In Perfect Shape: Certifiably Optimal 3D Shape Reconstruction from 2D Landmarks

Heng Yang and Luca Carlone
Laboratory for Information & Decision Systems (LIDS)
Massachusetts Institute of Technology
{hankyang, lcarlone}@mit.edu

Abstract

We study the problem of 3D shape reconstruction from 2D landmarks extracted in a single image. We adopt the 3D deformable shape model and formulate the reconstruction as a joint optimization of the camera pose and the linear shape parameters. Our first contribution is to apply Lasserre's hierarchy of convex Sums-of-Squares (SOS) relaxations to solve the shape reconstruction problem and show that the SOS relaxation of minimum order 2 empirically solves the original non-convex problem exactly. *Our second contribution is to exploit the structure of the polynomial in the objective function and find a reduced set of basis monomials for the SOS relaxation that significantly decreases the size of the resulting semidefinite program (SDP) without compromising its accuracy. These two contributions, to the best of our knowledge, lead to the* first certifiably optimal *solver for 3D shape reconstruction, that we name* Shape*. *Our third contribution is to add an outlier rejection layer to* Shape* *using a truncated least squares (TLS) robust cost function and leveraging* graduated non-convexity *to solve TLS* without initialization. *The result is a robust reconstruction algorithm, named* Shape#, *that tolerates a large amount of outlier measurements. We evaluate the performance of* Shape* *and* Shape# *in both simulated and real experiments, showing that* Shape* *outperforms local optimization and previous convex relaxation techniques, while* Shape# *achieves state-of-the-art performance and is robust against* 70% *outliers in the* FG3DCar *dataset.*

1. Introduction

3D object detection and pose estimation from a single image is a fundamental problem in computer vision. Despite the progress in semantic segmentation [11], depth estimation [20], and pose estimation [16, 43], reconstructing the 3D shape and pose of an object from a single image remains a challenging task [2, 49, 37, 42, 18, 35].

A typical approach for 3D shape reconstruction is to first detect 2D landmarks in a single image, and then solve a model-based optimization to lift the 2D landmarks to form a 3D model [48, 49, 35, 25, 40]. For the optimization to be well-posed, the unknown shape is assumed to be a *3D deformable model*, composed by a linear combination of *basis shapes*, handcrafted or learned from a large corpus of training data [8]. The optimization then seeks to jointly optimize the coefficients of the linear combination (*shape parameters*) and the *camera pose* to minimize the *reprojection errors* between the 3D model and the 2D landmarks. This model-based paradigm has been successful in several applications such as face recognition [4, 10], car model fitting [25, 12], and human pose estimation [49, 35].

Despite its long history and broad range of applications, there is still no *globally optimal* solver for the non-convex optimization problem arising in 3D shape reconstruction. Therefore, most existing solutions adopt a local optimization strategy, which alternates between solving for the camera pose and the shape parameters. These techniques, as shown in prior works [35, 12], require an initial guess for the solution and often get stuck in local minima. In addition, 2D landmark detectors are prone to produce *outliers*, causing existing methods to be brittle [40]. Therefore, the motivation for this paper is two-fold: (i) to develop a *certifiably optimal* shape reconstruction solver, and (ii) to develop a robust reconstruction algorithm that is insensitive to a large amount of outlier 2D measurements (*e.g.*, 70%).

Contributions. Our first contribution is to formulate the shape reconstruction problem as a *polynomial optimization* problem and apply Lasserre's hierarchy of *Sums-of-Squares* (SOS) relaxations to relax the non-convex polynomial optimization into a convex *semidefinite program* (SDP). We show the SOS relaxation of minimum order 2 empirically solves the non-convex shape reconstruction problem exactly and provides a global optimality certificate. The second contribution is to apply *basis reduction*, a technique that exploits the sparse structure of the polynomial in the objective function, to reduce the size of the resulting SDP. We show that basis reduction significantly improves the efficiency of the SOS relaxation without compromising global optimality. To the best of our knowledge, this is the first certifiably optimal solver for shape reconstruction, and we name it Shape*. Our third contribution is to robustify Shape* by adopting a truncated least squares (TLS) robust cost function and solving the resulting robust estimation problem us-

ing graduated non-convexity [3]. The resulting algorithm, named Shape#, is robust against 70% outliers and does not require an initial guess.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 introduces notation and preliminaries on SOS relaxations. Section 4 introduces the shape reconstruction problem. Section 5 develops our SOS solver (Shape*). Section 6 presents an algorithm (Shape#) to robustify the SOS relaxation against outliers. Section 7 provides experimental results in both simulations and real datasets, while Section 8 concludes the paper.

2. Related Work

We limit our review to optimization-based approaches for 3D shape reconstruction from 2D landmarks. The interested reader can find a review of end-to-end shape and pose reconstruction using deep learning in [18, 37, 17].

Local Optimization. Most existing methods resort to local optimization to solve the non-convex joint optimization of shape parameters and camera pose. Blanz and Vetter [4] propose a method for face recognition by fitting a morphable model of the 3D face shape and texture to a single image using stochastic Newton's method to escape local minima. Gu and Kanade [10] align a deformable point-based 3D face model by alternatively deforming the 3D model and updating the 3D pose. Using similar alternating optimization, Ramakrishna *et al.* [35] tackle 3D human pose estimation by finding a sparse set of basis shapes from an over-complete human shape dictionary using projected matching pursuit; the approach is further improved by Fan *et al.* [9] to include pose locality constraints. Lin *et al.* [25] demonstrate joint 3D car model fitting and fine-grained classification; car model fitting in cluttered images is investigated in [12]. To mitigate the impact of outlying 2D landmarks, Li *et al.* [24] propose a RANSAC-type method for car model fitting and Wang *et al.* [40] replace the least squares estimation with an ℓ_1-norm minimization.

Convex Relaxation. More recently, Zhou *et al.* [48] develop a convex relaxation, where they first over-parametrize the 3D deformable shape model by associating one rotation with each basis and then relax the resulting Stiefel manifold constraint to its convex envelope. Although showing superior performance compared to local optimization, the convex relaxation in [48] comes with no optimality guarantee and is typically loose in practice. In addition, Zhou *et al.* [49] model outliers using a sparse matrix and augment the optimization with an ℓ_1 regularization to achieve robustness against 40% outliers. In contrast, we will show that our convex relaxation comes with certifiable optimality, and our robust reconstruction approach can handle 70% outliers.

3. Notation and Preliminaries

We use $\mathcal{S}^n$ to denote the set of $n \times n$ symmetric matrices. We write $\boldsymbol{A} \in \mathcal{S}^n_+$ (resp. $\boldsymbol{A} \in \mathcal{S}^n_{++}$) to denote that the matrix $\boldsymbol{A} \in \mathcal{S}^n$ is positive semidefinite (PSD) (resp. positive definite (PD)). Given $\boldsymbol{x} = [x_1, \dots, x_n]^\mathsf{T}$, we let $\mathbb{R}[\boldsymbol{x}]$ (resp. $\mathbb{R}[\boldsymbol{x}]_d$) be the ring of polynomials in n variables with real coefficients (resp. with degree at most d), and $[\boldsymbol{x}]_d$ be the vector of all $\binom{n+d}{d}$ monomials with degree up to d.

We now give a brief summary of SOS relaxations for polynomial optimization. Our review is based on [5, 30, 22]. We first introduce the notion of SOS polynomial.

Definition 1 (SOS Polynomial [5]) *A polynomial $p(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]_{2d}$ is said to be a sums-of-squares (SOS) polynomial if there exist polynomials $q_1, \dots, q_m \in \mathbb{R}[\boldsymbol{x}]_d$ such that:*

$$p(\boldsymbol{x}) = \sum_{i=1}^{m} q_i^2(\boldsymbol{x}). \tag{1}$$

We use Σ_n (resp. $\Sigma_{n,2d}$) to denote the set of SOS polynomials in n variables (resp. with degree at most $2d$). A polynomial $p(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]_{2d}$ is SOS if and only if there exists a PSD matrix $\boldsymbol{Q} \in \mathcal{S}^{N_Q}_+$ with $N_Q = \binom{n+d}{d}$, such that:

$$p(\boldsymbol{x}) = [\boldsymbol{x}]_d^\mathsf{T} \boldsymbol{Q} [\boldsymbol{x}]_d, \tag{2}$$

and $\boldsymbol{Q}$ is called the Gram matrix of $p(\boldsymbol{x})$.

Now consider the following polynomial optimization:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ s.t. \quad & h_i(\boldsymbol{x}) = 0, i = 1, \dots, m, \\ & g_k(\boldsymbol{x}) \geq 0, k = 1, \dots, l, \end{aligned} \tag{3}$$

where $f, h_i, g_k \in \mathbb{R}[\boldsymbol{x}]$ are all polynomials and let $\mathcal{X}$ be the feasible set defined by h_i, g_k. For convenience, denote $\boldsymbol{h} := (h_1, \dots, h_m)$, $g_0 := 1$ and $\boldsymbol{g} = (g_0, \dots, g_l)$. We call

$$\langle \boldsymbol{h} \rangle := \{h \in \mathbb{R}[\boldsymbol{x}] : h = \textstyle\sum_{i=1}^m \lambda_i h_i, \lambda_i \in \mathbb{R}[\boldsymbol{x}]\}, \tag{4}$$

$$\langle \boldsymbol{h} \rangle_{2\beta} := \{h \in \langle \boldsymbol{h} \rangle : \deg(\lambda_i h_i) \leq 2\beta\}, \tag{5}$$

the *ideal* and the *2β-th truncated ideal* of $\boldsymbol{h}$, where $\deg(\cdot)$ is the degree of a polynomial. The *ideal* is simply a summation of polynomials with polynomial coefficients, a construct that will simplify the notation later on. We call

$$Q(\boldsymbol{g}) := \{g \in \mathbb{R}[\boldsymbol{x}] : g = \textstyle\sum_{k=0}^m s_k g_k, s_k \in \Sigma_n\}, \tag{6}$$

$$Q_\beta(\boldsymbol{g}) := \{g \in Q(\boldsymbol{g}) : \deg(s_k g_k) \leq 2\beta\}, \tag{7}$$

the *quadratic module* and the *β-th truncated quadratic module* generated from $\boldsymbol{g}$. Note that the quadratic module is similar to the *ideal*, except now we require the polynomial coefficients to be SOS. Apparently, if $p(\boldsymbol{x}) \in \langle \boldsymbol{h} \rangle + Q(\boldsymbol{g})$,

then $p(\boldsymbol{x})$ is nonnegative on $\mathcal{X}$[1]. Putinar's Positivstellensatz [34] describes when the reverse is also true.

Theorem 2 (Putinar's Positivstellensatz [34]) *Let $\mathcal{X}$ be the feasible set of problem (3). Assume $\langle \boldsymbol{h} \rangle + Q(\boldsymbol{g})$ is Archimedean,* i.e., *$M - \|\boldsymbol{x}\|_2^2 \in \langle \boldsymbol{h} \rangle_{2\beta} + Q_\beta(\boldsymbol{g})$ for some $\beta \in \mathbb{N}$ and $M > 0$. If $p(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}]$ is positive on $\mathcal{X}$, then $p(\boldsymbol{x}) \in \langle \boldsymbol{h} \rangle + Q(\boldsymbol{g})$.*

Based on Putinar's Positivstellensatz, Lasserre [21] derived a sequence of SOS relaxations that approximates the global minimum of problem (3) with increasing accuracy. The key insight behind Lasserre's hierarchy is twofold. The first insight is that problem (3), which we can write succinctly as $\min_{\boldsymbol{x} \in \mathcal{X}} f(x)$, can be equivalently written as $\max_{\boldsymbol{x}, \gamma} \gamma, s.t. f(x) - \gamma \geq 0$ on $\mathcal{X}$ (intuition: the latter pushes the lower bound γ to reach the global minimum of $f(\boldsymbol{x})$). The second intuition is that we can rewrite the condition $f(x) - \gamma \geq 0$ on $\mathcal{X}$, using Putinar's Positivstellensatz (Theorem 2), leading to the following hierarchy of Sums-of-Squares relaxations.

Theorem 3 (Lasserre's Hierarchy [21]) *Lasserre's hierarchy of order β is the following SOS program:*

$$\max \quad \gamma, \quad s.t. \;\; f(\boldsymbol{x}) - \gamma \in \langle \boldsymbol{h} \rangle_{2\beta} + Q_\beta(\boldsymbol{g}), \tag{8}$$

which can be written as a standard SDP. Moreover, let $f^\star$ be the global minimum of (3) and $f_\beta^\star$ be the optimal value of (8), then $f_\beta^\star$ monotonically increases and $f_\beta^\star \to f^\star$ when $\beta \to \infty$. More recently, Nie [30] proved that under Archimedeanness, Lasserre's hierarchy has finite convergence generically (i.e., *$f_\beta^\star = f^\star$ for some finite β).*

In computer vision, Lasserre's hierarchy was first used by Kahl and Henrion [15] to minimize rational functions arising in geometric reconstruction problems, and more recently by Probst *et al.* [33] as a framework to solve a set of 3D vision problems. In this paper we will show that the SOS relaxation as written in eq. (8) allows using *basis reduction* to exploit the *sparsity pattern* of polynomials and leads to significantly smaller semidefinite programs.

4. Problem Statement: Shape Reconstruction

Assume we are given N pixel measurements $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N] \in \mathbb{R}^{2 \times N}$ (the 2D *landmarks*), generated from the projection of points belonging to an unknown 3D shape $\boldsymbol{S} \in \mathbb{R}^{3 \times N}$ onto an image. Further assume the shape $\boldsymbol{S}$ that can be represented as a linear combination of K predefined basis shapes $\boldsymbol{B}_k \in \mathbb{R}^{3 \times N}$, *i.e.* $\boldsymbol{S} = \sum_{k=1}^K c_k \boldsymbol{B}_k$, where $\{c_k\}_{k=1}^K$ are (unknown) *shape coefficients*. Then, the generative model of the 2D landmarks reads:

$$\boldsymbol{z}_i = \Pi \boldsymbol{R} \left(\sum_{k=1}^K c_k \boldsymbol{B}_{ki} \right) + \boldsymbol{t} + \boldsymbol{\epsilon}_i, i = 1, \ldots, N, \tag{9}$$

where $\boldsymbol{B}_{ki}$ denotes the i-th 3D point on the k-th basis shape, $\boldsymbol{\epsilon}_i \in \mathbb{R}^2$ models the measurement noise, and Π is the (known) weak perspective projection matrix:

$$\Pi = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \end{bmatrix}, \tag{10}$$

with s_x and s_y being constants[2]. In eq. (9), $\boldsymbol{R} \in \mathrm{SO}(3)$ and $\boldsymbol{t} \in \mathbb{R}^2$ model the (unknown) rotation and translation of the shape $\boldsymbol{S}$ relative to the camera (only a 2D translation can be estimated). *The shape reconstruction problem consists in the joint estimation of the shape parameters $\{c_k\}_{k=1}^K$ and the camera pose $(\boldsymbol{R}, \boldsymbol{t})$*[3].

Without loss of generality, we adopt the *nonnegative sparse coding* (NNSC) convention [49] and assume all the coefficients c_k are nonnegative[4]. Due to the existence of noise, we solve the following *weighted least squares optimization* with *Lasso (ℓ_1-norm) regularization*:

$$\min_{\substack{c_k \geq 0, k=1,\ldots,K \\ \boldsymbol{t} \in \mathbb{R}^2, \boldsymbol{R} \in \mathrm{SO}(3)}} \sum_{i=1}^N w_i \left\| \boldsymbol{z}_i - \Pi \boldsymbol{R} \left(\sum_{k=1}^K c_k \boldsymbol{B}_{ki} \right) - \boldsymbol{t} \right\|^2 + \alpha \sum_{k=1}^K |c_k| \tag{11}$$

The ℓ_1-norm regularization (controlled by a given constant α) encourages the coefficients c_k to be sparse when the shape $\boldsymbol{S}$ is generated from only a subset of the basis shapes [49] (note that the ℓ_1-norm becomes redundant when using the NNSC convention). Contrary to previous approaches [49, 35], we explicitly associate a given weight $w_i \geq 0$ to each 2D measurement $\boldsymbol{z}_i$ in eq. (11). On the one hand, this allows accommodating heterogeneous noise in the 2D landmarks (*e.g.*, $w_i = 1/\sigma_i^2$ when the noise $\boldsymbol{\epsilon}_i$ is Gaussian, $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_i^2 \mathbf{I}_2)$). On the other hand, as shown in Section 6, the weighted least squares framework is useful to robustify (11) against outliers.

5. Certifiably Optimal Shape Reconstruction

This section shows how to develop a *certifiably optimal* solver for problem (11). Our first step is to algebraically eliminate the translation $\boldsymbol{t}$ and obtain a translation-free shape reconstruction problem, as shown below.

[1] If $p \in \langle \boldsymbol{h} \rangle + Q(\boldsymbol{g})$, then $p = h + g$, with $h \in \langle \boldsymbol{h} \rangle$ and $g \in Q(\boldsymbol{g})$. For any $\boldsymbol{x} \in \mathcal{X}$, since $h_i(\boldsymbol{x}) = 0$, so $h(\boldsymbol{x}) = \sum \lambda_i h_i = 0$; since $g_k(\boldsymbol{x}) \geq 0$ and $s_k(\boldsymbol{x}) \geq 0$, so $g = \sum s_k g_k \geq 0$. Therefore, $p = h + g \geq 0$

[2] The weak perspective camera model is a good approximation of the full perspective camera model when *the distance from the object to the camera* is much larger than *the depth of the object itself [48]*. [50] showed that the solution obtained using the weak perspective model provides a good initialization when refining the pose for the full perspective model.

[3] Shape reconstruction in the case of a single 3D model, *i.e.*, $K = 1$, is called shape alignment and has been solved recently in [44].

[4] The general case of real coefficients is equivalent to the NNSC case where for each basis $\boldsymbol{B}_k$ we also add the basis $-\boldsymbol{B}_k$.

Theorem 4 (Translation-free Shape Reconstruction) *The shape reconstruction problem (11) is equivalent to the following translation-free optimization:*

$$\min_{\substack{c_k\geq 0,k=1,\dots,K\\ \boldsymbol{R}\in\mathrm{SO}(3)}} \sum_{i=1}^{N}\left\|\tilde{\boldsymbol{z}}_i-\Pi\boldsymbol{R}\left(\sum_{k=1}^{K}c_k\tilde{\boldsymbol{B}}_{ki}\right)\right\|^2+\alpha\sum_{k=1}^{K}|c_k| \quad (12)$$

where $\tilde{\boldsymbol{z}}_i$ and $\tilde{\boldsymbol{B}}_{ki}$ can be computed as follows:

$$\tilde{\boldsymbol{z}}_i=\sqrt{w_i}(\boldsymbol{z}_i-\bar{\boldsymbol{z}}^w),\ \textit{with}\ \bar{\boldsymbol{z}}^w=\frac{\sum_{i=1}^{N}w_i\boldsymbol{z}_i}{\sum_{i=1}^{N}w_i}, \quad (13)$$

$$\tilde{\boldsymbol{B}}_{ki}=\sqrt{w_i}(\boldsymbol{B}_{ki}-\bar{\boldsymbol{B}}_k^w),\ \textit{with}\ \bar{\boldsymbol{B}}_k^w=\frac{\sum_{i=1}^{N}w_i\boldsymbol{B}_{ki}}{\sum_{i=1}^{N}w_i}. \quad (14)$$

Further, let $\boldsymbol{R}^\star$ and $c_k^\star, k=1,\dots,K$, be the global minimizer of the above translation-free optimization (12), then the optimal translation $\boldsymbol{t}^\star$ can be recovered as:

$$\boldsymbol{t}^\star=\bar{\boldsymbol{z}}^w-\Pi\boldsymbol{R}^\star\left(\sum_{k=1}^{K}c_k^\star\bar{\boldsymbol{B}}_k^w\right). \quad (15)$$

A formal proof of Theorem 4 can be found in the Supplementary Material. The intuition behind Theorem 4 is that if we express the landmark coordinates and 3D basis shapes with respect to their (weighted) centroids $\bar{\boldsymbol{z}}^w$ and $\bar{\boldsymbol{B}}_k^w, k=1,\dots,K$, we can remove the dependence on the translation $\boldsymbol{t}$. This strategy is inspired by Horn's method for point cloud registration [14], and generalizes [49] to the weighted and non-centered case.

5.1. SOS Relaxation

This section applies Lasserre's hierarchy as described in Theorem 3 to solve the translation-free shape reconstruction problem (12). We do this in two steps: we first show problem (12) can be formulated as a polynomial optimization in the form (3); and then we add valid constraints to make the feasible set Archimedean.

Polynomial Optimization Formulation. Denote $\boldsymbol{c}=[c_1,\dots,c_k]^\mathsf{T}\in\mathbb{R}^K$, $\boldsymbol{r}=\mathrm{vec}(\boldsymbol{R})=[\boldsymbol{r}_1^\mathsf{T},\boldsymbol{r}_2^\mathsf{T},\boldsymbol{r}_3^\mathsf{T}]^\mathsf{T}\in\mathbb{R}^9$, with $\boldsymbol{r}_i, i=1,2,3$ being the i-th column of $\boldsymbol{R}$, then $\boldsymbol{x}:=[\boldsymbol{c}^\mathsf{T},\boldsymbol{r}^\mathsf{T}]^\mathsf{T}\in\mathbb{R}^{K+9}$ is the unknown decision vector in (3). Consider the first term in the objective function of (12). We can write:

$$q_i(\boldsymbol{x}):=\left\|\tilde{\boldsymbol{z}}_i-\Pi\boldsymbol{R}\left(\textstyle\sum_{k=1}^{K}c_k\tilde{\boldsymbol{B}}_{ki}\right)\right\|^2=\left\|\tilde{\boldsymbol{z}}_i-\Pi\textstyle\sum_{k=1}^{K}c_k\boldsymbol{R}\tilde{\boldsymbol{B}}_{ki}\right\|^2, \quad (16)$$

then it becomes clear that $q_i(\boldsymbol{x})$ is a polynomial function of $\boldsymbol{x}$ with degree 4. Because the Lasso regularization is linear in $\boldsymbol{c}$, the objective function $f(\boldsymbol{x})$ is a degree-4 polynomial.

Now we consider the feasible set of (12). The inequality constraints $c_k\geq 0$ are already in generic form (3) with $g_k(\boldsymbol{x})=c_k, k=1,\dots,K$, being degree-1 polynomials. As for the $\boldsymbol{R}\in\mathrm{SO}(3)$ constraint, it has already been shown in related work [38, 6] that enforcing $\boldsymbol{R}\in\mathrm{SO}(3)$ is equivalent to imposing 15 quadratic equality constraints.

Lemma 5 (Quadratic Constraints for $\mathrm{SO}(3)$ [38, 6]) *For a matrix $\boldsymbol{R}\in\mathbb{R}^{3\times 3}$, the constraint $\boldsymbol{R}\in\mathrm{SO}(3)$ (where $\mathrm{SO}(3):=\{\boldsymbol{R}:\boldsymbol{R}^\mathsf{T}\boldsymbol{R}=\mathbf{I}_3,\det\boldsymbol{R}=+1\}$ is the set of proper rotation matrices) is equivalent to the following set of degree-2 polynomial equality constraints ($h_i(\boldsymbol{x})=0, i=1,\dots,15$):*

$$\begin{cases} h_1=1-\|\boldsymbol{r}_1\|^2, h_2=1-\|\boldsymbol{r}_2\|^2, h_3=1-\|\boldsymbol{r}_3\|^2\\ h_4=\boldsymbol{r}_1^\mathsf{T}\boldsymbol{r}_2,\quad h_5=\boldsymbol{r}_2^\mathsf{T}\boldsymbol{r}_3,\quad h_6=\boldsymbol{r}_3^\mathsf{T}\boldsymbol{r}_1\\ h_{7,8,9}=\boldsymbol{r}_1\times\boldsymbol{r}_2-\boldsymbol{r}_3\\ h_{10,11,12}=\boldsymbol{r}_2\times\boldsymbol{r}_3-\boldsymbol{r}_1\\ h_{13,14,15}=\boldsymbol{r}_3\times\boldsymbol{r}_1-\boldsymbol{r}_2 \end{cases} \quad (17)$$

where $\boldsymbol{r}_i\in\mathbb{R}^3, i=1,2,3$, denotes the i-th column of $\boldsymbol{R}$ and "$\times$" represents the vector cross product.

In eq. (17), $h_{1,2,3}$ constrain the columns to be unit vectors, $h_{4,5,6}$ constrain the columns to be mutually orthogonal, and h_{7-15} constrain the columns to satisfy the right-hand rule (*i.e.*, the determinant constraint)[5].

In summary, the translation-free problem (12) is equivalent to a polynomial optimization with a degree-4 objective $f(\boldsymbol{x})$, constrained by 15 quadratic equalities $h_i(\boldsymbol{x})$ (eq. (17)) and K linear inequalities $g_k(\boldsymbol{x})=c_k$.

Archimedean Feasible Set. The issue with the feasible set defined by inequalities $c_k\geq 0$ and equalities (17) is that $\langle\boldsymbol{h}\rangle+Q(\boldsymbol{g})$ is *not* Archimedean, which can be easily seen from the *unboundedness* of the linear inequality $c_k\geq 0$[6]. However, we know the linear coefficients must be bounded because the pixel measurement values $\boldsymbol{Z}$ lie in a bounded set (the 2D image). Therefore, we propose to *normalize* the 2D measurements and the 3D basis shapes: (i) for 2D measurements $\boldsymbol{Z}$, we first divide them by s_x and s_y (eq. (10)), and then scale them such that they lie inside a unit circle; (ii) for each 3D basis shape $\boldsymbol{B}_k$, we scale $\boldsymbol{B}_k$ such that it lies inside a unit sphere. With this proper normalization, we can add the following degree-2 inequality constraints ($c_k^2\leq 1$) that bound the linear coefficients:

$$g_{K+k}(\boldsymbol{x})=1-c_k^2, k=1,\dots,K. \quad (18)$$

Now we can certify the Archimedeanness of $\langle\boldsymbol{h}\rangle+Q(\boldsymbol{g})$:

$$K+3-\|\boldsymbol{x}\|_2^2=\underbrace{\sum_{k=1}^{K}1\cdot g_{K+k}}_{\in Q_1(\boldsymbol{g})}+\underbrace{h_1+h_2+h_3}_{\in\langle\boldsymbol{h}\rangle_2}, \quad (19)$$

with $M=K+3$ and $\beta=1$ (*cf.* Theorem 2).

Apply Lasserre's Hierarchy. With Archimedeanness, we can now apply Lasserre's hierarchy of SOS relaxations.

[5] We remark that the 15 equality constraints in (17) are redundant. For example, $h_{1,2,3,7,8,9}$ are sufficient to fully constrain $\boldsymbol{R}\in\mathrm{SO}(3)$. We also found that, empirically, choosing $h_{1,2,3}$ and h_{7-15} yields similar tightness results as choosing all 15 constraints.

[6] $M-\|\boldsymbol{x}\|_2^2\geq 0$ requires $\boldsymbol{x}$ to have bounded ℓ_2-norm.

Proposition 6 (SOS Relaxations for Shape Reconstruction) *The SOS relaxation of order β ($\beta \geq 2$)[7] for the translation-free shape reconstruction problem (12) is the following convex semidefinite program:*

$$\max_{\substack{\gamma\in\mathbb{R},\boldsymbol{S}_0\in\mathcal{S}_+^{N_0}\\ \boldsymbol{S}_k\in\mathcal{S}_+^{N_s},\ k=1,\ldots,2K\\ \boldsymbol{\lambda}_i\in\mathbb{R}^{N_\lambda},i=1,\ldots,15}} \quad \gamma \tag{20}$$

$$\begin{aligned} s.t. \quad & f(\boldsymbol{x}) - \gamma = [\boldsymbol{x}]_\beta^\mathsf{T}\boldsymbol{S}_0[\boldsymbol{x}]_\beta + \\ & \textstyle\sum_{k=1}^{2K}\left([\boldsymbol{x}]_{\beta-1}^\mathsf{T}\boldsymbol{S}_k[\boldsymbol{x}]_{\beta-1}\right)g_k(\boldsymbol{x}) + \\ & \textstyle\sum_{i=1}^{15}\left(\boldsymbol{\lambda}_i^\mathsf{T}[\boldsymbol{x}]_{2\beta-2}\right)h_i(\boldsymbol{x}), \end{aligned} \tag{21}$$

where $f(\boldsymbol{x})$ is the objective function defined in (12), $g_k(\boldsymbol{x}), k = 1,\ldots,2K$ are the inequality constraints $c_k, 1-c_k^2$, $h_i(\boldsymbol{x}), i = 1,\ldots,15$ are the equality constraints defined in (17), and $N_0 := \binom{K+9+\beta}{\beta}$, $N_s := \binom{K+8+\beta}{\beta-1}$, $N_\lambda := \binom{K+7+2\beta}{2\beta-2}$ are the sizes of matrices and vectors.

While a formal proof of Proposition 6 is given in the Supplementary Material, we observe that (20) immediately results from the application of Lasserre's hierarchy to (8), by parametrizing $Q_\beta(\boldsymbol{g})$ with monomial bases $[\boldsymbol{x}]_{\beta-1}$, $[\boldsymbol{x}]_\beta$ and PSD matrices $\boldsymbol{S}_0, \boldsymbol{S}_k, k = 1,\ldots,2K$ (one for each g_k), and by parametrizing $\langle\boldsymbol{h}\rangle_{2\beta}$ with monomial basis $[\boldsymbol{x}]_{2\beta-2}$ and coefficient vectors $\boldsymbol{\lambda}_i, i = 1,\ldots,15$ (one for each h_i). Problem (20) can be written as an SDP and solved globally using standard convex solvers (*e.g.* YALMIP [26]). We call the SDP written in (20) the *primal* SDP. The *dual* SDP of (20) can be derived using *moment relaxation* [21, 23, 22], which is readily available in GloptiPoly 3 [13].

Extract Solutions from SDP. After solving the SDP (20), we can extract solutions to the original non-convex problem (12), a procedure we call *rounding*.

Proposition 7 (Rounding and Duality Gap) *Let $f_\beta^\star = \gamma^\star$ and $\boldsymbol{S}_0^{\beta\star}, \boldsymbol{S}_k^{\beta\star}, \boldsymbol{\lambda}_i^{\beta\star}$ be the optimal solutions to the SDP (20) at order β; compute $\boldsymbol{v}^{\beta\star}$ as the eigenvector corresponding to the minimum eigenvalue of $\boldsymbol{S}_0^{\beta\star}$, and then normalize $\boldsymbol{v}^{\beta\star}$ such that the first entry is equal to 1. Then an approximate solution to problem (12) can be obtained as:*

$$\hat{\boldsymbol{c}}^\beta = proj_{\boldsymbol{g}}([\boldsymbol{v}^{\beta\star}]_{\boldsymbol{c}}); \quad \hat{\boldsymbol{r}}^\beta = proj_{\boldsymbol{h}}([\boldsymbol{v}^{\beta\star}]_{\boldsymbol{r}}), \tag{22}$$

where $[\boldsymbol{v}^{\beta\star}]_{\boldsymbol{c}}$ (resp. $[\boldsymbol{v}^{\beta\star}]_{\boldsymbol{r}}$) denotes the entries of $\boldsymbol{v}^{\beta\star}$ corresponding to monomials $\boldsymbol{c}$ (resp. $\boldsymbol{r}$), and $proj_{\boldsymbol{g}}$ (resp. $proj_{\boldsymbol{h}}$) denotes projection to the feasible set defined by $\boldsymbol{g}$ (resp. $\boldsymbol{h}$). Specifically for problem (12), $proj_{\boldsymbol{g}}$ is rounding each coefficient c_k to the $[0,1]$ interval, and $proj_{\boldsymbol{h}}$ is the projection to $\mathrm{SO}(3)$. Moreover, let $\hat{f}_\beta$ be the value of the objective function evaluated at the approximate solution $\hat{\boldsymbol{x}}^\beta := [(\hat{\boldsymbol{c}}^\beta)^\mathsf{T}, (\hat{\boldsymbol{r}}^\beta)^\mathsf{T}]^\mathsf{T}$, then the following inequality holds (weak duality):

$$f_\beta^\star \leq f^\star \leq \hat{f}_\beta, \tag{23}$$

where $f^\star$ is the true (unknown) global minimum of problem (12). We define the relative duality gap η_β as:

$$\eta_\beta = (\hat{f}_\beta - f_\beta^\star)/\hat{f}_\beta, \tag{24}$$

which quantifies the quality of the SOS relaxation.

Certifiable Global Optimality. Besides extracting solutions to the original problem, we can also verify when the SOS relaxation solves the original problem *exactly*.

Theorem 8 (Certificate of Global Optimality) *Let $f_\beta^\star = \gamma^\star$ and $\boldsymbol{S}_0^{\beta\star}$ be the optimal solutions to the SDP (20) at order β. If corank$(\boldsymbol{S}_0^{\beta\star}) = 1$ (the corank is the dimension of the null space of $\boldsymbol{S}_0^{\beta\star}$), then $f_\beta^\star$ is the global minimum of problem (12), and the relaxation is said to be* tight *at order β. Moreover, the relative duality gap $\eta_\beta = 0$ and the solution $\hat{\boldsymbol{x}}^\beta$ extracted using Proposition 7 is the unique global minimizer of problem (12).*

The proof of Theorem 8 is given in the Supplementary Material. Empirically (Section 7), we observed that the relaxation is always tight at the minimum relaxation order $\beta = 2$. Note that even when the relaxation is not tight, one can still obtain an approximate solution using Proposition 7 and quantify how suboptimal the approximate solution is using the relative duality gap η_β.

5.2. Basis Reduction

Despite the theoretical soundness and finite convergence at order $\beta = 2$, the size of the SDP (20) is $N_0 = \binom{K+9+\beta}{\beta}$, which for $\beta = 2$ becomes $\binom{K+11}{2}$, implying that the size of the SDP grows *quadratically* in the number of bases K. Although there have been promising advances in improving the scalability of SDP solvers (see [28] for a thorough review), such as exploiting sparsity [41, 39, 29] and low-rankness [7, 36], in this section we demonstrate a simple yet effective approach, called *basis reduction*, that exploits the structure of the objective function to significantly reduce the size of the SDP in (20).

In a nutshell, basis reduction methods seek to find a smaller, but still expressive enough, subset of the full vector of monomials $[\boldsymbol{x}]_\beta$ on the right-hand side (RHS) of eq. (21), to explain the objective function $f(\boldsymbol{x})$ on the left-hand side (LHS). There exist standard approximation algorithms for basis reduction, discussed in [31, 32] and implemented in YALMIP [27]. However, in practice we found the basis selection method in YALMIP failed to find any reduction

[7] The minimum relaxation order is 2 because $f(\boldsymbol{x})$ has degree 4.

for the SDP (20). Therefore, here we propose a problem-specific reduction, which follows from the examination of which monomials appear on the LHS of (21).

Proposition 9 (SOS Relaxation with Basis Reduction) *The SOS relaxation of order $\beta = 2$ with basis reduction for the translation-free shape reconstruction problem (12) is the following convex semidefinite program:*

$$\begin{aligned} \max_{\substack{\gamma\in\mathbb{R},\boldsymbol{S}_0\in\mathcal{S}_+^{N_0'} \\ \boldsymbol{S}_k\in\mathcal{S}_+^{N_s'},k=1,\dots,2K, \\ \boldsymbol{\lambda}_i\in\mathbb{R}^{N_\lambda'},i=1,\dots,15}} \quad & \gamma \qquad (25) \\ s.t. \quad & f(\boldsymbol{x})-\gamma = m_2(\boldsymbol{x})^\mathsf{T}\boldsymbol{S}_0 m_2(\boldsymbol{x}) + \\ & \textstyle\sum_{k=1}^{2K}([\boldsymbol{r}]_1^\mathsf{T}\boldsymbol{S}_k[\boldsymbol{r}])g_k(\boldsymbol{x}) + \\ & \textstyle\sum_{i=1}^{15}(\boldsymbol{\lambda}_i^\mathsf{T}[\boldsymbol{c}]_2)h_i(\boldsymbol{x}), \qquad (26) \end{aligned}$$

where $N_0' = 10K+10$, $N_s' = 10$, $N_\lambda' = \binom{K+2}{2}$, and $m_2(\boldsymbol{x}) = [1, \boldsymbol{c}^\mathsf{T}, \boldsymbol{r}^\mathsf{T}, \boldsymbol{c}^\mathsf{T}\otimes\boldsymbol{r}^\mathsf{T}]^\mathsf{T} \in \mathbb{R}^{N_0}$, and where $\otimes$ is the Kronecker product.

Comparing the SDP (25) and (20), the most significant change is replacing the full monomial basis $[\boldsymbol{x}]_\beta$ in (20) with a much smaller monomial basis $m_2(\boldsymbol{x})$ that excludes degree-2 monomials purely supported in $\boldsymbol{c}$ and $\boldsymbol{r}$. This reduction is motivated by analyzing the monomial terms in $f(\boldsymbol{x})$. Although a formal proof of the equivalence between (20) and (25) remains open, we provide an intuitive explanation in the Supplementary Material. After basis reduction, the size of the SDP (25) is $N_0' = 10K+10$, which is *linear* in K and much smaller than the size of the original SDP (20) $N_k = \binom{K+11}{2}$[8]. Section 7 numerically shows that the SDP after basis reduction gives the same (tight) solution as the original SDP.

5.3. Shape*: Algorithm Summary

To summarize the derivation in this section, our solver for the shape reconstruction problem (11), named Shape*, works as follows. It first solves the SDP (25) and applies the rounding described in Proposition 7 to compute an estimate of the shape parameters c_k and rotation $\boldsymbol{R}$ and possibly certify its optimality. Then, Shape* uses the closed-form expression (15) to retrieve the translation estimate $\boldsymbol{t}$.

6. Robust Outlier Rejection

Section 5 proposed a certifiably optimal solver for problem (11). However, the least squares formulation (11) tends to be sensitive to outliers: the pixel measurements $\boldsymbol{Z}$ in eq. (9) are typically produced by learning-based or handcrafted detectors [49], which might produce largely incorrect measurements (*e.g.* due to wrong data association $\boldsymbol{z}_i \leftrightarrow \boldsymbol{B}_{ki}$), which in turn leads to poor shape reconstruction results. This section shows how to regain robustness by iteratively solving the weighted least squares problem (11) and adjusting the weights w_i to reject outliers.

The key insight is to substitute the least square penalty in (11) with a robust cost function, namely the *truncated least squares (TLS)* cost [46, 45, 19, 47]. Hence, we propose the following *TLS shape reconstruction* formulation:

$$\min_{\substack{c_k\geq 0, \\ k=1,\dots,K \\ \boldsymbol{t}\in\mathbb{R}^2,\boldsymbol{R}\in\mathrm{SO}(3)}} \sum_{i=1}^{N}\rho_{\bar{c}}\left(r_i(c_k,\boldsymbol{R},\boldsymbol{t})\right) + \alpha\sum_{k=1}^{K}c_k \qquad (27)$$

where $r_i(c_k,\boldsymbol{R},\boldsymbol{t}) := \left\|\boldsymbol{z}_i - \Pi\boldsymbol{R}\left(\sum_{k=1}^{K}c_k\boldsymbol{B}_{ki}\right) - \boldsymbol{t}\right\|$ (introduced for notational convenience), and $\rho_{\bar{c}}(r) = \min(r^2, \bar{c}^2)$ implements a truncated least squares cost, which is quadratic for small residuals and saturates to a constant value for residuals larger than a maximum error $\bar{c}$.

Our second insight is that $\rho_{\bar{c}}(r)$ can be written as $\rho_{\bar{c}}(r) = \min_{w\in\{0,1\}} wr^2 + (1-w)\bar{c}^2$, by introducing extra slack binary variables $w \in \{0,1\}$. Therefore, we can write problem (27) equivalently as:

$$\min_{\substack{c_k\geq 0,k=1,\dots,K, \\ w_i\in\{0,1\},i=1,\dots,N \\ \boldsymbol{t}\in\mathbb{R}^2,\boldsymbol{R}\in\mathrm{SO}(3)}} \sum_{i=1}^{N} w_i\left(r_i(c_k,\boldsymbol{R},\boldsymbol{t})\right) + (1-w_i)\bar{c}^2 + \alpha\sum_{k=1}^{K}c_k \qquad (28)$$

The final insight is that now we can minimize (28) by iteratively minimizing (i) over $c_k, \boldsymbol{R}, \boldsymbol{t}$ (with fixed weights w_i), and (ii) over the weights w_i (with fixed $c_k, \boldsymbol{R}, \boldsymbol{t}$). The rationale for this approach is that step (i) can be implemented using Shape* (since in this case the weights are fixed), and step (ii) can be implemented in closed-form. To improve convergence of this iterative algorithm, we adopt graduated non-convexity [3, 44], which starts with a convex approximation of problem (28) and uses a control parameter μ to gradually increase the amount of non-convexity, till (for large μ) one solves (28). The resulting algorithm named Shape# is given in Algorithm 1. We refer the reader to the Supplementary Material and [44] for a complete derivation of Algorithm 1 and for the closed-form expression of the weight update in line 6 of the algorithm.

Shape# is deterministic and does not require an initial guess. We remark that the graduated non-convexity scheme in Shape# (contrarily to Shape*) is not guaranteed to converge to an optimal solution of (28), but we show in the next section that it is empirically robust to 70% outliers.

7. Experiments

Implementation details. Both Shape* and Shape# are implemented in Matlab, with both SOS relaxations (20) and (25) implemented using YALMIP [26] and the resulting SDPs solved using MOSEK [1].

[8] For $K = 5, 10, 20$, $N_0 = 120, 210, 465$, while $N_0' = 60, 110, 210$.

Algorithm 1: Robust Shape Reconstruction.

input : measurements $\boldsymbol{z}_i, i = 1, \dots, N$,
basis shapes $\boldsymbol{B}_k, k = 1, \dots, M$
maximum error $\bar{c}$, regularization constant α
output: shape reconstruction: $c_k^\star, \boldsymbol{R}^\star, \boldsymbol{t}^\star$

/* Initialization */
$w_i^{(0)} = 1, \quad i = 1, \dots, N$
$\mu^{(0)} = 10^{-4}$
/* Iterations */
for $\tau = 1 : \mathit{maxIter}$ **do**
  /* Variable update */
  $c_k^{(\tau)}, \boldsymbol{R}^{(\tau)}, \boldsymbol{t}^{(\tau)} = \mathsf{Shape}^\star(\boldsymbol{z}_i, \boldsymbol{B}_k, \alpha, w_i^{(\tau-1)}, \mu^{(\tau-1)})$
  /* Compute residual errors */
  $r_i^{(\tau)} = \|\boldsymbol{z}_i - \Pi \boldsymbol{R}^{(\tau)} \left(\sum_{k=1}^K c_k^{(\tau)} \boldsymbol{B}_{ki}\right) - \boldsymbol{t}^{(\tau)}\|$
  /* Weight update */
  $w_i^{(\tau)} = \text{weightUpdate}(r_i^{(\tau)}, \bar{c}, \mu^{(\tau-1)})$
  /* Compute objective function */
  $f^{(\tau)} = \text{computeObjective}(r_i^{(\tau)}, w_i^{(\tau)}, \mu^{(\tau-1)}, \alpha, \bar{c})$
  /* Check convergence ($\tau > 1$) */
  if $|f^{(\tau)} - f^{(\tau-1)}| < 10^{-10}$ **then**
    break
  /* Update control parameter μ */
  $\mu^{(\tau)} = 2 \cdot \mu^{(\tau-1)}$
return $c_k^{(\tau)}, \boldsymbol{R}^{(\tau)}, \boldsymbol{t}^{(\tau)}$.

7.1. Efficiency Improvement by Basis Reduction

We first evaluate the efficiency improvement due to basis reduction in simulation. We fix the number of correspondences $N = 100$, and increase the number of basis shapes $K = 5, 10, 20$. At each K, we first randomly generate K basis shapes $\boldsymbol{B}_1, \dots, \boldsymbol{B}_K \in \mathbb{R}^{3\times N}$, with entries sampled independently from a Normal distribution $\mathcal{N}(0, 1)$. Then K linear coefficients $\boldsymbol{c} = [c_1, \dots, c_K]^\mathsf{T}$ are uniformly sampled from the interval $[0, 1]$, and a rotation matrix $\boldsymbol{R}$ is randomly chosen. The 2D measurements $\boldsymbol{Z}$ are computed from the generative model (9) with $\boldsymbol{t} = \boldsymbol{0}$, $s_x = s_y = 1$ for Π, and additive noise $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\boldsymbol{0}, 0.01^2)$. For shape reconstruction, we feed the noisy $\boldsymbol{Z}$ and bases $\boldsymbol{B}_k$ to (i) the SOS relaxation (20) without basis reduction, and (ii) the SOS relaxation (25) with basis reduction, both at relaxation order $\beta = 2$ and with no Lasso regularization ($\alpha = 0$).

To evaluate the effects of introducing basis reduction, we compute the following statistics for each choice of K: (i) solution time for the SDP; (ii) tightness of the SOS relaxation, including corank$(\boldsymbol{S}_0^{2\star})$ and relative duality gap η_2; (iii) accuracy of reconstruction, including the coefficients estimation error (ℓ_2 norm of the difference between estimated and ground-truth coefficients) and the rotation estimation error (the geodesic distance between estimated and ground-truth rotation). Table 1 shows the resulting statistics. We see that the SOS relaxation without basis reduction quickly becomes intractable at $K = 20$ (mean solution time is 2440 seconds), while the relaxation with basis reduction can still be solved in a reasonable amount of time (107 seconds)[9]. In addition, from the co-rank of $(\boldsymbol{S}_0^{2\star})$ and the relative duality gap η_2, we see basis reduction has no negative impact on the quality of the relaxation, which remains tight at order $\beta = 2$. This observation is further validated by the identical accuracy of $\boldsymbol{c}$ and $\boldsymbol{R}$ estimation before and after basis reduction (last two rows of Table 1).

# of Bases K	$K = 5$	$K = 10$	$K = 20$
SDP Time [s]	3.52	47.0	2440
	0.550	**5.28**	**107**
corank$(\boldsymbol{S}_0^{2\star})$	1	1	1
	1	**1**	**1**
Duality Gap η_2	5e−6	7e−6	4e−5
	1e−5	**2e−5**	**1e−5**
$\boldsymbol{c}$ Error	1.3e−3	2.3e−3	3.2e−3
	1.3e−3	**2.3e−3**	**3.2e−3**
$\boldsymbol{R}$ Error [deg]	0.0690	0.0487	0.0298
	0.0690	**0.0487**	**0.0298**

Table 1. Efficiency improvement by basis reduction. **Bold** text represent mean values computed by solving the SOS relaxation ***with*** basis reduction (25), while normal text represent mean values computed by solving the SOS relaxation *without* basis reduction (20). Statistics are computed over 20 Monte Carlo runs.

7.2. Shape* for Outlier-Free Reconstruction

In this section, we compare the performance of Shape* against state-of-art optimization techniques for shape reconstruction. We follow the same protocol as in Section 7.1, but only generate 2D measurements from a *sparse* set of $p = 2$ basis shapes. This is done by only sampling p out of K *nonzero* shape coefficients, *i.e.*, $c_{p+1}, \dots, c_K = 0$. We then compare the performance of Shape*, setting $\alpha = 0.01$ to encourage sparseness, against three state-of-the-art optimization techniques: (i) the projective matching pursuit method [35] (label: PMP), which uses principal component analysis to first obtain a set of orthogonal bases from $\{\boldsymbol{B}_k\}$ and then locally optimizes the shape parameters and camera pose using the mean shape as an initial guess; (ii) the alternative optimization method [48] (label: Altern), which locally optimizes problem (11) by alternatively updating $\boldsymbol{c}$ and $\boldsymbol{R}$, initialized at the mean shape; and (iii) the convex relaxation with refinement proposed in [49] (label: convex+refine), which uses a convex relaxation and then refines the solution to obtain $\boldsymbol{c}$ and $\boldsymbol{R}$. Fig. 1 shows the boxplots of the 3D shape estimation error (mean ℓ_2 distance between the reconstructed shape and the ground-truth shape) and the rotation estimation error for $K = 5, 10, 20$ basis shapes and 20 Monte Carlo runs. We observe that Shape* has the highest accuracy in estimating the 3D shape and camera pose, though the other three methods also perform quite well. In all the Monte Carlo runs, Shape* achieves corank$(\boldsymbol{S}_0^{2\star}) = 1$ and mean relative duality gap $\eta_2 = 6.3\text{e}{-5}$, indicating that Shape* was able to obtain an optimal solution.

[9] Our basis reduction can potentially be *combined* with other scalability improvement techniques reviewed in [28], such as low-rank SDP solvers.

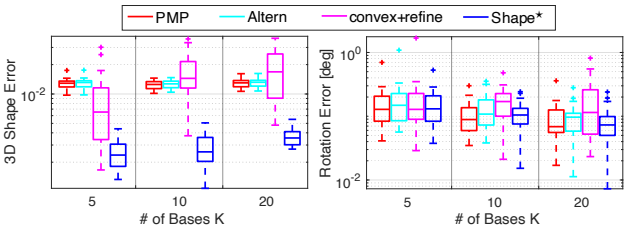


Figure 1. 3D shape estimation error (left) and rotation estimation error (right) by Shape* compared with PMP [35], Altern [48] and convex+refine [49], for increasing basis shapes $K = 5, 10, 20$.

7.3. Shape# for Robust Reconstruction

This section shows that Shape# achieves state-of-the-art performance on the FG3DCar [25] dataset. The FG3DCar dataset contains 300 car images with ground-truth 2D landmarks $\boldsymbol{Z} \in \mathbb{R}^{2\times N}, N = 256$. It also contains $K = 15$ 3D mesh models of different cars $\{\boldsymbol{B}_k\}_{k=1}^{K}$. To generate outliers, we randomly change $10\% - 70\%$ of the N ground-truth 2D landmarks $\boldsymbol{Z}$ to be arbitrary positions inside the image. We then evaluate the robustness of Shape# compared with two other robust methods based on the assumption of sparse outliers in [49]: (i) robustified alternative optimization (label: Altern+Robust) and (ii) robustified convex optimization (label: Convex+Robust). Fig. 2 boxplots the shape estimation and rotation estimation error[10] under increasing outlier rates computed over 40 randomly chosen images in the FG3DCar dataset. We can see that Shape# is insensitive to 70% outliers, while the accuracy of both Altern+Robust and Convex+Robust decreases with respect to higher outlier rates and they fail at 60% outliers. Fig. 3 shows two examples of qualitative results, where we see Shape# gives high-quality model fitting at 70% outliers, while the quality of Altern+Robust and Convex+Robust starts decreasing at 40% outliers. More qualitative results are given in the Supplementary Material.

8. Conclusions

We presented Shape*, the first certifiably optimal solver for 3D shape reconstruction from 2D landmarks in a single image. Shape* is developed by applying Lasserre's hierarchy of SOS relaxations combined with basis reduction to improve efficiency. Experimental results show that the SOS relaxation of order 2 always achieves global optimality. To handle outlying measurements, we also proposed Shape#, which solves a truncated least squares robust estimation problem by iteratively running Shape* without the need for an initial guess. We show that Shape# achieves robustness against 70% outliers on the FG3DCar dataset and outperforms state-of-the-art solvers.

[10] Although there is no ground-truth reconstruction for each image, the original paper [25] uses local optimization (with full perspective camera model) to reconstruct high-quality 3D shapes for all images, and we use their reconstructions as ground-truth.

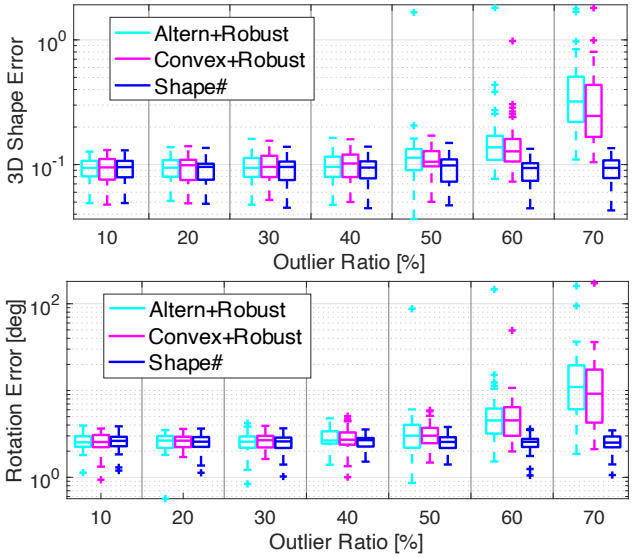
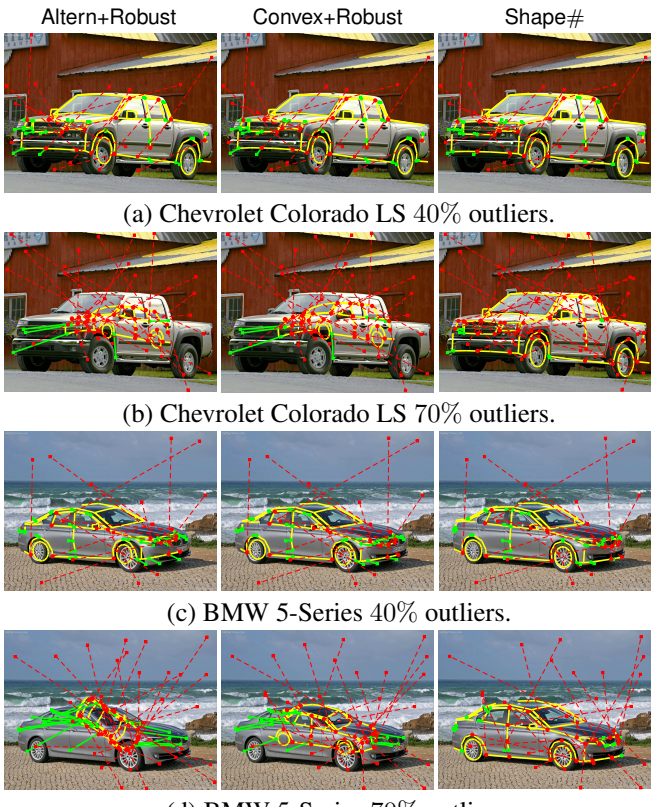


Figure 2. 3D shape estimation error (top) and rotation estimation error (bottom) by Shape# compared with Altern+Robust [49] and Convex+Robust [49] under increasing outlier rates.



(a) Chevrolet Colorado LS 40% outliers.

(b) Chevrolet Colorado LS 70% outliers.

(c) BMW 5-Series 40% outliers.

(d) BMW 5-Series 70% outliers.

Figure 3. Selected qualitative results on the FG3DCar dataset [25] under 40% and 70% outlier rates using Altern+Robust [49], Convex+Robust [49], and Shape#. (a)-(b): results on the Chevrolet Colorado LS; (c)-(d): results on the BMW 5-Series. Green: inliers. Red: outliers. Circle: 3D landmark. Square: 2D landmark. [Best viewed electronically.]

References

[1] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.1.*, 2017. 6

[2] Mathieu Aubry, Daniel Maturana, Alexei A Efros, Bryan C Russell, and Josef Sivic. Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 3762–3769, 2014. 1

[3] Andrew Blake and Andrew Zisserman. *Visual reconstruction*. MIT Press, 1987. 2, 6

[4] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(9):1063–1074, 2003. 1, 2

[5] Grigoriy Blekherman, Pablo A Parrilo, and Rekha R Thomas. *Semidefinite optimization and convex algebraic geometry*. SIAM, 2012. 2

[6] Jesus Briales and Javier Gonzalez-Jimenez. Convex Global 3D Registration with Lagrangian Duality. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017. 4

[7] Burer, Samuel and Monteiro, Renato D C. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003. 5

[8] Timothy F. Cootes, Christopher J. Taylor, David H. Cooper, and Jim Graham. Active shape models - their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, January 1995. 1

[9] Xiaochuan Fan, Kang Zheng, Youjie Zhou, and Song Wang. Pose locality constrained representation for 3D human pose reconstruction. In *European Conf. on Computer Vision (ECCV)*, pages 174–188. Springer, 2014. 2

[10] Lie Gu and Takeo Kanade. 3D alignment of face in a single image. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 1305–1312, 2006. 1, 2

[11] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *Intl. Conf. on Computer Vision (ICCV)*, pages 2980–2988, 2017. 1

[12] Mohsen Hejrati and Deva Ramanan. Analyzing 3D Objects in Cluttered Images. In *Advances in Neural Information Processing Systems (NIPS)*, pages 593–601, 2012. 1, 2

[13] Didier Henrion, Jean-Bernard Lasserre, and Johan Löfberg. GloptiPoly 3: moments, optimization and semidefinite programming. *Optim. Methods. Softw.*, 24(4-5):761–779, 2009. 5

[14] Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Amer.*, 4(4):629–642, Apr 1987. 4

[15] Fredrik Kahl and Didier Henrion. Globally optimal estimates for geometric reconstruction problems. *Intl. J. of Computer Vision*, 74(1):3–15, 2007. 3

[16] Laurent Kneip, Hongdong Li, and Yongduek Seo. UPnP: An optimal o(n) solution to the absolute pose problem with universal applicability. In *European Conf. on Computer Vision (ECCV)*, pages 127–142. Springer, 2014. 1

[17] Nikos Kolotouros, Georgios Pavlakos, Michael J Black, and Kostas Daniilidis. Learning to reconstruct 3d human pose and shape via model-fitting in the loop. In *Intl. Conf. on Computer Vision (ICCV)*, pages 2252–2261, 2019. 2

[18] Nikos Kolotouros, Georgios Pavlakos, and Kostas Daniilidis. Convolutional mesh regression for single-image human shape reconstruction. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2

[19] Pierre-Yves Lajoie, Siyi Hu, Giovanni Beltrame, and Luca Carlone. Modeling perceptual aliasing in SLAM via discrete-continuous graphical models. *IEEE Robotics and Automation Letters (RA-L)*, 2019. 6

[20] Katrin Lasinger, René Ranftl, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *arXiv preprint arXiv:1907.01341*, 2019. 1

[21] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2001. 3, 5

[22] Jean-Bernard Lasserre. *Moments, positive polynomials and their applications*, volume 1. World Scientific, 2010. 2, 5

[23] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009. 5

[24] Yan Li, Leon Gu, and Takeo Kanade. Robustly aligning a shape model and its application to car alignment of unknown pose. *IEEE transactions on pattern analysis and machine intelligence*, 33(9):1860–1876, 2011. 2

[25] Yen-Liang Lin, Vlad I. Morariu, Winston H. Hsu, and Larry S. Davis. Jointly optimizing 3D model fitting and fine-grained classification. In *European Conf. on Computer Vision (ECCV)*, 2014. 1, 2, 8

[26] Johan Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, volume 3. Taipei, Taiwan, 2004. 5, 6

[27] Johan Lofberg. Pre-and post-processing sum-of-squares programs in practice. *IEEE Transactions on Automatic Control*, 54(5):1007–1011, 2009. 5

[28] Anirudha Majumdar, Georgina Hall, and Amir Ali Ahmadi. A survey of recent scalability improvements for semidefinite programming with applications in machine learning, control, and robotics. *arXiv preprint arXiv:1908.05209*, 2019. 5, 7

[29] Joshua G Mangelson, Jinsun Liu, Ryan M Eustice, and Ram Vasudevan. Guaranteed globally optimal planar pose graph and landmark SLAM via sparse-bounded sums-of-squares programming. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 9306–9312. IEEE, 2019. 5

[30] Jiawang Nie. Optimality conditions and finite convergence of lasserre's hierarchy. *Mathematical programming*, 146(1-2):97–121, 2014. 2, 3

[31] Frank Permenter and Pablo Parrilo. Partial facial reduction: simplified, equivalent SDPs via approximations of the PSD cone. 171(1-2):1–54, 2018. 5

[32] Frank Permenter and Pablo A Parrilo. Basis selection for sos programs via facial reduction and polyhedral approximations. In *53rd IEEE Conference on Decision and Control*, pages 6615–6620. IEEE, 2014. 5

[33] Thomas Probst, Danda Pani Paudel, Ajad Chhatkuli, and Luc Van Gool. Convex relaxations for consensus and non-

minimal problems in 3D vision. In *Intl. Conf. on Computer Vision (ICCV)*, 2019. 3

[34] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993. 3

[35] Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Reconstructing 3D human pose from 2D image landmarks. In *European Conf. on Computer Vision (ECCV)*, 2012. 1, 2, 3, 7, 8

[36] 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. *Intl. J. of Robotics Research*, 38(2-3):95–125, 2019. 5

[37] Denis Tome, Chris Russell, and Lourdes Agapito. Lifting from the deep: Convolutional 3d pose estimation from a single image. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2500–2509, 2017. 1, 2

[38] Roberto Tron, David M Rosen, and Luca Carlone. On the inclusion of determinant constraints in lagrangian duality for 3D SLAM. In *Robotics: Science and Systems (RSS), Workshop "The problem of mobile sensors: Setting future goals and indicators of progress for SLAM"*. 4

[39] Hayato Waki, Sunyoung Kim, Masakazu Kojima, and Masakazu Muramatsu. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM J. Optim.*, 17(1):218–242, 2006. 5

[40] Chunyu Wang, Yizhou Wang, Zhouchen Lin, Alan Loddon Yuille, and Wen Gao. Robust estimation of 3D human poses from a single image. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2369–2376, 2014. 1, 2

[41] Tillmann Weisser, Jean B Lasserre, and Kim-Chuan Toh. Sparse-BSOS: a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity. *Math. Program. Comput.*, 10(1):1–32, 2018. 5

[42] Jiajun Wu, Tianfan Xue, Joseph J Lim, Yuandong Tian, Joshua B Tenenbaum, Antonio Torralba, and William T Freeman. Single image 3d interpreter network. In *European Conf. on Computer Vision (ECCV)*, pages 365–382. Springer, 2016. 1

[43] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes. In *Robotics: Science and Systems (RSS)*, 2018. 1

[44] Heng Yang, Pasquale Antonante, Vasileios Tzoumas, and Luca Carlone. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. *IEEE Robotics and Automation Letters (RA-L)*, 2020. 3, 6

[45] Heng Yang and Luca Carlone. A polynomial-time solution for robust registration with extreme outlier rates. In *Robotics: Science and Systems (RSS)*, 2019. 6

[46] Heng Yang and Luca Carlone. A quaternion-based certifiably optimal solution to the Wahba problem with outliers. In *Intl. Conf. on Computer Vision (ICCV)*, 2019. 6

[47] Heng Yang, Jingnan Shi, and Luca Carlone. TEASER: Fast and Certifiable Point Cloud Registration. *arXiv preprint arXiv:2001.07715*, 2020. 6

[48] Xiaowei Zhou, Spyridon Leonardos, Xiaoyan Hu, and Kostas Daniilidis. 3D shape reconstruction from 2D landmarks: A convex formulation. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015. 1, 2, 3, 7, 8

[49] Xiaowei Zhou, Menglong Zhu, Spyridon Leonardos, and Kostas Daniilidis. Sparse representation for 3D shape estimation: A convex relaxation approach. *IEEE Trans. Pattern Anal. Machine Intell.*, 39(8):1648–1661, 2017. 1, 2, 3, 4, 6, 7, 8

[50] Xiaowei Zhou, Menglong Zhu, Georgios Pavlakos, Spyridon Leonardos, Konstantinos G Derpanis, and Kostas Daniilidis. Monocap: Monocular human motion capture using a cnn coupled with a geometric prior. *IEEE Trans. Pattern Anal. Machine Intell.*, 41(4):901–914, 2018. 3