a closed-form solution for $\lambda_5(L - \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i)$ that does not depend on the point cloud spatial distribution, as desired. Since $\mathbb{1}_n = \mathbb{1}\mathbb{1}^T$ and $\mathbb{1}$ belongs to the null space of the noise-free matrix, we get that, for the no missing data case,

$$\lambda_5(L - \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i) = \lambda_4(m(I - Y^T \tilde{C}^{-1} Y)) \tag{29}$$

The term $Y^T \tilde{C}^{-1} Y$ has eigenvalues $\{0, \ldots, 0\, 1, 1, 1\}$, which means that $m(I - Y^T \tilde{C}^{-1} Y)$ will have eigenvalues $\{0, 0, 0, m, \ldots, m\}$. Recall that we have shown that Y belongs to the null-space of the noise-free matrix, by adding $\mu Y^T Y$ we get that the eigenvalues of $m(I - Y^T \tilde{C}^{-1} Y) + \mu Y^T Y$ will be $\{m, \ldots, m, ?, ?, ?\}$, where by ? we represent the eigenvalues of $\mu Y^T Y$ for a sufficiently large μ. This means that the smallest non-zero eigenvalue will be

$$\lambda_5(L - \sum_i \Theta_i Y^T \tilde{C}_i^{-1} Y \Theta_i) = m. \tag{30}$$

The expression for (27) also simplifies as

$$\eta'_0 := \| m Y^T \tilde{C}^{-1} Y - \sum_i X_i^T R_i^T C_i^{-1} R_i X_i \|, \tag{31}$$

resulting in the following corollary.

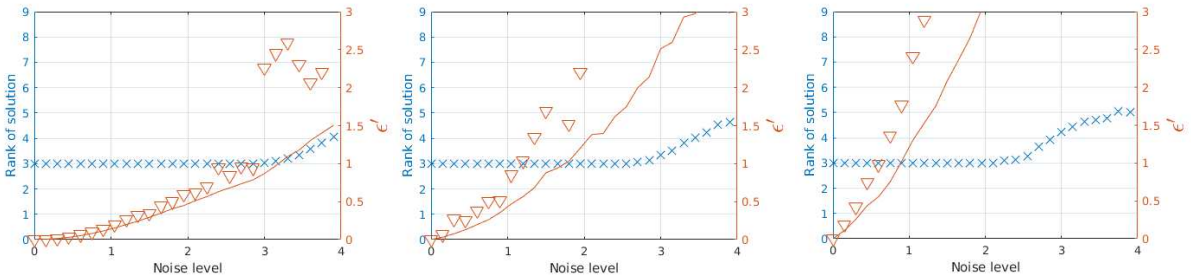
Corollary 2.1 (No Missing Data) *If all points are visible in all views, i.e., $w_{i,j} = 1$, $\forall i, j$, given a candidate solution R^* that fulfills (9), if $R_k X_k Y^T \succeq 0$ for all $k = 1, \ldots, m$, then there is no duality gap between (P) and (D) if*

$$\epsilon'_0 := \frac{\eta'_0}{m} \leq 1. \tag{32}$$

5.3. Interpretation and Evaluation

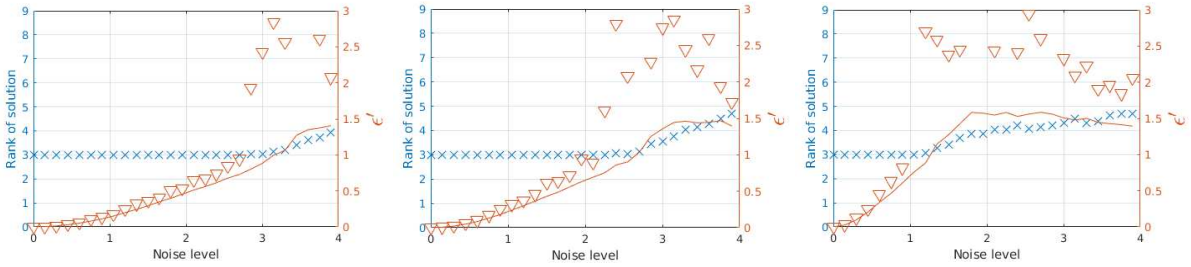
By comparing the results from Theorem 2 and Corollary 2.1 with Theorem 1 and Corollary 1.1, respectively, one can draw some connections. First, and as desired, the noise-free component has now no dependence on the point cloud distribution. Second, while in Theorem 1 we have η as the difference between correlation matrices, weighted by P (graph of observable data), in Theorem 2 we have η' as the difference between the Gram matrices of Y and the source point sets, normalized by the correlation matrices $\tilde{C}_i$ and C_i respectively.

To evaluate the result from Theorem 2 we apply it to the PSR problem instances generated in Section 4.3, under the same noise, missing data and point cloud covariance conditions. The results of the two experiments with the generated data are shown in Figures 3 and 4. We have sufficient conditions for global optimality of the candidate solution if $\epsilon' \leq 1$ and since the bound is tight, we also plot the maximum ϵ' over the instances for each noise level (red triangles) to clarify that in some instances we got $\epsilon' > 1$ even though the average is below 1.



(a) $p = 0$ (b) $p = 0.25$ (c) $p = 0.5$

Figure 3: Evaluation of the result from Theorem 2 under the same conditions as in Figure 1. The continuous red line shows the average of ϵ' over the instances and the red triangles the maximum ϵ' obtain for each noise level.



(a) $q = 1$ (b) $q = 0.5$ (c) $q = 0.1$

Figure 4: Evaluation of the result from Theorem 2 under the same conditions as in Figure 2. The continuous red line shows the average of ϵ' over the instances and the red triangles the maximum ϵ' obtain for each noise level.

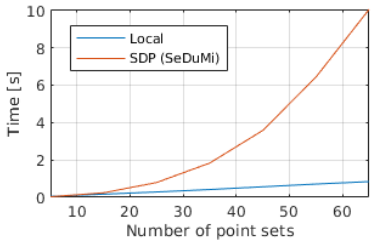


Figure 5: Comparison between the running time of our local optimizer and solving the SDP using SeDuMi [23].

As desired, the result from Theorem 2 is more robust to changes in the target point cloud's shape, giving a tighter bound on ϵ'. Additionally, there is also some improvement in terms of the tightness of the bound for different missing data percentages when compared with the result from Lemma 1.

6. Experiments

Given Lemma 1 and Theorems 1, and 2 we will now apply it to real data. We also give a brief explanation of the local optimizer used for obtaining the candidate solutions.

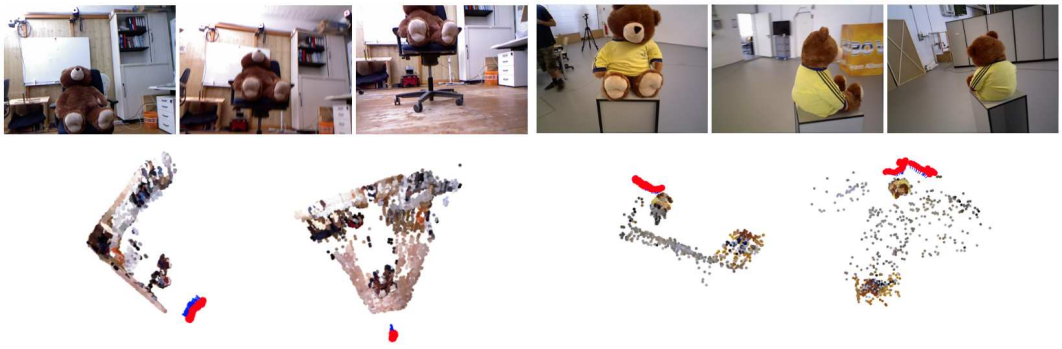
6.1. Local Optimization

One of the main applications of our results is enabling the test for global optimality of a local minimum obtained through some local optimization algorithm. This is particularly advantageous since local optimization provides faster convergence compared to solving the SDP. Figure 5 shows a comparison between the running time of the local optimization method using in our experiments and solving the SDP, for increasing values of m (number of source point sets).

In our simple local optimizer, we linearize rotations as $R \approx (I + a_1 S_1 + a_2 S_2 + a_3 S_3)R$, with

$$S_1 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; S_2 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \; S_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{33}$$

and where a_1, a_2, and a_3 parameterize each rotation. We update the parameters of each rotation independently using the gradient of the linearized cost function of the primal (P), and then project the solution into $\mathcal{SO}(3)$ using the exponential map $R \leftarrow e^{a_1 S_1 + a_2 S_2 + a_3 S_3} R$. As a stopping criterion, we used a threshold on the average magnitude of the gradient. Note that this is a simplistic local optimizer for this problem, and that more efficient methods can (and should) be used to solve it.



(a) (b)

Figure 6: (a) Evaluation of our results in the *fr1/teddy* sequence of the TUM RGB-D dataset ($m = 47$, $n = 3606$). The solution obtained with the local optimizater is the global minimum of (P) since it resulted in a psd M_Λ and $\epsilon' = 0.1769$. We also got $\epsilon = 2.0963$; (b) Evaluation of our results in the *fr3/teddy* sequence of the TUM RGB-D dataset ($m = 50$, $n = 2812$). The solution obtained with the local optimizater resulted in a non-psd M_Λ, $\epsilon = 2157.6$ and $\epsilon' = 714.84$.

6.2. Evaluation of Results using Real Data

In order to show that the results from Lemma 1 and Theorems 1 and 2 can be applied to real problems, we use the local optimizer to reconstruct part of two sequences of the TUM RGB-D dataset [22]: *fr1/teddy* and *fr3/teddy*. The correspondences between images in each sequence are obtained by matching SIFT features [15], using MATLAB [17] and VLFeat library [27]. Three of the images used and two different points of view of the reconstruction scene are shown in Figures 6a and 6b.

For the *fr1/teddy* sequence we use the first 47 images to get the point clouds and correspondences, resulting in $n = 3606$ points. The result obtained by using our local optimizer resulted in $M_\Lambda \succeq 0$, which by Lemma 1 directly implies that it is a global maximum of (P). We also evaluate the results from Theorems 1 and 2, which give $\epsilon = 2.0963$ and $\epsilon' = 0.1769$, respectively, and the latter also provide sufficient conditions for global optimality of the candidate solution. The fact that the sufficient condition related to ϵ is not fulfilled shows, even though valid, how sensible this condition is to missing data and spatial distribution of the point cloud.

For the *fr3/teddy* sequence we take 1 in every 6 images, perform the SIFT matching and end up with $n = 2812$ points. For our candidate solution obtained with the local optimizer we got a non-psd M_Λ, $\epsilon = 2157.6$ and $\epsilon' = 714.84$. We also solved the SDP and the solution obtained had rank 4, which shows that the SDP relaxation for this sequence is non-tight. This sequence differs from the previous by having more movement and flow of points, resulting in more missing data and less shared points between source point sets, and as it is possible to see in Figure 1 and 2 these factors make the SDP relaxation less tight for lower amounts of noise.

7. Conclusions

In this paper we study global optimality conditions for the Point Set Registration problem through SDP relaxation. We use Lagragian duality to determine a sufficient condition, Lemma 1, that guarantee that a candidate solution is a global minimum of the primal problem. Then, through spectral analysis, we decouple the problem into noise-free and noisy components, and define bounds (Theorems 1 and 2) that can also be used to test for global optimality of a candidate solution. In particular, for the no missing data scenario, we present Corollaries 1.1 and 2.1 that can be efficiently computed.

The experiments show that the SDP relaxation is tight for relatively high amounts of noise in the source point sets, but it degrades as the target point cloud becomes close to planar and as we increase the amount of missing data in the problem. This restricts the use of SDP relaxation for PSR from scenarios where the reconstructed scene is roughly planar or where there are few shared points between several source points sets.

However, for scenarios where most of the points are observable, the SDP relaxation is tight and our results can be applied as a certificate for global optimality of a candidate solution.

Acknowledgements

This work was supported by the Swedish Research Council (grants no. 2016-04445 and no. 2018-05375) and the Swedish Foundation for Strategic Research (Semantic Mapping and Visual Navigation for Smart Robots).

References

[1] Mica Arie-Nachimson, Shahar Z Kovalsky, Ira Kemelmacher-Shlizerman, Amit Singer, and Ronen Basri. Global motion estimation from point matches. In *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2012. 1

[2] Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello. Global registration of 3d point sets via lrs decomposition. In *ECCV*, 2016. 1

[3] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(5):698–700, Sep. 1987. 1

[4] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, Feb 1992. 1

[5] Uttaran Bhattacharya and Venu Madhav Govindu. Efficient and robust registration on the 3d special euclidean group. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 1

[6] Jesus Briales and Javier Gonzalez-Jimenez. Convex global 3d registration with lagrangian duality. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, jul 2017. 1

[7] Kunal N Chaudhury, Yuehaw Khoo, and Amit Singer. Global registration of multiple point clouds using semidefinite programming. *SIAM Journal on Optimization*, 25(1):468–501, 2015. 1, 2

[8] Anders P. Eriksson, Carl Olsson, Fredrik Kahl, and Tat-Jun Chin. Rotation averaging and strong duality. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 127–135, 2018. 1, 2, 3, 4, 5

[9] Johan Fredriksson and Carl Olsson. Simultaneous multiple rotation averaging using Lagrangian duality. In *Asian Conference on Computer Vision*, 2012. 1

[10] V. Govindu. Combining two-view constraints for motion estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2001. 1

[11] V. M. Govindu and A. Pooja. On averaging multiview relations for 3d scan registration. *IEEE Transactions on Image Processing*, 23(3):1289–1302, 2014. 1

[12] R. Hartley, J. Trumpf, and Y. Dai. Rotation averaging and weak convexity. In *International Symposium on Mathematical Theory of Networks and Systems*, 2010. 1

[13] Berthold K. P. Horn, Hugh M. Hilden, and Shahriar Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *J. Opt. Soc. Am. A*, 5(7):1127–1135, Jul 1988. 1

[14] H. Li and R. Hartley. The 3d-3d registration problem revisited. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, Oct 2007. 1

[15] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov 2004. 8

[16] D. Martinec and T. Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007. 1

[17] MATLAB. *version 9.5.0 (R2018b)*. The MathWorks Inc., Natick, Massachusetts, 2018. 8

[18] Carl Olsson and Anders P. Eriksson. Solving quadratically constrained geometrical problems using lagrangian duality. In *International Conference on Pattern Recognition (ICPR)*, 2008. 1

[19] David M. Rosen, Luca Carlone, Afonso S. Bandeira, and John J. Leonard. SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group. *CoRR*, abs/1612.07386, 2016. 1

[20] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. *Proc. 3DIM*, 2001, 10 2001. 1

[21] G. C. Sharp, S. W. Lee, and D. K. Wehe. Multiview registration of 3d scenes by minimizing error between coordinate frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1037–1050, Aug 2004. 1

[22] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d slam systems. In *Proc. of the International Conference on Intelligent Robot Systems (IROS)*, Oct. 2012. 8

[23] Jos F. Sturm and Gerardo A. Guerra. A matlab toolbox for optimization over symmetric cones. 1999. 7

[24] A. Torsello, E. Rodolà, and A. Albarelli. Multiview registration via graph diffusion of dual quaternions. In *CVPR 2011*, pages 2441–2448, 2011. 1

[25] P. Van Mieghem, K. Devriendt, and H. Cetinay. Pseudoinverse of the laplacian and best spreader node in a network. *Phys. Rev. E*, 96:032311, Sep 2017. 3

[26] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Rev.*, 38(1):49–95, Mar. 1996. 1

[27] Andrea Vedaldi and Brian Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. volume 3, pages 1469–1472, 10 2010. 8

[28] Heng Yang and Luca Carlone. A quaternion-based certifiably optimal solution to the wahba problem with outliers. pages 1665–1674, 10 2019. 1

[29] Lei Zhang, Ligang Liu, Craig Gotsman, and Steven J. Gortler. An as-rigid-as-possible approach to sensor network localization. *ACM Trans. Sen. Netw.*, 6(4):35:1–35:21, July 2010. 1

[30] Y. Zhong and N. Boumal. Near-optimal bounds for phase synchronization. *ArXiv e-prints*, Mar. 2017. 1