Estimation Contracts for Outlier-Robust Geometric Perception

Other titles in Foundations and Trends® in Robotics

Social Interactions for Autonomous Driving: A Review and Perspectives
Wenshuo Wang, Letian Wang, Chengyuan Zhang, Changliu Liu and Lijun Sun
ISBN: 978-1-63828-128-3

Interactive Imitation Learning in Robotics: A Survey
Carlos Celemin, Rodrigo Pérez-Dattari, Eugenio Chisari, Giovanni Franzese, Leandro de Souza Rosa, Ravi Prakash, Zlatan Ajanović, Marta Ferraz, Abhinav Valada and Jens Kober
ISBN: 978-1-63828-126-9

A Survey on the Integration of Machine Learning with Sampling-based Motion Planning
Troy McMahon, Aravind Sivaramakrishnan, Edgar Granados and Kostas E. Bekris
ISBN: 978-1-63828-134-4

Adoption of Robots for Disasters: Lessons from the Response to COVID-19
Robin R. Murphy, Vignesh B.M. Gandudi, Justin Adams, Angela Clendenin and Jason Moats
ISBN: 978-1-68083-862-6

Cybersecurity in Robotics: Challenges, Quantitative Modeling, and Practice
Quanyan Zhu, Stefan Rass, Bernhard Dieber and Víctor Mayoral Vilches
ISBN: 978-1-68083-860-2

Estimation Contracts for Outlier-Robust Geometric Perception

Luca Carlone
MIT
lcarlone@mit.edu

the essence of knowledge
Boston — Delft

Foundations and Trends® in Robotics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

L. Carlone. *Estimation Contracts for Outlier-Robust Geometric Perception*. Foundations and Trends® in Robotics, vol. 11, no. 2-3, pp. 90–224, 2023.

ISBN: 978-1-63828-223-5

Foundations and Trends® in Robotics
Volume 11, Issue 2-3, 2023

Editorial Scope

Topics

Foundations and Trends® in Robotics publishes survey and tutorial articles in the following topics:

- Mathematical modelling
- Kinematics
- Dynamics
- Estimation Methods
- Robot Control
- Planning
- Artificial Intelligence in Robotics
- Software Systems and Architectures
- Mechanisms and Actuators
- Sensors and Estimation
- Planning and Control
- Human-Robot Interaction
- Industrial Robotics
- Service Robotics

Information for Librarians

Foundations and Trends® in Robotics, 2023, Volume 11, 4 issues. ISSN paper version 1935-8253. ISSN online version 1935-8261. Also available as a combined paper and online subscription.

Contents

Estimation Contracts for Outlier-Robust Geometric Perception

Luca Carlone

Laboratory for Information & Decision Systems (LIDS) and the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, USA; lcarlone@mit.edu

---

ABSTRACT

Outlier-robust estimation is a fundamental problem and has been extensively investigated by statisticians and practitioners. The last few years have seen a convergence across research fields towards "algorithmic robust statistics", which focuses on developing tractable outlier-robust techniques for high-dimensional estimation problems. Despite this convergence, research efforts across fields have been mostly disconnected from one another. This monograph bridges recent work on certifiable outlier-robust estimation for geometric perception in robotics and computer vision with parallel work in robust statistics. In particular, we adapt and extend recent results on *robust linear regression* (applicable to the low-outlier regime with $\ll 50\%$ outliers) and *list-decodable regression* (applicable to the high-outlier regime with $\gg 50\%$ outliers) to the setup commonly found

This work was partially funded by the NSF CAREER award "Certifiable Perception for Autonomous Cyber-Physical Systems" and by ARL DCIST CRA W911NF-17-2-0181.

Luca Carlone (2023), "Estimation Contracts for Outlier-Robust Geometric Perception", Foundations and Trends® in Robotics: Vol. 11, No. 2-3, pp 90–224. DOI: 10.1561/2300000077.

Online Appendix is available at http://dx.doi.org/10.1561/2300000077_app.

in robotics and vision, where (i) variables (*e.g.,* rotations, poses) belong to a non-convex domain, (ii) measurements are vector-valued, and (iii) the number of outliers is not known a priori. The emphasis here is on performance guarantees: rather than proposing radically new algorithms, we provide conditions on the input measurements under which modern estimation algorithms (possibly after small modifications) are guaranteed to recover an estimate close to the ground truth in the presence of outliers. These conditions are what we call an "estimation contract". The monograph also provides numerical experiments to shed light on the applicability of the theoretical results and to showcase the potential of list-decodable regression algorithms in geometric perception. Besides the proposed extensions of existing results, we believe the main contributions of this monograph are (i) to unify parallel research lines by pointing out commonalities and differences, (ii) to introduce advanced material (*e.g.,* sum-of-squares proofs) in an accessible and self-contained presentation for the practitioner, and (iii) to point out a few immediate opportunities and open questions in outlier-robust geometric perception.

1

Introduction

Geometric perception is the problem of estimating unknown geometric models (*e.g.,* poses, rotations, 3D structure) from sensor data (*e.g.,* camera images, lidar scans, inertial data, wheel odometry). Geometric perception has been at the center stage of robotics and computer vision research since their inception, and includes problems such as object pose (and possibly shape) estimation [131], [134], robot or camera motion estimation [106], sensor calibration [59], Simultaneous Localization And Mapping (SLAM) [22], and Structure from Motion (SfM) [114], to mention a few.

At its core, geometric perception solves an estimation problem, where, given measurements $\boldsymbol{y}_i, i = 1, \ldots, n$, one has to compute a variable of interest $\boldsymbol{x}^\circ$ (the "ground truth"). For instance, in an object pose estimation problem, $\boldsymbol{x}^\circ$ is the to-be-computed 3D pose of the object (say, a car), while the $\boldsymbol{y}_i$'s might be observations of relevant points on the object (*e.g.,* the wheels and the headlights of the car). The unknown $\boldsymbol{x}^\circ$ and the measurements $\boldsymbol{y}_i$ are related by a *measurement (or generative) model*. In this monograph, we focus our attention on the common case where the measurements are vector-valued, *i.e.,* $\boldsymbol{y}_i \in \mathbb{R}^{d_y}$, and the noise is additive, leading to measurement models in the form:

$$\boldsymbol{y}_i = f_i(\boldsymbol{x}^\circ) + \boldsymbol{\epsilon}, \quad \text{with} \quad \boldsymbol{y}_i \in \mathbb{R}^{d_y} \quad \text{and} \quad \boldsymbol{x}^\circ \in \mathbb{X} \subseteq \mathbb{R}^{d_x}, \tag{1.1}$$

where $f_i(\cdot)$ is a known function, $\boldsymbol{\epsilon}$ is the measurement noise, and $\mathbb{X}$ is the domain of $\boldsymbol{x}^\circ$ (*e.g.,* the set of 3D poses in an object pose estimation problem). As we will see in Section 3, many geometric perception problems have measurement models in the form of eq. (1.1).[1]

When the noise in (1.1) is zero-mean and Gaussian, the maximum likelihood estimate of $\boldsymbol{x}^\circ$ can be computed via standard least squares:[2]

$$\boldsymbol{x}_{\mathsf{LS}} = \underset{\boldsymbol{x}\in\mathbb{X}}{\arg\min} \sum_{i=1}^{n} \|\boldsymbol{y}_i - f_i(\boldsymbol{x})\|_2^2 \,. \tag{LS}$$

While problem (LS) can be still hard to solve (*e.g.,* due to potential non-convexity of $f_i(\cdot)$ or $\mathbb{X}$), its structure —at least for common geometric perception problems— has been extensively studied in robotics and vision, and the literature offers a broad range of solvers, including closed-form solutions [63], iterative local solvers [37], minimal solvers [78], and convex relaxations [24], [25], [68], [104], [109], [134].

Outlier-robust estimation. In practice, many of the measurements fed to the estimation process are *outliers*, *i.e.,* they largely deviate from the measurement model (1.1) and possibly do not carry any information about $\boldsymbol{x}^\circ$. In robotics and vision, the measurements $\boldsymbol{y}_i$ are the result of a pre-processing of the raw sensor data; such preprocessing is often referred to as the *perception front-end*, while the estimation algorithms that compute $\boldsymbol{x}^\circ$ from the $\boldsymbol{y}_i$'s are referred to as the *perception back-end*. For instance, in an object pose estimation problem, the perception front-end extracts the position of relevant features $\boldsymbol{y}_i$ on the object from raw image pixels (typically using a neural network), while the

[1] Assuming additive noise comes at a small loss of generality, *e.g.,* in SLAM and rotation averaging the measurements belong to a smooth manifold rather than a vector space and the noise is multiplicative. However, even in these cases, the resulting outlier-free formulations —under suitable noise assumptions— lead to standard least squares [59], [104], hence we believe adapting the results in this monograph to those setups is indeed possible, see Section 9.

[2] Without loss of generality, we assume $\boldsymbol{\epsilon}$ to have an isotropic Gaussian distribution with identity covariance, but arbitrary covariances can be easily accommodated by rescaling $\boldsymbol{y}_i$ and $f_i(\cdot)$ by the square root of the inverse covariance.

Full text available at: http://dx.doi.org/10.1561/2300000077

back-end computes the object pose given the $\boldsymbol{y}_i$'s. The perception front-end is prone to errors (*e.g.,* the network may mis-detect the wheels of the car in the image), resulting in measurements $\boldsymbol{y}_i$ with large errors. In the presence of outliers, the least squares estimator (LS) is known to produce grossly incorrect results, hence it is desirable to adopt an outlier-robust estimator that can correctly estimate $\boldsymbol{x}^\circ$ in the presence of many outliers. In this monograph, we do not make assumptions on the nature of the outliers and consider the worst case where a fraction β of measurements is arbitrarily corrupted, a setup commonly referred to as the *strong adversary model* in statistics and learning [71].

The robust statistics lens. Classical robust statistics [65], [94], [105], [117] provides many alternative formulations to (LS) that allow regaining robustness to outliers. For instance, if the number of outliers is known, say a fraction β of the n measurements is corrupted, we can use the *Least Trimmed Squares* (LTS) estimator [105] to compute an outlier-robust estimate:

$$\boldsymbol{x}_{\mathsf{LTS}} = \underset{\substack{\boldsymbol{\omega}\in\{0;1\}^n \\ \boldsymbol{x}\in\mathbb{X}}}{\arg\min} \sum_{i=1}^{n} \omega_i \cdot \|\boldsymbol{y}_i - f_i(\boldsymbol{x})\|_2^2 \ , \quad \text{subject to} \quad \sum_{i=1}^{n} \omega_i = \alpha\, n \ , \tag{LTS}$$

where we defined the *inlier rate* $\alpha \triangleq 1 - \beta$, and introduced binary variables $\boldsymbol{\omega} \in \{0;1\}^n$ which are in charge of selecting the best $\alpha\, n$ measurements (when $\omega_i = 1$, the i-th measurement is selected as an inlier by (LTS), while $\omega_i = 0$ otherwise); in words, (LTS) selects the $\alpha\, n$ measurements that induce the smallest error for some estimate $\boldsymbol{x}$ and disregards the remaining measurements as outliers. Unfortunately, the optimization problem (LTS), as well as many other popular outlier-robust formulations, are NP-hard [13] and for a long while no tractable algorithm was available for high-dimensional outlier-robust estimation problems (*e.g.,* in the problems we discuss in Section 3 and Section 9, $\boldsymbol{x}$'s dimension ranges from 9 to potentially more than a thousand). In recent years, *algorithmic robust statistics* came to the rescue, by proposing polynomial-time algorithms for outlier-robust estimation with strong performance guarantees, including [15], [41], [46], [71], [102]. For instance, while not explicitly recognized in the paper, the algorithm by Klivans *et al.* [71] can be understood as a convex relaxation for problem (LTS) for the case where $f_i(\cdot)$ is a real-valued linear function. Many of these works use Lasserre's moment relaxation [81] as an algorithmic workhorse, and adopt the dual view of sum-of-squares relaxations [98] to prove bounds on the quality of the estimates.

The computer vision lens. In typical robotics and vision applications, the number of outliers is unknown, therefore outlier-robust estimators have to simultaneously look for a suitable estimate of $\boldsymbol{x}^\circ$ while searching for a large set of inliers. In computer vision, a common formulation for outlier-robust estimation with unknown number of outliers is *consensus maximization* [32], which searches for the largest set of inliers such that the measurements selected as inliers have a low error with respect to some estimate:

$$\boldsymbol{x}_{\mathsf{MC}} = \underset{\substack{\boldsymbol{\omega}\in\{0;1\}^n \\ \boldsymbol{x}\in\mathbb{X}}}{\arg\max} \sum_{i=1}^{n} \omega_i \,, \qquad \text{subject to} \quad \omega_i \cdot \|\boldsymbol{y}_i - f_i(\boldsymbol{x})\|_2^2 \leq \bar{c}^2 \,, \tag{MC}$$

where the given constant $\bar{c} \geq 0$ is the maximum error for a measurement to be considered an inlier. Problem (MC) has been shown to be inapproximable [2], [31], and the literature has been traditionally split between fast heuristics (which do not provide performance guarantees) and globally optimal solvers (which can compute optimal solutions but run in worst-case exponential time). The recent work [129] shows that for common geometric perception problems, (MC) can be written as a polynomial optimization problem (POP) and relaxed via Lasserre's moment relaxation. The key insight behind [129], reviewed in Section 3, is that —for common perception problems— the domain $\mathbb{X}$ is a basic semi-algebraic set (*i.e.*, it can be written as a set of polynomial inequalities), while with a suitable parametrization, the function $f_i(\cdot)$ becomes a (vector-valued) linear function.

The robotics lens. In robotics, the go-to approach for outlier-robust estimation has been the use of M-estimators [65], which replace the least squares cost in (LS) with a robust loss function. In this monograph we focus on a particular choice of robust loss function, the *truncated least squares (or truncated quadratic)* cost:

$$
\begin{aligned}
\boldsymbol{x}_{\mathsf{TLS}} &= \underset{\boldsymbol{x}\in\mathbb{X}}{\arg\min} \sum_{i=1}^{n} \min\left(\|\boldsymbol{y}_i - f_i(\boldsymbol{x})\|_2^2 \ , \ \bar{c}^2\right) \\
&= \underset{\substack{\boldsymbol{\omega}\in\{0;1\}^n \\ \boldsymbol{x}\in\mathbb{X}}}{\arg\min} \sum_{i=1}^{n} \omega_i \cdot \|\boldsymbol{y}_i - f_i(\boldsymbol{x})\|_2^2 + (1-\omega_i)\cdot \bar{c}^2 \ , \qquad \text{(TLS)}
\end{aligned}
$$

where the objective is the pointwise minimum of a quadratic and a constant function, *i.e.,* it is quadratic for small *residuals* $\|\boldsymbol{y}_i - f_i(\boldsymbol{x})\|_2 \leq \bar{c}$, and becomes constant for large residuals. In the second line in (TLS) we noticed that the truncated least squares cost can be equivalently rewritten using auxiliary binary variables $\boldsymbol{\omega}$, by observing that for two numbers a, b, $\min(a, b) = \min_{\omega\in\{0;1\}} \omega\cdot a+(1-\omega)\cdot b$. Also problem (TLS) has been shown to be inapproximable in the worst case [2]. While traditionally problem (TLS) has been attacked using local solvers [113] or continuation schemes [125], recent work [77], [126], [128], [129], [131] has shown that for common perception problems, (TLS) can be written as a POP and relaxed via Lasserre's moment relaxation. More surprisingly, many works have empirically observed the relaxation to be tight [126], [128], [129], at least for reasonable levels of noise and outliers, with very recent work [101] providing initial theoretical results to support such empirical evidence, at least for the specific problem of rotation search. However, the performance of these estimators is commonly demonstrated via empirical evaluation, and the literature is still lacking more general theoretical guarantees on the quality of the resulting estimates.

Catalyst, convergence, and contribution. Despite the heterogeneity of the formulations reviewed above, we observe that recent years have witnessed a convergence across fields towards designing tractable algorithms for high-dimensional outlier-robust estimation using moment relaxations. A few examples include [23], [70], [71], [77], [101], [126], [128], [129], [131]. Such a convergence has been triggered by the progress in polynomial optimization via moment and sum-of-squares relaxations, starting from the seminal works [81], [89], [98], [99], [111]. At the same time, research across fields has remained disconnected, with researchers being mostly unaware of the parallel work in other areas.

The goal of this monograph is to bridge this gap and connect geometric perception problems in robotics and vision to novel tools in outlier-robust statistics. Towards this goal we adapt and extend recent results from robust statistics to the setup and formulations commonly found in robotics and vision. For the case with low outlier rates (*i.e.,* $\beta \ll 0.5$), we adapt results from [71], which considers outlier-robust regression using least trimmed squares (LTS) with scalar linear measurements, to the robotics setup where the measurements are vector valued and the variables belong to a non-convex domain; we also develop a simple bound on the distance of the estimate from the ground truth (while [71] focuses on bounding the residual errors for the inliers). Then, we extend these results to the case where the number of outliers is unknown. In particular, we compute bounds on the estimation error (*i.e.,* the distance between the estimate and $\boldsymbol{x}^{\circ}$) for (MC) and (TLS). These results constitute the first general performance guarantees for the convex relaxations [126], [128], [129], going beyond the empirical observations in [129] and the problem-specific optimality guarantees in [101], [126].

Then we consider the case with high outlier rates (*i.e.,* $\beta \gg 0.5$), where a majority of the measurements are outliers. While in robotics and vision it has been observed that with random (*i.e.,* non-adversarial) outliers, the point estimators (MC) and (TLS) are still able to retrieve good estimates for $\boldsymbol{x}^{\circ}$ [126], [128], [129], [131], in the presence of adversarial outliers, the estimate resulting from (LTS), (MC), and (TLS) can be arbitrarily far from the ground truth: intuitively, since the outliers constitute the majority of the measurements, they can agree on an arbitrary $\boldsymbol{x}$ and form a large set of mutually consistent measurements that are picked as solution to (LTS), (MC), and (TLS).[3] In robotics

[3]Note that the case with a high number of adversarial outliers is often the one encountered in practice in robotics and vision: think about a motion estimation problem where the robot has to estimate its motion from point features detected by the camera [106]: if there is a large moving object in front of the camera, most features may fall on the moving object (rather that on the static portion of the scene), leading to incorrect motion estimates.

and related fields, this setup has been recognized to require computing multiple estimates, in order to find one that is close to the ground truth, ranging from early work on multi-hypothesis target tracking [6] and particle filters [36], to recent work on multi-hypothesis smoothing [56], [64]. However, none of these works simultaneously provide tractable algorithms and performance guarantees for the resulting estimates. In this monograph, we connect to the recent literature on *list-decodable regression* [70], which proposes polynomial-time estimators that return a small list of estimates such that with high probability at least one of the estimates is close to the ground truth. In particular, we provide an adaptation of the results in [70] to account for vector-valued measurements.

Finally, we present numerical experiments on a canonical geometric perception problem to shed light on the theoretical results. The experiments provide encouraging evidence that many of the assumptions supporting the theoretical analysis (*e.g.,* certifiable hypercontractivity) are often satisfied by real data. At the same time, they reveal a large gap between theory (which mostly guarantees performance for high-order, computationally expensive moment relaxations) and practice (where low-order relaxations already exhibit impressive performance). Our numerical evaluation also provides the first empirical evidence that a sparse low-order moment relaxation for list-decodable regression (based on a modified version of the algorithm proposed in [70]) is able to accurately recover estimates in geometric perception problems with high outlier rates, where (LTS), (MC), and (TLS) are doomed to fail. Moreover, the experiments show that if the measurements are generated by multiple estimates (*e.g.,* different subsets of measurements are generated by different variables $\boldsymbol{x}^{\circ}$), then our sparse moment relaxation for list-decodable regression is able to simultaneously recover *all* the estimates generating the data.[4] We release open-source code to reproduce our numerical experiments, including an implementation of key algorithms covered in this monograph at https://github.com/MIT-SPARK/estimation-contracts.

[4]With reference to the motion estimation example in footnote 3, such an algorithm would simultaneously recover the motion of all the objects in the scene, rather than just the motion with respect to the object capturing most point features, which would be quite useful in practical applications.

We remark that the emphasis in this monograph is on performance guarantees. We do not present new algorithms (we mostly propose small modifications to existing algorithms) but rather try to address the question: under which conditions on the input measurements can we guarantee that modern outlier-robust estimation algorithms based on moment relaxations recover an estimate close to the ground truth in the presence of outliers? These conditions are what we call an "estimation contract". Besides the proposed extensions of existing results, we believe the main contributions of this monograph are (i) to unify parallel research lines by pointing out commonalities and differences, (ii) to introduce advanced material (*e.g.,* sum-of-squares proofs) in an accessible and self-contained presentation for the practitioner, and (iii) to point out a few immediate opportunities and open questions in outlier-robust geometric perception. This "unification" is expected to benefit both practitioners and researchers in robust statistics.

On the robotics and computer vision side, this monograph provides new and fairly general performance guarantees for robust estimation algorithms based on moment relaxations, applied to geometric perception problems. Moreover, the monograph reviews a new proof system (based on sum-of-squares proofs) that provides a richer language to discuss properties of moment relaxations beyond the typical analysis based on a manual design of dual certificates [52], [101], [104], [126]. Furthermore, it positions list-decodable regression based on moment relaxations as a useful and computationally tractable tool for multi-hypotheses estimation. On the robust statistics side, we hope the reader will be intrigued by the remarks about the practical performance and the empirical tightness of the moment relaxation of (TLS) (discussed in greater detail in [129]) and the practical performance of a low-order relaxation for list-decodable regression, which we believe deserve further investigation. We also hope to attract further attention towards the case where the number of outlier is unknown and the variables are confined to semi-algebraic sets, which is the setup commonly encountered in robotics and vision problems.

Monograph structure. Section 2 starts by reviewing related works across fields. Section 3 showcases the fact that many estimation problems in robotics and vision can be formulated using a linear measurement model with variables belonging to a basic semi-algebraic set. Section 4 introduces notation and preliminaries (while postponing as many details as possible to the appendices). Section 5 succinctly states the problem of outlier-robust estimation and our quest for estimation contracts. Section 6 studies the case with low outlier rates and provides error bounds for (LTS), (MC), and (TLS). Section 7 studies the case with high outlier rates and adapts results from list-decodable regression. Section 8 presents numerical experiments on a rotation search problem. Section 9 discusses opportunities and open problems, and Section 10 concludes the monograph.

The appendices provide background information regarding moment relaxations (Appendix A), pseudo-distributions (Appendix B), sum-of-squares proofs (Appendix C), and algorithmic details about sparse list-decodable estimation (Appendix D). The technical proofs are contained in the Online Appendix.[5]

Learning paths. The main body of this monograph (up to Section 10) is designed to be self-contained, and readers interested in getting a bird's-eye view of the technical tools underlying our results, the corresponding performance guarantees, and sparse list-decodable estimation can focus their attention on these sections. Expert readers from statistics might find the numerical experiments of particular interest. Junior researchers in computer vision and robotics interested in advancing this research are recommended to carefully read Appendices A-C, which should serve as a gentle introduction to the key technical tools. Beyond this monograph, the book [60] and monograph [37] provide necessary background information about geometric perception, the book [17] provides an in-depth introduction to semidefinite programming and sum-of-squares relaxations, and the notes [9] and monograph [55] provide a more formal introduction to sum-of-squares proofs.

[5] http://dx.doi.org/10.1561/2300000077_app.

Acknowledgments

We thank Pablo Parrilo for pointing out relevant work in robust statistics and for suggesting potential connections that led to the development of this monograph. We also thank Heng Yang for useful discussion about the relation between Putinar's and Schmudgen's Positivstellensätze and for documenting and sharing relevant code about rotation search and image stitching. Finally, we thank Sushrut Karmalkar for useful discussions on certifiable anti-concentration and list-decodable regression, Pravesh Kothari for useful discussions on outlier-robust estimation and help with the proof of Theorem 12, and Tat-Jun Chin for early feedback on this monograph.

Appendices

A

An Algorithmic View of Lasserre's Hierarchy of Moment Relaxations

Here we provide an algorithmic (and somewhat unorthodox) view of Lasserre's hierarchy of moment relaxations [81]; we refer the reader to [79], [80] for a more standard introduction.

Lasserre's hierarchy provides a systematic way to relax a polynomial optimization problem (POP) into a semidefinite (convex) program. We start by restating (POP):

$$p^\star \triangleq \min_{\boldsymbol{x} \in \mathbb{R}^{d_x}} \left\{ p(\boldsymbol{x}) \;\middle|\; \begin{array}{l} h_i(\boldsymbol{x}) = 0, i = 1, \dots, l_h \\ g_j(\boldsymbol{x}) \geq 0, j = 1, \dots, l_g \end{array} \right\}, \tag{POP}$$

where $p(\boldsymbol{x}), h_i(\boldsymbol{x}), g_j(\boldsymbol{x})$ are polynomials in the variable $\boldsymbol{x} \in \mathbb{R}^{d_x}$.

The key idea behind Lasserre's hierarchy of moment relaxations is to (i) rewrite the polynomial optimization problem (POP) using the moment matrix $\boldsymbol{X}_{2r}$,[1] (ii) relax the (non-convex) rank-1 constraint on $\boldsymbol{X}_{2r}$, and (iii) add redundant constraints that are trivially satisfied in (POP) but might still improve the quality of the relaxation; as shown below, this leads to a semidefinite program.

[1]Recall that the moment matrix is defined as $\boldsymbol{X}_{2r} \triangleq [\boldsymbol{x}]_r[\boldsymbol{x}]_r^\mathsf{T}$, where $[\boldsymbol{x}]_r$ is the vector of monomials of degree up to r. For instance, for $\boldsymbol{x} = [x_1\,;\,x_2]$ and $r = 2$, the matrix $\boldsymbol{X}_{2r}$ takes the form in eq. (4.2).

(i) Rewriting (POP) using $\boldsymbol{X}_{2r}$. Recall that any polynomial of degree up to $2r$ can be written as a linear function of the moment matrix $\boldsymbol{X}_{2r}$ (*cf.* Section 4.2). Therefore, we pick a positive integer r (the *order* of the relaxation) such that $2r \geq \max\{\deg(p), \deg(h_1), \ldots, \deg(h_{l_h}), \deg(g_1), \ldots, \deg\left(g_{l_g}\right)\}$, such that we can express both objective function and constraints as a linear function of $\boldsymbol{X}_{2r}$. With this choice of r, we can rewrite the objective and the equality constraints in (POP) as:

$$\text{objective}: \quad \langle \boldsymbol{C}_1, \boldsymbol{X}_{2r} \rangle \tag{A.1}$$

$$\text{equality constraints}: \langle \boldsymbol{A}_{\mathrm{eq},j}, \boldsymbol{X}_{2r} \rangle = 0,\ j = 1, \ldots, l_h, \tag{A.2}$$

for suitable matrices $\boldsymbol{C}_1$ and $\boldsymbol{A}_{\mathrm{eq},j}$.

(ii) Relaxing the (non-convex) rank-1 constraint on $\boldsymbol{X}_{2r}$. At the previous point we noticed we can rewrite objective and constraints in (POP) as linear (hence convex) functions of $\boldsymbol{X}_{2r}$. However, $\boldsymbol{X}_{2r}$ still belongs to the set of positive-semidefinite rank-1 matrices (since it is defined as $[\boldsymbol{x}]_r[\boldsymbol{x}]_r^\mathsf{T}$, where $[\boldsymbol{x}]_r$ is a vector of monomials), which is a non-convex set due to the rank constraint. Therefore, we simply relax the rank constraint and only enforce:

$$\text{pseudo-moment matrix}: \quad \boldsymbol{X}_{2r} \succeq 0. \tag{A.3}$$

(iii) Adding redundant constraints. Since we have relaxed (POP) by re-parametrizing it using $\boldsymbol{X}_{2r}$ and dropping the rank constraint, the final step to obtain Lasserre's relaxation consists in adding extra constraints to make the relaxation tighter. First of all, we observe that there are multiple repeated entries in the moment matrix (*e.g.,* in (4.2), the entry x_1x_2 appears 4 times in the matrix). Therefore, we can enforce these entries to be the same. In general, this leads to $m_{\mathrm{mom}} = \mathfrak{t}(\underline{d}_r) - \underline{d}_{2r} + 1$ linear constraints, where $\underline{d}_{2r} \triangleq \binom{d_x+2r}{2r}$ (the size of the monomial basis of degree up to $2r$, *i.e.,* $[\boldsymbol{x}]_{2r}$) and $\mathfrak{t}(n) \triangleq \frac{n(n+1)}{2}$ is the dimension of $\mathbb{S}^n$. These constraints are typically called *moment constraints*:

$$\begin{aligned}\text{moment constraints}: \quad & \langle \boldsymbol{A}_{\mathrm{mom},0}, \boldsymbol{X}_{2r} \rangle = 1, \\ & \langle \boldsymbol{A}_{\mathrm{mom},j}, \boldsymbol{X}_{2r} \rangle = 0, \\ & j = 1, \ldots, \mathfrak{t}(\underline{d}_r) - \underline{d}_{2r},\end{aligned} \tag{A.4}$$

where $\boldsymbol{A}_{\text{mom},0}$ is all-zero except $[\boldsymbol{A}_{\text{mom},0}]_{11} = 1$, and it is used to define the constraint $[\boldsymbol{X}_{2r}]_{11} = 1$, following from the definition of the moment matrix (see eq. (4.2)).

Second, we can also add *redundant* equality constraints. Simply put, if $h_i = 0$, then also $h_i \cdot x_1 = 0$, $h_i \cdot x_2 = 0$, and so on, for any monomial we multiply by h_i. Since via $\boldsymbol{X}_{2r}$ we can represent any polynomial of degree up to $2r$, we can write as linear constraints any polynomial equality in the form $h_i \cdot [\boldsymbol{x}]_{2r-\deg(h_i)} = \boldsymbol{0}$ (the degree of the monomials is chosen such that the product does not exceed degree $2r$). These new equalities can again be written linearly as:

$$\begin{aligned}\text{(redundant) equality constraints} : \langle \boldsymbol{A}_{\text{req},ij}, \boldsymbol{X}_{2r} \rangle = 0,\\ i = 1, \ldots, l_h, \quad j = 1, \ldots, \underline{d}_{2r-\deg(h_i)},\end{aligned} \tag{A.5}$$

for suitable $\boldsymbol{A}_{\text{req},ij}$. Since the first entry of $[\boldsymbol{x}]_{2r-\deg(h_i)}$ is always 1 (*i.e.*, the monomial of order zero), eq. (A.5) already includes the original equality constraints in (A.2).

Finally, we observe that if $g_j \geq 0$, then for any positive semidefinite matrix $\boldsymbol{M}$, it holds $g_j \cdot \boldsymbol{M} \succeq 0$. Since we can represent any polynomial of order up to $2r$ as a linear function of $\boldsymbol{X}_{2r}$, we can add redundant constraints in the form $g_j \cdot \boldsymbol{X}_{2(r-\lceil \deg(g_j)/2 \rceil)} \succeq 0$ (by construction $g_j \cdot \boldsymbol{X}_{2(r-\lceil \deg(g_j)/2 \rceil)}$ only contains polynomials of degree up to $2r$). To phrase the resulting relaxation in the standard form (SDP), it is common to add extra matrix variables $\boldsymbol{X}_{g_j} = g_j \cdot \boldsymbol{X}_{2(r-\lceil \deg(g_j)/2 \rceil)}$ for $j = 1, \ldots, l_g$ (the *localizing matrices* [79, §3.2.1]) and then force these matrices to be a linear function of $\boldsymbol{X}_{2r}$:

$$\text{localizing matrices} : \quad \boldsymbol{X}_{g_j} \succeq 0, \quad j = 1, \ldots, l_g, \tag{A.6}$$

$$\begin{aligned}\text{localizing constraints} : \langle \boldsymbol{A}_{\text{loc},jkh}, \boldsymbol{X}_{2r} \rangle = [\boldsymbol{X}_{g_j}]_{hk},\\ j = 1, \ldots, l_g, \quad 1 \leq h \leq k \leq \underline{d}_{r-\lceil \deg(g_j)/2 \rceil},\end{aligned} \tag{A.7}$$

where the linear constraints (for some matrix $\boldsymbol{A}_{\text{loc},jkh}$) enforce each entry of $\boldsymbol{X}_{g_j}$ to be a linear combination of entries of the matrix $\boldsymbol{X}_{2r}$.

Following steps (i)-(iii) above, it is straightforward to obtain the following semidefinite program:

$$f^\star_{2r} = \min_{\boldsymbol{X} = (\boldsymbol{X}_{2r}, \{\boldsymbol{X}_{g_j}\}_{j \in [l_g]})} \{ \langle \boldsymbol{C}_1, \boldsymbol{X}_{2r} \rangle \mid \mathcal{A}(\boldsymbol{X}) = \boldsymbol{b}, \boldsymbol{X} \succeq 0 \}, \tag{A.8}$$

where the variable $\boldsymbol{X} = (\boldsymbol{X}_{2r}, \{\boldsymbol{X}_{g_j}\}_{j\in[l_g]})$ is a collection of positive-semidefinite matrices (*cf.* (A.3) and (A.9)), the objective is the one given in (A.1), and the linear constraints $\mathcal{A}(\boldsymbol{X}) = \boldsymbol{b}$ collect all the constraints in (A.4), (A.5), and (A.7). Problem (A.8) can be readily formulated as a multi-block SDP in the primal form (SDP), which matches the data format used by common SDP solvers. The matrix $\boldsymbol{X}_{2r}$ solving (A.8) is typically referred to as the *pseudo-moment matrix*.[2] One can solve the relaxation for different choices of r, leading to a *hierarchy* of convex relaxations.

While we presented Lasserre's hierarchy in a somewhat procedural way, the importance of the hierarchy lies in its stunning theoretical properties, that we review below.

Theorem A.1 (Lasserre's Hierarchy [79], [81], [91])**.** Let $-\infty < p^\star < \infty$ be the optimum of (POP) and $f^\star_{2r}$ (resp. $\boldsymbol{X}^\star_{2r}$) be the optimum (resp. one optimizer) of (A.8), and assume (POP) is explicitly bounded (*i.e.,* it satisfies the Archimedeanness condition in [17, Definition 3.137]), then

(i) (lower bound and convergence) $f^\star_{2r}$ converges to $p^\star$ from below as $r \to \infty$, and convergence occurs at a finite r under suitable technical conditions [91];

(ii) (rank-one solutions) if $f^\star_{2r} = p^\star$ at some finite r, then for every global minimizer $\boldsymbol{x}^\star$ of (POP), $\boldsymbol{X}^\star_{2r} \triangleq [\boldsymbol{x}^\star]_r[\boldsymbol{x}^\star]_r^\mathsf{T}$ is optimal for (A.8), and every rank-one optimal solution $\boldsymbol{X}^\star_{2r}$ of (A.8) can be written as $[\boldsymbol{x}^\star]_r[\boldsymbol{x}^\star]_r^\mathsf{T}$ for some $\boldsymbol{x}^\star$ that is optimal for (POP);

(iii) (optimality certificate) if $\operatorname{rank}(\boldsymbol{X}^\star_{2r}) = 1$ at some finite r, then $f^\star_{2r} = p^\star$.

Theorem A.1 states that (A.8) provides a hierarchy of lower bounds for (POP). When the relaxation is exact ($p^\star = f^\star_{2r}$), global minimizers of (POP) correspond to rank-one solutions of (A.8). Moreover, after solving the convex SDP (A.8), one can check the rank of the optimal solution $\boldsymbol{X}^\star_{2r}$ to obtain a *certificate* of global optimality.

[2]The rationale behind this name will become apparent in Appendix B.

Further tightening the relaxation. As we discussed above, in the standard presentation of Lasserre's hierarchy, one adds a localizing matrix for each inequality constraint to enforce constraints such as $g_j \cdot \boldsymbol{X}_{2(r-\lceil \deg(g_j)/2 \rceil)} \succeq 0$. However, in principle, we could also add constraints enforcing $g_{j_1} \cdot g_{j_2} \cdot \boldsymbol{M} \succeq 0$, for any pair of inequality constraints $g_{j_1} \geq 0$ and $g_{j_2} \geq 0$, for $j_1, j_2 \in [l_g]$. More generally, we can add constraints $\prod_{j \in \mathcal{S}} g_j \cdot \boldsymbol{M} \succeq 0$, for any subset $\mathcal{S} \subseteq [l_g]$ as long as $\deg\left(\prod_{j \in \mathcal{S}} g_j\right)$ has degree no larger than $2r$. After adding those extra constraints, we can still phrase the resulting relaxation in the standard form (SDP), by adding extra matrix variables $\boldsymbol{X}_{\mathcal{S}} = \prod_{j \in \mathcal{S}} g_j \cdot \boldsymbol{X}_{2\left(r - \left\lceil \sum_{j \in \mathcal{S}} \deg(g_j)/2 \right\rceil\right)}$, and then forcing these matrices to be a linear function of $\boldsymbol{X}_{2r}$:

$$\text{localizing matrices}: \quad \boldsymbol{X}_{\mathcal{S}} \succeq 0, \quad \mathcal{S} \subseteq [l_g], \tag{A.9}$$

$$\begin{aligned} \text{localizing constraints}: \langle \boldsymbol{A}_{\text{loc},\mathcal{S},kh}, \boldsymbol{X}_{2r} \rangle &= [\boldsymbol{X}_{\mathcal{S}}]_{hk}, \\ \mathcal{S} \subseteq [l_g], \quad 1 \leq h \leq k \leq &\, \underline{d}_{r - \left\lceil \sum_{j \in \mathcal{S}} \deg(g_j)/2 \right\rceil}, \end{aligned} \tag{A.10}$$

where, similarly to the standard Lasserre's relaxation, the linear constraints (for some suitable matrices $\boldsymbol{A}_{\text{loc},\mathcal{S},kh}$) enforce each entry of $\boldsymbol{X}_{\mathcal{S}}$ to be a linear combination of the entries in $\boldsymbol{X}_{2r}$.

The additional constraints in eq. (A.10) make the relaxation tighter compared to the standard presentation of Lasserre's relaxation, but are not necessary to obtain the convergence result in Theorem A.1, which holds regardless for explicitly bounded constraint sets. However, these constraints become necessary to obtain convergence results akin to Theorem A.1 for the case where the set of constraints is *not* explicitly bounded (see [55, Section 3.3] and [17, p. 115] for a more extensive discussion). For this reason, in order to maintain generality, related work following the "proofs to algorithms" paradigm typically assumes those constraints to be present, see, *e.g.,* [70], [71]. These terms will indeed appear in the definitions of sos proofs and constrained pseudo-distribution, see Appendix B. In order to keep the definitions in our monograph consistent with [70], [71], we will also assume these terms to be present, even though they are not strictly necessary under Assumption 3.2.

B

Pseudo-distributions and Moment Relaxations

The start of this appendix follows standard introductions about pseudo-distributions given in related work [9], [70], [71], [75], while later we attempt to draw more explicit connections with the optimization machinery in Appendix A. Although such a connection is self-evident to the expert reader (indeed pseudo-distributions are the language traditionally used to justify the moment relaxation [79]), such a connection is often less immediate for the practitioner, in particular when taking the algorithmic view of moment relaxations presented in Appendix A.

Pseudo-distributions. Pseudo-distributions are a generalization of the concept of probability distribution. A standard probability distribution μ with finite support in $\mathbb{R}^{d_x}$ is simply a function $\mu : \mathbb{R}^{d_x} \mapsto \mathbb{R}$ such that $\sum_{\boldsymbol{x} \in \text{support}(\mu)} \mu(\boldsymbol{x}) = 1$ and $\mu(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$. In other words, if support(μ) is a finite collection a points in $\mathbb{R}^{d_x}$, μ assigns a non-negative probability mass to each of these points, such that those probabilities sum up to 1. Similarly, a pseudo-distribution $\tilde{\mu}$ is a finitely supported function such that $\sum_{\boldsymbol{x} \in \text{support}(\tilde{\mu})} \tilde{\mu}(\boldsymbol{x}) = 1$ but in this case the non-negativity condition is replaced by a milder condition (*i.e.,* a pseudo-distribution can assume negative values over its support).

In order to formally introduce the notion of pseudo-distribution, we start by defining the *pseudo-expectation* of a function $f : \mathbb{R}^{d_x} \mapsto \mathbb{R}$ under a finitely supported function $\tilde{\mu}$:

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[f(\boldsymbol{x})\right] \doteq \sum_{\boldsymbol{x} \in \text{support}(\tilde{\mu})} f(\boldsymbol{x}) \cdot \tilde{\mu}(\boldsymbol{x}). \tag{B.1}$$

We are now ready to formally define a pseudo-distribution.

Definition B.1 (Pseudo-distribution). A finitely supported function $\tilde{\mu} : \mathbb{R}^{d_x} \mapsto \mathbb{R}$ is a *level-ℓ pseudo-distribution* if $\tilde{\mathbb{E}}_{\tilde{\mu}}[1] = 1$ and $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[f(\boldsymbol{x})^2\right] \geq 0$ for all polynomials f of degree $\deg(f) \leq \ell/2$.

In words, $\tilde{\mu}$ is a function that is allowed to become negative as long as its "expectation" (more precisely, pseudo-expectation) with respect to every squared polynomials $f(\boldsymbol{x})^2$ of sufficiently low degree remains positive. It is possible to show that a level-∞ pseudo-distribution is an actual probability distribution, since the condition $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[f(\boldsymbol{x})^2\right] \geq 0$ would enforce $\tilde{\mu}$ to remain positive (in this case the pseudo-expectation becomes the traditional expectation of the distribution).

Towards reconnecting pseudo-distributions with the optimization machinery in Appendix A, we start by observing the following link between pseudo-distributions and pseudo-moment matrices.

Lemma B.1 (Pseudo-moment matrix [9]). Let $\tilde{\mu} : \mathbb{R}^{d_x} \mapsto \mathbb{R}$ be a finitely supported function with $\tilde{\mathbb{E}}_{\tilde{\mu}}[1]{=}1$. Then, $\tilde{\mu}$ is a level-ℓ pseudo-distribution iff the pseudo-moment matrix $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[[\boldsymbol{x}]_{\ell/2}[\boldsymbol{x}]_{\ell/2}^{\mathsf{T}}\right]$ is positive semidefinite, where $[\boldsymbol{x}]_{\ell/2}$ is the vector of monomials of degree up to $\ell/2$.

Now we define what it means for a pseudo-distribution to satisfy a set of polynomial constraints.

Definition B.2 (Constrained pseudo-distribution). Let $\mathcal{A} \doteq \{f_1 \geq 0, \ldots, f_m \geq 0\}$ be a set of polynomial constraints over $\mathbb{R}^{d_x}$. Let $\tilde{\mu} : \mathbb{R}^{d_x} \mapsto \mathbb{R}$ be a level-ℓ pseudo-distribution. We say that $\tilde{\mu}$ *satisfies* $\mathcal{A}$ at degree k, denoted as $\tilde{\mu} \models^{k} \mathcal{A}$, if every set $\mathcal{S} \subset [m]$ and every sum-of-squares polynomial h on $\mathbb{R}^{d_x}$ with $\deg(h) + \sum_{i \in \mathcal{S}} \max\{\deg(f_i), k\} \leq \ell$ satisfies:

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[h \cdot \prod_{i \in \mathcal{S}} f_i\right] \geq 0. \tag{B.2}$$

Moreover, we say that $\tilde{\mu} \stackrel{k}{\models} \mathcal{A}$ holds *approximately* if the above inequalities are satisfied up to an error of $2^{-d_x^{\ell}} \cdot \|h\| \cdot \prod_{i \in \mathcal{S}} \|f_i\|$, where $\|\cdot\|$ denotes the Euclidean norm of the coefficients of the polynomial.

The notion of pseudo-distributions *approximately* satisfying a set of constraints is useful to account for the practical observation that numerical SDP solvers (which we are going to use to find pseudo-distributions, as discussed later in this section) will only satisfy the constraints up to some numerical tolerance, and we have to make sure that such numerical errors do not lead us to draw incorrect conclusions using the sos proof system (see Appendix C).

In this monograph, we make use of the following facts about pseudo-distributions.

Fact B.2 (Linearity [8]). Let f, g be polynomials of degree at most ℓ in indeterminate $\boldsymbol{x} \in \mathbb{R}^{d_x}$ and take $\alpha, \beta \in \mathbb{R}$. Then, for any level-$\ell$ pseudo-distribution $\tilde{\mu}$,

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\alpha\, f(\boldsymbol{x}) + \beta\, g(\boldsymbol{x})\right] = \alpha \tilde{\mathbb{E}}_{\tilde{\mu}}\left[f(\boldsymbol{x})\right] + \beta \tilde{\mathbb{E}}_{\tilde{\mu}}\left[g(\boldsymbol{x})\right]. \tag{B.3}$$

Fact B.3 (Cauchy-Schwarz for pseudo-distributions [70]). Let f, g be polynomials of degree at most ℓ in indeterminate $\boldsymbol{x} \in \mathbb{R}^{d_x}$. Then, for any level-$\ell$ pseudo-distribution $\tilde{\mu}$,

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[f \cdot g\right] \leq \sqrt{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[f^2\right]} \cdot \sqrt{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[g^2\right]}, \tag{B.4}$$

and (specializing the result above to $g = 1$):

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[f\right]^2 \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[f^2\right]. \tag{B.5}$$

Fact B.4 (Hölder's inequality for pseudo-distributions [71]). Let f, g be sos polynomials. Let p, q be positive integers such that $1/p + 1/q = 1$. Then, for any pseudo-distribution $\tilde{\mu}$ of level $\ell \geq pq \cdot \deg(f) \cdot \deg(g)$, we have:

$$\left(\tilde{\mathbb{E}}_{\tilde{\mu}}\left[f \cdot g\right]\right)^{pq} \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[f^p\right]^q \cdot \tilde{\mathbb{E}}_{\tilde{\mu}}\left[g^q\right]^p. \tag{B.6}$$

In particular, for all even integers $k \geq 2$, and polynomial f with $\deg(f) \cdot k \leq \ell$:

$$\left(\tilde{\mathbb{E}}_{\tilde{\mu}}\left[f\right]\right)^k \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[f^k\right]. \tag{B.7}$$

Fact B.5 (Norm inequality for pseudo-distributions)**.** Let $\boldsymbol{v}$ be an m-vector with polynomial entries of degree at most $\ell/2$ in indeterminate $\boldsymbol{x} \in \mathbb{R}^{d_x}$. Then, for any degree-$\ell$ pseudo-distribution $\tilde{\mu}$,

$$\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\boldsymbol{v}\right]\right\|_2^2 \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\|\boldsymbol{v}\|_2^2\right]. \tag{B.8}$$

Proof. By definition, $\|\boldsymbol{v}\|_2^2 = \sum_{i=1}^m v_i^2$. Moreover, by (B.7), $\left(\tilde{\mathbb{E}}_{\tilde{\mu}}\left[v_i\right]\right)^2 \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[v_i^2\right]$. Therefore:

$$\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\boldsymbol{v}\right]\right\|_2^2 = \sum_{i=1}^m \left(\tilde{\mathbb{E}}_{\tilde{\mu}}\left[v_i\right]\right)^2 \leq \sum_{i=1}^m \tilde{\mathbb{E}}_{\tilde{\mu}}\left[v_i^2\right] \overbrace{=}^{linearity} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\sum_{i=1}^m v_i^2\right] = \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\|\boldsymbol{v}\|_2^2\right], \tag{B.9}$$

proving the claim. ∎

Making the connection with moment relaxations explicit. The non-expert reader might still be confused about the relation between pseudo-distributions and moment relaxations. To shed some light, let us restate our (POP):

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^{d_x}} \quad & p(\boldsymbol{x}) \\ \text{subject to} \quad & h_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, l_h \\ & g_j(\boldsymbol{x}) \geq 0, \quad j = 1, \ldots, l_g. \end{aligned} \tag{B.10}$$

Now, we start by relaxing (B.10) using pseudo-distributions, and show that this leads back to the relaxation presented in Appendix A. In particular, we relax (B.10) to:

$$\begin{aligned} \min_{\tilde{\mu}} \quad & \tilde{\mathbb{E}}_{\tilde{\mu}}\left[p(\boldsymbol{x})\right] && \text{(B.11)} \\ \text{subject to} \quad & \tilde{\mu} \text{ is a level-}\ell \text{ pseudo-distribution} && \text{(B.12)} \\ & \tilde{\mathbb{E}}_{\tilde{\mu}}\left[h_i(\boldsymbol{x}) \cdot q(\boldsymbol{x})\right] = 0, && \text{(B.13)} \\ & \text{for all } i = 1, \ldots, l_h \text{ and for all } q \in \mathbb{R}[\boldsymbol{x}], \text{ such that } \deg\left(h_i \cdot q\right) \leq \ell \\ & \tilde{\mathbb{E}}_{\mu}\left[\textstyle\prod_{j\in\mathcal{S}} g_j(\boldsymbol{x}) \cdot s(\boldsymbol{x})^2\right] \geq 0, && \text{(B.14)} \\ & \text{for all } \mathcal{S} \subseteq [l_g] \text{ and for all } s \in \mathbb{R}[\boldsymbol{x}] \text{ such that } \deg\left(\textstyle\prod_{j\in\mathcal{S}} g_j \cdot s^2\right) \leq \ell. \end{aligned}$$

Despite the complexity of (B.11), it is apparent that (B.11) is a relaxation of (B.10): for any $\boldsymbol{x}$ that is feasible for (B.10) (*i.e.*, that satisfies

$h_i(\boldsymbol{x}) = 0$ and $g_j(\boldsymbol{x}) \geq 0$), we can define a (pseudo-)distribution μ_x supported on $\boldsymbol{x}$ (*i.e.*, $\mu_x(\boldsymbol{x}) = 1$ and zero elsewhere) which is also feasible for (B.11) (such pseudo-distribution μ_x is such that $\tilde{\mathbb{E}}_{\mu_x}[p(\boldsymbol{x})] = p(\boldsymbol{x})$ for any polynomial p, hence also preserving the objective of (B.10)). Indeed, it is possible to show that if we require $\tilde{\mu}$ to be an actual distribution, and replace the pseudo-expectations with actual expectations, then (B.11) becomes equivalent to (B.10), see [80] for a more extensive discussion. The advantage of the relaxation (B.11) is its tractability: while (B.10) is NP-hard [80], the relaxation (B.11) can be written as a semidefinite program (SDP) and solved in polynomial time. Indeed, in the following we show that rewriting (B.11) as an SDP leads us back to the same moment relaxation we procedurally introduced in Appendix A. Towards this goal, we will need some extra notation.

Preliminaries to connect problem (B.11) *with Appendix A:* Recall that $[\boldsymbol{x}]_{\ell/2}$ is the vector of monomials of degree up to $\ell/2$ and therefore the moment matrix $\boldsymbol{X}_\ell \triangleq [\boldsymbol{x}]_{\ell/2}[\boldsymbol{x}]_{\ell/2}^{\mathsf{T}}$ contains all monomials of degree up to ℓ. It will be useful to define (and visualize) the pseudo-expectation of the moment matrix: $\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{X}_\ell]$. For instance, for the case with $\boldsymbol{x} = [x_1\,;\,x_2]$ and $\ell = 4$:

$$\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{X}_4] \triangleq \begin{bmatrix} 1 & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_2^2] \\ \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^3] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2^2] \\ \tilde{\mathbb{E}}_{\tilde{\mu}}[x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_2^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_2^3] \\ \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^3] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^4] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^3x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2x_2^2] \\ \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^3x_2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2x_2^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2^3] \\ \tilde{\mathbb{E}}_{\tilde{\mu}}[x_2^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_2^3] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1^2x_2^2] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_1x_2^3] & \tilde{\mathbb{E}}_{\tilde{\mu}}[x_2^4] \end{bmatrix} \tag{B.15}$$

In the following, we will also need a more convenient way to index the monomials in $[\boldsymbol{x}]_\ell$ (and, as a consequence, the entries of $\boldsymbol{X}_\ell$ and $\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{X}_\ell]$). Using standard notation, for a vector $\boldsymbol{\alpha} \in \mathbb{N}^{d_x}$, we write $\boldsymbol{x}^{\boldsymbol{\alpha}}$ to denote the monomial with exponents $\boldsymbol{\alpha}$ (for instance, for $\boldsymbol{\alpha} = [1\,;\,3\,;\,0\,;\,5]$, $\boldsymbol{x}^{\boldsymbol{\alpha}} = x_1x_2^3x_4^5$). We also denote $|\boldsymbol{\alpha}| \triangleq \sum_{i=1}^{d_x} \alpha_i$, which is the degree of the monomial. Using this notation, we can index with $\boldsymbol{\alpha}$ the monomials appearing in $[\boldsymbol{x}]_\ell$. For instance, for $\ell = 2$:

$$[\boldsymbol{x}]_2 \triangleq [\overbrace{1}^{[0\,;\,0]}\,;\,\overbrace{x_1}^{[1\,;\,0]}\,;\,\overbrace{x_2}^{[0\,;\,1]}\,;\,\overbrace{x_1^2}^{[2\,;\,0]}\,;\,\overbrace{x_1x_2}^{[1\,;\,1]}\,;\,\overbrace{x_2^2}^{[0\,;\,2]}], \tag{B.16}$$

where for each monomial, we reported the corresponding "index" $\boldsymbol{\alpha}$.

We can similarly index the rows and columns of $\boldsymbol{X}_\ell$ and $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\boldsymbol{X}_\ell\right]$ using two indices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. For instance, the entry of the matrix indexed by row $\boldsymbol{\alpha} = [2\,;\,0]$ and column $\boldsymbol{\alpha} = [0\,;\,1]$ in (B.15) will be $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[x_1^2 x_2\right]$. Note that the monomial appearing in row $\boldsymbol{\alpha}$ and column $\boldsymbol{\beta}$ of the moment matrix $\boldsymbol{X}_\ell$ will always have exponent $\boldsymbol{\alpha}+\boldsymbol{\beta}$, since, due to the definition of the moment matrix ($\boldsymbol{X}_\ell \triangleq [\boldsymbol{x}]_{\ell/2}[\boldsymbol{x}]_{\ell/2}^\mathsf{T}$) its entry $[\boldsymbol{X}_\ell]_{\boldsymbol{\alpha}\boldsymbol{\beta}} = \boldsymbol{x}^{\boldsymbol{\alpha}} \cdot \boldsymbol{x}^{\boldsymbol{\beta}}$ and for two monomials $\boldsymbol{x}^{\boldsymbol{\alpha}}$ and $\boldsymbol{x}^{\boldsymbol{\beta}}$, it holds:[1]

$$\boldsymbol{x}^{\boldsymbol{\alpha}} \cdot \boldsymbol{x}^{\boldsymbol{\beta}} = \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}. \tag{B.17}$$

Finally, we will conveniently use the following representation of a polynomial $f(\boldsymbol{x})$ of degree ℓ:

$$f(\boldsymbol{x}) = \textstyle\sum_{\boldsymbol{\alpha}:|\boldsymbol{\alpha}|\leq\ell} \bar{f}_{\boldsymbol{\alpha}}\, \boldsymbol{x}^{\boldsymbol{\alpha}}, \tag{B.18}$$

where we simply observed that the polynomial is the sum of monomials $\boldsymbol{x}^{\boldsymbol{\alpha}}$ of degree $|\boldsymbol{\alpha}|\leq \ell$, and with suitable coefficients $\bar{f}_{\boldsymbol{\alpha}}$, again indexed by $\boldsymbol{\alpha}$.

We are now ready to show that both objective and constraints (B.11) can be rewritten in a way that leads back to the moment relaxation in Appendix A.

Rewriting the objective (B.11)*:* To simplify the objective $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[p(\boldsymbol{x})\right]$, we note that the pseudo-expectation is a linear operator, hence:

$$\begin{aligned}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[p(\boldsymbol{x})\right] &\overbrace{=}^{\text{using (B.18)}} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\sum_{\boldsymbol{\alpha}:|\boldsymbol{\alpha}|\leq\ell} \bar{p}_{\boldsymbol{\alpha}}\, \boldsymbol{x}^{\boldsymbol{\alpha}}\right] \overbrace{=}^{\text{using Fact B.2}} \sum_{\boldsymbol{\alpha}:|\boldsymbol{\alpha}|\leq\ell} \bar{p}_{\boldsymbol{\alpha}}\, \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\boldsymbol{x}^{\boldsymbol{\alpha}}\right] \\ &\overbrace{=}^{\text{for a suitable matrix } \boldsymbol{C}_1} \left\langle \boldsymbol{C}_1, \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\boldsymbol{X}_\ell\right] \right\rangle,\end{aligned} \tag{B.19}$$

which indeed produces the same structure as the objective of the moment relaxation in (A.1) with $\ell = 2r$; as we will see in a while, $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\boldsymbol{X}_\ell\right]$ will become the main matrix variable in the optimization.

Rewriting the equality constraints (B.13)*:* To simplify the constraint $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[h_i(\boldsymbol{x})q(\boldsymbol{x})\right] = 0$ (which has to hold for polynomials q of degree

[1]For instance, the product between the monomial $x_1 x_2^3 x_4^5$ (namely, $\boldsymbol{x}^{\boldsymbol{\alpha}}$ with $\boldsymbol{\alpha} = [1\,;\,3\,;\,0\,;\,5]$) and the monomial $x_1^2 x_2 x_3$ (namely, $\boldsymbol{x}^{\boldsymbol{\beta}}$ with $\boldsymbol{\beta} = [2\,;\,1\,;\,1\,;\,0]$) is $(x_1 x_2^3 x_4^5) \cdot (x_1^2 x_2 x_3) = x_1^3 x_2^4 x_3 x_4^5$, which corresponds to the exponent vector $[3\,;\,4\,;\,1\,;\,5]$.

$\deg(h_i \cdot q) \le \ell$) we note that it suffices to require $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[h_i(\boldsymbol{x})\boldsymbol{x}^{\boldsymbol{\beta}}\right] = 0$ for $|\boldsymbol{\beta}| \le \ell - \deg(h_i)$; this follows from the fact that any polynomial is a sum of monomials and the pseudo-expectation is a linear function. Let us now manipulate $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[h_i(\boldsymbol{x})\boldsymbol{x}^{\boldsymbol{\beta}}\right] = 0$ as follows:

$$
\begin{aligned}
\tilde{\mathbb{E}}_{\tilde{\mu}}\left[h_i(\boldsymbol{x})\boldsymbol{x}^{\boldsymbol{\beta}}\right] = 0 \quad &\overbrace{\iff}^{\text{using (B.18) for } h_i(\boldsymbol{x})} \quad \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\sum_{\boldsymbol{\alpha}} \bar{h}_{i,\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\beta}}\right] = 0 \\
\overbrace{\iff}^{\text{using (B.17)}} \quad \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\sum_{\boldsymbol{\alpha}} \bar{h}_{i,\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}\right] = 0 \quad &\overbrace{\iff}^{\text{using Fact B.2}} \quad \sum_{\boldsymbol{\alpha}} \bar{h}_{i,\boldsymbol{\alpha}} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}\right] = 0 \\
&\overbrace{\iff}^{\text{for a suitable matrix } \boldsymbol{A}_{i,\boldsymbol{\beta}}} \quad \left\langle \boldsymbol{A}_{i,\boldsymbol{\beta}}, \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\boldsymbol{X}_\ell\right]\right\rangle = 0,
\end{aligned}
\tag{B.20}
$$

which has to be imposed for each $\boldsymbol{\beta}$ such that $|\boldsymbol{\beta}| \le \ell - \deg(h_i)$. Note that the constraints in (B.20) capture both the equality constraints in (A.2) (for $|\boldsymbol{\beta}| = 0$) as well as the redundant constraints (A.5) (for $0 < |\boldsymbol{\beta}| \le \ell - \deg(h_i)$).

Rewriting the inequality constraints in (B.36)*:* We simplify the constraint $\tilde{\mathbb{E}}_{\mu}\left[\prod_{j\in\mathcal{S}} g_j(\boldsymbol{x}) \cdot s(\boldsymbol{x})^2\right] \ge 0$, which has to hold for all $\mathcal{S} \subseteq [l_g]$ and for all $s \in \mathbb{R}[\boldsymbol{x}]$ such that $\deg\left(\prod_{j\in\mathcal{S}} g_j \cdot s^2\right) \le \ell$. Towards this goal, we use the representation (B.18) for $s(\boldsymbol{x})$ and write $s(\boldsymbol{x}) = \sum_{\boldsymbol{\beta}:|\boldsymbol{\beta}|\le t} \bar{s}_{\boldsymbol{\beta}} \boldsymbol{x}^{\boldsymbol{\beta}}$, where $t \triangleq \left\lfloor \frac{\ell - \deg(\prod_{j\in\mathcal{S}} g_j)}{2} \right\rfloor$. Therefore, we obtain:

$$
\tilde{\mathbb{E}}_{\mu}\left[g_j(\boldsymbol{x}) \cdot s(\boldsymbol{x})^2\right] \ge 0 \overbrace{\iff}^{\text{expanding } s^2} \tilde{\mathbb{E}}_{\mu}\left[g_j(\boldsymbol{x}) \cdot \sum_{\boldsymbol{\alpha}:|\boldsymbol{\alpha}|\le t} \bar{s}_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}:|\boldsymbol{\beta}|\le t} \bar{s}_{\boldsymbol{\beta}} \boldsymbol{x}^{\boldsymbol{\beta}}\right] \ge 0
\tag{B.21}
$$

$$
\overbrace{\iff}^{\text{rearranging}} \tilde{\mathbb{E}}_{\mu}\left[\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}:|\boldsymbol{\alpha}|,|\boldsymbol{\beta}|\le t} \bar{s}_{\boldsymbol{\alpha}} \bar{s}_{\boldsymbol{\beta}} \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_j(\boldsymbol{x})\right] \ge 0
\tag{B.22}
$$

$$
\overbrace{\iff}^{\text{using (B.18) on } \prod_{\mathcal{S}\subseteq[l_g]} g_j(\boldsymbol{x})} \tilde{\mathbb{E}}_{\mu}\left[\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}:|\boldsymbol{\alpha}|,|\boldsymbol{\beta}|\le t} \bar{s}_{\boldsymbol{\alpha}} \bar{s}_{\boldsymbol{\beta}} \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \sum_{\boldsymbol{\gamma}:|\boldsymbol{\gamma}|\le\deg\left(\prod_{j\in\mathcal{S}} g_j(\boldsymbol{x})\right)} \bar{g}_{\mathcal{S},\boldsymbol{\gamma}} \boldsymbol{x}^{\boldsymbol{\gamma}}\right] \ge 0
\tag{B.23}
$$

$$\overset{\text{rearranging}}{\Longleftrightarrow} \tilde{\mathbb{E}}_\mu \left[\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}:|\boldsymbol{\alpha}|,|\boldsymbol{\beta}|\leq t} \bar{s}_{\boldsymbol{\alpha}} \bar{s}_{\boldsymbol{\beta}} \sum_{\boldsymbol{\gamma}:|\boldsymbol{\gamma}|\leq \deg\left(\prod_{j\in\mathcal{S}} g_j(\boldsymbol{x})\right)} \bar{g}_{\mathcal{S},\boldsymbol{\gamma}} \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} \right] \geq 0 \tag{B.24}$$

$$\overset{\text{using Fact B.2}}{\Longleftrightarrow} \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}:|\boldsymbol{\alpha}|,|\boldsymbol{\beta}|\leq t} \bar{s}_{\boldsymbol{\alpha}} \bar{s}_{\boldsymbol{\beta}} \sum_{\boldsymbol{\gamma}:|\boldsymbol{\gamma}|\leq \deg\left(\prod_{j\in\mathcal{S}} g_j(\boldsymbol{x})\right)} \bar{g}_{\mathcal{S},\boldsymbol{\gamma}} \tilde{\mathbb{E}}_\mu \left[\boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}\right] \geq 0. \tag{B.25}$$

Now note that $|\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}|\leq \ell$ by construction, and hence we can write, for a given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, each $\sum_{\gamma:|\gamma|\leq \deg\left(\prod_{j\in\mathcal{S}} g_j(\boldsymbol{x})\right)} \bar{g}_{\mathcal{S},\gamma} \tilde{\mathbb{E}}_\mu \left[\boldsymbol{x}^{\boldsymbol{\alpha}+\beta+\gamma}\right]$ as a linear function of $\tilde{\mathbb{E}}_{\tilde{\mu}} [\boldsymbol{X}_\ell]$. Moreover, since $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are such that $|\boldsymbol{\alpha}|, |\boldsymbol{\beta}|\leq t$, we can group these entries into an $t\times t$ matrix $\boldsymbol{X}_{\mathcal{S}}$, which is such that:

$$[\boldsymbol{X}_{\mathcal{S}}]_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \left\langle \boldsymbol{A}_{\text{loc},\mathcal{S},\boldsymbol{\alpha}\boldsymbol{\beta}}, \tilde{\mathbb{E}}_{\tilde{\mu}} [\boldsymbol{X}_\ell] \right\rangle, \tag{B.26}$$

for some suitable matrix $\boldsymbol{A}_{\text{loc},\mathcal{S},\boldsymbol{\alpha}\boldsymbol{\beta}}$, such that $\left\langle \boldsymbol{A}_{\text{loc},\mathcal{S},\boldsymbol{\alpha}\boldsymbol{\beta}}, \tilde{\mathbb{E}}_{\tilde{\mu}} [\boldsymbol{X}_\ell] \right\rangle = \sum_{\gamma:|\gamma|\leq \deg\left(\prod_{j\in\mathcal{S}} g_j\right)} \bar{g}_{i,\gamma} \tilde{\mathbb{E}}_\mu \left[\boldsymbol{x}^{\boldsymbol{\alpha}+\beta+\gamma}\right]$. Using the matrix $\boldsymbol{X}_{\mathcal{S}}$ and defining a vector $\bar{\boldsymbol{s}} \in \mathbb{R}^t$ with entries $\bar{s}_{\boldsymbol{\alpha}}$ for $|\boldsymbol{\alpha}|\leq t$, we rewrite (B.25) as:

$$\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}:|\boldsymbol{\alpha}|,|\boldsymbol{\beta}|\leq t} \bar{s}_{\boldsymbol{\alpha}} \bar{s}_{\boldsymbol{\beta}} [\boldsymbol{X}_{\mathcal{S}}]_{\boldsymbol{\alpha},\boldsymbol{\beta}} \geq 0 \iff \bar{\boldsymbol{s}}^{\mathsf{T}} \boldsymbol{X}_{\mathcal{S}} \bar{\boldsymbol{s}} \geq 0. \tag{B.27}$$

Since this has to hold for any $\bar{\boldsymbol{s}}$ (*i.e.*, any polynomial $s(\boldsymbol{x})$ of appropriate degree), we conclude the constraint above is equivalent to:

$$\boldsymbol{X}_{\mathcal{S}} \succeq 0. \tag{B.28}$$

Now we can easily see that (B.26) and (B.28) match the localizing constraints we wrote in (A.7).

Rewriting (B.12)*:* Finally, the constraint (B.12) imposes that $\tilde{\mu}$ must be a level ℓ pseudo-distribution. However, we know from Theorem B.1 that $\tilde{\mu}$ is a level-ℓ pseudo-distribution if and only if the pseudo-moment matrix $\tilde{\mathbb{E}}_{\tilde{\mu}} \left[[\boldsymbol{x}]_{\ell/2}[\boldsymbol{x}]_{\ell/2}^{\mathsf{T}}\right]$ is positive semidefinite and $\tilde{\mathbb{E}}_{\tilde{\mu}} [1] = 1$. Therefore, we can reparametrize the objective (B.19) and constraints (B.20), (B.26), (B.28) with a matrix variable (in place of $\tilde{\mathbb{E}}_{\tilde{\mu}} [\boldsymbol{X}_\ell]$) that is constrained to be positive semidefinite and to have the top-left entry equal to 1 (*cf.* (B.15)). This yields back the relaxation described in Appendix A as expected.

Constrained pseudo-distributions: A practical view So far we have shown that taking suitable pseudo-expectations over the objective and constraints in a polynomial optimization problem leads to a convex relaxation, known as the moment relaxation. Now we want to shed some light on Definition B.2 by showing that the condition (B.2) is indeed the same as the inequality constrains in (B.36) and hence admits the same transcription as an SDP.

Towards this goal, let us consider the following *feasibility* POP:

$$\begin{aligned} \text{find} \quad & \boldsymbol{x} \in \mathbb{R}^{d_x} && \text{(B.29)} \\ \text{subject to} \quad & h_i(\boldsymbol{x}) = 0, i = 1, \ldots, l_h && \text{(B.30)} \\ & g_j(\boldsymbol{x}) \geq 0, j = 1, \ldots, l_g. && \text{(B.31)} \end{aligned}$$

This is similar to (POP), with the exception that we are looking for a feasible solution rather than optimizing a cost function. Now note that we can write a polynomial equality $h_i(\boldsymbol{x}) = 0$ as two inequality constraints $h_i(\boldsymbol{x}) \leq 0$ and $-h_i(\boldsymbol{x}) \leq 0$. Hence, without loss of generality we rewrite (B.29) as:

$$\begin{aligned} \text{find} \quad & \boldsymbol{x} \in \mathbb{R}^{d_x} && \text{(B.32)} \\ \text{subject to} \quad & f_j(\boldsymbol{x}) \geq 0, \quad j = 1, \ldots, m, && \text{(B.33)} \end{aligned}$$

for suitable polynomials f_i, $i = 1, \ldots, m$. Similarly to what we did earlier in this section, we relax (B.32) by using pseudo-expectations:

$$\begin{aligned} \text{find} \quad & \tilde{\mu} && \text{(B.34)} \\ \text{subject to} \quad & \tilde{\mu} \text{ is a level-}\ell \text{ pseudo-distribution} && \text{(B.35)} \\ & \tilde{\mathbb{E}}_{\mu}\left[s(\boldsymbol{x})^2 \cdot \textstyle\prod_{i \in \mathcal{S}} f_i(\boldsymbol{x})\right] \geq 0, && \text{(B.36)} \\ & \text{for every set } \mathcal{S} \in [m] \text{ and every } s \in \mathbb{R}[\boldsymbol{x}] \text{ such that } \deg\left(s^2 \cdot \textstyle\prod_{i \in \mathcal{S}} f_i\right) \leq \ell. \end{aligned}$$

First of all, we note that (B.36) matches the definition of constrained pseudo-distribution in Definition B.2 for $k = 0$. Moreover, following the same derivation as above, we can easily show that (i) (B.34) can be transcribed as a standard SDP, and (ii) every pseudo-distribution solving Lasserre's relaxation of a (POP) satisfies the set of constraints in the (POP) in the sense of Definition B.2.

Now we note that Definition B.2 allows some extra slack through the parameter k, *i.e.*, $\tilde{\mu}$ satisfies $\mathcal{A}$ at degree k, if every set $\mathcal{S} \subset$

$[m]$ and every sum-of-squares polynomial h on $\mathbb{R}^{d_x}$ with $\deg\left(s^2\right) + \sum_{i\in\mathcal{S}} \max\{\deg\left(f_i\right), k\} \leq \ell$ satisfies $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[s^2 \cdot \prod_{i\in\mathcal{S}} f_i\right] \geq 0$. This essentially means that the inequality $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[s^2 \cdot \prod_{i\in\mathcal{S}} f_i\right] \geq 0$ is enforced for a smaller number of subsets $\mathcal{S}$.

C

Sum-of-Squares Proofs

Sum-of-squares proofs provide an advanced way to reason about polynomial constraints and to infer properties of pseudo-distributions, or, equivalently, properties of the moment relaxation in Appendix A. The presentation in this section builds on [9], but also collects inference rules from other papers, which we cite as we present the results.

Let us denote with $f(\boldsymbol{x})$ a polynomial in variables $\boldsymbol{x} = [x_1; x_2; \ldots; x_{d_x}]$ and let $\mathcal{A} = \{f_1(\boldsymbol{x}) \geq 0, \ldots, f_m(\boldsymbol{x}) \geq 0\}$ be a system of polynomial constraints over $\mathbb{R}^{d_x}$. In the following, we omit the argument when clear from the context and write f instead of $f(\boldsymbol{x})$. A polynomial p is sum-of-squares (sos) if there exist polynomials $q_1, \ldots, q_t$ such that $p = q_1^2 + \ldots + q_t^2$.

The key idea is to relate two sets of polynomial constraints using a "sum-of-squares proof" (the definition below is the same as Definition 4.1 in the main monograph).

Definition C.1 (Sum-of-squares proof)**.** Given a system of polynomial constraint $\mathcal{A}$ and a polynomial g, a sum-of-square (sos) proof that the system $\mathcal{A}$ implies $g \geq 0$ consists of sum-of-squares polynomials $\{p_{\mathcal{S}}\}_{\mathcal{S} \subseteq [m]}$ such that:

$$g = \sum_{\mathcal{S} \subseteq [m]} p_{\mathcal{S}} \cdot \prod_{i \in \mathcal{S}} f_i. \tag{C.1}$$

We say that the proof has degree k if for every $\mathcal{S} \subseteq [m]$, $\deg\left(p_{\mathcal{S}} \cdot \prod_{i \in \mathcal{S}} f_i\right) \leq k$ where $\deg(\cdot)$ denotes the degree of a polynomial. We use the notation:

$$\mathcal{A}(\boldsymbol{x}) \,\Big|\frac{k}{\boldsymbol{x}}\, \{g(\boldsymbol{x}) \geq 0\} \qquad \text{or} \qquad \{f_i(\boldsymbol{x}) \geq 0, \ldots, f_m(\boldsymbol{x}) \geq 0\} \,\Big|\frac{k}{\boldsymbol{x}}\, \{g(\boldsymbol{x}) \geq 0\} \tag{C.2}$$

to denote that there is a proof of degree at most k of the fact that $\mathcal{A} = \{f_i(\boldsymbol{x}) \geq 0, \ldots, f_m(\boldsymbol{x}) \geq 0\}$ implies $g \geq 0$ (*i.e.*, any $\boldsymbol{x}$ that satisfies $\mathcal{A}(\boldsymbol{x})$ is such that $g(\boldsymbol{x}) \geq 0$). We omit the variables and write $\mathcal{A}(\boldsymbol{x}) \,\Big|\!\frac{k}{}\, \{g(\boldsymbol{x}) \geq 0\}$, when they are clear from the context. Moreover, we write

$$\Big|\frac{k}{\boldsymbol{x}}\, \{g(\boldsymbol{x}) \geq 0\} \tag{C.3}$$

if there is a sum-of-squares proof that $g(\boldsymbol{x}) \geq 0$ for any $\boldsymbol{x} \in \mathbb{R}^{d_x}$ (*i.e.*, $g(\boldsymbol{x})$ is sum-of-squares).

From eq. (C.1), it is clear why the polynomials $p_{\mathcal{S}}$ are a "proof" of $g \geq 0$ for any $\boldsymbol{x}$ satisfying $\mathcal{A}$: for any $\boldsymbol{x} \in \mathcal{A}$, $\prod_{i \in \mathcal{S}} f_i \geq 0$ by definition, hence if we can write g as the product of a sum-of-squares (hence non-negative) polynomial and $\prod_{i \in \mathcal{S}} f_i$, we automatically prove that $g \geq 0$ whenever $\boldsymbol{x} \in \mathcal{A}$.

Sum-of-squares proofs allow us to deduce properties of pseudo-distributions: in particular, if we have an sos proof relating two sets of constraints, we can conclude that any pseudo-distribution satisfying a set of constraints, must also satisfy the other. This is formalized below.

Fact C.1 (Soundness [70]). Consider a level-ℓ pseudo-distribution $\tilde{\mu}$ such that $\tilde{\mu} \,\Big|\!\!=\!\frac{k}{}\, \mathcal{A}$. If there exists a sum-of-squares proof that $\mathcal{A} \,\Big|\!\frac{k'}{}\, \mathcal{B}$, then $\tilde{\mu} \,\Big|\!\!=\!\frac{k \cdot k' + k'}{}\, \mathcal{B}$.

If the pseudo-distribution $\tilde{\mu}$ satisfies $\mathcal{A}$ only approximately, soundness continues to hold but we require an upper bound on the bit-complexity of the sum-of-squares proof $\mathcal{A} \,\Big|\!\frac{k'}{}\, \mathcal{B}$ (*i.e.*, the number of bits to write down the proof). In this monograph, we mostly disregard bit-complexity issues and refer the reader to [55, §3] for a more formal discussion. In other words, similarly to [70], [71], we assume that all numbers appearing in the input have bit complexity $d_x^{O(1)}$ and all sos proofs will have bit complexity $d_x^{O(\ell)}$, which is enough to claim soundness for the sos proof system.

Not only sos proofs allow us to infer properties of pseudo-distributions, but also the reverse is true. The following fact states that every property of a pseudo-distribution can be derived via a sum-of-squares proof.

Fact C.2 (Completeness [70]). Suppose $\ell \geq k' \geq k$ and $\mathcal{A}$ is a system of explicitly bounded polynomial constraints with degree at most k (*i.e.*, $\mathcal{A} \vdash \{\|\boldsymbol{x}\|_2^2 \leq M_x^2\}$ for some finite M_x). Let $\{g \geq 0\}$ be a polynomial constraint. If every level-ℓ pseudo-distribution that satisfies $\tilde{\mu} \overset{k}{\models} \mathcal{A}$ also satisfies $\tilde{\mu} \overset{k'}{\models} \mathcal{B}$, then for every $\epsilon > 0$ there is a sum-of-squares proof $\mathcal{A} \overset{\ell}{\vdash} \{g \geq -\epsilon\}$.

Sos rules. Sum-of-squares provide a proof system to reasons about polynomial constraints. For instance, if we have an sos proof that $\mathcal{A}$ implies $g \geq 0$, we may want to use such a proof system to infer if another implication also holds true, say $\mathcal{A}$ implies $g' \geq 0$ (for some other polynomial g'). Reasoning in this proof system is not immediate. For instance, the fact that $p(\boldsymbol{x}) \geq 0$ for some degree-k polynomial does not necessarily imply that there is a sum-of-squares proof $\vdash^{k}_{\boldsymbol{x}} \{p(\boldsymbol{x}) \geq 0\}$. Similarly, for some polynomial constraints $\mathcal{A}$ and polynomials $g(\boldsymbol{x})$ and $g(\boldsymbol{x})'$ with $g'(\boldsymbol{x}) \geq g(\boldsymbol{x})$ for every $\boldsymbol{x} \in \mathbb{R}^{d_x}$, the fact that $\mathcal{A} \vdash^{k}_{\boldsymbol{x}} \{g(\boldsymbol{x}) \geq 0\}$ does not necessarily imply that $\mathcal{A} \vdash^{k}_{\boldsymbol{x}} \{g'(\boldsymbol{x}) \geq 0\}$, since the latter fact might not admit a sum-of-squares proof. In this sense, the sos proof system is more restrictive than the typical algebraic manipulation we are used to. Fortunately, previous work provides a toolkit of inference rules that can be used to correctly reason over sos proofs. We collect key facts below, mostly drawing from [43], [62], [70], [71], [86].

Fact C.3 (Inference Rules [70]). The following inference rules hold for systems of polynomial constraints $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and polynomials $f, g : \mathbb{R}^{d_x} \mapsto \mathbb{R}$:

$$\text{addition:} \quad \frac{\mathcal{A} \vdash^{k} \{f \geq 0, g \geq 0\}}{\mathcal{A} \vdash^{k} \{f + g \geq 0\}} \tag{C.4}$$

$$\text{multiplication:} \quad \frac{\mathcal{A} \vdash^{k} \{f \geq 0\} \,,\; \mathcal{A} \vdash^{k'} \{g \geq 0\}}{\mathcal{A} \vdash^{k+k'} \{f \cdot g \geq 0\}} \tag{C.5}$$

$$\text{transitivity:} \quad \frac{\mathcal{A} \vdash^{k} \mathcal{B} \,,\; \mathcal{B} \vdash^{k'} \mathcal{C}}{\mathcal{A} \vdash^{k \cdot k'} \mathcal{C}} \tag{C.6}$$

where, for two logical statements A and B, we use the standard inference notation $\frac{\mathsf{A}}{\mathsf{B}}$ to denote that if A is true, then B must be true.

Fact C.4 (Basics, p. 59 in [17] and p. 70 in [55])**.** Let $p(\boldsymbol{x})$ be a degree-k polynomial such that $p(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^{d_x}$. Then:

$$\vdash^{k}_{\boldsymbol{x}} \{p(\boldsymbol{x}) \geq 0\} \tag{C.7}$$

(*i.e.*, $p(\boldsymbol{x})$ is sos) if:

- $d_x = 1$ (univariate case),
- $k = 2$ (quadratic polynomials), or
- $d_x = 2$ and $k = 4$ (bivariate, quartic polynomials).

Moreover, (C.7) holds whenever p is a function over the Boolean hypercube $p : \{0, 1\}^{d_x} \mapsto \mathbb{R}$.

Fact C.5 (Univariate polynomials over interval, Fact 3.7 in [70])**.** For any univariate degree k polynomial $p(x) \geq 0$ for $x \in [a, b]$,

$$\{x \geq a, x \leq b\} \vdash^{k}_{x} \{p(\boldsymbol{x}) \geq 0\} \,. \tag{C.8}$$

Fact C.6 (Sos generalized triangle inequality, Fact 4.8 in [71])**.** For any $a_1, a_2, \ldots, a_m$

$$\vdash^{k}_{a_1, a_2, \ldots, a_m} \left\{ \left(\sum_{i=1}^{m} a_i \right)^{k} \leq m^k \left(\sum_{i=1}^{m} a_i^k \right) \right\} . \tag{C.9}$$

Fact C.7 (Sos triangle inequality (same as Fact C.6 with $m = 2$ and $k = 2$))**.** For any a_1, a_2

$$\vdash^{2}_{a,b} \left\{ (a + b)^2 \leq 2^2 a^2 + 2^2 b^2 \right\} . \tag{C.10}$$

Fact C.8 (Sos triangle inequality 2.0, p. 18 in [71]). For any indeterminates a, b, scalar δ, and even integer k:

$$\Big|\frac{k}{a,b}\ \delta^k a^k \leq (2\delta)^k (a-b)^k + (2\delta)^k b^k. \tag{C.11}$$

Fact C.9 (Sos squaring). Let f, g be sos polynomials of degree at most k and $\mathcal{A} = \{f_1(\boldsymbol{x}) \geq 0, \ldots, f_m(\boldsymbol{x}) \geq 0\}$ be a system of polynomial inequalities. If $\mathcal{A}\Big|\frac{k'}{\boldsymbol{x}}\ \{f \geq g\}$, then $\mathcal{A}\Big|\frac{k'+k}{\boldsymbol{x}}\ \{f^2 \geq g^2\}$.

Proof. The assumption $\mathcal{A}\Big|\frac{k}{\boldsymbol{x}}\ f = 0$ implies that:

$$f - g = \sum_{\mathcal{S} \subseteq [m]} p_{\mathcal{S}} \cdot \prod_{i \in \mathcal{S}} f_i. \tag{C.12}$$

Now note that $f^2 - g^2 = (f-g)(f+g)$ hence:

$$f^2 - g^2 = (f-g)(f+g) = (f+g) \sum_{\mathcal{S} \subseteq [m]} p_{\mathcal{S}} \cdot \prod_{i \in \mathcal{S}} f_i. \tag{C.13}$$

Since f, g are sos polynomial, the previous relation proves $\mathcal{A}\Big|\frac{k'+k}{\boldsymbol{x}}\ \{f^2 \geq g^2\}$ with sos proof $(f+g) \cdot p_{\mathcal{S}}$ and by noting that the maximum degree appearing in (C.13) is $k' + k$. ■

Fact C.10 (Sos triangle inequality with norms, Fact A.2 in [62]). Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be n-length vectors of indeterminates. Then:

$$\Big|\frac{2}{\boldsymbol{x}_1, \boldsymbol{x}_2}\ \left\{ \|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_2^2 \leq 2\,\|\boldsymbol{x}_1\|_2^2 + 2\,\|\boldsymbol{x}_2\|_2^2 \right\}. \tag{C.14}$$

Fact C.11 (Sos generalized triangle inequality with norms). Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be n-length vectors of indeterminates and $k \in \mathbb{N}$ be even. Then:

$$\Big|\frac{k}{\boldsymbol{x}_1, \boldsymbol{x}_2}\ \left\{ \|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_2^k \leq 2^k\,\|\boldsymbol{x}_1\|_2^k + 2^k\,\|\boldsymbol{x}_2\|_2^k \right\}. \tag{C.15}$$

Proof.

$$\Big|\frac{k}{\boldsymbol{x}_1, \boldsymbol{x}_2}\ \|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_2^k = \left(\|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_2^2\right)^{\frac{k}{2}} \overbrace{\leq}^{using\ (C.14)} \left(2\,\|\boldsymbol{x}_1\|_2^2 + 2\,\|\boldsymbol{x}_2\|_2^2\right)^{\frac{k}{2}} \tag{C.16}$$

$$\overbrace{\leq}^{using\ (C.9)} 2^{\frac{k}{2}}\left(2\,\|\boldsymbol{x}_1\|_2^2\right)^{\frac{k}{2}} + 2^{\frac{k}{2}}\left(2\,\|\boldsymbol{x}_2\|_2\right)^{\frac{k}{2}} = 2^k\,\|\boldsymbol{x}_1\|_2^k + 2^k\,\|\boldsymbol{x}_2\|_2^k. \tag{C.17}$$

■

Fact C.12 (Sos Cauchy-Schwarz, Fact A.1 in [62])**.** Let $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ be polynomials in some indeterminates. Then:

$$\Big|\frac{4}{x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_n} \left\{ \left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) \right\}, \tag{C.18}$$

or, written in vector form, for two n-length vectors $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\Big|\frac{4}{\boldsymbol{x},\boldsymbol{y}} \left\{ \left(\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}\right)^2 \leq \left(\|\boldsymbol{x}\|_2^2\right)\left(\|\boldsymbol{y}\|_2^2\right) \right\}. \tag{C.19}$$

Fact C.13 (Sos Hölder's inequality, Fact 4.4 in [71])**.** Let f_i, g_i for $1 \leq i \leq n$ be sos polynomials. Let p, q be integers such that $\frac{1}{p} + \frac{1}{q} = 1$. Then:

$$\Big|\frac{pq}{} \left\{ \left(\frac{1}{n}\sum_{i=1}^{n} f_i g_i\right)^{pq} \leq \left(\frac{1}{n}\sum_{i=1}^{n} f_i^p\right)^q \left(\frac{1}{n}\sum_{i=1}^{n} g_i^q\right)^p \right\}. \tag{C.20}$$

Fact C.14 (Sos Hölder's inequality 2.0, Fact A.6 in [62])**.** Let $\omega_1, \ldots, \omega_n$ and $x_1, \ldots, x_n$ be indeterminates. Let $q \in \mathbb{N}$ be a power of 2. Then:

$$\left\{\omega_i^2 = \omega_i, \forall i \in [n]\right\} \Big|\frac{O(q)}{\omega_1,\ldots,\omega_n,x_1,\ldots,x_n} \left\{ \left(\sum_{i=1}^{n} \omega_i x_i\right)^q \leq \left(\sum_{i=1}^{n} \omega_i^2\right)^{q-1} \left(\sum_{i=1}^{n} x_i^q\right) \right\}, \tag{C.21}$$

and

$$\left\{\omega_i^2 = \omega_i, \forall i \in [n]\right\} \Big|\frac{O(q)}{\omega_1,\ldots,\omega_n,x_1,\ldots,x_n} \left\{ \left(\sum_{i=1}^{n} \omega_i x_i\right)^q \leq \left(\sum_{i=1}^{n} \omega_i^2\right)^{q-1} \left(\sum_{i=1}^{n} \omega_i x_i^q\right) \right\}. \tag{C.22}$$

Fact C.15 (Sos Hölder's inequality 3.0, Fact A.3 in [43])**.** Let f_i, g_i for $1 \leq i \leq n$ be indeterminates. Then:

$$\Big|\frac{2}{} \left\{ \left(\frac{1}{n}\sum_{i=1}^{n} f_i g_i\right)^2 \leq \left(\frac{1}{n}\sum_{i=1}^{n} f_i^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} g_i^2\right) \right\}. \tag{C.23}$$

Fact C.16 (Lemma A.3 in [86])**.** Let x be indeterminate and a be a positive real number. Then:

$$\{x^2 \leq a^2\} \vdash \{x \leq a, x \geq -a\}. \tag{C.24}$$

Fact C.17 (Lemma A.2 in [86]). Let $\boldsymbol{x}$ be indeterminate and $\boldsymbol{a}$ be a unit vector. Let $\mathcal{A}\{\|\boldsymbol{x}\|^2 = 1, (\boldsymbol{x}^\top \boldsymbol{a})^2 \le \tau\}$. Then, for any $\boldsymbol{b}$ such that $\|\boldsymbol{a} - \boldsymbol{b}\|_2 \le 2\delta$, we have:

$$\mathcal{A} \vdash \left\{ (\boldsymbol{x}^\top \boldsymbol{b})^2 \le (\sqrt{\tau} + \sqrt{\delta})^2 \right\}. \tag{C.25}$$

We conclude with a self-evident fact that reassures us that certain manipulations of polynomials are easy to reason over, even in the sos proof system.

Fact C.18 (Equalities). Let f, g be polynomials and $\mathcal{A}$ be a system of polynomial inequalities. If $f = g$ and $\mathcal{A} \vdash_{\boldsymbol{x}}^{k} f = 0$, then $\mathcal{A} \vdash_{\boldsymbol{x}}^{k} g = 0$.

D

Sparse LIst-Decodable Estimation (SLIDE)

In this appendix, we present a variant of Algorithm 5 that empirically returns an accurate list of estimates, as shown in Section 8.5. We start by stating the algorithm, whose pseudocode is given in Algorithm 6.

The proposed algorithm is very close to Algorithm 5 (and is still based on the key insights from [70]), but includes three small but important changes. First of all, instead of solving a moment relaxation of order $r = 2$, which is still expensive for large n,[1] we develop a sparse relaxation (line 1). The sparse relaxation uses the following sparse monomial basis

$$\boldsymbol{m}(\boldsymbol{\omega}, \boldsymbol{x}) \triangleq [1\,;\, \omega_1\,;\, \ldots\,;\, \omega_n\,;\, \boldsymbol{x}\,;\, \omega_1\boldsymbol{x}\,;\, \ldots\,;\, \omega_n\boldsymbol{x}], \tag{D.1}$$

which neglects other degree-2 monomials (*e.g.*, $\omega_i \cdot \omega_j$) that do not appear in problem (LDR) while still giving access to the pseudo-expectations used in Algorithm 5. Note that the sparse relaxation leads to SDPs of more manageable size $(n+1)(d_x+1)$.[2] Note that the idea of using

[1] An order-2 moment relaxation of (LDR) entails solving an SDP with a matrix of size $\binom{n+d_x+2}{2}$, which is already as large as 1830 for $d_x = 9$ and $n = 50$, which is the typical setup considered in our experiments. In our tests, MOSEK [88] runs out of memory when fed an SDP of size larger than 1000.

[2] For instance, when $d_x = 9$ and $n = 50$, the sparse relaxation leads to a more compact SDP with a moment matrix of size 510×510.

Algorithm 6: Sparse LIst-Decodable Estimation (SLIDE).

Input: input data $(\boldsymbol{y}_i, \boldsymbol{A}_i)$, $i \in [n]$, inlier rate α.
Output: list of estimates of $\boldsymbol{x}^\circ$.

`/* Algorithm solves a relaxation of the following problem: */`

`/*`
$$\min_{\boldsymbol{\omega},\boldsymbol{x}} \|\boldsymbol{\omega}\|_2^2, \quad \texttt{s.t.} \quad \mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}} \triangleq \left\{ \begin{array}{c} \omega_i^2 = \omega_i, \quad i = [n] \\ \sum_{i=1}^n \omega_i = \alpha n \\ \omega_i \cdot \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}\right\|_2^2 \le \bar{c}^2, \quad i = [n] \\ \boldsymbol{x} \in \mathbb{X} \end{array} \right\} \tag{LDR}$$

`/* Compute matrix` $\boldsymbol{X}^\star$ `by solving SDP from sparse moment relaxation */`

1 $\boldsymbol{X}^\star = \text{solve\_sparse\_moment\_relaxation\_at\_order\_2}\,(\text{LDR})$

`/* Compute list of estimates */`

2 create empty list $\mathcal{L} = \emptyset$

3 **for** $i \in [n]$ **do**

4 $\quad \boldsymbol{v}_i = \begin{cases} \dfrac{\boldsymbol{X}^\star_{[\omega_i \boldsymbol{x}]}}{\boldsymbol{X}^\star_{[\omega_i]}} & \text{if } \boldsymbol{X}^\star_{[\omega_i]} > 0 \\ \boldsymbol{0} & \text{otherwise} \end{cases}$

5 $\quad \boldsymbol{x}_i = \text{project\_to\_}\mathbb{X}(\boldsymbol{v}_i)$

6 $\quad$ add $\boldsymbol{x}_i$ to $\mathcal{L}$

7 **end**

8 **return** $\mathcal{L}$.

a sparse moment relaxation is not new (see Remark 10 in [129]), but our relaxation is slightly different from [126], [129] and tailored to list-decodable estimation. Later in this section, we provide a derivation of the sparse moment relaxation for the rotation search problem.

The second modification is to round the estimates to the domain $\mathbb{X}$ (line 5). The latter is a consequential change: we empirically noticed that the original approach in Algorithm 5 (with our sparse relaxation) produces estimates with norm close to zero, hence leading to large estimation errors. Projecting the estimates to $\mathbb{X}$ has the effect of re-normalizing the result and correcting scaling problems, enabling the compelling results in Section 8.5. At the end of this section we show that projecting to the domain $\mathbb{X}$ is straightforward in the rotation search problem.

Finally, the third modification with respect to Algorithm 5, is that Algorithm 6 always returns n hypotheses (line 3), rather than sampling; this makes the result deterministic and independent on the choice of number of hypotheses (which can no longer be guided by the guarantees in Theorem 7.1).[3]

Sparse moment relaxation for rotation search. Here we provide an example of sparse relaxation arising when applying SLIDE to the rotation search problem. Let us start by tailoring the polynomial optimization problem (LDR) to rotation search:

$$\min_{\boldsymbol{\omega},\boldsymbol{R}} \|\boldsymbol{\omega}\|_2^2, \quad \text{s.t.} \quad \left\{ \begin{array}{l} \omega_i^2 = \omega_i, \quad i = [n] \\ \sum_{i=1}^n \omega_i = \alpha n \\ \omega_i \cdot \|\boldsymbol{b}_i - \boldsymbol{R}\boldsymbol{a}_i\|^2 \leq \bar{c}^2, \quad i = [n] \\ \boldsymbol{R} \in \mathrm{SO}(3) \end{array} \right\} \tag{D.2}$$

where we substituted the residual errors $\left(\left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\top \boldsymbol{x}\right\|_2^2\right)$ with their expression in the rotation search problem $(\|\boldsymbol{b}_i - \boldsymbol{R}\boldsymbol{a}_i\|^2)$, and where we made explicit that the domain is SO(3). Before presenting the relaxation, we reparametrize (D.2) using unit quaternions: while we could directly relax (D.2) following the approach we describe below, using quaternions has the benefit of (i) leading to an even smaller relaxation (since the quaternion is parametrized by $d_x = 4$ variables instead of 9 variables needed to write a rotation matrix) and (ii) admitting a straightforward projection to the domain $\mathbb{X}$.

Proposition D.1 (Quaternion-based reformulation of (D.2))**.** The polynomial optimization problem (D.2) can be equivalently written as:

$$\min_{\boldsymbol{\omega},\boldsymbol{q}} \|\boldsymbol{\omega}\|_2^2, \quad \text{s.t.} \quad \left\{ \begin{array}{l} \omega_i^2 = \omega_i, \quad i = [n] \\ \sum_{i=1}^n \omega_i = \alpha n \\ \omega_i \cdot \left(\|\boldsymbol{b}_i\|^2 + \|\boldsymbol{a}_i\|^2 - 2\operatorname{tr}\left(\boldsymbol{M}_{ij}^\top \boldsymbol{q}\boldsymbol{q}^\top\right)\right) \leq \bar{c}^2, \quad i = [n] \\ \|\boldsymbol{q}\|^2 = 1 \end{array} \right\} \tag{D.3}$$

[3]Note that we can safely discard the hypotheses corresponding to $\boldsymbol{X}^\star_{[\omega_i]} = 0$ since those are uninformative (*i.e.*, they always correspond to $\boldsymbol{v}_i = \boldsymbol{0}$). We only keep them in Algorithm 6 for the sake of simplicity, such that the output list $\mathcal{L}$ has always size n.

such that any optimal solution of (D.3) can be mapped back to an optimal solution of (D.2) and vice-versa. In (D.3), $\boldsymbol{M}_{ij}$ is a constant matrix whose expression depends on $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$.

Proof. The proof proceeds by inspection, by reparametrizing the rotation $\boldsymbol{R}$ with the corresponding unit quaternion. Each unit quaternion corresponds to a unique rotation, hence we replace the domain $\boldsymbol{R} \in \mathrm{SO}(3)$ with the constraint that the quaternion must have unit norm (*i.e.,* $\|\boldsymbol{q}\|^2 = 1$). Comparing (D.2) and (D.3), we realize we only have to rewrite the maximum-residual inequality constraint in (D.2) in the quaternion-based form in (D.3). This derivation is largely inspired by [126] (which presents a similar reformulation applied to a different polynomial optimization problem), but here we present a simpler proof. We start by observing that the rotation matrix associated to the quaternion $\boldsymbol{q} = [q_1\,;\,q_2\,;\,q_3\,;\,q_4]$ (in our notation, q_4 is the scalar part of the quaternion) is:

$$\boldsymbol{R} = \begin{bmatrix} 2(q_1^2+q_4^2)-1 & 2(q_1q_2-q_3q_4) & 2(q_1q_3+q_2q_4) \\ 2(q_1q_2+q_3q_4) & 2(q_2^2+q_4^2)-1 & 2(q_2q_3-q_1q_4) \\ 2(q_1q_3-q_2q_4) & 2(q_2q_3+q_1q_4) & 2(q_3^2+q_4^2)-1 \end{bmatrix} = \tag{D.4}$$

$$\begin{bmatrix} q_1^2+q_4^2-q_2^2-q_3^2 & 2(q_1q_2-q_3q_4) & 2(q_1q_3+q_2q_4) \\ 2(q_1q_2+q_3q_4) & q_2^2+q_4^2-q_1^2-q_3^2 & 2(q_2q_3-q_1q_4) \\ 2(q_1q_3-q_2q_4) & 2(q_2q_3+q_1q_4) & q_3^2+q_4^2-q_1^2-q_2^2 \end{bmatrix} \tag{D.5}$$

where the expression in (D.5) is obtained by substituting $\|\boldsymbol{q}\|^2 = q_1^2 + q_2^2 + q_3^2 + q_4^2 = 1$ (instead of 1) in the diagonal entries of the expression in (D.4). Now, by inspection from (D.5), we note that:

$$vec(\boldsymbol{R}) = \boldsymbol{P} \cdot vec(\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}) \tag{D.6}$$

where:

$$
\boldsymbol{P} \triangleq \left[\begin{array}{cccc|cccc|cccc|cccc}
1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
\hline
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\
-1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
\hline
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1
\end{array}\right],
$$

$$
vec(\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}) = \left[\begin{array}{c}
q_1^2 \\ q_2 q_1 \\ q_3 q_1 \\ q_4 q_1 \\ \hline
q_2 q_1 \\ q_2^2 \\ q_3 q_2 \\ q_4 q_2 \\ \hline
q_3 q_1 \\ q_3 q_2 \\ q_3^2 \\ q_4 q_3 \\ \hline
q_4 q_1 \\ q_4 q_2 \\ q_4 q_3 \\ q_4^2
\end{array}\right] \tag{D.7}
$$

Equipped with these relations, we are now ready to rewrite the inequality constraint in (D.2) as in (D.3). We develop the squared residual $\|\boldsymbol{b}_i - \boldsymbol{R}\boldsymbol{a}_i\|^2$ in (D.2) as follows:

$$
\|\boldsymbol{b}_i - \boldsymbol{R}\boldsymbol{a}_i\|^2 = (\text{ developing the squares }) \tag{D.8}
$$

$$
\|\boldsymbol{b}_i\|^2 + \|\boldsymbol{a}_i\|^2 - 2\boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{R}\boldsymbol{a}_i = (\text{ recalling that for a scalar } a = vec(a) \text{ }) \tag{D.9}
$$

$$
\|\boldsymbol{b}_i\|^2 + \|\boldsymbol{a}_i\|^2 - 2vec(\boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{R}\boldsymbol{a}_i) = (\text{using } vec(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathsf{T}} \otimes \boldsymbol{A})vec(\boldsymbol{B})) \tag{D.10}
$$

$$
\|\boldsymbol{b}_i\|^2 + \|\boldsymbol{a}_i\|^2 - 2(\boldsymbol{a}_i^{\mathsf{T}} \otimes \boldsymbol{b}_i^{\mathsf{T}})vec(\boldsymbol{R}) = (\text{using (D.6)}) \tag{D.11}
$$

$$
\|\boldsymbol{b}_i\|^2 + \|\boldsymbol{a}_i\|^2 - 2(\boldsymbol{a}_i^{\mathsf{T}} \otimes \boldsymbol{b}_i^{\mathsf{T}})\boldsymbol{P}\, vec(\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}) = (\text{using tr}\left(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{B}\right) = vec(\boldsymbol{A})^{\mathsf{T}}vec(\boldsymbol{B})) \tag{D.12}
$$

$$
\|\boldsymbol{b}_i\|^2 + \|\boldsymbol{a}_i\|^2 - 2\text{tr}\left(\boldsymbol{M}_{ij}^{\mathsf{T}}\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}\right) \tag{D.13}
$$

where $\boldsymbol{M}_{ij}$ is a 4×4 matrix such that $vec(\boldsymbol{M}_{ij}) = ((\boldsymbol{a}_i^{\mathsf{T}} \otimes \boldsymbol{b}_i^{\mathsf{T}})\boldsymbol{P})^{\mathsf{T}} = \boldsymbol{P}^{\mathsf{T}}(\boldsymbol{a}_i \otimes \boldsymbol{b}_i)$ (in other words, $\boldsymbol{M}_{ij}$ simply rearranges the 16 entries of the vector $\boldsymbol{P}^{\mathsf{T}}(\boldsymbol{a}_i \otimes \boldsymbol{b}_i)$ into a 4×4 matrix). Replacing $\|\boldsymbol{b}_i - \boldsymbol{R}\boldsymbol{a}_i\|^2$ in (D.2) with (D.13) yields the inequality in (D.3), hence proving the claim. ■

Proposition D.2 (Sparse moment relaxation of (D.3))**.** The following SDP is a convex relaxation of the non-convex optimization problem (D.3):

$$\min_{\boldsymbol{X}\in\mathbb{S}^{5(n+1)}} \quad \sum_{i=1}^{n} \boldsymbol{X}^2_{[\omega_i]} \tag{D.14}$$

$$\text{s.t.} \quad \boldsymbol{X}_{[\omega_i\,,\,\omega_i]} = \boldsymbol{X}_{[\omega_i]}, \quad i = [n] \tag{D.15}$$

$$\sum_{i=1}^{n} \boldsymbol{X}_{[\omega_i]} = \alpha n \tag{D.16}$$

$$\boldsymbol{X}_{[\omega_i]} \cdot \left(\|\boldsymbol{b}_i\|^2 + \|\boldsymbol{a}_i\|^2\right) - 2\mathrm{tr}\left(\boldsymbol{M}_{ij}^{\mathsf{T}} \boldsymbol{X}_{[\boldsymbol{q}\,,\,\omega_i\boldsymbol{q}^{\mathsf{T}}]}\right) \le \bar{c}^2, \quad i = [n] \tag{D.17}$$

$$\mathrm{tr}\left(\boldsymbol{X}_{[\boldsymbol{q}\,,\,\boldsymbol{q}^{\mathsf{T}}]}\right) = 1 \tag{D.18}$$

$$\boldsymbol{X} \succeq 0 \tag{D.19}$$

$$\boldsymbol{X}_{[1]} = 1 \tag{D.20}$$

$$\boldsymbol{X}_{[\omega_i\boldsymbol{q}\,,\,\omega_i\boldsymbol{q}^{\mathsf{T}}]} = \boldsymbol{X}_{[\boldsymbol{q}\,,\,\omega_i\boldsymbol{q}^{\mathsf{T}}]}, \quad i = [n] \tag{D.21}$$

$$\boldsymbol{X}_{[\omega_i\boldsymbol{q}^{\mathsf{T}}]} = \boldsymbol{X}_{[\omega_i\,,\,\boldsymbol{q}^{\mathsf{T}}]}, \quad i = [n] \tag{D.22}$$

$$\boldsymbol{X}_{[\omega_i\boldsymbol{q}^{\mathsf{T}}]} = \boldsymbol{X}_{[\omega_i\,,\,\omega_i\boldsymbol{q}^{\mathsf{T}}]}, \quad i = [n] \tag{D.23}$$

$$\boldsymbol{X}_{[\omega_i\boldsymbol{q}\,,\,\omega_j\boldsymbol{q}^{\mathsf{T}}]} = \boldsymbol{X}^{\mathsf{T}}_{[\omega_i\boldsymbol{q}\,,\,\omega_j\boldsymbol{q}^{\mathsf{T}}]}, \quad i,j = [n] \tag{D.24}$$

$$\mathrm{tr}\left(\boldsymbol{X}_{[\omega_i\boldsymbol{q}\,,\,\omega_j\boldsymbol{q}^{\mathsf{T}}]}\right) = \boldsymbol{X}_{[\omega_i\,,\,\omega_j]}, \quad i,j = [n] \tag{D.25}$$

where we index the rows of the matrix $\boldsymbol{X}$ according to the monomials $\boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q})$, index the columns of $\boldsymbol{X}$ according to the monomials $\boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q})^{\mathsf{T}} \triangleq [1\,,\,\omega_1\,,\,\ldots\,,\,\omega_n\,,\,\boldsymbol{q}^{\mathsf{T}}\,,\,\omega_1\boldsymbol{q}^{\mathsf{T}}\,,\,\ldots\,,\,\omega_n\boldsymbol{q}^{\mathsf{T}}]$, and use the notation $\boldsymbol{X}_{[i,j]}$ to access entries of the matrix with row indexed by monomial i and column indexed by monomial j; we also overload the notation and write as $\boldsymbol{X}_{[i]}$ to denote $\boldsymbol{X}_{[i,1]}$.

Proof. While the SDP appears to be quite complicated, its constraints should become apparent from the structure of the moment matrix built on the sparse monomial basis $\boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q}) \triangleq [1\,;\,\omega_1\,;\,\ldots\,;\,\omega_n\,;\,\boldsymbol{q}\,;\,\omega_1\boldsymbol{q}\,;\,\ldots\,;\,\omega_n\boldsymbol{q}]$:

$$\boldsymbol{X} = \boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q})\boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q})^{\mathsf{T}} =$$

$$\begin{array}{c} \\ \begin{array}{c} 1 \\ \omega_1 \\ \vdots \\ \omega_n \\ \boldsymbol{q} \\ \omega_1\boldsymbol{q} \\ \vdots \\ \omega_n\boldsymbol{q} \end{array} \end{array} \begin{array}{c} \begin{array}{cccc|c|ccc} 1 & \omega_1 & \dots & \omega_n & \boldsymbol{q}^{\mathsf{T}} & \omega_1\boldsymbol{q}^{\mathsf{T}} & \dots & \omega_n\boldsymbol{q}^{\mathsf{T}} \end{array} \\ \left[\begin{array}{cccc|c|ccc} 1 & \omega_1 & \dots & \omega_n & \boldsymbol{q}^{\mathsf{T}} & \omega_1\boldsymbol{q}^{\mathsf{T}} & \dots & \omega_n\boldsymbol{q}^{\mathsf{T}} \\ * & \omega_1^2 & \dots & \vdots & \omega_1\boldsymbol{q}^{\mathsf{T}} & \omega_1^2\boldsymbol{q}^{\mathsf{T}} & \dots & \omega_1\omega_n\boldsymbol{q}^{\mathsf{T}} \\ * & * & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & \dots & \omega_n^2 & \omega_n\boldsymbol{q}^{\mathsf{T}} & \omega_1\omega_n\boldsymbol{q}^{\mathsf{T}} & \dots & \omega_n^2\boldsymbol{q}^{\mathsf{T}} \\ \hline * & * & \dots & * & \boldsymbol{q}\boldsymbol{q}^{\mathsf{T}} & \omega_1\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}} & \dots & \omega_n\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}} \\ \hline * & * & \dots & * & * & \omega_1^2\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}} & \dots & \omega_1\omega_n\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}} \\ * & * & \dots & * & * & * & \ddots & \vdots \\ * & * & \dots & * & * & * & \dots & \omega_n^2\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}} \end{array}\right] \end{array} \tag{D.26}$$

where we also reported in gray the row and column indices described in the statement of the proposition. We prove the proposition in two steps. First, we show how to rewrite (D.3) using the moment matrix $\boldsymbol{X}$ in (D.26), which leads to the objective and constraints in (D.14)-(D.18). Second, we show that any moment matrix with the structure in (D.26) satisfies the constraints (D.19)-(D.25), hence the feasible set of (D.14) contains the feasible set of (D.3). Let us start by rewriting (D.3) using the moment matrix $\boldsymbol{X}$:

$$\min_{\boldsymbol{\omega},\boldsymbol{q},\boldsymbol{X}} \quad \sum_{i=1}^{n} \boldsymbol{X}^2_{[\omega_i]} \tag{D.27}$$

$$\text{s.t.} \quad \boldsymbol{X}_{[\omega_i\,,\,\omega_i]} = \boldsymbol{X}_{[\omega_i]}, \quad i = [n] \tag{D.28}$$

$$\sum_{i=1}^{n} \boldsymbol{X}_{[\omega_i]} = \alpha n \tag{D.29}$$

$$\boldsymbol{X}_{[\omega_i]} \cdot \left(\|\boldsymbol{b}_i\|^2 + \|\boldsymbol{a}_i\|^2\right) - 2\mathrm{tr}\left(\boldsymbol{M}_{ij}^{\mathsf{T}}\boldsymbol{X}_{[\boldsymbol{q}\,,\,\omega_i\boldsymbol{q}^{\mathsf{T}}]}\right) \leq \bar{c}^2, \quad i = [n] \tag{D.30}$$

$$\mathrm{tr}\left(\boldsymbol{X}_{[\boldsymbol{q}\,,\,\boldsymbol{q}^{\mathsf{T}}]}\right) = 1 \tag{D.31}$$

$$\boldsymbol{X} = \boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q}) \cdot \boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q})^{\mathsf{T}} \tag{D.32}$$

where $\boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q}) \triangleq [1\,;\,\omega_1\,;\,\dots\,;\,\omega_n\,;\,\boldsymbol{q}\,;\,\omega_1\boldsymbol{q}\,;\,\dots\,;\,\omega_n\boldsymbol{q}]$, and we simply noticed (from inspection of (D.26)) that $\boldsymbol{X}_{[\omega_i\,,\,\omega_i]} = \omega_i^2$, $\boldsymbol{X}_{[\omega_i]} = \omega_i$, $\boldsymbol{X}_{[\boldsymbol{q}\,,\,\omega_i\boldsymbol{q}^{\mathsf{T}}]} = \omega_i\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}$, and $\mathrm{tr}\left(\boldsymbol{X}_{[\boldsymbol{q}\,,\,\boldsymbol{q}^{\mathsf{T}}]}\right) = \mathrm{tr}\left(\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}\right) = \boldsymbol{q}^{\mathsf{T}}\boldsymbol{q} = \|\boldsymbol{q}\|^2$, hence (D.27) just rewrites objective and constraints in (D.3) using the entries of the moment matrix $\boldsymbol{X}$ in (D.1). Problem (D.27) is equivalent to (D.3) and is still non-convex due to the non-convexity of the constraint $\boldsymbol{X} = \boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q}) \cdot \boldsymbol{m}(\boldsymbol{\omega},\boldsymbol{q})^{\mathsf{T}}$.

Now we are only left to prove that the feasible set of (D.14) contains the feasible set of (D.27). More precisely, we prove that any matrix that satisfies $\boldsymbol{X} = \boldsymbol{m}(\boldsymbol{\omega}, \boldsymbol{q}) \cdot \boldsymbol{m}(\boldsymbol{\omega}, \boldsymbol{q})^{\mathsf{T}}$ also satisfies constraints (D.19)-(D.25) in (D.14). Clearly, any $\boldsymbol{X} = \boldsymbol{m}(\boldsymbol{\omega}, \boldsymbol{q}) \cdot \boldsymbol{m}(\boldsymbol{\omega}, \boldsymbol{q})^{\mathsf{T}}$ is such that $\boldsymbol{X} \succeq 0$. The rest of the constraints can be also seen to hold by simple inspection of the entries of the moment matrix (D.1) and recalling that our constraint set also imposes $\omega_i^2 = \omega_i$ (for all $i \in [n]$) and $\operatorname{tr}\left(\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}\right) = \|\boldsymbol{q}\|^2 = 1$. Therefore, since (D.14) has the same objective of (D.27), but its feasible set includes the feasible set of (D.27), problem (D.14) is a relaxation of (D.27). Finally, we observe that (D.14) is a convex program, since it minimizes a convex cost function over the cone of positive-semidefinite matrices and subject to a set of linear constraints. ■

Rounding for rotation search. According to Algorithm 6, after solving the sparse moment relaxation and obtaining the matrix $\boldsymbol{X}^{\star}$, we build the vectors $\boldsymbol{v}_i$ from the entries of the matrix $\boldsymbol{X}^{\star}$, and then project those vectors to the domain $\mathbb{X}$. In our quaternion-based formulation of the rotation search problem (Proposition D.1), $\boldsymbol{v}_i$ are 4-dimensional vectors, while $\mathbb{X}$ is the set of unit quaternions. Hence projecting onto the domain $\mathbb{X}$ (line 5 in Algorithm 6) only requires normalizing the vectors $\boldsymbol{v}_i$ to have unit norm, *i.e.*, $\boldsymbol{x}_i = \boldsymbol{v}_i / \|\boldsymbol{v}_i\|$. In particular, we add $\boldsymbol{x}_i = \boldsymbol{v}_i / \|\boldsymbol{v}_i\|$ whenever $\|\boldsymbol{v}_i\| > 0$, while we mark an estimate as invalid when $\|\boldsymbol{v}_i\| = 0$ and disregard it from the evaluation.

References

[1] P. Agarwal, G. D. Tipaldi, L. Spinello, C. Stachniss, and W. Burgard, "Robust map optimization using dynamic covariance scaling," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2013.

[2] P. Antonante, V. Tzoumas, H. Yang, and L. Carlone, "Outlier-robust estimation: Hardness, minimally tuned algorithms, and applications," *IEEE Trans. Robotics*, vol. 38, no. 1, 2021, pp. 281–301. URL: https://arxiv.org/pdf/2007.15109.pdf.

[3] P. Awasthi, M. F. Balcan, and P. M. Long, "The power of localization for efficiently learning linear separators with noise," *J. ACM*, vol. 63, no. 6, 2017.

[4] A. Bakshi and P. K. Kothari, "List-decodable subspace recovery: Dimension independent error in polynomial time," in *Proc. of the Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, ser. SODA '21, pp. 1279–1297, 2021.

[5] M.-F. Balcan, A. Blum, and S. Vempala, "A discriminative framework for clustering via similarity functions," in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, ser. STOC '08, pp. 671–680, 2008.

[6] Y. Bar-Shalom and X. Li, *Estimation and Tracking: principles, techniques and software*. Boston, London: Artech House, 1993.

[7] B. Barak, F. Brandao, A. Harrow, J. Kelner, D. Steurer, and Y. Zhou, "Hypercontractivity, sum-of-squares proofs, and their applications," in *Proc. of the Annual ACM Symp. on Theory of Computing*, pp. 307–326, 2012.

[8] B. Barak, J. A. Kelner, and D. Steurer, "Rounding sum-of-squares relaxations," in *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, ser. STOC '14, pp. 31–40, 2014.

[9] B. Barak and D. Steurer, "Proofs, beliefs, and algorithms through the lens of sum-of-squares," in *Lecture notes in preparation*, 2016. URL: http://sumofsquares.org.

[10] J. T. Barron, "A general and adaptive robust loss function," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4331–4339, 2019.

[11] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in *European Conf. on Computer Vision (ECCV)*, 2006.

[12] J. C. Bazin, Y. Seo, and M. Pollefeys, "Globally optimal consensus set maximization through rotation search," in *Asian Conference on Computer Vision*, Springer, pp. 539–551, 2012.

[13] T. Bernholt, "Robust estimators are hard to compute," Dortmund, Technical Report 2005, 52, 2006. URL: http://hdl.handle.net/10419/22645.

[14] K. Bhatia, P. Jain, and P. Kar, "Robust regression via hard thresholding," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 721–729, 2015.

[15] K. Bhatia, P. Jain, P. Kamalaruban, and P. Kar, "Consistent robust regression," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 30, Curran Associates, Inc., 2017.

[16] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Intl. J. of Computer Vision*, vol. 19, no. 1, 1996, pp. 57–91.

[17] G. Blekherman, P. A. Parrilo, and R. R. Thomas, *Semidefinite optimization and convex algebraic geometry*. SIAM, 2012.

[18] N. Boumal, V. Voroninski, and A. Bandeira, "The non-convex Burer-Monteiro approach works on smooth semidefinite programs," *arXiv*, 2016.

[19] J. Briales and J. Gonzalez-Jimenez, "Cartan-sync: Fast and global SE(d)-synchronization," *IEEE Robot. Autom. Lett*, vol. 2, no. 4, 2017, pp. 2127–2134.

[20] J. Briales and J. Gonzalez-Jimenez, "Convex Global 3D Registration with Lagrangian Duality," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[21] J. Briales, L. Kneip, and J. Gonzalez-Jimenez, "A certifiably globally optimal solution to the non-minimal relative pose problem," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[22] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robotics*, vol. 32, no. 6, 2016, pp. 1309–1332. DOI: 10.1109/TRO.2016.2624754.

[23] L. Carlone and G. Calafiore, "Convex relaxations for pose graph optimization with outliers," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 2, 2018, pp. 1160–1167. URL: https://arxiv.org/pdf/1801.02112.pdf.

[24] L. Carlone, G. Calafiore, C. Tommolillo, and F. Dellaert, "Planar pose graph optimization: Duality, optimal solutions, and verification," *IEEE Trans. Robotics*, vol. 32, no. 3, 2016, pp. 545–565. URL: https://www.dropbox.com/s/peoktkct0cw42av/2015j-TRO-dualityPGO2D.pdf?dl=0.

[25] L. Carlone and F. Dellaert, "Duality-based verification techniques for 2D SLAM," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 4589–4596, 2015. URL: https://www.dropbox.com/s/4wtxyp817hdfnna/2015c-ICRA-duality2D.pdf?dl=0.

[26] L. Carlone, D. Rosen, G. Calafiore, J. Leonard, and F. Dellaert, "Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 125–132, 2015. URL: https://doi.org/10.48550/arXiv.1506.00746.

[27] L. Carlone, R. Tron, K. Daniilidis, and F. Dellaert, "Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 4597–4604, 2015. URL: https://www.dropbox.com/s/x0n5r366u33fu7x/2015c-ICRA-initPGO3d.pdf?dl=0.

[28] M. Charikar, J. Steinhardt, and G. Valiant, "Learning from untrusted data," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2017, pp. 47–60, Montreal, Canada, 2017.

[29] N. Chebrolu, T. Läbe, O. Vysotska, J. Behley, and C. Stachniss, "Adaptive robust kernels for non-linear least squares problems," *arXiv preprint arXiv:2004.14938*, 2020.

[30] Y. Chen, C. Caramanis, and S. Mannor, "Robust sparse regression under adversarial corruption," in *Intl. Conf. on Machine Learning (ICML)*, vol. 28, pp. 774–782, 2013.

[31] T.-J. Chin, Z. Cai, and F. Neumann, "Robust fitting in computer vision: Easy or hard?" In *European Conf. on Computer Vision (ECCV)*, 2018.

[32] T. J. Chin and D. Suter, "The maximum consensus problem: Recent algorithmic advances," *Synthesis Lectures on Computer Vision*, vol. 7, no. 2, 2017, pp. 1–194.

[33] T.-J. Chin, P. Purkait, A. Eriksson, and D. Suter, "Efficient globally optimal consensus maximisation with tree search," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2413–2421, 2015.

[34] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Comput. Vis. Image Underst.*, vol. 61, no. 1, Jan. 1995, pp. 38–59.

[35] T. D'Orsi, G. Novikov, and D. D. Steurer, "Consistent regression when oblivious outliers overwhelm," in *Intl. Conf. on Machine Learning (ICML)*, vol. 139, pp. 2297–2306, 2021.

[36] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte Carlo Localization for mobile robots," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 1999.

[37] F. Dellaert and M. Kaess, "Factor graphs for robot perception," *Foundations and Trends in Robotics*, vol. 6, no. 1-2, 2017, pp. 1–139.

[38] F. Dellaert, D. Rosen, J. Wu, R. Mahony, and L. Carlone, "Shonan rotation averaging: Global optimality by surfing $SO(p)^n$," in *European Conf. on Computer Vision (ECCV)*, 2020. URL: https://dellaert.github.io/ShonanAveraging/.

[39] I. Diakonikolas, G. Kamath, D. Kane, J. Li, A. Moitra, and A. Stewart, "Robust estimators in high dimensions without the computational intractability," in *IEEE 57th Annual Symposium on Foundations of Computer Science*, IEEE, pp. 655–664, 2016.

[40] I. Diakonikolas, G. Kamath, D. Kane, J. Li, A. Moitra, and A. Stewart, "Robust estimators in high-dimensions without the computational intractability," *SIAM Journal on Computing*, vol. 48, no. 2, 2019, pp. 742–864. DOI: 10.1137/17M1126680.

[41] I. Diakonikolas, G. Kamath, D. Kane, J. Li, J. Steinhardt, and A. Stewart, "Sever: A robust meta-algorithm for stochastic optimization," in *Intl. Conf. on Machine Learning (ICML)*, K. Chaudhuri and R. Salakhutdinov, Eds., ser. Proceedings of Machine Learning Research, vol. 97, pp. 1596–1606, 2019.

[42] I. Diakonikolas, G. Kamath, D. M. Kane, J. Li, A. Moitra, and A. Stewart, "Robustly learning a gaussian: Getting optimal error, efficiently," in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '18, pp. 2683–2702, New Orleans, Louisiana, 2018.

[43] I. Diakonikolas, D. M. Kane, S. Karmalkar, A. Pensia, and T. Pittas, "Robust sparse mean estimation via sum of squares," in *Proceedings of Thirty Fifth Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, vol. 178, pp. 4703–4763, PMLR, 2022.

[44] I. Diakonikolas, D. M. Kane, and A. Stewart, "Learning geometric concepts with nasty noise," in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2018, pp. 1061–1073, 2018.

[45] I. Diakonikolas, D. M. Kane, and A. Stewart, "List-decodable robust mean estimation and learning mixtures of spherical gaussians," in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2018, pp. 1047–1060, 2018.

[46] I. Diakonikolas, W. Kong, and A. Stewart, "Efficient algorithms and lower bounds for robust linear regression," in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '19, pp. 2745–2754, San Diego, California, 2019.

[47] K. Doherty, D. M. Rosen, and J. Leonard, "Performance guarantees for spectral initialization in rotation averaging and pose-graph SLAM," *CoRR*, vol. abs/2201.03773, 2022. URL: https://arxiv.org/abs/2201.03773.

[48] S. S. Du, S. Balakrishnan, and A. Singh, "Computationally efficient robust estimation of sparse functionals," *CoRR*, vol. abs/1702.07709, 2017. URL: https://arxiv.org/abs/1702.07709.

[49] O. Enqvist, E. Ask, F. Kahl, and K. Åström, "Robust fitting for multiple view geometry," in *European Conf. on Computer Vision (ECCV)*, Springer, pp. 738–751, 2012.

[50] O. Enqvist, E. Ask, F. Kahl, and K. Åström, "Tractable algorithms for robust model estimation," *Intl. J. of Computer Vision*, vol. 112, no. 1, 2015, pp. 115–129.

[51] O. Enqvist and F. Kahl, "Robust optimal pose estimation," in *European Conf. on Computer Vision (ECCV)*, Springer, pp. 141–153, 2008.

[52] A. Eriksson, C. Olsson, F. Kahl, and T.-J. Chin, "Rotation averaging and strong duality," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[53] T. Fan, H. Wang, M. Rubenstein, and T. Murphey, "Efficient and guaranteed planar pose graph optimization using the complex number representation," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1904–1911, 2019.

[54] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography," *Commun. ACM*, vol. 24, 1981, pp. 381–395.

[55] N. Fleming, P. Kothari, and T. Pitassi, "Semialgebraic proofs and efficient algorithm design," *Foundations and Trends in Theoretical Computer Science*, vol. 14, no. 1-2, 2019, pp. 1–221.

[56] D. Fourie, J. Leonard, and M. Kaess, "A nonparametric belief solution to the bayes tree," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 2189–2196, 2016.

[57] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv e-print: 1412.6572*, 2014.

[58] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming*. URL: http://cvxr.com/cvx.

[59] R. Hartley, J. Trumpf, Y. Dai, and H. Li, "Rotation averaging," *IJCV*, vol. 103, no. 3, 2013, pp. 267–305.

[60] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Second. Cambridge University Press, 2004.

[61] T. Hitchcox and J. R. Forbes, "Mind the gap: Norm-aware adaptive robust loss for multivariate least-squares problems," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, 2022, pp. 7116–7123.

[62] S. B. Hopkins and J. Li, "Mixture models, robustness, and sum of squares proofs," in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2018, pp. 1021–1034, Los Angeles, CA, USA, 2018.

[63] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Amer.*, vol. 4, no. 4, Apr. 1987, pp. 629–642.

[64] M. Hsiao and M. Kaess, "MH-iSAM2: multi-hypothesis iSAM using Bayes Tree and Hypo-tree," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 1274–1280, 2019.

[65] P. Huber, *Robust Statistics*. John Wiley & Sons, New York, NY, 1981.

[66] M. Ikram, S. Khaliq, M. Anjum, and W. Hussain, "Perceptual aliasing++: Adversarial attack for visual SLAM front-end and back-end," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, 2022, pp. 4670–4677.

[67] G. Izatt, H. Dai, and R. Tedrake, "Globally optimal object pose estimation in point clouds with mixed-integer programming," in *Proc. of the Intl. Symp. of Robotics Research (ISRR)*, 2017.

[68] F. Kahl and D. Henrion, "Globally optimal estimates for geometric reconstruction problems," *Intl. J. of Computer Vision*, vol. 74, no. 1, 2007, pp. 3–15.

[69] S. Karmalkar and E. Price, "Compressed sensing with adversarial sparse noise via L1 regression," *CoRR*, vol. abs/1809.08055, 2018. URL: http://arxiv.org/abs/1809.08055.

[70] S. Karmalkar, A. Klivans, and P. Kothari, "List-decodable linear regression," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 32, 2019.

[71] A. R. Klivans, P. K. Kothari, and R. Meka, "Efficient algorithms for outlier-robust regression," *CoRR*, vol. abs/1803.03241, 2018. URL: http://arxiv.org/abs/1803.03241.

[72] A. R. Klivans, P. M. Long, and R. A. Servedio, "Learning halfspaces with malicious noise," in *Automata, Languages and Programming*, S. Albers, A. Marchetti-Spaccamela, Y. Matias, S. Nikoletseas, and W. Thomas, Eds., pp. 609–621, 2009.

[73] L. Kneip, H. Li, and Y. Seo, "UPnP: An optimal o(n) solution to the absolute pose problem with universal applicability," in *European Conf. on Computer Vision (ECCV)*, Springer, pp. 127–142, 2014.

[74] P. K. Kothari and J. Steinhardt, "Better agnostic clustering via relaxed tensor norms," *CoRR*, vol. abs/1711.07465, 2017. URL: http://arxiv.org/abs/1711.07465.

[75] P. K. Kothari, J. Steinhardt, and D. Steurer, "Robust moment estimation and improved clustering via sum of squares," in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2018, pp. 1035–1046, Los Angeles, CA, USA, 2018.

[76] K. A. Lai, A. B. Rao, and S. Vempala, “Agnostic estimation of mean and covariance,” in *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 665–674, IEEE Computer Society, 2016. DOI: 10.1109/FOCS.2016.76.

[77] P. Lajoie, S. Hu, G. Beltrame, and L. Carlone, “Modeling perceptual aliasing in SLAM via discrete-continuous graphical models,” *IEEE Robotics and Automation Letters (RA-L)*, 2019.

[78] V. Larsson, M. Oskarsson, K. Astrom, A. Wallis, T. Pajdla, and Z. Kukelova, “Beyond Grobner bases: Basis selection for minimal solvers,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 3945–3954, 2018.

[79] J. Lasserre, *Moments, positive polynomials and their applications*, vol. 1. World Scientific, 2010.

[80] J. Lasserre, “The Moment-SoS hierarchy,” in *Int. Cong. of Math.*, vol. 4, pp. 3791–3815, 2018.

[81] J. B. Lasserre, “Global optimization with polynomials and the problem of moments,” *SIAM J. Optim.*, vol. 11, no. 3, 2001, pp. 796–817.

[82] H. M. Le, T.-J. Chin, A. Eriksson, T.-T. Do, and D. Suter, “Deterministic approximate methods for maximum consensus robust fitting,” *IEEE Trans. Pattern Anal. Machine Intell.*, 2019.

[83] S. H. Lee and J. Civera, “Robust single rotation averaging,” *CoRR*, vol. abs/2004.00732, 2020. URL: https://arxiv.org/abs/2004.00732.

[84] H. Li, “Consensus set maximization with guaranteed global optimality for robust geometry estimation,” in *Intl. Conf. on Computer Vision (ICCV)*, pp. 1074–1080, 2009.

[85] H. Li and R. Hartley, “The 3D-3D registration problem revisited,” in *Intl. Conf. on Computer Vision (ICCV)*, IEEE, pp. 1–8, 2007.

[86] T. Ma, J. Shi, and D. Steurer, “Polynomial-time tensor decompositions with sum-of-squares,” in *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 438–446, 2016.

[87] G. Meneghetti, M. Danelljan, M. Felsberg, and K. Nordberg, "Image alignment for panorama stitching in sparsely structured environments," in *Scandinavian Conference on Image Analysis*, Springer, pp. 428–439, 2015.

[88] MOSEK ApS, *The mosek optimization toolbox for matlab manual. version 8.1.* 2017. URL: http://docs.mosek.com/8.1/toolbox/index.html.

[89] Y. Nesterov, "Squared functional systems and optimization problems, high performance optimization," *Appl. Optim.*, vol. 33, 2000, pp. 405–440.

[90] N. Nguyen and T. Tran, "Exact recoverability from dense corrupted observations via ℓ_1-minimization," *IEEE Trans. on Information Theory*, vol. 59, no. 4, 2013, pp. 2017–2035.

[91] J. Nie, "Optimality conditions and finite convergence of lasserre's hierarchy," *Mathematical programming*, vol. 146, no. 1-2, 2014, pp. 97–121.

[92] T. Norman, N. Weinberger, and K. Levy, "Robust linear regression for general feature distribution," *ArXiv*, vol. abs/2202.02080, 2022.

[93] C. Olsson, O. Enqvist, and F. Kahl, "A polynomial-time bound for matching and registration with outliers," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 1–8, 2008.

[94] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection.* John Wiley & Sons, New York, NY, 1987.

[95] A. Papachristodoulou, J. Anderson, G. Valmorbida, S. Prajna, P. Seiler, and P. A. Parrilo, *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*, 2013. URL: http://arxiv.org/abs/1310.4716.

[96] Á. Parra Bustos and T. J. Chin, "Guaranteed outlier removal for rotation search," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2165–2173, 2015.

[97] Á. Parra Bustos and T. J. Chin, "Guaranteed outlier removal for point cloud registration with correspondences," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 40, no. 12, 2018, pp. 2868–2882.

[98] P. Parrilo, "Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization," Ph.D. dissertation, California Institute of Technology, 2000.

[99] P. A. Parrilo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical programming*, vol. 96, no. 2, 2003, pp. 293–320.

[100] D. P. Paudel, A. Habed, C. Demonceaux, and P. Vasseur, "Robust and optimal sum-of-squares-based point-to-plane registration of image sets and structured scenes," in *Intl. Conf. on Computer Vision (ICCV)*, pp. 2048–2056, 2015.

[101] L. Peng, M. Fazlyab, and R. Vidal, *Towards understanding the semidefinite relaxations of truncated least-squares in robust rotation search*, 2022. DOI: 10.48550/ARXIV.2207.08350.

[102] A. Prasad, A. S. Suggala, S. Balakrishnan, and P. Ravikumar, "Robust estimation via robust gradient estimation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 82, 2020.

[103] P. Raghavendra and M. Yau, "List decodable learning via sum of squares," in *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '20, pp. 161–180, Salt Lake City, Utah, 2020.

[104] D. Rosen, L. Carlone, A. Bandeira, and J. Leonard, "SE-Sync: A certifiably correct algorithm for synchronization over the Special Euclidean group," *Intl. J. of Robotics Research*, 2018. URL: https://arxiv.org/abs/1611.00128.

[105] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, NY, 1981.

[106] D. Scaramuzza and F. Fraundorfer, "Visual odometry: Part I the first 30 years and fundamentals," 2011.

[107] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113, 2016.

[108] G. Schweighofer and A. Pinz, "Globally optimal O(n) solution to the PnP problem for general camera models.," in *British Machine Vision Conf. (BMVC)*, pp. 1–10, 2008.

[109] J. Shi, H. Yang, and L. Carlone, "Optimal pose and shape estimation for category-level 3D object perception," in *Robotics: Science and Systems (RSS)*, 2021.

[110] J. Shi, H. Yang, and L. Carlone, "Optimal and robust category-level perception: Object pose and shape estimation from 2D and 3D semantic keypoints," *arXiv preprint: 2206.12498*, 2022. URL: https://arxiv.org/pdf/2206.12498.pdf.

[111] N. Shor, "Nondifferentiable optimization and polynomial problems," *Nonconvex Optimization and its Applications*, vol. 24, 1998.

[112] N. Sunderhauf and P. Protzel, "Towards a robust back-end for pose graph SLAM," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 1254–1261, 2012.

[113] K. M. Tavish and T. D. Barfoot, "At all costs: A comparison of robust cost functions for camera correspondence outliers," in *Conf. Computer and Robot Vision*, IEEE, pp. 62–69, 2015.

[114] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *International workshop on vision algorithms*, Springer, pp. 298–372, 1999.

[115] R. Tron, D. Rosen, and L. 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"*, 2015. URL: https://www.dropbox.com/s/859umrdf7ldd2kv/2015ws-rss-duality3Ddet.pdf?dl=0.

[116] E. Tsakonas, J. Jaldén, N. Sidiropoulos, and B. Ottersten, "Convergence of the huber regression M-estimate in the presence of dense outliers," *IEEE Signal Processing Letters*, vol. 21, no. 10, 2014, pp. 1211–1214.

[117] J. W. Tukey, "Mathematics and the picturing of data," *Proceedings of the International Congress of Mathematicians*, vol. 2, 1975, pp. 523–531. URL: https://cir.nii.ac.jp/crid/1573950399770196096.

[118] R. H. Tütüncü, K.-C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical programming*, vol. 95, no. 2, 2003, pp. 189–217.

[119] G. Wahba, “A least squares estimate of satellite attitude,” *SIAM review*, vol. 7, no. 3, 1965, pp. 409–409.

[120] J. Wang, V. Magron, J. B. Lasserre, and N. H. A. Mai, “CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization,” *arXiv preprint arXiv:2005.02828*, 2020.

[121] J. Wang, V. Magron, and J.-B. Lasserre, “Chordal-TSSOS: a moment-SOS hierarchy that exploits term sparsity with chordal extension,” *SIAM Journal on Optimization*, vol. 31, no. 1, 2021, pp. 114–141.

[122] J. Wang, V. Magron, and J.-B. Lasserre, “TSSOS: A Moment-SOS hierarchy that exploits term sparsity,” *SIAM Journal on Optimization*, vol. 31, no. 1, 2021, pp. 30–58.

[123] T. Weisser, J. B. Lasserre, and K.-C. Toh, “Sparse-BSOS: A bounded degree SOS hierarchy for large scale polynomial optimization with sparsity,” *Math. Program. Comput.*, vol. 10, no. 1, 2018, pp. 1–32.

[124] J. Wright and Y. Ma, “Dense error correction via ℓ^1-minimization,” *IEEE Trans. on Information Theory*, vol. 56, no. 7, 2010, pp. 3540–3560.

[125] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, “Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection,” *IEEE Robotics and Automation Letters (RA-L)*, vol. 5, no. 2, 2020, pp. 1127–1134. URL: https://arxiv.org/pdf/1909.08605.pdf.

[126] H. Yang and L. Carlone, “A quaternion-based certifiably optimal solution to the Wahba problem with outliers,” in *Intl. Conf. on Computer Vision (ICCV)*, 2019. URL: https://arxiv.org/pdf/1905.12536.pdf.

[127] H. Yang and L. Carlone, “In perfect shape: Certifiably optimal 3D shape reconstruction from 2D landmarks,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020. URL: https://arxiv.org/pdf/1911.11924.pdf.

[128] H. Yang and L. Carlone, “One ring to rule them all: Certifiably robust geometric perception with outliers,” in *Conf. on Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 18 846–18 859, 2020. URL: https://arxiv.org/pdf/2006.06769.pdf.

[129] H. Yang and L. Carlone, "Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, 2022. URL: https://arxiv.org/pdf/2109.03349.pdf.

[130] H. Yang, L. Liang, L. Carlone, and K. Toh, "An inexact projected gradient method with rounding and lifting by nonlinear programming for solving rank-one semidefinite relaxation of polynomial optimization," *Mathematical Programming (MAPR)*, 2022. URL: https://arxiv.org/pdf/2105.14033.pdf.

[131] H. Yang, J. Shi, and L. Carlone, "TEASER: Fast and Certifiable Point Cloud Registration," *IEEE Trans. Robotics*, vol. 37, no. 2, 2020, pp. 314–333. URL: https://arxiv.org/pdf/2001.07715.pdf.

[132] H. Yang, C. Doran, and J.-J. Slotine, "Dynamical pose estimation," in *Intl. Conf. on Computer Vision (ICCV)*, pp. 5926–5935, 2021.

[133] J. Yang, H. Li, and Y. Jia, "Optimal essential matrix estimation via inlier-set maximization," in *European Conf. on Computer Vision (ECCV)*, Springer, pp. 111–126, 2014.

[134] X. Zhou, S. Leonardos, X. Hu, and K. Daniilidis, "3D shape reconstruction from 2D landmarks: A convex formulation," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015.