Stable Rank-One Matrix Completion is Solved by the Level 2 Lasserre Relaxation

As Published: https://doi.org/10.1007/s10208-020-09471-y

Publisher: Springer US

Persistent URL: https://hdl.handle.net/1721.1/136886

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Stable Rank-One Matrix Completion is Solved by the Level 2 Lasserre Relaxation

Cite this article as: Augustin Cosse and Laurent Demanet, Stable Rank-One Matrix Completion is Solved by the Level 2 Lasserre Relaxation, Foundations of Computational Mathematics https://doi.org/10.1007/s10208-020-09471-y

Foundations of Computational Mathematics manuscript No.
(will be inserted by the editor)

Stable rank one matrix completion is solved by the level 2 Lasserre relaxation.

Augustin Cosse · Laurent Demanet

Received: date / Accepted: date

Abstract This paper studies the problem of deterministic rank-one matrix completion. It is known that the simplest semidefinite programming relaxation, involving minimization of the nuclear norm, does not in general return the solution for this problem. In this paper, we show that in every instance where the problem has a unique solution, one can provably recover the original matrix through the level 2 Lasserre relaxation with minimization of the trace norm. We further show that the solution of the proposed semidefinite program is Lipschitz-stable with respect to perturbations of the observed entries, unlike more basic algorithms such as nonlinear propagation or ridge regression. Our proof is based on recursively building a certificate of optimality corresponding to a dual sum-of-squares (SoS) polynomial. This SoS polynomial is built from the polynomial ideal generated by the completion constraints and the monomials provided by the minimization of the trace. The proposed relaxation fits in the framework of the Lasserre hierarchy, albeit with the key addition of the trace objective function. Finally, we show how to represent and manipulate the moment tensor in favorable complexity by means of a hierarchical low-rank factorization.

Keywords Matrix completion · Convex optimization · Semidefinite programming · Semidefinite programming hierarchies · Duality in optimization · Sum-of-Squares polynomials

Communicated by Thomas Strohmer

✉ Augustin Cosse
Département de Mathématiques et Applications,
Ecole Normale Supérieure, Ulm, Paris.
PSL Research University.
E-mail: augustin.cosse@ens.fr

Laurent Demanet
Department of Mathematics,
Massachusetts Institute of Technology, MA.
E-mail: laurent@math.mit.edu

Mathematics Subject Classification (2010) Primary 90C22 · 90C25 · 65K05 · Secondary 90C59

Acknowledgements Both authors were supported by a grant from the MISTI MIT-Belgium seed fund. AC is grateful to the FNRS, FSMP, BAEF and Francqui Foundations and is supported by EOARD grant FA9550-18-1-7007. LD is supported by AFOSR grant FA9550-17-1-0316, ONR grant N00014-16-1-2122, and NSF grant DMS-1255203. AC is grateful to MIT Math, Harvard IACS, the University of Chicago as well as NYU Courant Institute and Center for Data Science for hosting him during this work.

1 Introduction

Low rank matrix completion has been studied extensively throughout the last few years, among other reasons because of its practical interest in machine learning and data science. Matrix completion provides a useful tool to compress and manipulate large databases such as in genomics and finance, and to infer information from a few measurements such as in collaborative filtering or triangulation. Good introductions along with recovery results for random designs and arbitrary ranks can be found in [10, 22].

The objective of this paper is to provide an algorithm that solves the rank one case in a stable and comprehensive way. Let $\mathcal{M}(1; m \times n)$ denote the set of rank-1 matrices of size $m \times n$; we consider the problem of recovering an unknown rank one matrix $\boldsymbol{X}_0 \in \mathcal{M}(1; m \times n)$, $\boldsymbol{X}_0 = \boldsymbol{x}_0 \boldsymbol{y}_0^T$ when we are given $O(m+n)$ entries from this matrix, and when those entries are possibly corrupted by an additive noise ε. We do not make any assumption on the noise. In the noiseless case, this problem reads

$$\begin{array}{rl} \text{find} & \boldsymbol{X} \in \mathbb{R}^{m\times n} \\ \text{subject to} & \operatorname{rank}(\boldsymbol{X}) = 1 \\ & \boldsymbol{X}_{ij} = (\boldsymbol{X}_0)_{ij} \quad (i,j) \in \Omega. \end{array} \tag{1.1}$$

Here Ω denotes the set of measurements. As a slight abuse, we will also speak of constraints $\{\boldsymbol{X}_{ij} - (\boldsymbol{X}_0)_{ij} = 0\}_{(i,j)\in\Omega}$ as belonging to the set Ω. In the noisy case, the second constraint in (1.1) is relaxed to $\|\boldsymbol{X}_{ij} - ((\boldsymbol{X}_0)_{ij} + \varepsilon_{ij})\| \leq \sigma$ in a standard fashion.

Clearly, one cannot always solve problem (1.1). For example, if no information is known on a given column (resp. row), it becomes impossible to recover the entries corresponding to this column (resp. row). Another limitation occurs when the rank-1 matrix has a zero entry; then the corresponding row or column will be zero, and the completion problem will generically lack injectivity. As an illustration of the issue with zero entries, consider the problem where the first row and last column are known and are both trivial. The number of measurements is $(m+n-1)$, which corresponds to the number of unknowns in the problem if we assume $(x_0)_1 = 1$. However, in this case, any matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ of the form $\boldsymbol{X} = \boldsymbol{v}\boldsymbol{w}^*$ with $v_1 = w_n = 0$ is a valid solution of the problem. For this reason, we consider the completion problem on $\mathcal{M}^*(1, m \times n)$, where $\mathcal{M}^*(1, m \times n)$ denotes the restriction of $\mathcal{M}(1; m \times n)$ to matrices for which none of the entries are zero.

To formalize the notion of injectivity, we introduce the mapping $\mathcal{R}_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^{|\Omega|}$ that corresponds to extracting the observed entries of the matrix. We let $\mathcal{R}_\Omega^1$ denote the restriction of $\mathcal{R}_\Omega$ to matrices of rank-1 that have no zero rows/columns. Invertibility of this restriction $\mathcal{R}_\Omega^1$ corresponds to asking whether one can uniquely recover the matrix $\boldsymbol{X}$ from the knowledge of $\mathcal{R}_\Omega(\boldsymbol{X})$ and the fact that $\boldsymbol{X}$ has rank 1. Let us denote by $\mathcal{V}_1$, $\mathcal{V}_2$ the sets of row and column indices of $\boldsymbol{X}$. We consider the bipartite graph $\mathcal{G}(\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$ associated to problem (1.1), where the set of edges is defined by $(i, j) \in \mathcal{E}$ iff $(i, j) \in \Omega$. The vertices of the bipartite graph $\mathcal{G}$ corresponding to $\boldsymbol{X}$ are labeled by the corresponding row and column indices. The conditions for the recovery of the matrix $\boldsymbol{X}$ from the set Ω are related to the properties of this bipartite graph as expressed by the following lemma which can be found, for example, in [24]:

Lemma 1 (Rank-1 completion) *The mask $\mathcal{R}_\Omega$ is injective on $\mathcal{M}^*(1; m \times n)$ if and only if $\mathcal{G}$ is connected.*

Lemma 1 has an interesting consequence. Within the noiseless framework, rank one matrix completion can be solved exactly through a nonlinear propagation approach. To understand this, let us write $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{y}^T \in \mathbb{R}^{m\times n}$ with $x_1 = 1$. We now use K to denote the number of variables in the problem, i.e $K = m + n - 1$. We also use $\mathbf{z} \in \mathbb{R}^K$ to denote the concatenation of $[x_2, \dots, x_m] \in \mathbb{R}^{m-1}$ and $\boldsymbol{y} = [y_1, \dots, y_n] \in \mathbb{R}^n$, $\mathbf{z} = (\boldsymbol{x}_{2\to m}, \boldsymbol{y})$. When we deal with the rank one case, an implication of lemma 1 is that for all x_i, y_j, the bipartite graph corresponding to the mask Ω always contains at least one connected path starting with an edge corresponding to an element of the first row and for which the series of existing edges corresponds to running through X according to chains of constraints of the form

$$y_{i_1} \to y_{i_1}x_{i_2} \to x_{i_2}y_{i_3} \to \dots y_{i_{L+1}}x_n \qquad \text{(to reach } x_n\text{)} \tag{1.2}$$

$$y_{i'_1} \to y_{i'_1}x_{i'_2} \to x_{i'_2}y_{i'_3} \to \dots x_{i'_{L'+1}}y_m \qquad \text{(to reach } y_m\text{)} \tag{1.3}$$

More generally, the two chains (1.2) and (1.3) can read, using the vector $\mathbf{z} \in \mathbb{R}^{m+n-1}$, as

$$z_{i_1} \to z_{i_1}z_{i_2} \to z_{i_2}z_{i_3} \to \dots z_{i_{L+1}}z_n. \tag{1.4}$$

In other words, each of the entries of $\boldsymbol{x}$ and $\boldsymbol{y}$ can always be related to an element of the first row whose value is known because of the normalization $x_1 = 1$. Each of the elements making up the bilinear constraints can then be obtained in the absence of noise by iteratively propagating the value of the elements of the first row through (1.4). As we explain in the sequel, such a propagation scheme however lacks robustness to noise, especially when the magnitude of the entries is on the order of the magnitude of the noise.

When the measurements are corrupted by noise, a popular approach is to turn to minimization of the nuclear norm as a proxy for the rank (see [16, 35] for early references). However, the nuclear norm does not always guarantee recovery of the rank one matrix $\boldsymbol{X}_0$ when the noise vanishes. An important gap regarding rank-one matrix completion has thus been the lack of an algorithm providing a proper (deterministic) stability estimate of the form

$$\|\boldsymbol{X} - \boldsymbol{X}_0\| \leq \omega(\|\mathcal{R}_\Omega(\boldsymbol{X}) - \mathcal{R}_\Omega(\boldsymbol{X}_0)\|). \tag{1.5}$$

for some Lipschitz function $\omega(\tau)$ obeying $\omega(\tau) \to 0$ when $\tau \to 0$.

We can now state the main contributions of the paper.

- First, we show that rank-one matrix completion can be solved through the level 2 Lasserre relaxation. Our result is sharp in terms of measurements; recovery is always possible as soon as the nonlinear problem has a unique solution. This is in contrast to previous results that required a random measurement set [9]. It also confirms that there exist instances of rank minimization problems that can be solved in a comprehensive way (without constraints of incoherence and/or randomness) using higher levels (> 1) of semidefinite programming relaxation.
- Second, we show that when the measurements are corrupted by noise, the solution to the semidefinite relaxation remains proportional to the noise level. In particular, this solution is shown to be Lipschitz-stable with respect to the noise level. This is in contrast with nonlinear approaches such as [24, 25].
- Finally, our proof system, based on constructing a dual polynomial, incidentally reveals two important facts: First, minimization of the trace norm helps certify recovery because it provides additional squares of monomials that are useful in constructing the certificate of optimality. Second, recovery can be related to the possibility of propagating known information through the graph by means of polynomial equations.

The next sections discuss the limitations of propagation, minimization of the nuclear norm, and ridge regression. We illustrate these limitations on the simple problem of completing the rank-one matrix $\boldsymbol{X}_0$,

$$\boldsymbol{X}_0 = \begin{pmatrix} 1 & ? \\ \delta & 1 \end{pmatrix}, \tag{1.6}$$

when δ is a small parameter, and for which, given the rank one constraint, the only missing entry is obviously given by $1/\delta$.

1.1 Propagation is unstable

We start by discussing the simple propagation scheme. In the noiseless framework, this scheme can be efficiently applied by writing $\boldsymbol{X}_0$ as

$$\boldsymbol{X}_0 = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}, \tag{1.7}$$

and simply deriving X_{12} as $X_{12} = \frac{X_{22}X_{11}}{X_{21}}$. Now assume that the entries are corrupted by a noise $\boldsymbol{\epsilon}$ so that the measurements are given by $\tilde{X}_{11} = (X_0)_{11} + \varepsilon_{11}$, $\tilde{X}_{21} = (X_0)_{21} + \varepsilon_{21}$ and $\tilde{X}_{22} = (X_0)_{22} + \varepsilon_{22}$ with $\varepsilon = (\varepsilon_{11}, \varepsilon_{21}, \varepsilon_{22}) \in \mathbb{R}^3$. Taking a noise ε with $\|\varepsilon\|$ on the order of δ, such as for example $\varepsilon_{21} = -.9\delta$, will result in important errors when using propagation as shown below,

$$\tilde{X}_{12} = \frac{\tilde{X}_{22}\tilde{X}_{11}}{\tilde{X}_{21}} = \frac{1 + O(\delta)}{\delta + \varepsilon_{21}} \sim 10\frac{1}{\delta} \qquad \frac{\tilde{X}_{12} - (X_0)_{12}}{(X_0)_{12}} = 900\% \tag{1.8}$$

The estimates derived through the propagation algorithm are thus unreliable when the entries are corrupted by an unknown noise ε of magnitude comparable to the smallest entries in the matrix. In addition, there is no effective, general method to select the propagation path optimally.

Another elementary method consists in taking the logarithm of the constraints, and solving the resulting system to obtain the logarithm of the unknowns. Although the method is reasonable for some positive matrices, it is easy to see that it suffers from a similar kind of instability.

1.2 Nuclear norm fails

In this section, we briefly study how nuclear norm minimization would perform in the framework of problem (1.6) as before. Nuclear norm minimization was first formalized in [16, 35], and guarantees were given, for the matrix completion problem, in a probabilistic framework, in [9]. Nuclear norm minimization relies on solving the convex program

$$\text{minimize} \quad \|\boldsymbol{X}\|_* \tag{1.9a}$$

$$\text{subject to} \quad \boldsymbol{X}_{ij} = (\boldsymbol{X}_0)_{ij}, \quad (i,j) \in \Omega. \tag{1.9b}$$

Where $\|\boldsymbol{X}\|_*$ is used as a proxy for the rank. In the case of (1.6), for a sufficiently small δ, if we let

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1/(2\delta) \\ \delta & 1 \end{pmatrix} \tag{1.10}$$

it can be easily verified that $\|\boldsymbol{X}\|_* < \|\boldsymbol{X}_0\|_*$, $\operatorname{rank}(\boldsymbol{X}) = 2$, and the nuclear norm minimization (1.9) thus does not return the unknown matrix $\boldsymbol{X}_0$ despite the fact that a sufficient amount of measurements are provided.

In the case of a symmetric, positive semidefinite matrix $\boldsymbol{X}$, program (1.9) reduces to the minimization of the trace of the matrix under the constraint $\boldsymbol{X} \succeq 0$. When the diagonal is fully measured, it is known that this formulation succeeds at recovering the original matrix $\boldsymbol{X}_0$ as soon as the completion problem is well-posed [15, 19].

1.3 Ridge regression has local minima

For completeness, this section briefly discusses ridge regression (i.e. Tikhonov regularization) for rank one matrix completion. This approach has gained in popularity over the last years and is equivalent to solving a regularized version of the non convex problem (1.1). In fact, it is natural to wonder whether the semidefinite programming formulation of this paper which relies on the minimization of the trace norm is not simply a form of ridge regression. This section precisely refutes this idea. The ridge regression formulation reads as

$$\min \quad \|\mathcal{R}_\Omega(\boldsymbol{x}\boldsymbol{y}^T) - \mathcal{R}_\Omega(\boldsymbol{X}_0 + \varepsilon)\|_F^2 + \lambda(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2) \tag{1.11}$$

The most popular way to solve formulation (1.11) is through gradient descent.

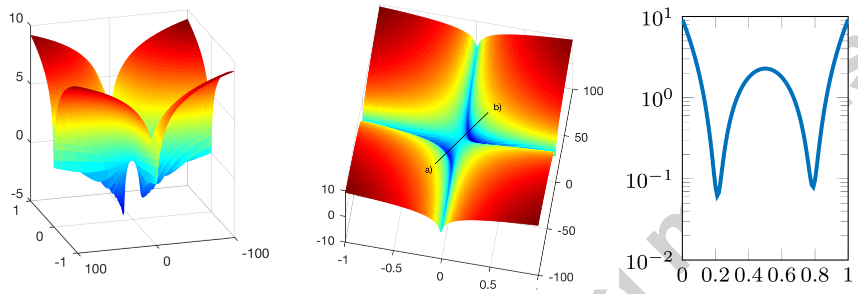

Fig. 1 Representation of the ridge regression energy landscape for the minimization problem $f(x, y) = \|xy - 1\|^2 + \|y - \delta\|^2$, plus regularization terms such as in (1.11). Without regularization, there is one infimum at $(-\infty, 0)$, and one minimum at $(\delta, 1/\delta)$ (In the figure above, we take $\delta = .1$). With regularization the infimum located at $(-\infty, 0)$ becomes a minimum and moves closer to the origin such as shown above. The evolution of the loss function on the line joining the two minima is displayed on the right.

As shown by Fig. 1, the landscape underlying this formulation suffers from a clear lack of convexity. Moreover, when several measurements are given and no convergence guarantee is known, it is not clear how to initialize the algorithm. In the numerical simulations of section 4, we follow standard practice and initialize $\boldsymbol{x}$ and $\boldsymbol{y}$ from the top singular vectors of the matrix $\sum_{ij}(\mathcal{P}_\Omega(\boldsymbol{X}))_{ij}\boldsymbol{e}_i\boldsymbol{e}_j^T$ weighted by their corresponding singular value. As the resulting experiments show, however, even in the absence of noise and with the aforementioned initialization, when the matrix size is sufficiently large and the number of measurements is sufficiently close to the recoverability limit, ridge regression will fail to return the global minimizer.

1.4 Lasserre hierarchy

Let $\mathbf{z} = (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^K$. When dealing with algebraic problems like (1.1), it will be useful to write those problems as general polynomial optimization problems

(POP) on $\boldsymbol{x}$ and $\boldsymbol{y}$ (or equivalently on $\boldsymbol{z}$) of the form

$$\mathbf{z}^* = \underset{\mathbf{z}}{\text{argmin}} \quad p_0(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^n \tag{1.12a}$$

$$\text{subject to} \quad h_1(\mathbf{z}) \geq 0,\ h_2(\mathbf{z}) \geq 0, \ldots, h_K(\mathbf{z}) \geq 0. \tag{1.12b}$$

For some polynomials $h_0(\mathbf{z}), h_1(\mathbf{z}), \ldots h_K(\mathbf{z}) \in \mathbb{R}[\mathbf{z}]$ where $\mathbb{R}[\boldsymbol{z}] = \mathbb{R}[z_1, \ldots, z_{m+n-1}]$ is used to denote the polynomial ring in K variables over the reals. In the case of problem (1.1), $h_1(z),\ h_2(z), \ldots, h_K(z)$, for $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$, thus denote the polynomial versions of the completion constraints, i.e. they encode polynomial equations of the form $x_i y_j = (X_0)_{ij}$ for $(i,j) \in \Omega$.

We will sometimes use the concise notation $\{h_j(\mathbf{z})\}_{j\in[J]}$ to denote the set of polynomial constraints. This set of constraints defines a semialgebraic set S of feasible points, which we write as

$$S = \{\mathbf{z} \in \mathbb{R}^K,\ |\ h_1(\mathbf{z}) = 0, \ldots, h_K(\mathbf{z}) = 0\} \tag{1.13}$$

For $\mathbf{z} \in \mathbb{R}^K$ and $\boldsymbol{\alpha} \in \mathbb{N}^K$, we introduce the multi-index notation $\boldsymbol{z}^{\boldsymbol{\alpha}} = z_1^{\alpha_1} z_2^{\alpha_2} \ldots z_K^{\alpha_K}$ with $|\boldsymbol{\alpha}| = \alpha_1 + \alpha_2 + \ldots + \alpha_K$, the degree of the monomial $\mathbf{z}^{\boldsymbol{\alpha}}$. In particular, we use $\mathbb{N}_t^K$ to denote the set of K-tuples $(\alpha_1, \alpha_2, \ldots, \alpha_K)$ on $\mathbb{N}$, with $|\boldsymbol{\alpha}| \leq t$. We will use $\mathbf{z}_{\mathcal{B}}$ to denote the sequence of all monomials in $\mathbf{z} \in \mathbb{R}^K$ for some standard ordering. Hence $\mathbf{z}_{\mathcal{B}} := (1, z_1, z_2\ , \ldots, z_1^2, z_1 z_2, \ldots, \mathbf{z}^{\boldsymbol{\alpha}}, \ldots)$. Similarly, we will use $\mathbf{z}_{\mathcal{B}}^d$ to denote the vector of all monomials $\mathbf{z}^{\boldsymbol{\alpha}}$ from the standard basis with degree bounded by d, i.e. $|\boldsymbol{\alpha}| \leq d$.

In this paper, polynomials will be alternatively represented as

- Weighted sums of monomials $p(z) = \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} \mathbf{z}^{\boldsymbol{\alpha}}$, where the $p_{\boldsymbol{\alpha}}$ thus denotes the multiplicative coefficient of the monomial $\mathbf{z}^{\boldsymbol{\alpha}}$ in $p(z)$.
- Vectors/sequences of coefficients $p = (p_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^n}$ such that $p(z) = \boldsymbol{p}^T \mathbf{z}_{\mathcal{B}}$
- Gram matrices, $\boldsymbol{Q}$, such that $p(z) = \langle \boldsymbol{Q}, \mathbf{z}_{\mathcal{B}} \mathbf{z}_{\mathcal{B}}^T \rangle = \mathbf{z}_{\mathcal{B}}^T \boldsymbol{Q} \mathbf{z}_{\mathcal{B}}$. In this case we will use the notation $\mathbf{Z}_{\mathcal{B}}$ (resp. $\mathbf{Z}_{\mathcal{B}}^d$) to represent the matrix generated from the standard basis as $\mathbf{Z}_{\mathcal{B}} = \mathbf{z}_{\mathcal{B}} \mathbf{z}_{\mathcal{B}}^T$ (resp. $\mathbf{Z}_{\mathcal{B}}^{2d} = \mathbf{z}_{\mathcal{B}}^d (\mathbf{z}_{\mathcal{B}}^d)^T$).

For a maximal degree d, and a probability measure $d\mu$, we let $m_{\boldsymbol{\alpha}} = \int \boldsymbol{z}^\alpha d\mu$ denote the moments associated to the measure $d\mu$. The moments matrix $\boldsymbol{M} = \boldsymbol{M}(m)$ associated to the measure $d\mu$ is the infinite dimensional matrix with rows and columns indexed by the K-tuples of $\mathbb{N}^K$ and such that $(\boldsymbol{M}(m))_{\boldsymbol{\alpha}+\boldsymbol{\beta}} = m_{\boldsymbol{\alpha}+\boldsymbol{\beta}}$ for any two multi-indices $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^K$. One can always truncate this matrix to moments of order at most $2d$. We denote this truncation as $\boldsymbol{M}^{(2d)}(m)$. In particular, the moments matrix $\boldsymbol{M}_0 = \boldsymbol{M}_0^{(4)} = \boldsymbol{m}_0 \boldsymbol{m}_0^T$, where $\boldsymbol{m}_0$ is the truncated Zeta vector, $(\boldsymbol{z}_0^{\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}|\leq 2}$, is simply given by $\int \mathbf{Z}_{\mathcal{B}}^4 d\mu$ for the one atomic measure $d\mu = \delta(\mathbf{z} - \mathbf{z}_0) d\mathbf{z}$. To further highlight the connection to [26, 28], for given polynomials $h_\ell(\mathbf{z})$, we also introduce the notion of "shifted" sequence $h_\ell m$ as $h_\ell m = \boldsymbol{M} \boldsymbol{h}_\ell$ where $\boldsymbol{h}_\ell$ is the infinite sequence of coefficients of the polynomial $h_\ell(z)$. The shifted sequence is thus the sequence obtained as $(h_\ell m)_{\boldsymbol{\alpha}} = \int h_\ell(z) \mathbf{z}^\alpha d\mu$ for any k-tuple $\boldsymbol{\alpha} \in \mathbb{N}^K$. Just as for the sequence of moments, m, one can define the moment matrix (i.e localizing matrix) associated to the shifted sequence, $\boldsymbol{M}(h_\ell m)$ and one can consider its truncation to elements of order at most $2d$, $\boldsymbol{M}^{(2d)}(h_\ell m)$. As before the matrix is defined as $(M(h_\ell m))_{\boldsymbol{\alpha}+\boldsymbol{\beta}} = (h_\ell m)_{\boldsymbol{\alpha}+\boldsymbol{\beta}}$.

For the general (finite) set of polynomials $h_1(z), h_2(z), \ldots, h_K(z)$, we let $\mathcal{I}$ denote the ideal generated by those polynomials, i.e.

$$\mathcal{I}(h_1,\ldots,h_K) := \left\{ \sum_{j\in[m]} p_j(z)h_j(z), \text{ for polynomials } p_j(z) \in \mathbb{R}[\mathbf{z}] \right\} \tag{1.14}$$

Equivalently, we will use $\mathcal{I}_d$ to denote the *truncated* ideal, whose maximal degree is bounded by d,

$$\mathcal{I}_d(h_1,\ldots,h_K) := \left\{ \sum_{j\in[m]} p_j(z)h_j(z),\ p_j(z) \in \mathbb{R}[\mathbf{z}],\, \deg(p_j) \leq d - \deg(h_j) \right\} \tag{1.15}$$

Given the matrix $\mathbf{Z}_{\mathcal{B}}^d = \mathbf{z}_{\mathcal{B}}^d(\mathbf{z}_{\mathcal{B}}^d)^T$ used to represent monomials of degree at most $2d$, when writing polynomials in matrix form, we will need to access monomials of a given degree. As an example, consider the univariate monomial basis $(1, z, z^2, z^3, \ldots)$. The corresponding matrix for the monomial basis truncated at degree 4 reads

$$\mathbf{Z}_{\mathcal{B}}^2 = \begin{pmatrix} 1 & z & z^2 \\ z & z^2 & z^3 \\ z^2 & z^3 & z^4 \end{pmatrix} \tag{1.16}$$

Now consider the polynomial $p(z) := z^2 - 1$. This polynomial can be defined from the matrix $\mathbf{Z}_{\mathcal{B}}^2$ by introducing appropriate matrices to access the monomials. Those matrices are simply assembled from the product of two canonical basis vectors. That is, for any degree γ, one access the monomial z^γ in $\mathbf{Z}$ by means of the matrices $\boldsymbol{e}_\alpha \boldsymbol{e}_\beta^T$ for any α, β such that $\gamma = \alpha + \beta$. For any such matrix, we have

$$\langle \mathbf{Z}_{\mathcal{B}}^2, \boldsymbol{e}_\alpha \boldsymbol{e}_\beta^T \rangle = z^\gamma.$$

In particular, using those matrices, the polynomial $p(z)$, $z \in \mathbb{R}$, with associated Gram matrix $\boldsymbol{Q}$ can read as

$$p(z) = \langle \boldsymbol{Q}, \mathbf{Z}_{\mathcal{B}}^2 \rangle = \sum_{\gamma \geq 0} \sum_{\alpha+\beta=\gamma} C_\gamma p_\gamma \langle \mathbf{Z}_{\mathcal{B}}^2, \boldsymbol{e}_\alpha \boldsymbol{e}_\beta^T \rangle. \tag{1.17}$$

C_γ is a normalizing constant defined for each degree γ as

$$C_\gamma = \frac{1}{|\{(\alpha,\beta)|\ \alpha+\beta=\gamma\}|}$$

To write expression (1.17) compactly, we introduce auxiliary matrices $\boldsymbol{B}_\gamma$ relative to each of the monomials z^γ, defined as

$$\boldsymbol{B}_\gamma = \sum_{\alpha+\beta=\gamma} \boldsymbol{e}_\alpha \boldsymbol{e}_\beta^T. \tag{1.18}$$

Using those matrices, the polynomial $p(\mathbf{z})$ can now read directly as

$$p(z) = \sum_{\gamma} C_\gamma p_\gamma \langle \mathbf{Z}_{\mathcal{B}}^2, \boldsymbol{B}_\gamma \rangle.$$

The constant C_γ can now also be expressed more simply as $C_\gamma = \frac{1}{\|\boldsymbol{B}_\gamma\|_F^2}$. As an example, the matrix $\boldsymbol{B}_2$ used to access the monomial z^2 in (1.16) reads

$$\boldsymbol{B}_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{1.19}$$

When dealing with polynomials on $\mathbb{R}^K$, the same idea applies and we will denote the corresponding matrices as $\boldsymbol{B}_{\boldsymbol{\gamma}}$ where $\boldsymbol{\gamma}$ is the multi-index used to access the monomial $\mathbf{z}^{\boldsymbol{\gamma}}$. Finally, note that the coefficients $p_{\boldsymbol{\gamma}}$ can read from the matrices $\boldsymbol{B}_{\boldsymbol{\gamma}}$ and the Gram matrix $\boldsymbol{Q}$ of $p(z)$ as $p_\gamma = \langle \boldsymbol{Q}, \boldsymbol{B}_\gamma \rangle$.

We are now ready to introduce the formulation of the Lasserre hierarchy as it appears in [26, 28].

Definition 1 The level-d Lasserre relaxation, $\mathbb{Q}_K^{2d}$, for a set of polynomial constraints $h_1(z) = 0, \ldots, h_K(z) = 0$ with minimization of the Trace norm, is the following semidefinite programming relaxation

$$\mathbb{Q}_K^{2d} \begin{cases} \min \operatorname{Tr}(\boldsymbol{M}_K^{2d}(m)), \\ \boldsymbol{M}_K^{2d}(m) \succeq 0, \\ \boldsymbol{M}_K^{2d-d_\ell}(h_\ell m) = 0, \; \ell = 1, \ldots, L \end{cases} \tag{1.20}$$

The set of constraints $\boldsymbol{M}_K^{2d-d_j}(h_j m) = 0$ can be expressed from the moments matrix $\boldsymbol{M}_K^{2d}$ and the matrices $\boldsymbol{B}_{\boldsymbol{\gamma}}$ as

$$\left\{ \boldsymbol{M}_K^{2d-d_\ell}(h_\ell m) = 0, \; 1 \leq \ell \leq K \right\} \tag{1.21}$$

$$= \left\{ \sum_{\boldsymbol{\zeta} \in \mathbb{N}^K} \frac{(h_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2} \langle \boldsymbol{M}^{2d}, \boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}} \rangle = 0, \; \boldsymbol{\kappa} \in \mathbb{N}_{2d-d_\ell}^K, 1 \leq \ell \leq L \right\} \tag{1.22}$$

Formulation (1.20) should be understood as an extension to nuclear norm minimization. When minimizing the nuclear norm of rank one matrices, one only enforces constraints on monomials/moments of degree at most two on the entries of the generating vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. The nuclear norm was shown in [16] to be equivalent to the following semidefinite program,

$$\begin{aligned} \text{minimize} \quad & \operatorname{Tr}(\boldsymbol{W}) \\ \text{subject to} \quad & \boldsymbol{W} = \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{X}_{21} & \boldsymbol{X}_{22} \end{bmatrix}, \; \boldsymbol{W} \succeq 0, \\ & (\boldsymbol{X}_{12})_{ij} = (X_0)_{ij}, \quad (i,j) \in \Omega. \end{aligned} \tag{1.23}$$

When $\boldsymbol{X}_0 = \boldsymbol{x}_0\boldsymbol{y}_0^T$, the matrix $\boldsymbol{W}$ is thus a proxy for the rank one matrix

$$\boldsymbol{W}_0 = \begin{bmatrix} \boldsymbol{x}_0\boldsymbol{x}_0^T & \boldsymbol{x}_0\boldsymbol{y}_0^T \\ \boldsymbol{y}_0\boldsymbol{x}_0^T & \boldsymbol{y}_0\boldsymbol{y}_0^T \end{bmatrix} \tag{1.24}$$

The positive semidefinite constraint on $\boldsymbol{X}$ is used in combination with the trace norm, as a convex relaxation of the rank one constraint. As said earlier, formulation (1.23) only optimizes over monomials of bidegree $(1, 1)$. The key idea of the level 2 Lasserre relaxation is to extend this type of formulation to monomials of higher degree in the original unknowns. Introducing $\mathbf{z}_0 = (\boldsymbol{x}_0, \boldsymbol{y}_0)$ and $\mathbf{z}_0^{(2)} = \text{vec}(\mathbf{z}_0^{\otimes^2}) = \text{vec}(\mathbf{z}_0 \otimes \mathbf{z}_0)$, the second level of (1.20) considers the larger matrix $\boldsymbol{M}_K^4(m_0) = \boldsymbol{M}_0$ defined as

$$\boldsymbol{M}_0 = \begin{bmatrix} 1 & \mathbf{z}_0 & \mathbf{z}_0^{(2)} \\ \mathbf{z}_0 & \mathbf{z}_0^{\otimes^2} & \mathbf{z}_0 \otimes \mathbf{z}_0^{(2)} \\ \mathbf{z}_0^{(2)} & \mathbf{z}_0^{(2)} \otimes \mathbf{z}_0 & \mathbf{z}_0^{(2)} \otimes \mathbf{z}_0^{(2)} \end{bmatrix}, \tag{1.25}$$

In (1.25), we thus have $\boldsymbol{W}_0 = \mathbf{z}_0^{\otimes^2}$. At order $d = 2$, the semidefinite relaxation (1.20) considers as unknowns all the entries of a positive semidefinite proxy $\boldsymbol{M}$ of the same structure as $\boldsymbol{M}_0$, but without the explicit link to a vector $\mathbf{z}_0$. Instead, two categories of linear constraints are intended to force the matrix $\boldsymbol{M}$ to inherit the structure of $\boldsymbol{M}_0$:

- *Structural constraints/ total symmetry.* Due to the additional monomials that appear in (1.25), there now exist corresponding additional relations between the entries of $\boldsymbol{M}_0$ (and thus $\boldsymbol{M}$ as well). In particular, all monomials in $\mathbf{z}_0^{(2)}$ find an exact match in the block $\mathbf{z}_0^{\otimes^2}$. Within the block $\mathbf{z}_0^{(2)} \otimes \mathbf{z}_0^{(2)}$, one must also list all the total symmetry constraints of a tensor of order 4. More generally, the structural constraints enforce, on the unknown matrix $\boldsymbol{M}$, equality of the entries that are identical in the rank-one matrix $\boldsymbol{M}_0$. Note that in (1.20), those constraints are implicitly enforced through the "moments" structure of the matrix $\boldsymbol{M}_K^{2d}$
- *Higher-order affine constraints.* Similarly, for any of the original affine constraints involving the elements of $\mathbf{z}_0^{\otimes^2}$, one can now define higher order constraints that are jointly enforced on the elements of $\mathbf{z}_0$ and the elements of the block $\mathbf{z}_0^{(2)} \otimes \mathbf{z}_0$. As an example, consider that one is given the constraint $X_{ij} = x_i y_j = (X_0)_{ij}$. It is now possible to enforce the constraints $x_i y_j x_k - (X_0)_{ij} x_k$ for any monomial x_k. More generally, the higher-order constraints are obtained by multiplying the original constraints by any product of the entries of x and y of degree at most two. They are encoded in (1.22)

Note that since formulation (1.20) is a relaxation of problem (1.1) (see [26, 28]), the moments matrix $\boldsymbol{M}_0$ generated from $\mathbf{z}_0$ and introduced in (1.25) is always a feasible solution for this formulation. Moreover, since any of the formulations in this paper follow a minimization of the trace over the PSD cone, the infimum is always attained (i.e the recession direction corresponding to the trace minimization goes towards the origin (0 trace) and leaves the PSD cone, whereas any recession direction of the PSD cone necessarily moves away from the origin, in the opposite direction). As indicated in [36], Theorem 27.3, the objective and the feasible set do not share any recession direction and the infimum is always achieved on the feasible set.

We also introduce the robust version of (1.20)

$$\min \operatorname{Tr}(\boldsymbol{M}_K^{(2d)}), \text{ s.t. } \boldsymbol{M}_K^{2d}(m) \succeq 0, \sqrt{\sum_{\ell=1}^{K} \left\| \boldsymbol{M}_K^{2d-d_\ell}(\tilde{h}_\ell m) \right\|_F^2} \leq \eta. \tag{1.26}$$

The constraints in formulation (1.26) are thus decomposed into a structural part (implicitely encoded in the moments structure of $\boldsymbol{M}_K^{(2d)}$) which is unaffected by the noise, and the affine constraints encoded by the localizing matrices $\boldsymbol{M}_K^{2d-2d_\ell}(\tilde{h}_\ell m)$ which are affected by noise, hence the notation $\tilde{h}_\ell$. The hierarchy (1.26) thus uses those noisy constraints in an essential way, not just as a right handside, as they are being multiplied by monomials to generate the entries of $\boldsymbol{M}_K^{2d-d_\ell}(\tilde{h}_\ell m)$. Just as for (1.23), as soon as we take η sufficiently large (i.e. $\eta \geq \|\varepsilon\|_2\sqrt{(1+\|\mathbf{z}_0\|_2^2+\|\mathbf{z}_0\|_2^4|)}$), the rank one moments matrix $\boldsymbol{M}_0$ is a feasible point for (1.26).

We are now ready to state the main results of the paper.

1.5 Main Result

The main result of this paper only requires the necessary and sufficient conditions of lemma 1 to be satisfied. Our first theorem states that in the noiseless framework, the level 2 Lasserre relaxation completely solves problem (1.1).

Theorem 1 *Consider problem (1.1), with $\boldsymbol{X} \in \mathcal{M}^*(1; m \times n)$, rank one, and a mask $\mathcal{R}_\Omega$ that is injective on $\mathcal{M}^*(1; m \times n)$. This problem is solved exactly through the level 2 Lasserre relaxation with minimization of the trace norm, i.e the relaxation always outputs the rank one matrix $\boldsymbol{X}_0$.*

The first novelty of Theorem 1 is that it is completely independent of the magnitude of the entries of $\boldsymbol{X}_0$. In particular, it does not require any incoherence assumptions.

The interest of semidefinite programs lies in their robustness vis-a-vis corruption of the data. This is what Theorem 2 below makes precise. It shows that when considering observations that are corrupted by a noise ε with bounded ℓ_2 norm satisfying $\|\varepsilon\|_2\sqrt{(1+\|\mathbf{z}_0\|_2^2+\|\mathbf{z}_0\|_2^4|)} \leq \eta$, so that $(\tilde{X}_0)_{ij} = (X_0)_{ij} + \varepsilon_{ij}$, the solution to the semidefinite program (1.26) remains within the noise level. We use $\delta_L > 0$ and $\delta_U > 0$ to denote the positive lower and upper bounds on the entries of $\mathbf{z}_0$. That is to say $\delta_L \leq |(z_0)_i| \leq \delta_U$. We prove the following estimate

Theorem 2 *Let $\boldsymbol{M}_0$ denote the rank one matrix introduced in (1.25) for $\mathbf{z}_0 \in \mathbb{R}^K$. Assume that the necessary and sufficient conditions of lemma 1 are satisfied.*

There exists a constant $C(\delta_L, \delta_U)$ *such that the solution* $\boldsymbol{M}$ *to the semidefinite program* (1.26) *obeys*

$$\frac{\|\boldsymbol{M} - \boldsymbol{M}_0\|_F}{\|\boldsymbol{M}_0\|_F} \leq C(\delta_L, \delta_U) \cdot (m+n)^{5/2} \eta. \tag{1.27}$$

The constant C *depends on the magnitude of the entries of* $\boldsymbol{X}_0$*, but not on* Ω*,* m, n*, or* ε*.*

Once $\boldsymbol{M}$ is found, one can read off $\boldsymbol{X}$ from the entries of $\boldsymbol{M}$ corresponding to $|\boldsymbol{\alpha}| = 1$. Note that for the propagation and log-system algorithms, a similar error bound can only be expected to hold provided $\|\boldsymbol{\varepsilon}\|_2 \leq C \min_{i,j} |(X_0)_{ij}|$ for some $C < 1$, and would otherwise become unbounded.

Most of the $(m+n)^{5/2}$ multiplicative factor in Theorem 2 arises because of the propagation of noise though the certificate (i.e. the fact that the certificate relies on a chain of length $m+n$). In the particular case of problem (1.1), it is possible to reduce this factor by relying on the sparsity of the constraints and by considering a smaller semidefinite program that only retains the order 2 monomials which appear in those constraints. Moreover, if the path between the root nodes (i.e the nodes corresponding to the degree one constraints) and any other vertex in the graph, is explicitly given, it is possible to bound the noise along those paths only rather than on the whole set of constraints of the form $h_\ell(\mathbf{z})\mathbf{z}^{\boldsymbol{\alpha}}$ for all $|\boldsymbol{\alpha}| \leq 2$.

Those ideas are summarized through Corollary 1 below. For any of the chains $\mathcal{P}_i$ of the form (1.4), connecting a root node (i.e. a degree one constraint) to any of the leaf nodes in the bipartite graph $\mathcal{G}(\mathcal{V}_1, \mathcal{V}_2)$, we label the constraints appearing in the chain as $h_{i_1,i_2}, h_{i_2,i_3}, \ldots$, or equivalently $h_{\ell_1}, h_{\ell_2}, \ldots$ For any of those constraints, we further let $\mathcal{H}_{i_k,i_{k+1}} = \mathcal{H}_{\ell_k} = \left\{z_{i_{k-1}}, z_{i_{k+2}}\right\}$. We also define $\mathcal{H}^0_{\ell_k}$ as $\mathcal{H}^0_{\ell_k} = \left\{(z_0)_{i_{k-1}}, (z_0)_{i_{k+2}}\right\}$. We observe entries $(\tilde{X}_0)_{ij} = (X_0)_{ij} + \varepsilon_{ij}$ where ε is a noise term with bounded ℓ_2 norm satisfying

$$\max_i \sqrt{\sum_{\ell \in \mathcal{P}_i} |\varepsilon_\ell|^2 (1 + \sum_{a \in \mathcal{H}^0_\ell} |a|^2)} \leq \eta' \tag{1.28}$$

Corollary 1 *Assume that the paths relating any of the root nodes to the leaf nodes in* $\mathcal{G}(\mathcal{V}_1, \mathcal{V}_2)$ *are explicitly given and that the noise along those paths is bounded by* η'*. There exists a reduced version of the level-2 Lasserre hierarchy and a constant* $C(\delta_L, \delta_U)$ *such that the solution to the corresponding semidefinite program obeys*

$$\frac{\|\boldsymbol{M} - \boldsymbol{M}_0\|_F}{\|\boldsymbol{M}_0\|_F} \leq C(\delta_L, \delta_U) \cdot (m+n)^2 \eta'. \tag{1.29}$$

The constant C *depends on the magnitude of the entries of* $\boldsymbol{X}_0$*, but not on* Ω*,* m, n*, or* ε*.*

1.6 Connections with existing work

Low-rank matrix completion has led to a number of results over the last few years. Among those results, nuclear norm minimization in particular, has been used to generate important recovery guarantees [7, 9, 10, 12, 34]. Other notable progress on the question includes the results of Keshavan et al. [20, 21, 22] which certify recovery with high probability (w.h.p.) given $O((m+n)\log(m+n))$ measurements with an error bound which scales linearly in the noise as $O(m+n)$ provided that both the magnitude of the entries and the number of measurements are sufficiently large. Singer et al. [40] investigate matrix completion with a non random sampling mask based on the structure of the measurements. Their paper is interested in determining whether completion is possible or not in the general rank-r case using rigidity theory. Other papers that focus on characterizing the sampling patterns enabling matrix completion include [23, 24, 25]. Those papers show that feasibility and uniqueness of the completion only depends on the structure of the measurements. They propose an algorithm based on the completion of the k-by-k minors (circuit polynomials). The derivation of stable, deterministic recovery guarantees on the one hand, and the connection between tractable recovery/completion and the structure of the measurements graph on the other, remain open problems. Along that line, [15, 19] solves the symmetric rank one completion problem when the diagonal entries are given. The noiseless result of this paper was presented in the introductory note [13].

The use of semidefinite programming as a relaxation to hard/non convex problems was popularized through the pioneering work of Goemans and Williamson on MaxCut [17]. The semidefinite program which we use in this paper is in fact an instance of the more general hierarchies of semidefinite programs [28] which were introduced through the work of Parrilo [32, 33], Shor [37, 38, 39], Nesterov [29], and Lasserre [26] as an extension to the original semidefinite programming relaxation used on MaxCut. An updated formulation based on the Jacobian of the polynomial constraints was later presented in [30]. Semidefinite hierarchies are based on making the semidefinite programs gradually tighter by adding more variables and constraints, resulting in optimization in gradually higher dimensional subspaces. Examples of successful developments based on semidefinite programming or nuclear norm relaxations of hard/non convex problems can be found in [2, 8, 10, 11, 15].

An important downside of those hierarchies has been the lack of convincing instances for which levels higher than one were leading to noticeable enhancements. A few improvements have however been made over the last few years. On the first point, in a paper which is related to this one, Tang et al. [41] show that the tensor decomposition problem can be solved through a semidefinite programming relaxation corresponding to the lowest level of a semidefinite hierarchy. In [3], Barak et al. use the Rademacher complexity to show that the tensor completion problem can be solved with high probability with $\widetilde{O}(n^{3/2})$ measurements through a level-6 semidefinite programming relaxation. In [14] Nie et al. show that computation of the real eigenvalues of symmetric tensors can be achieved at a finite level of the hierarchy. Finally in [18], Gouveia et al. discuss the connection between the k^{th}

theta body of an ideal $\mathcal{I}$ and the set of solutions of the k^{th} Lasserre relaxation of $\text{conv}(\mathcal{V}_{\mathbb{R}}(\mathcal{I}))$.

A few papers address possible reductions in the complexity of semidefinite programming hierarchies based on the structure of the polynomial constraints. Among those papers, Lasserre [27] as well as Nie et al. [31] introduce tailored relaxations for problems where sparsity occurs in the constraints and Ahmadi [1] discusses possible reductions in the complexity of the hierarchies by means of the chordal extension of the graph whose cliques are defined from the polynomial constraints.

2 Proof of Theorem 1.

To ensure unique recovery of the matrix $\boldsymbol{M}_0$ from the second level ($d = 2$) of the semidefinite hierarchy (1.20), traditional convex optimization proofs are based on satisfying first order optimality conditions[1] by exhibiting a dual vector λ such that $-\mathcal{A}^*\lambda - I \in \partial\imath_{\mathcal{K}}(\boldsymbol{M}_0)$ where $\imath_{\mathcal{K}}$ denotes the indicator function of the positive semidefinite (PSD) cone (see for example [10]) and $\mathcal{A}$ is the linear map encoding all the constraints. In section 2.1 below, we start by giving a more detailed characterization of problem (1.20), for $d = 2$, in terms of the constraints. We then provide the general conditions for the existence of such a certificate. In section 2.2, we show how satisfiability of these conditions can be reduced to the construction of a dual polynomial with particular structure. Section 2.3 finally shows how such a dual polynomial can be constructed.

2.1 Dual certificate

In this section, we give an explicit expression for the general condition $-\mathcal{A}^*\lambda - \boldsymbol{I} \in \partial\imath_{\mathcal{K}}(\boldsymbol{M}_0)$ on the dual vector λ certifying optimality in the case of problem (1.20) with $d = 2$. We then show how this condition can be extended to ensure uniqueness in addition to optimality at $\boldsymbol{M}_0$. We now give the detailed expression of the constraints in (1.20). As before, we let $h_1(z),\ h_2(z), \ldots, h_K(z)$, for $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$, to denote the polynomial versions of the completion constraints. Those constraints therefore encode polynomial equations of the form $x_i y_j = (X_0)_{ij}$ for $(i, j) \in \Omega$.

Recall that each matrix $\boldsymbol{B}_{\boldsymbol{\gamma}}$ can be decomposed into a sum of elementary matrices $\boldsymbol{E}_{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2}$ with only a single non zero entry, for multi-indices $\boldsymbol{\gamma}_1,\ \boldsymbol{\gamma}_2 \in \mathbb{N}^K$, i.e., $\boldsymbol{B}_{\boldsymbol{\gamma}} = \sum_{\boldsymbol{\gamma}_1+\boldsymbol{\gamma}_2=\boldsymbol{\gamma}} \boldsymbol{E}_{\boldsymbol{\gamma}_1,\gamma_2}$ with $\boldsymbol{E}_{\boldsymbol{\gamma}_1,\gamma_2} = \boldsymbol{e}_{\boldsymbol{\gamma}_1}\boldsymbol{e}_{\boldsymbol{\gamma}_2}^T$. We solve the following

[1] Note that in the case of convex optimization those conditions are necessary and sufficient.

particular instance of (1.20),

$$\mathbb{Q}_K^4 \begin{cases} \min \ \operatorname{Tr}(\boldsymbol{M}_K^4(m)), \\ \qquad \boldsymbol{M}_K^4(m) \succeq 0, \\ \boldsymbol{M}_K^{4-d_\ell}(h_\ell m) = 0, \ \ell = 1, \dots, L \end{cases} \tag{2.1}$$

The constraints $\boldsymbol{M}_K^{4-d_\ell}(h_\ell m)$ can expand using the $\boldsymbol{B}_{\boldsymbol{\gamma}}$ as

$$\sum_{\boldsymbol{\zeta} \in \mathbb{N}^K} \frac{(h_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2} \langle \boldsymbol{M}_K^4, \boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}} \rangle = 0, \ \boldsymbol{\kappa} \in \mathbb{N}^K_{4-d_\ell}, 1 \le \ell \le L \tag{2.2}$$

In the proof of Theorem 1, we will write the structural constraints (i.e the implicit constraints $(\boldsymbol{M}_K^4)_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} = (\boldsymbol{M}_K^4)_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}$ for any $\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$ resulting from the fact that $\boldsymbol{M}_K^4$ is a moments matrix) together with the normalization constraint $\boldsymbol{M}_{11} = 1$ compactly as $\boldsymbol{M} = \sum_\gamma m_\gamma \boldsymbol{B}_\gamma + \boldsymbol{e}_1 \boldsymbol{e}_1^T$ by introducing additional variables m_γ. This enables us to explicitly enforce those structural constraints. The resulting structure of $\boldsymbol{M}$ is Hankel-type (and would be exactly Hankel in the one-dimensional case as we saw earlier). The sum in (2.2) is taken over all the coefficients of each constraint $h_\ell(x) = 0$, $h_\ell(x) = \sum_{\boldsymbol{\zeta}} (h_\ell)_{\boldsymbol{\zeta}} \mathbf{x}^{\boldsymbol{\zeta}}$.

We now derive the first order optimality conditions $-\mathcal{A}^*\lambda - \boldsymbol{I} \in \partial \imath_{\mathcal{K}}(\boldsymbol{M}_0)$ for problem (2.1) in terms of the Lagrangian dual function $\mathcal{L}$. Introducing multipliers for each of the polynomial constraints, the Lagrangian can be written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{M}, \boldsymbol{m}, \boldsymbol{\lambda}, \boldsymbol{\xi}) &= \operatorname{Tr}(\boldsymbol{M}) + \langle \boldsymbol{M} - \sum_{\boldsymbol{\gamma}} m_{\boldsymbol{\gamma}} \boldsymbol{B}_{\boldsymbol{\gamma}} - \boldsymbol{e}_1 \boldsymbol{e}_1^T, \boldsymbol{\xi} \rangle \\ &+ \sum_\ell \sum_{\boldsymbol{\kappa} \in \mathbb{N}^K_{2(t-d_{h_\ell})}} \lambda_{\ell,\boldsymbol{\kappa}} \left(\sum_{\boldsymbol{\zeta} \in \mathbb{N}^K} \frac{(h_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2} \langle \boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}, \boldsymbol{M} \rangle \right) \\ &+ \imath_{\mathcal{K}}(\boldsymbol{M}). \end{aligned} \tag{2.3}$$

The multipliers $\lambda_{\ell,\boldsymbol{\kappa}}$ correspond to each of the original and shifted polynomial constraints while $\boldsymbol{\xi}$ encode the Hankel-type structure of the matrix $\boldsymbol{M}$. Let T denote the tangent space

$$T = \left\{ \boldsymbol{m}_0 \boldsymbol{v}^T + \boldsymbol{v} \boldsymbol{m}_0^T, \quad \boldsymbol{v} \in \mathbb{R}^{|\mathbb{N}_2^K|} \right\}, \tag{2.4}$$

$T^\perp$ being its orthogonal complement, and let Y_T denote the projection of the matrix Y onto the subspace T. Usual convex optimization theory states that $\boldsymbol{M}_0 = \boldsymbol{m}_0 \boldsymbol{m}_0^T$ is a minimizer for problem (2.1) if and only if one can find dual vectors $(\boldsymbol{\xi}, \boldsymbol{\lambda})$ such that $0 \in \partial \mathcal{L}(\boldsymbol{M}_0, \boldsymbol{m}_0, \boldsymbol{\lambda}, \boldsymbol{\xi})$. The dual variables $\boldsymbol{\xi}, \boldsymbol{\lambda}$ combine into a dual certificate Z, and must obey the following three conditions.

1) $Y = \boldsymbol{I} - \boldsymbol{\xi} - \sum_\ell \sum_{\boldsymbol{\kappa} \in \mathbb{N}^K_{2(t-d_{h_\ell})}} \lambda_{\ell\boldsymbol{\kappa}} \left(\sum_{\boldsymbol{\zeta}} \frac{(h_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2} \boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}} \right)$

2) $Y_T = 0, \quad Y_{T^\perp} \succeq 0$

3) $\langle \boldsymbol{B}_{\boldsymbol{\gamma}}, \boldsymbol{\xi}\rangle = 0, \qquad \forall\, \boldsymbol{\gamma} \neq 0.$

Conditions 1 and 2 are obtained by requiring that the derivative of the Lagrangian with respect to the moments matrix $\boldsymbol{M}$ belongs to the normal cone (subdifferential of the indicator of the PSD cone) at $\boldsymbol{M}_0$. Condition 3 is obtained by requiring that the derivative of the Lagrangian with respect to the vector of moments, $\boldsymbol{m}$, vanishes. Note that this last condition only requires orthogonality of $\boldsymbol{\xi}$ with respect to the $\boldsymbol{B}_{\boldsymbol{\gamma}}$ for $\boldsymbol{\gamma} \neq 0$. This implies that $\boldsymbol{\xi}$ can contain any constant term of the form $\rho \boldsymbol{e}_1 \boldsymbol{e}_1^T$ for any constant ρ. Together, conditions 1) to 3) can be understood as a version of the Kuhn-Tucker conditions which read

$$\mathcal{A}(\boldsymbol{M}) = b,\ (I - \mathcal{A}^*\boldsymbol{\lambda})\boldsymbol{M} = 0,\ \boldsymbol{M} \succeq 0,\ \boldsymbol{I} - \mathcal{A}^*\boldsymbol{\lambda} \succeq 0. \tag{2.5}$$

We decide to write the Lagrangian with the indicator function of the PSD cone which simply leads to the replacement of the complementary slackness conditions (second condition in (2.5)) by the subspace conditions $\boldsymbol{Y}_T = 0$ and $\boldsymbol{Y}_T^\perp \succeq 0$.

The following proposition guarantees unique recovery in addition to the optimality ensured by conditions 1) to 3) above.

Proposition 1 *To ensure unique recovery of $\boldsymbol{M}_0$ through the relaxation* (2.1)*, in addition to the conditions 1), 2), and 3) mentioned above, it is sufficient to require $Y_{T^\perp} \succ 0$ along with the injectivity on T of all the linear constraints $\mathcal{A}(\boldsymbol{M}) = \boldsymbol{b}$ arising from the measure version of the polynomial constraints $h_\ell(\boldsymbol{z}) = 0$ and the structure of the moments matrix.*

Proof We now use the decompositions $Y = \boldsymbol{I} - Y_2 = I - Y_1 - \boldsymbol{\xi}$, where Y_1 and Y_2 are defined as

$$Y_2 = \boldsymbol{\xi} + \sum_\ell \sum_{\boldsymbol{\kappa} \in \mathbb{N}^K_{2(t-d_{h_\ell})}} \lambda_{\ell\boldsymbol{\kappa}} \left(\sum_{\boldsymbol{\zeta}\in\mathbb{N}^K} \frac{(h_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2} \boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}} \right) = Y_1 + \boldsymbol{\xi} \tag{2.6}$$

From this, one can lower bound the trace of any solution (2.1) as

$$\begin{aligned}
\operatorname{Tr}(\boldsymbol{M}_0) &= \langle I, \boldsymbol{M}_0\rangle = \langle I_T, \boldsymbol{M}_0\rangle = \langle (Y_2)_T, \boldsymbol{M}_0\rangle && (2.7)\\
&= \langle Y_2, \boldsymbol{M}_0 - \boldsymbol{M}\rangle + \langle Y_2, \boldsymbol{M}\rangle = \langle Y_2, \boldsymbol{M}\rangle && (2.8)\\
&= \langle I_T, \boldsymbol{M}_T\rangle + \langle (Y_2)_{T^\perp}, \boldsymbol{M}\rangle && (2.9)\\
&= \operatorname{Tr}(\boldsymbol{M}_T) + \langle (Y_2)_{T^\perp}, \boldsymbol{M}\rangle && (2.10)\\
&< \operatorname{Tr}(\boldsymbol{M}) \qquad \text{for } \boldsymbol{M}_{T^\perp} \neq 0 && (2.11)
\end{aligned}$$

In (2.8), we use $\langle \boldsymbol{\xi}, \boldsymbol{M} - \boldsymbol{M}_0\rangle = 0$ allong with the fact that Y_1 belongs to the range of $\mathcal{A}^*$ and $\mathcal{A}(\boldsymbol{M}) = \mathcal{A}(\boldsymbol{M}_0)$. The last inequality follows from $(Y_2)_{T^\perp} \prec I_{T^\perp}$ which, since we take $\boldsymbol{M} \succeq 0$, implies $\langle (Y_2)_{T^\perp}, \boldsymbol{M}\rangle < \operatorname{Tr}(\boldsymbol{M}_{T^\perp})$ for $\boldsymbol{M}_{T^\perp} \neq 0$ for any solution $\boldsymbol{M}$ to the semidefinite program (2.1). This last inequality in turn implies that any optimal solution must satisfy $\boldsymbol{M}_{T^\perp} = 0$. Finally $\boldsymbol{M}_T = (\boldsymbol{M}_0)_T$ by injectivity of the constraints on T.

Note that to satisfy $Y_{T^\perp} \succ 0$ and $Y_T = 0$, it is sufficient to ask for $\boldsymbol{m}_0 \in \mathrm{Null}(Y)$ and to require Y to be positive semidefinite and exact rank $|\mathbb{N}_2^K| - 1$.

To see this, note that if Y is rank $|\mathbb{N}_2^K| - 1$ with a dimension 1 nullspace given by $\boldsymbol{m}_0$, we can write $Y = \sum_{j=1}^{|\mathbb{N}_2^K|-1} \lambda_j \boldsymbol{v}_j \boldsymbol{v}_j^T$ with $\langle \boldsymbol{v}_j, \boldsymbol{m}_0\rangle = 0$. Let $\tilde{\boldsymbol{m}}_0 = \frac{\boldsymbol{m}_0}{\|\boldsymbol{m}_0\|}$. From this we have

$$Y_T = \mathcal{P}_T(Y) = \tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^* Y + Y\tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^* - \tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^* Y\tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^* = 0 \tag{2.12}$$

which follows from the nullspace assumption. Since $\mathrm{rank}(Y) = |\mathbb{N}_2^K| - 1$, any other vector $\boldsymbol{w}$ with $\langle \boldsymbol{w}, \boldsymbol{m}_0\rangle = 0$ can always read as $\boldsymbol{w} = \sum_{j=1}^{|\mathbb{N}_2^K|-1} \alpha_j \boldsymbol{v}_j$ where $\boldsymbol{v}_j$ are the non trivial eigenvectors of Y. More generally, for any vector $\boldsymbol{x}$, we also have $\boldsymbol{x} = \beta_0\tilde{\boldsymbol{m}}_0 + \sum_{j=1}^{|\mathbb{N}_2^K|-1} \beta_j \boldsymbol{v}_j = \beta_0\tilde{\boldsymbol{m}}_0 + \boldsymbol{w}$. From this, we have

$$\begin{aligned}\langle Y_T^\perp, \boldsymbol{x}\boldsymbol{x}^T\rangle &= \langle Y, \mathcal{P}_T^\perp(\boldsymbol{x}\boldsymbol{x}^T)\rangle && (2.13)\\ &= \langle Y, \boldsymbol{x}\boldsymbol{x}^T - \tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^T\boldsymbol{x}\boldsymbol{x}^T - \boldsymbol{x}\boldsymbol{x}^T\tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^T + \tilde{\boldsymbol{m}}_0|\langle\tilde{\boldsymbol{m}}_0, \boldsymbol{x}\rangle|^2\tilde{\boldsymbol{m}}_0^T\rangle && (2.14)\end{aligned}$$

Now use

$$\begin{aligned}\langle Y, \boldsymbol{x}\boldsymbol{x}^T\rangle &= \langle Y, \beta_0^2\tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^T + \beta_0\tilde{\boldsymbol{m}}_0\boldsymbol{w}^T + \beta_0\boldsymbol{w}\tilde{\boldsymbol{m}}_0^T + \boldsymbol{w}\boldsymbol{w}^T\rangle && (2.15)\\ -\langle Y, \tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^T\boldsymbol{x}\boldsymbol{x}^T\rangle &= -\langle Y, \beta_0^2\tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^T + \beta_0\tilde{\boldsymbol{m}}_0\boldsymbol{w}^T\rangle && (2.16)\\ -\langle Y, \boldsymbol{x}\boldsymbol{x}^T\tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^T\rangle &= -\langle Y, \beta_0^2\tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^T + \beta_0\boldsymbol{w}\tilde{\boldsymbol{m}}_0^T\rangle && (2.17)\\ \langle Y, \tilde{\boldsymbol{m}}_0|\langle\tilde{\boldsymbol{m}}_0, \boldsymbol{x}\rangle|^2\tilde{\boldsymbol{m}}_0^T\rangle &= \beta_0^2\langle Y, \tilde{\boldsymbol{m}}_0\tilde{\boldsymbol{m}}_0^T\rangle && (2.18)\end{aligned}$$

Combining (2.15) to (2.18) and substituting those equations into (2.14), we can write $\langle Y_T^\perp, \boldsymbol{x}\boldsymbol{x}^T\rangle$ as

$$\langle Y_T^\perp, \boldsymbol{x}\boldsymbol{x}^T\rangle = \langle Y, \boldsymbol{w}\boldsymbol{w}^T\rangle = \langle \sum_{j=1}^{|\mathbb{N}_2^K|-1} \lambda_j\boldsymbol{v}_j\boldsymbol{v}_j^T, \sum_{j=1}^{|\mathbb{N}_2^K|-1} \beta_j^2\boldsymbol{v}_j\boldsymbol{v}_j^T\rangle = \sum_{j=1}^{|\mathbb{N}_2^K|-1} \beta_j^2\lambda_j > 0 \tag{2.19}$$

In the next section, we show how the duality between sum-of-squares polynomials and positive semidefinite matrices can help us construct a dual certificate satisfying the conditions $Y_T = 0$ and $Y_T^\perp \succ 0$.

2.2 Sums-of-squares and positive semidefinite matrices

We call sum-of-squares (SoS) polynomial, any polynomial $p(z)$ for which there exists a decomposition $p(z) = \sum_{j=1}^m s_j^2(z)$ for some polynomials $s_j \in \mathbb{R}[\boldsymbol{z}]$. Introducing a polynomial version of Proposition 1 requires the following lemma from [28] relating SoS and semidefinite programming (SDP).

Proposition 2 (Equivalence between SoS and SDP) *Let $\mathbb{N}^K_{2d}$ denote the set of K-tuples $\boldsymbol{\alpha} \in \mathbb{N}^K$ such that $\sum_i \alpha_i \leq 2d$ and let $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_K^{\alpha_K}$. Let $p(\mathbf{z}) \in \mathbb{R}[\boldsymbol{z}]$ with $p(\mathbf{z}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^K_{2d}} p_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$ be a polynomial of degree $\leq 2d$, the following assertions are equivalent,*

1) $p(\mathbf{z})$ is a sum-of-squares polynomial
2) There exists a positive semidefinite (Gram) matrix $\boldsymbol{A}$ such that

$$p(\mathbf{z}) = \mathbf{z}_{\mathcal{B}}^T \boldsymbol{A} \mathbf{z}_{\mathcal{B}}, \tag{2.20}$$

It is important to notice that Proposition 2 doesn't provide a strict equivalence between a matrix certificate and a polynomial certificate. Observe that the existence of a sum-of-squares polynomial $p(\boldsymbol{z})$ such that $p(\boldsymbol{z}) = \mathbf{z}_{\mathcal{B}}^T \boldsymbol{A} \mathbf{z}_{\mathcal{B}}$ does not imply that the Gram matrix $\boldsymbol{A}$ is positive semidefinite. In other words, not all matrices encoding sum-of-squares polynomials are PSD. As an illustration, consider the following example:

Example 1 (sum-of-squares and positive semidefiniteness)

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \boldsymbol{B} = \begin{pmatrix} 0 & 0 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 0 \end{pmatrix}, \ \boldsymbol{C} = \begin{pmatrix} 0 & 0 & 1/4 \\ 0 & 1/2 & 0 \\ 1/4 & 0 & 0 \end{pmatrix}.$$

For the vector of monomials $\mathbf{z}_{\mathcal{B}} = (1 \quad z \quad z^2)$. All those matrices are encoding the same SoS polynomial $p(z)$

$$\mathbf{z}_{\mathcal{B}}^T \boldsymbol{A} \mathbf{z}_{\mathcal{B}} = \mathbf{z}_{\mathcal{B}}^T \boldsymbol{B} \mathbf{z}_{\mathcal{B}} = \mathbf{z}_{\mathcal{B}}^T \boldsymbol{C} \mathbf{z}_{\mathcal{B}} = p(z) = z^2.$$

However, only one of them is positive semidefinite. The second and third Gram matrices can therefore not be used as the matrix form of a SoS-type certificate. However, note that there exists a matrix $\boldsymbol{\xi}$ such that $\langle \boldsymbol{\xi}, \boldsymbol{B}_\gamma \rangle = 0$ for all $\boldsymbol{\gamma}$, satisfying $\boldsymbol{C} + \boldsymbol{\xi} = \boldsymbol{A}$ or equivalently $\boldsymbol{B} + \boldsymbol{\xi} = \boldsymbol{A}$. Indeed, for $\boldsymbol{C}$ it suffices to take

$$\boldsymbol{\xi} = \begin{pmatrix} 0 & 0 & -1/4 \\ 0 & 1/2 & 0 \\ -1/4 & 0 & 0 \end{pmatrix}$$

This is the point of the following lemma which formalizes and closes the gap between matrix and polynomial certificate.

Lemma 2 below proves equivalence of the matrix certificates up to a $\boldsymbol{\xi}$ provided that the corresponding polynomials are the same.

Lemma 2 *Let Y_1 and Y_2 be two matrices such that $\mathbf{z}_{\mathcal{B}}^T Y_1 \mathbf{z}_{\mathcal{B}} = \mathbf{z}_{\mathcal{B}}^T Y_2 \mathbf{z}_{\mathcal{B}}$ for all z, i.e., the polynomials corresponding to Y_1 and Y_2 are identical. Then there exists a matrix $\boldsymbol{\xi}$ with $\langle \boldsymbol{\xi}, \boldsymbol{B}_{\boldsymbol{\gamma}} \rangle = 0$ for all $\boldsymbol{\gamma}$ and such that $Y_1 = Y_2 + \boldsymbol{\xi}$.*

Proof $\mathbf{z}_{\mathcal{B}}^T(Y_1 - Y_2)\mathbf{z}_{\mathcal{B}} = 0 \quad \forall z \Leftrightarrow \langle Y_1 - Y_2, \mathbf{z}_{\mathcal{B}}\mathbf{z}_{\mathcal{B}}^T\rangle = 0 \quad \forall z \quad \Rightarrow \langle Y_1 - Y_2, \boldsymbol{B}_{\boldsymbol{\gamma}}\rangle = 0 \quad \forall \boldsymbol{\gamma}$. This last implication holds in the reverse direction: if a polynomial $p(\mathbf{z})$ has all zero coefficients, then it must be the zero polynomial.

The conditions of Proposition 1, Proposition 2 and Lemma 2 imply the following result, arising from the polynomial nature of problem (1.1),

Proposition 3 (Polynomial Form) *To ensure unique recovery of* $\boldsymbol{M}_0 = \boldsymbol{m}_0\boldsymbol{m}_0^T$ *with* $(\boldsymbol{m}_0)_{\boldsymbol{\gamma}} = \mathbf{z}_0^{\gamma}$ *the Zeta vector associated to the one atomic measure* $d\mu = \delta(\mathbf{z} - \mathbf{z}_0)$*, it is sufficient, in addition to the injectivity of the constraints on* T*, to find a sum of squares polynomial* $\sum_j s_j^2(z)$ *of degree less than or equal to* 4*, whose associated Gram matrix (i.e. the matrix* $\boldsymbol{Q}$ *such that* $\mathbf{z}_{\mathcal{B}}^T\boldsymbol{Q}\mathbf{z}_{\mathcal{B}} = \sum_{j=1}^N |\boldsymbol{s}_j(\mathbf{z})|^2$*) has rank exactly* $|\mathbb{N}_2^K| - 1$*, polynomials* $\lambda_\ell(z)$ *of degree less than or equal to* $4 - d_\ell$ *and constant* ρ *such that*

$$q(z) = \sum_j s_j^2(z) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}_2^K} \mathbf{z}^{2\boldsymbol{\alpha}} - \rho + \sum_\ell h_\ell(z)\lambda_\ell(z), \tag{2.21}$$

with $q(z)$ *satisfying* $q(z_0) = 0$.

Proof The form of the polynomial $q(z)$ in (2.21) implies the existence of a matrix Y_1 in the range of $\mathcal{A}^*$ such that $\mathbf{z}_{\mathcal{B}}^T(I - Y_1 - \rho\boldsymbol{e}_1\boldsymbol{e}_1^T)\mathbf{z}_{\mathcal{B}} = \sum_j s_j^2(z)$ (condition 1 in Proposition 1). By lemma 2, we can then add a matrix $\boldsymbol{\xi}$ to $I - Y_1 - \rho\boldsymbol{e}_1\boldsymbol{e}_1^T$ to get the positive semidefinite matrix $Y = \sum_j \boldsymbol{s}_j\boldsymbol{s}_j^T = I - Y_2 = I - Y_1 - \rho\boldsymbol{e}_1\boldsymbol{e}_1^T - \boldsymbol{\xi} \succeq 0$ which now satisfies the condition $Y_T^\perp \succeq 0$ of section 2.1. Note that such a $\boldsymbol{\xi}$ always exists, by lemma 2, as we have

$$\begin{aligned} q(z) &= \mathbf{z}_{\mathcal{B}}^T \sum_j \boldsymbol{s}_j\boldsymbol{s}_j^T\mathbf{z}_{\mathcal{B}} &(2.22)\\ &= \mathbf{z}_{\mathcal{B}}^T(I - Y_2)\mathbf{z}_{\mathcal{B}} &(2.23)\\ &= \mathbf{z}_{\mathcal{B}}^T(I - Y_1 - \rho\boldsymbol{e}_1\boldsymbol{e}_1^T - \boldsymbol{\xi})\mathbf{z}_{\mathcal{B}} &(2.24)\\ &= \mathbf{z}_{\mathcal{B}}^T(I - Y_1 - \rho\boldsymbol{e}_1\boldsymbol{e}_1^T)\mathbf{z}_{\mathcal{B}} &(2.25)\end{aligned}$$

As indicated by Proposition 3, since $q(z)$ is SoS, to satisfy the last condition, $Y_T = 0$, it suffices to require $s_j(z_0) = 0$. Indeed, for $(\boldsymbol{m}_0)_{\boldsymbol{\gamma}} = \mathbf{z}_0^{\gamma}$, we have

$$\{s_j(z_0) = 0,\ \forall j\} \iff \{\langle \sum_j \boldsymbol{s}_j\boldsymbol{s}_j^T, \boldsymbol{m}_0\boldsymbol{y}^T + \boldsymbol{y}\boldsymbol{m}_0^T\rangle,\ \forall \boldsymbol{y} \in \mathbb{R}^{|\mathbb{N}_2^K|}\} = 0. \tag{2.26}$$

The forward implication follows from $\boldsymbol{s}_j^T\boldsymbol{m}_0 = s_j(z_0) = 0$. The backward implication follows from taking $\boldsymbol{y} = \boldsymbol{m}_0$ and $\sum_j |\langle \boldsymbol{s}_j, \boldsymbol{m}_0\rangle|^2 = \sum_j |s_j(z_0)|^2$. Note that the second part of (2.26) is precisely the condition $Y_T = 0$ (see Proposition 1).

The last term on the RHS of (2.21) is a contribution of degree ≤ 4 from the ideal $\mathcal{I}(h_1, \ldots, h_K)$ generated by the constraints $h_j(z)$. This contribution thus vanishes at $z = z_0$ and the value of the constant ρ in (2.21), which derives from the one degree of freedom of $\boldsymbol{B}_0$, is thus fixed by enforcing $q(z_0) = 0$.

By assumption, the Gram matrix $\boldsymbol{s}_j\boldsymbol{s}_j^T = I - \boldsymbol{Y}_1 - \rho \boldsymbol{e}_1\boldsymbol{e}_1^T - \boldsymbol{\xi} \succeq 0$ is of rank $|\mathbb{N}_2^K| - 1$. As we have shown that $Y_T = 0$, $\tilde{\boldsymbol{m}}_0$ is in the nullspace of $Y = \sum_j \boldsymbol{s}_j\boldsymbol{s}_j^T$. For any vector $\boldsymbol{w}$ with $\langle \boldsymbol{w}, \tilde{\boldsymbol{m}}_0\rangle = 0$, we must then have $\langle \boldsymbol{w}\boldsymbol{w}^T, Y\rangle > 0$. This implies $Y_T^\perp = (I - \boldsymbol{Y}_1 - \rho\boldsymbol{e}_1\boldsymbol{e}_1^T - \boldsymbol{\xi})_{T^\perp} \succ 0$ (see the discussion (2.13) to (2.19)).

Because of Proposition 2 and Lemma 2, we can now just focus on finding a (dual) polynomial $q(z)$ with the structure (2.21).

2.3 Construction of the dual polynomial

In this section we show how to construct the dual polynomial satisfying the decomposition (2.21). As explained above, such a polynomial implies the existence of a matrix Y satisfying the conditions 1 to 3 from Proposition 1 and serves as the first part of the proof of Theorem 1. We then show injectivity on T to conclude this proof.

Remember that $\boldsymbol{z}$ is given by the concatenation $\boldsymbol{z} = (\boldsymbol{x},\ \boldsymbol{y})$ of all first order monomials arising in problem (1.1). Our construction of the certificate is based on choosing the squares on the LHS of (2.21) to be the canonical polynomials $(\mathbf{z}^{\boldsymbol{\alpha}} - \mathbf{z}_0^{\boldsymbol{\alpha}})^2$ for all $|\boldsymbol{\alpha}| \leq 2$ and to show that those canonical squares can be obtained from the ideal; the squared monomials arising from the trace norm and the constant $\rho = \sum_{\boldsymbol{\alpha}} \mathbf{z}_0^{2\boldsymbol{\alpha}}$. The resulting expression for the certificate is simply

$$q(z) = \sum_{|\boldsymbol{\gamma}|\leq 2} (\mathbf{z}^{\boldsymbol{\gamma}} - \mathbf{z}_0^{\boldsymbol{\gamma}})^2.$$

First, let us show that for all monomials $\mathbf{z}^{\boldsymbol{\alpha}}$ with $|\boldsymbol{\alpha}| = 1$ one can build the polynomial $-2\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\alpha}} + 2(\mathbf{z}_0^{\boldsymbol{\alpha}})^2$ by using a decomposition from the ideal of degree at most 3.

- Either the constraint $\mathbf{z}^{\boldsymbol{\alpha}} = \mathbf{z}_0^{\boldsymbol{\alpha}}$ is present explicitly ($\mathbf{z}^{\boldsymbol{\alpha}} = y_\ell$ corresponds to an element of the first row of $\boldsymbol{X}$ hence $h_\ell(z) \equiv y_\ell - (y_0)_\ell$ is a constraint in Ω) and one can then just multiply this constraint by $-2(\mathbf{z}_0^{\boldsymbol{\alpha}})$ to get the desired polynomial $-2(\mathbf{z}_0)^{\boldsymbol{\alpha}}\mathbf{z}^{\boldsymbol{\alpha}} + 2(\mathbf{z}_0^{\boldsymbol{\alpha}})^2$
- Or, since the bipartite graph is connected, the first order monomial $\mathbf{z}^{\boldsymbol{\alpha}}$, $|\boldsymbol{\alpha}| = 1$, appears in a chain like (1.4), such that if we denote the corresponding numerical values by $(z_0)_{i_1}$, $(z_0)_{i_1}(z_0)_{i_2}$, $\ldots$,$(z_0)_{i_{\ell-1}}(z_0)_\ell$, the constraints $z_{i_1} - (z_0)_{i_1}, \ldots, z_{i_\ell - 1}z_\ell - (z_0)_\ell(z_0)_{i_{\ell-1}}$ belong to Ω and thus to the ideal $\mathcal{I}$. Using (1.4), one can thus recursively combine the elements of the chain in the following way,

$$\begin{aligned}(z_0)_{i_{\ell-2}}(z_0)_{i_{\ell-1}}(z_\ell - (z_0)_\ell) = &(z_\ell z_{i_{\ell-1}} - (z_0)_\ell(z_0)_{i_{\ell-1}})z_{i_{\ell-2}}\\ &- (z_{i_{\ell-2}}z_{i_{\ell-1}} - (z_0)_{i_{\ell-2}}(z_0)_{i_{\ell-1}})z_\ell\\ &+ (z_0)_\ell(z_0)_{i_{\ell-1}}(z_{i_{\ell-2}} - (z_0)_{i_{\ell-2}}).\end{aligned} \quad (2.27)$$

This telescoping relation holds for all ℓ throughout the chain until the second element, (z_{i_2}), for which we have $(z_0)_{i_1}(z_{i_2} - (z_0)_{i_2}) = (z_{i_2}z_{i_1} - (z_0)_{i_2}(z_0)_{i_1}) - z_{i_2}(z_{i_1} - (z_0)_{i_1}) \in \mathcal{I}$. The key here is that one can make use of the bilinear constraints to get a propagation argument which remains degree-3 since the multiplicative factor $(z_0)_\ell(z_0)_{i_{\ell-1}}$ in front of the propagation term $(z_{i_{\ell-2}} - (z_0)_{i_{\ell-2}})$ remains constant. In particular, note that we never use the third order constraints $z^{\boldsymbol{\alpha}}(z_{i_1} - (z_0)_{i_1})$ for $|\boldsymbol{\alpha}| = 2$. The highest degree of the monomials multiplying the first order constraints is one. This will be important later when establishing the stability result.

Now that we can build the polynomials $-2(z_0)_k z_k + 2(z_0)_k^2$ for all k as degree-3 decompositions from the ideal $\mathcal{I}$, one can just add those polynomials to the trace and constant ρ contributions $z_k^2 - (z_0)_k^2$ in order to get the squares $(z_k - (z_0)_k)^2$. We thus get $|\mathbb{N}_1^K| - 1$ of the required squares. The remaining $\binom{K}{2}$ decompositions for the second order squared polynomials $(\mathbf{z}^{\boldsymbol{\alpha}} - \mathbf{z}_0^{\boldsymbol{\alpha}})^2$ for $|\boldsymbol{\alpha}| = 2$, are built from the first order decompositions, the trace, and constant ρ as follows. $\forall \boldsymbol{\alpha}, \boldsymbol{\beta}$ with $|\boldsymbol{\alpha}|, |\boldsymbol{\beta}| = 1$,

$$\begin{aligned}(\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}^{\boldsymbol{\beta}} - \mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}})^2 &= (\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}^{\boldsymbol{\beta}})^2 - (\mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}})^2\\ &\quad - 2\mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}}(\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}^{\boldsymbol{\beta}} - \mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}}),\end{aligned} \tag{2.28}$$

where the first two terms arise from the contribution of the trace and ρ, and the third one can be expressed from the ideal $\mathcal{I}$ with degree at most 4, as

$$\begin{aligned}-2\mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}}(\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}^{\boldsymbol{\beta}} - \mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}}) &= (-2\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\alpha}} + 2(\mathbf{z}_0^{\boldsymbol{\alpha}})^2)(\mathbf{z}_0^{\boldsymbol{\beta}})^2\\ &\quad + (\mathbf{z}^{\boldsymbol{\beta}} - \mathbf{z}_0^{\boldsymbol{\beta}})(-2\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\alpha}})\mathbf{z}_0^{\boldsymbol{\beta}}\end{aligned} \tag{2.29}$$

The first term is of degree at most 3 and the second one is of degree at most 4.

Together, the telescoping relations (2.27), (2.28) and (2.29) show that one can build a SoS polynomial of the form $\sum_{0<|\boldsymbol{\alpha}|\leq 2}(\mathbf{z}^{\boldsymbol{\alpha}} - \mathbf{z}_0^{\boldsymbol{\alpha}})^2$, which contains the square of each of the $|\mathbb{N}_2^K| - 1$ polynomials $\mathbf{z}^{\alpha} - \mathbf{z}_0^{\boldsymbol{\alpha}}$, for $0 < |\boldsymbol{\alpha}| \leq 2$, and which can be written from the ideal generated by the completion constraints. This polynomial can be encoded by the rank $|\mathbb{N}_2^K| - 1$ Gram matrix $\boldsymbol{Y} = \boldsymbol{Q} = \sum_{0<|\boldsymbol{\alpha}|\leq 2} \boldsymbol{v}_{\boldsymbol{\alpha}}\boldsymbol{v}_{\boldsymbol{\alpha}}^T$ where the $\boldsymbol{v}_{\boldsymbol{\alpha}}$ are vectors indexed over the K-tuples of $\mathbb{N}_2^K$ and whose only non zero entries are the first one, which we set to $-\mathbf{z}_0^{\boldsymbol{\alpha}}$, and the entry corresponding to the K-tuple $\boldsymbol{\alpha}$ which we set to 1. In other words, we have $\boldsymbol{v}_{\boldsymbol{\alpha}} = (-\mathbf{z}_0^{\boldsymbol{\alpha}}, 0, \ldots, 0, 1, 0, \ldots, 0)$. Each of those $|\mathbb{N}_2^K| - 1$ vectors $\boldsymbol{v}_{\boldsymbol{\alpha}}$ is linearly independent from the others as it is the only vector for which the entry corresponding to the K-tuple $\boldsymbol{\alpha}$ is non zero. The Gram matrix $\boldsymbol{Q}$ thus has rank $|\mathbb{N}_2^K| - 1$ as desired. In fact, the nullspace of this matrix is given by the vector $\boldsymbol{m}_0$ encoding all the monomials of degree at most two in the entries of $\mathbf{z}_0$. For any other vector $\boldsymbol{x}$, the equality $\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} = 0$ necessarily implies $\sum_{0<|\boldsymbol{\alpha}|\leq 2} |\langle \boldsymbol{v}_{\boldsymbol{\alpha}}, \boldsymbol{x}\rangle|^2 = 0$ and hence $x_{\boldsymbol{\alpha}} = x_0\mathbf{z}_0^{\boldsymbol{\alpha}}$ for all $\boldsymbol{\alpha} \neq \mathbf{0}$. This in turn implies $\boldsymbol{x} = x_0\boldsymbol{m}_0$. Finally, note that if the rank $|\mathbb{N}_2^K| - 1$ Gram matrix $\boldsymbol{Y} = \boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{Y}_2$ encoding the SoS does not match the Gram matrix $\boldsymbol{I} - \boldsymbol{Y}_1$ encoding the RHS in (2.21), Lemma 2 ensures that there always exists a matrix $\boldsymbol{\xi}$ such that $\boldsymbol{I} - \boldsymbol{Y}_2 = \boldsymbol{Q} = \boldsymbol{Y} = \boldsymbol{I} - \boldsymbol{Y}_1 - \boldsymbol{\xi}$. As a consequence, the first condition of Proposition 1 holds for the PSD matrix $\boldsymbol{Y} = \boldsymbol{Q}$ which vanishes on T. Since the rank condition implies $\boldsymbol{Q}_T^\perp = \boldsymbol{Y}_T^\perp \succ 0$ through Proposition 3, the first three conditions of Proposition 1 hold.

To conclude the proof of Theorem 1, we show that the linear map $\mathcal{A}$ grouping the linear constraints derived from the polynomials h_ℓ and the structure of the moments matrix, is injective on T. For this purpose, let us show that the nullspace of $\mathcal{A}$ is empty on T. Let us consider any $\boldsymbol{H} = \boldsymbol{m}_0\boldsymbol{v}^T + \boldsymbol{v}\boldsymbol{m}_0^T$. The condition $\boldsymbol{H}_{11} = 1$ (probability measure) implies $v_1 = 0$ and reduces $\boldsymbol{H}$ to a matrix for which the first column equals the first row and is given by $(v_2 \ldots, v_{|\mathbb{N}_2^K|})$. From the connectivity of the graph, there must exist at least one additional constraint setting to zero another element of the first row or column. So there exists $\ell \neq 1$ s.t. $v_\ell = 0$.

Accordingly, the whole corresponding row and column reduce to $((z_0)_\ell v_k)_{k \leq |\mathbb{N}_2^K|}$. Since $(z_0)_\ell \neq 0$[2], one can then apply the next constraint $z_\ell z_m = 0$ which implies $v_m = 0$ for some $m \neq \ell$. By applying this idea recursively and using the fact that the bipartite graph is connected, and hence that every v_ℓ must appear in the chain, one can show that all the entries of $\boldsymbol{v}$ corresponding to first order moments must vanish.

For the higher order moments, we first rely on the fact that $\boldsymbol{H}$ is rank two and symmetric which implies $\boldsymbol{H}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} = 0$ for all $|\boldsymbol{\alpha}_1|, |\boldsymbol{\alpha}_2| \leq 1$. We then use the fact that $\boldsymbol{H}$ is required to satisfy the structural constraints (i.e. $\boldsymbol{H}$ is a pseudo-moments matrix), we thus have $\boldsymbol{H}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} = \boldsymbol{H}_{\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2,\boldsymbol{0}}$ for any $|\boldsymbol{\alpha}_1|, |\boldsymbol{\alpha}_2| \leq 1$. From this, the whole first row/column of $\boldsymbol{H}$ must be zero. Since $(\boldsymbol{m}_0)_1 = 1$, this finally gives $\boldsymbol{v} = 0$ and hence $\boldsymbol{H} = \boldsymbol{0}$.

3 Stability

In this section, we prove Theorem 2 and Corollary 1. We consider measurements $(\tilde{X}_0)_{ij} = (X_0)_{ij} + \varepsilon_{ij}$ that are corrupted by a noise ε_{ij} with bounded ℓ_2 norm $\|\varepsilon\|\sqrt{1 + \|\mathbf{z}_0\|_2^2 + \|\mathbf{z}_0\|_2^4} \leq \eta$. If $h_\ell(\mathbf{z}) := \mathbf{z}^{\boldsymbol{\alpha}} - \mathbf{z}_0^{\boldsymbol{\alpha}}$ denotes a constraint in Ω, we let $\tilde{h}_\ell(\mathbf{z}) := \mathbf{z}^{\boldsymbol{\alpha}} - \tilde{\mathbf{z}}_0^{\boldsymbol{\alpha}}$ denote the corresponding noisy constraint with $\tilde{\mathbf{z}}_0^{\boldsymbol{\alpha}} = \mathbf{z}_0^{\boldsymbol{\alpha}} + \varepsilon_{ij}$. Hence, $\tilde{h}_\ell(\mathbf{z}) = h_\ell(\mathbf{z}) - \varepsilon_{ij}$ with (i,j) relative to the constraint indexed by ℓ, or with a slight abuse of notation, $\tilde{h}_\ell(\mathbf{z}) = h_\ell(\mathbf{z}) - \varepsilon_\ell$.

Let $\eta \geq \|\varepsilon\|_2\sqrt{(1 + \|\mathbf{z}_0\|_2^2 + \|\mathbf{z}_0\|_2^4|)}$ with $\|\varepsilon\|_2 = \sqrt{\sum_{ij\in\Omega}\varepsilon_{ij}^2}$. The stable version of (2.1) reads,

$$\min \operatorname{Tr}(\boldsymbol{M}_K^{(4)}), \text{ s.t. } \boldsymbol{M}_K^4(m) \succeq 0, \sqrt{\sum_{\ell=1}^K \left\|\boldsymbol{M}_K^{4-d_\ell}(\tilde{h}_\ell m)\right\|_F^2} \leq \eta. \tag{3.1}$$

[2] Recall that we assumed $(\boldsymbol{X}_0)_{ij} \neq 0$ for all (i,j)

As indicated above, the last inequality in (3.1) can expand from the moments matrix $\boldsymbol{M}_K^4$ and the matrices $\boldsymbol{B}_{\boldsymbol{\gamma}}$ as

$$\sqrt{\sum_{\ell} \sum_{\boldsymbol{\kappa} \in \mathbb{N}^K_{2(2-d_{h_\ell})}} \left| \sum_{\boldsymbol{\zeta}} \frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2} \langle \boldsymbol{M}_K^4, \boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}} \rangle \right|^2} \leq \eta. \text{ for } 1 \leq \ell \leq L \tag{3.2}$$

The left handside in (3.2) encodes the ℓ_2 norm of all affine constraints, including those that are obtained by multiplying the original constraints $h_\ell(\mathbf{z}) = \mathbf{0}$ by monomials of degree less than or equal to 2. The first sum that appears in (3.2) is taken over all the different noisy constraints $\tilde{h}_\ell$. The second sum (over $\boldsymbol{\kappa}$) is taken over all the "shifts" of those polynomials. For a given $\boldsymbol{\kappa}$, the corresponding shifted polynomial is simply obtained by multiplying $\tilde{h}_\ell(\mathbf{z})$ by the corresponding monomial $\mathbf{z}^{\boldsymbol{\kappa}}$.

It is worth pointing out that formulation (3.1) is not unit-independent, since the moment matrix M mixes different powers of the original variables. This can be remedied by assigning dimensional weights w_γ to the $\boldsymbol{B}_\gamma$ matrices – an operation that modifies the numerics and the theory in an obvious way. Formulation (3.1) leads to the recovery result of Theorem 2 which is restated below for clarity.

Theorem 2 *Let $\boldsymbol{M}_0$ denote the rank one matrix introduced in* (1.25) *for $\mathbf{z}_0 \in \mathbb{R}^K$. Assume that the necessary and sufficient conditions of lemma 1 are satisfied. There exists a constant $C(\delta_L, \delta_U)$ such that the solution $\boldsymbol{M}$ to the semidefinite program* (1.26) *obeys*

$$\frac{\|\boldsymbol{M} - \boldsymbol{M}_0\|_F}{\|\boldsymbol{M}_0\|_F} \leq C(\delta_L, \delta_U) \cdot (m+n)^{5/2} \eta. \tag{1.27}$$

The constant C depends on the magnitude of the entries of $\boldsymbol{X}_0$, but not on Ω, m, n, or ε.

As explained earlier, the stability result of Theorem 2 can be improved by considering a reduced moments matrix that only retains the second order monomials appearing in the constraints. Further improvement in the Lipschitz stability constant can be achieved if the path relating the root node(s) (i.e the nodes encoding the degree 1 constraints) to any of the leaf nodes in the graph $\mathcal{G}(\mathcal{V}_1, \mathcal{V}_2)$ can be explicitly given. In this case, we can focus on constraining the noise along each of the paths. We recall the resulting estimate below.

Corollary 1 *Assume that the paths relating any of the root nodes to the leaf nodes in $\mathcal{G}(\mathcal{V}_1, \mathcal{V}_2)$ are explicitly given and that the noise along those paths is bounded by η'. There exists a reduced version of the level-2 Lasserre hierarchy and a constant $C(\delta_L, \delta_U)$ such that the solution to the corresponding semidefinite program obeys*

$$\frac{\|\boldsymbol{M} - \boldsymbol{M}_0\|_F}{\|\boldsymbol{M}_0\|_F} \leq C(\delta_L, \delta_U) \cdot (m+n)^2 \eta'. \tag{1.29}$$

The constant C depends on the magnitude of the entries of $\boldsymbol{X}_0$, but not on Ω, m, n, or ε.

Now that we have introduced the dual certificate, we can give more details on the reduced semidefinite program underlying Corollary 1. This semidefinite program is identical to (1.26) except that the moments matrix $\boldsymbol{M}_K^4$ is now a proxy for the (reduced) rank one matrix $\boldsymbol{M}_0 = \left[1, \mathbf{z}_0, ((x_0)_i(y_0)_j)_{(i,j)\in\Omega}\right]^{\otimes 2}$. As explained above, we also replace the affine constraints in (1.26) by a path related bound. For any of the sets $\mathcal{H}_{\ell_k}$ introduced in section 1.5, we let $\mathcal{M}_{\ell_k}$ to denote the set of multi-indices corresponding to the variables in $\mathcal{H}_{\ell_k}$. That is to say, $\boldsymbol{\alpha} \in \mathcal{M}_{\ell,k}$ iff $\mathbf{z}^{\boldsymbol{\alpha}} \in \mathcal{H}_{\ell,k}$. The affine constraints for the reduced semidefinite program read as

$$\max_i \sqrt{\sum_{\ell\in\mathcal{P}_i}\sum_{\boldsymbol{\kappa}\in\mathcal{M}_\ell}\left|\sum_{\boldsymbol{\zeta}}\frac{\tilde{h}_\ell}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2}\langle \boldsymbol{M}, \boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\rangle\right|^2} \leq \eta'. \tag{3.3}$$

The next section proceeds with the proof of Theorem 2.

3.1 Proof of Theorem 2

Let $\|\boldsymbol{M}\|_p$ denote the Schatten p-norm of $\boldsymbol{M}$,

$$\|\boldsymbol{M}\|_p = \left(\sum_k \sigma_k^p\right)^{1/p}.$$

We therefore have $\|\boldsymbol{M}\|_1 = \|\boldsymbol{M}\|_*$ which denotes the nuclear norm of $\boldsymbol{M}$, $\|\boldsymbol{M}\|_2 = \|\boldsymbol{M}\|_F$ which is used to denote the Frobenius norm of $\boldsymbol{M}$ and $\|\boldsymbol{M}\|_\infty = \|\boldsymbol{M}\|$ which denotes the operator norm of $\boldsymbol{M}$.

Any optimal solution $\boldsymbol{M}$ to (3.1) can read $\boldsymbol{M} = \boldsymbol{H} + \boldsymbol{M}_0$. To prove stability of the recovery, we first highlight the following,

- $\mathrm{Tr}(\boldsymbol{M}_0 + \boldsymbol{H}) \leq \mathrm{Tr}(\boldsymbol{M}_0)$ and therefore $\mathrm{Tr}(\boldsymbol{H}) \leq 0$.
- Both $\boldsymbol{M}$ and $\boldsymbol{M}_0$ are feasible points for (3.1), hence both satisfy the normalization constraint $\boldsymbol{M}_{11} = (\boldsymbol{M}_0)_{11} = 1$ which can be exactly enforced. As a consequence, $\boldsymbol{H}_{11} = (\boldsymbol{M}_0)_{11} - \boldsymbol{M}_{11} = 0$, and all constant terms in the constraints $\tilde{h}_1, \dots, \tilde{h}_L$ vanish when those constraints are applied to $\boldsymbol{H}$. We have

$$\frac{(\tilde{h}_\ell)_{\mathbf{0}}}{\|\boldsymbol{B}_{\mathbf{0}}\|_F^2}\langle \boldsymbol{H}, \boldsymbol{B}_{\mathbf{0}}\rangle = 0, \qquad \ell = 1, \dots, L. \tag{3.4}$$

More generally, for both $\boldsymbol{M}$ and $\boldsymbol{M}_0$, as η is bounding the ℓ_2 norm of the vector $(\varepsilon_\ell \mathbf{z}_0^{\boldsymbol{\kappa}})_{\ell,|\boldsymbol{\kappa}|\leq 2}$ of weighted residuals, we must have,

$$\sum_{\ell,\kappa}\left|\sum_\zeta (\tilde{h}_\ell)_\zeta \langle \boldsymbol{B}_{\zeta+\kappa}, \boldsymbol{M}_0\rangle\right|^2 = \sum_{\ell,\kappa}\left|\sum_\zeta (h_\ell)_\zeta\langle \boldsymbol{B}_{\zeta+\kappa}, \boldsymbol{M}_0\rangle - \varepsilon_\ell \mathbf{z}_0^\kappa\right|^2 \tag{3.5}$$

$$= \sum_{\ell,\kappa}(\varepsilon_\ell \mathbf{z}_0^{\boldsymbol{\alpha}})^2 \leq \eta^2 \tag{3.6}$$

$$\sum_{\ell,\kappa}\left|\sum_{\zeta}(\tilde{h}_\ell)_\zeta\langle \boldsymbol{B}_{\zeta+\kappa},\boldsymbol{M}\rangle\right|^2 \leq \sum_{\ell,\kappa}(\varepsilon_\ell \mathbf{z}_0^{\boldsymbol{\alpha}})^2 \leq \eta^2$$

From those relations we can derive a similar bound on $\boldsymbol{H}$,

$$\begin{aligned}&\sqrt{\sum_{\boldsymbol{\kappa}}\sum_{\ell}\left|\sum_{\boldsymbol{\zeta}}\frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2}\langle \boldsymbol{H},\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\rangle\right|^2}\\ &=\sqrt{\sum_{\boldsymbol{\kappa}}\sum_{\ell}\left|\sum_{\boldsymbol{\zeta}}\frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2}\langle \boldsymbol{M}-\boldsymbol{M}_0,\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\rangle\right|^2}\\ &\leq \sqrt{2}\eta.\end{aligned}$$

- Finally, note that $\boldsymbol{H}+\boldsymbol{M}_0 \succeq 0$ implies $\langle \boldsymbol{H}+\boldsymbol{M}_0,\boldsymbol{W}\rangle \geq 0$ for all $\boldsymbol{W}\succeq 0$ including all $\boldsymbol{W}\in T^\perp$ which implies $\boldsymbol{H}_{T^\perp}\succeq 0$.

The matrix form of $\boldsymbol{Y}_2$ belongs to the range of $\mathcal{A}^*$ (i.e, its polynomial form belongs to the ideal $\mathcal{I}$) modulo a $\boldsymbol{\xi}$ and $\boldsymbol{Y}_2$ can read as

$$\boldsymbol{Y}_2 = \boldsymbol{\xi} + \sum_{\ell}\sum_{\boldsymbol{\kappa}\in\mathbb{N}^K_{2(t-d_{h_\ell})}}\lambda_{\ell\boldsymbol{\kappa}}\left(\sum_{\boldsymbol{\zeta}}\frac{(h_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2}\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\right), \tag{3.7}$$

for some $\lambda_{\ell\kappa}$, and $\boldsymbol{\xi}$ orthogonal to the $\boldsymbol{B}_{\boldsymbol{\gamma}}$.

The next section derives a bound on $\|\boldsymbol{H}_T^\perp\|$. For this, we start by bounding $|\langle \boldsymbol{H},\boldsymbol{Y}_1\rangle|$. Note that $|\langle \boldsymbol{Y}_2,\boldsymbol{H}\rangle| = |\langle \boldsymbol{Y}_1,\boldsymbol{H}\rangle|$, as $\langle \boldsymbol{\xi},\boldsymbol{H}\rangle = 0$.

3.1.1 Bound on $\boldsymbol{H}_T^\perp$

The certificate (3.7) is built from the noiseless constraints h_ℓ, while the solutions $\boldsymbol{M}$, $\boldsymbol{M}_0$ and thus $\boldsymbol{H}$ are bounded with respect to the corrupted constraints from (3.1). As we saw earlier, the noisy constraints relate to the noiseless constraints as

$$\tilde{h}_\ell(z) = \sum_{\boldsymbol{\zeta}}(\tilde{h}_\ell)_{\boldsymbol{\zeta}}\mathbf{z}^{\boldsymbol{\zeta}} = \sum_{\boldsymbol{\zeta}\neq 0}(h_\ell)_{\boldsymbol{\zeta}}\mathbf{z}^{\boldsymbol{\zeta}} + (h_\ell)_0 - \varepsilon_\ell = h_\ell(z) - \varepsilon_\ell. \tag{3.8}$$

Let $\boldsymbol{Y}_1^{(1)}$ and $\boldsymbol{Y}_1^{(2)}$ denote the contributions to Y_1 corresponding to the first and second order squares in the SoS certificate of section 2.3 respectively. For both $\boldsymbol{M}$ and $\boldsymbol{M}_0$, we further introduce the decomposition

$$\boldsymbol{M} = \begin{bmatrix} M_{11} & M_{12}\\ M_{21} & M_{22}\end{bmatrix}. \tag{3.9}$$

For those matrices, because of the structural constraints and PSD constraint, one can write $\mathrm{Tr}(\boldsymbol{M}) \geq \|\boldsymbol{M}_{11}\|_F^2 + \mathrm{Tr}(\boldsymbol{M}_{11})$. Moreover, we have $\mathrm{Tr}(\boldsymbol{M}_0) \geq \mathrm{Tr}(\boldsymbol{M})$ so in particular,

$$\|\boldsymbol{M}_{11}\|_F \leq \sqrt{\mathrm{Tr}(\boldsymbol{M}) - \mathrm{Tr}(\boldsymbol{M}_{11})} \leq \sqrt{\mathrm{Tr}(\boldsymbol{M})} \leq \sqrt{\mathrm{Tr}(\boldsymbol{M}_0)} = \|\boldsymbol{m}_0\|. \tag{3.10}$$

For any constraint h_ℓ, using the recursion (2.27), the difference between h_ℓ and $\tilde{h}_\ell$ will only affect the entries in $\boldsymbol{H}$ corresponding to first and zero order moments. Since $\boldsymbol{H}_{11} = 0$ (see the discussion above), this difference will thus only affect first order entries. Let $W_{\ell,\kappa}$ denote the multiplicative coefficient of the pair $h_\ell(\mathbf{z})\mathbf{z}^{\boldsymbol{\kappa}}$, for $|\boldsymbol{\kappa}| \leq 1$, in the expression of $\boldsymbol{Y}_1^{(1)}$

$$|\langle \boldsymbol{Y}_1^{(1)}, \boldsymbol{H}\rangle| = \left|\sum_\kappa \sum_\ell W_{\ell,\kappa} \sum_\zeta (h_\ell)_\zeta \langle \boldsymbol{B}_{\zeta+\kappa}, \boldsymbol{H}\rangle\right| \tag{3.11}$$

$$= \left|\sum_\kappa \sum_\ell W_{\ell,\kappa} \sum_\zeta (\tilde{h}_\ell)_\zeta \langle \boldsymbol{B}_{\zeta+\kappa}, \boldsymbol{M}\rangle - \sum_{|\kappa|\leq 1} \sum_\ell W_{\ell,\kappa}\varepsilon_\ell \langle \boldsymbol{B}_\kappa, \boldsymbol{M}\rangle\right| \tag{3.12}$$

$$\leq \left|\sum_\kappa \sum_\ell W_{\ell,\kappa} \sum_\zeta (\tilde{h}_\ell)_\zeta \langle \boldsymbol{B}_{\zeta+\kappa}, \boldsymbol{M}\rangle\right| + \left|\sum_{|\kappa|\leq 1} \sum_\ell W_{\ell,\kappa}\varepsilon_\ell \langle \boldsymbol{B}_\kappa, \boldsymbol{M}\rangle\right| \tag{3.13}$$

$$\leq \left|\sum_\kappa \sum_\ell W_{\ell,\kappa} \sum_\zeta (\tilde{h}_\ell)_\zeta \langle \boldsymbol{B}_{\zeta+\kappa}, \boldsymbol{M}\rangle\right| + C_1(m+n)^{3/2}\|\boldsymbol{M}_{11}\|_F\|\varepsilon\|_\infty \tag{3.14}$$

$$\leq \left|\sum_\kappa \sum_\ell W_{\ell,\kappa} \sum_\zeta (\tilde{h}_\ell)_\zeta \langle \boldsymbol{B}_{\zeta+\kappa}, \boldsymbol{M}\rangle\right| + C_1(m+n)^{3/2}\|\boldsymbol{m}_0\|\|\varepsilon\|_\infty \tag{3.15}$$

$$\leq (m+n)^{3/2} C_2(\delta_L, \delta_U) \max\{\eta, \|\boldsymbol{m}_0\|\|\varepsilon\|_\infty\} \tag{3.16}$$

To get (3.14), first note that since the second term in (3.13) only contains first order monomials, the total contribution $\sum_{\boldsymbol{\kappa}} W_{\ell\boldsymbol{\kappa}}\varepsilon_\ell \boldsymbol{B}_{\boldsymbol{\kappa}}$ necessarily has the form $\boldsymbol{v}\boldsymbol{e}_1^T + \boldsymbol{e}_1\boldsymbol{v}^T$ with $\boldsymbol{v}$ having at most $(m+n)$ non zero entries. We can also replace $\boldsymbol{M}$ by $\boldsymbol{M}_{11}$. Then note that for any given path between the root node and any of the leaf nodes, the monomial associated to the vertex t_i in the path will multiply the constraints $h_{(i-2,i-1)}(\mathbf{z})$ and $h_{(i+1,i+2))}(\mathbf{z})$ according to (2.27). Since those constraints will be used to express all the monomials associated to vertices that appear between any of the edges associated to those constraints and the leaf node, the products $h_{(i-2,i-1)}(\mathbf{z})\mathbf{z}_{t_i}$ and $h_{(i+1,i+2)}(\mathbf{z})\mathbf{z}_{t_i}$ will thus appear at most $4(m+n)$ times in the construction of the certificate. Moreover, when being used in the construction of the certificate, the product $h_{(i-2,i-1)}(\mathbf{z})\mathbf{z}_{t_i}$ and $h_{(i+1,i+2)}(\mathbf{z})\mathbf{z}_{t_i}$ will be multiplied by a ratio of the entries of $\mathbf{z}_0$. For lower and upper bounds δ_L, δ_U on those entries, and a constant $C_1(\delta_L, \delta_U)$ that only depends on those entries, we thus have $\|\sum_\ell \sum_{\boldsymbol{\kappa}} \varepsilon_\ell \boldsymbol{B}_{\boldsymbol{\kappa}}\|_F \leq C_1(\delta_L, \delta_U)\cdot(m+n)^{3/2}\|\varepsilon\|_\infty$.

The bound (3.15) follows from (3.10). Finally, (3.16) follows from the Cauchy-Schwarz inequality, noting that every constraint $\tilde{h}_\ell$ will only be multiplied by two distinct monomials according to (2.27). The resulting pairs $(\ell, \boldsymbol{\kappa})$ will then appear at most $2(m+n)$ times in the certificate, to express the subsequent order 1 monomials. This gives the upper bound $|W_{\ell,\boldsymbol{\kappa}}| \leq 4(m+n)C_2(\delta_L, \delta_U)$. Since only $2(m+n)$ $(\ell, \boldsymbol{\kappa})$ pairs are used, the vector $\boldsymbol{w}$ whose entries are indexed over the pairs $(\ell, \boldsymbol{\kappa})$ and defined by $W_{\ell,\boldsymbol{\kappa}}$ satisfies $\|\boldsymbol{w}\| \leq (m+n)^{3/2}C_2(\delta_L, \delta_U)$. On the other hand, note that the vector $\boldsymbol{u}$, indexed on the pairs $(\ell, \boldsymbol{\kappa})$ and whose $(\ell, \boldsymbol{\kappa})$ entry reads as $u_{\ell,\boldsymbol{\kappa}} = \sum_{\boldsymbol{\zeta}} (\tilde{h}_\ell)_{\boldsymbol{\zeta}} \langle \boldsymbol{B}_{\boldsymbol{\zeta},\boldsymbol{\kappa}}, \boldsymbol{M}\rangle$ can be bounded from (3.2) as $\|\boldsymbol{u}\|_2 \leq \eta$. Together, those bounds give $|\langle \boldsymbol{w}, \boldsymbol{u}\rangle| \leq (m+n)^{3/2}C_2(\delta_L, \delta_U)\cdot\eta$.

We now bound the second order contributions gathered in $\boldsymbol{Y}_1^{(2)}$. From (2.29), those contributions can be decomposed as $\boldsymbol{Y}_1^{(2)} = \boldsymbol{V}_1 + \boldsymbol{V}_2$, where $\boldsymbol{V}_1$ only involves the decomposition of degree one polynomials (first term on the RHS of (2.29)), and $\boldsymbol{V}_2$ denotes the higher order contributions (second term on the RHS of (2.29)).

Consider the first term $-2(\mathbf{z}^{\boldsymbol{\alpha}} - \mathbf{z}_0^{\boldsymbol{\alpha}})\mathbf{z}_0^{\boldsymbol{\alpha}}|\mathbf{z}_0^{\boldsymbol{\beta}}|^2$ on the RHS of (2.29). The first order polynomial $\mathbf{z}^{\boldsymbol{\alpha}} - \mathbf{z}_0^{\boldsymbol{\alpha}}$ is expressed by at most $2(m+n)$ of the pairs $(\ell, \boldsymbol{\kappa})$ with associated constraints $|h_\ell(\mathbf{z})\mathbf{z}^{\boldsymbol{\kappa}}| \leq \mathbf{z}_0^{\boldsymbol{\kappa}}\|\varepsilon\|_\infty$. Now let $W_{\ell,\boldsymbol{\kappa},\boldsymbol{\alpha}}$ denote the multiplicative coefficient of the polynomial $h_\ell(\mathbf{z})\mathbf{z}^{\boldsymbol{\kappa}}\mathbf{z}_0^{\boldsymbol{\alpha}}$ in the decomposition of $-2(\mathbf{z}^{\boldsymbol{\alpha}} - \mathbf{z}_0^{\boldsymbol{\alpha}})\mathbf{z}_0^{\boldsymbol{\alpha}}$. Using this notation we can write the contribution from $\boldsymbol{V}_1$ as

$$\langle \boldsymbol{H}, \boldsymbol{V}_1\rangle = \sum_{|\boldsymbol{\alpha}|\leq 1}\sum_{|\boldsymbol{\beta}|\leq 1}\sum_{\boldsymbol{\kappa}}\sum_{\ell} W_{\ell,\boldsymbol{\kappa},\boldsymbol{\alpha}}\, |\mathbf{z}_0^{\boldsymbol{\beta}}|^2 \sum_{\boldsymbol{\zeta}} (h_\ell)_{\boldsymbol{\zeta}} \langle \boldsymbol{B}_{\zeta+\boldsymbol{\kappa}}, \boldsymbol{H}\rangle \tag{3.17}$$

Since all of the pairs $(\ell, \boldsymbol{\kappa})$ in any given chain are distinct, and since $\mathbf{z}^{\boldsymbol{\kappa}}\mathbf{z}^{\boldsymbol{\alpha}} \neq \mathbf{z}^{\boldsymbol{\kappa}'}\mathbf{z}^{\boldsymbol{\alpha}'}$ except for $\boldsymbol{\alpha}' = \boldsymbol{\kappa}$ and $\boldsymbol{\kappa}' = \boldsymbol{\alpha}$, we can write $|W_{\ell,\boldsymbol{\kappa},\boldsymbol{\alpha}}| \leq C(\delta_L, \delta_U)$ for some constant C which only depends on the magnitude of the entries in $\mathbf{z}_0$. Moreover, distinct triples $(\ell, \boldsymbol{\kappa}, \boldsymbol{\alpha})$ are upper bounded by distinct entries in the vector $\mathbf{z}_0 \otimes \mathbf{z}_0 \otimes \boldsymbol{\varepsilon}$. Using Cauchy-Schwartz, we can thus bound (3.17) as

$$\left|\sum_{|\boldsymbol{\alpha}|\leq 1}\sum_{|\boldsymbol{\beta}|\leq 1}\sum_{\boldsymbol{\kappa}}\sum_{\ell} W_{\ell,\boldsymbol{\kappa},\boldsymbol{\alpha}}\, |\mathbf{z}_0^{\boldsymbol{\beta}}|^2 \sum_{\boldsymbol{\zeta}} (h_\ell)_{\boldsymbol{\zeta}} \langle \boldsymbol{B}_{\zeta+\boldsymbol{\kappa}}, \boldsymbol{H}\rangle\right| \tag{3.18}$$

$$\leq \sum_{|\boldsymbol{\beta}|\leq 1}\left|\sum_{|\boldsymbol{\alpha}|\leq 1}\sum_{\boldsymbol{\kappa}}\sum_{\ell} W_{\ell,\boldsymbol{\kappa},\boldsymbol{\alpha}} \sum_{\boldsymbol{\zeta}} (h_\ell)_{\boldsymbol{\zeta}} \langle \boldsymbol{B}_{\zeta+\boldsymbol{\kappa}}, \boldsymbol{H}\rangle\right| |\mathbf{z}_0^{\boldsymbol{\beta}}|^2 \tag{3.19}$$

$$\leq \|\mathbf{z}_0\|^2 \cdot \eta \cdot (m+n) \cdot C_3(\delta_L, \delta_U) \tag{3.20}$$

A similar reasoning holds for $|\langle \boldsymbol{H}, \boldsymbol{V}_2\rangle|$, noting that every α in $(\mathbf{z}^\beta - \mathbf{z}_0^\beta)\mathbf{z}^{\boldsymbol{\alpha}}$ gives a different constraint in (3.1). From this we can write

$$|\langle \boldsymbol{H}, \boldsymbol{Y}_1^{(2)}\rangle| \leq C_4(\delta_L, \delta_U)\cdot(m+n)^{3/2}\cdot\eta\cdot\|\boldsymbol{z}_0\|_2 \tag{3.21}$$

Consider the sum-of-squares certificate of section 2.3. In polynomial form, we have seen that this certificate reads $\sum_{|\boldsymbol{\gamma}|\leq 2}(\mathbf{z}^{\boldsymbol{\gamma}} - \mathbf{z}_0^{\boldsymbol{\gamma}})^2 = \sum_{|\boldsymbol{\gamma}|\leq 2}(\mathbf{z}^{\boldsymbol{\gamma}} - (\boldsymbol{m}_0)_{\boldsymbol{\gamma}})^2$. One possible matrix representation[3] of this certificate is thus given by $\bar{Z} = \sum_j \boldsymbol{s}_j\boldsymbol{s}_j^T$

[3] An alternative representation would be given by the decomposition Trace + ideal of section 2.3 and encoded as $\boldsymbol{I} - \boldsymbol{Y}_1$.

where each $\boldsymbol{s}_j$ denotes a vector of the form $-(m_0)_{\boldsymbol{\gamma}}\boldsymbol{e}_1 + \boldsymbol{e}_{\boldsymbol{\gamma}}$. Using this form, we get

$$\bar{Z} = \boldsymbol{I} - \boldsymbol{m}_0\boldsymbol{e}_1^T - \boldsymbol{e}_1\boldsymbol{m}_0^T + \|\boldsymbol{m}_0\|^2\boldsymbol{e}_1\boldsymbol{e}_1^T = \boldsymbol{I} - \boldsymbol{Y}_2. \tag{3.22}$$

For $\boldsymbol{m} \perp \boldsymbol{m}_0$, with $\|\boldsymbol{m}\| = 1$ and $m_1 = \boldsymbol{e}_1^T\boldsymbol{m}$, we have

$$\langle \boldsymbol{Y}_2, \boldsymbol{m}\boldsymbol{m}^T\rangle = -m_1^2\|\boldsymbol{m}_0\|^2 \leq 0.$$

This last equation implies that $(\boldsymbol{Y}_2)_{T^\perp} \preceq 0$.

Let $\boldsymbol{I} - \boldsymbol{Y}_1$ denote the matrix form of the polynomial certificate constructed in section 2.3. As we have $\mathbf{z}_{\mathcal{B}}^*(\boldsymbol{I} - \boldsymbol{Y}_1)\mathbf{z}_{\mathcal{B}} = \mathbf{z}_{\mathcal{B}}^*(\boldsymbol{I} - \boldsymbol{Y}_2)\mathbf{z}_{\mathcal{B}} = \sum_{|\boldsymbol{\gamma}|\leq 2}(\mathbf{z}^{\boldsymbol{\gamma}} - \mathbf{z}_0^{\boldsymbol{\gamma}})^2 = \sum_j s_j(\mathbf{z})^2$, Proposition 2 applies and there exists a matrix $\boldsymbol{\xi}$ satisfying $\boldsymbol{\xi} + Y_1 = Y_2$. Recall that the certificate reads

$$Y = \boldsymbol{I} - \boldsymbol{Y}_2 = \boldsymbol{I} - \boldsymbol{\xi} - \boldsymbol{Y}_1 = \boldsymbol{I} - \boldsymbol{\xi} - \sum_{\ell} \sum_{\boldsymbol{\kappa}\in\mathbb{N}^K_{2(t-d_{h_\ell})}} \lambda_{\ell\boldsymbol{\kappa}} \left(\sum_{\boldsymbol{\zeta}} \frac{(h_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\|_F^2}\boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}\right) \tag{3.23}$$

Now using $\mathrm{Tr}(\boldsymbol{H}) \leq 0$, we can write,

$$\begin{aligned}
0 &\geq \mathrm{Tr}(\boldsymbol{H}_T) + \mathrm{Tr}(\boldsymbol{H}_{T^\perp}) && (3.24)\\
&= \langle \boldsymbol{H}, \boldsymbol{I}_T\rangle + \langle \boldsymbol{H}, \boldsymbol{I}_{T^\perp}\rangle && (3.25)\\
&= \langle \boldsymbol{H}, \boldsymbol{I}_T\rangle - \langle \boldsymbol{H}, \boldsymbol{Y}_2\rangle + \langle \boldsymbol{H}, \boldsymbol{Y}_2\rangle + \langle \boldsymbol{H}, \boldsymbol{I}_{T^\perp}\rangle && (3.26)\\
&= \langle \boldsymbol{H}_T, \boldsymbol{I}_T - (\boldsymbol{Y}_2)_T\rangle - \langle \boldsymbol{H}_{T^\perp}, (\boldsymbol{Y}_2)_{T^\perp}\rangle + \langle \boldsymbol{H}, \boldsymbol{Y}_2\rangle + \langle \boldsymbol{H}, \boldsymbol{I}_{T^\perp}\rangle && (3.27)\\
&\geq -\langle \boldsymbol{H}_{T^\perp}, (\boldsymbol{Y}_2)_{T^\perp}\rangle - |\langle \boldsymbol{H}, \boldsymbol{Y}_2\rangle| + \langle \boldsymbol{H}, \boldsymbol{I}_{T^\perp}\rangle && (3.28)\\
&\geq -\langle \boldsymbol{H}_{T^\perp}, (\boldsymbol{Y}_2)_{T^\perp}\rangle - |\langle \boldsymbol{H}, \boldsymbol{Y}_1\rangle| + \langle \boldsymbol{H}, \boldsymbol{I}_{T^\perp}\rangle && (3.29)\\
&\geq -|\langle \boldsymbol{H}, \boldsymbol{Y}_1\rangle| + \mathrm{Tr}(\boldsymbol{H}_{T^\perp}) && (3.30)
\end{aligned}$$

As explained above, $\boldsymbol{Y}_1$ is used to denote the component of the dual certificate which is in the range of $\mathcal{A}^*$, i.e. $\boldsymbol{Y}_1 = \mathcal{A}^*\lambda$. (3.27) follows from $\boldsymbol{Y}_T = 0$. In (3.29), we use the fact that $\boldsymbol{Y}_1 = \boldsymbol{Y}_2 + \boldsymbol{\xi}$ and $\langle \boldsymbol{\xi}, \boldsymbol{B}_{\boldsymbol{\gamma}}\rangle = 0$, for all $\boldsymbol{\gamma}$. Since both $\boldsymbol{M}$ and $\boldsymbol{M}_0$ are solutions to problem (3.1), both of these matrices satisfy the structural constraints exactly, and read $\boldsymbol{M} = \sum_{\boldsymbol{\gamma}} m_\gamma \boldsymbol{B}_{\boldsymbol{\gamma}}$, $\boldsymbol{M}_0 = \sum_{\boldsymbol{\gamma}} (m_0)_{\boldsymbol{\gamma}}\boldsymbol{B}_\gamma$ for some m_γ. Together with lemma 2, this implies $\langle \boldsymbol{H}, \boldsymbol{\xi}\rangle = \langle \boldsymbol{M} - \boldsymbol{M}_0, \boldsymbol{\xi}\rangle = 0$. Finally, in (3.30), we use the fact that for a positive semidefinite matrix $\boldsymbol{H}_{T^\perp}$, and a matrix $(\boldsymbol{Y}_2)_{T^\perp}$ such that $(\boldsymbol{Y}_2)_{T^\perp} \preceq 0$, $\langle \boldsymbol{H}_{T^\perp}, (\boldsymbol{Y}_2)_{T^\perp}\rangle \leq 0$.

From (3.16) and (3.21), we then have

$$\mathrm{Tr}(\boldsymbol{H}_{T^\perp}) \leq |\langle \boldsymbol{H}, \boldsymbol{Y}_1\rangle| \leq C_5(\delta_L, \delta_U) \cdot (m+n)^{3/2} \cdot \eta \cdot \|\boldsymbol{z}_0\|_2 \tag{3.31}$$

3.1.2 Bound on $\boldsymbol{H}_T$

We now use the injectivity of the linear map $\mathcal{A}$, encoding the polynomial constraints, on the tangent space T to derive a bound on $\boldsymbol{H}_T$. Let $\boldsymbol{H}_T$ be expressed as $\boldsymbol{H}_T = \boldsymbol{y}\boldsymbol{m}_0^T + \boldsymbol{m}_0\boldsymbol{y}^T$ for some $\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}_2^K}$ (see (2.4)).

Using this decomposition for $\boldsymbol{H}_T$, and letting $h_{i_1} \to h_{i_2} \to \dots$ denote the ordered series of constraints making the chain (1.4), we have

$$
\begin{aligned}
y_{i_1} + y_1(\boldsymbol{z}_0)_{i_1} &= \left\{\tilde{\mathcal{A}}(\boldsymbol{H}_T)\right\}_1 = \sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_{i_1})_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{H}_T, \boldsymbol{B}_{\boldsymbol{\zeta}}\rangle\\
y_{i_1}(\boldsymbol{z}_0)_{i_2} + y_{i_2}(\boldsymbol{z}_0)_{i_1} &= \left\{\tilde{\mathcal{A}}(\boldsymbol{H}_T)\right\}_2 = \sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_{i_2})_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{H}_T, \boldsymbol{B}_{\boldsymbol{\zeta}}\rangle\\
y_{i_2}(\boldsymbol{z}_0)_{i_3} + y_{i_3}(\boldsymbol{z}_0)_{i_2} &= \left\{\tilde{\mathcal{A}}(\boldsymbol{H}_T)\right\}_3 = \sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_{i_3})_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{H}_T, \boldsymbol{B}_{\boldsymbol{\zeta}}\rangle\\
y_{i_3}(\boldsymbol{z}_0)_{i_4} + y_{i_4}(\boldsymbol{z}_0)_{i_3} &= \dots
\end{aligned}
$$

To derive a bound for $\boldsymbol{H}_T$, we then isolate each of the entries in $\boldsymbol{y}$ as,

$$
\begin{aligned}
y_{i_1} &= \sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_{i_1})_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{H}_T, \boldsymbol{B}_{\boldsymbol{\zeta}}\rangle - y_1(z_0)_{i_1}\\
y_{i_2} &= \frac{1}{(z_0)_{i_1}}\left[\sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_{i_2})_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{H}_T, \boldsymbol{B}_{\boldsymbol{\zeta}}\rangle - (z_0)_{i_2} y_{i_1}\right]\\
y_{i_3} &= \frac{1}{(z_0)_{i_2}}\left[\sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_{i_3})_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{H}_T, \boldsymbol{B}_{\boldsymbol{\zeta}}\rangle - y_{i_2}(z_0)_{i_3}\right]\\
y_{i_4} &= \dots
\end{aligned}
\tag{3.32}
$$

The sequence of equalities in (3.32) thus gives a general expression for every first order entry y_{i_ℓ} of $\boldsymbol{y}$, as a weighted combination of the constraints

$$
y_{i_\ell} = \frac{1}{(\boldsymbol{z}_0)_{i_{\ell-1}}}\sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_{i_\ell})_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{H}_T, \boldsymbol{B}_{\boldsymbol{\zeta}}\rangle \tag{3.33}
$$

$$
\pm \sum_{j=1}^{\ell-1} \frac{(\boldsymbol{z}_0)_{i_\ell}}{(\boldsymbol{z}_0)_{i_{\ell-j-1}}(\boldsymbol{z}_0)_{i_{\ell-j}}} \sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_j)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{H}_T, \boldsymbol{B}_{\boldsymbol{\zeta}}\rangle \tag{3.34}
$$

$$
\pm \frac{(z_0)_{i_\ell}}{(z_0)_{i_1}} \sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_{i_1})_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{H}_T, \boldsymbol{B}_{\boldsymbol{\zeta}}\rangle \pm (z_0)_{i_\ell}(\boldsymbol{H}_{T^\perp})_{11}. \tag{3.35}
$$

The last term in (3.35) follows from $y_1(z_0)_1 = (H_T)_{11} = H_{11} - (H_T^\perp)_{1,1} = -(H_T^\perp)_{1,1}$. Let $(\boldsymbol{y}_{\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}|\leq 1}$ denote the entries in $\boldsymbol{y}$ corresponding to the multi-indices that give rise to degree one monomials. Let C_6 bound each of the weights appearing in front of the constraints making up the chain in (3.35). The first order part of $\boldsymbol{y}$, $(\boldsymbol{y}_{\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}|\leq 1}$, has length $m+n$ and each of the terms appearing in this sum is generated by a different constraint. We can thus write

$$\begin{aligned}\|(\boldsymbol{y}_{\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}|\leq 1}\|_2 &\leq C_6(\delta_L,\delta_U)\cdot(m+n)^{1/2}\left(\sum_\ell\left|\sum_{|\boldsymbol{\zeta}|>0}\frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle\boldsymbol{H}_T,\boldsymbol{B}_{\boldsymbol{\zeta}}\rangle\right|+\|\boldsymbol{H}_{T^\perp}\|_1\right)\\ &\leq C_6(\delta_L,\delta_U)\cdot(m+n)^{1/2}\left(\sum_\ell\left|\sum_{|\boldsymbol{\zeta}|>0}\frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle\boldsymbol{H}_T,\boldsymbol{B}_{\boldsymbol{\zeta}}\rangle\right|+\|\boldsymbol{H}_{T^\perp}\|_1\right).\end{aligned} \tag{3.36}$$

The constant $C_6(\delta_L,\delta_U)$ depends on the entries of $\boldsymbol{X}_0$ but not on m or n. To bound the second order entries in $\boldsymbol{y}$, we rely on the structural constraints $(\boldsymbol{H})_{\boldsymbol{\alpha},\boldsymbol{\beta}} = (\boldsymbol{H})_{\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{0}}$ for $|\boldsymbol{\alpha}|=|\boldsymbol{\beta}|=1$. Those constraints are unaffected by the noise. We further use $\boldsymbol{H}=\boldsymbol{H}_T+\boldsymbol{H}_T^\perp$, as well as $\boldsymbol{H}_T=\boldsymbol{y}\boldsymbol{m}_0^T+\boldsymbol{m}_0\boldsymbol{y}^T$ for some $\boldsymbol{y}$.

$$\begin{aligned}0 &= (\boldsymbol{H})_{\boldsymbol{\alpha},\boldsymbol{\beta}}-(\boldsymbol{H})_{\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{0}} && (3.37)\\ &= (\boldsymbol{H}_T)_{\boldsymbol{\alpha},\boldsymbol{\beta}}-(\boldsymbol{H}_T)_{\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{0}} && (3.38)\\ &+ (\boldsymbol{H}_T^\perp)_{\boldsymbol{\alpha},\boldsymbol{\beta}}-(\boldsymbol{H}_T^\perp)_{\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{0}} && (3.39)\\ &= \boldsymbol{y}_{\boldsymbol{\alpha}}(\boldsymbol{m}_0)_{\boldsymbol{\beta}}+\boldsymbol{y}_{\boldsymbol{\beta}}(\boldsymbol{m}_0)_{\boldsymbol{\alpha}}-\boldsymbol{y}_{\boldsymbol{\alpha}+\boldsymbol{\beta}}(\boldsymbol{m}_0)_0 && (3.40)\\ &-(\boldsymbol{y})_{\boldsymbol{0}}(\boldsymbol{m}_0)_{\boldsymbol{\alpha}+\boldsymbol{\beta}}+(\boldsymbol{H}_T^\perp)_{\boldsymbol{\alpha},\boldsymbol{\beta}}-(\boldsymbol{H}_T^\perp)_{\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{0}} && (3.41)\end{aligned}$$

Using those two lines, we can then bound the second order entries in $\boldsymbol{y}$ as

$$|\boldsymbol{y}_{\boldsymbol{\alpha}+\boldsymbol{\beta}}(\boldsymbol{m}_0)_{\boldsymbol{0}}| \leq C_7\left\{\|\boldsymbol{m}_0\|_\infty\left(\|\boldsymbol{H}_T^\perp\|_1+\sup|\boldsymbol{y}_{\boldsymbol{\alpha}}|\right)+\left|\langle\boldsymbol{H}_T^\perp,\boldsymbol{E}_{\boldsymbol{\alpha},\boldsymbol{\beta}}-\boldsymbol{E}_{\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{0}}\rangle\right|\right\} \tag{3.42}$$

$$\leq 2C_8(\delta_L,\delta_U)\cdot\left(\sum_\ell\left|\sum_{|\boldsymbol{\zeta}|>0}\frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle\boldsymbol{H}_T,\boldsymbol{B}_{\boldsymbol{\zeta}}\rangle\right|+\|\boldsymbol{H}_T^\perp\|_1\right) \tag{3.43}$$

In the lines above, we used

$$|\boldsymbol{y}_{\boldsymbol{0}}(\boldsymbol{m}_0)_{\boldsymbol{0}}| = |(\boldsymbol{H}_T)_{11}| = \left|\boldsymbol{H}_{11}-(\boldsymbol{H}_T^\perp)_{11}\right| \leq \left|(\boldsymbol{H}_T^\perp)_{11}\right| \leq \|\boldsymbol{H}_T^\perp\|_1 \tag{3.44}$$

From this we can thus control the norm of $\boldsymbol{y}$ as

$$\|\boldsymbol{y}\| \leq (m+n)\cdot C_9(\delta_L,\delta_U)\left(\sum_\ell\left|\sum_{|\boldsymbol{\zeta}|>0}\frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle\boldsymbol{H}_T,\boldsymbol{B}_{\boldsymbol{\zeta}}\rangle\right|+\|\boldsymbol{H}_T^\perp\|_1\right) \tag{3.45}$$

and hence

$$\|\boldsymbol{H}_T\|_F \leq \|\boldsymbol{m}_0\|_2\|\boldsymbol{y}\| \tag{3.46}$$

$$\leq \|\boldsymbol{m}_0\|(m+n)\cdot C_9(\delta_L,\delta_U)\left(\sum_\ell\left|\sum_{|\boldsymbol{\zeta}|>0}\frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle\boldsymbol{H}_T,\boldsymbol{B}_{\boldsymbol{\zeta}}\rangle\right|+\|\boldsymbol{H}_T^\perp\|_1\right) \tag{3.47}$$

To bound the last factor, we use $\boldsymbol{H}_T = \boldsymbol{H} - \boldsymbol{H}_T^\perp$ together with

$$\sum_\ell \left| \sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2} \langle \boldsymbol{H}, \boldsymbol{B}_{\boldsymbol{\zeta}} \rangle \right| \leq \eta \cdot (m+n)^{1/2} \tag{3.48}$$

which follows from the second constraint in (1.26). Then we have

$$\sum_\ell \left| \sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2} \langle \boldsymbol{H}_T^\perp, \boldsymbol{B}_{\boldsymbol{\zeta}} \rangle \right| = \sum_\ell \sum_{|\boldsymbol{\zeta}|>0} \frac{|(\tilde{h}_\ell)_{\boldsymbol{\zeta}}|}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2} \sum_{\boldsymbol{\alpha}-\boldsymbol{\beta}=\boldsymbol{\zeta}} |\boldsymbol{H}_T^\perp|_{\boldsymbol{\alpha},\boldsymbol{\beta}} \tag{3.49}$$

$$\leq C_{10} \cdot (m+n)^{1/2} \|\boldsymbol{H}_T^\perp\|_F \tag{3.50}$$

$$\leq C_{10} \cdot (m+n)^{1/2} \|\boldsymbol{H}_T^\perp\|_1 \tag{3.51}$$

In (3.49), we use $|\boldsymbol{H}_T^\perp|$ to denote the entry-wise absolute value of $\boldsymbol{H}_T^\perp$. The last line follows from the fact that we consider $(m+n)$ constraints, each accessing distinct elements in $\boldsymbol{H}_T^\perp$. The result follows from an application of the Cauchy-Schwarz inequality. Grouping (3.47) as well as (3.48) and (3.51), with $C_{11} = C_{11}(\delta_L, \delta_U)$, we have

$$\|(\boldsymbol{H}_T)\|_F \leq C_{11} \cdot (m+n) \|\boldsymbol{m}_0\|_2 \left((m+n)^{1/2} \|\boldsymbol{H}_T^\perp\|_1 + 2\eta (m+n)^{1/2} \right) \tag{3.52}$$

$$\leq C_{11} \cdot (m+n)^{3/2} \|\boldsymbol{m}_0\|_2 \left(|\langle \boldsymbol{H}, \mathbf{Y}_1 \rangle| + 2\eta \right) \tag{3.53}$$

$$\leq C'_{11} \cdot (m+n)^{5/2} \|\boldsymbol{m}_0\|^2 \max\{\eta, \|\boldsymbol{z}_0\| \|\boldsymbol{\varepsilon}\|\} \tag{3.54}$$

In (3.54), we use $\|\boldsymbol{m}_0\|_2 \leq C(\delta_L, \delta_U) \cdot (m+n)$ for a constant $C(\delta_L, \delta_U)$ that only depends on the magnitude of $|x_0|, |y_0|$. Using this last bound along with (3.31), we finally get,

$$\begin{aligned} \|\boldsymbol{H}\|_F &\leq \|\boldsymbol{H}_T\|_F + \|\boldsymbol{H}_{T^\perp}\|_F \\ &\leq \|\boldsymbol{H}_T\|_F + \|\boldsymbol{H}_{T^\perp}\|_1 \\ &\leq C(\delta_L, \delta_U) \cdot (m+n)^{5/2} \eta \cdot \|\boldsymbol{M}_0\|_F. \end{aligned}$$

The constant $C(\delta_L, \delta_U)$ only depends on the magnitude of the entries of x_0 and y_0 and not on m or n.

The next section shows how the Lipschitz stability constant can be reduced when the paths connecting the root nodes (corresponding to constraints of the form $h_\ell(\mathbf{z}) \equiv z_\ell - (z_0)_\ell = 0$) and the leaf nodes are explicitly given.

3.2 Proof of Corollary 1

The proof of Corollary 1 follows the idea of section 3.1 with the difference that we now consider a reduced semidefinite program with a path specific bound of the form (3.3).

The moments matrix has now size $O(m+n) \times O(m+n)$. The part of the certificate expressing first order squares remains unchanged. The second order squares can be written directly from the constraints, trace and constant ρ without the need for any propagation, i.e. $(\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}^{\boldsymbol{\beta}} - \mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}})^2 = -2\mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}}(\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}^{\boldsymbol{\beta}} - \mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}}) + (\mathbf{z}^{\boldsymbol{\alpha}}\mathbf{z}^{\boldsymbol{\beta}})^2 - (\mathbf{z}_0^{\boldsymbol{\alpha}}\mathbf{z}_0^{\boldsymbol{\beta}})^2$. The certificate thus becomes much sparser. Conditions 1) to 3) still hold for this certificate as it still has the exact same structure $\sum_j \boldsymbol{s}_j\boldsymbol{s}_j^*$ as before except that the number of such squares is reduced. The squared polynomials are now given by $(\mathbf{z}^{\boldsymbol{\alpha}} - \mathbf{z}_0^{\boldsymbol{\alpha}})^2$, for all $|\boldsymbol{\alpha}| \leq 1$ and $(\mathbf{z}^{\boldsymbol{\gamma}} - \mathbf{z}_0^{\boldsymbol{\gamma}})^2$ for all $|\boldsymbol{\gamma}| = 2$ such that $\mathbf{z}^{\boldsymbol{\gamma}} - \mathbf{z}_0^{\boldsymbol{\gamma}}$ appears in the constraints. As a consequence, the matrix has now size $|\{\boldsymbol{\alpha},\ |\boldsymbol{\alpha}| \leq 1\}| + |\{\boldsymbol{\gamma} \in \Omega\}|$. The rank condition in 2) still holds as well.

As for the proof of Theorem 1, we let $\boldsymbol{Y}_1^{(1)}$ and $\boldsymbol{Y}_1^{(2)}$, with $\boldsymbol{Y}_1 = \boldsymbol{Y}_1^{(1)} + \boldsymbol{Y}_1^{(2)}$ denote the contributions of first and second order squares to the certificate $\boldsymbol{Y}_1$. To bound the inner product $|\langle \boldsymbol{H}, \boldsymbol{Y}\rangle|$, we once again replace the noiseless constraints h_ℓ appearing in the expression of the certificate with their noisy versions, $\tilde{h}_\ell$, that are bounded through (3.3). For the first order contribution $\boldsymbol{Y}_1^{(1)}$, we have,

$$|\langle \boldsymbol{Y}_1^{(1)}, \boldsymbol{H}\rangle| \leq \left|\sum_{\kappa}\sum_{\ell} W_{\ell,\kappa}\sum_{\boldsymbol{\zeta}}(\tilde{h}_\ell)_{\boldsymbol{\zeta}}\langle \boldsymbol{B}_{\boldsymbol{\zeta}+\boldsymbol{\kappa}}, \boldsymbol{M}\rangle\right| + \left|\sum_{|\boldsymbol{\kappa}|\leq 1}\sum_{\ell} W_{\ell,\boldsymbol{\kappa}}\varepsilon_\ell\langle \boldsymbol{B}_{\boldsymbol{\kappa}}, \boldsymbol{M}\rangle\right| \tag{3.55}$$

$$\leq (m+n)^{3/2} C_1(\delta_L, \delta_U) \max\left\{\eta', \|\boldsymbol{m}_0\| \|\varepsilon\|_\infty\right\} \tag{3.56}$$

The bound (3.56) follows the exact same reasoning as (3.15). For the second term, following the proof of Theorem 1, noting that we now only use second order moments appearing in the constraints, we can write $|\langle \boldsymbol{Y}_1^{(2)}, \boldsymbol{H}\rangle| \leq C(\delta_L, \delta_U)\cdot(m+n)^{1/2}\|\varepsilon\| \leq C(\delta_L, \delta_U)\cdot(m+n)^{1/2}\eta'$. From this we write

$$|\langle \boldsymbol{H}, \boldsymbol{Y}\rangle| \leq (m+n)^{3/2} C_1(\delta_L, \delta_U) \max\left\{\eta', \|\boldsymbol{m}_0\| \|\varepsilon\|_\infty\right\} \tag{3.57}$$

In a similar way, the expression for the y_{i_k} in (3.32) also relies on the first order constraints making up the path from the root node to y_{i_k} so that the various relations and bounds in (3.32) and (3.36) can now be reduced to

$$\|\boldsymbol{y}_{|\boldsymbol{\alpha}|\leq 1}\|_2 \leq C_2\cdot\sqrt{m+n}\left(\sup_{\mathcal{P}_i}\sum_{\ell\in\mathcal{P}_i}\left|\left(\sum_{|\boldsymbol{\zeta}|>0}\frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{B}_{\boldsymbol{\zeta}}, \boldsymbol{H}_T\rangle\right)\right| + \|\boldsymbol{H}_T^\perp\|_1\right). \tag{3.58}$$

Relations (3.40) and (3.41) still hold. Since we consider a reduced moments matrix, we only need to account for the $O(m+n)$ second order monomials appearing in the telescoping relations, and one can thus simply bound the second order part of $\boldsymbol{y}$ as

$$\|\boldsymbol{y}_{\boldsymbol{\alpha}+\boldsymbol{\beta}}\|_2 \leq C_3(\delta_L, \delta_U)\cdot\sqrt{m+n}\sup_{\mathcal{P}_i}\sum_{\ell\in\mathcal{P}_i}\left|\left(\sum_{|\boldsymbol{\zeta}|>0}\frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2}\langle \boldsymbol{B}_{\boldsymbol{\zeta}}, \boldsymbol{H}_T\rangle\right)\right| \tag{3.59}$$

$$+ C_3(\delta_L, \delta_U)\cdot\sqrt{m+n}\cdot\|\boldsymbol{H}_{T^\perp}\|_1 \tag{3.60}$$

Grouping (3.58) and (3.60), the bound on $\boldsymbol{H}_T$ can therefore read,

$$\|\boldsymbol{H}_T\|_F \leq \sqrt{m+n}\|\boldsymbol{m}_0\|_2 \cdot C_4(\delta_L, \delta_U) \cdot \sup_{\mathcal{P}_i} \sum_{\ell \in \mathcal{P}_i} \left| \left(\sum_{|\boldsymbol{\zeta}|>0} \frac{(\tilde{h}_\ell)_{\boldsymbol{\zeta}}}{\|\boldsymbol{B}_{\boldsymbol{\zeta}}\|_F^2} \langle \boldsymbol{B}_{\boldsymbol{\zeta}}, \boldsymbol{H}_T \rangle \right) \right| \tag{3.61}$$

$$+ \sqrt{m+n}\|\boldsymbol{m}_0\|_2 \cdot C_4'(\delta_L, \delta_U) \cdot \|\boldsymbol{H}_{T^\perp}\|_1 \tag{3.62}$$

Using (3.57) together with (3.30), we have

$$\|(\boldsymbol{H}_T)\|_F \leq C_5 \cdot (m+n)\|\boldsymbol{m}_0\|_2 \left(\eta' + (m+n)^{3/2} \max\left\{\eta', \|\boldsymbol{m}_0\|\|\varepsilon\|_\infty\right\}\right) \tag{3.63}$$

$$\leq C_6 \cdot (m+n)^2 \eta' \|\boldsymbol{M}_0\|_F \tag{3.64}$$

Where $C_6 = C_6(\delta_L, \delta_U)$ is a constant that only depends on the entries of $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$.

4 Numerical Simulations

In this section we illustrate the stability of the semidefinite program (2.1) and discuss its scalability through low rank factorizations. Section 4.1 starts by providing a comparison of the stability and recovery guarantees of the level 2 Lasserre relaxation against traditional approaches such as nuclear norm minimization, nonlinear propagation, and ridge regression. Sections 4.2 to 4.5 then discuss the efficiency of scalable numerical schemes. Simply listing the moments up to order 4 has complexity $O(K^4)$, and hence is not a scalable representation of the moments matrix. The traditional remedy is the factorized gradient approach due to Burer and Monteiro [5, 6]. Our first numerical observation will however not be a surprise to the specialist: difficult instances of matrix completion lead to the presence of spurious local minimizers. With adequate compression of the variables and constraints, and provided convergence is to the global minimizer, we show how the problem can be solved in an empirical $O(K^2)$ complexity.

The conclusions of this section can be summarized as follows.

- Factorization approaches sometimes introduce spurious minimizers for sufficiently difficult (small δ) problems. When convergence to such minimizers occurs, it is sometimes possible to extract the solution from a subset of the moments rather than to consider the whole matrix (see Fig. 5).
- Using a simple low rank factorization for the moments matrix still leads to a storage complexity $O(K^2)$, hence is not fully scalable. We propose instead to view the moments matrix as a tensor, and upgrade to a more efficient hierarchical low-rank factorization with storage complexity $O(K)$. This factorization seems to always work when the simpler factorized gradient works.

- The hierarchical factorization is in itself not sufficient to guarantee scalability. Indeed, formulation (2.1), and in particular moments (i.e. total symmetry) constraints, still require encoding a combinatorial ($O(K^4)$) number of equalities. To resolve this difficulty, Section 4.4 introduces three different trace relations, which are derived from the third and fourth order total symmetry constraints. Enforcing those relations in place of the original total symmetry constraints reduces the computational cost from $O(K^4)$ to $O(1)$ in the best case. Empirically, we observe that the trace relations can be used as a substitute for the more expensive total symmetry constraints as soon as the traditional factorized gradient method works.
- Given the $O(1)$ trace relations and the hierarchical low rank factorization of the moments tensor, a last bottleneck that prevents reducing the global computational cost from $O(K^3)$ to $O(K^2)$ is given by the Higher Order Affine constraints (i.e. the shifted moments constraints obtained by multiplying any of the original polynomial equations $h_1(z) = 0$ to $h_K(z) = 0$, by any monomial of degree at most two). Enforcing those constraints requires storing matrices of size $O(K^3)$. We suggest to encode these constraints through random sampling, minimizing over distinct batches of size $\mathcal{O}(K)$ iteratively. Such a formulation does not seem to modify the convergence properties and enables us to apply the semidefinite program (2.1) to matrices $\boldsymbol{X}$ of size 100×100.

4.1 Lipschitz stability

To illustrate how the noise can affect a nonlinear reconstruction in the propagation framework, we conduct the following experiment. We consider a noise vector $\boldsymbol{\varepsilon} = \gamma \boldsymbol{n}/\|\boldsymbol{n}\|$ for $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_3)$. We gradually increase the amplitude γ of the noise vector. For this noise vector, we let $\boldsymbol{M}_P$ denote the solution obtained through propagation and $\boldsymbol{M}_L$ the solution obtained through the stable semidefinite relaxation (1.26) for $d = 2$. We consider the matrix of example (1.6) for which we let $\delta = .01$. The numerical experiments are then repeated as follows.

- We randomly draw the noise vector $n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_3)$.
- The noise vector is multiplied by the scaling coefficient γ taking values between .001 and .01, so that the corruption is at most 100% of the signal. The noise is added to the entries $(X_0)_{11}$, $(X_0)_{22}$ and $(X_0)_{21}$ of $\boldsymbol{X}_0$ to define the (noisy) measurements.
- Our semidefinite programming relaxation is then solved with CVX[4] for the noisy measurements. We compute the difference between the returned solution $\boldsymbol{M}_L$ and the optimal solution to the noiseless problem $\boldsymbol{M}_0$ through the Frobenius norm as $\|\boldsymbol{M}_L - \boldsymbol{M}_0\|_F/\|\boldsymbol{M}_0\|_F$.
- The equivalent solution obtained through nonlinear propagation is computed and compared to $\boldsymbol{M}_0$ as $\|\boldsymbol{M}_P - \boldsymbol{M}_0\|_F/\|\boldsymbol{M}_0\|_F$.

[4] `http://cvxr.com/about/`

Those various steps are repeated for the various noise levels and for a collection of random vectors $\boldsymbol{n}$. Note that, as we are in the framework of example (1.6), nuclear norm fails even in the absence of noise. For each choice of γ, the relative errors $\|\boldsymbol{M}_P - \boldsymbol{M}_0\|/\|\boldsymbol{M}_0\|$ and $\|\boldsymbol{M}_L - \boldsymbol{M}_0\|_F/\|\boldsymbol{M}_0\|_F$ are averaged over all the noise vectors. The results are shown in Fig. 2. This figure thus illustrates the evolution of the averaged relative errors $\mathbb{E}_{\boldsymbol{n}}\|\boldsymbol{M}_L - \boldsymbol{M}_0\|/\|\boldsymbol{M}_0\|$ and $\mathbb{E}_{\boldsymbol{n}}\|\boldsymbol{M}_P - \boldsymbol{M}_0\|_F/\|\boldsymbol{M}_0\|_F$ for the semidefinite programming relaxation, and for nonlinear propagation in an instance where nuclear norm minimization fails. The SNR is measured in [dB] as $20\log(\delta/\gamma)$.

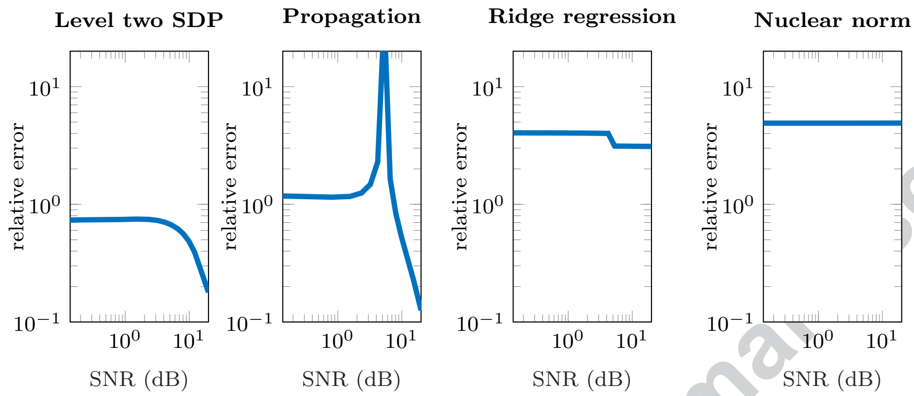

Fig. 2 Evolution of the relative error $\|\boldsymbol{M} - \boldsymbol{M}_0\|_F/\|\boldsymbol{M}_0\|_F$ as a function of the noise level (SNR [dB]) for the semidefinite program; nonlinear propagation; ridge regression; and nuclear norm minimization. Blowup can occur for nonlinear propagation whenever the noise takes on values that are close, yet opposite in sign, to the small entries in the matrix; as explained in section 1.1. Both ridge regression and nuclear norm minimization are known to fail, even in the absence of noise.

4.2 Toward scalability: low-rank factorization

Despite its interest in terms of stability, the semidefinite program (3.1) remains difficult to implement for practical problems because of the size of the second order moments matrix involved. Solving the completion problem on a matrix of size $K \times K$ through (3.1) requires storing a matrix of size K^4, which is difficult when considering sufficiently interesting instances. In this section, we introduce and discuss more scalable numerical methods based on low rank factorizations of the moment matrix (1.25). As is common in semidefinite programming, the recovery guarantees are however lost when passing to such formulations. This phenomenon is illustrated by Figs 3 to 5.

Among the most popular approaches of the last few years, one the of the most efficient, popularized by [5] encodes the unknown positive semidefinite matrix $\boldsymbol{M}$ from (2.1) as the product $\boldsymbol{M} \approx \boldsymbol{T}\boldsymbol{T}^*$, with $\boldsymbol{T}$ of size $(K + K^2) \times r$ for small r,

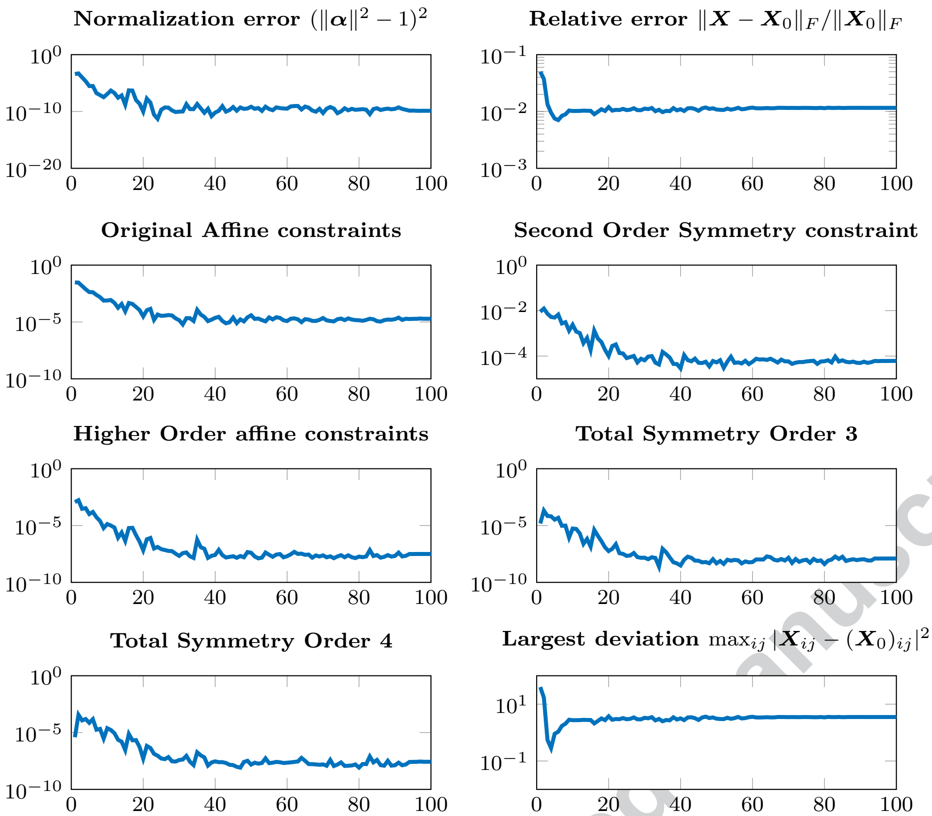

Fig. 3 Evolution of each of the error terms appearing in the augmented Lagrangian formulation (4.1) on the benchmark problem (1.6) with $\delta = 0.3$. The relative error (top right), together with the largest deviation, highlight the convergence to a local minimizer. All the other error terms, certifying feasibility, have already reached small thresholds. The plots above should also be compared to the evolution of the trace and global misfit shown in Fig. 4, and the structure of the local minimizer shown in Fig. 5.

and then minimizes the augmented Lagrangian over the factor $\boldsymbol{T}$. Note that in our case, $\boldsymbol{T}$ is of the form $\boldsymbol{T} = [\boldsymbol{\alpha}^T, \boldsymbol{R}^T, \boldsymbol{\Pi}^T]$ where $\boldsymbol{\alpha} \in \mathbb{R}^r$ encodes the normalizing constant, $\boldsymbol{R} \in \mathbb{R}^{K \times r}$ and $\boldsymbol{\Pi} = [\boldsymbol{\Pi}_1, \dots \boldsymbol{\Pi}_r] \in \mathbb{R}^{K^2 \times r}$ with $\boldsymbol{\Pi}_j \in \mathbb{R}^{N^2}$

The rank r of each of the factors can be constrained, and is increased when reaching local minimizers. If we let r denote the rank of the compressed matrix $\boldsymbol{M}$, such a formulation thus results in a reduction of the number of unknowns from $O(K^4)$ to only $O(K^2 \cdot r)$. The gradient steps are done with respect to $\boldsymbol{\alpha}, \boldsymbol{R}$ and $\boldsymbol{\Pi}$. For a set of constraints defined as $\langle \boldsymbol{A}_i, \boldsymbol{X} \rangle = b_i$ and encoded in the linear map $\mathcal{A} : \boldsymbol{X} \mapsto \mathcal{A}(\boldsymbol{X}) = \{\langle \boldsymbol{A}_i, \boldsymbol{X} \rangle\}_{i=1}^m$, a vector of multipliers $\boldsymbol{\lambda} \in \mathbb{R}^m$ and penalty term $\sigma \in \mathbb{R}$, the augmented Lagrangian function corresponding to a minimization

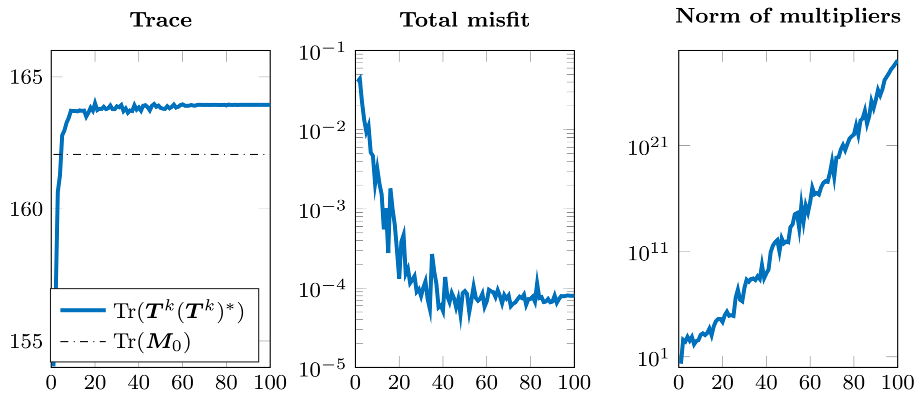

Fig. 4 Evolution of the Trace, misfit and norm of the Lagrange multipliers. The blowup in the norm of the multipliers confirms that we leave the regime in which Theorem 5.4 in [6] works, and thus lose the recovery guarantees that are following from this theorem.

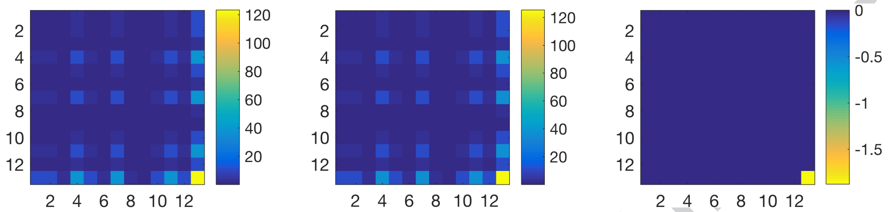

Fig. 5 Despite the existence of local minima for sufficiently difficult formulations (e.g. small δ in (1.6)), the solutions returned by the minimization of the low rank or hierarchical low rank Lagrangian, sometimes remain relatively close to the global solution except for the highest order moments. This figure illustrates this phenomenon. The exact moments matrix M_0 for the simple example of (1.6) with $\delta = 0.3$ is shown on the left. The moments matrix M recovered through low rank factorization is shown in the center and the difference $M - M_0$ is shown on the right. In practice, for a problem that is not too difficult (i.e. a sufficiently large value of δ in (1.6)) it is possible to recover the global solution by simply extracting the lower order moments.

of the trace under the linear constraints $\mathcal{A}(TT^*) = b$ reads

$$\mathcal{L}(T, \lambda, \sigma) = \|T\|_F^2 - 2\sum_{i=1}^{m} \lambda_i \left(\langle A_i, TT^*\rangle - b_i\right) + \sigma \sum_{i=1}^{m} |\langle A_i, TT^*\rangle - b_i|^2 \quad (4.1)$$

We then minimize this objective following the approach introduced in [5].

4.3 From low rank to hierarchical low rank

In order to further improve the efficiency, it is possible to consider an additional level of factorization for the columns of Π. Since those columns are encoding the moments of order 2, we can view them as vectorized matrices, and encode each of those matrices as low rank factorizations. This leads to a multi-level or *hierarchical encoding* of the moments matrix.

The resulting (i.e Tucker) factorization now relies on two dynamic ranks. The first rank, r_1, controls the factorization of the moments matrix as a whole. The second rank, r_2, controls the factorization of the fourth order moments which are stored as the factors $\{\boldsymbol{S}_{k,\ell}\}_{(k,\ell)\in[r_2]\times[r_1]}$ satisfying

$$\boldsymbol{\Pi}_\ell = \sum_{k=1}^{r_2} \boldsymbol{S}_{k,\ell}\boldsymbol{S}_{k,\ell}^T, \qquad \ell = 1,\dots,r_1. \tag{4.2}$$

Optimization is performed on the objective obtained by substituting this multi-level factorization in (4.1). The power of the hierarchical low-rank idea lies in its scalability, and the fact that it can be applied recursively to higher-degree moments matrices, thus potentially enabling scalable optimization over higher levels of semidefinite programming relaxation. The next section discusses how the combinatorial total symmetry constraints can be enforced efficiently as well.

4.4 Replacing total symmetry with trace relations

In this section, we discuss three trace relations which can be used as scalable substitutes to the more computationally expensive third and fourth order total symmetry constraints. We provide numerical evidence that whenever the factorized gradient method works, enforcing those trace relations in place of total symmetry works just as well, yet reduces the computational (combinatorial) cost of those constraints from $O(K^4)$ to $O(1)$. Those relations seem to work best when applied to the multilevel low rank factorization.

The total symmetry constraints are used to encode equality of the entries of $\boldsymbol{M}$ that correspond to the same moments (see the discussion in section 1.4). When applied to the third and fourth order moments tensors, those constraints enforce equality between any permutation of the multi-index. I.e if $\boldsymbol{M}^{(3)}$ and $\boldsymbol{M}^{(4)}$ encode the third and fourth order blocks in $\boldsymbol{M}$, then those constraints require that for any 3-tuple (i,j,k) and permutation π, $(\boldsymbol{M}^{(3)})_{i,j,k} = (\boldsymbol{M}^{(3)})_{\pi(i,j,k)}$. Similarly, on the fourth order block, for any 4-tuple (i,j,k,ℓ) and any permutation π, the moments matrix must satisfy $(\boldsymbol{M}^{(4)})_{i,j,k,\ell} = (\boldsymbol{M}^{(4)})_{\pi(i,j,k,\ell)}$.

One of the implications of the total symmetry constraints is that the contraction of any fourth order block does not depend on the indices over which this contraction is taken. In other words, the sum $\sum_{i=1}^N \boldsymbol{M}^{(4)}_{iijk}$ is the same as the sum $\sum_{i=1}^N \boldsymbol{M}^{(4)}_{jiik}$, and so it is for any of the sums $\sum_{i=1}^N \boldsymbol{M}^{(4)}_{\pi(iijk)}$ for any permutation operator $\pi : K^4 \mapsto K^4$. When assuming that the tensor $\boldsymbol{M}^{(4)}$ is rank one, that is $\boldsymbol{M}_0^{(4)} = \text{vec}(\mathbf{z}_0 \otimes \mathbf{z}_0)\text{vec}(\mathbf{z}_0 \otimes \mathbf{z}_0)^T$, those constraints can be used to derive interesting trace relations on the second order tensor $\boldsymbol{M}^{(2)}$. For $\boldsymbol{M}^{(4)} = \boldsymbol{M}^{(2)} \otimes \boldsymbol{M}^{(2)}$,

$\sum_{i=1}^N \boldsymbol{M}^{(4)}_{iijk} = \sum_{i=1}^N \boldsymbol{M}^{(4)}_{jiik}$ in particular implies the following relation on $\boldsymbol{M}^{(2)}$,

$$\mathrm{Tr}(\boldsymbol{M}^{(2)})\boldsymbol{M}^{(2)} = (\boldsymbol{M}^{(2)})^2. \tag{4.3}$$

Linearizing this relation brings us back to enforcing equality of the contractions $\sum_{i=1}^N \boldsymbol{M}^4_{iijk} - \sum_{i=1}^N \boldsymbol{M}^4_{jiik} = 0$. Moreover, this first contraction can be enforced very efficiently on the low rank factorization of $\boldsymbol{M}^{(4)}$,

$$\sum_{k=1}^{r_1} \mathrm{Tr}(\mathrm{Mat}(\boldsymbol{\Pi}_k))\mathrm{Mat}(\boldsymbol{\Pi}_k) = \sum_{k=1}^{r} \mathrm{Mat}(\boldsymbol{\Pi}_k)\mathrm{Mat}(\boldsymbol{\Pi}_k) \tag{4.4}$$

A similar formulation holds for the hierarchical factorization. Using $\mathrm{Mat}(\boldsymbol{\Pi}_k) = \boldsymbol{S}_k\boldsymbol{S}_k^*$, we get

$$\sum_{k=1}^{r_1} \|\boldsymbol{S}_k\|_F^2 \boldsymbol{S}_k\boldsymbol{S}_k^* = \sum_{k=1}^{r_1} \boldsymbol{S}_k(\boldsymbol{S}_k^*\boldsymbol{S}_k)\boldsymbol{S}_k \tag{4.5}$$

$$= \sum_{k=1}^{r_1} \boldsymbol{S}_k \boldsymbol{\Delta}_{k,k} \boldsymbol{S}_k^* \tag{4.6}$$

The natural extension to (4.3) is to take a second contraction with respect to the indices remaining in this first constraint. This gives a second trace relation that requires the trace of the squared matrix to match the square of this matrix trace,

$$\mathrm{Tr}(\boldsymbol{M}^{(2)})^2 = \mathrm{Tr}((\boldsymbol{M}^{(2)})^2). \tag{4.7}$$

Relation (4.7) reduces the whole set of symmetry constraints to a single equality that can again be enforced efficiently on the hierarchical low rank factors. Note that when enforced on positive semidefinite matrices, (4.7) is in fact equivalent to enforcing an exact rank one constraint, as it requires $(\sum_i \lambda_i)^2 = \sum_i \lambda_i^2$ for $\lambda_i \geq 0$. I.e.

$$\left\{\boldsymbol{X} \in \mathbb{S}_N^+ \;:\; \mathrm{Tr}(\boldsymbol{X})^2 = \mathrm{Tr}(\boldsymbol{X}^2)\right\} = \left\{\boldsymbol{X} \in \mathbb{S}_N^+ \;:\; \mathrm{rank}(\boldsymbol{X}) \leq 1\right\} \tag{4.8}$$

It is finally possible to consider a trace contraction for third order total symmetry constraints. At order 3, following from the constraints $\boldsymbol{M}^{(3)}_{ijk} = \boldsymbol{M}^{(3)}_{\pi(i,j,k)}$, taking a contraction of the indices gives

$$\mathrm{Tr}(\boldsymbol{M}^{(2)})\boldsymbol{M}^{(1)} = \boldsymbol{M}^{(2)}\boldsymbol{M}^{(1)} \tag{4.9}$$

Here we again let $\boldsymbol{M}^{(2)}$ denote the matrix encoding the second order moments $\boldsymbol{M}^{(2)} \approx \boldsymbol{M}^{(1)} \otimes \boldsymbol{M}^{(1)}$ and $\boldsymbol{M}^{(1)}$ encode the first order moments. This third relation can again be expressed compactly on the low rank factorizations,

$$\sum_{r=1}^{r_1} \mathrm{Tr}(\mathrm{Mat}(\boldsymbol{\Pi}_r))(\boldsymbol{R}_r) = \sum_{r=1}^{r_1} \mathrm{Mat}(\boldsymbol{\Pi}_r)\boldsymbol{R}_r \tag{4.10}$$

To study the result of replacing third and fourth order total symmetry constraints by the trace relations above, we apply those relations on the simple example (1.6) for $\delta = .5$ with the single low rank and hierarchical low rank factorizations respectively. The results are shown in Figs. 9 (low rank) and 10 (hierarchical low rank). When considering the simpler factorized gradient approach, it seems that replacing the full set of 4^{th} order symmetry constraints with the corresponding contraction can lead to a slight reduction in the accuracy. The total symmetry constraints are not entirely satisfied as highlighted by Fig. 9 and this results in a partial recovery of the fourth order tensor. It remains possible to extract the solution $\boldsymbol{X}_0$ from the second order moments. A comparison of the iterations of Fig. 9 and 10 suggests that replacing total symmetry by the relations (4.3), (4.7) and (4.9) performs best when used on the hierarchical low rank factorization.

Generally speaking, it again seems that when the factorized gradient approach converges, which typically happens on problems that are not too difficult (i.e δ not too small), it always seems possible to replace the combinatorial Total Symmetry constraints by the more tractable trace relations on both the 3^{rd} and 4^{th} order tensors, and to recover the solution in both the low rank and hierarchical low rank frameworks. As we do not have empirical evidence that choosing the contraction (4.3) over the contraction (4.7) will lead to stronger recovery guarantees, we will always favor the former over the latter, as this one reduces to a single equation. On the remaining large scale examples of this paper, we thus always replace total symmetry constraints with relation (4.9) (third order moments) and relation (4.7) (fourth order moments).

To illustrate the interest of the combination of a multi-level low rank decomposition and of the trace relations (4.7) and (4.9) on practical instances, we now apply this combination to a first large scale example. On this example, both nuclear norm and ridge regression fail to recover the solution. We take the entries to be bounded as $\delta \leq \mathbf{z}_0 \leq 1$, with $\delta = .1$. The bipartite graph defining the measurements is represented in Fig. 8. This graph is generated at random while enforced to span the $(m+n)$ vertices with a minimal number of edges. For this particular problem, the solution returned by nuclear norm minimization gives a relative error $\|\boldsymbol{X} - \boldsymbol{X}_0\|_F / \|\boldsymbol{X}_0\|_F = 0.7382$. The iterates returned by the ridge regression formulation (1.11) are displayed in Fig. 6 (relative error and data misfit). We do not consider any regularization in this case).

The iterations following from optimizing over the hierarchical low rank formulation with the trace relations (4.7) and (4.9) are displayed in Fig. 7.

4.5 Subsampling the higher order affine constraints

A last computational bottleneck which prevents the application of the hierarchical formulation of section (4.3) to larger matrices comes from the higher-order affine constraints. Those constraints can be written as the matrix product $\boldsymbol{A}\boldsymbol{T}\boldsymbol{T}^* = 0$ where $\boldsymbol{T}$ encodes the low rank factor of size $O(K^2)$ and $\boldsymbol{A}$ simply applies the affine constraints to the second order part of $\boldsymbol{T}$, columnwise. To further reduce the computational cost, we propose to draw smaller $O(K)$ "batches" of monomials $\mathcal{S}_i \subseteq [K]\times[K]$ from the full set of second order monomials appearing in $\boldsymbol{T}$. We then minimize the resulting augmented Lagrangian resulting from each $\mathcal{S}_i$ sequentially.

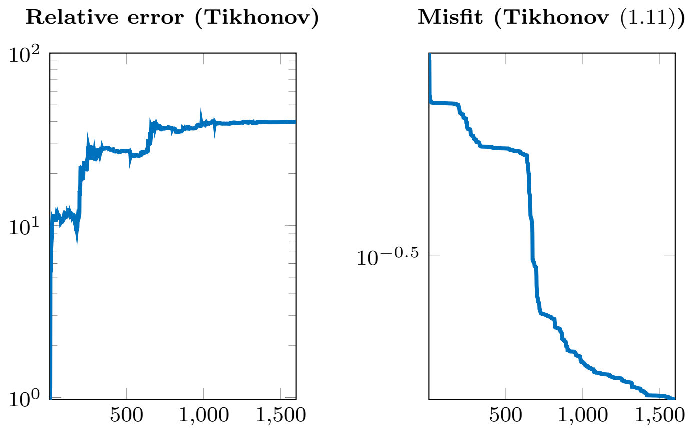

Fig. 6 It is possible to find instances of the rank one matrix completion problem for which even the ridge regression formulation will not be able to return the global minimizer. Such instances are more frequent when considering matrices of large size. In this particular case, ridge regression is clearly shown to converge to the wrong minimizer. The level 2 Lasserre relaxation, on the other hand, returns the true solution, even when considering a hierarchical low rank factorization with strong constraint on the rank (2 in this case) (see Fig. 7). The matrix considered here is 20×20 with entries bounded as $\delta \leq (X_0)_{ij} \leq 1$ for $\delta = 0.1$ and a mask Ω whose underlying bipartite graph is shown in Fig. 8. The corresponding rank constrained iterations for the level 2 Lasserre relaxation are displayed in Fig. 7.

To illustrate the effect of the sampling, we provide numerical experiments on a 100×100 matrix. For the algorithm to be fully efficient, we combine the trace relations (4.7) (on the fourth order block) and (4.9) (on the third order block), and follow the "minibatch" approach discussed above. On a 100×100 matrix, the factorized gradient method would require storing matrices of size at least $O(K^3) = 8e6$, thus leading to poor performance in terms of runtime. On convex solvers such as CVX or GLOPTIPOLY, this example would require storing matrices of size $O(K^4)$ ($\mathcal{O}(K^3)$ in the reduced formulation of Corollary 1). The iterations on this example are displayed in Fig. 11, and the corresponding bipartite graph used as a mask Ω for the completion problem is shown in Fig. 12. Solving this problem takes no more than 10 mins on a laptop with 2 GHz Intel Core i5.

5 Stable completion of rank-one tensors

Theorems 1 and 2 above both have direct extensions to the tensor completion problem. To illustrate this, we now consider a rank one d-tensor $T \in \mathbb{R}^{n^d}$, $T = \boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \ldots \otimes \boldsymbol{x}_d$, $\boldsymbol{x}_\ell \in \mathbb{R}^{N_\ell}$, and let $K = \sum_{\ell=1}^d N_\ell - (d-1)$. To this tensor T, one can associate a d-uniform hypergraph $\mathcal{H}(\mathcal{V}, \mathcal{E})$ whose set of vertices $\mathcal{V}$ is given by the set of multi-indices $(i_1, i_2, \ldots, i_d)$, and whose set of hyperedges $\mathcal{E}$ is defined from the measurements appearing in $\mathcal{P}_\Omega(T)$. We say that the hypergraph $\mathcal{H}$ has a propagation sequence when there exists a sequence of hyperedges such that any given hyperedge E_ℓ in the sequence shares $d-1$ vertices with the previous edges $\cup_{k \leq \ell-1} E_k$ in that same sequence. This idea is summarized by Definition 2 below.

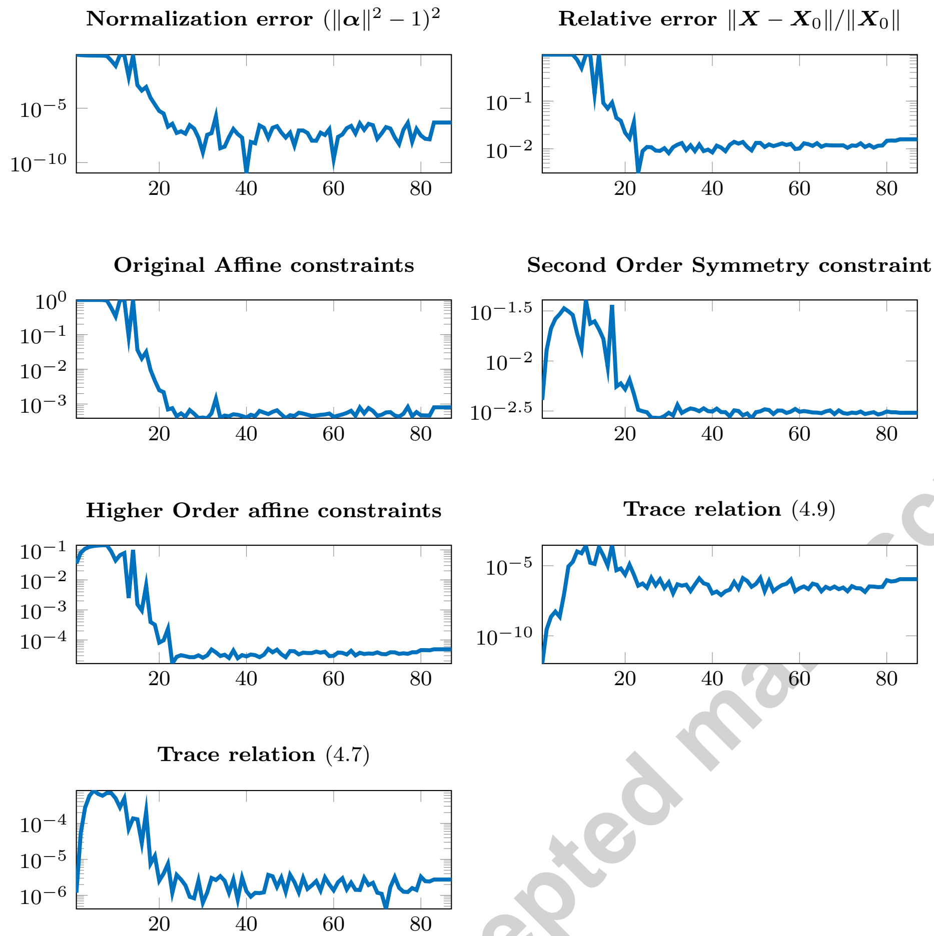

Fig. 7 Evolution of the various constraints and error terms appearing in the augmented Lagrangian, for the hierarchical low rank factorization (here r_1 and r_2 are set to 4) with contractions (4.7) and (4.9) for the 20×20 example of Fig. 6, with measurements defined from the bipartite graph of Fig 8.

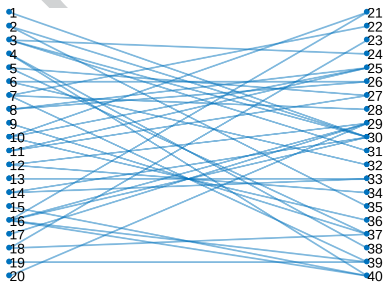

Fig. 8 Bipartite graph corresponding to the 20×20 example of Figs 6 and 7 used to illustrate the failure of nuclear norm minimization and ridge regression.

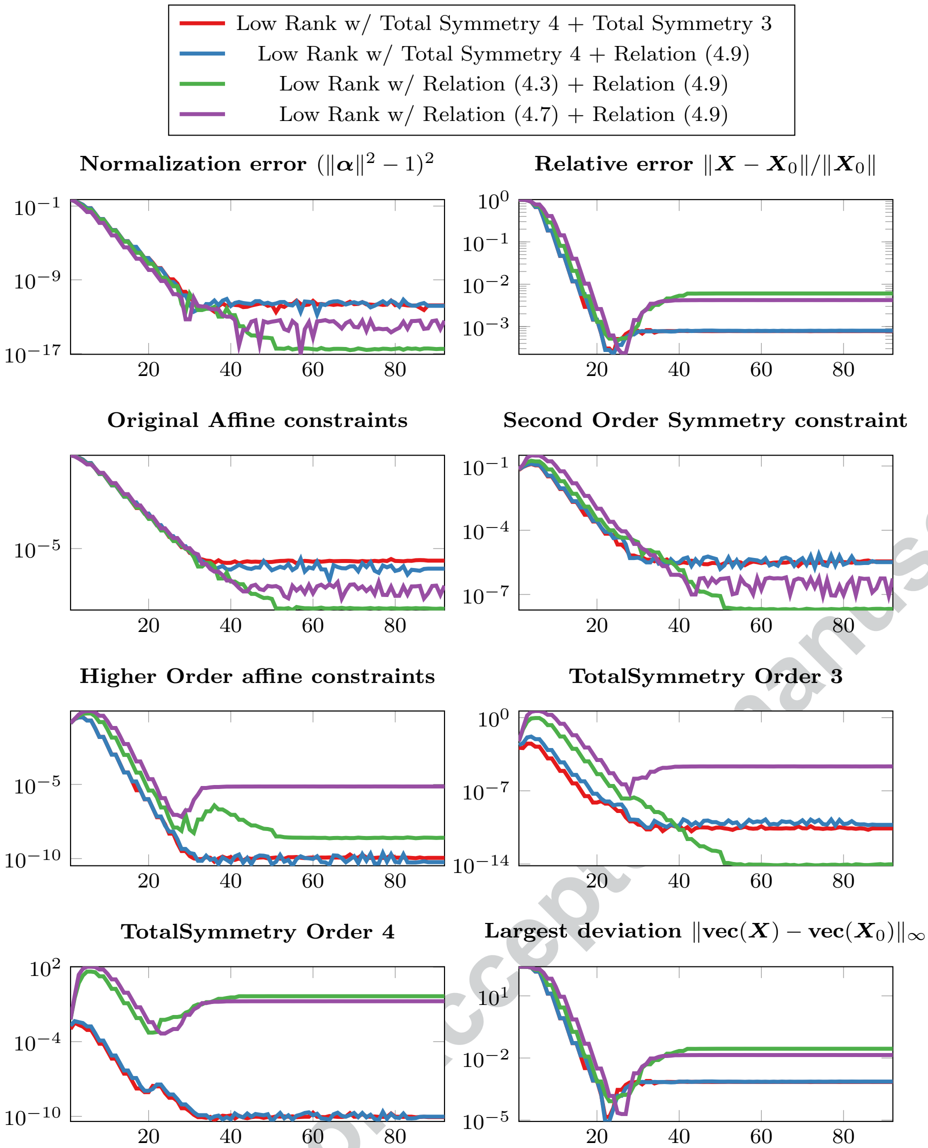

Fig. 9 Comparison of the relations (4.3), (4.7) and (4.9) on the simple low rank decomposition, as substitutes for the total symmetry constraints. As shown by the evolution of the relative error, all 4 approaches recover estimates that are very close to the optimal solution M_0. Adding some of the 4^{th} order Total symmetry might help improve those estimates as indicated by the relative error.

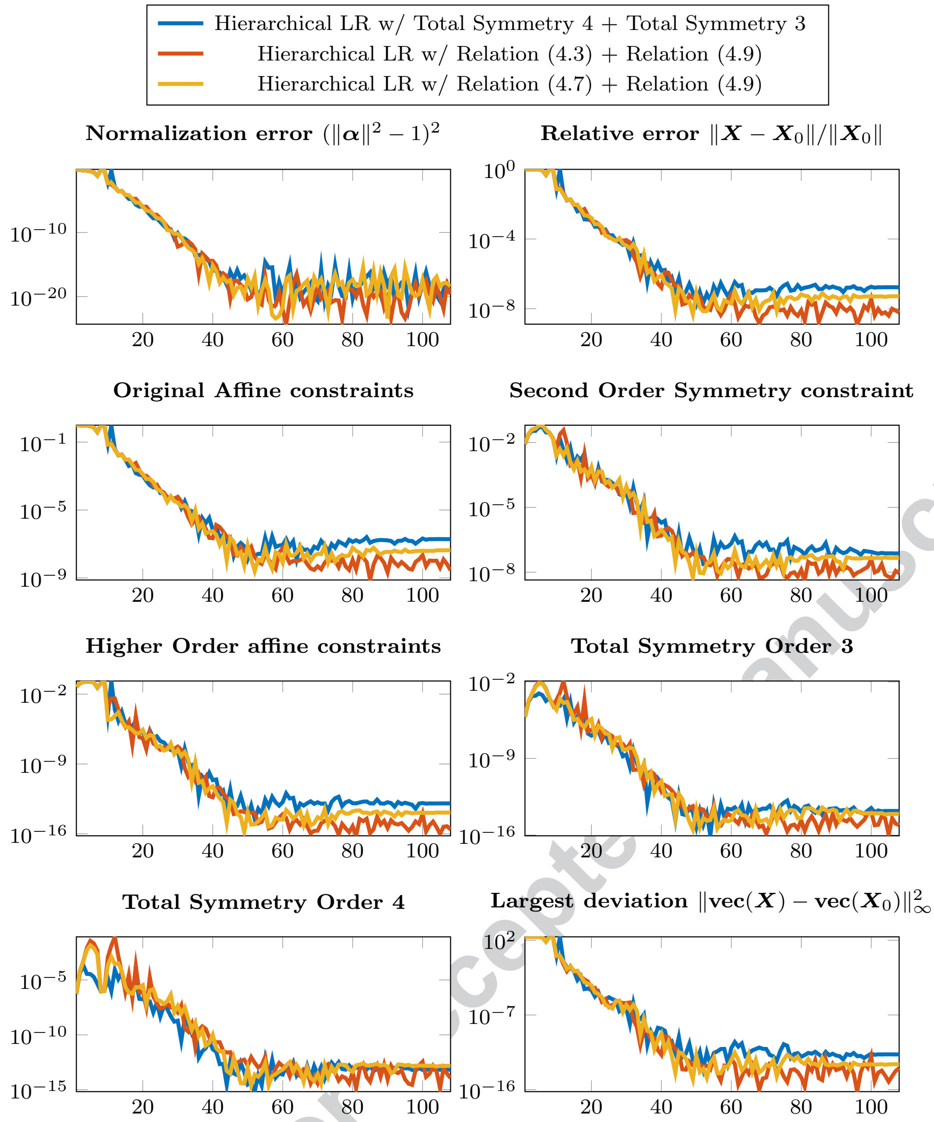

Fig. 10 Minimization of the augmented Lagrangian, on the hierarchical low rank factorization (here we consider ranks 4 and 2) for various combinations of the total symmetry constraints and trace relations. The evolution of the relative error seems to suggest that on the hierarchical factorization, enforcing the trace relations of section 4.4 is sometimes exactly equivalent to requiring total symmetry by means of the combinatorial constraints.

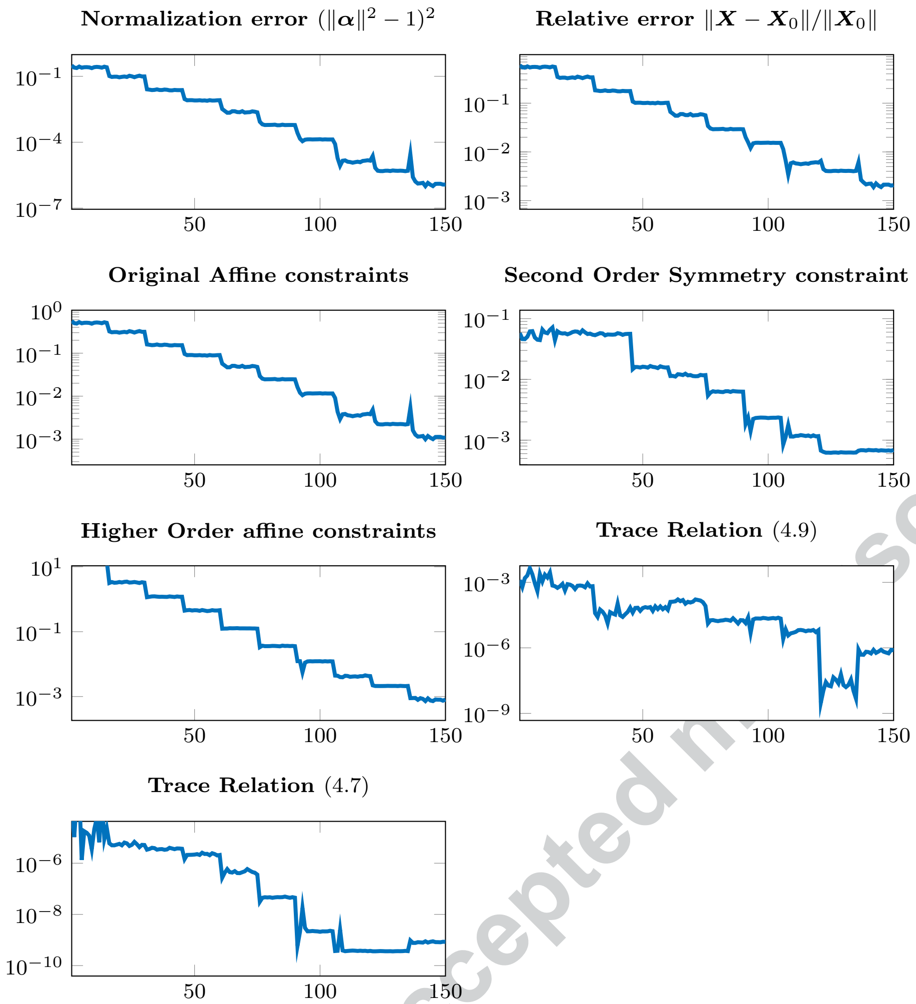

Fig. 11 When considering sufficiently large (e.g. 100×100) completion problems, minimizing the whole set of higher order affine constraints is not efficient anymore because those constraints require storing matrices of size $O(N^3)$ while the hierarchical low rank decomposition only requires storing matrices of size $O(N)$. For this reason, we divide the set of second order monomials into smaller batches $\mathcal{S}_i$ of size $O(m+n)$ sampled at random, and we minimize the resulting augmented Lagrangians sequentially as explained in section 4.5. The jumps in the figures above correspond to the resampling of the monomials. For each resampling, we reset the relative weight of the trace with respect to the misfit term, whence the jump occurring at the transition between two batches. The entries of $\boldsymbol{X}_0$ are controlled as $\delta \leq (X_0)_{ij} \leq 1$ with $\delta = 0.25$ and the ranks of the hierarchical factorization are set to $r_1 = r_2 = 2$. For large matrices, the total symmetry constraints are too expensive, and we have no choice but to rely on relations (4.9) and (4.7).

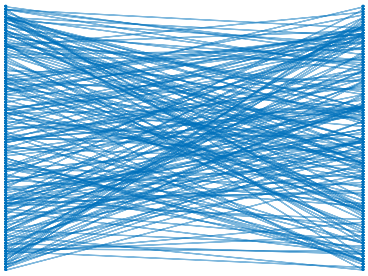

Fig. 12 Bipartite graph corresponding to the 100×100 example of Fig. 11. The number of edges is $O(N) \approx 200$. The graph is generated at random while required to span the whole set of vertices minimally.

Definition 2 (Definition 1 in [4]) Let $\mathcal{H}(\mathcal{V}, \mathcal{E})$ be a d-uniform hypergraph on $N_1 \times N_2 \times \ldots N_d = |\mathcal{V}|$ vertices. A sequence $E_1, \ldots, E_K \in \mathcal{E}$ of hyperedges is called a propagation sequence if, for any $1 \leq \ell \leq K$, $|E_{\ell+1} \cap \bigcup_{i=1}^{\ell} E_\ell| = d - 1$. If the hypergraph has a propagation sequence, then it is called propagation connected.

Provided that there exists a propagation sequence in $\mathcal{H}$ spanning the whole set of vertices, the ideas of Theorem 1 can be extended to the completion of rank one tensors. The general idea is the following. We first let $\boldsymbol{z} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d)$ to denote the concatenation of the generating vectors $\boldsymbol{x}_1$ to $\boldsymbol{x}_d$ as we did for the matrix completion problem. To each hyperedge E_k corresponds a monomial $\prod_{\ell \in E_k} z_\ell$. If we let $E_1, E_2, \ldots, E_M$ to denote the propagation sequence, and use $(e_1^1, e_2^1, \ldots, e_d^1)$ to denote the d-tuple associated to E_1, we can always re-normalize the vectors $\boldsymbol{x}_1$ to $\boldsymbol{x}_d$ as $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i/(x_i)_{e_i^1}$ for $i = 1, \ldots, d-1$ and $\tilde{\boldsymbol{x}}_d = \boldsymbol{x}_d \prod_{i=1}^{d-1} (\boldsymbol{x}_\ell)_{e_\ell^1}$.

From the propagation sequence, we then know that there exists at least one edge in Ω with $d-1$ vertices in common with E_1. In particular, this implies that the monomial associated to E_2 must be of one of the following forms: either we have $\prod_{\ell \in E_2} z_\ell = z_{i_1} z_{i_2}$ where $z_{i_1} \in E_1$, or we have $\prod_{\ell \in E_2} z_\ell = z_{i_2}$ (i.e. a degree one monomial).

From E_1, one can derive the expression for $\tilde{x}_{d,i_1} = z_{i_1}$ (or equivalently the polynomial equation $z_{i_1} - (z_0)_{i_1} = 0$). From E_2, one can derive the expression for z_{i_2} either directly (if the associated monomial is degree 1), or by combining the polynomial equations from E_1 and E_2 as

$$-(z_0)_{i_1}(z_{i_2} - (z_0)_{i_2}) = -(z_{i_2} z_{i_1} - (z_0)_{i_2}(z_0)_{i_1}) + (z_{i_1} - (z_0)_{i_1}) z_{i_2} \tag{5.1}$$

Iterating on this step, one can build a sum of squares polynomial which extends the result of Theorem 1 and 2 to the tensor case. This idea is summarized by the following corollary (a corresponding stability result can be derived).

Corollary 2 *Let $T \in \mathbb{R}^{N_1 \times \ldots \times N_d}$ denote an order d rank-one tensor. Assume that we are given the entries $T_{i_1,\ldots,i_d}$ for $(i_1, i_2, \ldots, i_d) \in \Omega$ and let $\mathcal{H}(\mathcal{V}, \mathcal{E})$ denote the associated hypergraph with vertices defined by the d-tuples of indices from T and edges defined by the entries which appear in Ω. That is $e = (v_1, \ldots, v_d) \in \mathcal{E}$*

iff $(v_1, \ldots, v_d) \in \Omega$. *Assume that there exists at least one propagation sequence* $E_1, E_2, \ldots, E_M$ *that spans the whole set of hypervertices* $\mathcal{V}$. *Then the tensor can be efficiently completed through a level* $d+1$ *semidefinite programming relaxation with minimization of the trace norm of the moments matrix.*

Proof The proof is very similar to the proof of Theorem 1. We therefore only focus on the construction of the dual sum of squares polynomial $\sum_j \boldsymbol{s}_j \boldsymbol{s}_j^T$. By assumption, we know that there exists a connected sequence $E_1, E_2, \ldots, E_M$ corresponding to measurements in Ω. For each E_i, we label the vertices of E_i as $(v_{i1}, v_{i2}, \ldots, v_{id})$. To each such d-tuple of vertices also corresponds a given d-tuple of variables $(z_{i1}, z_{i2}, \ldots, z_{id})$. Given E_1, as explained above, we rewrite the tensor as $T = \tilde{x}_1 \otimes \tilde{x}_2 \otimes \ldots \tilde{x}_d$ where $\tilde{x}_\ell = x_\ell / z_{1\ell}$, for $1 \leq \ell \leq d-1$ and $\tilde{x}_d = x_d \prod_{k=1}^{d-1} z_{1k}$. The entry in the tensor corresponding to the first hyperedge E_1 thus now corresponds to a d-tuple $(1, 1, \ldots, (z_{i_1})_0)$ whose associated polynomial is of degree one and can read as $h_1(z_{i_1}) \equiv z_{i_1} - (z_0)_{i_1} = 0$ as $E_1 \in \Omega$.

Now since the sequence is propagation connected, we can always find an hyperedge E_2 such that E_2 shares $d-1$ vertices with E_1. In particular, this implies that the d-tuple associated with E_2 must have one the following forms: either it has $d-2$ unitary elements and includes z_{i_1}, $(1, \ldots, 1, z_{i_1}, z_{i_2})$ for some z_{i_2}, or it has $d-1$ unitary elements, and hence is of the form $(1, \ldots, 1, z_{i_2})$ for some i_2. Pursuing like this, since the sequence of hyperedges spans the whole set of vertices, we can build a sequence of polynomial equations of the form

$$\begin{aligned} 0 &= z_{i_1} - (z_0)_{i_1} \\ 0 &= z_{i_1} z_{i_2} - (z_0)_{i_1} (z_0)_{i_2} \\ 0 &= z_{i_1} z_{i_2} z_{i_3} - (z_0)_{i_1} (z_0)_{i_2} (z_0)_{i_3} \\ &\vdots \\ 0 &= z_{i_1} z_{i_2} \ldots z_{i_d} - (z_0)_{i_1} (z_0)_{i_2} \ldots (z_0)_{i_d} \end{aligned} \tag{5.2}$$

Note that the propagation sequence might also contain constant degree connections of the form

$$\begin{aligned} 0 &= z_{i_1} z_{i_2} - (z_0)_{i_1} (z_0)_{i_2} \\ 0 &= z_{i_1} z_{i_3} - (z_0)_{i_1} (z_0)_{i_3} \end{aligned} \tag{5.3}$$

Let us label the equations appearing in the chain as $h_1(z) = 0, \ldots, h_M(z) = 0$. As before, the completion problem can be written as the following feasibility problem

$$\text{find } z \text{ s.t. } h_1(z) = 0, h_2(z) = 0, \ldots, h_M(z) = 0. \tag{5.4}$$

We then consider the semidefinite relaxation,

$$\min \ \operatorname{Tr}(\boldsymbol{M}_K^{(d+1)}) \text{ s.t. } \boldsymbol{M}_K^{(d+1)}(m) \succeq 0 \ \text{ and } \ \boldsymbol{M}_K^{2(d+1)-d_\ell}(h_\ell m) = 0 \tag{5.5}$$

To prove that the moments matrix corresponding to the one atomic measure $d\mu = \delta(\mathbf{z} - \mathbf{z}_0) d\mathbf{z}$ is the unique solution to this problem, we need to show that it is possible to build a sum of squares $\sum_j s_j^2(z)$ with associated Gram matrix $\boldsymbol{Q} = \sum_j \boldsymbol{s}_j \boldsymbol{s}_j^T$ satisfying $\operatorname{rank}(\boldsymbol{Q}) = |\mathbb{N}_K^{d+1}| - 1$, from the ideal $\mathcal{I} = \langle h_1(z), \ldots, h_M(z) \rangle$ and the trace. As for the matrix case, we build the SoS $\boldsymbol{s}_j^2$ by finding decompositions for all monomials of degree $\leq d+1$ from low degree combinations of the polynomials appearing in the propagation sequence.

$$\begin{aligned} z_{i_1} - (z_0)_{i_1} &= h_1(z) && (5.6)\\ -(z_0)_{i_1}(z_{i_2} - (z_0)_{i_2}) &= h_1(z) z_{i_2} - h_2(z) && (5.7)\\ -(z_0)_{i_1}(z_0)_{i_2}(z_{i_3} - (z_0)_{i_3}) &= h_2(z) z_{i_3} - h_3(z) && (5.8)\\ &\vdots && (5.9)\end{aligned}$$

More generally, we can express any of the monomials from a degree $2(d+1)$ combination of the constraints appearing in the propagation sequences as follows.

- Either the polynomial $h_{\ell-1}(z)$ has a lower degree than $h_\ell(z)$, i.e.

$$\begin{aligned} h_{\ell-1} &\equiv z_{i_1} z_{i_2} \dots z_{i_{\ell-1}} - (z_0)_{i_1}(z_0)_{i_2} \dots (z_0)_{i_{\ell-1}} = 0\\ h_\ell &\equiv z_{i_1} z_{i_2} \dots z_{i_{\ell-1}} z_{i_\ell} - (z_0)_{i_1}(z_0)_{i_2} \dots (z_0)_{i_{\ell-1}}(z_0)_{i_\ell} = 0 \end{aligned} \tag{5.10}$$

- Or the two consecutive polynomials $h_\ell(z)$ and $h_{\ell-1}(z)$ have the same degree. In this case because the sequence $E_1, \dots, E_M$ is propagation connected, the two equations must share $d-1$ variables. Without loss of generality, we label the variables that appear in $E_{\ell-1} \cap E_\ell$ as $z_{j_1}, \dots, z_{j_{d-1}}$. We then for example have

$$\begin{aligned} h_{\ell-1} &\equiv \prod_{k=1}^{d} z_{j_k} - \prod_{k=1}^{d} (z_0)_{j_k} = 0\\ h_\ell &\equiv z_{i_d} \prod_{k=1}^{d-1} z_{j_k} - (z_0)_{i_d} \prod_{k=1}^{d-1} (z_0)_{j_k} = 0 \end{aligned} \tag{5.11}$$

In the first case (5.10), the monomial z_{i_L} can be written from the two equations as

$$-\left(z_{i_\ell} - (z_0)_{i_\ell}\right) \prod_{k=1}^{\ell-1} (z_0)_{i_k} = h_{\ell-1}(z) z_{i_\ell} - h_\ell(z) \tag{5.12}$$

In the second case,

$$\begin{aligned} (z_{i_d} - (z_0)_{i_d}) \prod_{i \in E_{\ell-1}} (z_0)_i &= z_{E_{\ell-1} \setminus (E_\ell \cap E_{L-1})} \left(\prod_{i \in E_\ell} z_i - \prod_{i \in E_\ell} (z_0)_i \right)\\ &\quad - z_{i_d} \left(\prod_{i \in E_{\ell-1}} z_i - \prod_{i \in E_{\ell-1}} (z_0)_i \right)\\ &\quad + z_{E_{\ell-1} \setminus (E_\ell \cap E_{\ell-1})} \prod_{i \in E_\ell} (z_0)_i \end{aligned} \tag{5.13}$$

Note that the third term on the RHS of (5.13) can be obtained from the ideal by combining the previous polynomials in the chain, as a contribution of degree at most $d+1$. Equations (5.11) and (5.13) together imply that every first order square $(z_{i_k} - (z_0)_{i_k})^2$ can be written as a combination of degree at most $d+1$ from the ideal $\langle h_1, h_2, \dots, h_M\rangle$ and the trace of the moment matrix. I.e., we have $(z_{i_k} - (z_0)_{i_k})^2 = z_{i_k}^2 - (z_0)_{i_k}^2 + \mathcal{I}_{d+1}$ for every $\ell = 1, \dots, M$. For all the other monomials of degree up to $d+1$ we simply write

$$\begin{aligned} \mathbf{z}^{\boldsymbol{\alpha}} - \mathbf{z}_0^{\boldsymbol{\alpha}} &= z_{i_1} z_{i_2} \dots z_{i_K} - (z_0)_{i_1} (z_0)_{i_2} \dots (z_0)_{i_K} \\ &= (z_{i_1} - (z_0)_{i_1})\, z_{i_2} \dots z_{i_K} \\ &+ (z_0)_{i_1} (z_{i_2} - (z_0)_{i_2}) z_{i_3} \dots z_{i_K} \\ &+ \dots \end{aligned} \tag{5.14}$$

Each of the terms appearing in the decomposition (5.14) is of degree at most $(d+1)(d+1)$ and can thus be written from a moments matrix of order $d+1$. The rest of the proof then follows the exact same idea as the proof of Theorem 1.

References

1. A. A. Ahmadi, G. Hall, A. Papachristodoulou, J. Saunderson, and Y. Zheng. Improving efficiency and scalability of sum of squares optimization: Recent advances and limitations. *arXiv preprint arXiv:1710.01358*, 2017.
2. A. S. Bandeira. Convex relaxations for certain inverse problems on graphs. 2015.
3. B. Barak and A. Moitra. Tensor prediction, Rademacher complexity and random 3-XOR. *arXiv preprint arXiv:1501.06521*, 2015.
4. R. Berke and M. Onsjö. Propagation connectivity of random hypergraphs. In *Stochastic Algorithms: Foundations and Applications*, pages 117–126. Springer, 2009.
5. S. Burer and R. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.
6. S. Burer and R. D. Monteiro. Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming*, 103(3):427–444, 2005.
7. E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM review*, 57(2):225–251, 2015.
8. E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.
9. E. J. Candès and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.
10. E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.
11. E. J. Candès, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.
12. E. J. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *Information Theory, IEEE Transactions on*, 56(5):2053–2080, 2010.

13. A. Cosse and L. Demanet. Rank-one matrix completion is solved by the sum-of-squares relaxation of order two. In *Proceedings of the 6th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP'15)*. IEEE, 2015.
14. C.-F. Cui, Y.-H. Dai, and J. Nie. All real eigenvalues of symmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 35(4):1582–1601, 2014.
15. L. Demanet and V. Jugnon. Convex recovery from interferometric measurements. *arXiv preprint arXiv:1307.6864*, 2013.
16. M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, March 2002.
17. M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.
18. J. Gouveia, P. A. Parrilo, and R. R. Thomas. Theta bodies for polynomial ideals. *SIAM Journal on Optimization*, 20(4):2097–2118, 2010.
19. V. Jugnon and L. Demanet. Interferometric inversion: a robust approach to linear inverse problems. In *Proceedings of SEG Annual Meeting, Houston*, pages 5180–5184, 2013.
20. R. Keshavan, A. Montanari, and S. Oh. Matrix completion from noisy entries. In *Advances in Neural Information Processing Systems*, pages 952–960, 2009.
21. R. H. Keshavan, A. Montanari, and S. Oh. Learning low rank matrices from $O(n)$ entries. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 1365–1372. IEEE, 2008.
22. R. H. Keshavan, A. Montanari, and S. Oh. Matrix completion from a few entries. *Information Theory, IEEE Transactions on*, 56(6):2980–2998, 2010.
23. F. Kiraly and L. Theran. Error-minimizing estimates and universal entry-wise error bounds for low-rank matrix completion. In *Advances in Neural Information Processing Systems*, pages 2364–2372, 2013.
24. F. Király and R. Tomioka. A combinatorial algebraic approach for the identifiability of low-rank matrix completion. *arXiv preprint arXiv:1206.6470*, 2012.
25. F. J. Király, L. Theran, and R. Tomioka. The algebraic combinatorial approach for low-rank matrix completion. *Journal of Machine Learning Research*, 16:1391–1436, 2015.
26. J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.
27. J. B. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity. *SIAM Journal on Optimization*, 17(3):822–843, 2006.
28. M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009.
29. Y. Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, pages 405–440. Springer, 2000.
30. J. Nie. An exact jacobian SDP relaxation for polynomial optimization. *Mathematical Programming*, 137(1-2):225–255, 2013.
31. J. Nie and J. Demmel. Sparse SoS relaxations for minimizing functions that are summations of small polynomials. *SIAM Journal on Optimization*, 19(4):1534–1558, 2008.
32. P. A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, Citeseer, 2000.

33. P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming*, 96(2):293–320, 2003.
34. B. Recht. A simpler approach to matrix completion. *The Journal of Machine Learning Research*, 12:3413–3430, 2011.
35. B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.
36. R. T. Rockafellar. *Convex analysis*. Number 28. Princeton university press, 1970.
37. N. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics and Systems Analysis*, 23(6):731–734, 1987.
38. N. Z. Shor. Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences*, 25(6):1–11, 1987.
39. N. Z. Shor. An approach to obtaining global extremums in polynomial mathematical programming problems. *Cybernetics*, 23(5):695–700, 1988.
40. A. Singer and M. Cucuringu. Uniqueness of low-rank matrix completion by rigidity theory. *SIAM Journal on Matrix Analysis and Applications*, 31(4):1621–1641, 2010.
41. G. Tang and P. Shah. Guaranteed tensor decomposition: A moment approach. In *Proceedings of The 32nd International Conference on Machine Learning*, pages 1491–1500, 2015.