Statistically Consistent Total Least-Squares Estimation of Object Scales*

Arne Hasselbring[0000-0003-0350-9817] and Udo Frese[0000-0001-8325-6324]

University of Bremen, Faculty 3 – Mathematics and Computer Science,
Postfach 330 440, 28334 Bremen, Germany
{arha,ufrese}@uni-bremen.de

Abstract. Estimating object poses is a fundamental problem in computer vision in general as well as for robotic manipulation in particular. Most approaches require a known 3D model of the object. One step towards a more general formulation is to estimate the object's width, height and depth with the pose, e. g. consider a generic box, cylinder or plate instead of one with known dimensions.
This paper investigates the last stage of such a pipeline, namely least-squares estimating pose and scales from point correspondences aggregated into a fixed size matrix. Therefore it encapsulates the scaled $SO(3)$ manifold in a so-called $\boxplus$-operator and derives a Gauss-Newton based optimizer with initial guess on that.
We find that the resulting estimator is strongly biased towards small scales. This is due to the structure of the least-squares loss: Noise in recognized object points is multiplied with the to be estimated transformation matrix, violating the additive noise assumption. It has no effect in the prevalent use of this loss for pose estimation but affects the scale. We propose a solution to this bias based on an approximation of total least-squares that preserves the advantage of a fixed size representation and show that it provides relatively consistent uncertainty estimates.

Keywords: Least-Squares Estimation · 9D Object Pose Estimation · RGB-D Perception

A Well-Definedness of ρ and σ

σ is well-defined since $\mathbf{Q}^\top\mathbf{Q}$ is diagonal and positive (by definition of $\mathcal{Q}$), and thus its square-root is well-defined and still diagonal and positive. ρ is well-defined since the inverse of a positive diagonal matrix always exists. The result of ρ is indeed in $SO(3)$ since

$$\rho(\mathbf{Q})^\top\rho(\mathbf{Q}) = \sigma(\mathbf{Q})^{-1}\mathbf{Q}^\top\mathbf{Q}\sigma(\mathbf{Q})^{-1} = \sigma(\mathbf{Q})^{-1}\sigma(\mathbf{Q})^2\sigma(\mathbf{Q})^{-1} = \mathbf{I} \tag{1}$$

$$\det(\rho(\mathbf{Q})) = \frac{\det(\mathbf{Q})}{\det(\sigma(\mathbf{Q}))} = \frac{\det(\mathbf{Q})}{\sqrt{\det(\mathbf{Q}^\top\mathbf{Q})}} = \frac{\det(\mathbf{Q})}{|\det(\mathbf{Q})|} = 1. \tag{2}$$

* This is the supplemental material of a paper accepted at the ROBOVIS 2025 conference. The entire paper will be made available after the embargo period of 12 months.

B Proof of the $\boxplus$-Axioms for the Scaled $SO(3)$

The original publication [1] required the axiom that the zero-vector is the neutral element of $\boxplus$. This can actually be derived from the other axioms and is trivial in this case anyway.

The first axiom requires surjectivity of $\boxplus$ (all states can reach any other state by a single $\boxplus$-step).

$$\forall \mathbf{Q}_1, \mathbf{Q}_2 \in \mathcal{Q}\colon \mathbf{Q}_1 \boxplus (\mathbf{Q}_2 \boxminus \mathbf{Q}_1) = \mathbf{Q}_2 \tag{3}$$

Proof.

$$\begin{aligned}
&\mathbf{Q}_1 \boxplus (\mathbf{Q}_2 \boxminus \mathbf{Q}_1)\\
=&\exp_\times (\mathbf{Q}_2 \boxminus \mathbf{Q}_1)_\times \mathbf{Q}_1 \exp_\circ (\mathbf{Q}_2 \boxminus \mathbf{Q}_1)_\circ\\
=&\exp_\times \log_\times (\rho(\mathbf{Q}_2)\rho(\mathbf{Q}_1)^{-1}) \mathbf{Q}_1 \exp_\circ \log_\circ (\sigma(\mathbf{Q}_1)^{-1}\sigma(\mathbf{Q}_2))\\
=&\mathbf{Q}_2 \sigma(\mathbf{Q}_2)^{-1}\sigma(\mathbf{Q}_1)\mathbf{Q}_1^{-1}\mathbf{Q}_1\sigma(\mathbf{Q}_1)^{-1}\sigma(\mathbf{Q}_2)\\
=&\mathbf{Q}_2
\end{aligned}$$

The second axiom requires local injectivity of $\boxplus$ (within some neighborhood V of a state, the tangent vector transferring to another state is unique).

$$\forall \mathbf{Q} \in \mathcal{Q}, \delta \in V\colon (\mathbf{Q} \boxplus \delta) \boxminus \mathbf{Q} = \delta \tag{4}$$

For this, we need the following lemma

$$\sigma(\mathbf{Q} \boxplus \delta) = \sigma(\mathbf{Q}) \exp_\circ \delta_\circ \tag{5}$$

which is proven by using the diagonality of $\exp_\circ$ and $\mathbf{Q}^\top \mathbf{Q}$ and orthogonality of $\exp_\times$:

$$\begin{aligned}
&\sigma(\mathbf{Q} \boxplus \delta)\\
=&\sigma(\exp_\times \delta_\times \mathbf{Q} \exp_\circ \delta_\circ)\\
=&\sqrt{(\exp_\circ \delta_\circ)^\top \mathbf{Q}^\top (\exp_\times \delta_\times)^\top \exp_\times \delta_\times \mathbf{Q} \exp_\circ \delta_\circ}\\
=&\sqrt{\exp_\circ \delta_\circ \mathbf{Q}^\top \mathbf{Q} \exp_\circ \delta_\circ}\\
=&\sqrt{\exp_\circ \delta_\circ} \sigma(\mathbf{Q}) \sqrt{\exp_\circ \delta_\circ}\\
=&\sigma(\mathbf{Q}) \exp_\circ \delta_\circ
\end{aligned}$$

Proof.

$$\begin{aligned}
&(\mathbf{Q} \boxplus \delta) \boxminus \mathbf{Q}\\
=&(\log_\times (\rho(\mathbf{Q} \boxplus \delta)\rho(\mathbf{Q})^{-1}), \log_\circ (\sigma(\mathbf{Q})^{-1}\sigma(\mathbf{Q} \boxplus \delta)))^\top\\
=&(\log_\times (\rho(\mathbf{Q} \boxplus \delta)\rho(\mathbf{Q})^{-1}), \log_\circ \exp_\circ (\delta_\circ))^\top\\
=&(\log_\times (\rho(\mathbf{Q} \boxplus \delta)\rho(\mathbf{Q})^{-1}), \delta_\circ)^\top
\end{aligned}$$

$$
\begin{aligned}
&=(\log_\times((\mathbf{Q}\boxplus\delta)\sigma(\mathbf{Q}\boxplus\delta)^{-1}\sigma(\mathbf{Q})\mathbf{Q}^{-1}),\delta_\circ)^\top\\
&=(\log_\times((\mathbf{Q}\boxplus\delta)(\exp_\circ\delta_\circ)^{-1}\sigma(\mathbf{Q})^{-1}\sigma(\mathbf{Q})\mathbf{Q}^{-1}),\delta_\circ)^\top\\
&=(\log_\times(\exp_\times\delta_\times\mathbf{Q}\exp_\circ\delta_\circ(\exp_\circ\delta_\circ)^{-1}\mathbf{Q}^{-1}),\delta_\circ)^\top\\
&=(\log_\times\exp_\times\delta_\times,\delta_\circ)^\top\\
&=(\delta_\times,\delta_\circ)^\top\\
&=\delta
\end{aligned}
$$

The cancellation of $\log_\times$ against $\exp_\times$ in the penultimate step is what requires V to be restricted to angles up to π.

The third axiom requires 1-Lipschitzness of the family of functions $f_{\mathbf{Q}}\colon\mathbb{R}^6\to\mathcal{Q};\quad\delta\mapsto\mathbf{Q}\boxplus\delta$:

$$
\forall\mathbf{Q}\in\mathcal{Q},\delta_1,\delta_2\in\mathbb{R}^9\colon\|(\mathbf{Q}\boxplus\delta_1)\boxminus(\mathbf{Q}\boxplus\delta_2)\|\leq\|\delta_1-\delta_2\| \tag{6}
$$

Proof.

$$
\begin{aligned}
&\|(\mathbf{Q}\boxplus\delta_1)\boxminus(\mathbf{Q}\boxplus\delta_2)\|^2\\
=&\|\log_\times(\rho(\mathbf{Q}\boxplus\delta_1)\rho(\mathbf{Q}\boxplus\delta_2)^{-1})\|^2\\
+&\|\log_\circ(\sigma(\mathbf{Q}\boxplus\delta_2)^{-1}\sigma(\mathbf{Q}\boxplus\delta_1))\|^2\\
=&\|\log_\times((\mathbf{Q}\boxplus\delta_1)\sigma(\mathbf{Q}\boxplus\delta_1)^{-1}\sigma(\mathbf{Q}\boxplus\delta_2)(\mathbf{Q}\boxplus\delta_2)^{-1})\|^2\\
+&\|\log_\circ((\exp_\circ\delta_{2,\circ})^{-1}\sigma(\mathbf{Q})^{-1}\sigma(\mathbf{Q})\exp_\circ\delta_{1,\circ}))\|^2\\
=&\|\log_\times((\mathbf{Q}\boxplus\delta_1)(\exp_\circ\delta_1)^{-1}\sigma(\mathbf{Q})^{-1}\sigma(\mathbf{Q})(\exp_\circ\delta_2)(\mathbf{Q}\boxplus\delta_2)^{-1})\|^2\\
+&\|\log_\circ((\exp_\circ-\delta_{2,\circ})\exp_\circ\delta_{1,\circ})\|^2\\
=&\|\log_\times((\mathbf{Q}\boxplus\delta_1)(\exp_\circ\delta_1)^{-1}(\exp_\circ\delta_2)(\mathbf{Q}\boxplus\delta_2)^{-1})\|^2\\
+&\|\log_\circ\exp_\circ(\delta_{1,\circ}-\delta_{2,\circ})\|^2\\
=&\|\log_\times(\exp_\times\delta_1\mathbf{Q}\exp_\circ\delta_1(\exp_\circ\delta_1)^{-1}(\exp_\circ\delta_2)(\exp_\circ\delta_2)^{-1}\mathbf{Q}^{-1}(\exp_\times\delta_2)^{-1})\|^2\\
+&\|\delta_{1,\circ}-\delta_{2,\circ}\|^2\\
=&\|\log_\times(\exp_\times\delta_1(\exp_\times\delta_2)^{-1})\|^2+\|\delta_{1,\circ}-\delta_{2,\circ}\|^2\\
\leq&\|\delta_{1,\times}-\delta_{2,\times}\|^2+\|\delta_{1,\circ}-\delta_{2,\circ}\|^2\\
=&\|\delta_1-\delta_2\|^2
\end{aligned}
$$

The $\leq$ in the penultimate step is justified by the fact that $SO(3)$ is a $\boxplus$-manifold [1].

In conclusion, the scaled $SO(3)$ with the given operators is a $\boxplus$-manifold.

C Jacobian of $\boxplus$

The Jacobian of the $\boxplus$-operator on $\mathcal{T}$, as used in the Gauss-Newton iteration, is the following:

$$\mathbf{J}_{\boxplus,\delta}(\mathbf{T}) = \begin{pmatrix} 0 & T_{31} & -T_{21} & T_{11} & 0 & 0 & 0 & 0 & 0 & T_{11} \\ 0 & T_{32} & -T_{22} & 0 & T_{12} & 0 & 0 & 0 & 0 & T_{12} \\ 0 & T_{33} & -T_{23} & 0 & 0 & T_{13} & 0 & 0 & 0 & T_{13} \\ -T_{31} & 0 & T_{11} & T_{21} & 0 & 0 & 0 & 0 & 0 & T_{21} \\ -T_{32} & 0 & T_{12} & 0 & T_{22} & 0 & 0 & 0 & 0 & T_{22} \\ -T_{33} & 0 & T_{13} & 0 & 0 & T_{23} & 0 & 0 & 0 & T_{23} \\ T_{21} & -T_{11} & 0 & T_{31} & 0 & 0 & 0 & 0 & 0 & T_{31} \\ T_{22} & -T_{12} & 0 & 0 & T_{32} & 0 & 0 & 0 & 0 & T_{32} \\ T_{23} & -T_{13} & 0 & 0 & 0 & T_{33} & 0 & 0 & 0 & T_{33} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & T_{14} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & T_{24} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & T_{34} \end{pmatrix} \tag{7}$$

Recall that the columns correspond to the three rotation, scaling and translation parameters, respectively, and the final column captures the constant part of the linearization. The rotation parameters affect $\mathbf{Q}$ according to the cross product matrix pattern. The scaling parameters act on the columns of $\mathbf{Q}$. The translation parameters are direct offsets to the translation of $\mathbf{T}$.

D Flattening certain Expressions Involving T

To derive the denominator expression in the total least squares approach, in all sensor models we needed to express $\operatorname{tr} \mathbf{W}\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}^\top\mathbf{W}^\top$ as $\bar{\mathbf{T}}^\top\boldsymbol{\Omega}^L\bar{\mathbf{T}}$, i. e. flatten it into our fixed sized representation.

Lemma 1. *Let* $\mathbf{W} \in \mathbb{R}^{d\times 4}$, $\mathbf{T} \in \mathbb{R}^{4\times 4}$ *with* $\mathbf{T}_{4\bullet} = \begin{pmatrix} 0 \; 0 \; 0 \; 1 \end{pmatrix}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{4\times 4}$, *symmetric positive semidefinite with* $\boldsymbol{\Sigma}_{44} = 0$. *Then*

$$\operatorname{tr}\left(\mathbf{W}\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}^\top\mathbf{W}^\top\right) = \bar{\mathbf{T}}^\top\boldsymbol{\Omega}^L\bar{\mathbf{T}}, \quad \textit{with} \quad \begin{pmatrix} (\mathbf{W}^\top\mathbf{W})_\blacksquare \otimes \boldsymbol{\Sigma}_\blacksquare & \mathbf{0}_{9\times 1} & \mathbf{0}_{9\times 3} \\ \mathbf{0}_{1\times 9} & 0 & \mathbf{0}_{1\times 3} \\ \mathbf{0}_{3\times 9} & \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 3} \end{pmatrix}. \tag{8}$$

Proof.

$$\bar{\mathbf{T}}^\top\boldsymbol{\Omega}^L\bar{\mathbf{T}} = \operatorname{tr}\left(\mathbf{W}\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}^\top\mathbf{W}^\top\right) = \operatorname{tr}\left(\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}^\top\mathbf{W}^\top\mathbf{W}\right) \tag{9}$$

$$= \sum_{k,l,m,n=1}^{4,4,4,4} \mathbf{T}_{kl}\boldsymbol{\Sigma}_{lm}\mathbf{T}_{nm}(\mathbf{W}^\top\mathbf{W})_{nk} \tag{10}$$

$$= \sum_{k,l,m,n=1}^{3,3,3,3} \mathbf{T}_{kl}\boldsymbol{\Sigma}_{lm}\mathbf{T}_{nm}(\mathbf{W}^\top\mathbf{W})_{nk} \tag{11}$$

The last step holds, because from positive definiteness, the whole fourth row and column of $\mathbf{\Sigma}$ is zero, so $l = 4$ and $m = 4$ can be omitted. The same holds for $k = 4$ and $n = 4$, because $\mathbf{T}_{kl}$ or $\mathbf{T}_{nm}$ are zero.

We consider the coefficients for different products of $\mathbf{T}$ elements. There is no constant and no linear term. Each quadratic term $\mathbf{T}_{kl}\mathbf{T}_{nm}$ is multiplied with $\mathbf{\Sigma}_{lm}(\mathbf{W}^\top\mathbf{W})_{kn}$, using symmetry of $\mathbf{W}^\top\mathbf{W}$.

By flattening $\mathbf{T}$ into $\bar{\mathbf{T}}$, its row-indices k and n have stride 3 in the rows and columns of $\mathbf{\Omega}^L$ respectively. Both address an element of $\mathbf{\Sigma}_i^O$. The column-indices l and m have stride 1 and address an element of $\mathbf{W}^\top\mathbf{W}$. The two elements are multiplied. This is conveniently expressed with a Kronecker product, leading to the formula (8).

References

1. Hertzberg, C., Wagner, R., Frese, U., Schröder, L.: Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds. Information Fusion **14**(1), 57–77 (Jan 2013)