Lecture Notes on Sum-of-Squares Optimization

Dmitriy Kunisky

Spring 2022 (last updated April 12, 2022)

Contents

List of Open Problems

1 | Invitation to Sum-of-Squares Proofs for Computer Scientists

As we will see, SOS algorithms and proofs apply to an enormous range of problems—*any* optimization problem that can be expressed in terms of polynomial constraints and a polynomial objective. This generality is a big part of the power and appeal of SOS algorithms, both in theory and in practice. However, starting in full generality can also make the whole enterprise seem quite daunting and not as concrete as you, presumably interested in actual algorithms to solve specific problems, might have hoped. For that reason, before moving on to that general theory, we will start by seeing how some first hints of SOS show up in a natural interpretation of an algorithm you may have seen before.

1.1 Maximum Cut and Approximation Ratios

That algorithm is a relaxation of the problem of finding the largest *cut* in a graph.

Definition 1.1 (MaxCut)**.** *A* cut *of a graph $G = (V, E)$ is a subset of the edges, $C \subseteq V$. The* size *of the cut is the number of edges $\{v, w\} \in E$ with $v \in C$ and $w \notin C$. We denote by* $\mathsf{MaxCut}(G)$ *the size of the largest cut in G.*

You can check that this problem admits the following linear-algebraic encoding. We introduce the *graph Laplacian* of $G = (V, E)$, $\boldsymbol{L} = \boldsymbol{L}_G = \frac{1}{4}(\boldsymbol{D}_G - \boldsymbol{A}_G)$, where $\boldsymbol{D}_G \in \mathbb{R}^{V \times V}$ is a diagonal matrix with diagonal entries equal to the vertex degrees in G and $\boldsymbol{A}_G$ is the adjacency matrix.[1] (The factor of $\frac{1}{4}$ is not usually included but will be convenient for our purposes.) Then, we have

$$\mathsf{MaxCut}(G) = \max_{\boldsymbol{x} \in \{\pm 1\}^V} \boldsymbol{x}^\top \boldsymbol{L} \boldsymbol{x}, \tag{1.1}$$

because, as you can also check, $\boldsymbol{x}^\top \boldsymbol{L} \boldsymbol{x}$ is the size of the cut where the partition of vertices is specificed by the signs in $\boldsymbol{x}$ (see Exercise 1.1).

MaxCut has occasional practical applications (see the chapter notes), though not as many as we sometimes pretend. Our honest reasons for studying it are mostly theoretical: MaxCut is an NP-complete problem (one of Karp's original list in [Kar72]) that is especially simple to formulate and, unlike other problems like satisfiability, has the linear-algebraic flavor we saw in (1.1). It is also maybe the simplest hard case of optimization over Boolean variables:

[1] When we write $\mathbb{R}^{V \times V}$, $\{\pm 1\}^V$, or other similar notations, we assume the vertices of G have some implicit ordering to identify elements of these sets with concrete matrices or vectors.

maximizing a linear objective function over $\boldsymbol{x} \in \{\pm 1\}^n$ is easy, but MaxCut, a special case (since we assume a particular structure of the matrix $\boldsymbol{L}$) of maximizing a quadratic objective function, is already hard.

Since it is hard to compute $\mathsf{MaxCut}(G)$ or to find a maximizer $\boldsymbol{x}^\star$, we can downgrade our algorithmic hopes to finding *approximate* solutions. One can ask to approximate either the quantity $\mathsf{MaxCut}(G)$, or the maximizer $\boldsymbol{x}^\star$. Both problems are interesting, but the latter has been studied more. So, given a function $\hat{\boldsymbol{x}}(G) \in \{\pm 1\}^V$, we say $\hat{\boldsymbol{x}}$ is an *α-approximation* of MaxCut if, for *all* graphs G,

$$\hat{\boldsymbol{x}}(G)^\top \boldsymbol{L} \hat{\boldsymbol{x}}(G) \geq \alpha \cdot \mathsf{MaxCut}(G). \tag{1.2}$$

If $\hat{\boldsymbol{x}}(G)$ is random, then we say the same if (1.2) holds with an expectation taken on the left-hand side.[2]

The following is a seemingly naive benchmark. We sketch the simple proof; see Exercise 1.1 for a guide to the details.

Proposition 1.2. *There is a randomized polynomial-time $\frac{1}{2}$-approximation of* MaxCut.

Proof. Choose $\hat{x}_v \sim \mathsf{Unif}(\{\pm 1\})$ independently at random. Then, by linearity of expectation $\mathbb{E}[\hat{\boldsymbol{x}}(G)^\top \boldsymbol{L} \hat{\boldsymbol{x}}(G)] = \frac{1}{2}|E| \geq \frac{1}{2}\mathsf{MaxCut}(G)$. □

This does not seem like a good idea: we are just choosing a uniformly random cut among all partitions of the vertices of the graph, without using any information about G at all! A bit less trivially, by using some information about G, we can remove the randomness from this approximation (see Exercise 1.2) in a variation dating back to a 1967 result of Erdős [Erd67]. That is an improvement, but it is still tempting look for an α-approximation with $\alpha > \frac{1}{2}$.

There were many attempts to find such algorithms. One branch of work in this direction focused on combinatorial algorithms (similar to Erdős' "greedy" algorithm detailed in Exercise 1.2), including [Vit81, PT82, HV91, HL96]. These achieved approximations with ratios of the form $\frac{1}{2} + \epsilon(G)$, but with the term $\epsilon(G) \to 0$ as G grows in some suitable sense—in the number of vertices, the maximum degree, or some similar quantity.

Another promising direction considered *linear programming (LP)* relaxations of MaxCut. The most common such relaxation is the *metric relaxation*, which is derived as follows. We first expand the linear-algebraic formulation of MaxCut:

$$\mathsf{MaxCut}(G) = \max_{\boldsymbol{x} \in \{\pm 1\}^V} \boldsymbol{x}^\top \boldsymbol{L} \boldsymbol{x} = \max_{\boldsymbol{x} \in \{\pm 1\}^V} \sum_{i,j=1}^{n} L_{ij} x_i x_j. \tag{1.3}$$

We then introduce a matrix variable $\boldsymbol{X} \in \mathbb{R}^{n \times n}$, which we think of as $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top$, i.e., having $X_{ij} = x_i x_j$. But, we do not impose the constraint that $\boldsymbol{X}$ have rank one, but rather only some collection of *linear* constraints that must hold for all $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top$. It turns out that there are very many independent such constraints (exponentially many in n), and no small subset of those suffices to describe the polytope of $\boldsymbol{X}$ that is the convex hull of the $\boldsymbol{x}\boldsymbol{x}^\top$. That polytope is called the *cut polytope*, and if it were simple to describe, then we could solve

[2] As we will mention later, both of the random algorithms we consider can be *derandomized* to achieve the same approximation ratio deterministically.

MaxCut efficiently with linear programming! See, e.g., [DL09] for lots of information about the cut polytope and its geometry.

Instead, we choose only a small tractable subset of these constraints. For the metric relaxation, we optimize:

$$\mathsf{LP}(G) := \left\{ \begin{array}{ll} \text{maximize} & \langle \boldsymbol{L}, \boldsymbol{X} \rangle = \sum_{i,j} L_{ij} X_{ij} \\ \text{subject to} & -1 \le X_{ij} \le 1 \text{ for all } i,j \in V, \\ & X_{ii} = 1 \text{ for all } i \in V, \\ & X_{ij} + X_{jk} + X_{ik} \ge -1 \text{ for all } i,j,k \in V, \\ & X_{ij} + X_{jk} - X_{ik} \le 1 \text{ for all } i,j,k \in V \end{array} \right\}. \tag{1.4}$$

It is less obvious, but you can convince yourself that the latter two collections of inequalities, called *triangle inequalities*,[3] must hold whenever $X_{ij} = x_i x_j$ for $\boldsymbol{x} \in \{\pm 1\}^n$.

Analyzing this, [BM86] showed that $\mathsf{LP}(G)$ is *tight*—solving MaxCut exactly!—for all graphs that do not contain the complete graph on five vertices as a minor (including, for instance, all planar graphs). However, outside this case, it remained unclear how to reason. Works such as [PT94] studied how well the value of this LP relaxation approximated the value $\mathsf{MaxCut}(G)$, and showed that it achieved a non-trivial approximation for dense random graphs. But this was unsatisfactory on two counts: first, it did not consider a *worst-case* approximation ratio (indeed, later work would show that the worst-case approximation ratio of small LPs like this is again $\frac{1}{2}$); and second, it only produced a bound on $\mathsf{MaxCut}(G)$, not a good cut $\hat{\boldsymbol{x}}(G)$.

These were the best results and algorithms known until roughly thirty years after the greedy algorithm of [Erd67]. We next describe the breakthrough that improved dramatically on this state of affairs.

1.2 The Goemans-Williamson Relaxation

It turns out that the key to an improved approximation is to use *semidefinite programming (SDP)* instead of LP. Goemans and Williamson in [GW95] proposed starting by solving the following SDP:

$$\mathsf{SDP}(G) := \left\{ \begin{array}{ll} \text{maximize} & \langle \boldsymbol{L}, \boldsymbol{X} \rangle = \sum_{i,j} L_{ij} X_{ij} \\ \text{subject to} & \boldsymbol{X} \succeq \boldsymbol{0}, \\ & X_{ii} = 1 \text{ for all } i \in V \end{array} \right\}. \tag{1.5}$$

Let us see a few ways of understanding what this program does and how one could derive it from first principles. Since we are now closer to topics relevant to SOS, we will give more precise names to these interpretations.

[3] There are two interpretations of the name. First, the index pairs involved form a triangle, and indeed there are analogous larger *cycle inequalities* involving cycles of indices of greater length. Second, another viewpoint on this relaxation is that $\rho(i,j) = \frac{1 - x_i x_j}{2}$ forms a *discrete metric* associated to the cut defined by $\boldsymbol{x}$, where $\rho(i,j) = 1$ if i and j are on opposite sides of the cut and $\rho(i,j) = 0$ otherwise. This, being a metric, must satisfy the triangle inequality $\rho(i,k) \le \rho(i,j) + \rho(j,k)$, which you can check implies the second set of triangle inequalities. When this metric interpretation is especially valuable, some work considers instead the weaker LP relaxation not including the first set of inequalities.

ALGEBRAIC INTERPRETATION The first interpretation is of the same kind as we gave for $\mathsf{LP}(G)$, a straightforward way of obtaining a relaxation by discarding a rank constraint from a linearization. As before, for any $\boldsymbol{x} \in \{\pm 1\}^V$, $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top$ is feasible for $\mathsf{SDP}(G)$, and $\langle \boldsymbol{L}, \boldsymbol{x}\boldsymbol{x}^\top \rangle = \boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}$, the size of the cut achieved by $\boldsymbol{x}$. Conversely, if $\boldsymbol{X}$ is feasible for $\mathsf{SDP}(G)$ and also has rank one, then $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top$ for some $\boldsymbol{x} \in \{\pm 1\}^V$. Thus $\mathsf{SDP}(G)$ is, suitably interpreted, the optimization of (1.1) adjusted to make the feasible set bigger (also making the optimization tractable to carry out). In particular,

$$\mathsf{MaxCut}(G) \leq \mathsf{SDP}(G). \tag{1.6}$$

To come up with $\mathsf{SDP}(G)$ by this reasoning, you would start by rewriting

$$\mathsf{MaxCut}(G) = \max_{\boldsymbol{x} \in \{\pm 1\}^V} \boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \left\{ \begin{array}{ll} \text{maximize} & \langle \boldsymbol{L}, \boldsymbol{X} \rangle \\ \text{subject to} & \boldsymbol{X} \succeq \boldsymbol{0}, \\ & X_{ii} = 1 \text{ for all } i \in V, \\ & \mathsf{rank}(\boldsymbol{X}) = 1 \end{array} \right\}, \tag{1.7}$$

and then discard the rank constraint to obtain an SDP. This is the same "keep tractable constraints" operation as for $\mathsf{LP}(G)$, only we have expanded our idea of what counts as "tractable" to include the semidefinite constraint $\boldsymbol{X} \succeq \boldsymbol{0}$.[4]

GEOMETRIC INTERPRETATION Another way to understand $\mathsf{SDP}(G)$ is more geometric: while the entries of $\boldsymbol{x}\boldsymbol{x}^\top$ for $\boldsymbol{x} \in \{\pm 1\}^V$ are $(\boldsymbol{x}\boldsymbol{x}^\top)_{ij} = x_i x_j$, for $\boldsymbol{X}$ feasible for $\mathsf{SDP}(G)$, there exist some *vectors* $\boldsymbol{v}_i$ for $i \in V$ that have unit length and with $X_{ij} = \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$. (See Proposition A.4 for this equivalent characterization of psd matrices.) Conversely, any such *Gram matrix* of unit vectors is feasible for $\mathsf{SDP}(G)$; we may also without loss of generality suppose $\boldsymbol{v}_i \in \mathbb{R}^V$ to specify a dimension. So, $\mathsf{SDP}(G)$ can be seen as computing the largest *vector cut* of G, an analog of a cut where each vertex is assigned not just a binary membership on one side of the cut or the other, but a continuous membership vector:

$$\mathsf{SDP}(G) = \left\{ \begin{array}{ll} \text{maximize} & \sum_{i,j} L_{ij} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \\ \text{subject to} & \boldsymbol{v}_i \in \mathbb{R}^V, \\ & \|\boldsymbol{v}_i\| = 1 \text{ for all } i \in V \end{array} \right\}. \tag{1.8}$$

PROBABILISTIC INTERPRETATION Finally, and what will be most valuable for us later, you might think probabilistically. Let us write $\mathcal{M}(\{\pm 1\}^V)$ for the set of probability measures over $\{\pm 1\}^V$. Then, we may "convexify" MaxCut as

$$\mathsf{MaxCut}(G) = \max_{\boldsymbol{x} \in \{\pm 1\}^V} \boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \max_{\mu \in \mathcal{M}(\{\pm 1\}^V)} \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mu} [\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}] = \max_{\mu \in \mathcal{M}(\{\pm 1\}^V)} \left\langle \boldsymbol{L}, \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mu} [\boldsymbol{x}\boldsymbol{x}^\top] \right\rangle. \tag{1.9}$$

In the first step we use that the maximum will be achieved by μ a point mass at an optimal $\boldsymbol{x}$, and in the second we use the linearity of expectation.[5] The set of matrices in the last

[4] If you are wondering why we have eliminated the triangle inequalities and whether adding them back in might be helpful—excellent question! We are just following [GW95] for now; see below in Section 1.4 for discussion of precisely this enhancement.

[5] To be clear about notation, the expectation of a random matrix is the matrix of expectations of its entries: $(\mathbb{E}[\boldsymbol{M}])_{ij} = \mathbb{E}[M_{ij}]$.

optimization forms the *cut polytope* mentioned above. Said probabilistically, we are optimizing over the set of matrices of degree 2 moments of distributions over the hypercube; since $\boldsymbol{x}\boldsymbol{x}^\top$ is unchanged by negating $\boldsymbol{x}$, we may also restrict our attention to μ that are *centered* with $\mathbb{E}_{\boldsymbol{x}\sim\mu}[\boldsymbol{x}] = \boldsymbol{0}$, in which case these are equivalently covariance matrices. We may then arrive at $\mathsf{SDP}(G)$ by imposing some computationally-tractable collection of constraints we can come up with on such a matrix. Indeed, if $\boldsymbol{X} = \mathbb{E}_{\boldsymbol{x}\sim\mu}[\boldsymbol{x}\boldsymbol{x}^\top]$ for some μ, then $X_{ii} = \mathbb{E}[x_i^2] = \mathbb{E}[1] = 1$ since $x_i^2 = 1$ for any $\boldsymbol{x} \in \{\pm 1\}^V$. And, $\boldsymbol{X} \succeq \boldsymbol{0}$, as any moment matrix must be; to see this explicitly note that $\boldsymbol{a}^\top \boldsymbol{X} \boldsymbol{a} = \mathbb{E}_{\boldsymbol{x}\sim\mu}[\langle \boldsymbol{a}, \boldsymbol{x}\rangle^2] \geq 0$. In this last step we perform basically the same deductions as in the algebraic interpretation; however, we will see that the extra probabilistic language will come in handy when we try to generalize this reasoning later.

Let us now return to the actual result of [GW95], which shows how $\mathsf{SDP}(G)$ helps us to approximate $\mathsf{MaxCut}(G)$. The key idea is that any feasible point of $\mathsf{GW}(G)$ may be *rounded* to a feasible point of $\mathsf{MaxCut}(G)$ (recall that when we discussed $\mathsf{LP}(G)$ we did not have this idea in hand; we only obtained a $\hat{\boldsymbol{x}}(G)$ when $\mathsf{LP}(G)$ solved $\mathsf{MaxCut}(G)$ exactly).

Theorem 1.3 (Goemans-Williamson rounding)**.** *For any $\boldsymbol{X}$ feasible for $\mathsf{SDP}(G)$, there exists a random $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}(\boldsymbol{X})$ computable from $\boldsymbol{X}$ in polynomial time such that*

$$\mathbb{E}[\hat{\boldsymbol{x}}^\top \boldsymbol{L} \hat{\boldsymbol{x}}] \geq \alpha^{\mathsf{GW}} \cdot \langle \boldsymbol{L}, \boldsymbol{X}\rangle, \tag{1.10}$$

where

$$\alpha^{\mathsf{GW}} := \frac{2}{\pi} \min_{\rho\in[-1,1]} \frac{\arccos\rho}{1-\rho} = 0.87856^+. \tag{1.11}$$

If you have not seen the proof, Exercise 1.3 guides you through the analysis of the randomized rounding. The idea of the construction works with the geometric interpretation of $\mathsf{SDP}(G)$: we view $\boldsymbol{X}$ as corresponding to a vector cut $\{\boldsymbol{v}_i\}_{i\in V}$, and from this we construct a genuine cut by partitioning the $\boldsymbol{v}_i$ in space by a random hyperplane. For this reason, the Goemans-Williamson rounding is sometimes called *hyperplane rounding*.

There are two important corollaries. The first is the approximation result we have been building up to.

Corollary 1.4 (Goemans-Williamson approximation)**.** *For any $\epsilon > 0$, there is a randomized polynomial-time $(\alpha^{\mathsf{GW}} - \epsilon)$-approximation of* MaxCut*.*

Proof. Using standard SDP solvers, in polynomial time we may compute $\boldsymbol{X}$ with $\langle \boldsymbol{L}, \boldsymbol{X}\rangle \geq \mathsf{SDP}(G) - \epsilon$. For $\hat{\boldsymbol{x}}$ computed from this $\boldsymbol{X}$, we then have

$$\mathbb{E}[\hat{\boldsymbol{x}}^\top \boldsymbol{L} \hat{\boldsymbol{x}}] \geq \alpha^{\mathsf{GW}}(\mathsf{SDP}(G) - \epsilon) \geq (\alpha^{\mathsf{GW}} - \epsilon)\mathsf{MaxCut}(G), \tag{1.12}$$

and the result follows. □

As with the $\frac{1}{2}$-approximation from Proposition 1.2, the randomness may be removed from the Goemans-Williamson rounding through an iterative procedure called the *method of conditional expectations* [MR95].

The second corollary is that we may efficiently compute, in $\mathsf{SDP}(G)$, a fairly tight upper *bound* on $\mathsf{MaxCut}(G)$. This is perhaps less practically useful without a rounding procedure,

but measuring the tightness of such bounds will be a valuable way of directly measuring the performance of more complicated relaxation algorithms (if, say, we do not know how to round them or how to analyze a rounding procedure).

Corollary 1.5 (Goemans-Williamson certificate)**.** *For all graphs G, we have*

$$\mathsf{SDP}(G) \geq \mathsf{MaxCut}(G), \tag{1.13}$$

$$\mathsf{SDP}(G) \leq \frac{1}{\alpha^{\mathsf{GW}}} \cdot \mathsf{MaxCut}(G), \tag{1.14}$$

where $1/\alpha^{\mathsf{GW}} = 1.139^{-}$.

Let us introduce some more of the language we will use to talk about such algorithms: we say that SDP *certifies* a bound on MaxCut, because (1.13) holds for *all* input graphs G—we may take the output of $\mathsf{SDP}(G)$ as a *certificate* of such an upper bound. In principle certification algorithms could be arbitrary so long as they obey this property, but in practice all the ones we know are based on convex relaxation in the fashion of $\mathsf{LP}(G)$ and $\mathsf{SDP}(G)$. For these algorithms, we call the constant in (1.14), controlling the relative error in the upper bound, the *integrality gap* (because those $\boldsymbol{X}$ in $\mathsf{SDP}(G)$ with $\mathsf{rank}(\boldsymbol{X}) = 1$ are often called the *integral solutions*).

1.3 Duality and Sum-of-Squares Proofs

The above sections have been a standard presentation of the Goemans-Williamson approximation algorithm. Now, we move into the more idiosyncratic perspective that will lead us to SOS algorithms. The key step is to take the *dual* of this SDP,[6] which, in this case, is always equal to the original one and thus just gives a different way of expressing $\mathsf{SDP}(G)$:

$$\mathsf{SDP}(G) = \left\{ \begin{array}{ll} \text{minimize} & \mathsf{Tr}(\boldsymbol{D}) \\ \text{subject to} & \boldsymbol{D} \text{ diagonal}, \\ & \boldsymbol{D} \succeq \boldsymbol{L} \end{array} \right\}. \tag{1.15}$$

Let us expand the definition of the relation $\boldsymbol{D} \succeq \boldsymbol{L}$ in a particular way. This relation means $\boldsymbol{D} = \boldsymbol{L} + \boldsymbol{A}$ for some $\boldsymbol{A} \succeq \boldsymbol{0}$, which in turn is equivalent (see Proposition A.4) to there existing some $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$ with

$$\boldsymbol{D} = \boldsymbol{L} + \sum_{a=1}^{N} \boldsymbol{v}_a \boldsymbol{v}_a^{\top}. \tag{1.16}$$

Now, let us translate this linear-algebraic equation into a polynomial equation. In general, any symmetric matrix $\boldsymbol{H} \in \mathbb{R}^{n \times n}_{\mathsf{sym}}$ is uniquely determined by the associated quadratic form, $\boldsymbol{y}^{\top} \boldsymbol{H} \boldsymbol{y} = \sum_{i,j=1}^{n} H_{ij} y_i y_j$, viewed as a polynomial in the entries of $\boldsymbol{y}$—from the quadratic form, we can read off each entry H_{ij} as the coefficient of y_{ij}, divided by two if $i \neq j$. Thus, introducing variables $\boldsymbol{y} = (y_i)_{i \in V}$, (1.16) is equivalent to the equation of polynomials

$$\sum_{i \in V} D_{ii} y_i^2 = \sum_{i,j \in V} L_{ij} y_i y_j + \sum_{a=1}^{N} \left(\sum_{i \in V} (\boldsymbol{v}_a)_i y_i \right)^2. \tag{1.17}$$

[6] See Appendix B.1 for a general discussion of how to carry out this procedure.

Massaging this expression a little bit more, we call $v_{a,i} := (\boldsymbol{v}_a)_i$, $d_i := D_{ii}$, $c := \sum_{i \in V} d_i$ and we replace $d_i y_i^2 = d_i + d_i(y_i^2 - 1)$ and rearrange, obtaining

$$c = \sum_{i,j \in V} L_{ij} y_i y_j + \underbrace{\sum_{i \in V} d_i (1 - y_i^2)}_{\textcircled{\text{Con}}} + \underbrace{\sum_{a=1}^{N} \left(\sum_{i \in V} v_{a,i} y_i \right)^2}_{\textcircled{\text{SOS}}}. \tag{1.18}$$

Note here that for such an equation to hold we *must* have $c = \sum_{i \in V} d_i$ by equating the constant terms of either side.

Such a polynomial equation is what we will call a *sum-of-squares (SOS) proof*. We claim that this gives a simple proof that, whenever $\boldsymbol{y} \in \{\pm 1\}^V$, we must have $\sum_{i,j \in V} L_{ij} y_i y_j \leq c$. In other words, such an equation proves that $\mathsf{MaxCut}(G) \leq c$. That is because, first, $\boldsymbol{y} \in \{\pm 1\}^V$ is equivalent to $1 - y_i^2 = 0$ for all $i \in V$, and second, any real number squared is non-negative. Therefore, whenever we evaluate above with $\boldsymbol{y} \in \{\pm 1\}^V$ we will have

$$\textcircled{\text{Con}} = 0, \tag{1.19}$$

$$\textcircled{\text{SOS}} \geq 0, \tag{1.20}$$

and our claim follows. (The names of the two terms stand for "constraints" and "sum-of-squares," respectively.) Summarizing this reasoning, we reach the following fourth interpretation of $\mathsf{SDP}(G)$, a wildly different one from the three we saw above.

PROOF SYSTEM INTERPRETATION $\mathsf{SDP}(G)$ may be seen as optimizing upper bounds on $\mathsf{MaxCut}(G)$ that are obtained within the *SOS proof system* where polynomial reasoning of the above kind is available. Formally, we have

$$\mathsf{SDP}(G) = \left\{ \begin{array}{ll} \text{minimize} & c \\ \text{subject to} & c = \sum_{i,j} L_{ij} y_i y_j + \sum_i d_i (1 - y_i^2) + \sum_a (\sum_i v_{a,i} y_i)^2 \\ & \text{for some } d_i, v_{a,j} \in \mathbb{R} \end{array} \right\}, \tag{1.21}$$

where we emphasize that the y_i are symbolic indeterminate variables, and the constraint equation must be viewed as an equality of polynomials, coefficient by coefficient. In fact, we will see that this way of viewing $\mathsf{SDP}(G)$ is one of the most fruitful to try to generalize, and most of this course will be dedicated to understanding the vastly more general family of algorithms that resembles the proof system interpretation of $\mathsf{SDP}(G)$.

1.4 TOWARDS IMPROVEMENTS

At the time of writing, we are coming up on the thirtieth anniversary of [GW95]—the same amount of time it took for an improvement on the trivial $\frac{1}{2}$-approximation algorithm for MaxCut to materialize. Is there reason to hope to see further improvements on Goemans and Williamson's $\alpha^{\mathsf{GW}} \approx 0.878^+$-approximation?

On the one hand, a beautiful line of work has connected the possibility of improving on the α^{GW} approximation ratio to Khot's *Unique Games Conjecture (UGC)*, first introduced in [Kho02]. We will not get into the details here, but this conjecture states that it is NP-hard solve the following decision problem: for a particular form of constraint satisfaction problem (CSP), for any $\epsilon > 0$, distinguish cases where at most an ϵ fraction of the constraints are satisfiable from ones where at least a $1 - \epsilon$ fraction of the constraints are satisfiable. The UGC, if true, implies limitations on approximating many other CSPs. One of its most striking consequences is that, conditional on UGC, it is NP-hard to approximate MaxCut with any $\alpha^{\mathsf{GW}} + \epsilon$ approximation ratio [KKMO07]. Similar results on optimality of semidefinite programming would also follow for other CSPs [Rag08].[7]

On the other hand, it is tempting to try to improve various aspects of the Goemans-Williamson algorithm. To begin, could the analysis be improved without changing the underlying algorithm? The results of [FS02] give a strong negative answer. First, they show that there is a sequence of graphs G_n so that $\mathsf{MaxCut}(G_n)/\mathsf{SDP}(G_n) \to \alpha^{\mathsf{GW}}$, i.e., the Goemans-Williamson analysis of the tightness of the bound $\mathsf{SDP}(G)$ on $\mathsf{MaxCut}(G)$ (as in Corollary 1.5) is optimal. And second, they show that there is a sequence of graphs G_n and a sequence of $\boldsymbol{X}_n$ optimal for $\mathsf{SDP}(G_n)$ so that, if C_n is the cut value achieved by the *best* hyperplane rounding of $\boldsymbol{X}_n$ (not just a random one), then $C_n/\mathsf{MaxCut}(G_n) \to \alpha^{\mathsf{GW}}$ (similar but weaker results were obtained earlier by [Kar99]).

Next, could adding some straightforward families of constraints to $\mathsf{SDP}(G)$ improve the performance? In particular, we might consider adding the triangle inequalities that we included in the metric relaxation $\mathsf{LP}(G)$:

$$\mathsf{SDP}^{\triangle}(G) := \left\{ \begin{array}{ll} \text{maximize} & \langle \boldsymbol{L}, \boldsymbol{X} \rangle \\ \text{subject to} & \boldsymbol{X} \succeq \boldsymbol{0}, \\ & X_{ii} = 1 \text{ for all } i \in V, \\ & X_{ij} + X_{jk} + X_{ik} \geq -1 \text{ for all } i,j,k \in V, \\ & X_{ij} + X_{jk} - X_{ik} \leq 1 \text{ for all } i,j,k \in V \end{array} \right\}. \tag{1.22}$$

You will show in Exercise 1.6 that the triangle inequalities are *not* automatically satisfied by any $\boldsymbol{X}$ feasible for $\mathsf{SDP}(G)$ and thus that $\mathsf{SDP}^{\triangle}(G) < \mathsf{SDP}(G)$ for some G. However, [KV05] (the much older conference version of [KV15]; some results with a weaker approximation ratio were also shown by [FS02]) showed that, again, there is a sequence of graphs G_n so that $\mathsf{MaxCut}(G_n)/\mathsf{SDP}^{\triangle}(G_n) \to \alpha^{\mathsf{GW}}$.

But the most tantalizing option of all is to pursue generalizations of our proof system interpretation of $\mathsf{SDP}(G)$. Indeed, we will see in later chapters that, again using semidefinite programming, we can optimize over a broader type of polynomial proof of a bound on $\mathsf{MaxCut}(G)$, of the form

$$c = \sum_{i,j} L_{ij} y_i y_j + \sum_i (1 - y_i^2) \cdot p_i(y_1, \ldots, y_n) + \sum_a q_a(y_1, \ldots, y_n)^2, \tag{1.23}$$

and that we can do this optimization in polynomial time in n so long as we fix some constant polynomial degree D and enforce that that is the greatest degree of any polynomial

[7] One especially simple problem for which hardness of approximation would follow from UGC is the vertex cover problem [KR08].

appearing in the above equation: $\deg(1-y_i^2)p_i, \deg q_a^2 \le D$. Modulo some modest technicalities that you will address in Exercise 1.7, above we have discussed the simplest case $D = 2$. For any larger D, essentially nothing more is known.

Open Problem 1.1 (Higher-degree SOS for MaxCut)**.** *Does optimizing over SOS proofs of any fixed degree D improve upon the worst-case SDP integrality gap $1/\alpha^{GW}$ for* MaxCut *(thereby disproving UGC) or not? Even for $D = 4$ the answer remains unknown!*

While MaxCut is a convenient concrete problem to discuss, the state of affairs is the same for other CSPs where the UGC implies hardness of approximation beyond a ratio achieved by a simple SDP.

Conversely, much of the perception that SOS is such a powerful family of algorithms stems from work showing that higher-degree SOS (still of polynomial runtime) successfully distinguishes unique games instances that are not distinguished by other, weaker convex programming hierarchies until super-polynomial runtime [BBH+12, OZ13, KOTZ14]. Thus, finding integrality gap instances for higher-degree SOS matching the degree 2 gaps (a negative answer to Open Problem 1.1) should be seen as strong evidence in favor of UGC.

1.5 What Is This Course About?

We have seen in this quick overview two of the key motivations and themes that will guide us through the rest of the course. To give you an idea of what is to come, let us pick out these main points.

First, we will focus on how SOS is an algorithmic *paradigm*. If you have not seen the Goemans-Williamson approximation algorithm before and are not used to its ideas, which have since become fairly widespread, it likely seems—quite reasonably—ingenious. Some earlier applications of SDP to combinatorial problems are similarly striking in their originality. Perhaps most surprising is Lovászś application of the so-called ϑ *function*, an SDP relaxation of the independent set problem quite similar in flavor to our $\mathsf{SDP}(G)$, to determine the Shannon capacity of the 5-cycle, a long-standing open problem at the time [Lov79]. What is so surprising is that algorithms for efficient semidefinite programming did not exist when Lovász was writing, and would not for more than a decade! Lovász was thinking, therefore, of an SDP not as an algorithm but just as a related optimization problem, in particular with objects similar to our "vector cuts" mentioned above. But, as [BKM19] recently noticed, Lovász's ϑ function, too, is none other than an instance of optimization over SOS proofs. In this way, SOS optimization gives us a unified lens on many brilliant algorithmic developments, and reduces them to the same routine construction.

Second, one of the key challenges we will be looking at is how to analyze not just degree 2 SOS but also degrees 4 and higher. The older algorithms mentioned above, for instance, correspond to SOS proofs of degree 2. As we have seen, higher degree proofs give a promising direction to search for improved algorithms and approximations. But it will gradually become clear that degree 2 is quite often deceptively simple, and, alas, much less is known about degree 4 SOS and above (as Problem 1.1 already suggests). Still, we will survey what is known, and in particular will reach exciting recent developments that, for certain problems

(especially *random* ones where we are interested in an average-case rather than worst-case analysis), have started to clarify the power and limitations of higher-degree SOS.

Exercises

Exercise 1.1. *Show that the graph Laplacian $\boldsymbol{L}$ of a graph $G = (V, E)$, as we have defined it, corresponds to the quadratic form*

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \frac{1}{4} \sum_{\{v,w\} \in E} (x_v - x_w)^2. \tag{1.24}$$

Infer from this the description (1.1) *of* $\mathsf{MaxCut}(G)$ *and verify the calculation in Proposition 1.2.*

Exercise 1.2 ([Erd67])**.** *Derandomize the approximation algorithm of Proposition 1.2: give a deterministic polynomial-time $\frac{1}{2}$-approximation of* MaxCut*.*

> Hint: *Go through the vertices one by one, in any order, choosing prudently on which side of the cut to put each one.*

Exercise 1.3 ([GW95])**.** *Prove Theorem 1.3. Use the following choice of a random $\hat{\boldsymbol{x}}$: let $\boldsymbol{v}_i$ for $i \in V$ be unit vectors so that $X_{ij} = \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$ (the vectors of the "vector cut" in our geometric interpretation of the Goemans-Williamson SDP). Draw a standard Gaussian random vector $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, and set $\hat{x}_i := \mathsf{sgn}(\langle \boldsymbol{g}, \boldsymbol{v}_i \rangle)$.*

> Hint: *By linearity of expectation, it is enough to compute quantities of the form*
>
> $$f(\boldsymbol{v}, \boldsymbol{w}) = \mathop{\mathbb{E}}_{\boldsymbol{g}} [\mathsf{sgn}(\langle \boldsymbol{g}, \boldsymbol{v} \rangle) \, \mathsf{sgn}(\langle \boldsymbol{g}, \boldsymbol{w} \rangle)], \tag{1.25}$$
>
> *for $\boldsymbol{v}, \boldsymbol{w}$ unit vectors. Reduce this to a geometric calculation in the two-dimensional plane, carry out that calculation with some trigonometry, and use the result to complete the proof.*

Exercise 1.4 (Nesterov's approximation algorithm)**.** *We consider the same approximation algorithm as in Theorem 1.3 and Exercise 1.3, but now with $\boldsymbol{L} \succeq \boldsymbol{0}$ an arbitrary psd matrix. Again, let $\boldsymbol{X}$ be a feasible point for the Goemans-Williamson SDP in this setting.*

1. *Prove the matrix inequality*

   $$\mathbb{E}[\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^\top] \succeq \frac{2}{\pi} \boldsymbol{X}. \tag{1.26}$$

   Use the Schur product theorem: if $\boldsymbol{S}, \boldsymbol{T} \succeq \boldsymbol{0}$, then the matrix $\boldsymbol{M}$ with $M_{ij} = S_{ij} T_{ij}$ also has $\boldsymbol{M} \succeq \boldsymbol{0}$.

2. *Consider the* little Grothendieck problem*: given $\boldsymbol{A} \succeq \boldsymbol{0}$, we want to solve the optimization*

   $$\begin{array}{ll} \textit{maximize} & \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \\ \textit{subject to} & \boldsymbol{x} \in \{\pm 1\}^n. \end{array} \tag{1.27}$$

Use the previous parts to describe a randomized polynomial-time $\frac{2}{\pi}$-approximation algorithm for this problem.[8] *Why doesn't your algorithm give the same approximation for arbitrary $\boldsymbol{A}$?*

Exercise 1.5 (Weak integrality gap lower bound)**.** *Compute* $\mathsf{SDP}(G)$ *explicitly for G equal to the cycle on five vertices. Compare this with* $\mathsf{MaxCut}(G)$ *and infer a lower bound on the integrality gap of the Goemans-Williamson SDP. Your bound should be within 0.01 of* $1/\alpha^{\mathsf{GW}}$*.*

> HINT: *It might be easier to reason in the dual of* $\mathsf{SDP}(G)$*.*

Exercise 1.6 (Power of triangle inequalities)**.** *Show that there exists a graph G and $\boldsymbol{X}$ such that:*

1. *$\boldsymbol{X}$ is optimal for* $\mathsf{SDP}(G)$*, i.e., has $\boldsymbol{X} \succeq \boldsymbol{0}$, $X_{ii} = 1$, and $\langle \boldsymbol{L}, \boldsymbol{X} \rangle = \mathsf{SDP}(G)$; and*
2. *$\boldsymbol{X}$ does* not *satisfy the triangle inequalities, i.e., there are i, j, k such that $X_{ij} + X_{jk} + X_{ik} < -1$; and use this to deduce that*
3. $\mathsf{SDP}^{\triangle}(G) < \mathsf{SDP}(G)$*.*

> HINT: *It suffices to consider G the complete graph on three vertices and $\boldsymbol{X} \in \mathbb{R}^{3\times 3}_{\succeq \boldsymbol{0}}$ that is the Gram matrix of three vectors in $\mathbb{R}^2$. Draw a picture of three such unit vectors. How can you make the off-diagonal entries of such $\boldsymbol{X}$ as negative as possible?*

Exercise 1.7 (Heterogeneous degree 2 SOS)**.** *Show that*

$$\mathsf{SDP}(G) = \left\{ \begin{array}{ll} \textit{minimize} & c \\ \textit{subject to} & c = \sum_{i,j} L_{ij} y_i y_j + \sum_i d_i (1 - y_i^2) + \sum_a (w_a + \sum_i v_{a,i} y_i)^2 \\ & \textit{for some } d_i, v_{a,j}, w_a \in \mathbb{R} \end{array} \right\}, \quad (1.28)$$

where we allow constant terms w_a in the polynomials of the "SOS term" of an SOS proof.

NOTES

OTHER SOURCES The presentation in this chapter is heavily inspired by the recent survey article [Moi20] and the earlier notes [Ban15].

APPLICABILITY OF MAXCUT The citation often given for applications in statistical physics and Very Large Scale Integrated (VLSI) circuit design is [BGJR88]. However, calculating exact ground states of spin systems is less relevant than calculating free energies and sampling from Gibbs distributions at positive temperature in statistical physics, which in practice is mostly done with Monte Carlo algorithms. Finding ground states is also more often done by annealing such samplers. Also, practical VLSI optimization typically involves much more

[8] Note that MaxCut is a special case of the little Grothendieck problem, since any graph Laplacian is psd. The Goemans-Williamson analysis for MaxCut uses special properties of graph Laplacians that we do not use in the analysis outlined here, and by taking advantage of those gets the stronger 0.878^{+}-approximation.

complicated constraints than MaxCut-like problems encode conveniently. On the other hand, there are tricks for "moving" constraints into the objective function. And, recent developments suggesting that MaxCut and its generalization to *quadratic unconstrained binary optimization (QUBO)*—where the matrix $\boldsymbol{L}$ can be arbitrary—can be solved efficiently with quantum algorithms has driven more work on such methods. See [KHG+14, GKD18] for surveys.

GREEDY ALGORITHM While the greedy algorithm for MaxCut is clearly implicit in [Erd67], it is sometimes (including by [GW95]) credited to [SG74]. Surprisingly, some aspects of its performance were unknown until quite recently. In particular, [CST11] answered an open question posed by [MS08], showing that randomizing the order in which vertices are processed by the greedy algorithm does not improve its approximation ratio to exceed $\frac{1}{2}$.

MORE ON THE APPROXIMATION RATIO It is broadly interesting what kinds of algorithms can or cannot improve on the $\frac{1}{2}$ approximation ratio. Trevisan in [Tre12] proposed an interesting algorithm that is a $(\frac{1}{2}+\epsilon)$-approximation but avoids solving an SDP, instead only working with the spectrum of G (the Goemans-Williamson SDP is invoked only in the analysis of the algorithm, not the algorithm itself). Several works have shown, on the other hand, that LPs (including those stronger than $\mathsf{LP}(G)$) cannot improve upon the $\frac{1}{2}$ ratio [dlVKM07, STT07] until they run in super-polynomial time [OS18, HST19]. There is a definite intuition suggested by these works that some sort of "access" to the spectrum of G is needed for an algorithm to improve on the $\frac{1}{2}$ approximation ratio, but this has not (to my knowledge) ever been made precise.

Open Problem 1.2 (Non-spectral approximation algorithms for MaxCut)**.** *Is there any approximation algorithm for* MaxCut *that achieves an approximation ratio greater than $\frac{1}{2}$ that does not depend on the spectrum of G or the "spectral reasoning" implicit in SDPs? If not, can we prove a general result making that precise?*

STRONG HARDNESS OF APPROXIMATION In addition to the hardness of approximation conditional on UGC discussed above, it is known that it is NP-hard to approximate MaxCut with a ratio of $\frac{16}{17} \approx 0.941$ [Hås01]. See the paper for a history of such results; there are analogs for various other problems, where the best-known approximation is conjectured optimal under UGC while a weaker hardness result holds under $\mathsf{P} \neq \mathsf{NP}$.

GAIN APPROXIMATION There are other interesting approximation questions besides the approximation ratio as we have defined it. One is the question of *gain* approximation, which, in the context of MaxCut, asks for what $f(\epsilon)$ we can give an algorithm that, if G has a cut of size $(\frac{1}{2}+\epsilon)|E|$, will produce a cut of size at least $(\frac{1}{2}+f(\epsilon))|E|$. The reason for the name is that this question concerns how the gain over a random cut (which cuts on average $\frac{1}{2}|E|$ edges) differs between the optimal cut and the one an algorithm produces. The different rounding of the Goemans-Williamson SDP given by [FL06, CW04] is known to achieve this for $f(\epsilon) = \Omega(\epsilon/\log(\epsilon^{-1}))$, which is optimal if UGC is true [KO06].

LOSS APPROXIMATION Similarly, one may ask for what $g(\epsilon)$ we can give an algorithm that, if G has a cut of size $(1-\epsilon)|E|$, will produce a cut of size at least $(1-g(\epsilon))|E|$. It is not a standard name, but it seems intuitive to call this *loss* approximation, by analogy with gain approximation. Here, the Goemans-Williamson algorithm achieves $g(\epsilon) = O(\epsilon^{1/2})$, and this is again optimal if UGC is true [MOO05].

GROTHENDIECK PROBLEMS A number of similar approximation problems have been treated in the literature and connected to so-called *Grothendieck problems*. [AN04] treated approximation for problems of the form

$$\max_{\substack{\boldsymbol{x}\in\{\pm 1\}^m \\ \boldsymbol{y}\in\{\pm 1\}^n}} \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} \tag{1.29}$$

for $\boldsymbol{A}$ an arbitrary rectangular matrix. The fact that the SDP relaxation of such a problem bounds the original problem by a constant factor independent of the dimensions m, n is known as *Grothendieck's inequality*, first discovered in [Gro56] in a more abstract functional-analytic setting and later further explored and reformulated in our simpler matrix language by [LP68]. Similarly, [Nes98] treated problems of the form

$$\max_{\boldsymbol{x}\in\{\pm 1\}^n} \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \tag{1.30}$$

for $\boldsymbol{A}$ either arbitrary or psd. The latter is a direct generalization of the Goemans-Williamson setting where $\boldsymbol{A}$ must further be a Laplacian matrix; this variant is known as the *little Grothendieck problem*. For $\boldsymbol{A} \succeq \boldsymbol{0}$ the Goemans-Williamson rounding also gives an effective approximation algorithm, while for $\boldsymbol{A}$ arbitrary [CW04] studied approximation algorithms, albeit ones that do not achieve a constant approximation ratio.

Part I

General Theory

2 | Algebraic Proof Systems

We have seen in Chapter 1 the first glimmer of a powerful *proof system* that manipulates polynomials to bound polynomial optimization problems. We have also suggested that we may optimize over such proofs using semidefinite programming. Before proceeding to these algorithmic applications, we will review the line of mathematical work that led to such a proof system. In doing so, we will look at questions that basically amount to the *completeness* of the SOS proof system—when is it true that a certain polynomial being bounded implies that there exists an SOS proof of that boundedness?

2.1 Obstructions and Certificates: A Simple Example

To illustrate the general structure of the kinds of statements we will look at, let us first give a completely elementary example. Suppose we are interested in whether two integers $a, b \in \mathbb{Z}$ have a non-trivial common divisor (one that is greater than 1). The following is a result of elementary number theory.

Proposition 2.1 ("Weak" Bézout's identity)**.** *Suppose $a, b \in \mathbb{Z}$. Then, exactly one of the following holds:*

1. *There exists $d > 1$ with $d \mid a$ and $d \mid b$.*
2. *There exist $x, y \in \mathbb{Z}$ with $ax + by = 1$.*

A linear equation as in Condition 2 in integer variables is often called a *linear Diophantine equation (LDE)*.

We may note immediately that the two cases are mutually exclusive: if both Conditions 1 and 2 held then we would have $d \mid 1$, a contradiction to $d > 1$. The content of the result is that *exactly* one of the two holds. We might view this as saying that the only *obstruction* to the existence of a common divisor of a and b is a solution to the LDE $ax + by = 1$. Alternatively, we might say that any such LDE is a *refutation* of a common divisor or a simple *proof* that there exists no common divisor. The Proposition tells us that such a proof system of refuting common divisors is *complete*: whenever a and b share no common divisor, there is an "LDE proof" of this fact.

Thinking more algorithmically, several further questions are natural. First, are there algorithms to efficiently search for LDE proofs? Naively, we might simply try brute force, enumerating $(x, y) \in \mathbb{Z}^2$ in order of, say, $|x| + |y|$. How long would such an approach take? This leads us to a *proof complexity* question: when a and b share no divisor, do there exist

x, y with $ax + by = 1$ and $|x| + |y| \leq F(a,b)$ for some small value $F(a,b)$? What are the a, b that achieve this bound, the "hardest" instances for LDE refutation of common divisors?

You might be aware that all of these questions have fairly simple answers in elementary number theory (if not, see the chapter notes). But, below, we will proceed to precisely analogous questions over *polynomials* rather than integers, and will see that mysteries quickly crop up.

2.2 Hilbert's Nullstellensatz

The following foundational result of algebraic geometry, due to Hilbert in [Hil93a], may be thought of as a direct analogue of Proposition 2.1 for polynomials.[1]

Theorem 2.2 (Weak Nullstellensatz)**.** *Suppose $p_1, \ldots, p_m \in \mathbb{C}[x_1, \ldots, x_n]$. Then, exactly one of the following holds:*

1. *The p_i have a common zero: there exists $z \in \mathbb{C}^n$ with $p_1(z) = \cdots = p_m(z) = 0$; or*
2. *1 belongs to the ideal generated by the p_i: there exist $q_1, \ldots, q_m \in \mathbb{C}[x_1, \ldots, x_n]$ such that $\sum_{i=1}^m p_i q_i = 1$.*

Without going too far afield, let us say a few words about the proof ideas. Like Proposition 2.1, Theorem 2.2 should be viewed as a structure theorem about the *ideals* in a particular commutative ring. An *ideal* I is a set closed under addition and "contagious" under multiplication, so that if $a \in I$ then $ab \in I$ for any b in the ring in question. The typical example is $(a_1, \ldots, a_n) := \{\sum_{i=1}^n a_i b_i\}$, the set of linear combinations of some collection of a_i. A ring is *Noetherian* when all ideals are of this form, and this is the case for all of $\mathbb{Z}$, $\mathbb{R}[x_1, \ldots, x_n]$, and $\mathbb{C}[x_1, \ldots, x_n]$, the only rings we will look at.[2]

Proposition 2.1 treats the ring $\mathbb{Z}$, whose ideals are quite simple: each one is *principal*, equal to (n), the set of multiples of some number n. The stronger Bézout's identity says that the ideal generated by *two* numbers a, b is actually equal to the ideal generated by *one* number, their greatest common divisor.

Proof Sketch of Nullstellensatz. The situation is not so simple for $\mathbb{C}[x_1, \ldots, x_n]$, but there is an analogous structure theorem, where common zeroes play the role of common divisors: the *maximal* ideals (under inclusion) of $\mathbb{C}[x_1, \ldots, x_n]$ are those equal to $(x_1 - z_1, \ldots, x_n - z_n)$, which are the polynomials that are zero at a particular $z = (z_1, \ldots, z_n)$. Taking this for granted, the weak Nullstellensatz follows since, if $1 \notin (p_1, \ldots, p_m)$, then $(p_1, \ldots, p_m) \neq \mathbb{C}[x_1, \ldots, x_n]$, so it is contained in some maximal ideal, $(p_1, \ldots, p_m) \subseteq (x_1 - z_1, \ldots, x_n - z_n)$ for some z, and so the p_i have the common zero z.

The reason for the structure theorem on maximal ideals is deeper and related to the algebraic closedness of $\mathbb{C}$ (that is, the fundamental theorem of algebra). The idea is that if

[1] The ordinary or "strong" Nullstellensatz concerns solutions of systems with a further constraint $r(z) \neq 0$, but can be derived easily from the weak Nullstellensatz.

[2] For polynomial rings, this is a consequence of another famous commutative algebra result of Hilbert's, his *basis theorem* [Hil90].

I is a maximal ideal of $\mathbb{C}[x_1, \ldots, x_n]$, then the quotient $\mathbb{C}[x_1, \ldots, x_n]/I$ is a field.[3] This field contains $\mathbb{C}$, and is finitely generated over $\mathbb{C}$ by the images of $x_1, \ldots, x_n$ under the quotient. Moreover, crucially, one can show that these images are *algebraic* over $\mathbb{C}$: they are roots of some polynomial with coefficients in $\mathbb{C}$. But all such polynomials have all their roots in $\mathbb{C}$ already! So $\mathbb{C}[x_1, \ldots, x_n]/I$ is isomorphic to $\mathbb{C}$ again. Then, first under the quotient and then through this isomorphism, we may view x_i as mapping to some $z_i \in \mathbb{C}$. In particular, $x_i - z_i$ maps to zero, so I contains $(x_1 - z_1, \ldots, x_n - z_n)$, which one can show is itself a maximal ideal, and so $I = (x_1 - z_1, \ldots, x_n - z_n)$. □

2.2.1 Nullstellensatz Effectivization and Proof Systems

We give a proof sketch of the Nullstellensatz as it is typically presented in commutative algebra textbooks to emphasize why it is often called "non-constructive." This statement refers to the fact that, even if we are given $p_1, \ldots, p_m$ with no common zero, the proof does not give us a way to construct the certificate in Condition 2 of Theorem 2.2. This is because of the step where, assuming $(p_1, \ldots, p_m)$ is a proper ideal, we find a maximal ideal containing it. In general, the fact that any proper ideal in a commutative ring is contained in a maximal ideal requires the Axiom of Choice for its proof, and indeed is logically equivalent to the Axiom of Choice, and therefore fundamentally cannot be made constructive.

Fortunately, for the specific ring $\mathbb{C}[x_1, \ldots, x_n]$ this is not the case, and accordingly there are alternative proofs that are constructive and correspond to algorithms for testing whether $p_1, \ldots, p_m$ have a common zero. These come in two flavors.

Gröbner Basis Construction First, every ideal I has a *Gröbner basis*, a set of generators satisfying certain consistency properties with respect to a fixed ordering of monomials. One consequence of the restrictions on Gröbner bases is that they allow direct testing of whether $1 \in I$; indeed, this is the case if and only if 1 is one of the generators in the basis. And, a Gröbner basis can be computed using *Buchberger's algorithm*, an analog for polynomials of Euclid's algorithm for finding LDE proofs over integers. The algorithm always terminates, but its worst-case runtime is $(\max \deg p_m)^{2^{\Omega(n)}}$. On the other hand, the runtime can be fast in practice, and its speed can be dramatically improved by a good choice of monomial ordering, so Buchberger's algorithm can be practically useful. Both Gröbner bases and Buchberger's algorithm date to Buchberger's 1965 thesis [Buc06]; the books [CLOS94, Stu02] are good resources for this approach.

Linear System Refutations More relevant to the approach we will pursue, suppose we restrict the degrees of all polynomials appearing in a Nullstellensatz refutation by $\deg p_i q_i \leq D$, and ask for an equality $\sum_{i=1}^{m} p_i q_i = 1$. Then, by equating coefficients on either side, this may be written as a linear system, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. While we will see that this approach is ultimately equal in power to Buchberger's algorithm, it is much more flexible, in the sense that even if we have a small computational budget that would not allow Buchberger's algorithm to

[3] To get some intuition, remember the case of $\mathbb{Z}$: all ideals there are of the form (n) for some integer n, and you can check that the maximal ones are (p) where p is prime. And indeed, $\mathbb{Z}/(p)$ is the finite field on p elements of integers modulo p, while $\mathbb{Z}/(n)$ for composite n is only the ring of integers modulo n.

terminate, we can still *try* to look for a low-degree refutation, trading power in our proof system for a faster runtime.

Example 2.3. *Suppose we wish to refute the system of linear equations*

$$p_1(x,y) = x + y = 0, \tag{2.1}$$
$$p_2(x,y) = 2x + 3y - 1 = 0, \tag{2.2}$$
$$p_3(x,y) = 3x - y - 2 = 0 \tag{2.3}$$

with a degree 2 Nullstellensatz certificate. Then, we would introduce nine scalar variables (a,b,c,d,e,f,g,h,j) *and the polynomials*

$$q_1(x,y) = ax + by + c, \tag{2.4}$$
$$q_2(x,y) = dx + ey + f, \tag{2.5}$$
$$q_3(x,y) = gx + hy + j \tag{2.6}$$

and expand the polynomial equality,

$$\begin{aligned} 1 &= p_1(x,y)q_1(x,y) + p_2(x,y)q_2(x,y) + p_3(x,y)q_3(x,y) \\ &= (ax+by+c)(x+y) + (2x+3y-1)(dx+ey+f) + (3x-y-2)(gx+hy+j) \\ &= (a+2d+3g)x^2 + (b+3e-h)y^2 + (a+2e+3h)xy \\ &\qquad + (c-d-2g)x + (c-e-j)y + (-f-2j). \end{aligned} \tag{2.7}$$

Equating coefficients gives the new system

$$a + 2d + 3g = 0, \tag{2.8}$$
$$b + 3e - h = 0, \tag{2.9}$$
$$a + 2e + 3h = 0, \tag{2.10}$$
$$c - d - 2g = 0, \tag{2.11}$$
$$c - e - f = 0, \tag{2.12}$$
$$-f - 2j = 1, \tag{2.13}$$

or, in matrix notation,

$$\begin{bmatrix} 1 & 0 & 0 & 2 & 0 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 0 & 3 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 2 & 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ j \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{2.14}$$

We can just repeat this with ever increasing degrees D—if the initial polynomial system has no solution, then the Nullstellensatz promises that we will eventually find a certificate in this way. But how long will it take? And can we ever stop searching for a refutation and be assured that the polynomial system actually *does* have a solution?

These questions ask for an effectivization of the Nullstellensatz in terms of the degree: what is the largest possible degree of the lowest-degree refutation of a given polynomial system? The following result of Hermann in 1925 was the first one to this effect.[4]

Theorem 2.4 ([Her98])**.** *There exists a Nullstellensatz refutation of an infeasible polynomial system with* $D = \max \deg p_i q_i \leq (\max \deg p_i)^{2^{O(n)}}$.

A similar result holds for the general *ideal membership* problem of asking whether some $g(\boldsymbol{x})$ can be written as $g = \sum_{i=1}^{m} p_i q_i$, and in this case the doubly-exponential bound is unfortunately optimal [MM82]. However, the following improvement holds for the weak Nullstellensatz.

Theorem 2.5 ([Bro87, Kol88])**.** *There exists a Nullstellensatz refutation of an infeasible polynomial system with* $D = \max \deg p_i q_i \leq (\max \deg p_i)^n$.

This latter bound was also shown by [Kol88] to be optimal. So, in principle, we may find Nullstellensatz refutations by solving sufficiently large linear systems, where "large" is exponential in the number of variables and polynomial in the underlying degree.

Example 2.6 (Nullstellensatz refutation of MaxCut)**.** *Consider encoding* MaxCut*, as we considered in Chapter 1, in such a polynomial system. It is most convenient to encode the statement "there is a cut of size k in G," which is equivalent to there being a $\boldsymbol{x} \in \mathbb{C}^{|V|}$ that satisfies the system*

$$x_i^2 - x_i = 0 \text{ for all } i \in V, \tag{2.15}$$

$$\sum_{i,j=1}^{n} L_{ij} x_i x_j - k = 0 \tag{2.16}$$

where $\boldsymbol{L}$ is the graph Laplacian as usual. Here, all polynomials are quadratic and the number of variables is $|V|$, so Theorem 2.5 says that it suffices to consider certificates of degree $D = 2^{|V|}$. A polynomial of degree D in n variables has $\theta(n^D)$ coefficients, so these would be linear systems of $|V|^{2^{|V|}}$ equations in roughly as many variables—completely impractical to solve!

For other interesting combinatorial problems the general bounds from Theorem 2.5 are equally useless.

For this reason, more recently, another line of work sought, for *specific* polynomial systems of interest, often encoding computational or combinatorial problems, to prove upper or lower bounds on the degree required for Nullstellensatz refutations. These are just a weaker (though computationally cheaper) version of SOS proofs, so we will not get into the details here, but some important references are [BIK+96, GV01] and Chapter 6 of [FKP19] gives a thorough survey. Besides being a natural stop on the way to SOS proofs, this research direction fits into a much broader and longer story about proof complexity and restricted logical proof systems, which we discuss in the chapter notes.

[4] See the cited translation for some discussion of Hermann's remarkable reasoning, which has a quite algorithmic quality despite coming decades before the development of the general theory of computation.

2.3 The Real Case

A natural follow-up question to the Nullstellensatz is whether we can say something similar for real polynomials, $p_1, \ldots, p_m \in \mathbb{R}[x_1, \ldots, x_n]$. We can immediately see that, even when $n = 1$, the Nullstellensatz proof system is no longer complete: the polynomial $p(x) = 1 + x^2$ has no real zeros, but nor is there any $q(x)$ so that $p(x)q(x) = 1$. Thus in the real case there must be some other obstruction to polynomials having zeros besides the "ideal obstruction" from the Nullstellensatz.

2.3.1 Real Nullstellensatz

Indeed, generalizing the $1 + x^2$ example, we see that any $1 + \sum s_i(\boldsymbol{x})^2$ for $s_i \in \mathbb{R}[x_1, \ldots, x_n]$ will be a polynomial having no zeroes. We may therefore add this as a new obstruction in the real case.

Definition 2.7 (SOS polynomials)**.** $\mathsf{SOS} \subset \mathbb{R}[x_1, \ldots, x_n]$ *(with the dependence on n left implicit) is the set of $\sum_{i=1}^m s_i(\boldsymbol{x})^2$ for some $m \in \mathbb{N}$. We also say that p* is SOS *when $p \in \mathsf{SOS}$.*

A stronger refutation of a shared zero of $p_1, \ldots, p_m$ is then an equality of the form

$$\sum_{i=1}^{m} p_i(\boldsymbol{x}) q_i(\boldsymbol{x}) = 1 + \underbrace{\sum_{i=1}^{p} s_i(\boldsymbol{x})^2}_{\in \mathsf{SOS}}. \tag{2.17}$$

It turns out that a proof system based on this kind of refutation is in fact complete for systems of real algebraic equations.

Theorem 2.8 (Weak real Nullstellensatz)**.** *Suppose $p_1, \ldots, p_m \in \mathbb{R}[x_1, \ldots, x_n]$. Then, exactly one of the following holds:*

1. *The p_i have a common zero: there exists $\boldsymbol{z} \in \mathbb{R}^n$ with $p_1(\boldsymbol{z}) = \cdots = p_m(\boldsymbol{z}) = 0$; or*
2. *$1 + \mathsf{SOS}$ intersects the ideal generated by the p_i: there exist $q_1, \ldots, q_m, s_1, \ldots, s_p \in \mathbb{R}[x_1, \ldots, x_n]$ such that $\sum_{i=1}^m p_i q_i = 1 + \sum_{j=1}^p s_j^2$.*

In fact, historically this result (as well as a "strong" real Nullstellensatz) arose, first in the work of Krivine [Kri64] and later more explicitly in that of Stengle [Ste74], only as a special case of much more general *Positivstellensatz* results that we will discuss below, where we also allow *inequality* constraints $q_i(\boldsymbol{x}) \geq 0$ in the system we seek to refute. Before getting to those results, let us step back and first review some results that considered the simpler question of whether, as we have implicitly suggested above, sum-of-squares polynomials adequately represent all non-negative polynomials. This will prepare us for the several different strengths of Positivstellensatz available under different assumptions, which are useful to varying extents for certification of bounds on polynomial optimization problems.

2.3.2 Sums of Squares and Hilbert's 17th Problem

We consider a similar, but not quite identical, question to the case $m = 1$ of the weak real Nullstellensatz. Suppose $p \in \mathbb{R}[x_1, \ldots, x_n]$. The difference is just that we suppose only $p \geq 0$ rather than $p > 0$, and ask, does p belong to SOS?

We can give a few positive results in special cases.

Proposition 2.9 (Univariate polynomials)**.** *Suppose $p \in \mathbb{R}[x]$ has $p(x) \geq 0$ for all $x \in \mathbb{R}$. Then, p is SOS.*

Proof 1. $|p(x)| \to \infty$ as $x \to \infty$ and as $x \to -\infty$, so the minimum of $p(x)$ is achieved at some $c \in \mathbb{R}$. Let $t = p(c) \geq 0$ and let $q(x) = p(x) - t \geq 0$. $q(x)$ then has a zero of even order at c, say of order $2k$. Thus $q(x) = (x-c)^{2k} r(x)$ for some $r(x)$ with $\deg r < \deg p$ and $r \geq 0$. We then have $p(x) = (x-c)^{2k} r(x) + t$, and repeating this inductively gives an SOS decomposition. □

A more careful proof also shows that only two squares suffice.

Proof 2. The leading coefficient of p must be positive, so we may factor this out and assume without loss of generality that p is monic. Let $r_1, \ldots, r_m$ be the distinct real roots of p and $a_1 \pm b_1 i, \ldots, a_n \pm b_n i$ be the complex roots of i, counted with multiplicity, which must come in conjugate pairs. Each r_i must be a root of even order, say some $2k_i$. Then, we have

$$p(x) = \prod_{i=1}^{m} (x - r_i)^{2k_i} \prod_{j=1}^{n} (x - a_j - b_j i)(x - a_j + b_j i) \tag{2.18}$$

Let us write $s_1(x) = \prod_{i=1}^{m} (x - r_i)^{k_i}$ and $s_2(x) + i s_3(x) = \prod_{j=1}^{n} (x - a_j - b_j i)$ upon grouping real and complex coefficients. We then have

$$\begin{aligned} p(x) &= s_1(x)^2 (s_2(x) + i s_3(x))(s_2(x) - i s_3(x)) \\ &= s_1(x)^2 (s_2(x)^2 + s_3(x)^2) \\ &= (s_1(x) s_2(x))^2 + (s_1(x) s_3(x))^2, \end{aligned} \tag{2.19}$$

completing the proof. □

Proposition 2.10 (Quadratic forms)**.** *Suppose $p \in \mathbb{R}[x_1, \ldots, x_n]$ is homogeneous of degree 2 and has $p(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Then, there exist $s_1, \ldots, s_n \in \mathbb{R}[x_1, \ldots, x_n]$ with $p = \sum_{i=1}^{n} s_i^2$.*

Proof. Since p is homogeneous, we may express it as $p(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x}$ for some $\boldsymbol{A} \in \mathbb{R}^{n \times n}_{\mathsf{sym}}$. By the spectral theorem, there are $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and an orthonormal basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of $\mathbb{R}^n$ so that $\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top$, whereby $p(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i \langle \boldsymbol{v}_i, \boldsymbol{x} \rangle^2$. Then, $p \geq 0$ if and only if $\lambda_i \geq 0$ for all $i \in [n]$, and in this case we have an SOS representation $p(\boldsymbol{x}) = \sum_{i=1}^{n} (\sqrt{\lambda_i} \langle \boldsymbol{v}_i, \boldsymbol{x} \rangle)^2$. □

In Exercise 2.2 you will show that the same holds under the weaker condition $\deg p \leq 2$, with no homogeneity assumption.

A result of Hilbert settled all remaining cases: just other one degree and arity leads to SOS and non-negative polynomials being equivalent, while all others do not.

Theorem 2.11 (Hilbert [Hil88])**.** *The following hold:*

1. *Every non-negative polynomial of degree 4 in 2 variables is SOS.*

2. *If $n \geq 3$ and $d \geq 4$ or $n = 2$ and $d \geq 6$, for even d, there exists a non-negative polynomial of degree d in n variables that is not SOS.*

(Note that there do not exist non-negative polynomials of odd degree, since such polynomials will have a leading term that is negative and dominates the others going to infinity along a suitable ray from the origin.)

Hilbert's proof of the second part was not constructive, and it took many years for concrete examples to materialize, though simple ones exist. The best-known example is the following.[5]

Theorem 2.12 (Motzkin's polynomial [Mot67])**.** *The polynomial $p(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ is non-negative for all $(x,y) \in \mathbb{R}^2$, but is not SOS.*

Proof. To see that $p(x,y) \geq 0$, we apply the arithmetic-geometric mean inequality:

$$x^4y^2 + x^2y^4 + 1 \geq 3(x^6y^6)^{1/3} = 3x^2y^2. \tag{2.20}$$

To see that p is not SOS, suppose we have $p = \sum_{i=1}^m s_i^2$. No monomial with a power of x or y greater than or equal to 3 cannot occur in any s_i, or else p would have a leading term with at least a sixth power. Similarly, the monomial x^2y^2 cannot occur in any s_i, or else x^4y^4 would appear in p, then by the same argument the monomials x^2 and y^2 cannot occur in any s_i, and then the monomials x and y cannot occur either. The only remaining monomials that can appear in each s_i are x^2y, xy^2, xy, and 1. But then, the coefficient of x^2y^2 in p must be non-negative (since it can only come from squared xy terms), which is a contradiction. □

Thus to fully represent all non-negative polynomials we must somehow expand our notion of sums of squares. Hilbert proposed such an expanded definition, asking in the 17th problem of his famous 1900 address to the International Congress of Mathematicians: is every non-negative polynomial a sum of *rational* squares? Hilbert had already showed that this is the case for bivariate polynomials (with $n = 2$) in [Hil93b], generalizing the quartic case from Theorem 2.11. Artin soon resolved the rest of the problem in the affirmative.

Theorem 2.13 (Artin [Art27])**.** *Suppose $p \in \mathbb{R}[x_1, \ldots, x_n]$ with $p \geq 0$. Then, there exist $r_1, \ldots, r_m, s_1, \ldots, s_m \in \mathbb{R}[x_1, \ldots, x_n]$ such that $p = \sum_{i=1}^m (r_i/s_i)^2$.*

As you will show in Exercise 2.5, the conclusion is equivalent to p being a quotient of SOS polynomials, which is to say to there existing $q \in \mathsf{SOS}$ such that $pq \in \mathsf{SOS}$.

Example 2.14. *Motzkin's polynomial can be written as a sum of four rational squares as follows:*

$$x^4y^2 + x^2y^4 + 1 - 3x^2y^2 = \frac{(1 + x^2 + y^2)x^2y^2(x^2 + y^2 - 2)^2 + (x^2 - y^2)^2}{(x^2 + y^2)^2}, \tag{2.21}$$

which gives an SOS representation upon distributing in the numerator.

[5] While it is the example most often given, there is a misconception that the Motzkin polynomial is a very special and "finely tuned" example. That is not so: we show that it can be "perturbed" in Exercise 2.4 and discuss other, unrelated examples in the chapter notes.

A few mysteries persist surrounding Artin's result. The following question highlights the inconvenience of variable denominators in Artin's theorem, which we will discuss further in the next section.

Open Problem 2.1 (Convexity of Artin cones [AH19]). *Given $r, d, n \geq 1$, let C be the cone of homogeneous $p \in \mathbb{R}[x_1, \ldots, x_n]$ so that $\deg p = 2d$ and there exists a homogeneous $q \in \mathbb{R}[x_1, \ldots, x_n]$ that is SOS and has $\deg q = 2r$ so that $pq \in \mathsf{SOS}$. (In words, C is the cone of quotients of SOS polynomials with homogeneous numerator and denominator of fixed degree.) Is C convex for all choices of r, d, n?*

There is also the following striking bound on how many summands are needed. Perhaps surprisingly, it depends only on the number of variables.

Theorem 2.15 (Pfister [Pfi67]). *$m = 2^n$ suffices in Theorem 2.13.*

One part of Pfister's argument is particularly charming and worth highlighting: there are classical formulas, starting with

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2, \tag{2.22}$$

showing that sums of 1 (trivially), 2, 4, and 8 squares are closed under multiplication, where the terms getting squared on the right-hand side are bilinear in the variables showing up in the two sums of squares on the left-hand side (as above).[6] These have a connection to division algebras equipped with well-behaved norms, corresponding to the multiplicativity of the squared norm for real numbers, complex numbers, quaternions, and octonions, respectively. A classical result of Hurwitz shows that such a formula cannot hold for any other number of squares [Hur98]. However, Pfister showed the surprising result that such a formula holds for *any* number of squares that is a power of two—provided we allow rational functions on the right-hand side! See [Ben17] for more discussion, Conrad's notes [Con] for an exposition of this part of Pfister's argument, or [Pfi95] for a book-length treatment.

One may ask if this is optimal. For $n = 1$ it is because the polynomial $1 + x^2$ requires two squares, while for $n = 2$ it is because the Motzkin polynomial requires four squares—though this latter is a non-trivial result of [CEP71].

Open Problem 2.2 (Number of summands in Artin's theorem). *For $n \geq 3$, what is the maximum number of rational square summands needed to express a non-negative polynomial of $\mathbb{R}[x_1, \ldots, x_n]$? The best known general lower bound is $n + 1$ (see Exercise 2.8), while the best known upper bound is 2^n from Theorem 2.15.*

Relatedly, [DLV04] proved lower bounds on the *complexity* of rational SOS representations in terms of both the number of summands and circuits computing the polynomials getting squared.

[6] The cases of four squares, for instance, may be used to reduce Lagrange's famous theorem that every non-negative integer is a sum of four integer squares to the special case of primes.

2.3.3 Dealing with Denominators

When we later consider SOS optimization more formally, given $p(\boldsymbol{x})$, we will be interested (in the simplest case) in solving problems of the form

$$\begin{array}{ll} \text{minimize} & c \\ \text{subject to} & c - p(\boldsymbol{x}) = s(\boldsymbol{x}), \\ & s(\boldsymbol{x}) \in \mathsf{SOS} \end{array} \tag{2.23}$$

subject to degree constraints. This will prove a global bound $p(\boldsymbol{x}) \leq c$. We will see that this is possible with semidefinite programming.

Motivated by Artin's theorem, we might be tempted instead to try to solve a problem of the form

$$\begin{array}{ll} \text{minimize} & c \\ \text{subject to} & s(\boldsymbol{x})(c - p(\boldsymbol{x})) = t(\boldsymbol{x}), \\ & s(\boldsymbol{x}), t(\boldsymbol{x}) \in \mathsf{SOS}, \end{array} \tag{2.24}$$

again subject to degree constraints.[7] However, this kind of problem, where both c and s are variables to be optimized over, can no longer be solved with semidefinite programming (and more generally is not convex), since we now have a nonlinear constraint involving the product $c \cdot s(\boldsymbol{x})$.

There are two reasonable ways around this. First, for the given type of problem, we could solve feasibility SDPs repeatedly for a fixed c and perform binary search to try to minimize c. On the other hand, it also seems that, so long as we are solving many SDPs, many search iterations might be saved by sometimes letting s be fixed and minimizing c.

Open Problem 2.3 (Denominator pursuit for rational function SOS)**.** *Investigate the efficacy of algorithms that, in the setting of* (2.24)*, alternate solving feasibility problems over* $(s(\boldsymbol{x}), t(\boldsymbol{x}))$ *with varying* c *and minimizing over* $(c, t(\boldsymbol{x}))$*, especially as compared to binary search over* c*.*

Alternatively, we can take advantage of an interesting class of results on *uniform denominators* in Artin's theorem. These results state that, under some assumptions, Artin's theorem still holds with a specific form of denominator shared among all terms, and, in effect, suggest a specific class of $s(\boldsymbol{x})$ that we might use above.

The following is the earliest result on uniform denominators.

Theorem 2.16 (Pólya [Pól28])**.** *Suppose that* $p \in \mathbb{R}[x_1, \ldots, x_n]$ *has* $p(\boldsymbol{x}) > 0$ *for all* $\boldsymbol{x} \neq \mathbf{0}$*,* p *is homogeneous, and* p *even, i.e.* $p(-\boldsymbol{x}) = p(\boldsymbol{x})$*. Then, for all sufficiently large* N*,* $(\sum_{i=1}^n x_i^2)^N \cdot p(\boldsymbol{x})$ *has only positive coefficients (and therefore is a sum of squares of monomials).*

Later, [PR01] made Pólya's theorem effective, describing how large N may be taken.

Using Pólya's theorem, the following gives a direct proof of a result close to the full power of Artin's theorem.

[7] As mentioned earlier, while the theorem expresses polynomials as sums of quotients of squares, this is equivalent to an expression as a quotient of sums of squares, as invoked here. Exercise 2.5 asks you to prove this.

Theorem 2.17 (Habicht [Hab39])**.** *Suppose that $p \in \mathbb{R}[x_1, \ldots, x_n]$ has $p(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ and p is homogeneous of even degree. Then, there exist $r, s \in \mathbb{R}[x_1, \ldots, x_n]$ that are homogeneous of even degree and have only positive coefficients (therefore each being sums of squares of monomials) such that $p = r/s$.*

This result is of some historical interest, as it admits, through Pólya's theorem (which has an elementary proof), a proof without the complicated machinery of ordered fields involved in Artin's theorem.

Finally, the strongest and most directly useful of these results is the following more recent one. See also the citation for quantitative bounds on N in terms of properties of the polynomial p.

Theorem 2.18 (Reznick [Rez95])**.** *Suppose that $p \in \mathbb{R}[x_1, \ldots, x_n]$ has $p(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ and p is homogeneous of even degree. Then, there exists $N \in \mathbb{N}$ and $q \in \mathsf{SOS}$ homogeneous so that $p = q/(\sum_{i=1}^n x_i^2)^N$.*

This combines useful features of Pólya's and Habicht's theorems: like Pólya's theorem it has a uniform denominator, while like Habicht's theorem it does not need the restrictive assumption that p be an even polynomial.

In any case, to the best of my knowledge almost nothing is known about the kinds of lower bounds we will look at later in the course when we look at rational function SOS proofs rather than polynomial ones.

Open Problem 2.4 (Lower bounds on rational function SOS)**.** *Investigate the power of SOS with low-degree rational functions or quotients of low-degree sums of squares for any SOS degree lower bounds proved on polynomial problems in the computer science literature.*

The formulation is slightly facetious—there are a few exceptions, like the results of [BGP16] for symmetric functions on the hypercube—but for less structured problems I believe this question is wide open.

The work of [AH19] proposes hierarchies of numerical methods, similar to the "standard" SOS hierarchy we introduce in Chapter 4, based on Pólya's and Reznick's uniform denominator theorems as well. The one based on Pólya's theorem is especially intriguing since it does not require any convex optimization but rather only computing and checking the signs of coefficients in polynomial products. It remains unclear how these compare to other SOS methods for concrete problems, however.

2.3.4 Positivstellensätze

We now return to the weak real Nullstellensatz mentioned earlier, and describe a further broad class of results to which it belongs, the *Positivstellensätze*.[8] We will see that while these results in their most general form resemble the weak real Nullstellensatz, in fact they also have corollaries that are useful for certifying positivity and non-negativity in the way of the results on SOS representations above.

[8] *Satz* is German for *theorem*, and *sätze* is German for *theorems*.

Theorem 2.19 (Krivine-Stengle "weak" Positivstellensatz [Kri64, Ste74])**.** *Let* $p_1, \dots, p_m \in \mathbb{R}[x_1, \dots, x_n]$*. Then, exactly one of the following holds:*

1. *There exists* $\boldsymbol{z} \in \mathbb{R}^n$ *such that* $p_i(\boldsymbol{z}) \geq 0$ *for all* $i \in [m]$*.*
2. *There exist* $q_S \in \mathsf{SOS}$ *for each* $S \subseteq [m]$ *such that*

$$\sum_{S \subseteq [m]} q_S(\boldsymbol{x}) \prod_{i \in S} p_i(\boldsymbol{x}) = -1. \tag{2.25}$$

A set defined by a collection of polynomial inequalities is called a *semialgebraic set*. Thus the Krivine-Stengle weak Positivstellensatz gives a condition for the non-emptiness of a semialgebraic set, in the same way that the Nullstellensatz gives a condition for the non-emptiness of a complex variety and the weak real Nullstellensatz gives a condition for the non-emptiness of a real variety.

Actually, a real variety is just a special case of a semialgebraic set, and so, as promised, the weak real Nullstellensatz is an easy consequence of the weak Positivstellensatz.

Proof of Theorem 2.8. Apply Theorem 2.19 with the polynomials $\pm p_i$. Then, Condition 1 is equivalent to $p_i(\boldsymbol{z}) = 0$ for all $i \in [m]$, which is Condition 1 of Theorem 2.8. Since any polynomial is a difference of two SOS polynomials (for example, $p = \frac{1}{4}(1+p)^2 - \frac{1}{4}(1-p)^2$), Condition 2 is equivalent to there existing some $q_1, \dots, q_m \in \mathbb{R}[x_1, \dots, x_n]$ and $q_0 \in \mathsf{SOS}$ so that

$$q_0(\boldsymbol{x}) + \sum_{i=1}^{m} p_i(\boldsymbol{x}) q_i(\boldsymbol{x}) = -1, \tag{2.26}$$

rearranging which gives Condition 2 of Theorem 2.8. □

More generally, by including both $\pm p$ for some polynomial p in the constraints, Theorem 2.19 allows for refuting systems with some polynomial equalities and some polynomial inequalities.

We can also derive sufficient forms of certificates for the positivity and non-negativity of a polynomial over a semialgebraic set. Traditionally these are called ordinary or "strong" Positivstellensätze, though the naming is misleading because they follow from the weak Positivstellensatz (as for the weak and strong Nullstellensatz).

Corollary 2.20 (Krivine-Stengle Positivstellensatz)**.** *Suppose that* $p_1, \dots, p_m \in \mathbb{R}[x_1, \dots, x_n]$ *and define the sets*

$$\mathcal{K} := \{\boldsymbol{x} \in \mathbb{R}^n : p_i(\boldsymbol{x}) \geq 0 \text{ for all } i \in [m]\},$$
$$S := \left\{ \sum_{S \subseteq [m]} q_S(\boldsymbol{x}) \prod_{i \in S} p_i(\boldsymbol{x}) : q_S \in \mathsf{SOS} \right\}.$$

Let $r \in \mathbb{R}[x_1, \dots, x_n]$ *be a further polynomial. Then, the following hold:*

1. $r(\boldsymbol{x}) > 0$ *for all* $\boldsymbol{x} \in \mathcal{K}$ *if and only if there exist* $s_1, s_2 \in S$ *such that* $s_1 r = 1 + s_2$*.*
2. $r(\boldsymbol{x}) \geq 0$ *for all* $\boldsymbol{x} \in \mathcal{K}$ *if and only if there exist* $s_1, s_2 \in S$ *and* $a \in \mathbb{N}$ *such that* $s_1 r = r^{2a} + s_2$*.*

Actually, the second result is only a corollary of a slightly expanded Krivine-Stengle weak Positivstellensatz, which we do not include here. We mention this result just because, with its inclusion, Artin's theorem is an easy further corollary of this corollary.

Proof of Theorem 2.13. We take $m = 1$ and set $p_1 = 1$. Then, $\mathcal{K} = \mathbb{R}^n$ and $S = \mathsf{SOS}$. So, Result 2 of Corollary 2.20 says that, if $r(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$, then for some $a \in \mathbb{Z}$ and $s_1, s_2 \in \mathsf{SOS}$, we have $r = (r^{2a} + s_2)/s_1$, which upon applying Exercise 2.5 gives the desired representation. □

It is best to think of all of the results on real-valued certificates and refutations we have seen so far (except those on uniform denominators) as being mostly of historical interest. The Krivine-Stengle weak Positivstellensatz is really the "mother theorem" that captures all of these in one statement.

While the strong Positivstellensatz results move us away from refuting systems and towards certifying bounds, they are still results about "rational SOS" involving denominators (s_1 above), which we have seen present some challenges and ambiguities when implementing algorithms. Fortunately, there is also a branch of Positivstellensatz results about "denominator-free SOS," which will be the most relevant to our future pursuits.

The general idea of these results to keep in mind is that the price of denominator-free SOS certificates is the assumption of *compactness* of the underlying semialgebraic set.

Theorem 2.21 (Schmüdgen Positivstellensatz [Sch91])**.** *Let $p_1, \ldots, p_m \in \mathbb{R}[x_1, \ldots, x_n]$ and $\mathcal{K}$ and S be as in Corollary 2.20. Suppose further that $\mathcal{K}$ is compact. If $r \in \mathbb{R}[x_1, \ldots, x_n]$ has $r(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{K}$, then $r \in S$. That is, there exist $q_S \in \mathsf{SOS}$ for each $S \subseteq [m]$ such that*

$$r(\boldsymbol{x}) = \sum_{S \subseteq [m]} q_S(\boldsymbol{x}) \prod_{i \in S} p_i(\boldsymbol{x}). \tag{2.27}$$

It is worth noting that the compactness assumption is very often satisfied in practice, especially for applications in computer science and combinatorial optimization. That is because, in those cases, we are usually optimizing over x_i Boolean variables, typically encoded as $x_i^2 - 1 = 0$ (forcing $\boldsymbol{x} \in \{\pm 1\}^n$) or $x_i^2 - x_i = 0$ (forcing $\boldsymbol{x} \in \{0, 1\}^n$).

The remaining frustration for applications is the summation over $S \subseteq [m]$, having 2^m terms, in the description of our certificates. In the above Boolean setting, for instance, we will have at least one constraint per variable and thus $m \geq n$, so if we want to optimize over $q_S(\boldsymbol{x})$ as SDP variables, then we will have $\Omega(2^n)$ such variables, a prohibitively fast rate of growth. Fortunately, under a stronger "explicit compactness" assumption called the *Archimedean property* we may actually allow for a restriction to only those S with $|S| \leq 1$ in the certificate.

Theorem 2.22 (Putinar Positivstellensatz [Put93])**.** *Let $p_1, \ldots, p_m \in \mathbb{R}[x_1, \ldots, x_n]$, let $\mathcal{K}$ be as in Corollary 2.20, and define*

$$S^{(0)} := \left\{ q_0(\boldsymbol{x}) + \sum_{i=1}^{m} q_i(\boldsymbol{x}) p_i(\boldsymbol{x}) : q_0, \ldots, q_m \in \mathsf{SOS} \right\}. \tag{2.28}$$

Suppose that, for some $R > 0$, we have the Archimedean property

$$R - \sum_{i=1}^{n} x_i^2 \in S^{(0)}. \tag{2.29}$$

Then, if $r \in \mathbb{R}[x_1, \ldots, x_n]$ has $r(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{K}$, then $r \in S^{(0)}$. That is, there exist $q_0, \ldots, q_m \in \mathsf{SOS}$ such that

$$r(\boldsymbol{x}) = q_0(\boldsymbol{x}) + \sum_{i=1}^{m} q_i(\boldsymbol{x}) p_i(\boldsymbol{x}). \tag{2.30}$$

By the same token as compactness, the Archimedean property will also be satisfied automatically for many of the problems we look at later.

2.3.5 Positivstellensatz Effectivization and Proof Systems

Finally, let us again discuss how to make the abstract Positivstellensatz results "effective" in a way that is useful for algorithms. The ways to do this again come in two flavors, which exactly parallel the two approaches to Nullstellensatz effectivization from Section 2.2.1.

Cylindrical Algebraic Decomposition Developed by Collins in [Col75], *cylindrical algebraic decomposition (CAD)* is in a very rough sense a real-valued analog of Gröbner bases and Buchberger's algorithm. CAD computes, given a sequence of polynomials $p_1, \ldots, p_m \in \mathbb{R}[x_1, \ldots, x_n]$, a decomposition of $\mathbb{R}^n$ into "cells" $C_i \subseteq \mathbb{R}^n$ where for every i and $j \in [m]$, p_j is either strictly positive, strictly negative, or identically zero on C_i. The cells must also satisfy a certain consistency property with respect to projections to coordinate subspaces of $\mathbb{R}^n$, which, crucially, makes it tractable to answer whether a cell is empty or not. Thus using such a decomposition we can determine emptiness or non-emptiness of semialgebraic sets, just like with Positivstellensatz certificates. In fact, the algorithm implementing CAD can do much more, as it allows for *quantifier elimination*, allowing us to describe sets like

$$\{\boldsymbol{z} \in \mathbb{R}^n : \text{for all } \boldsymbol{x} \in \mathbb{R}^n, \text{ there exists } \boldsymbol{y} \in \mathbb{R}^n \text{ such that } F(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \geq 0\} \tag{2.31}$$

for any polynomial F. This is a subject unto itself related to the logical properties of the axioms of $\mathbb{R}$; see [BPR06] for a standard reference. Like Buchberger's algorithm, CAD can be useful but has very slow (doubly exponential in n) worst-case runtime. And, as in the complex-valued case, if one tries CAD and it grinds to a halt, it is useful to have a more flexible option that lets us trade proof system power for a faster algorithm.

Semidefinite Programming As with the Nullstellensatz, we may also consider constraining the degrees of all polynomials involved and trying to solve the resulting system of equations. It is no longer obvious how to do this, since even though we are left with finitely many degrees of freedom we also have a system involving constraints of the form "$q \in \mathsf{SOS}$," which does not boil down to merely a linear constraint on the coefficients of q. Not worrying about this yet, we start again by asking in general how large of a degree is necessary. The answer, alas, is even more disappointing than for the Nullstellensatz.

Theorem 2.23 (Effective Artin's theorem [LPR14])**.** *If $p \in \mathbb{R}[x_1, \ldots, x_n]$ has $p \geq 0$, then there exist $r, s \in \mathsf{SOS}$ with $p = r/s$ and $\deg r, \deg s \leq D$ for*

$$D = 2^{2^{2^{(\deg p)^{4^n}}}}. \tag{2.32}$$

The same work proves a similar, though slightly more complicated, bound for the polynomials appearing in a Krivine-Stengle weak Positivstellensatz certificate (in particular the bound is again a tower of five exponentials).

Open Problem 2.5 (Optimal effective Positivstellensatz)**.** *What is the optimal degree bound (in terms of the height of the tower of exponentials) of an effective Artin's theorem or Positivstellensatz? The best known lower bound is the singly-exponential $2^{\Omega(n)}$, while the best known upper bound, from Theorem 2.23, is a tower of five exponentials.*

This is an interesting but quite theoretical question—for any remotely practical purposes, we are again in the position of bounding the degrees by some small number and trying to solve the resulting system. Perhaps surprisingly, despite the challenging "$q \in \mathsf{SOS}$" constraints, it is often possible to do this efficiently. Historically, after the Nullstellensatz proof system had been intensely studied for some years, the Positivstellensatz proof system was proposed as an enhancement in [GV01] in the proof complexity literature without mention of algorithms. But, very soon thereafter, Parrilo and Lassere [Par00, Las01], both working in the quite separate optimization literature, concurrently realized that it is in fact possible to "automatize" the search for such proofs using SDP. We leave a detailed description of these developments to Chapter 4, since those SDPs will be our focus in the rest of the course.

Exercises

Exercise 2.1 (1969 Putnam Competition, Problem A1)**.** *Find a bivariate polynomial whose range is $(0, +\infty)$. Conclude that $p > 0$ does not imply $p \geq c$ for some $c > 0$.*[9] *Give an explicit weak real Nullstellensatz certificate for your polynomial.*

> Hint: *Two variables suffice, and you can choose your polynomial to be a sum of squares (though that does not constitute a real Nullstellensatz certificate!).*

Exercise 2.2. *Extend Proposition 2.10 to arbitrary, not necessarily homogeneous polynomials of degree 2.*

Exercise 2.3. *Show the following results illustrating that the nuance in Motzkin's "encoding" of the arithmetic mean–geometric mean inequality is necessary.*

1. *$f(x, y, z) = x^3 + y^3 + z^3 - 3xyz$ is not non-negative on $\mathbb{R}^3$.*
2. *$f(w, x, y, z) = w^4 + x^4 + y^4 + z^4 - 4wxyz$ is non-negative on $\mathbb{R}^4$, but is also SOS.*

[9] I learned of this exercise from this MathOverflow question. Only look if you want the answer!

Exercise 2.4 (Better than Motzkin [Sch12]). *Let $f(x,y) = x^4y^2 + x^2y^4 + 1 - x^2y^2$. Show that $f(x,y) \geq \frac{26}{27} > 0$ for all $(x,y) \in \mathbb{R}^2$, but, mimicking the proof of Theorem 2.12, show that f is not SOS.*

Exercise 2.5. *Let $p \in \mathbb{R}[x_1,\dots,x_n]$. Show that there exist $r_1,\dots,r_m,s_1,\dots,s_m \in \mathbb{R}[x_1,\dots,x_n]$ such that $p = \sum_{i=1}^m (r_i/s_i)^2$ if and only if there exist $R,S \in \mathsf{SOS}$ such that $p = R/S$.*

Exercise 2.6. *Find $q_i(\boldsymbol{x})$ for $i \in [3]$ and $\boldsymbol{x} = (x_1,x_2,x_3)$ such that*

$$x_1x_2 + x_2x_3 + x_1x_3 + 1 \in q_1(\boldsymbol{x})(x_1^2-1) + q_2(\boldsymbol{x})(x_2^2-1) + q_3(\boldsymbol{x})(x_3^2-1) + \mathsf{SOS}. \quad (2.33)$$

Recall that these correspond to inequalities we included in the metric LP relaxation of MaxCut*, but which were omitted from the Goemans-Williamson relaxation. This problem shows that a sufficiently high degree SOS relaxation (degree 4 will suffice) does automatically satisfy these inequalities.*

Exercise 2.7 (Schmüdgen but not Putinar). *Devise constraints $p_1,\dots,p_m$ so that the semialgebraic set $\mathcal{K}$ is compact, but the constraints do not satisfy the Archimedean property.*

> HINT: *Work backwards from knowing that Putinar's Positivstellensatz must fail to hold for such constraints. A very small example should suffice!*

Exercise 2.8. *Show that $1 + x_1^2 + \cdots + x_n^2$ is not a sum of n squares in $\mathbb{R}(x_1,\dots,x_n)$, proving the lower bound cited in Open Problem 2.2.*

> HINT: *Start by substituting $x_i = y_i/y_{n+1}$ and multiply by y_{n+1}^2 to homogenize. Then, clear the denominator in a rational SOS expression and consider terms on either side.*

Exercise 2.9 (Certificates by hand). *Suppose that $p \in \mathbb{R}[x_1,\dots,x_n]$ satisfies that $p(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \{\pm 1\}^n$. Without using Putinar's Positivstellensatz, show how, from p, you could compute $q_1,\dots,q_n,s_1,\dots,s_m \in \mathbb{R}[x_1,\dots,x_n]$ so that $p(\boldsymbol{x}) = \sum_{i=1}^n q_i(\boldsymbol{x})(x_i^2-1) + \sum_{i=1}^m s_i(\boldsymbol{x})^2$ for all $\boldsymbol{x} \in \{\pm 1\}^n$ (that is, show that such exist and describe how you would actually construct them).*

> HINT: *At least for the most straightforward solution, you will need to know a few facts about "Boolean Fourier analysis." At this level, it is just the observation that every function on the hypercube agrees with a unique multilinear polynomial. See Chapter 1 of [O'D14] if this is not familiar.*

NOTES

OTHER SOURCES The field we have briskly surveyed, often called *real algebraic geometry*, is vast, has deep historical roots, and has very many written treatments from different perspectives. Some directly relevant book-length treatments include [Mar08, PD13, Pow21], the second being remarkable for its careful treatment of history and bibliography. The survey [Pow11a] is also useful. The survey [Tau70], while older, gives a broad overview of the role of all kinds of sum-of-squares expressions throughout mathematics (including Pfister's results and the connection to normed division algebras).

LDE Proofs Linear Diophantine equations are straightforward to solve using *Euclid's algorithm* and its extensions, which amounts to repeatedly performing division with remainder. Both the outputs and all intermediate data are also integers not exceeding the size of the input, so the algorithm is tractable and simple to analyze. The worst-case number of *iterations* in Euclid's algorithm is, in a delightful twist, achieved by the Fibonacci numbers and related to the continued fraction expansion of the golden ratio, as discussed in, e.g., [HW79].

History and Hilbert's Motivations The reader might notice that the chronology of results does not coincide with the conceptual progression we are suggesting: in reality, Hilbert's early results on sums of polynomial squares precede even the Nullstellensatz. Hilbert seems to have had several simultaneous motivations. As detailed in [PD13], the sum of polynomial squares results (Theorem 2.11) were motivated by a problem posed in Minkowski's dissertation and thesis defense, for which Hilbert was an examiner. But the 17th problem—with its sums of rational squares—seems to have come up not as a generalization of these results but rather in a quite different question of what lengths can be constructed geometrically using a straightedge and, instead of a compass, a "scale" or a marked unit interval on the straightedge (sometimes also called a "gauge"). The Nullstellensatz belongs to another strand of work concerning the structure of ideals in commutative rings. The synthesis of these results as all being "about algebraic proofs" implicit here and in much recent discussion of these topics is therefore probably more of a modern perspective.

Concrete Non-Negative Non-SOS Polynomials After Motzkin's example, many further non-negative polynomials that are not SOS appeared in the literature. One nice example, given in [LL78], is the polynomial

$$\sum_{i=1}^{5} \prod_{j \neq i} (x_i - x_j). \tag{2.34}$$

Remarkably, only for exactly 5 variables is this polynomial both non-negative and not SOS! See [Rez00] for a history of examples, Chapter 3 of [Pow21] for the "Gram matrix method" sometimes used to show that polynomials are not SOS, and also [Rez07] for a deeper reconsideration of Hilbert's method for Theorem 2.11 yielding many more examples.

Rational Coefficients The question of Positivstellensatz certificates having rational coefficients for polynomial systems also having rational coefficients is natural to ask, especially if one is interested in using SOS and Positivstellensatz refutation as a computational proof assistant, in which case it is important to be able to verify certificates symbolically rather than just numerically. Artin's result, part of which we cite in Theorem 2.13, actually applies to fields other than $\mathbb{R}$, and in particular proves the same statement with $\mathbb{R}$ replaced by $\mathbb{Q}$ throughout. On the other hand, without denominators, [Sch16] showed that there are polynomials with coefficients in $\mathbb{Q}$ that are SOS over $\mathbb{R}$ but not SOS over $\mathbb{Q}$. For the Positivstellensatz, [Pow11b] developed analogs of Schmüdgen's and Putinar's versions over $\mathbb{Q}$. That work raises the following intriguing question about the applicability of Putinar's Positivstellensatz in the real and rational cases.

Open Problem 2.6 (Real vs. rational Archimedean property [Pow11b]). *Let* $p_1, \ldots, p_m \in \mathbb{Q}[x_1, \ldots, x_n]$*, let* $\mathcal{K}$ *be as in Corollary 2.20, and define*

$$S_{\mathbb{F}}^{(0)} := \left\{ q_0(\boldsymbol{x}) + \sum_{i=1}^{m} q_i(\boldsymbol{x}) p_i(\boldsymbol{x}) : q_0, \ldots, q_m \in \mathbb{F}[x_1, \ldots, x_n] \cap \mathsf{SOS} \right\} \tag{2.35}$$

for $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}\}$*. If for some* $R \in \mathbb{R}$ *with* $R > 0$ *we have*

$$R - \sum_{i=1}^{n} x_i^2 \in S_{\mathbb{R}}^{(0)}, \tag{2.36}$$

then do we necessarily have for some $R' \in \mathbb{Q}$ *with* $R' > 0$ *that*

$$R' - \sum_{i=1}^{n} x_i^2 \in S_{\mathbb{Q}}^{(0)} \ ? \tag{2.37}$$

Finally, computational aspects of finding $\mathbb{Q}$-valued SOS proofs with and without denominators are treated in [PP08, KLYZ12]. Chapter 11 of [Pow21] is useful for an overview of all of these matters.

MORE ON PROOF COMPLEXITY Measuring the difficulty of proving various propositions in various proof systems is the main question of the field of *proof complexity*. We have seen two proof systems: Nullstellensatz and Positivstellensatz proofs. Both belong to the subclass of *algebraic* proof systems, where we use polynomial manipulations to prove statements. Other proof systems that work this way include the Sherali-Adams, Lovász-Schrijver, Cutting Planes, and Ideal Proof systems. A further class of *logical* proof systems work instead over logical formulas and manipulate them using various permissible Boolean manipulations. These include the Resolution and Frege proof systems.

There is also a distinction between *static* and *dynamic* proof systems. In a static system we write a proof as a single derivation that combines initial axioms, e.g. as a single polynomial sentence in the Nullstellensatz or Positivstellensatz. In a dynamic system, more like writing proofs in real life, we write a proof sequentially and can accumulate a "supply" of statements we have proved already (e.g. polynomials that must be zero or non-negative) that we can then reuse without rewriting their proofs every time. This lets dynamic proofs be shorter than equivalent static proofs in some situations. The dynamic versions of Nullstellensatz and Positivstellensatz proofs are called *Polynomial Calculus* and *Positivstellensatz Calculus*, respectively.

Finally, the impetus behind much of the work in proof complexity from computer scientists is the seminal work [CR79], which described a general class of sound and complete *propositional proof systems* for Boolean formulas. If any such system had polynomially short proofs of all propositional tautologies (with proof "length" having a suitable technical meaning) it would follow that $\mathsf{NP} = \mathsf{coNP}$, an equivalence of complexity classes generally believed to be false. Thus, proving super-polynomial lower bounds against different, increasingly powerful proof systems gives evidence for $\mathsf{NP} \neq \mathsf{coNP}$, and this is a useful approach since this non-equality itself is believed to be a hard problem on par with $\mathsf{P} \neq \mathsf{NP}$.

We present this laundry list just to give an idea of the great variety of ideas present in this field; see the beginning of [FKP19] for much more discussion and bibliography.

3 | Moment Problems

To come.

4 | Lasserre-Parrilo Semidefinite Programming Relaxations

We now finally discuss how to implement the search for SOS proofs with concrete convex optimization algorithms. We will also see a dual formulation of hese programs that generalizes the original Goemans-Williamson SDP from Chapter 1 in its "probabilistic interpretation."

For the sake of simplicity, we will limit our discussion to the kinds of SOS proofs handled by Putinar's Positivstellensatz—this is not the only choice or even necessarily the best one, but other variants can be handled with straightforward adjustments of what we present, and this choice is most typical of the computer science literature. We also focus on certificates of bounds on optimization problems rather than refutations of systems of polynomial inequalities, again because this will be most relevant in applications. Thus: concretely, we will study bounds on an optimization problem of the form

$$\mathsf{Opt} := \left\{ \begin{array}{ll} \text{maximize} & p(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n, \\ & f_i(\boldsymbol{x}) = 0 \text{ for all } i \in [a], \\ & g_j(\boldsymbol{x}) \geq 0 \text{ for all } j \in [b], \end{array} \right\} \tag{4.1}$$

where $p, f_i, g_j \in \mathbb{R}[x_1, \ldots, x_n]$. As we have seen, it is not necessary to distinguish equality and inequality constraints, but we include this distinction because often the problems we look at later will only have equality constraints, so it will be convenient to describe them explicitly.

4.1 Parrilo Proof Relaxation

The first relaxation we consider, written in algebraic language, is the following *degree D sum-of-squares relaxation*:

$$\mathsf{Parr}_D := \left\{ \begin{array}{ll} \text{minimize} & c \\ \text{subject to} & c - p(\boldsymbol{x}) \overset{(\mathrm{p})}{=} \sum_{i=1}^{a} f_i(\boldsymbol{x}) q_i(\boldsymbol{x}) + r_0(\boldsymbol{x}) + \sum_{j=1}^{b} g_j(\boldsymbol{x}) r_j(\boldsymbol{x}), \\ & q_i \in \mathbb{R}[x_1, \ldots, x_n], \\ & \deg f_i q_i \leq D, \\ & r_j \in \mathsf{SOS} \subset \mathbb{R}[x_1, \ldots, x_n], \\ & \deg g_j r_j \leq D \end{array} \right\}. \tag{4.2}$$

That such a program could be written as an SDP was partly recognized in the early publications [Sho87, Nes00], but was fully developed by Parrilo in his dissertation and early publications [Par00, Par03], and is usually attributed to him.

Writing this as an SDP is not difficult, but rather notationally heavy. However, we will be quite explicit about this rewriting to make sure the reader is convinced that one can, with some work but without confusion, write a program translating a problem specified with polynomials into input that can be taken by a convex optimization solver (see Exercise 4.3 for a challenge to this effect).

Definition 4.1 (Multisets)**.** *We write $\left(\!\binom{S}{k}\!\right)$ for the multisets of k elements from the set S. These may be identified with those $\boldsymbol{a} \in \mathbb{N}^S$ with $\sum a_i = k$. We also write $\left(\!\binom{S}{\leq k}\!\right)$ for the multisets of between 0 and k elements from S. Likewise, we write $\left(\!\binom{n}{k}\!\right) := |\left(\!\binom{[n]}{k}\!\right)|$ and $\left(\!\binom{n}{\leq k}\!\right) := |\left(\!\binom{[n]}{\leq k}\!\right)|$. Finally, we call the* lexicographical ordering *on multisets of $\mathbb{N}$ the ordering that first orders multisets by size, and then lexicographically among multisets of a fixed size.*

Definition 4.2 (Monomial vector)**.** *For $\boldsymbol{x} \in \mathbb{R}^n$, write $\boldsymbol{x}^{\otimes \leq d}$ for the vector of $\boldsymbol{x}^S$ over all $S \in \left(\!\binom{[n]}{\leq d}\!\right)$, taken in lexicographical order.*

After a suitable reordering, this is just the concatenation of certain subsets of elements of $\boldsymbol{x}^{\otimes 0}, \boldsymbol{x}^{\otimes 1}, \ldots, \boldsymbol{x}^{\otimes d}$—hence the notation. For example, if $n = 2$, then

$$\boldsymbol{x}^{\otimes \leq 3} = \begin{bmatrix} 1 \\ \hline x_1 \\ x_2 \\ \hline x_1^2 \\ x_1 x_2 \\ x_2^2 \\ \hline x_1^3 \\ x_1^2 x_2 \\ x_1 x_2^2 \\ x_2^3 \end{bmatrix}, \tag{4.3}$$

where the lines separate the terms of each degree (0, 1, 2, 3).

Definition 4.3 (Coefficient vector)**.** *For a polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ with $\deg p \leq D$, let $\boldsymbol{v}^{(p,D)} \in \mathbb{R}^{\left(\!\binom{[n]}{\leq D}\!\right)}$ be the vector of coefficients of p with indices in lexicographical order.*

The coefficient vector is also the unique vector satisfying the property

$$p(\boldsymbol{x}) \overset{(\mathrm{p})}{=} \langle \boldsymbol{v}^{(p,D)}, \boldsymbol{x}^{\otimes \leq D} \rangle. \tag{4.4}$$

With this tool in hand, we begin to start expressing polynomial operations in terms of matrix and vector operations. We do not give proofs of results like the below; you may verify them as an exercise if you are not convinced.

Proposition 4.4 (Polynomial multiplication)**.** *For a polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ and $D \geq \deg p$, there is a matrix $\boldsymbol{M}^{(p,D)}$ such that, for any $q \in \mathbb{R}[x_1, \ldots, x_n]$ with $\deg q \leq D - \deg p$, $\boldsymbol{M}^{(p,D)} \boldsymbol{v}^{(q, D - \deg p)} = \boldsymbol{v}^{(pq, D)}$.*

Using this device, the polynomial equation

$$c - p(\boldsymbol{x}) = \sum_{i=1}^{a} f_i(\boldsymbol{x}) q_i(\boldsymbol{x}) + r_0(\boldsymbol{x}) + \sum_{j=1}^{b} g_j(\boldsymbol{x}) r_j(\boldsymbol{x}) \tag{4.5}$$

is equivalent to the vector equation

$$\boldsymbol{v}^{(c,D)} - \boldsymbol{v}^{(p,D)} = \sum_{i=1}^{a} \boldsymbol{M}^{(f_i,D)} \boldsymbol{v}^{(q_i, D-\deg f_i)} + \boldsymbol{v}^{(r_0,D)} + \sum_{j=1}^{b} \boldsymbol{M}^{(g_j,D)} \boldsymbol{v}^{(r_j, D-\deg g_j)}. \tag{4.6}$$

(Note that $\boldsymbol{v}^{(c)}$ views c as a constant polynomial, so we just have $\boldsymbol{v}^{(c)} = c\boldsymbol{e}_{\varnothing}$.)

We now turn to the issue of handling the property "belongs to SOS," which is more complicated. First, besides vectors of coefficients, we also introduce *matrices* representing polynomials.

Definition 4.5. *Let $D \geq 2$ be even and $p \in \mathbb{R}[x_1, \ldots, x_n]$ with $\deg p \leq D$. We say that $\boldsymbol{S} \in \mathbb{R}^{\binom{[n]}{\leq D/2} \times \binom{[n]}{\leq D/2}}$* represents *$p$ if*

$$p(\boldsymbol{x}) \overset{(\mathrm{p})}{=} \boldsymbol{x}^{\otimes \leq D/2 \top} \boldsymbol{S} \boldsymbol{x}^{\otimes \leq D/2}. \tag{4.7}$$

The following is the key relationship between polynomials belonging to SOS and matrices being psd that drives the SDP formulation of SOS.

Proposition 4.6. *$p \in \mathsf{SOS}$ if and only if there exists a $\boldsymbol{S}$ representing p such that $\boldsymbol{S} \succeq \boldsymbol{0}$.*[1]

Proof. Suppose $\deg p \leq D$ where D is even. Suppose first that $p \in \mathsf{SOS}$. Write $p(\boldsymbol{x}) = \sum_{i=1}^{m} s_i(\boldsymbol{x})^2$. Then, $\deg s_i \leq D/2$ for each i. We have

$$\begin{aligned}
p(\boldsymbol{x}) &= \sum_{i=1}^{m} s_i(\boldsymbol{x})^2 \\
&= \sum_{i=1}^{m} \langle \boldsymbol{v}^{(s_i, D/2)}, \boldsymbol{x}^{\otimes \leq D/2} \rangle^2 && \text{(by (4.4))} \\
&= \sum_{i=1}^{m} \boldsymbol{x}^{\otimes D/2 \top} \left(\boldsymbol{v}^{(s_i, D/2)} \boldsymbol{v}^{(s_i, D/2)\top} \right) \boldsymbol{x}^{\otimes \leq D/2} \\
&= \boldsymbol{x}^{\otimes D/2 \top} \left(\sum_{i=1}^{m} \boldsymbol{v}^{(s_i, D/2)} \boldsymbol{v}^{(s_i, D/2)\top} \right) \boldsymbol{x}^{\otimes D/2 \top}, && (4.8)
\end{aligned}$$

and the matrix in parentheses is a psd matrix representing p.

For the converse, if $\boldsymbol{S} \succeq \boldsymbol{0}$ represents p, then expanding $\boldsymbol{S} = \sum_{i=1}^{m} \boldsymbol{v}_i \boldsymbol{v}_i^\top$ and reversing the steps in the manipulation above shows that $p \in \mathsf{SOS}$. □

Finally, we will need the following result allowing us to translate between a matrix representing a polynomial and the coefficient vector of the polynomial.

[1] Recall that $\boldsymbol{S} \succeq \boldsymbol{0}$ means that $\boldsymbol{S}$ is also necessarily symmetric.

Definition 4.7. *For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, write $\mathsf{vec}(\boldsymbol{A}) \in \mathbb{R}^{mn}$ for the vector formed by concatenating the columns of $\boldsymbol{A}$.*

Proposition 4.8. *Suppose $D \geq 2$ is even. Then, there exists a matrix $\boldsymbol{V}^{(D)}$ such that, whenever $\boldsymbol{S} \in \mathbb{R}^{\binom{[n]}{\leq D/2}\times\binom{[n]}{\leq D/2}}$ represents $p(\boldsymbol{x}) \in \mathbb{R}[x_1,\ldots,x_n]$, then $\boldsymbol{V}^{(D)}\mathsf{vec}(\boldsymbol{S}) = \boldsymbol{v}^{(p,D)}$.*

Combining all these tools, we reach the following semidefinite programming formulation of Parr_D. The proof should be clear as a combination of the above Propositions.

Theorem 4.9 (Parrilo proof relaxation SDP)**.** *For any $p, f_1,\ldots,f_a,g_1,\ldots,g_b \in \mathbb{R}[x_1,\ldots,x_n]$ and $D \geq 2$ even with $\deg p, \deg f_i, \deg g_j \leq D$,*

$$\mathsf{Parr}_D = \left\{ \begin{array}{ll} \textit{minimize} & c \\ \textit{subject to} & c\boldsymbol{e}_\varnothing - \boldsymbol{v}^{(p,D)} = \sum_{i=1}^{a} \boldsymbol{M}^{(f_i,D)}\boldsymbol{v}^{(q_i,D-\deg f_i)} \\ & \qquad\qquad + \boldsymbol{V}^{(D)}\mathsf{vec}(\boldsymbol{R}_0) \\ & \qquad\qquad + \sum_{j=1}^{b} \boldsymbol{M}^{(g_j,D)}\boldsymbol{V}^{(D-\deg g_j)}\mathsf{vec}(\boldsymbol{R}_j), \\ & \boldsymbol{v}^{(q_i,D-\deg f_i)} \in \mathbb{R}^{\binom{[n]}{\leq D-\deg f_i}} \textit{ for each } i \in [a], \\ & \boldsymbol{R}_0 \in \mathbb{R}^{\binom{[n]}{\leq D/2}\times\binom{[n]}{\leq D/2}}, \\ & \boldsymbol{R}_j \in \mathbb{R}^{\binom{[n]}{\leq \lfloor (D-\deg g_j)/2\rfloor}\times\binom{[n]}{\leq \lfloor (D-\deg g_j)/2\rfloor}} \textit{ for each } j \in [b] \\ & \boldsymbol{R}_0,\ldots,\boldsymbol{R}_b \succeq \boldsymbol{0} \end{array} \right\}. \tag{4.9}$$

Let us point out a few features of this SDP that are especially salient for computational purposes. Below we think of n large and D fixed when we use $O(\cdot)$ notation (see Exercise 4.1 for justification of the bounds).

First, each inequality constraint $g_i(\boldsymbol{x}) \geq 0$ "costs" us one psd variable of dimension $n^{O(D)}$, while each equality constraint $f_i(\boldsymbol{x}) = 0$ only costs one vector variable of dimension $n^{O(D)}$. Thus inequality constraints are especially costly, and it is important to treat equality constraints separately if you are worried about runtime.

Second, the total number of linear constraints is $n^{O(D)}$. Linear constraints are again typically expensive for SDP solvers in practice. For this reason many of the most direct optimizations of solving SOS programs involves taking advantage of any redundancy or sparsity in the system of linear constraints that must be enforced.

Finally, these practicalities aside, the SDP in (4.9) involves $(a+b)n^{O(D)}$ scalar variables and $n^{O(D)}$ linear constraints, so, for D constant and a, b polynomial in n, we would expect to be able to solve the SDP in polynomial time in n. However, even this very coarse claim is not obvious and sometimes is simply false—see Section 4.5 for more discussion.

4.2 Lasserre Pseudomoment Relaxation

We saw back in Chapter 1 that SDP duality was an important tool for fully understanding all the possible interpretations of the Goemans-Williamson SDP. Thus it is natural to consider the dual of the SDP we have formed in Theorem 4.9. Actually, this admits an entirely separate general interpretation of SOS that we have not yet encountered, so, before writing out the dual SDP formulation, let us give this more "stylized" description. This will be a

generalization to arbitrary polynomial optimization problems of the "probabilistic interpretation" of the Goemans-Williamson SDP from Chapter 1. This formulation is due to Lasserre [Las01], developed concurrently with Parrilo's. Laurent's survey [Lau09] is also very useful for an overview of this perspective.

4.2.1 Stylized Description

The key object behind the Lasserre description of SOS is the following, which, as we will see later, is in some sense the correct "dual object" to an SOS proof. As in, for example, the duality between functions and measures in classical analysis, it is then not surprising that this dual object is something that "takes in" a polynomial and produces a number.

Definition 4.10 (Pseudoexpectation)**.** *A function* $\widetilde{\mathbb{E}} : \mathbb{R}[x_1, \ldots, x_n]_{\leq D} \to \mathbb{R}$ *is called a* degree D pseudoexpectation *for the family of constraints* $\{f_i(\boldsymbol{x}) = 0\}_{i=1}^{a}, \{g_j(\boldsymbol{x}) \geq 0\}_{j=1}^{b}$ *if the following properties hold:*

1. $\widetilde{\mathbb{E}}$ *is linear.*
2. $\widetilde{\mathbb{E}}[1] = 1$.
3. $\widetilde{\mathbb{E}}[f_i(\boldsymbol{x})q(\boldsymbol{x})] = 0$ *for all* $i \in [a]$ *and* q *with* $\deg f_i q \leq D$.

4A. $\widetilde{\mathbb{E}}[s(\boldsymbol{x})^2] \geq 0$ *whenever* $\deg s^2 \leq D$.

4B. $\widetilde{\mathbb{E}}[g_j(\boldsymbol{x})s(\boldsymbol{x})^2] \geq 0$ *for all* $j \in [b]$ *and* s *with* $\deg g_j s^2 \leq D$.

The *degree D Lasserre relaxation* of our polynomial optimization problem is then

$$\mathsf{Lass}_D := \left\{ \begin{array}{ll} \text{maximize} & \widetilde{\mathbb{E}}[p(\boldsymbol{x})] \\ \text{subject to} & \widetilde{\mathbb{E}} \text{ is a degree } D \text{ pseudoexpectation} \end{array} \right\}. \tag{4.10}$$

This looks simpler than the Parrilo relaxation, but this is of course only a superficial consequence of our having moved the complicated constraints into the definition of a pseudoexpectation.

The idea behind this relaxation is precisely along the lines of the probabilistic interpretation of the Goemans-Williamson SDP. Let us set

$$\mathcal{K} := \{\boldsymbol{x} \in \mathbb{R}^n : f_i(\boldsymbol{x}) = 0 \text{ for all } i \in [a], g_j(\boldsymbol{x}) \geq 0 \text{ for all } j \in b\}, \tag{4.11}$$

which is the semialgebraic set defined by our constraints. Then, letting $\mathcal{M}(\mathcal{K})$ denote the set of probability measures supported on $\mathcal{K}$, we may convexify our initial problem by optimizing over these measures instead of points in $\mathcal{K}$:

$$\left\{ \begin{array}{ll} \text{maximize} & p(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{K} \end{array} \right\} = \left\{ \begin{array}{ll} \text{maximize} & \mathbb{E}_\mu[p(\boldsymbol{x})] \\ \text{subject to} & \mu \in \mathcal{M}(\mathcal{K}) \end{array} \right\}, \tag{4.12}$$

where $\mathbb{E}_\mu$ is the expectation operator with respect to a measure μ. A pseudoexpectation with respect to the constraints defining $\mathcal{K}$ is then just a relaxation of the convex set $\{\mathbb{E}_\mu : \mu \in \mathcal{M}(\mathcal{K})\}$: clearly the Conditions 1, 2, 3, 4A, and 4B are satisfied by any such $\mathbb{E}_\mu$. Indeed, any such $\mathbb{E}_\mu$ would satisfy these conditions for arbitrary functions q_i and s_j; in contrast, a pseudoexpectation only needs to satisfy these conditions for evaluations on low-degree polynomials.

4.2.2 Semidefinite Program Implementation

Let us now see how to implement the Lasserre relaxation as an SDP, as we did with the Parrilo relaxation. First, we observe that, because of the linearity property (Condition 1), a pseudoexpectation is specified by its *pseudomoment sequence*. We view this as being encoded in a vector $\boldsymbol{m} \in \mathbb{R}^{\binom{[n]}{\leq D}}$, with

$$m_S = \widetilde{\mathbb{E}}[\boldsymbol{x}^S]. \tag{4.13}$$

In particular, for $p(\boldsymbol{x}) \in \mathbb{R}[x_1, \ldots, x_n]_{\leq D}$, we may expand $p(\boldsymbol{x}) = \sum_{S \in \binom{[n]}{\leq D}} v_S^{(p,D)} \boldsymbol{x}^S$ by the definition of the coefficient vector. Thus, pseudoexpectations may be evaluated as

$$\begin{aligned}
\widetilde{\mathbb{E}}[p(\boldsymbol{x})] &= \widetilde{\mathbb{E}}\left[\sum_{S \in \binom{[n]}{\leq D}} v_S^{(p,D)} \boldsymbol{x}^S\right] \\
&= \sum_{S \in \binom{[n]}{\leq D}} v_S^{(p,D)} \widetilde{\mathbb{E}}[\boldsymbol{x}^S] && \text{(linearity of } \widetilde{\mathbb{E}}\text{)} \\
&= \sum_{S \in \binom{[n]}{\leq D}} v_S^{(p,D)} m_S \\
&= \langle \boldsymbol{v}^{(p,D)}, \boldsymbol{m} \rangle. && (4.14)
\end{aligned}$$

Using this, we may reformulate the remaining conditions in the definition of a pseudoexpectation in terms of $\boldsymbol{m}$. Condition 1 is the simplest, and amounts just to

$$\langle \boldsymbol{v}^{(1,D)}, \boldsymbol{m} \rangle = m_\varnothing = 1. \tag{4.15}$$

For Condition 2, it is convenient to use our description of polynomial multiplication through matrix multiplication from Proposition 4.4: since $\boldsymbol{v}^{(f_i q, D)} = \boldsymbol{M}^{(f_i, D)} \boldsymbol{v}^{(q, D - \deg f_i)}$, we may write the condition as

$$0 = \langle \boldsymbol{M}^{(f_i,D)} \boldsymbol{v}, \boldsymbol{m} \rangle = \langle \boldsymbol{v}, \boldsymbol{M}^{(f_i,D)^\top} \boldsymbol{m} \rangle \text{ for all } \boldsymbol{v} \in \mathbb{R}^{\binom{n}{D - \deg f_i}}. \tag{4.16}$$

But, having this hold for all $\boldsymbol{v}$ just amounts to the vector equation

$$\boldsymbol{M}^{(f_i,D)^\top} \boldsymbol{m} = \boldsymbol{0}. \tag{4.17}$$

As for the Parrilo SDP, it is the positivity constraints, Conditions 4A and 4B, that are trickier to implement. The main extra device we will need is a means of converting moment sequences into moment matrices, which we can constrain to be psd to enforce these conditions. Moment matrices constructed from the pseudomoment sequence are referred to as *pseudomoment matrices*.

Definition 4.11. *For $\boldsymbol{m} \in \mathbb{R}^{\binom{[n]}{\leq D}}$ for some D even, let $\mathsf{mmat}_D(\boldsymbol{m}) \in \mathbb{R}_{\mathsf{sym}}^{\binom{[n]}{\leq D/2} \times \binom{[n]}{\leq D/2}}$ have entries*

$$\mathsf{mmat}_D(\boldsymbol{m})_{ST} = m_{S+T}, \tag{4.18}$$

where $S + T$ denotes the union of multisets (keeping repetitions, so that $|S + T| = |S| + |T|$).

We will also need the following more precise description of the vectors appearing above in our treatment of Condition 2.

Proposition 4.12. *Suppose that $\boldsymbol{m}$ is the moment sequence of $\widetilde{\mathbb{E}}$. Then, $(\boldsymbol{M}^{(g,D)^\top}\boldsymbol{m})_S = \widetilde{\mathbb{E}}[g(\boldsymbol{x})\boldsymbol{x}^S]$ for all $S \in \mathbb{R}^{\left(\!\binom{[n]}{D-\deg g}\!\right)}$.*

Combining these two claims, it follows by the same reasoning as we have used before that Conditions 4A and 4B are, respectively, equivalent to having

$$\mathsf{mmat}_D(\boldsymbol{m}) \succeq \boldsymbol{0}, \tag{4.19}$$

$$\mathsf{mmat}_{2\lfloor (D-\deg g_j)/2 \rfloor}(\boldsymbol{M}^{(g_j,D)^\top}\boldsymbol{m}) \succeq \boldsymbol{0} \text{ for all } j \in [b], \tag{4.20}$$

where the expression $2\lfloor (D - \deg g_j)/2 \rfloor$ is just $D - \deg g_j$ if this quantity is even and $D - \deg g_j - 1$ if it is odd.

Combining all these tools, we reach the following semidefinite programming formulation of Lass_D. As for the Parrilo SDP, we do not give the proof, which just combines the previous observations.

Theorem 4.13 (Lasserre pseudomoment relaxation SDP)**.** *For any $p, f_1, \ldots, f_a, g_1, \ldots, g_b \in \mathbb{R}[x_1, \ldots, x_n]$ and $D \geq 2$ even with $\deg p, \deg f_i, \deg g_j \leq D$,*

$$\mathsf{Lass}_D = \left\{ \begin{array}{ll} \textit{maximize} & \langle \boldsymbol{v}^{(p,D)}, \boldsymbol{m} \rangle \\ \textit{subject to} & \boldsymbol{m} \in \mathbb{R}^{\left(\!\binom{[n]}{\leq D}\!\right)}, \\ & m_\varnothing = 1, \\ & \boldsymbol{M}^{(f_i,D)^\top}\boldsymbol{m} = \boldsymbol{0} \textit{ for all } i \in [a], \\ & \mathsf{mmat}_D(\boldsymbol{m}) \succeq \boldsymbol{0}, \\ & \mathsf{mmat}_{2\lfloor (D-\deg g_j)/2 \rfloor}(\boldsymbol{M}^{(g_j,D)^\top}\boldsymbol{m}) \succeq \boldsymbol{0} \end{array} \right\}. \tag{4.21}$$

4.2.3 Goemans-Williamson Redux

Before proceeding, since the Lasserre SDP is at this point entirely new to us, let us pause to collect some intuition. First, let us see how the Goemans-Williamson SDP is none other than the Lasserre SDP for the MaxCut problem with $D = 2$.

Recall that MaxCut, written as a polynomial problem, is

$$\begin{array}{ll} \text{maximize} & \boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} \\ \text{subject to} & x_i^2 - 1 = 0 \text{ for all } i \in [n], \end{array} \tag{4.22}$$

where $\boldsymbol{L} \in \mathbb{R}^{n\times n}_{\mathsf{sym}}$ is the graph Laplacian but for our purposes now can be any symmetric matrix. We formulate the degree 2 Lasserre relaxation below.

Our decision variable will be indexed by $\left(\!\binom{[n]}{\leq 2}\!\right)$, having entries $m_\varnothing = 1$ for degree 0, $m_{\{1\}}, \ldots, m_{\{n\}}$ for degree 1, and $m_{\{i,j\}}$ for degree 2, where $1 \leq i \leq j \leq n$ and we allow $i = j$, viewing the braces as defining a multiset. The linear constraints are easier to decode in terms of the pseudoexpectation operator: since the degree of $x_i^2 - 1$ is 2, the constraints just impose that

$$0 = \widetilde{\mathbb{E}}[x_i^2 - 1] = \widetilde{\mathbb{E}}[x_i^2] - 1 = m_{\{i,i\}} - 1, \tag{4.23}$$

or that $m_{\{i,i\}} = 1$, for all i.

There are no inequality constraints, so there will be just one psd constraint, with a pseudomoment matrix indexed by $\left(\!\binom{[n]}{\leq 1}\!\right)$, i.e., indexed by $\varnothing, \{1\}, \ldots, \{n\}$. This constraint is:

$$\mathbf{0} \preceq \left[\begin{array}{c|cccc} m_\varnothing & m_{\{1\}} & m_{\{2\}} & \cdots & m_{\{n\}} \\ \hline m_{\{1\}} & m_{\{1,1\}} & m_{\{1,2\}} & \cdots & m_{\{1,n\}} \\ m_{\{2\}} & m_{\{1,2\}} & m_{\{2,2\}} & \cdots & m_{\{2,n\}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{\{n\}} & m_{\{1,n\}} & m_{\{2,n\}} & \cdots & m_{\{n,n\}} \end{array}\right] \tag{4.24}$$

which, substituting in the linear constraints, we can rewrite as

$$= \left[\begin{array}{c|cccc} 1 & m_{\{1\}} & m_{\{2\}} & \cdots & m_{\{n\}} \\ \hline m_{\{1\}} & 1 & m_{\{1,2\}} & \cdots & m_{\{1,n\}} \\ m_{\{2\}} & m_{\{1,2\}} & 1 & \cdots & m_{\{2,n\}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{\{n\}} & m_{\{1,n\}} & m_{\{2,n\}} & \cdots & 1 \end{array}\right], \tag{4.25}$$

and we introduce a bit of notation,

$$=: \left[\begin{array}{c|c} 1 & \boldsymbol{y}^\top \\ \hline \boldsymbol{y} & \boldsymbol{Y} \end{array}\right], \tag{4.26}$$

where $\boldsymbol{y}$ is the vector of degree 1 pseudomoments and $\boldsymbol{Y}$ the matrix of degree 2 pseudomoments.

Finally, the objective function is

$$\widetilde{\mathbb{E}}[\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}] = \langle \boldsymbol{L}, \widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top]\rangle = \langle \boldsymbol{L}, \boldsymbol{Y}\rangle. \tag{4.27}$$

Thus, writing the Lasserre relaxation in these terms, we find something quite close to the Goemans-Williamson SDP:

$$L_2 = \left\{\begin{array}{ll} \text{maximize} & \langle \boldsymbol{L}, \boldsymbol{Y}\rangle \\ \text{subject to} & \boldsymbol{y} \in \mathbb{R}^n, \\ & \boldsymbol{Y} \in \mathbb{R}^{n\times n}_{\mathsf{sym}}, \\ & Y_{ii} = 1 \text{ for all } i \in [n], \\ & \left[\begin{array}{c|c} 1 & \boldsymbol{y}^\top \\ \hline \boldsymbol{y} & \boldsymbol{Y} \end{array}\right] \succeq \mathbf{0}. \end{array}\right\}. \tag{4.28}$$

To identify this with the original Goemans-Williamson SDP, it suffices to note that the psd constraint above implies $\boldsymbol{Y} \succeq \mathbf{0}$, and conversely if $\boldsymbol{Y} \succeq \mathbf{0}$ with $Y_{ii} = 1$, then we may set $\boldsymbol{y} = \mathbf{0}$ so that $\boldsymbol{Y}$ appears in the lower right block of a feasible point. Thus, the $\boldsymbol{y}$ variable is superfluous and we have

$$L_2 = \left\{\begin{array}{ll} \text{maximize} & \langle \boldsymbol{L}, \boldsymbol{Y}\rangle \\ \text{subject to} & \boldsymbol{Y} \in \mathbb{R}^{n\times n}_{\mathsf{sym}}, \\ & Y_{ii} = 1 \text{ for all } i \in [n], \\ & \boldsymbol{Y} \succeq \mathbf{0} \end{array}\right\} = \mathsf{SDP}(G), \tag{4.29}$$

recovering the Goemans-Williamson SDP.

4.3 Duality

One may check, as a rather tedious exercise, that the SDPs given in Theorems 4.9 and 4.13 are duals in the ordinary sense of SDP duality. On the other hand, the *weak duality* inequality between them is easy to check using the SOS proof and pseudoexpectation interpretations.

Theorem 4.14 (Weak duality)**.** *For any $D \geq 2$ even,* $\mathsf{Lass}_D \leq \mathsf{Parr}_D$.

Proof. Suppose $c, q_1(\boldsymbol{x}), \ldots, q_a(\boldsymbol{x}), r_1(\boldsymbol{x}), \ldots, r_b(\boldsymbol{x})$ form a feasible point for Parr_D, which we recall means that $r_j \in \mathsf{SOS}$ and we have

$$c - p(\boldsymbol{x}) \overset{(\mathrm{p})}{=} \sum_{i=1}^{a} f_i(\boldsymbol{x}) q_i(\boldsymbol{x}) + r_0(\boldsymbol{x}) + \sum_{j=1}^{b} g_j(\boldsymbol{x}) r_j(\boldsymbol{x}). \tag{4.30}$$

Suppose also that $\widetilde{\mathbb{E}}$ is a feasible point for Lass_D. Applying $\widetilde{\mathbb{E}}$ to either side of (4.30) and applying all the conditions in the definition of a pseudoexpectation, we have

$$c - \widetilde{\mathbb{E}}[p(\boldsymbol{x})] = \underbrace{\sum_{i=1}^{a} \widetilde{\mathbb{E}}[f_i(\boldsymbol{x}) q_i(\boldsymbol{x})]}_{=0} + \underbrace{\widetilde{\mathbb{E}}[r_0(\boldsymbol{x})] + \sum_{j=1}^{b} \widetilde{\mathbb{E}}[g_j(\boldsymbol{x}) r_j(\boldsymbol{x})]}_{\geq 0} \geq 0, \tag{4.31}$$

whereby $\widetilde{\mathbb{E}}[p(\boldsymbol{x})] \leq c$. □

There is are various straightforward condition for equality to hold between general dual SDPs, a property called *strong duality*, though this does not always hold for SDPs (unlike LPs). It is natural to ask how this plays out for the specific SDPs appearing in SOS. In fact, here the *Archimedean property* that we already saw as the main condition for Putinar's Positivstellensatz reappears.

Theorem 4.15 (Strong duality [JH16])**.** *Let $D \geq 2$ be even. Suppose that the system of constraints $\{f_i(\boldsymbol{x}) = 0\}_{i=1}^{a}, \{g_j(\boldsymbol{x}) \geq 0\}_{j=1}^{b}$ satisfies the* degree D Archimedean property*: there exists $R > 0$, $q_1, \ldots, q_a \in \mathbb{R}[x_1, \ldots, x_n]$, and $s_0, s_1, \ldots, s_b \in \mathsf{SOS}$ such that*

$$R - \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{a} f_i(\boldsymbol{x}) q_i(\boldsymbol{x}) + s_0(\boldsymbol{x}) + \sum_{j=1}^{b} g_j(\boldsymbol{x}) s_j(\boldsymbol{x}), \tag{4.32}$$

where each term on the right-hand side has degree at most D: $\deg f_i q_i, \deg s_0, \deg s_j g_j \leq D$. *Then,* $\mathsf{Lass}_D = \mathsf{Parr}_D$.

This is a useful general tool, though often generic results for SDP duality, in particular *Slater's condition*, clearly apply upon writing out concrete SDPs for an SOS program (the result of [JH16] applies in some cases where Slater's condition does not, however). In any case, when strong duality holds, we adopt the agnostic notation

$$\mathsf{SOS}_D := \mathsf{Lass}_D = \mathsf{Parr}_D. \tag{4.33}$$

4.4 Convergence

We note also that, combined with Putinar's Positivstellensatz, the Archimedean property implies the following further desirable property.

Theorem 4.16 (Convergence of SOS SDPs)**.** *Suppose that the system of constraints* $\{f_i(\boldsymbol{x}) = 0\}_{i=1}^{a}, \{g_j(\boldsymbol{x}) \geq 0\}_{j=1}^{b}$ *satisfies the Archimedean property. Then,*

$$\lim_{D\to\infty} \mathsf{SOS}_D = \mathsf{Opt} = \left\{ \begin{array}{ll} \textit{maximize} & p(\boldsymbol{x}) \\ \textit{subject to} & \boldsymbol{x} \in \mathbb{R}^n, \\ & f_i(\boldsymbol{x}) = 0 \textit{ for all } i \in [a], \\ & g_j(\boldsymbol{x}) \geq 0 \textit{ for all } j \in [b] \end{array} \right\}. \tag{4.34}$$

Proof. Let $\mathcal{K}$ denote the semialgebraic set satisfying the constraints,

$$\mathcal{K} := \{\boldsymbol{x} \in \mathbb{R}^n : f_i(\boldsymbol{x}) = 0 \text{ for all } i \in [a], g_j(\boldsymbol{x}) \geq 0 \text{ for all } j \in b\}. \tag{4.35}$$

Then, for any $\epsilon > 0$, $p(\boldsymbol{x}) < \mathsf{Opt} + \epsilon$ for all $\boldsymbol{x} \in \mathcal{K}$. By Putinar's Positivstellensatz, this statement admits an SOS proof, so for sufficiently large D we have $\mathsf{Opt} \leq \mathsf{Parr}_D \leq \mathsf{Opt} + \epsilon$. This holds for any $\epsilon > 0$, so the result follows. (We note also that by strong duality we have $\mathsf{Lass}_D = \mathsf{Parr}_D$ for all sufficiently large D.) □

An interesting further question is when there exists a finite D such that $\mathsf{SOS}_D = \mathsf{Opt}$, a property called *finite convergence*. In early literature on the Lasserre relaxations, it was noticed that this often held in numerical experiments, but for about 10 years no general theoretical justification of this was known [HL03, HL05]. Finally, Nie [Nie14] showed that, in general, so long as the Archimedean condition is satisfied, finite convergence holds for "almost all" collections of constraint polynomials f_i, g_j; specifically, it holds for coefficients of such polynomials avoiding a particular algebraic surface.

4.5 Tractability and O'Donnell's Caveat

To come.

Exercises

Exercise 4.1 (Multisets)**.** *Prove the identities*

$$\left(\!\!\binom{n}{k}\!\!\right) = \binom{n+k-1}{k} = n^{O(k)}, \tag{4.36}$$

$$\left(\!\!\binom{n}{\leq k}\!\!\right) = \binom{n+k}{k} = n^{O(k)}, \tag{4.37}$$

where the $O(\cdot)$ *bounds refer to* k *fixed while* $n \to \infty$*.*

Exercise 4.2 (Matrices representing polynomials)**.** *Let $D \geq 2$ be even and $p \in \mathbb{R}[x_1, \ldots, x_n]$ with $\deg p \leq D$. Let $S = \{\boldsymbol{S} \in \mathbb{R}^{\binom{[n]}{\leq D/2} \times \binom{[n]}{\leq D/2}} : \boldsymbol{S} \text{ represents } p\}$.*

1. *Show that S is an affine subspace of $\mathbb{R}^{\binom{[n]}{\leq D/2} \times \binom{[n]}{\leq D/2}}$.*
2. *What (in terms of quantities related to p) is the dimension of S?*

Let $S_{\mathsf{sym}} := S \cap \mathbb{R}^{\binom{[n]}{\leq D/2} \times \binom{[n]}{\leq D/2}}_{\mathsf{sym}}$.

3. *Show that $S_{\mathsf{sym}} \neq \emptyset$.*
4. *Show that S_{sym} is an affine subspace of $\mathbb{R}^{\binom{[n]}{\leq D/2} \times \binom{[n]}{\leq D/2}}_{\mathsf{sym}}$.*
5. *When (in terms of quantities related to p) is there only one matrix in S_{sym}?*
6. *Show that, whenever there is more than one matrix in S_{sym} and $p \in \mathsf{SOS}$, then there exists $S \in S_{\mathsf{sym}}$ such that $S \not\succeq \mathbf{0}$.*

Exercise 4.3. *If you know a programming language that has (1) a library for symbolic manipulation of polynomials and (2) an interface to an SDP solver (say, CVX and variants thereof), write a small "glue library" that takes an optimization problem specified with polynomials and a degree D and feeds the corresponding Parrilo or Lasserre SDP into the solver.*[2]

> NOTE: *If you experiment, you will find that a naive implementation of a general library of this kind is very slow to construct concrete SDPs (e.g., higher degree relaxations of* MaxCut*), much slower than constructing them by hand. This is because a naive implementation will not take advantage of various sparsity and redundancy structure in your constraints—making choices about how to handle this structure automatically and generally remains an important and largely open engineering problem.*

Exercise 4.4. *Suppose $\boldsymbol{P}$ represents $p(\boldsymbol{x})$ and $\boldsymbol{Q}$ represents $q(\boldsymbol{x})$ (in the sense of Definition 4.5). What operation of $\boldsymbol{P}$ and $\boldsymbol{Q}$ gives a matrix representing the product $p(\boldsymbol{x})q(\boldsymbol{x})$?*

Exercise 4.5 (Locally consistent probability distributions)**.** *Let $\widetilde{\mathbb{E}}$ be a degree D pseudoexpectation for the hypercube constraints, $\{x_i^2 - 1 = 0\}_{i=1}^n$. Show that, for any subset $S \subseteq [n]$ with $|S| \leq D/2$, there is a probability distribution μ over $\{\pm 1\}^S$ that agrees with $\widetilde{\mathbb{E}}$: setting $\boldsymbol{x}|_S$ to contain those indices of $\boldsymbol{x}$ that belong to S, for all polynomials $p(\boldsymbol{x}|_S)$ with $\deg p \leq D$,*

$$\widetilde{\mathbb{E}}[p(\boldsymbol{x}|_S)] = \mathbb{E}_{\boldsymbol{y} \sim \mu}[p(\boldsymbol{y})]. \tag{4.38}$$

> HINT: *Use the result of Exercise 2.9.*

Exercise 4.6 (Testing membership in SOS)**.** *Describe a feasibility SDP that determines whether, given $p \in \mathbb{R}[x_1, \ldots, x_n]$, there exist $s_1, \ldots, s_a \in \mathbb{R}[x_1, \ldots, x_n]$ such that $p(\boldsymbol{x}) = \sum_{j=1}^a s_j(\boldsymbol{x})^2$ and $\deg s_j(\boldsymbol{x})^2 \leq D$. Give a dual SDP and an intuitive description of the dual in terms of pseudoexpectations.*

[2] For example, in `Python` it is not too hard to do this using `sympy` for polynomials and `cvxpy` for solvers.

HINT: *It is possible to view this as a special case of the SDPs described in this chapter without having to rederive those formulations.*

A variant of this exercise actually leads to an intriguing direction that I do not believe has been studied before (though it is so natural that I might just be unaware of the right reference).

Open Problem 4.1 (Pseudomoment dual of rational SOS)**.** *Investigate the "pseudomoment-style" duals of feasibility semidefinite programs encoding rational SOS, asking if there exist* $r_1, \dots, r_a, s_1, \dots, s_b \in \mathbb{R}[x_1, \dots, x_n]$ *such that* $p(\boldsymbol{x}) = (\sum_{i=1}^{a} r_i(\boldsymbol{x})^2)/(\sum_{j=1}^{b} s_j(\boldsymbol{x})^2)$ *and having* $\deg r_i(\boldsymbol{x})^2, \deg s_j(\boldsymbol{x})^2 \leq D$*. Can you prove any interesting degree lower bounds by constructing such pseudomoments?*

NOTES

SOFTWARE As SOS developed to be an increasingly practical tool starting in the 2000s, various software packages appeared for formulating and solving SOS programs as SDPs more or less "automatically" from polynomial descriptions. Some packages that include such functionality are `YALMIP`, `SOSTOOLS`, `GloptiPoly` (all in `MATLAB`), and, more recently, `SumOfSquares.jl` (in `Julia`) and `SumsOfSquares.m2` (in `Macaulay2`, a more specialized computer algebra system). A helpful summary of how to use various packages for basic SOS tasks is given at this link.

CONVERGENCE RATES A recent line of work considers a more precise and practically relevant variant of finite convergence, asking with what rate SOS_D will converge to Opt for various constraint sets. For example, the results of [DW12, FF20] show that $\mathsf{Opt} - \mathsf{SOS}_D = O(n^2/D^2)$ when the only constraint is the spherical constraint $\sum_{i=1}^{n} x_i^2 = 1$ and when the maximum and minimum of the objective polynomial on the sphere are bounded. (In fact, Reznick's work on uniform denominators in [Rez95] used some similar ideas, as the latter paper discusses.) Even more recently, [SL22] study similar questions over the hypercube, in the regime $D \propto n$.

PUTINAR VS. SCHMÜDGEN In the presence of inequality constraints, different branches of the literature made different choices about which Positivstellensatz to base their SDP relaxations on: Parrilo's initial work used Schmüdgen's Positivstellensatz while Lasserre's used Putinar's Positivstellensatz. As we have noted, the Putinar version is more common in the computer science literature, though this too is not universal; there is a bit of discussion in, e.g., [OZ13]. It is known that Schmüdgen certificates of a given bound can have much lower degree than Putinar certificates; this is shown in, e.g., Table 1 of [Mag15]. However, I am not aware of especially clean examples illustrating this, and the details of these tradeoffs, especially at low degree, remain unclear.

Part II

Sum-of-Squares Algorithms

5 | The Proofs-to-Algorithms Framework

We now proceed to somewhat more advanced applications of SOS optimization. In particular, we will study several problems which are *not*, on the face of it, polynomial optimization problems, whereby it is not clear that SOS can be useful. However, we will see a general scheme by which we can take advantage of SOS tools to help us solve such problems, and—what is particularly convenient when working with SOS—to prove that the resulting algorithms work well.

The broad proof strategy that these problems share is sometimes called the *proofs to algorithms* approach (see, e.g., the survey [BS14]). This is a bit of a misnomer: what really happens is a conversion from purely mathematical proofs to *other proofs* that SOS algorithms work well. The following concise and accurate summary appears in [Hop18c] (which we will discuss later):

> "SOS SDPs in statistical settings are amenable to an analysis strategy which converts proofs of statistical identifiability into analysis of an SDP-based algorithm by phrasing the identifiability proof as a dual solution to the SDP."

To elaborate on this, generally speaking, we proceed by the following steps.

1. Formulate a polynomial optimization problem related to the task we are trying to perform. Generically, this will look like our usual

$$\mathsf{Opt} := \left\{ \begin{array}{ll} \text{maximize} & p(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n, \\ & f_i(\boldsymbol{x}) = 0 \text{ for all } i \in [a], \\ & g_j(\boldsymbol{x}) \geq 0 \text{ for all } j \in [b] \end{array} \right\}. \tag{5.1}$$

2. Prove that, if we could solve the polynomial optimization problem, then the optimal point $\boldsymbol{x}^\star$ or its value $p(\boldsymbol{x}^\star)$ would help us in our task.

3. *Convert* the above proof into a proof that if we can solve an SOS relaxation of the problem from Step 1, then the optimal pseudoexpectation $\widetilde{\mathbb{E}}^\star$ or its objective value $\widetilde{\mathbb{E}}^\star[p(\boldsymbol{x})]$ will *still* help us in our task.

A little more specifically, often there is some "hidden" information in the data we are given to work with that we seek to recover (especially in problems with a more statistical

flavor). We, the algorithm analysts, can measure how well $x^\star$ recovers this information by evaluating how large some polynomial $q(x^\star)$ is. However, in the coefficients of $q(x^\star)$ there is information that is not directly available to the algorithm itself, which is using p as a proxy for q.

In such a situation, Step 2 of the plan will be a proof that if $p(x^\star)$ is large and $x^\star$ is feasible for the problem Opt, then $q(x^\star)$ is also large. This deduction will look like:

$$\begin{gathered} p(x^\star) \text{ large}, \;\; f_i(x) = 0 \text{ for all } i \in [a], \;\; g_j(x) \geq 0 \text{ for all } j \in [b] \\ \Downarrow \\ q(x^\star) \text{ large}. \end{gathered} \tag{5.2}$$

The main point is that, if such a proof only uses fairly routine manipulations of polynomials or, as we will see below, even slightly more complicated algebraic operations on $x^\star$ and polynomials thereof, then we can show that $\widetilde{\mathbb{E}}^\star$ admits a parallel deduction:

$$\begin{gathered} \widetilde{\mathbb{E}}^\star[p(x)] \text{ large}, \;\; \widetilde{\mathbb{E}}^\star \text{ satisfies Lasserre constraints for Opt} \\ \Downarrow \\ \widetilde{\mathbb{E}}^\star[q(x)] \text{ large}. \end{gathered} \tag{5.3}$$

Once we have this kind of result, we can *round* $\widetilde{\mathbb{E}}^\star$ to extract an actual point $\hat{x}$ which is almost as good as $x^\star$ for our purposes. And, if (5.3) is satisfied by $\widetilde{\mathbb{E}}^\star$ of low degree, then $\hat{x}$ can be efficiently computed, giving us a useful algorithm.

There are two difficulties in executing this plan. The first is an exercise in reducing proof complexity: we want to encode the deduction of (5.2) in such a simple series of steps that those steps may be reproduced exactly in (5.3). Many useful analytic tools may be reformulated in this kind of "SOS-friendly" way; as the authors of [FH14], where such representations of several well-known inequalities are presented, write,

> "...if the inequality $f \geq 0$ is 'classical' and 'famous' enough, then f usually turns out to be representable as a sum of squares, although such a representation is not always easy to find."

The second, quite different, difficulty is the matter of how exactly to round to extract a useful point that looks enough like a sample from the "pseudodistribution" underlying a pseudoexpectation—concretely, a random point whose moments mimic the pseudomoments in some useful way. Below we present some general tools for these two tasks that will be useful in the sequel. In the remainder of this part of the course, we will examine several case studies of the proofs-to-algorithms framework.

5.1 Reasoning About Pseudoexpectations

In the three results below, we let $a(x), b(x) \in \mathbb{R}[x_1, \dots, x_n]^N_{\leq D}$. In words, these are vectors in N coordinates, where each coordinate is a polynomial of degree at most D.

Proposition 5.1 (SOS Cauchy-Schwarz inequality)**.** *If $\widetilde{\mathbb{E}}$ is a pseudoexpectation of degree at least $2D$, then*

$$\widetilde{\mathbb{E}}[\langle \boldsymbol{a}(\boldsymbol{x}), \boldsymbol{b}(\boldsymbol{x})\rangle] \leq \left(\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x})\|_2^2]\right)^{1/2} \left(\widetilde{\mathbb{E}}[\|\boldsymbol{b}(\boldsymbol{x})\|_2^2]\right)^{1/2}. \tag{5.4}$$

Proof. By dividing by the terms on the right-hand side, we may without loss of generality suppose that $\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x})\|_2^2] = \widetilde{\mathbb{E}}[\|\boldsymbol{b}(\boldsymbol{x})\|_2^2] = 1$. Under this assumption, we have

$$0 \leq \widetilde{\mathbb{E}}\left[\frac{1}{2}\|\boldsymbol{a}(\boldsymbol{x}) - \boldsymbol{b}(\boldsymbol{x})\|_2^2\right] = 1 - \widetilde{\mathbb{E}}[\langle \boldsymbol{a}(\boldsymbol{x}), \boldsymbol{b}(\boldsymbol{x})\rangle], \tag{5.5}$$

and the result follows after rearranging. □

Corollary 5.2 (SOS Jensen inequality)**.** *Let $p(\boldsymbol{x}) \in \mathbb{R}[x_1, \ldots, x_n]_{\leq D}$ and $\widetilde{\mathbb{E}}$ be a pseudoexpectation of degree at least $2D$. Then,*

$$\widetilde{\mathbb{E}}[p(\boldsymbol{x})^2] \geq (\widetilde{\mathbb{E}}[p(\boldsymbol{x})])^2. \tag{5.6}$$

Proof. Apply the SOS Cauchy-Schwarz inequality with $N = 1$ to the polynomials $a(\boldsymbol{x}) = p(\boldsymbol{x})$ and $b(\boldsymbol{x}) = 1$. □

Below, we write $\boldsymbol{a}(\boldsymbol{x})^{\circ k}$ for the polynomial vector given by raising each entry of $\boldsymbol{a}(\boldsymbol{x})$ to the kth power. Note also that in the two proofs below, unlike the first one above, we do not appeal to any direct SOS proofs! Instead, we use the properties of pseudoexpectations we have built up already to derive further properties.

Proposition 5.3 (SOS Hölder inequality)**.** *If $\widetilde{\mathbb{E}}$ is a pseudoexpectation of degree at least $4D$, then*

$$\widetilde{\mathbb{E}}[\langle \boldsymbol{a}(\boldsymbol{x}), \boldsymbol{b}(\boldsymbol{x})^{\circ 3}\rangle] \leq \left(\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x})\|_4^4]\right)^{1/4} \left(\widetilde{\mathbb{E}}[\|\boldsymbol{b}(\boldsymbol{x})\|_4^4]\right)^{3/4}. \tag{5.7}$$

Proof. We use the Cauchy-Schwarz inequality twice:

$$\begin{aligned}
\widetilde{\mathbb{E}}[\langle \boldsymbol{a}(\boldsymbol{x}), \boldsymbol{b}(\boldsymbol{x})^{\circ 3}\rangle] &\leq \left(\widetilde{\mathbb{E}}[\langle \boldsymbol{a}(\boldsymbol{x})^{\circ 2}, \boldsymbol{b}(\boldsymbol{x})^{\circ 2}\rangle]\right)^{1/2} \cdot \left(\widetilde{\mathbb{E}}[\|\boldsymbol{b}(\boldsymbol{x})\|_4^4]\right)^{1/2} \\
&\leq \left(\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x})\|_4^4]\right)^{1/4} \left(\widetilde{\mathbb{E}}[\|\boldsymbol{b}(\boldsymbol{x})\|_4^4]\right)^{1/4} \cdot \left(\widetilde{\mathbb{E}}[\|\boldsymbol{b}(\boldsymbol{x})\|_4^4]\right)^{1/2} \\
&= \left(\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x})\|_4^4]\right)^{1/4} \left(\widetilde{\mathbb{E}}[\|\boldsymbol{b}(\boldsymbol{x})\|_4^4]\right)^{3/4},
\end{aligned} \tag{5.8}$$

completing the proof. □

Proposition 5.4 (SOS L^4 Minkowski inequality)**.** *If $\widetilde{\mathbb{E}}$ is a pseudoexpectation of degree at least $4D$, then*

$$\left(\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})\|_4^4]\right)^{1/4} \leq \left(\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x})\|_4^4]\right)^{1/4} + \left(\widetilde{\mathbb{E}}[\|\boldsymbol{b}(\boldsymbol{x})\|_4^4]\right)^{1/4}. \tag{5.9}$$

Proof. We start by expanding,

$$\begin{aligned}
&\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})\|_4^4] \\
&\quad = \widetilde{\mathbb{E}}\left[\sum_{i=1}^N (a_i(\boldsymbol{x}) + b_i(\boldsymbol{x}))^4\right] \\
&\quad = \widetilde{\mathbb{E}}\left[\sum_{i=1}^N a_i(\boldsymbol{x})(a_i(\boldsymbol{x}) + b_i(\boldsymbol{x}))^3\right] + \widetilde{\mathbb{E}}\left[\sum_{i=1}^N b_i(\boldsymbol{x})(a_i(\boldsymbol{x}) + b_i(\boldsymbol{x}))^3\right]
\end{aligned}$$

and, using the SOS Hölder inequality on each term, we find

$$\leq \left(\left(\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x})\|_4^4]\right)^{1/4} + \left(\widetilde{\mathbb{E}}[\|\boldsymbol{b}(\boldsymbol{x})\|_4^4]\right)^{1/4}\right)\left(\widetilde{\mathbb{E}}[\|\boldsymbol{a}(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})\|_4^4]\right)^{3/4}, \tag{5.10}$$

and rearranging this gives the result. □

The reference [FH14] gives some more systematic development of these kinds of results, albeit working purely in terms of SOS proofs rather than the pseudoexpectation language we use here.

5.2 Rounding Tools

For now we give just one broadly useful rounding approach. This is perhaps the most naive method possible: we just match as many pseudomoments as we possibly can using a simple actual distribution.

Proposition 5.5 (Gaussian rounding)**.** *If $\widetilde{\mathbb{E}}$ is a pseudoexpectation on $(x_1, \dots, x_n)$ (of any degree at least 2), then there exists a measure μ on $\mathbb{R}^n$ that matches $\widetilde{\mathbb{E}}$ on moments of degree at most 2: for each $p \in \mathbb{R}[x_1, \dots, x_n]$ with $\deg p \leq 2$, we have*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[p(\boldsymbol{x})] = \widetilde{\mathbb{E}}[p(\boldsymbol{x})]. \tag{5.11}$$

Proof. We will take the Gaussian distribution $\mu = \mathcal{N}(\widetilde{\mathbb{E}}[\boldsymbol{x}], \widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top] - \widetilde{\mathbb{E}}[\boldsymbol{x}]\widetilde{\mathbb{E}}[\boldsymbol{x}]^\top)$. Clearly this will satisfy the condition claimed, so long as the "pseudocovariance matrix" $\widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top] - \widetilde{\mathbb{E}}[\boldsymbol{x}]\widetilde{\mathbb{E}}[\boldsymbol{x}]^\top$ is psd so that the Gaussian distribution is well-defined. To check this, we compute

$$\boldsymbol{v}^\top(\widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top] - \widetilde{\mathbb{E}}[\boldsymbol{x}]\widetilde{\mathbb{E}}[\boldsymbol{x}]^\top)\boldsymbol{v} = \widetilde{\mathbb{E}}[\langle \boldsymbol{v}, \boldsymbol{x}\rangle^2] - \widetilde{\mathbb{E}}[\langle \boldsymbol{v}, \boldsymbol{x}\rangle]^2 \geq 0, \tag{5.12}$$

the inequality following by the SOS Jensen inequality (Corollary 5.2). □

As a point of reference, the Goemans-Williamson rounding scheme from Chapter 1 is actually (at least in its first step) a special case of this. Recall that the Goemans-Williamson SDP may be viewed as the degree 2 SOS relaxation (in the Lasserre formulation) of the MaxCut problem. In this equivalence, the decision variable $\boldsymbol{X}$ is $\boldsymbol{X} = \widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top]$ and, as we saw in Section 4.2.3, we may assume without loss of generality that $\widetilde{\mathbb{E}}[\boldsymbol{x}] = \boldsymbol{0}$ in this relaxation.[1]

In our original description of the Goemans-Williamson rounding, we computed $\boldsymbol{V} \in \mathbb{R}^{r\times n}$ such that $\boldsymbol{X} = \boldsymbol{V}^\top\boldsymbol{V}$ (the "geometric description" of the relaxation), and then rounded $\boldsymbol{X}$ to $\hat{\boldsymbol{x}} = \mathsf{sgn}(\boldsymbol{V}^\top\boldsymbol{g})$ for $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_r)$. However, the distribution of the random vector $\boldsymbol{V}^\top\boldsymbol{g}$ is none other than $\mathcal{N}(\boldsymbol{0}, \boldsymbol{V}^\top\boldsymbol{V}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{X}) = \mathcal{N}(\widetilde{\mathbb{E}}[\boldsymbol{x}], \widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top] - \widetilde{\mathbb{E}}[\boldsymbol{x}]\widetilde{\mathbb{E}}[\boldsymbol{x}]^\top)$. Thus, the "hyperplane rounding" of Goemans-Williamson may equivalently be viewed as the Gaussian rounding followed by taking the sign to obtain a hypercube point.

A small variation on this will also be useful: often, we will show that $\widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top]$ is "close" to a rank one matrix $\boldsymbol{v}\boldsymbol{v}^\top$ in some sense, and we will want to recover an estimate of the

[1] Generally, whenever both the objective function and the constraints of the underlying polynomial optimization problem are invariant under the negation map $\boldsymbol{x} \mapsto -\boldsymbol{x}$, we may assume without loss of generality that all odd pseudomoments are zero.

vector $\boldsymbol{v}$. It is then natural to use as an estimator the top eigenvector of $\widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top]$; however, to analyze such rounding procedures, we will need a perturbation inequality that lets us convert distances between $\widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top]$ and $\boldsymbol{v}\boldsymbol{v}^\top$ to distances between the top eigenvector of $\widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top]$ and $\boldsymbol{v}$. The following inequality gives precisely such a guarantee; it is a special case of the Wedin or Davis-Kahan theorems.

Proposition 5.6 ([DK70, Wed72, YWS15])**.** *Suppose $\boldsymbol{M} \succeq \boldsymbol{0}$, and $\boldsymbol{\Delta}$ has the same dimensions as $\boldsymbol{M}$ with $\|\boldsymbol{\Delta}\| < \lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M})$. Let $\boldsymbol{v}$ be the top eigenvector of $\boldsymbol{M}$ and $\widetilde{\boldsymbol{v}}$ the top eigenvector of $\boldsymbol{M} + \boldsymbol{\Delta}$. Then,*

$$\langle \boldsymbol{v}, \widetilde{\boldsymbol{v}} \rangle^2 \geq 1 - \left(\frac{\|\boldsymbol{\Delta}\|}{\lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|} \right)^2 . \tag{5.13}$$

Exercises

Exercise 5.1 (Eigenvector perturbation bound)**.** *In this exercise, you will prove Proposition 5.6. Recall the setting: suppose $\boldsymbol{M} \succeq \boldsymbol{0}$, and $\boldsymbol{\Delta}$ has the same dimensions as $\boldsymbol{M}$ with $\|\boldsymbol{\Delta}\| < \lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M})$. Let $\boldsymbol{v}$ be the top eigenvector of $\boldsymbol{M}$ and $\widetilde{\boldsymbol{v}}$ the top eigenvector of $\boldsymbol{M} + \boldsymbol{\Delta}$ (so that both are unit vectors). You will show the perturbation inequality*

$$\langle \boldsymbol{v}, \widetilde{\boldsymbol{v}} \rangle^2 \geq 1 - \left(\frac{\|\boldsymbol{\Delta}\|}{\lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|} \right)^2 . \tag{5.14}$$

Follow these steps:

1. *Show that $\lambda_1(\boldsymbol{M}) - \lambda_i(\boldsymbol{M} + \boldsymbol{\Delta}) \geq \lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|$ for all $i \geq 2$.*
2. *Using Step 1, show that $\|\boldsymbol{\Delta}\boldsymbol{v}\| \geq (\lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M}) - \|\boldsymbol{\Delta}\|) \cdot \|(\boldsymbol{I} - \widetilde{\boldsymbol{v}}\widetilde{\boldsymbol{v}}^\top)\boldsymbol{v}\|$.*
3. *Complete the proof.*

6 | Case Study 1: Sparse Vectors in Subspaces

The first problem we will attack with the proofs-to-algorithms framework is that of recovering a *sparse* vector in a random subspace.

Definition 6.1. *We write $\|\boldsymbol{x}\|_0$ for the number of non-zero entries in $\boldsymbol{x}$, and call $\boldsymbol{x}$ s-*sparse *if $\|\boldsymbol{x}\|_0 \leq s$.*

We will consider the following setting. Suppose $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{x}^\star \in \mathbb{R}^n$ are unit vectors, where the $\boldsymbol{v}_i$ are orthonormal and $\boldsymbol{x}^\star$ is s-sparse. Define

$$V := \mathsf{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k), \tag{6.1}$$

$$W := \mathsf{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{x}^\star). \tag{6.2}$$

We also introduce for later the notation

$$\boldsymbol{V} := \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_k \\ | & | & & | \end{bmatrix}. \tag{6.3}$$

Though this last condition can be relaxed, we will also assume that $\boldsymbol{x}^\star$ is orthogonal to all of the $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$:

$$\langle \boldsymbol{v}_i, \boldsymbol{x}^\star \rangle = 0 \text{ for all } i \in [k]. \tag{6.4}$$

The subspace V and the vector $\boldsymbol{x}^\star$ will be unknown to us; we will observe W (through, say, an arbitrary basis) and our job will be to try to recover $\boldsymbol{x}^\star$. Finally, we will assume that V is *uniformly random*, by, say, taking $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ to be uniformly random vectors on the unit sphere in $\mathbb{R}^k$. This specific choice is also not essential; as we will see, our algorithm will work for any V that is "generic enough" in a particular technical sense. The idea is just that V should not contain any sparse or nearly-sparse vectors in order for us to have some hope of recovering $\boldsymbol{x}^\star$ from W, in which $\boldsymbol{x}^\star$ will then be the only sparse vector.

We make one last definition to start formulating optimization problems associated to this statistical problem.

Definition 6.2. *For a subspace $V \subset \mathbb{R}^n$, we write $\boldsymbol{P}_V \in \mathbb{R}^{n \times n}_{\mathsf{sym}}$ for the matrix of the orthogonal projection to V.*

In particular, the constraint "$\boldsymbol{x} \in V$" may be encoded as a polynomial constraint $\boldsymbol{P}_V \boldsymbol{x} = \boldsymbol{x}$, or equivalently $(\boldsymbol{I} - \boldsymbol{P}_V)\boldsymbol{x} = \boldsymbol{P}_{V^\perp}\boldsymbol{x} = \boldsymbol{0}$.

As foreshadowed in Chapter 5, the first thing to notice is that recovering $\boldsymbol{x}^\star$ is not a polynomial optimization problem in any obvious way. While we would like to solve a problem like

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{x}\|_0 \\ \text{subject to} & \boldsymbol{P}_W \boldsymbol{x} = \boldsymbol{x}, \end{array} \tag{6.5}$$

$\|\boldsymbol{x}\|_0$ is not a polynomial of $\boldsymbol{x}$, and the set $\{\boldsymbol{x} : \boldsymbol{x} \text{ is } s\text{-sparse}\}$ is not a semialgebraic set. Therefore, we need some *proxy* polynomial optimization problem that approximates the above.

6.1 Step 1: Polynomial Optimization Formulation

The way we will encode our task in a polynomial optimization problem is by comparing other, better behaved, ℓ^p norms of $\boldsymbol{x}$.[1] Generally, if $\|\boldsymbol{x}\|_2$ is fixed, then $\|\boldsymbol{x}\|_p$ is *larger* for sparse vectors when $p > 2$, and *smaller* for sparse vectors when $p < 2$. As a sanity check to convince yourself of this, note that when $\|\boldsymbol{x}\|_0 = 1$ then all of these norms equal 1, while when $\boldsymbol{x} \in \{\pm n^{-1/2}\}^n$, the "least sparse" possible vectors even in any approximate sense and having $\|\boldsymbol{x}\|_0 = n$, we have

$$\|\boldsymbol{x}\|_p = \left(n^{-p/2} \cdot n\right)^{1/p} = n^{1/p - 1/2}. \tag{6.6}$$

The simplest such proxy problem that is convenient to encode as polynomial optimization is to fix the ℓ^2 norm and maximize the ℓ^4 norm. We thus define

$$\mathsf{Opt} := \left\{ \begin{array}{ll} \text{maximize} & \|\boldsymbol{x}\|_4^4 = \sum_{i=1}^n x_i^4 \\ \text{subject to} & \|\boldsymbol{x}\|_2^2 = \sum_{i=1}^n x_i^2 = 1, \\ & \boldsymbol{P}_W \boldsymbol{x} = \boldsymbol{x} \end{array} \right\}. \tag{6.7}$$

This is not a perfect approximation of minimizing the sparsity of $\boldsymbol{x}$, since "nearly sparse" vectors with some very small non-zero entries will still have ℓ^4-norm about as large as exactly sparse vectors. But, if $\boldsymbol{x}^\star$ is essentially unique even among nearly sparse vectors in W, this will suffice to recover $\boldsymbol{x}^\star$.

Before proceeding, let us also record the objective value that $\boldsymbol{x}^\star$ has for this problem. As you can check, among s-sparse vectors with fixed ℓ^2 norm, the ℓ^4 norm is maximized by those with s non-zero entries that are all equal in magnitude, and therefore equal to $s^{-1/2}$. Thus we have

$$\|\boldsymbol{x}^\star\|_4^4 = \sum_{i=1}^n (x_i^\star)^4 \geq (s^{-1/2})^4 \cdot s = \frac{1}{s}. \tag{6.8}$$

[1] As a side note, while we use norm notation for $\|\boldsymbol{x}\|_0$, it is in fact not a norm, as you can check that it fails the triangle inequality.

6.2 Step 2: Analysis of Polynomial Optimization Problem

Following the outline from Chapter 5, the first order of business is to show that solutions $\boldsymbol{x}$ to Opt having large objective value will have a high correlation with $\boldsymbol{x}^\star$. Note that $\boldsymbol{x}^\star$ itself is a feasible point, so we may safely assume that the optimizer $\boldsymbol{x}$ of Opt will have $\|\boldsymbol{x}\|_4 \geq \|\boldsymbol{x}^\star\|_4$.

As mentioned earlier, we will need some sort of quantitative regularity condition on V. It turns out that the following is the notion that is useful here.

Definition 6.3 (Mixed matrix norms)**.** *For $\boldsymbol{M} \in \mathbb{R}^{n\times n}_{\mathsf{sym}}$, we define the $p \to q$* norm *of $\boldsymbol{M}$ as*

$$\|\boldsymbol{M}\|_{p\to q} := \max_{\boldsymbol{v}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}} \frac{\|\boldsymbol{M}\boldsymbol{v}\|_q}{\|\boldsymbol{v}\|_p}. \tag{6.9}$$

For $V \subset \mathbb{R}^n$ a subspace, we write

$$\|V\|_{p\to q} := \|\boldsymbol{P}_V\|_{p\to q} = \max_{\boldsymbol{v}\in V\setminus\{\boldsymbol{0}\}} \frac{\|\boldsymbol{v}\|_q}{\|\boldsymbol{v}\|_p}. \tag{6.10}$$

The reason for the name is that $\|\cdot\|_{p\to q}$ is an operator norm when the domain space $\mathbb{R}^n$ is endowed with the ℓ^p norm and the range space $\mathbb{R}^n$ with the ℓ^q norm.

Not surprisingly, it is the $2 \to 4$ norm that will be relevant for our purposes. The following is the key technical claim showing that Opt is a good approximation of the problem of recovering $\boldsymbol{x}^\star$.

Lemma 6.4. *For any V, for all $\boldsymbol{x} \in W$ with $\|\boldsymbol{x}\|_2 = 1$ and $\|\boldsymbol{x}\|_4 \geq \|\boldsymbol{x}^\star\|_4$, we have*

$$|\langle \boldsymbol{x}, \boldsymbol{x}^\star\rangle| \geq 1 - \frac{\|V\|_{2\to 4}}{\|\boldsymbol{x}\|_4} \geq 1 - \frac{\|V\|_{2\to 4}}{\|\boldsymbol{x}^\star\|_4}. \tag{6.11}$$

Proof. Since we assume $\boldsymbol{x}^\star$ is orthogonal to V, we have

$$\boldsymbol{x} = \langle \boldsymbol{x}, \boldsymbol{x}^\star\rangle \boldsymbol{x}^\star + \boldsymbol{P}_V \boldsymbol{x}. \tag{6.12}$$

Taking ℓ^4 norms on either side and using the triangle inequality (Minkowski's inequality),

$$\begin{aligned} \|\boldsymbol{x}\|_4 &\leq |\langle \boldsymbol{x}, \boldsymbol{x}^\star\rangle| \cdot \|\boldsymbol{x}^\star\|_4 + \|\boldsymbol{P}_V \boldsymbol{x}\|_4 \\ &\leq |\langle \boldsymbol{x}, \boldsymbol{x}^\star\rangle| \cdot \|\boldsymbol{x}\|_4 + \|V\|_{2\to 4}, \end{aligned} \tag{6.13}$$

and rearranging gives the result. □

To apply this to our particular situation, we must also control the $2 \to 4$ norm of V when it is random.

Lemma 6.5 (Corollary of Theorem 7.1 of [BBH+12])**.** *For an absolute constant $C > 0$, if $k = o(n^{1/2})$, then $\mathbb{P}[\|V\|_{2\to 4} \leq C/n^{1/4}] \to 1$.*

Corollary 6.6. *For an absolute constant $C > 0$, for V uniformly random, if $k = o(n^{1/2})$, then with high probability, for any $\boldsymbol{x} \in W$ with $\|\boldsymbol{x}\|_2 = 1$ and $\|\boldsymbol{x}\|_4 \geq \|\boldsymbol{x}^\star\|_4$ we have*

$$|\langle \boldsymbol{x}, \boldsymbol{x}^\star\rangle| \geq 1 - C\left(\frac{s}{n}\right)^{1/4}. \tag{6.14}$$

Let us give the proof of Lemma 6.5, not worrying for the moment about our future goal of making this "SOS-compatible" but rather applying the most immediately appealing technique. We will use the following technical tool.

Proposition 6.7 (ϵ-nets of the sphere)**.** *For an absolute constant $C > 0$, for any $n \in \mathbb{N}$ and $\epsilon > 0$, there exists a subset $S \subset \mathbb{S}^{n-1}$ such that:*

1. *$|S| \leq \exp(Cn/\epsilon)$, and*
2. *For any $\boldsymbol{x} \in \mathbb{S}^{n-1}$, there exists $\boldsymbol{y} \in S$ so that $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \epsilon$.*

Proof Sketch of Lemma 6.5. Fix $\epsilon \in (0,1)$. Let S be an ϵ-net of $\mathbb{S}^{k-1}$ as in Proposition 6.7. We consider the "discretization" of the $2 \to 4$ norm over $\boldsymbol{V}S$:

$$M := \max_{\boldsymbol{y} \in S} \|\boldsymbol{V}\boldsymbol{y}\|_4, \tag{6.15}$$

and compare to the $2 \to 4$ norm itself:

$$\|V\|_{2\to 4} = \max_{\boldsymbol{z} \in \mathbb{S}^{k-1}} \|\boldsymbol{V}\boldsymbol{z}\|_4. \tag{6.16}$$

For any $\boldsymbol{z} \in \mathbb{S}^{k-1}$, there exists $\boldsymbol{y} \in S$ so that $\|\boldsymbol{y} - \boldsymbol{z}\|_2 \leq \epsilon$. For this $\boldsymbol{y}$, we have by the reverse triangle inequality

$$\big|\, \|\boldsymbol{V}\boldsymbol{y}\|_4 - \|\boldsymbol{V}\boldsymbol{z}\|_4 \,\big| \leq \|\boldsymbol{V}\boldsymbol{y} - \boldsymbol{V}\boldsymbol{z}\|_4 = \|\boldsymbol{V}(\boldsymbol{y} - \boldsymbol{z})\|_4 \leq \|V\|_{2\to 4}\|\boldsymbol{y} - \boldsymbol{z}\|_2 \leq \epsilon\|V\|_{2\to 4}. \tag{6.17}$$

Thus M is a multiplicative approximation of $\|V\|_{2\to 4}$:

$$\|V\|_{2\to 4} \leq M + \epsilon\|V\|_{2\to 4}, \tag{6.18}$$

and in particular

$$\|V\|_{2\to 4} \leq \frac{1}{1-\epsilon} M. \tag{6.19}$$

So, it suffices to fix, say, $\epsilon := \frac{1}{2}$, in which case $|S| \leq \exp(Ck)$ for some $C > 0$, and to bound M with high probability. We then apply a union bound:

$$\mathbb{P}[M \geq t] = \mathbb{P}[\|\boldsymbol{V}\boldsymbol{y}\|_4 \geq t \text{ for some } \boldsymbol{y} \in S]$$

and, since for any fixed $\boldsymbol{y}$ the vector $\boldsymbol{V}\boldsymbol{y}$ is just a uniformly random unit vector in $\mathbb{R}^n$, writing $\boldsymbol{v}$ for such a vector we have

$$\begin{aligned} &\leq |S| \cdot \mathbb{P}[\|\boldsymbol{v}\|_4 \geq t] \\ &= \exp(Ck) \cdot \mathbb{P}\left[\sum_{i=1}^n v_i^4 \geq t^4\right]. \end{aligned} \tag{6.20}$$

Now we make a few heuristic leaps to avoid technicalities, but give the idea of the rest of the argument: let us view $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \frac{1}{n}\boldsymbol{I}_n)$, a Gaussian vector whose expected norm is $\mathbb{E}\|\boldsymbol{v}\|_2^2 = 1$. We then have $\mathbb{E}v_i^4 = 3/n^2$, so we certainly must have $t^4 \gtrsim n^{-1}$, or $t \gtrsim n^{-1/4}$, for the remaining probability to decay at all as $n \to \infty$. If $t \geq Kn^{-1/4}$ for a large constant K, then,

since the v_i^4 random variables are heavy-tailed, the probability above is dominated by some one v_i^4 being large. Neglecting terms of sub-leading order and constants in the exponents, we have

$$\mathbb{P}\left[\sum_{i=1}^{n} v_i^4 \geq t^4\right] \asymp \mathbb{P}\left[v_1^4 \geq t^4\right] \asymp \mathbb{P}\left[|v_1| \geq t\right] \lesssim \exp(-nt^2) \lesssim \exp(-\sqrt{n}). \tag{6.21}$$

Thus the probability bound is

$$\mathbb{P}[M \geq t] \lesssim \exp(Ck - \sqrt{n}), \tag{6.22}$$

which tends to zero since we assume $k = o(\sqrt{n})$. □

6.3 Step 3: Proofs-to-Algorithms Analysis of Sum-of-Squares Relaxation

Finally, to finish executing the plan from Chapter 5, we revisit the proof above and make each step "SOS-effective," showing that analogs of Lemma 6.4 and Lemma 6.5 hold in ways that apply to low-degree pseudoexpectations over the Opt constraints.

First of all, we will need to work not just with bounds on the $2 \to 4$ norm, but on such bounds that are *certifiable* by SOS. We therefore make the following definition.

Definition 6.8 (SOS $2 \to 4$ norm)**.** *For $V \subset \mathbb{R}^n$ a subspace, we define*

$$\|V\|_{2\to 4}^{\mathsf{SOS}} = \left(\min\left\{K \in \mathbb{R} : K\|\boldsymbol{P}_V \boldsymbol{x}\|_2^4 - \|\boldsymbol{P}_V \boldsymbol{x}\|_4^4 \in \mathsf{SOS}\right\}\right)^{1/4}. \tag{6.23}$$

This is just a natural degree 4 SOS relaxation of computing the $2 \to 4$ norm of V; in particular, we will always have

$$\|V\|_{2\to 4} \leq \|V\|_{2\to 4}^{\mathsf{SOS}} \tag{6.24}$$

In fact, this problem and its higher-degree analogs are themselves quite important in the SOS literature, because of connections to the Unique Games Conjecture and related problems. This connection was first explored by [BBH+12], whose major technical result (which we give below) [BKS14] repurposed for the analysis of the planted sparse vector problem.

Lemma 6.9 (SOS version of Lemma 6.4)**.** *For any V, for any pseudoexpectation $\widetilde{\mathbb{E}}$ of degree at least 4 that respects the constraints of* Opt *and has $\widetilde{\mathbb{E}}[\|\boldsymbol{x}\|_4^4] \geq \|\boldsymbol{x}^\star\|_4^4$, we have*

$$\widetilde{\mathbb{E}}[\langle \boldsymbol{x}, \boldsymbol{x}^\star\rangle^2] \geq 1 - 8\frac{\|V\|_{2\to 4}^{\mathsf{SOS}}}{(\widetilde{\mathbb{E}}[\|\boldsymbol{x}\|_4^4])^{1/4}} \geq 1 - 8\frac{\|V\|_{2\to 4}^{\mathsf{SOS}}}{\|\boldsymbol{x}^\star\|_4}. \tag{6.25}$$

Lemma 6.10 (SOS version of Lemma 6.5; Theorem 7.1 of [BBH+12])**.** *For an absolute constant $C > 0$, if $k = o(n^{1/2})$, then $\mathbb{P}[\|V\|_{2\to 4}^{\mathsf{SOS}} \leq C/n^{1/4}] \to 1$.*

From these, combined with a straightforward rounding scheme, our main result follows.

Theorem 6.11. *For $C > 0$ an absolute constant, if $k = o(n^{1/2})$ and $x^\star$ is s-sparse, there exists a polynomial-time randomized algorithm outputting $\hat{x}$ that, with high probability over V, satisfies*

$$\mathbb{E}\langle \hat{x}, x^\star\rangle^2 \geq 1 - C\left(\frac{s}{n}\right)^{1/4}. \tag{6.26}$$

Proof. Let $\widetilde{\mathbb{E}}^\star$ be the optimizer of the degree 4 SOS relaxation of Opt. Since $x^\star$ is a feasible point for Opt, this satisfies $\widetilde{\mathbb{E}}^\star[\|x\|_4^4] \geq \|x^\star\|_4^4$. So, Lemma 6.9 applies to $\widetilde{\mathbb{E}}^\star$. Applying it, using Lemma 6.10 to bound $\|V\|_{2\to 4}^{\mathsf{SOS}}$, and using that if $x^\star$ is s-sparse then $\|x^\star\|_4 \geq s^{-1/4}$ then shows that, with high probability over V, we have

$$\widetilde{\mathbb{E}}^\star[\langle x, x^\star\rangle^2] \geq 1 - C\left(\frac{s}{n}\right)^{1/4}. \tag{6.27}$$

Lastly, using Proposition 5.5 on Gaussian rounding, there is a Gaussian law μ for $\hat{x}$ whose moments match the pseudomoments of $\widetilde{\mathbb{E}}^\star$ to degree at most 2. In particular, we have

$$\mathop{\mathbb{E}}_{\hat{x}\sim\mu}[\langle x, x^\star\rangle^2] = \widetilde{\mathbb{E}}^\star[\langle x, x^\star\rangle^2] \geq 1 - C\left(\frac{s}{n}\right)^{1/4}, \tag{6.28}$$

as claimed. □

Proof of Lemma 6.9. We start by mimicking our original proof of Lemma 6.4 almost verbatim, using tools we developed for pseudoexpectations in Section 5.1 as needed.

Since we assume $x^\star$ is orthogonal to V, we have the polynomial equality

$$x \overset{(\mathrm{p})}{=} \langle x, x^\star\rangle x^\star + P_V x + (I - P_W)x. \tag{6.29}$$

We consider summing fourth powers on either side and applying $\widetilde{\mathbb{E}}$ that satisfies the Opt constraints.

$$\left(\widetilde{\mathbb{E}}[\|x\|_4^4]\right)^{1/4} = \left(\widetilde{\mathbb{E}}[\|\langle x, x^\star\rangle x^\star + P_V x + (I - P_W)x\|_4^4]\right)^{1/4}$$

and, expanding, using that $\widetilde{\mathbb{E}}[((I - P_W)x)_i q(x)] = 0$ for any $q(x)$ with $\deg q \leq 3$ and then combining back into a fourth power of a norm, we find that we can drop the last term,

$$= \left(\widetilde{\mathbb{E}}[\|\langle x, x^\star\rangle x^\star + P_V x\|_4^4]\right)^{1/4}$$

and, by the SOS L^4 Minkowski inequality (Proposition 5.4),

$$\begin{aligned}
&\leq \left(\widetilde{\mathbb{E}}[\|\langle x, x^\star\rangle x^\star\|_4^4]\right)^{1/4} + \left(\widetilde{\mathbb{E}}[\|P_V x\|_4^4]\right)^{1/4} \\
&= \left(\widetilde{\mathbb{E}}[\langle x, x^\star\rangle^4]\right)^{1/4}\|x^\star\|_4 + \left(\widetilde{\mathbb{E}}[\|P_V x\|_4^4]\right)^{1/4} \\
&\leq \left(\widetilde{\mathbb{E}}[\langle x, x^\star\rangle^4]\right)^{1/4}\left(\widetilde{\mathbb{E}}[\|x\|_4^4]\right)^{1/4} + \|V\|_{2\to 4}^{\mathsf{SOS}}.
\end{aligned} \tag{6.30}$$

Rearranging then gives

$$\left(\widetilde{\mathbb{E}}[\langle x, x^\star\rangle^4]\right)^{1/4} \geq 1 - \frac{\|V\|_{2\to 4}^{\mathsf{SOS}}}{\left(\widetilde{\mathbb{E}}[\|x\|_4^4]\right)^{1/4}}. \tag{6.31}$$

If the right-hand side is negative then the statement of the Lemma is trivial, so we may suppose the right-hand side is between 0 and 1. In this case, taking fourth powers and bounding powers of the second term gives

$$\widetilde{\mathbb{E}}[\langle \boldsymbol{x}, \boldsymbol{x}^\star \rangle^4] \geq 1 - 8 \frac{\|V\|_{2\to 4}^{\mathsf{SOS}}}{\left(\widetilde{\mathbb{E}}[\|\boldsymbol{x}\|_4^4]\right)^{1/4}}. \tag{6.32}$$

Finally, note that since $\boldsymbol{x}^\star$ is a unit vector, we have $\boldsymbol{x}^\star \boldsymbol{x}^{\star\top} \preceq \boldsymbol{I}_n$, whereby

$$\|\boldsymbol{x}\|_2^2 - \langle \boldsymbol{x}, \boldsymbol{x}^\star \rangle^2 = \boldsymbol{x}^\top (\boldsymbol{I} - \boldsymbol{x}^\star \boldsymbol{x}^{\star\top}) \boldsymbol{x} \in \mathsf{SOS}. \tag{6.33}$$

Thus we also have

$$\|\boldsymbol{x}\|_2^2 \langle \boldsymbol{x}, \boldsymbol{x}^\star \rangle^2 - \langle \boldsymbol{x}, \boldsymbol{x}^\star \rangle^4 \in \mathsf{SOS}, \tag{6.34}$$

and since $\widetilde{\mathbb{E}}$ respects the constraint $\|\boldsymbol{x}\|_2^2 = 1$ of Opt, applying $\widetilde{\mathbb{E}}$ on either side gives

$$\widetilde{\mathbb{E}}[\langle \boldsymbol{x}, \boldsymbol{x}^\star \rangle^2] \geq \widetilde{\mathbb{E}}[\langle \boldsymbol{x}, \boldsymbol{x}^\star \rangle^4] \geq 1 - 8 \frac{\|V\|_{2\to 4}^{\mathsf{SOS}}}{\left(\widetilde{\mathbb{E}}[\|\boldsymbol{x}\|_4^4]\right)^{1/4}}, \tag{6.35}$$

completing the proof. □

We note that the gymnastics at the end, though they may seem insignificant on the surface, are actually quite important for the general plan of the proof of Theorem 6.11: we need to obtain a degree 2 measurement of correlation with $\boldsymbol{x}^\star$ (not a degree 4 one) so that we can control the same expectation when we replace the pseudoexpectation with an expectation over a Gaussian distribution.

6.3.1 Matrix Concentration and Proof of Lemma 6.10

We now move on to the proof of Lemma 6.10, which is the most technical step of the argument. Before giving (most of) a careful proof, let us make some preliminary calculations. To bound $\|V\|_{2\to 4}^{\mathsf{SOS}}$, we will want to prove an SOS membership of the form

$$\frac{C}{n} \|\boldsymbol{P}_V \boldsymbol{x}\|_2^4 - \|\boldsymbol{P}_V \boldsymbol{x}\|_4^4 \in \mathsf{SOS}. \tag{6.36}$$

As a first convenience, we note (you may check this as an exercise) that this is equivalent to a memberhsip, for $\boldsymbol{y} = (y_1, \ldots, y_k)$, of the form

$$\frac{C}{n} \|\boldsymbol{y}\|_2^4 - \|\boldsymbol{V}\boldsymbol{y}\|_4^4 \in \mathsf{SOS}, \tag{6.37}$$

where we are just reparametrizing the enumeration of vectors in $\boldsymbol{V}$ from $\boldsymbol{P}_V \boldsymbol{x} = \boldsymbol{V}\boldsymbol{V}^\top \boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{R}^n$ to $\boldsymbol{V}\boldsymbol{y}$ for $\boldsymbol{y} \in \mathbb{R}^k$.

We will convert this into a linear algebra problem. For this special case of degree 4 homogeneous polynomials, let us slightly reformulate our notion of a matrix *representing* a polynomial: for $p \in \mathbb{R}[y_1, \ldots, y_k]$ homogeneous of degree 4, we say $\boldsymbol{A} \in \mathbb{R}_{\mathsf{sym}}^{k^2 \times k^2}$ represents p

if $p(\boldsymbol{y}) = \boldsymbol{y}^{\otimes 2^\top} \boldsymbol{A} \boldsymbol{y}^{\otimes 2}$. Then, the above is equivalent (by a minor variation on Proposition 4.6) to there existing $\boldsymbol{A}$ representing $\|\boldsymbol{y}\|_2^4$ and $\boldsymbol{B}$ representing $\|\boldsymbol{V}\boldsymbol{y}\|_4^4$ such that

$$\boldsymbol{B} \preceq \frac{C}{n} \boldsymbol{A}. \tag{6.38}$$

Let us try making the most natural choices of $\boldsymbol{B}$ and $\boldsymbol{A}$. Conveniently, we may take $\boldsymbol{A} := \boldsymbol{I}_{k^2}$, since $\boldsymbol{y}^{\otimes 2^\top} \boldsymbol{I}_{k^2} \boldsymbol{y}^{\otimes 2} = \|\boldsymbol{y}^{\otimes 2}\|_2^2 = \|\boldsymbol{y}\|_2^4$. For $\boldsymbol{B}$, we note that, if $\boldsymbol{r}_1, \dots, \boldsymbol{r}_n \in \mathbb{R}^k$ are the *rows* of $\boldsymbol{V}$, then $(\boldsymbol{V}\boldsymbol{y})_i = \langle \boldsymbol{w}_i, \boldsymbol{y} \rangle$, and thus

$$\|\boldsymbol{V}\boldsymbol{y}\|_4^4 = \sum_{i=1}^n (\boldsymbol{V}\boldsymbol{y})_i^4 = \sum_{i=1}^n \langle \boldsymbol{w}_i, \boldsymbol{y} \rangle^4 = \boldsymbol{y}^{\otimes 2^\top} \left(\sum_{i=1}^n \boldsymbol{w}_i^{\otimes 2} \boldsymbol{w}_i^{\otimes 2^\top} \right) \boldsymbol{y}^{\otimes 2}, \tag{6.39}$$

so it seems reasonable to take

$$\boldsymbol{B} := \sum_{i=1}^n \boldsymbol{w}_i^{\otimes 2} \boldsymbol{w}_i^{\otimes 2^\top}. \tag{6.40}$$

Recall that we previous approximated the law of the $\boldsymbol{v}_i$ as independent $\mathcal{N}(\boldsymbol{0}, \frac{1}{n}\boldsymbol{I}_n)$ vectors, so, following the same approximation, we may view the $\boldsymbol{w}_j$ as independent $\mathcal{N}(\boldsymbol{0}, \frac{1}{n}\boldsymbol{I}_k)$ vectors. Taking instead $\boldsymbol{h}_1, \dots, \boldsymbol{h}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_k)$ independent, we have that the law of $\boldsymbol{B}$ is then equivalently

$$\boldsymbol{B} := \frac{1}{n^2} \sum_{i=1}^n \boldsymbol{h}_i^{\otimes 2} \boldsymbol{h}_i^{\otimes 2^\top}. \tag{6.41}$$

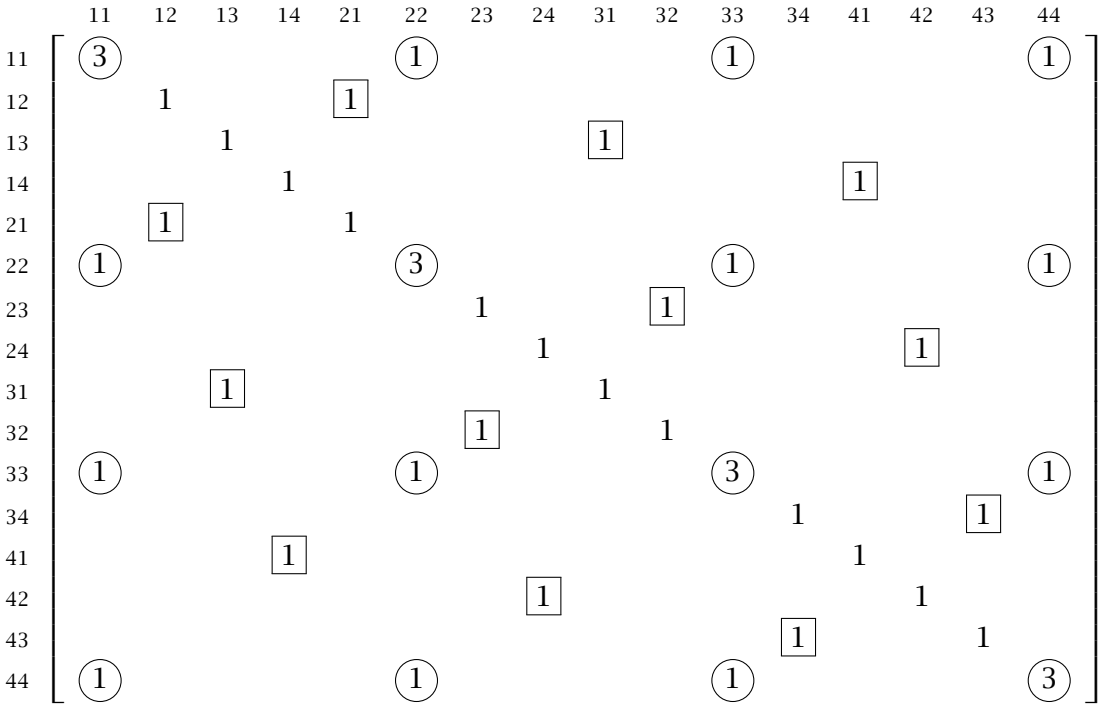
The bound we want is then equivalent to

$$\frac{1}{n} \sum_{i=1}^n \underbrace{\boldsymbol{h}_i^{\otimes 2} \boldsymbol{h}_i^{\otimes 2^\top}}_{=:\boldsymbol{H}_i} \preceq C \boldsymbol{I}_{k^2}, \tag{6.42}$$

which is just a bound on the largest eigenvalue of the left-hand side.

Since the left-hand side is an average of i.i.d. random matrices, it is natural to hope that it would converge in some sense to the expectation, $\mathbb{E}\boldsymbol{H}_1$. Supposing this happens, let us see if we at least have $\mathbb{E}\boldsymbol{H}_1 \preceq C\boldsymbol{I}_{k^2}$, or equivalently $\|\mathbb{E}\boldsymbol{H}_1\| \le C$. The matrix $\mathbb{E}\boldsymbol{H}_1$ will have

the following form, illustrated here with $k = 4$:



(6.43)

To process this: the rows and columns are indexed by pairs $(i, j) \in [k]^2$. The submatrix with rows and columns indexed by equal pairs (i, i) will have non-zero entries, with $\mathbb{E}[(\boldsymbol{h}_1)_i^4] = 3$ on the diagonal and $\mathbb{E}[(\boldsymbol{h}_1)_i^2(\boldsymbol{h}_1)_j^2] = 1$ on the off-diagonal—these are the circled entries. The rest of the diagonal will also be non-zero and equal 1. Finally, the entry indexed by pairs (i, j) and (j, i) will also equal 1—these are the entries marked with squares.

What is the largest eigenvalue of such a matrix? Let us write

$$\mathbb{E}\boldsymbol{H}_1 = \boldsymbol{I}_{k^2} + \boldsymbol{S}_1 + \boldsymbol{S}_2 + \boldsymbol{T}, \tag{6.44}$$

where $\boldsymbol{S}_1$ is a "blown up" version of the matrix $\boldsymbol{I}_k$ and $\boldsymbol{S}_2$ the same for $\mathbf{1}_k\mathbf{1}_k^\top$, both supported only on the submatrix indexed by (i, i) pairs (that is, $\boldsymbol{S}_1$ has a 1 everywhere where $\mathbb{E}\boldsymbol{H}_1$ has a 3, and $\boldsymbol{S}_2$ has a 1 in every circled entry), and where $\boldsymbol{T}$ consists of the entries in squares. We have $\|\boldsymbol{I}_{k^2}\| = 1$ and likewise $\|\boldsymbol{S}_1\| = 1$. Since $\boldsymbol{T}$ has only one non-zero entry per row, by the Gershgorin circle theorem $\|\boldsymbol{T}\| \le 1$. However, we have $\lambda_{\max}(\boldsymbol{S}_2) = \lambda_{\max}(\mathbf{1}_k\mathbf{1}_k^\top) = k$, so in fact we find, to our dismay, that

$$\lambda_{\max}(\mathbb{E}\boldsymbol{H}_1) \asymp k. \tag{6.45}$$

What went wrong? Here, we need to remember that our choice of representing matrices for $\|\boldsymbol{y}\|_2^4$ and $\|\boldsymbol{y}\|_4^4$ was not unique. We could try to adjust either $\boldsymbol{A}$ or $\boldsymbol{B}$; it is a little easier to work with $\boldsymbol{A}$. Indeed, it is useful to calculate what polynomial the "problematic" matrix $\boldsymbol{S}_2$ above represents:

$$\boldsymbol{y}^{\otimes 2^\top}\boldsymbol{S}_2\boldsymbol{y}^{\otimes 2} = \sum_{i=1}^k \sum_{j=1}^k y_i^2 y_j^2 = \|\boldsymbol{y}\|_2^4. \tag{6.46}$$

Thus, we may circumvent this challenge by making a different choice of $\boldsymbol{A}$, including a "barricade" in $\boldsymbol{A}$ against the term $\boldsymbol{S}_2$ that is making $\mathbb{E}\boldsymbol{H}_1$ exceed the identity in psd order.

For example, we can take

$$\boldsymbol{A} := \frac{1}{2}\boldsymbol{I}_{k^2} + \frac{1}{2}\boldsymbol{S}_2. \tag{6.47}$$

For sufficiently large C (e.g., $C = 6$ suffices), we will then have $\mathbb{E}\boldsymbol{H}_1 \preceq C\boldsymbol{A}$, so at least in expectation the bound we want will hold.

Thus it is enough to show that, with high probability, a bound holds of the form

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{H}_i \preceq C \cdot \mathbb{E}\boldsymbol{H}_1. \tag{6.48}$$

Equivalently, multiplying on the left and right by $(\mathbb{E}\boldsymbol{H}_1)^{-1/2}$, we may "whiten" this expression and, setting $\widehat{\boldsymbol{H}}_i := (\mathbb{E}\boldsymbol{H}_i)^{-1/2}\boldsymbol{H}_i(\mathbb{E}\boldsymbol{H}_i)^{-1/2}$, which satisfies $\mathbb{E}\widehat{\boldsymbol{H}}_i = \boldsymbol{I}_{k^2}$, it suffices to show

$$\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{H}}_i\right) \le C. \tag{6.49}$$

To show that this holds with high probability, we appeal to tools from the literature on *matrix concentration inequalities*. Specifically, we will use the following result.

Theorem 6.12 (Corollary of matrix Chernoff inequality; Theorem 5.1.1 of [Tro15])**.** *Suppose that $\boldsymbol{H}_1, \dots, \boldsymbol{H}_n \in \mathbb{R}^{m\times m}_{\mathsf{sym}}$ are i.i.d. random matrices that satisfy $\boldsymbol{0} \preceq \boldsymbol{H}_i \preceq L\boldsymbol{I}_m$ with probability 1. Then, for all $t \ge e$,*

$$\mathbb{P}\left[\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\boldsymbol{H}_i\right) \ge t\cdot\lambda_{\max}(\mathbb{E}\boldsymbol{H}_1)\right] \le m\exp\left(-\frac{\lambda_{\max}(\mathbb{E}\boldsymbol{H}_1)}{L}nt\log\left(\frac{e}{t}\right)\right). \tag{6.50}$$

To apply this properly, we would need to repeat our analysis with $\boldsymbol{h}_i$ having bounded norm, making them uniformly random vectors of norm exactly $\sqrt{k}$ rather than $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_k)$ vectors. Heuristically making this adjustment, we will have $\|\widehat{\boldsymbol{H}}_i\| \le \|\boldsymbol{H}_i\| = \|\boldsymbol{h}_i\|_2^4 = k^2$, so we may take $L = k^2$. We are working in dimension $m = k^2$. And lastly, we have $\lambda_{\max}(\mathbb{E}\widehat{\boldsymbol{H}}_1) = 1$ and it suffices for our purposes to take t a constant. So, we will have

$$\mathbb{P}\left[\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{H}}_i\right) \ge C\right] \le k^2\exp\left(-C\frac{n}{k^2}\right). \tag{6.51}$$

Thus, we need to take $n \gtrsim k^2\log k$ to make this tend to zero.

That is a bit weaker than what Lemma 6.10 promises—it should suffice to take $n \gg k^2$. In fact, the original work [BBH+12] uses a more specific tool for controlling the norm of the sum of i.i.d. *rank one* matrices, so long as the underlying vectors satisfy strong regularity assumptions. We do not get into this improvement here, but the reference where these bounds were first developed is [ALPTJ11].

Remark 6.13 (Nets to nets)**.** *It is interesting to note that the proof of [ALPTJ11] again uses the technique of discretizing over ϵ-nets and taking a union bound! Thus, in a rather roundabout way, we have actually constructed a genuine SOS version of our ϵ-net proof of Lemma 6.5. However, to do this we have had to solve the "matrix design" problem of choosing good representing matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and only then could we prove (via the calculations done for us in [ALPTJ11]) the requisite matrix inequality by discretizing.*

EXERCISES

Exercise 6.1 ($2 \to 4$ norm and small set expansion)**.** *Let $G = (V, E)$ be a d-regular graph with adjacency matrix $\boldsymbol{A}$. Fix some $\alpha \in [0,1]$, and let U be the subspace spanned by all eigenvectors of $\boldsymbol{A}$ with eigenvalue at least αd. Let $S \subseteq V$.*

Suppose we draw $\boldsymbol{x} \sim \mathsf{Unif}(S)$ and then $\boldsymbol{y}$ a uniformly random neighbor of $\boldsymbol{x}$. We then define

$$\Phi(S) := \mathbb{P}[\boldsymbol{y} \notin S], \tag{6.52}$$

a measure of the expansion *of S. Show that*

$$\Phi(S) \geq 1 - \alpha - \|U\|_{2\to 4}^2 \sqrt{|S|}. \tag{6.53}$$

> HINT: *First, show that $\|U\|_{2\to 4} = \|\boldsymbol{P}\|_{2\to 4} = \|bP\|_{4/3\to 2}$, where $\boldsymbol{P}$ is the projection matrix to U. Then, let $\boldsymbol{v}$ be the indicator vector of S, $v_i = \mathbb{1}\{i \in S\}$. Express $\Phi(S)$ in terms of a quadratic form with $\boldsymbol{v}$ and $\boldsymbol{A}$, and consider the decomposition of $\boldsymbol{v}$ into components in U and $U^\perp$.*

This shows that when the $2 \to 4$ norm of the top eigenspace of a graph is small then the graph is a small set expander*, i.e., sufficiently small subsets of vertices have large expansion. In fact, a converse is also true, which has been used to show that a good approximation of the $2 \to 4$ norm would refute the* Small Set Expansion Hypothesis*, which states, roughly speaking, that it is hard to distinguish good small set expanders from bad ones. This gives some evidence that it should be hard to approximate the $2 \to 4$ norm (and other similar quantities) in the worst case. See [BBH+12] for more details.*

NOTES

OTHER SOURCES Our treatment is quite close to that in Section 7.2 of the lecture notes [BS16], though we fill in some more details, especially in the analysis of the matrix concentration problem that arises at the end. As mentioned above, these results originate with [BKS14], who in this part of the paper were mostly finding a new application of earlier work of [BBH+12] when combined with their new ideas about rounding SOS relaxations.

7 | Case Study 2: Tensor Decomposition

The next application of SOS we will consider is that of decomposing low-rank tensors into components. Let $a_1, \ldots, a_r \in \mathbb{R}^n$ be unit vectors. We then observe the tensor

$$T := \sum_{i=1}^{r} a_i^{\otimes p} \tag{7.1}$$

and are interested in recovering the a_i. This is a much more subtle problem than the first one we saw, and the use of SOS in solving it is less obvious, so we will proceed slowly: first we will see when this problem is possible to solve and why using both some exact results and some heuristics, and then will understand how we can adapt the proofs-to-algorithms paradigm to help us solve it.

7.1 Rotation Problem and Benefit of Higher Moments

Let us consider the simplest case $p = 2$. Here, the problem is in fact generically *not* solvable: computational considerations aside, the a_i are simply not determined uniquely by T. In this case, T may be viewed as a matrix, given by $T = AA^\top$ where $A \in \mathbb{R}^{n \times r}$ has the a_i as its columns.

Suppose that we further assume the a_i are orthonormal. Then, T is the orthogonal projection matrix to the subspace spanned by the a_i. However, this subspace has many different bases, which can be completely uncorrelated with the a_i. Any a_i forming an orthonormal basis for the same subspace would yield the same observation T. In matrix language, whenever $Q \in \mathcal{O}(r)$ is orthogonal, we have

$$T = AA^\top = AQ(AQ)^\top, \tag{7.2}$$

so we cannot distinguish A from any such AQ on the basis of the observation T. This issue is sometimes called the *rotation problem* in the statistics literature.

Working with $p > 2$ has benefits and drawbacks: on the one hand, as we will see, this kind of "spectral obstruction" no longer appears, and for sufficiently small r the a_i are uniquely determined by T. On the other hand, we no longer have spectral tools at our disposal (like eigenvalue or singular value decompositions), so we must invent more sophisticated algorithms to try to recover the a_i.

Before proceeding, let us heuristically compute how large we might expect to be able to take r while being able to recover the a_i "information-theoretically" or without computational constraints. Note that the a_i have roughly rn degrees of freedom, while the number

of observations we make is roughly n^p, the number of entries in $\boldsymbol{T}$ (ignoring that $\boldsymbol{T}$ is symmetric, which will not affect the calculation at the level of precision we are interested in). A naive guess to the information-theoretic recoverability threshold might then be those r such that the number of degrees of freedom is at most the number of observations, $rn \leq n^p$ or

$$r \leq n^{p-1}. \tag{7.3}$$

Perhaps surprisingly, this very rough argument is correct once $p \geq 3$. In [BCO14] this is shown in a quite strong way: one consequence of their results is that, for a small constant c and sufficiently large n, whenever $r \leq cn^{p-1}$ then a "generic" low-rank tensor $\boldsymbol{T} \in (\mathbb{R}^n)^{\otimes p}$ uniquely identifies its components (for example, the set of $\boldsymbol{T}$ failing this has measure zero).

7.2 Verifiability and Injective Norm

Next, let us try to make a guess as to the growth of the rank r for which we might expect polynomial-time algorithms to be able to recover the $\boldsymbol{a}_i$. For this, we will adopt the average-case setting that our algorithms will operate in as well, assuming that the $\boldsymbol{a}_i$ are independent and uniformly distributed on the unit sphere.

The algorithms we will look at later will proceed by estimating one of the $\boldsymbol{a}_i$, and then (roughly speaking) repeating the process for a tensor $\boldsymbol{T}'$ where the $\boldsymbol{a}_i$ direction has been "projected away." For this reason, an important proxy problem is that of *verifying* a tensor component: given $\boldsymbol{x} \in \mathbb{R}^n$ another unit vector, can we check whether $\boldsymbol{x} \approx \boldsymbol{a}_i$ for some i?

A natural way to do this is to define

$$p(\boldsymbol{x}) = p_{\boldsymbol{T}}(\boldsymbol{x}) := \boldsymbol{T}[\boldsymbol{x}, \dots, \boldsymbol{x}] = \langle \boldsymbol{T}, \boldsymbol{x}^{\otimes p} \rangle = \sum_{i=1}^{r} \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle^p. \tag{7.4}$$

We might hope that $p(\boldsymbol{a}_i) \approx 1$ for each i, while $p(\boldsymbol{x}) \ll 1$ when $\boldsymbol{x}$ is not highly correlated with any of the $\boldsymbol{a}_i$. This in turn is similar to, though a bit stronger than, asking that the *injective norm* of $\boldsymbol{T}$ is close to 1.

Definition 7.1 (Injective norm)**.** *The* injective norm *of a symmetric tensor $\boldsymbol{T}$ is* $\|\boldsymbol{T}\|_{\mathsf{inj}} = \max_{\boldsymbol{x} \in \mathbb{S}^{n-1}} \boldsymbol{T}[\boldsymbol{x}, \dots, \boldsymbol{x}]$.

Even this question is rather tricky. As we can see from the form of $p(\boldsymbol{x})$, we have to consider the cases of even and odd p separately: for any fixed $\boldsymbol{x}$, the quantities $\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle$ are random, independent, and of magnitude about $n^{-1/2}$, but the behavior of their sum will depend on whether there are cancellations due to random signs or not. When p is even, then we will have

$$p(\boldsymbol{a}_i) \approx 1 + \sum_{i=1}^{r} (n^{-1/2})^p = 1 + \frac{r}{n^{p/2}}, \tag{7.5}$$

$$p(\boldsymbol{x}) \approx \frac{r}{n^{p/2}} \text{ for } \boldsymbol{x} \text{ fixed.} \tag{7.6}$$

Thus we are led to expect the algorithmic threshold

$$r \ll n^{p/2}. \tag{7.7}$$

Indeed, a straightforward formalization of this argument yields the following, for p both even and odd, which we will use later in our algorithms.

Proposition 7.2. *If $r \le n^{p/2}/\mathrm{polylog}(n)$, then, for any $\epsilon > 0$, with high probability we have*

$$\sup_{x \in \mathbb{S}^{n-1}} \left| p(x) - \max_{i=1}^{r} \langle a_i, x \rangle^p \right| \le \epsilon. \tag{7.8}$$

In words and more roughly, so long as $r \ll n^{p/2}$, the only way that $p(x)$ can be large is for x to be close to a tensor component.

When p is odd, a stronger claim holds, but to identify it we must carefully consider what drives large values of the sum we obtain in evaluating $p(x)$. Consider just the case of x fixed (as we saw above, the value of $p(a_i)$ will behave like this same sum plus 1). The quantities $n^{1/2}\langle a_i, x \rangle =: g_i$ behave like independent standard Gaussians, so we have

$$p(x) = n^{-p/2} \sum_{i=1}^{r} g_i^p. \tag{7.9}$$

While before it sufficed to just consider the mean value of $p(x)$, here we must more carefully consider the fluctuations in this random variable. Consider, therefore, the probability that $p(x) \ge t$ for some constant t, or equivalently $\mathbb{P}[\sum_{i=1}^{r} g_i^p \ge t n^{p/2}]$. We may then plug this into a union bound over an epsilon net of $\mathbb{S}^{n-1}$, so we are interested in whether or not this probability is of order $\exp(-\Omega(n))$.

The standard approach to analyzing this kind of probability is to use a Chernoff-type bound, which would involve computing $\mathbb{E}[\exp(\lambda g_i^p)]$. Note, however, that the g_i^p are i.i.d. random variables with a density that behaves like $\exp(-|x|^{2/p})$ (these are sometimes called *stretched exponential* densities), so these "exponential moments" are actually infinite. Thus the classical large deviations theory for sums of i.i.d. light-tailed random variables based on Chernoff bounds (and yielding, for instance, Cramér's large deviations theorem) does not apply. Fortunately, adjustments of that theory for this kind of random variable were derived by [Nag69a, Nag69b].

Let $\hat{t} := t n^{p/2}$. The basic idea of these results is that there are two regimes of $\hat{t}$, for which the deviation probability behaves differently. When $\hat{t} \lesssim r^{p/(2p-2)}$, then the "light-tailed" behavior holds: $\sum_{i=1}^{r} g_i^p$ behaves (after rescaling away a constant) like a Gaussian $\mathcal{N}(0, r)$, so $\mathbb{P}[\sum_{i=1}^{r} g_i^p \ge \hat{t}] \approx \exp(-\hat{t}^2/r)$, and this large deviation is driven by many terms in the sum being moderately large. When $\hat{t} \gtrsim r^{p/(2p-2)}$, then a "heavy-tailed" behavior takes over, where a large deviation of $\sum_{i=1}^{r} g_i^p$ is driven by one huge term dominating the sum. In this case, $\mathbb{P}[\sum_{i=1}^{r} g_i^p \ge \hat{t}] \approx \mathbb{P}[g_1^p \ge \hat{t}] \approx \exp(-\hat{t}^{2/p})$. Note that, as expected, the bounds match at the claimed threshold $\hat{t} = r^{p/(2p-2)}$.

Since we are interested in t constant, this threshold is when $n^{p/2} \sim r^{p/(2p-2)}$, or $r \sim n^{p-1}$, precisely the information-theoretic threshold of tensor decomposition! We then find that, so long as $r \ll n^{p-1}$, we have $\mathbb{P}[p(x) \ge t] \le \exp(-\Omega(n))$ and we expect (being informal about our union bound) that it should be possible to verify a tensor component and we expect to have $\|T\|_{\mathrm{inj}} = 1 + o(1)$ with high probability. (This may indeed be proved with a more careful analysis.)

We might then expect that, for odd p, tensor decomposition is possible all the way up to the information-theoretic threshold $r \sim n^{p-1}$. However, in a phenomenon that remains

somewhat mysterious, the threshold $r \sim n^{p/2}$ from the case of p even appears to be correct for p odd as well.

Open Problem 7.1 (Hardness of tensor decomposition)**.** *Give evidence of the computational hardness of order p tensor decomposition in the regime $n^{p/2} \ll r \ll n^{p-1}$ when p is odd.*

As we will see in Section 7.4.4, in fact it is a stronger property than the tensor injective norm just being small that will matter for us—we will also need for SOS to be able to *certify* that smallness. This too remains an open problem, which would be a more concrete way to give SOS-based evidence for the above.

Open Problem 7.2 (SOS lower bound for tensor injective norm)**.** *Show that, for p odd, low-degree SOS cannot certify the bound $\|\boldsymbol{T}\|_{\mathsf{inj}} \leq 1 + o(1)$ in the regime $n^{p/2} \ll r \ll n^{p-1}$.*

One natural direction for showing this kind of hardness addresses the simpler problem of for what values of r we can distinguish a random rank r tensor from a simpler random tensor with i.i.d. entries calibrated to have the same low-degree moments. While this seems too strong a notion of hardness to give evidence for the above claims, it remains an interesting problem that has received some attention independently.

Open Problem 7.3 (Distinguishing Wigner and Wishart tensors)**.** *Consider two distributions of random tensors, $\boldsymbol{T}^{(1)} = \frac{1}{\sqrt{r}} \sum_{i=1}^{r} \boldsymbol{a}_i^{\otimes p}$ for $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ and $\boldsymbol{T}^{(2)} \in (\mathbb{R}^n)^{\otimes p}$ with $T^{(2)}_{i_1 \cdots i_p} \sim \mathcal{N}(0, \mathbb{E}(T^{(1)}_{i_1 \cdots i_p})^2)$ drawn independently up to permutations of the indices. For what regimes of r are these distributions information-theoretically or computationally (in polynomial time) distinguishable as $n \to \infty$?*

The case that is better understood is $p = 2$, in which case we are asking about the distinguishability of the Wishart and Wigner random matrix *distributions. Here, [BDER16] showed both that when $r \gg n^3$ then $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$ are information-theoretically indistinguishable, and that when $r \ll n^3$ then a simple test examining the correlations of signs of "triangles" of entries in the matrices with high probability distinguishes the two (indeed, their motivation was studying random graph models formed as functions of these Gaussian constructions).*

For the tensor case $p \geq 3$, less is known. [NZ21, Mik20] have shown that if $r \gg n^{2p-1}$ then $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$ are information-theoretically indistinguishable (by proving quantitative bounds on the distance between their distributions). This should be tight, but I am not aware of any results on detection algorithms, or whether we should or should not expect a statistical-to-computational gap.

7.3 Jennrich Algorithm and Variants

Returning to the proofs-to-algorithms paradigm, let us familiarize ourselves with one kind of algorithm that can be used to decompose tensors with $p > 2$. For the sake of simplicity, let us consider $p = 3$. The basic idea is to use the tensor input to formulate a tractable spectral problem over matrices whose solution helps us identify the components $\boldsymbol{a}_i$. The algorithm we describe is attributed to Jennrich in the social science literature [Har70, LRA93], where it remained in relative obscurity until being rediscovered more recently.

The algorithm proceeds by drawing $g \sim \mathcal{N}(0, I_n)$ and forming the matrix

$$M_g := T[g, \cdot, \cdot] = \sum_{i=1}^{n} \langle a_i, g \rangle a_i a_i^\top, \tag{7.10}$$

where $T[g, \cdot, \cdot]$ denotes a "contraction" of T with g along one of its tensor "axes." If, as in our example from before, the a_i are orthonormal, then we can recover them by diagonalizing this matrix, since with probability 1 the $\langle a_i, g \rangle$ will be distinct and thus the eigendecomposition of M_g will be unique.

But, what is more surprising, we can complete the decomposition even if the a_i are only linearly independent. Let $D_g \in \mathbb{R}^{n \times n}$ be the diagonal matrix with diagonal entries $\langle a_i, g \rangle$ and let $A \in \mathbb{R}^{n \times n}$ have the a_i as its columns, so that $M_g = A D_g A^\top$. Let M_h be formed the same way with an independent $h \sim \mathcal{N}(0, I_n)$. Then, we observe that

$$M_g M_h^{-1} = (A D_g A^\top)(A D_h A^\top)^{-1} = A(D_g D_h^{-1}) A^{-1}, \tag{7.11}$$

and, while this is no longer a symmetric matrix, we can still recover A by computing its eigendecomposition, since the $(D_g D_h^{-1})_{ii} = \frac{\langle a_i, g \rangle}{\langle a_i, h \rangle}$ are again distinct with probability 1, so the a_i are the unique left eigenvectors of $M_g^{-1} M_h$.

Theorem 7.3. *If the a_i are linearly independent unit vectors, then Jennrich algorithm recovers them (with probability 1) from $T = \sum_{i=1}^{r} a_i^{\otimes 3}$.*

For a_i random in $\mathbb{S}^{n-1}$ or in general position, this shows that we can recover the a_i, when $p = 3$, for $r \le n$.

Remark 7.4. *It is reasonable to worry about numerical issues in Jennrich's algorithm; these are addressed by [BCMV14], who study the sensitivity of this procedure and show that, under a smoothed analysis model, the Jennrich algorithm indeed succeeds in recovering the a_i.*

We can also formulate a straightforward variant of this algorithm that, given higher degree moments, can recover the components for larger r.

Corollary 7.5. *If the $a_i^{\otimes k}$ are linearly independent for unit vectors a_i, then the adjusted Jennrich algorithm recovers them from $T = \sum_{i=1}^{r} a_i^{\otimes p}$ with $p = 2k + 1$.*

Proof. We again compute $T[g, \cdot, \ldots, \cdot]$, now with $2k$ free coordinates, but by grouping the free coordinates into two groups of k view the output as the matrix

$$M_g = \sum_{i=1}^{r} \langle a_i, g \rangle a_i^{\otimes k} a_i^{\otimes k^\top}. \tag{7.12}$$

We may then repeat the same argument from Jennrich's algorithm, using that the $a_i^{\otimes k}$ are linearly independent, to recover the $a_i^{\otimes k}$. Given these it is then straightforward to obtain the a_i, since, for example, from $a_i^{\otimes k}$ we may recover all ratios $(a_i)_j / (a_i)_k$ by taking ratios of suitable entries of the tensor. □

For a_i random or in general position, this shows that we can recover the a_i for

$$r \le n^k = n^{(p-1)/2}. \tag{7.13}$$

This threshold for r is lower than our prediction of the algorithmic threshold above by a factor of $n^{1/2}$. We are left with the natural question: is it possible to close this gap?

7.4 "Boosting" with Sum-of-Squares: Method of Pseudomoments

We now move towards a beautiful algorithmic idea of [BKS15], later elaborated on by [GM15, MSS16] and others. The idea is to use an SOS program to *imagine* or *hallucinate* higher degree moments than the ones we are actually given. Indeed, this is precisely the kind of data that comes with a high-degree SOS pseudoexpectation! We may then try to run the Jennrich algorithm and variants thereof on these imaginary moments, and hope to achieve the same performance these algorithms would have on higher degree moments. In a sense, in this strategy we trade computing time—needed to solve high-degree SOS relaxations—for data, and show that if we are willing to pay enough time, we can build "surrogate" moment data that, for our purposes, works just as well as "real" moment data.

We will focus on the case $p = 3$. Recall that in this case when $\boldsymbol{a}_i$ are random, Jennrich's algorithm recovers the $\boldsymbol{a}_i$ in polynomial time and with high probability when $r \lesssim n$. We will outline how it is possible to achieve the same when $r \ll n^{3/2}$ instead. This setup was treated by [GM15]'s sharpened analysis of an algorithm of [BKS15], giving a quasipolynomial time (i.e., time $n^{O(\log n)}$) algorithm, and later improved by [MSS16] to a truly polynomial time algorithm.

7.4.1 Step 1: Nuances in Polynomial Optimization

Before advancing to this elaborate algorithmic idea, let us see why simpler applications of proofs-to-algorithms would not work. To try to recover one of the $\boldsymbol{a}_i$, it might seem reasonable to try to solve the polynomial optimization problem

$$\mathsf{Opt} := \left\{ \begin{array}{ll} \text{maximize} & p(\boldsymbol{x}) = \boldsymbol{T}[\boldsymbol{x},\boldsymbol{x},\boldsymbol{x}] \\ \text{subject to} & \|\boldsymbol{x}\|^2 = 1 \end{array} \right\}, \tag{7.14}$$

the same problem defining the injective tensor norm.

In the case of the planted sparse vector problem (from Chapter 6), we could show that the output of the analogous problem was directly useful, being highly correlated with the object we wanted to recover. Here, we have to think a little more carefully—for $\boldsymbol{a}_i$ random, we do not expect the maximizer of this problem to be unique; indeed, we expect *each* of the $\boldsymbol{a}_1, \dots, \boldsymbol{a}_r$ to (approximately) maximize the objective.

On the surface, that still does not seem like a problem—if we could solve this optimization problem exactly, then the output would be one of the $\boldsymbol{a}_i$, just as we hoped. But this argument will not quite apply to SOS. To see why not, recall that the Lasserre formulation of SOS begins by convexifying the problem Opt itself to *measures* over the feasible set:

$$\mathsf{Opt} = \left\{ \begin{array}{ll} \text{maximize} & p(\boldsymbol{x}) = \boldsymbol{T}[\boldsymbol{x},\boldsymbol{x},\boldsymbol{x}] \\ \text{subject to} & \|\boldsymbol{x}\|^2 = 1 \end{array} \right\} = \left\{ \begin{array}{ll} \text{maximize} & \mathbb{E}_{\boldsymbol{x}\sim\mu}\, p(\boldsymbol{x}) \\ \text{subject to} & \mu \in \mathcal{M}(\mathbb{S}^{n-1}) \end{array} \right\}. \tag{7.15}$$

It is again helpful to compare with planted sparse vector: there, even after this step, we could show that any μ with high objective function value would have to mostly be concentrated on vectors close to the one we wanted to recover, and access to such μ amounts to

access to a good estimator of that vector. Here, in contrast, since there are several approximate optimizers $a_1, \dots, a_r$, a μ with a high objective function value will only need to be supported on the *set* $\{a_1, \dots, a_r\}$; roughly speaking, we may assume the optimizer is of the form $\mu^\star = \sum_{i=1}^r \lambda_i \delta_{a_i}$ for some $\lambda_i \in [0,1]$ with $\sum_{i=1}^r \lambda_i = 1$.

Would access to such a measure help us to find one of the a_i? It is not obvious that it would! Remember that, to make an argument we might convert to an SOS argument, we should use "polynomial reasoning," which in this setting amounts to information about the moments of $\mu^\star$. How can we use the moments of $\mu^\star$ to recover one or all of the a_i?

The key insight is to observe that this is just another tensor decomposition problem, albeit one where we have access to *more* moments than we started with in T. Namely, for any q, even $q \gg p$, under the above simplifications we can compute

$$\mathop{\mathbb{E}}_{x \sim \mu^\star} x^{\otimes q} = \sum_{i=1}^r \lambda_i a_i^{\otimes q}. \tag{7.16}$$

For sufficiently large odd q, the $a_i^{\otimes (q-1)/2}$ will be linearly independent, so, for instance, the modified Jennrich algorithm would recover the a_i.

Considering how to turn this into an SOS algorithm, we are led to the plan of using "imaginary moments" mentioned above, sometimes called (e.g. in [Shi19]) the *method of pseudomoments*: instead of computing a true measure $\mu^\star$, we will find a suitable pseudoexpectation $\widetilde{\mathbb{E}}$ and apply a Jennrich-type algorithm to $\widetilde{\mathbb{E}}[x^{\otimes q}]$ for some large q.

7.4.2 Step 2: Baby Jennrich Algorithm with True Moments

Following the proofs-to-algorithms plan, we now analyze a variant of Jennrich's algorithm in the above context assuming we receive an exact solution; that is, assuming that we have access to high-degree moments of some $\mu^\star$ as above. For the sake of simplicity, suppose $\mu^\star$ is the most balanced possible output, i.e.,

$$\mu^\star = \frac{1}{r} \sum_{i=1}^r \delta_{a_i}. \tag{7.17}$$

We call the algorithm we analyze the "baby Jennrich" algorithm (in [BS16] this is also referred to as the "brute data" algorithm since, as we will see, it succeeds at tensor decomposition provided we input very high-degree moments). We work with a tensor $\widetilde{T} = \mathbb{E}_{x \sim \mu^\star} x^{\otimes D}$ for some large D, which we assume is even. The algorithm is similar to Jennrich's algorithm, but contracts many of the axes of $\widetilde{T}$ with several different Gaussian random vectors.

Definition 7.6 (Baby Jennrich algorithm)**.** *Given a tensor $\widetilde{T} \in (\mathbb{R}^n)^{\otimes D}$ for D even, we define the randomized algorithm* $\mathsf{BJ}(\widetilde{T})$ *to output the top eigenvector (that with largest eigenvalue, and having unit norm*[1]*) of the matrix $M := \widetilde{T}[g_1, g_1, \dots, g_{D/2-1}, g_{D/2-1}, \cdot, \cdot]$ where $g_1, \dots, g_{D/2-1} \sim \mathcal{N}(0, I_n)$ independently.*

Proposition 7.7. $\mathsf{BJ}(\widetilde{T})$ *runs in time $n^{O(D)}$.*

[1] This eigenvector is only unique up to a sign flip; if v is one of these two vectors, to be fully precise we may output a uniform choice of v or $-v$ here.

As we will use later, the baby Jennrich algorithm can be applied to any tensor, but let us consider specifically the case $\widetilde{\boldsymbol{T}} = \mathbb{E}_{\boldsymbol{x}\sim\mu^\star}\boldsymbol{x}^{\otimes D}$. The matrix we form in computing $\mathsf{BJ}(\widetilde{\boldsymbol{T}})$ may then be written

$$\boldsymbol{M} = \sum_{i=1}^{r} W(\boldsymbol{a}_i)^2 \boldsymbol{a}_i \boldsymbol{a}_i^\top \tag{7.18}$$

where

$$W(\boldsymbol{x}) = \prod_{j=1}^{D/2-1} \langle \boldsymbol{g}_j, \boldsymbol{x}\rangle. \tag{7.19}$$

The idea is then that, with small probability but one that is bounded below, we have $\boldsymbol{M}$ is close to a multiple of some $\boldsymbol{a}_i\boldsymbol{a}_i^\top$, thanks to the $\boldsymbol{g}_j$ being unusually aligned with this particular $\boldsymbol{a}_i$. We can then repeatedly try such random choices of $\boldsymbol{g}_j$ until we successfully identify a component $\boldsymbol{a}_i$.

Remark 7.8. *One way to view the baby Jennrich rounding algorithm is that it tries many random "tiltings" or reweightings of the measure $\mu^\star$, giving weights $W(\boldsymbol{a}_i)^2$ to component $\boldsymbol{a}_i$. While most of the time these weights will be close to the same and the tilted measure will still look uniform over the $\boldsymbol{a}_i$, occasionally we will accidentally tilt the measure to strongly favor one component.*

To analyze the algorithm, we identify an event on which $W(\boldsymbol{a}_1)^2 \gg W(\boldsymbol{a}_i)^2$ for all $i \geq 2$. Namely, we have:

$$\begin{aligned}\mathbb{P}[W(\boldsymbol{a}_1)^2 \geq 2^{D-1}] &\geq \mathbb{P}\Big[\,|\langle \boldsymbol{g}_j, \boldsymbol{a}_1\rangle| \geq 2 \text{ for all } j \in [D/2-1]\Big] \\ &= \mathbb{P}[|\langle \boldsymbol{g}_1, \boldsymbol{a}_1\rangle| \geq 2]^{D/2-1} \\ &= e^{-O(D)},\end{aligned} \tag{7.20}$$

since $\langle \boldsymbol{g}_1, \boldsymbol{a}_1\rangle$ has the law $\mathcal{N}(0,1)$, so the remaining probability is just some small fixed positive number. On the other hand, one may show that, conditional on this rare event, the remaining $W(\boldsymbol{a}_i)^2$ are still small: for a small $\delta > 0$,

$$\mathbb{P}\left[\max_{2\leq i\leq r} W(\boldsymbol{a}_i)^2 \leq (2-\delta)^{D-1} \;\middle|\; W(\boldsymbol{a}_1)^2 \geq 2^{D-1}\right] \geq 1 - re^{-O(D)}. \tag{7.21}$$

We would like the probability $e^{-O(D)}$ of our event to be only polynomially small in n, so that in polynomially many trials we expect to observe this event at least once. We thus take D proportional to $\log n$.

Lemma 7.9. *Suppose $r \leq n^{3/2}/\mathrm{polylog}(n)$. Then, for any $\epsilon > 0$, there is a $C > 0$ and an algorithm that terminates in $n^{O(\log n)}$ with high probability (with C and the algorithm runtime depending only on ϵ and the polynomial bound on r) that, given $\boldsymbol{T} = \sum_{i=1}^r \boldsymbol{a}_i^{\otimes 3}$ and $\widetilde{\boldsymbol{T}} = \mathbb{E}_{\boldsymbol{x}\sim\mu^\star}\boldsymbol{x}^{\otimes D}$ for $D = C\log n$, outputs $\hat{\boldsymbol{a}}$ satisfying $\langle \hat{\boldsymbol{a}}, \boldsymbol{a}_i\rangle^2 \geq 1-\epsilon$ for some $i \in [r]$.*

We note that the algorithm runtime of $n^{O(D)} = n^{O(\log n)}$ comes just from the cost of, e.g., looking at every entry of $\widetilde{\boldsymbol{T}}$.

Proof. Let us choose C large enough so that, for some $B > 0$ and some $0 < a < A$ with $n^{A-a} \gg r$, (7.21) implies that

$$\mathbb{P}\left[W(\boldsymbol{a}_1)^2 \geq n^A \text{ and } \max_{2\leq i\leq r} W(\boldsymbol{a}_i)^2 \leq n^a\right] \geq n^{-B}. \tag{7.22}$$

On this event, we have

$$\frac{\|W(\boldsymbol{a}_1)^2 \boldsymbol{a}_1\boldsymbol{a}_1^\top\|}{\|\boldsymbol{M}\|} \geq \frac{n^A}{n^A + rn^a} = 1 - o(1). \tag{7.23}$$

Thus, taking $\hat{\boldsymbol{a}}$ to be the top eigenvector of $\boldsymbol{M}$, on this event we have that $\langle \hat{\boldsymbol{a}}, \boldsymbol{a}_1\rangle^2 \geq 1 - \epsilon$ by Proposition 5.6.

On the other hand, given any estimator $\hat{\boldsymbol{a}}$, we may check whether $\langle \hat{\boldsymbol{a}}, \boldsymbol{a}_i\rangle \geq 1 - \epsilon$ for some i by thresholding $\boldsymbol{T}[\hat{\boldsymbol{a}}, \hat{\boldsymbol{a}}, \hat{\boldsymbol{a}}]$ (by the "verifiability" property from Proposition 7.2, since $r \leq n^{3/2}/\mathrm{polylog}(n)$). Thus, the algorithm that repeatedly samples random W and checks whether $\hat{\boldsymbol{a}}$ satisfies this property satisfies the stated conditions, since in, e.g., n^{2B} trials at least one will with high probability satisfy the event above. □

Remark 7.10. *The result is intentionally suboptimal: in principle, we could use $\widetilde{\boldsymbol{T}}$ for the "verification" step, in which case we could make this algorithm work for a much larger r. Indeed, in this setting via $\widetilde{\boldsymbol{T}}$ we are allowing ourselves access, in effect, to $\sum_{i=1}^r \boldsymbol{a}_i^{\otimes D}$, so this is to be expected. However, in formulating the result as above we are anticipating that, in formulating the SOS version of the claim below, we will only be able to make much weaker assumptions on $\widetilde{\boldsymbol{T}}$, so we will only be able to use $\boldsymbol{T}$ itself for the verification step.*

7.4.3 Step 3A: Baby Jennrich Algorithm with Pseudomoments

To produce an SOS algorithm matching the performance discussed above, we will replace $\widetilde{\boldsymbol{T}}$ given by $\mathbb{E}_{\boldsymbol{x}\sim\mu^\star}\boldsymbol{x}^{\otimes D}$ with a tensor of *pseudomoments* coming from $\widetilde{\mathbb{E}}^\star$ the optimizer of a high-degree SOS relaxation of Opt:

$$\widetilde{\boldsymbol{T}} := \widetilde{\mathbb{E}}^\star[\boldsymbol{x}^{\otimes D}]. \tag{7.24}$$

Towards analyzing this idea, we first revisit the above argument about the baby Jennrich algorithm and give a version of it for pseudomoments. The main issue that needs to be addressed is that the above argument used that $\mu^\star$ was supported precisely on $\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_r\}$. When working with pseudomoments, we do not have access to such reasoning (we do not even have access to a genuine measure!), so we need to make a softer assumption. Reviewing the argument, we find that most of the structure of $\mu^\star$ is irrelevant, and that the baby Jennrich rounding procedure will work just as well for any μ that (1) is supported on $\mathbb{S}^{n-1}$ and (2) has a sufficiently large fraction of its mass on (or even near) some $\boldsymbol{a}_i$. These conditions are amenable to working with SOS pseudomoments, and indeed we have the following general result about applying the baby Jennrich algorithm to pseudomoments.

Lemma 7.11 (Baby Jennrich rounding; Theorem 5.1 of [BKS15])**.** *Suppose $\widetilde{\mathbb{E}}$ is a degree D pseudoexpectation (for D even) over the constraint $\|\boldsymbol{x}\|^2 = 1$ and satisfying $\widetilde{\mathbb{E}}\langle \boldsymbol{a}, \boldsymbol{x}\rangle^D \geq \exp(-\epsilon D)$ for some $\boldsymbol{a} \in \mathbb{S}^{n-1}$, where $D \geq (1/\epsilon)\log(1/\epsilon)$. Then,*

$$\mathbb{P}[\langle \mathsf{BJ}(\widetilde{\mathbb{E}}[\boldsymbol{x}^{\otimes D}]), \boldsymbol{a}\rangle^2 \geq 1 - O(\epsilon)] \geq \exp(-O_\epsilon(D)). \tag{7.25}$$

Proof Sketch. Let us again write $W(\boldsymbol{x}) := \prod_{j=1}^{D/2-1} \langle \boldsymbol{g}_j, \boldsymbol{x}\rangle$. The matrix $\boldsymbol{M}$ formed in computing $\mathsf{BJ}(\widetilde{\mathbb{E}}[\boldsymbol{x}^{\otimes D}])$ may then be expressed in terms of $\widetilde{\mathbb{E}}$ as

$$\boldsymbol{M} = \widetilde{\mathbb{E}}[W(\boldsymbol{x})^2 \boldsymbol{x}\boldsymbol{x}^\top]. \tag{7.26}$$

In particular, we have $\boldsymbol{M} \succeq \boldsymbol{0}$ and

$$\mathsf{Tr}(\boldsymbol{M}) = \widetilde{\mathbb{E}}[W(\boldsymbol{x})^2 \|\boldsymbol{x}\|^2] = \widetilde{\mathbb{E}}[W(\boldsymbol{x})^2], \tag{7.27}$$

since $\widetilde{\mathbb{E}}$ respects the constraint $\|\boldsymbol{x}\|^2 = 1$. By Proposition 5.6, it suffices to show that $\boldsymbol{a}^\top \boldsymbol{M}\boldsymbol{a} \geq (1 - O(\epsilon))\,\mathsf{Tr}(\boldsymbol{M})$ with the stated probability, which is equivalent to the relation of pseudoexpectations

$$\widetilde{\mathbb{E}}[W(\boldsymbol{x})^2 \langle \boldsymbol{x}, \boldsymbol{a}\rangle^2] \geq (1 - O(\epsilon))\widetilde{\mathbb{E}} W(\boldsymbol{x})^2. \tag{7.28}$$

Following the reasoning in the argument in Section 7.4.2, let E be the event that $\langle \boldsymbol{g}_j, \boldsymbol{a}\rangle \geq K$ for all $j \in [D/2 - 1]$, for some large $K = K(\epsilon)$ to be fixed. As before, we have $\mathbb{P}[E] \geq \exp(-O_\epsilon(D))$. Thus it suffices to show our original claim conditional on the event E.

In this proof sketch, we will only show that the *expectations* conditional on E of (7.28) holds; the full proof in [BKS15] requires a more intricate second moment calculation as well. Let us define

$$C := \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_n)} [\langle \boldsymbol{g}, \boldsymbol{a}\rangle^2 \mid \langle \boldsymbol{g}, \boldsymbol{a}\rangle \geq K] - 1 = \mathop{\mathbb{E}}_{g\sim\mathcal{N}(0,1)} [g^2 \mid g \geq K] - 1. \tag{7.29}$$

Clearly $C = C(K)$ is a strictly increasing function. We have

$$\mathbb{E}[\boldsymbol{g}_j \boldsymbol{g}_j^\top \mid E] = \mathbb{E}[((\boldsymbol{I}_n - \boldsymbol{a}\boldsymbol{a}^\top) + \boldsymbol{a}\boldsymbol{a}^\top)\boldsymbol{g}_j\boldsymbol{g}_j^\top((\boldsymbol{I}_n - \boldsymbol{a}\boldsymbol{a}^\top) + \boldsymbol{a}\boldsymbol{a}^\top) \mid E]$$

and, since conditioning on E only affects $\boldsymbol{g}_j$ in the $\boldsymbol{a}$ direction, we have $(\boldsymbol{I}_n - \boldsymbol{a}\boldsymbol{a}^\top)\boldsymbol{g}_j$ is centered and independent of $\boldsymbol{a}\boldsymbol{a}^\top\boldsymbol{g}_j$, so the cross terms cancel and we find

$$\begin{aligned} &= \mathbb{E}[\langle \boldsymbol{g}_j, \boldsymbol{a}\rangle^2 \mid E]\boldsymbol{a}\boldsymbol{a}^\top + \boldsymbol{I}_n - \boldsymbol{a}\boldsymbol{a}^\top \\ &= C\boldsymbol{a}\boldsymbol{a}^\top + \boldsymbol{I}_n. \end{aligned} \tag{7.30}$$

Let us write expectations over W as a shorthand for expectations over all $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{D/2-1}$. By linearity of the pseudoexpectation, we then have for the expectation on the right-hand side of (7.28),

$$\begin{aligned} (\text{RHS}) &= \mathop{\mathbb{E}}_{W} \left[\widetilde{\mathbb{E}}[W(\boldsymbol{x})^2] \mid E\right] \\ &= \widetilde{\mathbb{E}} \mathop{\mathbb{E}}_{W}[W(\boldsymbol{x})^2 \mid E] \end{aligned}$$

where we emphasize that the expectation is over the *coefficients* in W, and so, by independence,

$$= \widetilde{\mathbb{E}} \prod_{j=1}^{D/2-1} \mathbb{E}[\langle \boldsymbol{g}_j, \boldsymbol{x}\rangle^2 \mid E]$$

and using the above calculation

$$\begin{aligned} &= \widetilde{\mathbb{E}}\left(C\langle \boldsymbol{x}, \boldsymbol{a}\rangle^2 + \|\boldsymbol{x}\|^2\right)^{D/2-1} \\ &= \widetilde{\mathbb{E}}\left(C\langle \boldsymbol{x}, \boldsymbol{a}\rangle^2 + 1\right)^{D/2-1}. \end{aligned} \tag{7.31}$$

Similarly on the left-hand side we have

$$\begin{aligned} (\text{LHS}) &= \mathop{\mathbb{E}}_{W}\left[\widetilde{\mathbb{E}}[W(\boldsymbol{x})^2\langle \boldsymbol{x}, \boldsymbol{a}\rangle^2] \mid E\right] \\ &= \widetilde{\mathbb{E}}\left[\langle \boldsymbol{x}, \boldsymbol{a}\rangle^2\left(C\langle \boldsymbol{x}, \boldsymbol{a}\rangle^2 + 1\right)^{D/2-1}\right] \end{aligned}$$

This is now effectively a pseudoexpectation over the single scalar indeterminate $\langle \boldsymbol{x}, \boldsymbol{a}_i\rangle$. We recall that any degree D polynomial inequality in one variable admits a degree D SOS proof (see Proposition 2.9). In this case, we will use the univariate inequality $t^2(Ct^2+1)^{D/2-1} \geq (1-\frac{2}{C})(Ct^2+1)^{D/2-1} - (C-1)^{D/2-1}$ for all $t \in \mathbb{R}$, which may be checked by considering $t^2 \geq 1 - \frac{2}{C}$ and $t^2 \leq 1 - \frac{2}{C}$ separately. This gives:

$$\begin{aligned} &\geq \widetilde{\mathbb{E}}\left[\left(1-\frac{2}{C}\right)\left(C\langle \boldsymbol{x}, \boldsymbol{a}\rangle^2 + 1\right)^{D/2-1} - (C-1)^{D/2-1}\right] \\ &= \left(1-\frac{2}{C}\right)(\text{RHS}) - (C-1)^{D/2-1}. \end{aligned} \tag{7.32}$$

To finish the calculation, we make a series of estimates. We have:

$$\begin{aligned} (C-1)^{D/2-1} &= C^{D/2-1}(1-C^{-1})^{D/2-1} \\ &\leq C^{D/2-1}e^{-D/C} \\ &= C^{D/2-1}e^{-(1/C-\epsilon)D}e^{-\epsilon D} \end{aligned}$$

and, by the hypothesis of this Lemma,

$$\begin{aligned} &\leq C^{D/2-1}e^{-(1/C-\epsilon)D}\widetilde{\mathbb{E}}\langle \boldsymbol{a}, \boldsymbol{x}\rangle^D \\ &\leq C^{D/2-1}\widetilde{\mathbb{E}}\langle \boldsymbol{a}, \boldsymbol{x}\rangle^{D-2} \\ &\leq e^{-(1/C-\epsilon)D}(\text{RHS}). \end{aligned} \tag{7.33}$$

Thus we find

$$(\text{LHS}) \geq \left(1 - \frac{2}{C} - e^{-(1/C-\epsilon)D}\right)(\text{RHS}). \tag{7.34}$$

Taking, e.g., $C = 1/2\epsilon$ then gives the result, where rearranging our assumption that $D \geq (1/\epsilon)\log(1/\epsilon)$ we have $e^{-\epsilon D} \leq \epsilon$. □

7.4.4 Step 3B: SOS Version of Verifiability

The remaining ingredient of the analysis that we are missing is the following, which is in some sense an SOS analog of the "verifiability" property from Proposition 7.2.

Lemma 7.12. *For every $\epsilon > 0$, there exists $C > 0$ such that the following holds. Let $\widetilde{\mathbb{E}}$ be a degree $4D$ pseudoexpectation with $D = C \log n$ satisfying the constraints $\|\boldsymbol{x}\|^2 = 1$ and $p(\boldsymbol{x}) \geq 1 - 1/\mathrm{polylog}(n)$. Then, with high probability, there exists $i \in [r]$ such that $\widetilde{\mathbb{E}}[\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^D] \geq \exp(-\epsilon D)$.*

Intuitively, we should think of $\widetilde{\mathbb{E}} = \widetilde{\mathbb{E}}^\star$ the optimizer of the degree $4D$ Lasserre relaxation of Opt; we will explain below why the result needs to be stated slightly differently to facilitate a proof.

The crux of the matter is the following: since the objective value of $\widetilde{\mathbb{E}}^\star$ must be at least $p(\boldsymbol{a}_i) \approx 1$, we must have $\widetilde{\mathbb{E}}^\star[p(\boldsymbol{x})] = \widetilde{\mathbb{E}}^\star[\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^3] \gtrsim 1$. The difficulty is to show that this also implies, via an SOS proof, that $\widetilde{\mathbb{E}}^\star[\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^D] \gtrsim 1$. If we could do this, then by linearity there would exist some $i \in [r]$ for which $\widetilde{\mathbb{E}}^\star[\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^D] \gtrsim 1/r \geq e^{-O(\log n)}$, as desired.

This is perhaps the subtlest point of the analysis of [GM15], so we will not fully treat the details, but will give the main ideas. The tool at our disposal is the Hölder-like inequality of Exercise 7.1, which gives that, for $(k-2) \mid D$ and $kD/(k-2)$ an even number,

$$\|\boldsymbol{v}\|_D^D \left(\|\boldsymbol{v}\|_2^2\right)^{\frac{D}{k-2}} - \left(\|\boldsymbol{v}\|_k^k\right)^{\frac{D}{k-2}} \in \mathsf{SOS}. \tag{7.35}$$

Writing $\boldsymbol{A}$ for the matrix whose columns are the $\boldsymbol{a}_i$, we are interested in such an inequality with $\boldsymbol{v} = \boldsymbol{A}^\top \boldsymbol{x}$. Making this substitution and taking pseudoexpectations, we find something close to what we want if we take $k = 3$ and D a large even number:

$$\widetilde{\mathbb{E}}^\star \left[\left(\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^D \right) \left(\|\boldsymbol{A}^\top \boldsymbol{x}\|_2^2 \right)^D \right] \geq \widetilde{\mathbb{E}}^\star \left[\left(\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^3 \right)^{\frac{D}{k-2}} \right] \geq \left(\widetilde{\mathbb{E}}^\star \left[\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^3 \right] \right)^{\frac{D}{k-2}}, \tag{7.36}$$

the latter operation being by the SOS Jensen inequality.

The only problem is the remaining term on the left-hand side: we have $\|\boldsymbol{A}^\top \boldsymbol{x}\|_2^2 = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{x}$, and for $\boldsymbol{A}$ random under our model we expect $\boldsymbol{A}\boldsymbol{A}^\top \approx \frac{r}{n}\boldsymbol{I}_n$. Thus on the right-hand side we will have an extra factor of $\mathrm{poly}(n)^D = n^{\Omega(\log n)}$, thwarting our efforts.

To get around this, [GM15] go one step further: instead of evaluating (7.35) with $\boldsymbol{v} = \boldsymbol{A}^\top \boldsymbol{x}$, they evaluate it with the *coordinatewise square* of this vector. That gives, by the same manipulations,

$$\widetilde{\mathbb{E}}^\star \left[\left(\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^{2D} \right) \left(\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^4 \right)^D \right] \geq \left(\widetilde{\mathbb{E}}^\star \left[\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^6 \right] \right)^{\frac{D}{k-2}}. \tag{7.37}$$

Now, at least at an intuitive level, the situation is better: when $r \ll n^{3/2}$, we also have $r \ll n^2$ and $r \ll n^3$, so by the result of Proposition 7.2 for $p = 4, 6$, we expect that for $\widetilde{\mathbb{E}}^\star$ having $\widetilde{\mathbb{E}}^\star \sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^3 \approx 1$ we should also have $\widetilde{\mathbb{E}}^\star \sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^4, \widetilde{\mathbb{E}} \sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^6 \approx 1$, as we hope.

To actually implement this as an SOS argument requires a little bit more care: at this point, as we have done in the statement of Lemma 7.12, it is useful to adjust the optimization problem Opt to add $\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^3 \geq 1 - \gamma$ for some small γ as an additional constraint. That way, it suffices to produce SOS proofs deriving, from this constraint and $\|\boldsymbol{x}\|_2^2 = 1$, bounds

of the form

$$1 - \gamma' \le \sum_{i=1}^{r} \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle^4 \le 1 + \gamma', \tag{7.38}$$

$$1 - \gamma' \le \sum_{i=1}^{r} \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle^6 \le 1 + \gamma' \tag{7.39}$$

for some other small γ'. This is the main technical step in [GM15], using some rather intricate matrix concentration arguments. These are in the vein of, though more complicated than, what we have seen in Chapter 6; we will not give any more details here.

7.4.5 Step 3C: Full Quasipolynomial Time SOS Algorithm

We now put together the pieces of our putative algorithm.

Theorem 7.13. *Consider the following algorithm that takes $\boldsymbol{T}$ as input and outputs $\hat{\boldsymbol{a}} \in \mathbb{S}^{n-1}$: first, for a large absolute constant C, let $\widetilde{\mathbb{E}}^\star$ be the optimizer of the degree $4D$ Lasserre relaxation of* Opt. *Then, repeatedly set $\hat{\boldsymbol{a}} := \mathsf{BJ}(\widetilde{\mathbb{E}}[\boldsymbol{x}^{\otimes D}])$ until $p(\hat{\boldsymbol{a}}) = T[\hat{\boldsymbol{a}}, \dots, \hat{\boldsymbol{a}}] \ge 0.99$. Suppose $r \le n^{3/2}/\mathsf{polylog}(n)$. Then, this algorithm has the following properties:*

1. *With high probability, it terminates in quasipolynomial time $n^{O(\log n)}$.*
2. *With high probability, its output $\hat{\boldsymbol{a}}$ has $\langle \hat{\boldsymbol{a}}, \boldsymbol{a}_i \rangle \ge 0.9$ for some $i \in [r]$.*

The proof follows easily by combining our previous results.

Proof of Theorem 7.13. First, note that solving the requisite Lasserre relaxation takes time $n^{O(\log n)}$. Plugging Lemma 7.12 into Lemma 7.11, we find that, with high probability over the randomness in $\boldsymbol{T}$, there exists $i \in [r]$ such that

$$\mathbb{P}[\langle \mathsf{BJ}(\widetilde{\mathbb{E}}^\star[\boldsymbol{x}^{\otimes D}]), \boldsymbol{a}_i \rangle^2 \ge 0.999] \ge n^{-B} \tag{7.40}$$

for some $B > 0$, where this probability is over the randomness in the computation of $\mathsf{BJ}(\cdot)$. Thus, in $\mathsf{poly}(n)$ trials of computing $\hat{\boldsymbol{a}} := \mathsf{BJ}(\widetilde{\mathbb{E}}^\star[\boldsymbol{x}^{\otimes D}])$, each of which also takes time $n^{O(\log n)}$, we will find some $\hat{\boldsymbol{a}}$ with $\boldsymbol{T}[\hat{\boldsymbol{a}}, \hat{\boldsymbol{a}}, \hat{\boldsymbol{a}}] \ge 0.99$ by Proposition 7.2, giving the desired result. □

We note that the algorithm actually relies *both* on the original verifiability property, Proposition 7.2, and on its SOS variant, Lemma 7.12.

This algorithm has three apparent defects: first, it only runs in quasipolynomial rather than polynomial time; second, it only outputs one component $\hat{\boldsymbol{a}}$ rather than a full tensor decomposition; and third, the $\hat{\boldsymbol{a}}$ it outputs is only weakly correlated to an actual tensor component. The first issue takes significant further work to resolve, and we will discuss it in the next section. The latter two issues are easier to handle; we will not give a detailed analysis here, but let us sketch the basic ideas.

To handle the second issue, instead of just producing one approximate tensor component, we assemble a list $\mathcal{A} \subset \mathbb{S}^{n-1}$. While solving the Lasserre relaxation to find a new vector

to add to $\mathcal{A}$, we add the constraints $\{\langle a, x\rangle^2 \le \delta\}_{a\in\mathcal{A}}$ to the relaxation, for some small constant $\delta > 0$. At the end, $\mathcal{A} = \{\hat{a}_1, \dots, \hat{a}_r\}$ (provided we show that it is possible to repeat this procedure successfully for r steps) where, with high probability, $T[\hat{a}_i, \dots, \hat{a}_i] \ge 0.99$ for each $i \in [r]$ and $\langle \hat{a}_i, \hat{a}_j\rangle^2 \le \delta'$ for all $i \ne j$ for some other small, though larger than δ, constant $\delta' > 0$ (which also requires proof but should be intuitive). And to handle the third issue, there is a different fast algorithm for "refining" a weak estimate achieving some fairly large correlation (0.9 in our case) with the true tensor components to an estimate achieving an arbitrarily large correlation. This algorithm, analyzed by [AGJ14] in our setting, is a tensor variant of *power iteration* for matrices, and sets

$$\hat{a}_i := \frac{T[\hat{a}_i, \hat{a}_i, \cdot]}{\|T[\hat{a}_i, \hat{a}_i, \cdot]\|} \tag{7.41}$$

for each $i \in [r]$ repeatedly. Essentially, this work and related ones (see the chapter notes) imply that the main challenge of tensor decomposition is to find an initial "warm start" approximation as Theorem 7.13 provides; from there, straightforward local algorithms suffice to improve the approximation and converge to the true decomposition.

7.5 Polynomial Time with Jennrich Algorithm

Finally, let us sketch how the subsequent work of [MSS16] improved the quasipolynomial time algorithm we have given to a genuinely polynomial time one. The key idea is to use a rounding procedure much closer to the actual Jennrich algorithm rather than the baby Jennrich algorithm. They do not quite use the full Jennrich procedure we outlined, but rather the following slight variant.

Definition 7.14 (Spectral Jennrich algorithm)**.** *Given a tensor $\widetilde{T} \in (\mathbb{R}^n)^{\otimes 2d+1}$, we define the randomized algorithm $\mathsf{SJ}(\widetilde{T})$ to output the top eigenvector of the matrix $M := \widetilde{T}[g, \cdot, \dots, \cdot] \in \mathbb{R}^{n^d \times n^d}$ (where we reshape the partially contracted tensor appropriately) where $g \sim \mathcal{N}(0, I_n)$.*

This is just the higher-order variant of the Jennrich algorithm, adjusted to only build one matrix M and optimistically output the top eigenvector, assuming that enough bias will be introduced by the contraction with g that this eigenvector will be close to one of the a_i. The basic plan is again to repeatedly attempt this rounding $\hat{a} := \mathsf{SJ}(\widetilde{\mathbb{E}}^\star[x^{\otimes(2d+1)}])$ until a good estimate (verifying by finding that $p(\hat{a})$ is large) is obtained; it turns out that, in this scheme, it suffices to take $d = O(1)$, and correspondingly each rounding takes only polynomial time, and as before polynomially many rounding attempts suffice with high probability.

The more relevant issue, however, is that, even before forming an SOS relaxation, it is *not* the case that this rounding applied to high degree moments of a distribution achieving a large objective value in Opt will necessarily output a vector close to one of the a_i. For the sake of simplicity, we illustrate this with low degree, but it is straightforward to show that the same issue persists for higher degrees.

Proposition 7.15. *Let $a_1 = e_1, \dots, a_r = e_r$ for $r = n - 1$. Then, there exists a probability measure $\mu \in \mathcal{M}(\mathbb{S}^{n-1})$ satisfying the following properties:*

1. $\mathbb{E}_{x\sim\mu} \sum_{i=1}^r \langle a_i, x\rangle^3 = 1 - o(1)$.

2. *With probability at least* $1 - \exp(-n^\delta)$ *for some* $\delta > 0$, $\widehat{\boldsymbol{a}} = \mathsf{SJ}(\mathbb{E}_{\boldsymbol{x}\sim\mu}[\boldsymbol{x}^{\otimes 3}])$ *has* $\langle \widehat{\boldsymbol{a}}, \boldsymbol{a}_i\rangle^2 = o(1)$ *for all* $i \in [r]$.

Proof. We may take $\mu = \mathsf{Unif}(\{\sqrt{1-\epsilon^2}\boldsymbol{e}_i + \epsilon \boldsymbol{e}_n\}_{i=1}^{n-1})$. Then, loosely speaking, $\boldsymbol{M}$ typically has $\Theta(r\epsilon^3)$ mass on $\boldsymbol{e}_n\boldsymbol{e}_n^\top$ and only $O(1)$ in any other direction. Thus, since $r \sim n$, taking, e.g., $\epsilon = n^{-1/4}$ will give the desired behavior. □

In order for the spectral Jennrich algorithm to work well, we in fact *need* a kind of "spectral uniformity" property on μ. That the example above will cause problems, for example, may be ascribed to its having $\|\mathbb{E}_{\boldsymbol{x}\sim\mu}\boldsymbol{x}\boldsymbol{x}^\top\| \gg \frac{1}{r}$, while μ uniform over $\boldsymbol{a}_1, \dots, \boldsymbol{a}_r$, which would still achieve Condition 1, would have $\|\mathbb{E}_{\boldsymbol{x}\sim\mu}\boldsymbol{x}\boldsymbol{x}^\top\| = \frac{1}{r}$ as well, and would not exhibit the same behavior.

In [MSS16], this kind of property is referred to as high "spectral entropy" of μ, and it is shown that the spectral Jennrich rounding works well on measures of high spectral entropy when the tensor components are close to orthonormal. When translating the same idea to an SOS-based algorithm, the issue arises of how to enforce high spectral entropy on pseudomoment matrices. Fortunately, the flexibility of semidefinite programming helps us: to impose constraints of the form

$$\|\widetilde{\mathbb{E}}[\boldsymbol{x}^{\otimes k}\boldsymbol{x}^{\otimes k^\top}]\| \le C, \tag{7.42}$$

we may simply add the semidefinite constraints

$$-C\boldsymbol{I} \preceq \widetilde{\mathbb{E}}[\boldsymbol{x}^{\otimes k}\boldsymbol{x}^{\otimes k^\top}] \preceq C\boldsymbol{I} \tag{7.43}$$

to the SOS semidefinite program, a further semidefinite constraint on a principal submatrix of our decision variable. Implementing this idea, [MSS16] produce an improved version of our Theorem 7.13 that gives polynomial runtime.

Exercises

Exercise 7.1. *Show that the following inequalities admit SOS proofs.*

1. *(AM-GM inequality) Prove that, for any* $n \ge 1$,

$$\frac{x_1^{2n} + \cdots + x_n^{2n}}{n} - x_1^2 \cdots x_n^2 \in \mathsf{SOS}. \tag{7.44}$$

2. *(Hölder inequality variant) Suppose that* $(k-2) \mid D$ *with* $kD/(k-2)$ *an even integer. Show that*

$$\|\boldsymbol{x}\|_D^D \left(\|\boldsymbol{x}\|_2^2\right)^{\frac{D}{k-2}} - \left(\|\boldsymbol{x}\|_k^k\right)^{\frac{D}{k-2}} \in \mathsf{SOS}. \tag{7.45}$$

(This is an SOS-friendly rearrangement of the Hölder inequality $\|\boldsymbol{x}\|_k \le \|\boldsymbol{x}\|_1^{1-2/k}\|\boldsymbol{x}\|_2^{2/k} \le \|\boldsymbol{x}\|_D^{1-2/k}\|\boldsymbol{x}\|_2^{2/k}$*, where the latter holds since* $\|\boldsymbol{x}\|_1 \le \|\boldsymbol{x}\|_D$ *for any* $D \ge 1$*.)*

HINT: *For Part 1, let S_n denote the permutations of $[n]$, and write $\sigma(i)$ for $i \in [n]$ and $\sigma \in S_n$ as the image of i under σ. Define $p_k(\boldsymbol{x}) := \frac{1}{n!}\sum_{\sigma\in S_n} x_{\sigma(1)}^{2(n-k)} x_{\sigma(2)}^2 \cdots x_{\sigma(2+k-1)}^2$ for $k = 0,\dots,n-1$. Show that the left-hand side above is $p_0(\boldsymbol{x}) - p_{n-1}(\boldsymbol{x})$, and show that $p_{k-1}(\boldsymbol{x}) - p_k(\boldsymbol{x}) \in \mathsf{SOS}$ for each $k \in [n-1]$. For Part 2, use Part 1.*

NOTES

OTHER SOURCES As mentioned before, we have followed the refined analysis of [GM15] of [BKS15] in this chapter. A high-level overview of the same is included in [BS16] and [Moi20]; more details, especially on the improvement to a polynomial-time algorithm from [MSS16], may be found in Shi's thesis [Shi19].

APPLICATIONS Perhaps the main application of tensor decomposition algorithms is to the *method of moments*, a common paradigm in statistics. We might, for example, observe samples of some $\mu = \frac{1}{r}\sum_{i=1}^r \delta_{a_i}$, $\boldsymbol{y}_1,\dots,\boldsymbol{y}_m \sim \mu$, use this to form an empirical approximation of the moment tensor of μ, $\mathbb{E}_{x\sim\mu}\boldsymbol{x}^{\otimes p} = \frac{1}{r}\sum_{i=1}^r \boldsymbol{a}_i^{\otimes p} \approx \frac{1}{m}\sum_{j=1}^m \boldsymbol{y}_j^{\otimes p}$, and then apply tensor decomposition to this estimate to try to identify the $\boldsymbol{a}_i$. This requires a generalization of our analysis to the situation where we are given not $\boldsymbol{T} = \sum_{i=1}^r \boldsymbol{a}_i^{\otimes p}$, but a perturbation $\boldsymbol{T} + \boldsymbol{\Delta}$ where $\boldsymbol{\Delta}$ is "small" in some suitable sense. See, e.g., [BKS15] for such an application to dictionary learning, [Hop18a] for a similar approach to Gaussian mixture models, or [AGH+14, AGJ15] to "latent variable" statistical models. (These applications have different names in the statistics literature and slightly different assumptions and settings, but when formulated mathematically are ultimately very similar.)

OTHER TENSOR DECOMPOSITION ALGORITHMS Many algorithms other than SOS-based ones have been proposed in the literature for tensor decomposition. As mentioned in the main text, [AGJ14] analyzed tensor power iteration and showed that it successfully finds a good decomposition from a sufficiently "warm" initialization close to the true components. Indeed, the same holds for naive gradient descent; [GM17] (albeit only for 4-tensors) that in the region where $\sum_{i=1}^r \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle^4$ is slightly greater than the typical value of a random $\boldsymbol{x} \in \mathbb{S}^{n-1}$, the only local optima of this function of $\boldsymbol{x}$ are near $\pm\boldsymbol{a}_i$.

There are also some more elaborate variants of the idea of Jennrich's algorithm, such as the "FOOBI algorithm" specifically for $p = 4$ developed by [DLCC07] and achieving the $r \ll n^{p/2} = n^2$ threshold in this case; see also [MSS16] for discussion of how FOOBI may be thought of within the SOS framework.

In a different line of work, [HSSS16] adapted the analysis of SOS to produce a *spectral* algorithm for tensor decomposition, that builds a large matrix whose top eigenvector gives a good estimate of a tensor component. However, this result only worked for $r \le n^{4/3}/\mathsf{polylog}(n)$; more recently, [DdL+22] improved this approach to the optimal $r \le n^{3/2}/\mathsf{polylog}(n)$ by adjusting the construction of the requisite matrix.

TENSOR ORDER DEPENDENCE There is an unpleasant *ad hoc* dependence of some of the results we have seen on the tensor order p: the results of [GM15] only concern $p = 3$, the

results of [MSS16] concern $p = 3$ for the model of $\boldsymbol{a}_i$ random but $p = 4$ for the smoothed analysis model, the results of [GM17] on the optimization landscape and gradient descent only study $p = 4$ for the sake of convenience, and so forth. We would of course like a more unified picture for all p, but there appear to be many technical obstacles to doing this for the near-optimal SOS and spectral algorithms.

Open Problem 7.4 (Decomposition of tensors of general order)**.** *Unify and generalize the analyses and algorithms cited above to arbitrary tensor orders $p \geq 3$.*

8 | Case Study 3: Heavy-Tailed Mean Estimation

The last problem for which we will consider SOS algorithms is an extension of a very basic task in statistics, that of estimating the mean of an unknown distribution. As recent literature has shown, however, even this simple problem holds some interesting surprises, and we will see how SOS is useful for solving this problem effectively even for very poorly-behaved distributions.

8.1 Scalar Mean Estimation

The simple setting of the problem we consider is as follows: suppose $\rho \in \mathcal{M}(\mathbb{R})$ with $\mathbb{E}_{x\sim\rho}[x^2] < \infty$ and denote

$$\mu := \mathbb{E}_{x\sim\rho}[x], \tag{8.1}$$

$$\sigma^2 := \mathsf{Var}_{x\sim\rho}[x]. \tag{8.2}$$

Suppose that $x_1, \ldots, x_n \sim \rho$ independently. We are interested in estimating μ from these observations, i.e., in producing an *estimator* $\hat{\mu} : \mathbb{R}^n \to \mathbb{R}$ so that $\hat{\mu}(x_1, \ldots, x_n)$ that is typically close to μ.[1]

The simplest, oldest, and most standard idea for accomplishing this is the *empirical mean estimator*:

$$\hat{\mu}^{\mathsf{emp}} := \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{8.3}$$

This estimator also comes with a well-known *asymptotic* theory: by the central limit theorem, we have that $\sqrt{n}(\hat{\mu}^{\mathsf{emp}} - \mu)$ converges in distribution to $\mathcal{N}(0, \sigma^2)$, and in particular

$$\lim_{n\to\infty} \mathbb{P}\left[|\hat{\mu}^{\mathsf{emp}} - \mu| \geq \frac{t}{\sqrt{n}}\right] = \mathop{\mathbb{P}}_{g\sim\mathcal{N}(0,\sigma^2)}[|g| \geq t] \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{8.4}$$

Often this asymptotic relation is used as justification for *confidence intervals* for an estimate of μ: if we ignore the limit, then the above suggests that, with probability $1 - \delta$, we should have

$$\mu \in \left[\hat{\mu}^{\mathsf{emp}} - \sqrt{\frac{2\sigma^2 \log(1/\delta)}{n}}, \hat{\mu}^{\mathsf{emp}} + \sqrt{\frac{2\sigma^2 \log(1/\delta)}{n}}\right]. \tag{8.5}$$

[1] For us, "an estimator" is always actually a series of estimators for different sample sizes n.

However, the central limit theorem alone does not actually guarantee that this is the case. It is therefore of interest in statistics to ask what sizes of confidence intervals various estimators do actually achieve, a *non-asymptotic* question. The following terminology is not standard (as far as I know) but will be helpful.

Definition 8.1. *We say that an estimator $\hat{\mu}$ has* width $t = t(n, \delta)$ *if, for all $n \geq 1$ and $\delta > 0$,*

$$\mathbb{P}\left[|\hat{\mu}(x_1, \ldots, x_n) - \mu| > t(n, \delta)\right] \leq \delta. \tag{8.6}$$

We note that, in the special case of ρ a centered Gaussian distribution, the central limit theorem actually holds exactly for all n, so (8.5) indeed holds with probability at least $1 - \delta$. In fact, any subgaussian distribution shares the same behavior. For the sake of simplicity, from now on we assume $\sigma^2 = 1$ (and likewise for subgaussian variance proxies), but the dependence of various widths we discuss on σ^2 is easy to recover.

Proposition 8.2. *If ρ is 1-subgaussian, then $\hat{\mu}^{\mathrm{emp}}$ has width*

$$t \lesssim \sqrt{\frac{\log(1/\delta)}{n}}. \tag{8.7}$$

Proof. If ρ is 1-subgaussian, then $\hat{\mu}^{\mathrm{emp}}$ is $(1/n)$-subgaussian. In particular, we have the tail bound

$$\mathbb{P}[|\hat{\mu}^{\mathrm{emp}} - \mu| \geq t] \leq 2 \exp\left(-\frac{nt^2}{2}\right), \tag{8.8}$$

and rearranging gives the result. □

For this reason, we call (8.7) the *subgaussian width*.

It is natural to ask if we can achieve the same under weaker assumptions—indeed, the central limit theorem that motivated us holds so long as $\mathbb{E}_{x\sim\rho}[x^2] < \infty$, so it is reasonable to consider this much weaker condition.

Proposition 8.3. *If $\sigma^2 = 1$, then $\hat{\mu}^{\mathrm{emp}}$ has width*

$$t \lesssim \sqrt{\frac{1/\delta}{n}}. \tag{8.9}$$

Proof. If $\sigma^2 = 1$, then $\mathrm{Var}[\hat{\mu}^{\mathrm{emp}} - \mu] = 1/n$. In particular, by Chebyshev's inequality,

$$\mathbb{P}[|\hat{\mu}^{\mathrm{emp}} - \mu| \geq t] \leq \frac{1}{nt^2}, \tag{8.10}$$

and rearranging gives the result. □

In Exercise 8.1 you will show that this is optimal under the second moment assumption.

Surprisingly, however, this is not the end of the story. With the important caveat that we allow $\hat{\mu}$ to depend on the error probability δ, it is actually possible to do much better.

Theorem 8.4 ([Cat12])**.** *There is an estimator $\hat{\mu}^{\mathrm{mom}} = \hat{\mu}^{\mathrm{mom}}_{\delta}$ having subgaussian width for any ρ with $\sigma^2 = 1$. Moreover, $\hat{\mu}^{\mathrm{mom}}$ may be computed from $x_1, \ldots, x_n$ in linear time $O(n)$.*

It is known to be necessary that the estimator depend on δ for such an improvement; see [Cat12].

Definition 8.5. *The* median-of-means estimator $\hat{\mu}^{\mathsf{mom}} = \hat{\mu}_k^{\mathsf{mom}}$ *is defined as follows. For the sake of simplicity, suppose* $k \mid n$, *and set* $m := n/k$. *Partition* $[n]$ *into* $B_1, \dots, B_k$ *with* $|B_i| = m$. *Then, we take*

$$\hat{\mu}^{\mathsf{mom}} := \mathsf{med}\left(\left\{\frac{1}{m}\sum_{j\in B_i} x_j\right\}_{i=1}^{k}\right), \tag{8.11}$$

where $\mathsf{med}(\cdot)$ *is the median.*

In our situation we will take $k = k(\delta)$, producing the dependence on δ mentioned above.

The following is the main property of the median that we will use. We isolate it because, while in one dimension up to inconsequential "tie-breaking" changes this property uniquely characterizes the median, later we will be interested in higher-dimensional generalized medians, which admit subtle variations while still respecting a similar rule.

Proposition 8.6 (Majority property of median)**.** *Let* $y_1, \dots, y_k \in \mathbb{R}$ *and* $\nu := \mathsf{med}(y_1, \dots, y_k)$. *The following then hold:*

1. *If at least* $\frac{1}{2}k$ *of the* y_i *have* $y_i \geq s$, *then* $\nu \geq s$.
2. *If at least* $\frac{1}{2}k$ *of the* y_i *have* $y_i \leq s$, *then* $\nu \leq s$.
3. *If at least* $\frac{1}{2}k$ *of the* y_i *have* $|y_i - s| \leq t$, *then* $|\nu - s| \leq t$.

Proof. The first two follow because, by definition of the median, at least $\frac{1}{2}k$ of the y_i must have $y_i \leq \nu$ and at least $\frac{1}{2}k$ must have $y_i \geq \nu$. The third follows from the first two since $|a - s| \leq t$ is equivalent to $a \geq s - t$ and $a \leq s + t$. □

Proof of Theorem 8.4. Let $y_i := \frac{1}{m}\sum_{j\in B_i} x_j$ for $i \in [k]$. The main idea of the proof is that, while a few of the B_i might contain "outlier" samples from the tails of ρ that cause y_i to be poor estimates of μ, most of the y_i are good estimates.

More precisely, first, by the same Chebyshev's inequality calculation as in Proposition 8.3,

$$\mathbb{P}[|y_i - \mu| \geq t] \leq \frac{1}{mt^2}. \tag{8.12}$$

Note that these events are independent. Thus, we have

$$\begin{aligned}\mathbb{P}[|\hat{\mu}^{\mathsf{mom}} - \mu| \geq t] &\leq \mathbb{P}\left[\text{at least } \frac{k}{2} \text{ of the } y_i \text{ have } |y_i - \mu| \geq t\right] && \text{(Proposition 8.6)}\\ &\leq \mathop{\mathbb{P}}_{N\sim\mathsf{Bin}(k,1/mt^2)}\left[N \geq \frac{k}{2}\right]\end{aligned}$$

and, so long as $1/mt^2$ is bounded away from $\frac{1}{2}$, e.g., so long as $t \leq \sqrt{4/m} = \sqrt{4k/n}$, we have by Hoeffding's inequality

$$= \exp(-\Omega(k)). \tag{8.13}$$

Now, choosing $k = C\log(1/\delta)$ for a suitable C makes the last line equal δ, whereby we achieve a width of

$$t \le \sqrt{\frac{4k}{n}} \lesssim \sqrt{\frac{\log(1/\delta)}{n}}, \tag{8.14}$$

as desired. Exercise 8.2 shows that $\hat{\mu}^{\mathrm{mom}}$ can be computed in linear time as well. □

While there are some interesting more intricate questions left to consider about the scalar case (see the chapter notes), this is the end of the basic story. We now move on to the higher-dimensional case, where the situation is far subtler.

8.2 Vector Mean Estimation

The adjustment of our setting to d dimensions is straightforward: suppose $\rho \in \mathcal{M}(\mathbb{R}^d)$ with $\mathbb{E}_{x\sim\rho}[\|x\|^2] < \infty$ and denote

$$\boldsymbol{\mu} := \mathbb{E}_{x\sim\rho}[x]. \tag{8.15}$$

Generalizing our assumption above that $\sigma^2 = 1$, let us make an *isotropy* assumption to lighten the notation, supposing that

$$\mathrm{Cov}_{x\sim\rho}[x] = \mathbb{E}_{x\sim\rho}[xx^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top = \boldsymbol{I}_d. \tag{8.16}$$

Suppose that $x_1, \dots, x_n \sim \rho$ are i.i.d.; we then want to estimate $\boldsymbol{\mu}$ from these observations. The standard estimator is again

$$\hat{\boldsymbol{\mu}}^{\mathrm{emp}} = \hat{\boldsymbol{\mu}}^{\mathrm{emp}}(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{8.17}$$

To identify the analog of the subgaussian rate in this setting, consider $\rho = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$. We have $\boldsymbol{\mu} = \mathbf{0}$, so

$$\begin{aligned}\mathbb{P}[\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}} - \boldsymbol{\mu}\| \ge t] &= \mathbb{P}[\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}}\| \ge t] \\ &= \mathbb{P}[\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}}\| \ge \mathbb{E}\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}}\| + (t - \mathbb{E}\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}}\|)]\end{aligned}$$

and, viewing $\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}}\|$ as a function of the dn i.i.d. Gaussian random variables $(x_1, \dots, x_n)$, we have that this is a $\frac{1}{\sqrt{n}}$-Lipschitz function (see Exercise 8.3), whereby by Gaussian Lipschitz concentration we find

$$\le \exp(-2n(t - \mathbb{E}\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}}\|)^2). \tag{8.18}$$

Noting that, by Jensen's inequality and since the law of $\hat{\boldsymbol{\mu}}^{\mathrm{emp}}$ is $\mathcal{N}(\mathbf{0}, \frac{1}{n}\boldsymbol{I}_d)$,

$$\mathbb{E}\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}}\| \le (\mathbb{E}\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}}\|^2)^{1/2} = \sqrt{\frac{d}{n}}, \tag{8.19}$$

we find that the width of $\hat{\boldsymbol{\mu}}^{\mathrm{emp}}$ in the Gaussian case is

$$t \lesssim \mathbb{E}\|\hat{\boldsymbol{\mu}}^{\mathrm{emp}}\| + \sqrt{\frac{\log(1/\delta)}{n}} \le \frac{\sqrt{d} + \sqrt{\log(1/\delta)}}{\sqrt{n}}. \tag{8.20}$$

We note that this is identical to the one-dimensional subgaussian width, only with the $\sqrt{d}$ in the numerator representing the cost of higher dimensionality.

As in the one-dimensional case, it is possible with a bit more care to derive the same width for a natural definition of "subgaussian vectors." It is also the case that the above fails for heavy-tailed ρ. We will not go into these details here, and instead proceed directly to trying to achieve subgaussian width under general assumptions.

8.3 Strong Median Estimator

How can we generalize the one-dimensional median-of-means estimator to higher dimensions? It is natural to again form buckets $B_1, \dots, B_k$ of the $\boldsymbol{x}_i$, to take means $\boldsymbol{y}_i := \frac{1}{m}\sum_{j \in B_i} \boldsymbol{x}_j$, and to look for a suitable notion of "the median of the $\boldsymbol{y}_i$."

Reviewing our proof of Theorem 8.4, we see that the only property of the median we actually used was Property 3 of Proposition 8.6:

> "If at least $\frac{1}{2}k$ of the y_i have $|y_i - s| \le t$, then $|\mathsf{med}(y_1, \dots, y_k) - s| \le t$."

Not only that, but while for the one-dimensional median this holds for *all* t, we actually only needed to use that this property holds for a *specific* t chosen in our algorithm—thanks to the dependence of the algorithm on the confidence level δ. That motivates the following high-dimensional median.

Definition 8.7. *We say that $\boldsymbol{\nu}$ is a* t-median *of $\boldsymbol{y}_1, \dots, \boldsymbol{y}_k \in \mathbb{R}^d$ if, for strictly more than $\frac{1}{2}k$ of the $\boldsymbol{y}_i$, we have $\|\boldsymbol{y}_i - \boldsymbol{\nu}\| \le t$.*

While t-medians are not unique, they cannot be too far from one another.

Proposition 8.8. *If $\boldsymbol{\nu}, \boldsymbol{\nu}'$ are both t-medians of $\boldsymbol{y}_1, \dots, \boldsymbol{y}_k$, then $\|\boldsymbol{\nu} - \boldsymbol{\nu}'\| \le 2t$.*

Proof. By the pigeonhole principle, there is some $\boldsymbol{y}_i$ such that $\|\boldsymbol{\nu} - \boldsymbol{y}_i\| \le t$ and $\|\boldsymbol{\nu}' - \boldsymbol{y}_i\| \le t$. The result then follows by the triangle inequality. □

In one dimension, a softer version of our argument for Theorem 8.4 goes as follows: on the one hand, $\mathsf{med}(y_1, \dots, y_k)$ is a t-median of $y_1, \dots, y_k$. On the other hand, we showed that with high probability μ itself is a t-median of $y_1, \dots, y_k$. Thus, with the same probability $|\mu - \mathsf{med}(y_1, \dots, y_k)| \le 2t$.

The benefit of rephrasing the argument in this way is that we may repeat it essentially verbatim to get a result in higher dimensions, albeit one with suboptimal width.

Theorem 8.9. *Let $\hat{\boldsymbol{\mu}}$ output a t-median $\boldsymbol{y}_1, \dots, \boldsymbol{y}_k$ if one exists, or an arbitrary point (say, $\mathbf{0}$) if not. Then, for any isotropic ρ, this estimator achieves width*

$$t \lesssim \frac{\sqrt{d \log(1/\delta)}}{\sqrt{n}}. \tag{8.21}$$

Proof. By Chebyshev's inequality, we have

$$\mathbb{P}[\|\boldsymbol{y}_i - \boldsymbol{\mu}\| \geq t] \leq \frac{d}{mt^2}. \tag{8.22}$$

As in Theorem 8.4, we choose $k \sim \log(1/\delta)$ and $t \sim \sqrt{d/m} \sim \sqrt{d\log(1/\delta)/n}$ to make this probability at most, say, $1/4$. Then, for a suitable choice of the constants above we have

$$\mathbb{P}[\boldsymbol{\mu} \text{ is a } t\text{-median}] \geq \mathbb{P}\left[\text{more than } \frac{k}{2} \text{ of the } \boldsymbol{y}_i \text{ have } \|\boldsymbol{y}_i - \boldsymbol{\mu}\| \leq t\right] \geq 1 - \delta. \tag{8.23}$$

Whenever $\boldsymbol{\mu}$ is a t-median then $\hat{\boldsymbol{\mu}}$ is also a t-median, since some t-median (namely $\boldsymbol{\mu}$) exists. In this case, by Proposition 8.8, $\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| \leq 2t$, completing the proof. □

Finally, while it may not be immediately obvious, we note that, up to constants, it is possible to efficiently compute such an estimator.

Proposition 8.10. *There is an estimator $\hat{\boldsymbol{\mu}}'$ that achieves the same (up to constants) width as that of $\hat{\boldsymbol{\mu}}$ and can be computed in time $O(n + k^2) = O(n^2)$.*

Proof. We take $\hat{\boldsymbol{\mu}}'$ to search for a $2t$-median among the $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ themselves, which only requires scanning through the $k \times k$ matrix of their mutual distances once. If one exists, then $\hat{\boldsymbol{\mu}}'$ outputs it, and otherwise outputs an arbitrary point.

Whenever a t-median $\boldsymbol{\nu}$ exists, then any $\boldsymbol{y}_i$ with $\|\boldsymbol{y}_i - \boldsymbol{\nu}\| \leq t$ is a $2t$-median, since for any $\boldsymbol{y}_j$ with $\|\boldsymbol{y}_j - \boldsymbol{\nu}\| \leq t$ we have $\|\boldsymbol{y}_i - \boldsymbol{y}_j\| \leq 2t$ by triangle inequality. Thus, whenever $\boldsymbol{\mu}$ is a t-median of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, then in particular a t-median exists, so $\hat{\boldsymbol{\mu}}'$ outputs a $2t$-median. On the other hand, in this same case $\boldsymbol{\mu}$ is also a $2t$-median, so $\|\hat{\boldsymbol{\mu}}' - \boldsymbol{\mu}\| \leq 4t$. □

However, this estimator does not quite achieve subgaussian width. Can we do better?

8.4 Lugosi-Mendelson Weak Median Estimator

At the cost of computational tractability, [LM19] showed that this is indeed possible. The key is to loosen our notion of high-dimensional t-median so that Proposition 8.8 still holds, but so that μ is a t-median for smaller t. We follow the interpretation of this definition given in [Hop18c], where it is credited to Jerry Li.

Definition 8.11. *We say that $\boldsymbol{\nu}$ is* t-central *for $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k \in \mathbb{R}^d$ if, for all $\boldsymbol{u} \in \mathbb{S}^{d-1}$, for strictly more than $\frac{1}{2}k$ of the $\boldsymbol{y}_i$, we have $|\langle \boldsymbol{y}_i - \boldsymbol{\nu}, \boldsymbol{u}\rangle| \leq t$.*

We emphasize the order of the quantifiers: for a t-median, there must exist a single set $S \subseteq [n]$ with $|S| > \frac{k}{2}$ so that $\|\boldsymbol{y}_i - \boldsymbol{\nu}\| \leq t$ for all $i \in S$, and thus in particular $|\langle \boldsymbol{y}_i - \boldsymbol{\nu}, \boldsymbol{u}\rangle| \leq t$ for all $i \in S$ and $\boldsymbol{u} \in \mathbb{S}^{d-1}$. On the other hand, for a t-central point, there only needs to exist $S(\boldsymbol{u}) \subseteq [n]$ with $|S(\boldsymbol{u})| > \frac{k}{2}$, possibly a *different* set for different $\boldsymbol{u}$, so that $|\langle \boldsymbol{y}_i - \boldsymbol{\nu}, \boldsymbol{u}\rangle| \leq t$ for all $i \in S(\boldsymbol{u})$. Accordingly, every t-median is t-central, but not vice-versa.

Theorem 8.12 ([LM19])**.** *Let $\widehat{\mu}^{\mathsf{LM}}$ output a t-central point for $y_1, \dots, y_k$ if one exists, or an arbitrary point (say, $\mathbf{0}$) if not. Then, for any isotropic ρ, this estimator achieves the subgaussian width*

$$t \lesssim \frac{\sqrt{d} + \sqrt{\log(1/\delta)}}{\sqrt{n}}. \tag{8.24}$$

The proof follows the same outline as before. The main difficulty is in showing that μ is t-central for t with the subgaussian width scaling with probability $1 - \delta$, which is difficult because this t is much smaller than the smallest t for which μ is a t-median. However, once this is established, the result follows as before from the following.

Proposition 8.13. *If ν, ν' are both t-central for $y_1, \dots, y_k$, then $\|\nu - \nu'\| \le 2t$.*

Proof. Consider $u = (\nu - \nu')/\|\nu - \nu'\|$. The sets $S(u)$ of "close" points for ν and ν' in this direction again share some $i \in [n]$ by the pigeonhole principle. We then have

$$\|\nu - \nu'\| = \langle \nu - \nu', u \rangle \le |\langle \nu - y_i, u \rangle| + |\langle \nu' - y_i, u \rangle| \le 2t, \tag{8.25}$$

as desired. □

Unfortunately, unlike for the case of t-medians, it looks hard to find a t-central point, since it is no longer the case that one of the y_i must be $O(t)$-central when a t-central point exists. It even seems hard to *check* whether a given ν is t-central, since this involves quantifying over the infinitely many $u \in \mathbb{S}^{d-1}$, or, in practice, an exponential-size discretization of the possible u. These difficulties notwithstanding, we will see that we may use SOS to produce an effective version of the Lugosi-Mendelson estimator.

8.5 Hopkins' Sum-of-Squares Implementation

We now present the SOS-based algorithm devised by Hopkins in [Hop18c]. Hopkins proceeds in two steps. First, and what is not so difficult, he shows that it *is* in fact possible to effectively certify centrality using SOS; indeed, it is possible to do so using degree 2 SOS, giving a fairly simple SDP. ("Effectively" here means that we may certify that μ is t-central for t the subgaussian width.) Second, in a surprising twist, these SOS certificates are themselves embedded into *another* SOS program to search for "certifiably central" points. This allows us to apply the proofs-to-algorithms paradigm to the crucial result of Proposition 8.13, and to treat this problem following the same type of argument we have seen before.

8.5.1 Certifying Centrality

We first formulate centrality as a polynomial optimization problem. It will be useful to work with a slightly more restrictive definition of centrality, where we replace the majority condition in Definition 8.11 with a quantitative bound.

Definition 8.14. *We say that ν is* (t, ϵ)-central *for $y_1, \dots, y_k \in \mathbb{R}^d$ if, for all $u \in \mathbb{S}^{d-1}$, for at least $(1 - \epsilon)k$ of the y_i, we have $|\langle y_i - \nu, u \rangle| \le t$.*

Proposition 8.15. *Suppose $y_1, \ldots, y_k \in \mathbb{R}^d$, $t > 0$ is fixed, and $z \in \mathbb{R}^d$. Then, the smallest ϵ such that z is (t,ϵ)-central for $y_1, \ldots, y_k$ is given by either of the following equivalent polynomial optimization problems:*

$$\begin{aligned}\mathsf{Opt}(z, t, y_1, \ldots, y_k) &:= \left\{ \begin{array}{ll} \textit{maximize} & \frac{1}{k}\sum_{i=1}^k b_i \\ \textit{subject to} & b_i^2 - b_i = 0 \textit{ for all } i \in [k], \\ & \sum_{j=1}^d u_j^2 = 1, \\ & b_i \langle y_i - z, u \rangle \geq b_i t \textit{ for all } i \in [k] \end{array} \right\} \\ &= \left\{ \begin{array}{ll} \textit{maximize} & \frac{1}{k}\sum_{i=1}^k b_i \\ \textit{subject to} & b_i^2 \leq 1 \textit{ for all } i \in [k], \\ & \sum_{j=1}^d u_j^2 \leq 1, \\ & b_i \langle y_i - z, u \rangle \geq b_i t \textit{ for all } i \in [k] \end{array} \right\}. \end{aligned} \tag{8.26}$$

Proof. The first optimization problem may be equivalently rewritten

$$\begin{aligned}\mathsf{Opt}(z, t, y_1, \ldots, y_k) &= \left\{ \begin{array}{ll} \text{maximize} & \#\{i \in [k] : b_i = 1\}/k \\ \text{subject to} & b_i \in \{0,1\} \text{ for all } i \in [k], \\ & u \in \mathbb{S}^{d-1}, \\ & \langle y_i - z, u \rangle \geq t \text{ for all } i \in [k] \text{ with } b_i = 1 \end{array} \right\} \\ &= \left\{ \begin{array}{ll} \text{maximize} & \#\{i \in [k] : \langle y_i - z, u \rangle \geq t\}/k \\ \text{subject to} & u \in \mathbb{S}^{d-1} \end{array} \right\}, \end{aligned} \tag{8.27}$$

and in this last form this is by definition the smallest ϵ for which z is (t,ϵ)-central. The second form of the problem in the statement is equivalent to the first since if $b_i \in [-1,1]$ then we may replace every negative b_i with 0 and every positive b_i with 1, still have a feasible point, and only increase the objective function. Likewise if $\|u\| \leq 1$, then we may replace u with $u/\|u\|$ and only increase the objective function. □

With this in hand, we may address the question of certifying centrality by asking whether low-degree SOS proofs of centrality exist, in the form of SOS proofs of bounds on $\mathsf{Opt}_t(z)$. Let us define the shorthand

$$t^\star := \frac{\sqrt{d} + \sqrt{\log(1/\delta)}}{\sqrt{n}}. \tag{8.28}$$

Lemma 8.16. *For any $\epsilon > 0$, there exists $C = C(\epsilon) > 0$ such that, for any $\delta > 0$,*

$$\mathbb{P}\left[\textit{there is a degree 2 SOS proof that } \mathsf{Opt}(\mu, Ct^\star, y_1, \ldots, y_k) \leq \epsilon\right] \geq 1 - \delta. \tag{8.29}$$

Proof Sketch. Unlike most of our arguments about SOS algorithms, where we exhibited direct SOS proofs of desirable statements and thus only relied on *weak* duality between the Lasserre and Parrilo formulations, here we will use *strong* duality, the statement that the values of the Lasserre and Parrilo relaxations of $\mathsf{Opt}(z, t, y_1, \ldots, y_k)$ are equal (as follows from Theorem 4.15). In particular, letting $\mathsf{SOS}_2(\mu, t, y_1, \ldots, y_k)$ be the value of the degree 2

SOS relaxation of $\mathsf{Opt}(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k)$, we have

$$\mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k) = \left\{ \begin{array}{ll} \text{maximize} & \frac{1}{k}\sum_{i=1}^k b_i \\ \text{subject to} & \begin{bmatrix} 1 & \boldsymbol{b}^\top & \boldsymbol{u}^\top \\ \boldsymbol{b} & \boldsymbol{B} & \boldsymbol{W} \\ \boldsymbol{x} & \boldsymbol{W}^\top & \boldsymbol{U} \end{bmatrix} \succeq \boldsymbol{0}, \\ & B_{ii} \leq 1 \text{ for all } i \in [k], \\ & \mathsf{Tr}(\boldsymbol{U}) \leq 1, \\ & \boldsymbol{e}_i^\top \boldsymbol{W}(\boldsymbol{y}_i - \boldsymbol{\mu}) \geq t b_i \text{ for all } i \in [k] \end{array} \right\}. \tag{8.30}$$

We first address the question of concentration of this quantity around its mean. We note that, as a function of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, SOS_2 satisfies the following *bounded differences* property:

$$|\mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}, \boldsymbol{y}_i, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_k) - \mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}, \boldsymbol{y}_i', \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_k)| \leq \frac{1}{k}, \tag{8.31}$$

where $\boldsymbol{y}_i' \in \mathbb{R}^d$ is an arbitrary vector. This is because, from a feasible point of one relaxation, we may form a feasible point of the other by setting b_i, the ith row of $\boldsymbol{W}$, and the off-diagonal terms in the ith row and column of $\boldsymbol{B}$ all to zero, which changes the objective function by at most $\frac{1}{k}$. Since the $\boldsymbol{y}_i$ are independent, the bounded differences concentration inequality implies that

$$\mathbb{P}[|\mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k) - \mathbb{E}\mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k)| \geq \epsilon] \leq \exp(-k\epsilon^2) = \delta \tag{8.32}$$

for a suitable choice of $k \propto \log(1/\delta)$.

It remains to show that $\mathbb{E}\mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k)$ is not too large. First, we note that, applying the last constraint to the objective function, for any feasible point we have

$$\frac{1}{k}\sum_{i=1}^k b_i \leq \frac{1}{kt}\sum_{i=1}^k \boldsymbol{e}_i^\top \boldsymbol{W}(\boldsymbol{y}_i - \boldsymbol{\mu}) = \frac{1}{kt}\langle \boldsymbol{W}, \overline{\boldsymbol{Y}} \rangle, \tag{8.33}$$

where $\overline{\boldsymbol{Y}}$ has the $\boldsymbol{y}_i - \boldsymbol{\mu}$ as its rows. We may thus bound by the simpler SDP

$$\mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k) \leq \frac{1}{kt} \left\{ \begin{array}{ll} \text{maximize} & \langle \overline{\boldsymbol{Y}}, \boldsymbol{W} \rangle \\ \text{subject to} & \begin{bmatrix} \boldsymbol{B} & \boldsymbol{W} \\ \boldsymbol{W}^\top & \boldsymbol{U} \end{bmatrix} \succeq \boldsymbol{0}, \\ & B_{ii} \leq 1 \text{ for all } i \in [k], \\ & \mathsf{Tr}(\boldsymbol{U}) \leq 1 \end{array} \right\}. \tag{8.34}$$

This is a natural SDP relaxation of the $2 \to 1$ norm of $\overline{\boldsymbol{Y}}$,

$$\|\overline{\boldsymbol{Y}}\|_{2\to 1} := \max_{\|\boldsymbol{u}\|_2 \leq 1} \|\overline{\boldsymbol{Y}}\boldsymbol{u}\|_1 = \max_{\substack{\|\boldsymbol{u}\|_2 \leq 1, \\ \boldsymbol{b} \in \{\pm 1\}^k}} \boldsymbol{b}^\top \overline{\boldsymbol{Y}}\boldsymbol{u}. \tag{8.35}$$

The square of this norm is actually an instance of a hypercube optimization problem similar to what we have seen in Chapter 1:

$$\|\overline{\boldsymbol{Y}}\|_{2\to 1}^2 = \max_{\boldsymbol{b} \in \{\pm 1\}^k} \boldsymbol{b}^\top (\overline{\boldsymbol{Y}}\,\overline{\boldsymbol{Y}}^\top)\boldsymbol{b}. \tag{8.36}$$

Now, in Exercise 1.4 you showed that the integrality gap of the degree 2 SOS SDP for this optimization problem, which is the same as the Goemans-Williamson SDP, is at most $\pi/2$. Thus, though we will not show it carefully, it is not surprising that the integrality gap of the SDP in (8.34) as an approximation of $\|\overline{\boldsymbol{Y}}\|_{2\to 1}$ is at most $\sqrt{\pi/2}$. We thus find

$$\mathbb{E}\mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \dots, \boldsymbol{y}_k) \le \sqrt{\frac{\pi}{2}} \frac{1}{kt} \mathbb{E}\|\overline{\boldsymbol{Y}}\|_{2\to 1}. \tag{8.37}$$

To analyze the remaining expectation, we expand it and again extract an expectation and a fluctuation, as follows: letting $\overline{\boldsymbol{y}}_i := \boldsymbol{y}_i - \boldsymbol{\mu}$,

$$\begin{aligned} \mathbb{E}\|\overline{\boldsymbol{Y}}\|_{2\to 1} &= \mathbb{E} \sup_{\|\boldsymbol{u}\|=1} \sum_{i=1}^k |\langle \overline{\boldsymbol{y}}_i, \boldsymbol{u}\rangle| \\ &\le \mathbb{E} \sup_{\|\boldsymbol{u}\|=1} \sum_{i=1}^k (|\langle \overline{\boldsymbol{y}}_i, \boldsymbol{u}\rangle| - \mathbb{E}|\langle \overline{\boldsymbol{y}}_i, \boldsymbol{u}\rangle|) + k \sup_{\|\boldsymbol{u}\|=1} \mathbb{E}|\langle \overline{\boldsymbol{y}}_1, \boldsymbol{u}\rangle|. \end{aligned} \tag{8.38}$$

The second term is easy to analyze: by Jensen's inequality,

$$\begin{aligned} \sup_{\|\boldsymbol{u}\|=1} \mathbb{E}|\langle \overline{\boldsymbol{y}}_1, \boldsymbol{u}\rangle| &\le \sqrt{\sup_{\|\boldsymbol{u}\|=1} \mathbb{E}\langle \overline{\boldsymbol{y}}_1, \boldsymbol{u}\rangle^2} \\ &= \sqrt{\sup_{\|\boldsymbol{u}\|=1} \boldsymbol{u}^\top \mathsf{Cov}(\overline{\boldsymbol{y}}_1)\boldsymbol{u}} \\ &= \sqrt{\|\mathsf{Cov}(\overline{\boldsymbol{y}}_1)\|} \\ &= \sqrt{\frac{k}{n}}, \end{aligned} \tag{8.39}$$

since $\overline{\boldsymbol{y}}_1$ is an average of $m = n/k$ i.i.d. isotropic random vectors, so $\mathsf{Cov}(\overline{\boldsymbol{y}}_1) = \frac{k}{n}\boldsymbol{I}_d$. In the first term, at an intuitive level, the centering makes the sum comparable to the same sum with random signs to each term:

$$\begin{aligned} \mathbb{E} \sup_{\|\boldsymbol{u}\|=1} \sum_{i=1}^k (|\langle \overline{\boldsymbol{y}}_i, \boldsymbol{u}\rangle| - \mathbb{E}|\langle \overline{\boldsymbol{y}}_i, \boldsymbol{u}\rangle|) &\lesssim \mathop{\mathbb{E}}_{s_i \sim \mathsf{Unif}(\{\pm 1\})} \mathop{\mathbb{E}}_{\overline{\boldsymbol{y}}_i} \sup_{\|\boldsymbol{u}\|=1} \sum_{i=1}^k s_i \langle \overline{\boldsymbol{y}}_i, \boldsymbol{u}\rangle \\ &= \mathop{\mathbb{E}}_{s_i \sim \mathsf{Unif}(\{\pm 1\})} \mathop{\mathbb{E}}_{\overline{\boldsymbol{y}}_i} \left\| \sum_{i=1}^k s_i \overline{\boldsymbol{y}}_i \right\| \\ &\lesssim \sqrt{k} \cdot \mathbb{E}\|\overline{\boldsymbol{y}}_1\| \\ &= \sqrt{k} \cdot \sqrt{\mathsf{Tr}\,\mathsf{Cov}(\overline{\boldsymbol{y}}_1)} \\ &= \sqrt{k} \cdot \sqrt{\frac{k}{n} d} \\ &= k\sqrt{\frac{d}{n}}. \end{aligned} \tag{8.40}$$

One may verify this part more carefully by applying a symmetrization argument and then using some general concentration inequalities; see [Hop18c] for more details.

Putting the pieces together, we find

$$\mathbb{E}\mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \dots, \boldsymbol{y}_k) \lesssim \frac{1}{kt}\left(k\sqrt{\frac{k}{n}} + k\sqrt{\frac{d}{n}}\right) = \frac{\frac{\sqrt{k}+\sqrt{d}}{\sqrt{n}}}{t} = \frac{t^\star}{t}, \tag{8.41}$$

thus taking $t = \Theta(t^\star/\epsilon)$ we may ensure that $\mathbb{E}\mathsf{SOS}_2(\boldsymbol{\mu}, t, \boldsymbol{y}_1, \dots, \boldsymbol{y}_k) = O(\epsilon)$, as desired. □

8.5.2 Sum-of-Squares Squared

So far, we have only shown that SOS can *certify* centrality, in particular for the special point $\boldsymbol{\mu} \in \mathbb{R}^d$, with a low-degree proof. It is not immediately clear how this helps us use SOS to *find* a central point given the $\boldsymbol{y}_1, \dots, \boldsymbol{y}_k$ as input.

Definition 8.17. *We say that $\boldsymbol{z} \in \mathbb{R}^d$ is* certifiably (t, ϵ)-central *for $\boldsymbol{y}_1, \dots, \boldsymbol{y}_k$ if there is a degree 2 SOS proof that $\mathsf{Opt}(\boldsymbol{z}, t, \boldsymbol{y}_1, \dots, \boldsymbol{y}_k) \le \epsilon$, i.e., if there exist $\alpha_i, \beta_i, \gamma \ge 0$ and $S \in \mathsf{SOS}$ such that the following holds as a polynomial equality over $\mathbb{R}[b_1, \dots, b_k, u_1, \dots, u_d]$:*

$$\epsilon k - \sum_{i=1}^k b_i \overset{(\mathrm{p})}{=} \sum_{i=1}^k \alpha_i b_i(\langle \boldsymbol{y}_i - \boldsymbol{z}, \boldsymbol{u}\rangle - t) + \sum_{i=1}^k \beta_i(1 - b_i^2) + \gamma(1 - \|\boldsymbol{u}\|^2) + S(\boldsymbol{b}, \boldsymbol{u}). \tag{8.42}$$

In this language, Lemma 8.16 just says that $\boldsymbol{\mu}$ is certifiably $(O_\epsilon(t^\star), \epsilon)$-central for any $\epsilon > 0$ with probability $1 - \delta$.

The idea of [Hop18c] is then to search for a central point by solving an SOS relaxation *of the polynomial constraint* (8.42). That is, we solve an SOS relaxation of a polynomial system expressing the existence of an SOS proof.

To be more specific, we first note that, for any $S(\boldsymbol{b}, \boldsymbol{u}) \in \mathsf{SOS}$ of degree 2, we may write

$$S(\boldsymbol{b}, \boldsymbol{u}) = \sum_{i=1}^{d+k+1} \langle \boldsymbol{s}_i, (1\ \boldsymbol{b}\ \boldsymbol{u})^\top\rangle^2 \tag{8.43}$$

for some $\boldsymbol{s}_1, \dots, \boldsymbol{s}_{d+k+1} \in \mathbb{R}^{d+k+1}$. We will search for a degree D pseudoexpectation $\widetilde{\mathbb{E}}$ over the variables

$$x_1, \dots, x_d;\ \ \alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k, \gamma;\ \ (s_{ij})_{i,j=1}^{d+k+1}, \tag{8.44}$$

which satisfies the constraints

$$\epsilon k - \sum_{i=1}^k b_i \overset{(\mathrm{p})}{=} \sum_{i=1}^k \alpha_i b_i(\langle \boldsymbol{y}_i - \boldsymbol{x}, \boldsymbol{u}\rangle - t) + \sum_{i=1}^k \beta_i(1 - b_i^2) + \gamma(1 - \|\boldsymbol{u}\|^2) + \sum_{i=1}^{d+k+1} \langle \boldsymbol{s}_i, (1\ \boldsymbol{b}\ \boldsymbol{u})^\top\rangle^2, \tag{8.45}$$

$$\alpha_i, \beta_i, \gamma \ge 0, \tag{8.46}$$

where the first constraint is interpreted as a system of polynomial constraints in the variables in (8.44) coming from equating coefficients when the left- and right-hand sides are viewed as polynomials in $\boldsymbol{b}, \boldsymbol{u}$.

The following statement is the idea of the remaining claim; as we will mention below, it is not quite correct and requires a substantial but purely technical adjustment to obtain a truly correct version.

Lemma 8.18 ("Morally correct" SOS version of Proposition 8.13)**.** *For an absolute constant $D_0 \in 2\mathbb{N}$, if $z \in \mathbb{R}^d$ is certifiably (t,ϵ)-central and $\widetilde{\mathbb{E}}$ is a degree D_0 pseudoexpectation over the variables and constraints described above, then*

$$\widetilde{\mathbb{E}}\|x - z\|^2 \lesssim t^2. \tag{8.47}$$

Indeed, as the name suggests, the Lemma states that a sufficiently high-degree pseudoexpectation over certifiably central points satisfies the key "closeness property" from Proposition 8.13.

Corollary 8.19. *There is an estimator $\hat{\mu}$ having subgaussian width for any $\rho \in \mathcal{M}(\mathbb{R}^d)$ whose covariance is the identity.*

Proof. Fix some small $\epsilon > 0$, say $\epsilon = \frac{1}{100}$. Then, by Lemma 8.16, for some $C > 0$, with probability at least $1 - \delta$, μ is certifiably $(Ct^\star, \frac{1}{100})$-central.

Let $\hat{\mu}$ be the result of applying the Gaussian rounding of Proposition 5.5 to the x variables in a pseudoexpectation $\widetilde{\mathbb{E}}$ as above. Taking $z = \mu$ in Lemma 8.18, we find by the moment property of the Gaussian rounding and Jensen's inequality that $\mathbb{E}\|\hat{\mu} - \mu\| = O(t^\star)$, and the result follows. □

Let us sketch just the main ideas behind Lemma 8.18 and the necessary adjustments to the SOS program to actually make such a statement technically correct. This will be mostly a hand-waving discussion.

Of course, what we would like to do is to implement the "shared inlier" or "majority" argument from the proof of Proposition 8.13 in the SOS proof system. Unfortunately, such arguments using tools like the pigeonhole principle are precisely the kinds of arguments *not* directly expressible with SOS.

Let us recall the proof idea of Proposition 8.13 and start writing it in terms of certifiably central points and their certificates. Suppose $z, z' \in \mathbb{R}^d$ are two certifiably (t,ϵ)-central points for some small $\epsilon > 0$. This means that we have polynomial equations in indeterminates b, u of the form

$$\epsilon k - \sum_{i=1}^k b_i \overset{(\mathrm{p})}{=} \sum_{i=1}^k \alpha_i b_i(\langle y_i - z, u\rangle - t) + \sum_{i=1}^k \beta_i(1 - b_i^2) + \gamma(1 - \|u\|^2) + \mathsf{SOS}, \tag{8.48}$$

$$\epsilon k - \sum_{i=1}^k b_i \overset{(\mathrm{p})}{=} \sum_{i=1}^k \alpha_i' b_i(\langle y_i - z', u\rangle - t) + \sum_{i=1}^k \beta_i'(1 - b_i^2) + \gamma'(1 - \|u\|^2) + \mathsf{SOS}. \tag{8.49}$$

The proof of Proposition 8.13 amounts to setting $u := (z - z')/\|z - z'\|$ and using that, with this choice, there would have to exist a common $b_i = 0$ at optimality in the original problems $\mathsf{Opt}(z, t, y_1, \dots, y_k)$ and $\mathsf{Opt}(z', t, y_1, \dots, y_k)$.

What reflection of this fact can we see in just the certificates (8.48) and (8.49)? One tool we may use is *complementary slackness*: recall that the α_i in (8.48) are, in the original degree 2 SOS relaxation of $\mathsf{Opt}(z, t, y_1, \dots, y_k)$, Lagrange multipliers of constraints coming from the polynomial constraint $b_i\langle y_i - z, u\rangle \geq b_i t$. In particular, $\alpha_i > 0$ only when this associated constraint is "active" at optimality, so that the constraint is actually an equality This can happen either when $b_i = 0$ at optimality, so that y_i is an inlier for z in the worst

possible direction u, or when $\langle y_i - z, u\rangle = t$, so that y_i is on the threshold between an inlier and an outlier. In any case, very roughly speaking, we may take α_i to be some measurement of how much y_i is an inlier for z over nearly worst-case directions u, and likewise for α_i' and z'.

Another, more formal observation is that we would like to combine the certificates (8.48) and (8.49) in such a way that the dependence on the y_i will go away, so that we are left with only some relation between z and z'.

Combining these two ideas, we try the following: let $v := (z - z')/\|z - z'\|$. Evaluate (8.48) with $u = -v$ and $b_i = \alpha_i'/(\alpha_i + \alpha_i')$; evaluate (8.49) with $u = v$ and $b_i = \alpha_i/(\alpha_i + \alpha_i')$; and finally, add together the results. Using the non-negativity of the last three terms in each certificate with these choices, we find

$$-(1-2\epsilon)k \geq \sum_{i=1}^{k} \frac{\alpha_i \alpha_i'}{\alpha_i + \alpha_i'} (\|z - z'\| - 2t), \tag{8.50}$$

and rearranging this gives something close to what we want,

$$\|z - z'\| \leq 2t(1 + \Delta), \tag{8.51}$$

where

$$\Delta = \frac{k}{2t \sum_{i=1}^{k} \frac{\alpha_i \alpha_i'}{\alpha_i + \alpha_i'}} = \left(t \cdot \frac{1}{k} \sum_{i=1}^{k} \frac{2}{\frac{1}{\alpha_i} + \frac{1}{\alpha_i'}} \right)^{-1}. \tag{8.52}$$

To parse this equation, note that each term in the sum is a harmonic mean of α_i and α_i', which we may think of as a proxy for $\min\{\alpha_i, \alpha_i'\}$. Also, by considering the coefficients of b_i on either side of the certificates (8.48) and (8.49), we see that we expect α_i and α_i' to be roughly on the scale $\Theta(1/t)$. Thus, having Δ not too large amounts to having the average $t \cdot \min\{\alpha_i, \alpha_i'\}$ not too small. With our previous intuition, this is precisely a measurement of z and z' having a shared inlier in the v direction!

Unfortunately, to actually implement this idea as an SOS proof, we need to have some control over the "regularity" of the α_i. To do this, [Hop18c] just changes the underlying program: it turns out that a stronger version of Lemma 8.16 holds; not only is μ certifiably central, but the α_i in its certificate may be chosen to be somewhat "flat" without too many unusually large or small entries. Adjusting the SOS relaxation of certifiable centrality to only allow such certificates then allows us to prove a rigorous version of Lemma 8.18.

Exercises

Exercise 8.1 (Second moment width)**.** *For any $\delta > 0$ and $n \in \mathbb{N}$, show that there exists a centered $\rho \in \mathcal{M}(\mathbb{R})$ with variance 1 such that, when $x_1, \ldots, x_n \sim \rho$ are i.i.d.,*

$$\mathbb{P}\left[|\widehat{\mu}^{\mathrm{emp}}(x_1, \ldots, x_n)| \geq \sqrt{\frac{1/\delta}{n}} \right] \geq 1 - O(\delta). \tag{8.53}$$

Exercise 8.2 (Median-of-medians algorithm)**.** *Describe an algorithm that, given $x_1, \ldots, x_n \in \mathbb{R}$ and $k \in [n]$, finds the kth largest of the x_i in time $O(n)$.*

HINT: *Show that by forming small "buckets" of the x_i and computing the median of the medians of each bucket, you may find a good "pivot" element x_p: one where a fairly large fraction of elements is larger than x_p and a fairly large fraction is smaller. Use this to produce a divide-and-conquer algorithm.*

Exercise 8.3 (Lipschitz constant of empirical mean norm)**.** *Let $f : \mathbb{R}^{dn} \to \mathbb{R}$ be given by $f(\boldsymbol{x}) = \|\frac{1}{n}\sum_{i=0}^{n-1}(x_{di+1}, x_{di+2}, \ldots, x_{di+d})^\top\|$. Show that the Lipschitz constant of f is at most $1/\sqrt{n}$.*

HINT: *Form a $d \times n$ matrix from $\boldsymbol{x}$ and think in terms of matrix norms.*

NOTES

TUKEY MEDIANS The notion of a t-central point is somewhat similar to that of a Tukey median, a point $\boldsymbol{z}$ that maximizes the minimum over $\boldsymbol{u} \in \mathbb{S}^{d-1}$ of the smaller of the number of $i \in [k]$ with $\langle \boldsymbol{y}_i - \boldsymbol{z}, \boldsymbol{u}\rangle$ positive or negative. This is perhaps a more honest high-dimensional generalization of the median than those we work with, since it does not involve the fixed scale t and does not require the median to be *close* in any direction to the $\boldsymbol{y}_i$, but rather only for it to have a rank statistic close enough to $\frac{1}{2}k$ in all directions.

HODGES-LEHMANN ESTIMATORS AND SHARP CONSTANTS A variant of the median-of-means estimator in one dimension that has appeared in the literature is the *Hodges-Lehmann estimator* and its generalizations, which take the form

$$\hat{\mu} = \mathsf{med}\left(\left\{\frac{1}{m}\sum_{j\in B} x_j\right\}_{B\in\binom{[n]}{m}}\right). \tag{8.54}$$

Here, instead of taking the median of means over a *fixed* partition of $[n]$, we consider means over *all* subsets of a given size (the original Hodges-Lehmann estimator considered only $m = 2$). Interestingly, as shown in the recent paper [Min22], this kind of estimator can improve the constants in the width achieved by our $\hat{\mu}^{\mathsf{mom}}$ in the scalar case.

Open Problem 8.1 (Optimal constants in high-dimensional mean estimation)**.** *Investigate optimal constants in the subgaussian width in high-dimensional mean estimation. Are Hodges-Lehmann estimators still a good idea in high dimension?*

ADVERSARIAL AND CONTAMINATED MODELS A stronger requirement of robustness might ask that our algorithm work not only under poorly-behaved distributions ρ, but also when some of our data does not come from ρ, but is *contaminated* either by data from another, unrelated distribution, or, worse yet, can be provided by an adversary. Adaptations of semidefinite programming methods and other algorithms (developed after [Hop18c]) to this kind of setting are studied in, e.g., [DKP20, HLZ20, DL22].

PRIVACY Algorithmic approaches that are robust to outliers or poorly-behaved input distributions seem to also be amenable to being made *private*, not "leaking" very much information about individual samples. A few recent results in this direction are [KMV21, HKM21].

Part III

Sum-of-Squares Lower Bounds

9 | Case Study 4: Parity/Knapsack

We have seen in the previous part of the course that SOS is a powerful tool for algorithm design in various settings. Next, we will look at some results in the opposite direction, showing that SOS requires high degree (and thus long runtime) to solve certain problems. These results will come in two types:

1. Sometimes, limitations associated specifically to the SOS proof system will make it difficult for SOS to certify polynomial inequalities that are (to us, from outside the limited proof system) obviously true. These are *easy problems* that SOS cannot solve, and they tell us about faults in the SOS proof system.

2. Other times, we will show that SOS cannot solve problems that we have some other reason to believe are hard to solve (e.g., lower bounds against other algorithms). These are (conjecturally) *hard problems* that SOS cannot solve, and we use SOS lower bounds as a form of evidence for their hardness.

Not surprisingly, the first class of lower bound is usually much simpler. We will therefore start our study of SOS lower bounds with a simple example of a *deterministic* problem that SOS requires high degree to solve.

This problem will be formulated as an optimization problem over the hypercube $\{\pm 1\}^n$, as in the setting of Chapter 1:

$$\mathsf{Opt} := \left\{ \begin{array}{ll} \text{maximize} & p(x) \\ \text{subject to} & x_i^2 - 1 = 0 \text{ for all } i \in [n] \end{array} \right\}. \tag{9.1}$$

You have seen in Exercise 2.9 that the degree $2n$ relaxation solves any such problem *exactly*—at this degree, the relaxation of the hypercube is not a relaxation anymore, and every pseudoexpectation of this degree is an actual expectation with respect to a probability measure over the hypercube.

We will now show a converse result, giving a simple choice of $p(x)$ for which degree at least n is required to obtain a tight bound, due to [Gri01a, Lau03]. In fact, as [Lau03] conjectured and [FSP16] proved, the tightness of the degree $2n$ relaxation is suboptimal, and the degree $n + \mathbb{1}\{n \text{ is odd}\}$ relaxation is already exact; this result is substantially harder than the argument from Exercise 2.9, however.

In one sense, at least a lower bound of degree $d = \Omega(n)$ is to be expected, since looking at $p(x) = x^\top L x$ for L the class of graph Laplacian matrices encodes the NP-complete MaxCut problem in this class of polynomial optimization problems, and the exponential time hypothesis implies that time $\exp(\Omega(n))$ is required to solve MaxCut in the worst case.

But, we will actually show that a very simple specific example of a graph Laplacian gives our lower bound.

Namely, let $G = K_n$ be the complete graph on n vertices. We have $|E(G)| = \binom{n}{2}$, and may compute

$$\mathsf{MaxCut}(G) = \frac{1}{4}(n^2 - \mathbb{1}\{n \text{ is odd}\}). \tag{9.2}$$

We may manipulate the MaxCut objective function as follows:

$$\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x} = \frac{1}{4} \sum_{1 \le i < j \le n} (x_i - x_j)^2 = \frac{n(n-1)}{4} - \frac{1}{2} \sum_{1 \le i < j \le n} x_i x_j = \frac{n^2}{4} - \frac{1}{4} \left(\sum_{i=1}^n x_i \right)^2 . \tag{9.3}$$

Thus, we will consider Opt as above with $p(\boldsymbol{x}) = (\sum_{i=1}^n x_i)^2$:

$$\mathsf{Opt} := \left\{ \begin{array}{ll} \text{maximize} & (\sum_{i=1}^n x_i)^2 \\ \text{subject to} & x_i^2 - 1 = 0 \text{ for all } i \in [n] \end{array} \right\}. \tag{9.4}$$

Theorem 9.1 ([Gri01a, Lau03]). *If n is odd, then the value of the degree $n-1$ SOS relaxation of* Opt *is zero.*

In contrast, clearly $\sum_{i=1}^n x_i$ is a non-zero integer for any $\boldsymbol{x} \in \{\pm 1\}^n$, so we have $\mathsf{Opt} = 1$. It is also straightforward to show that this gives an integrality gap for $\mathsf{MaxCut}(G = K_n)$, showing that the degree $n-1$ relaxation has value $\frac{n^2}{4} > \frac{n^2-1}{4}$.

Example 9.2 ($n = 3$). *In this case, Theorem 9.1 says that degree 2 SOS cannot certify the true inequality $\widetilde{\mathbb{E}}(x_1 + x_2 + x_3)^2 \ge 1$. Setting $\boldsymbol{X} := \widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top]$ and rearranging, we see that this is equivalent to $X_{12} + X_{23} + X_{13} \ge -1$, which is precisely one of the triangle inequalities that Exercise 1.6 shows are not satisfied by degree 2 SOS (which is the same as the Goemans-Williamson relaxation). Indeed, we may take $\widetilde{\mathbb{E}}[\boldsymbol{x}] = \boldsymbol{0}$ and*

$$\widetilde{\mathbb{E}}[\boldsymbol{x}\boldsymbol{x}^\top] = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}, \tag{9.5}$$

in which case $\widetilde{\mathbb{E}}(x_1 + x_2 + x_3)^2 = \boldsymbol{1}^\top \boldsymbol{X} \boldsymbol{1} = 0$, as Theorem 9.1 claims.

9.1 Pseudoexpectation Values From Symmetry

To prove the result, we need to build a degree $n-1$ pseudoexpectation $\widetilde{\mathbb{E}}$ satisfying the hypercube constraints and having $\widetilde{\mathbb{E}}(\sum_{i=1}^n x_i)^2 = 0$. We simplify our task repeatedly by making some simple observations.

First, by linearity it suffices to specify the pseudoexpectations of monomials. Moreover, since $\widetilde{\mathbb{E}}[x_i^2 p(\boldsymbol{x})] = \widetilde{\mathbb{E}}[p(\boldsymbol{x})]$, it suffices to specify the pseudoexpectations of multilinear monomials $\boldsymbol{x}^S$ for $S \subset [n]$ with $|S| \le n-1$.

Next, given $\widetilde{\mathbb{E}}$ a pseudoexpectation over the hypercube constraints, we observe that $\widetilde{\mathbb{E}}'$ defined by $\widetilde{\mathbb{E}}'[p(\boldsymbol{x})] := \frac{1}{2}(\widetilde{\mathbb{E}}[p(\boldsymbol{x})] + \widetilde{\mathbb{E}}[p(-\boldsymbol{x})])$ is also a pseudoexpectation over the hypercube constraints, which has $\widetilde{\mathbb{E}}'(\sum_{i=1}^n x_i)^2 = \widetilde{\mathbb{E}}(\sum_{i=1}^n x_i)^2$. This pseudoexpectation also

satisfies $\widetilde{\mathbb{E}}' \boldsymbol{x}^S = 0$ for all sets S of *odd* size. Thus, by performing this symmetrization, we may assume without loss of generality that $\widetilde{\mathbb{E}} \boldsymbol{x}^S = 0$ for S having odd size.

Similarly, $\widetilde{\mathbb{E}}'[p(\boldsymbol{x})] := \frac{1}{n!} \sum_{\sigma \in S_n} \widetilde{\mathbb{E}}[p(x_{\sigma(1)}, \dots, x_{\sigma(n)})]$ is a pseudoexpectation over the hypercube constraints and has the same objective value as $\widetilde{\mathbb{E}}$ (here S_n is the symmetric group of permutations of $[n]$). This $\widetilde{\mathbb{E}}'$ has the important property that $\widetilde{\mathbb{E}}' \boldsymbol{x}^S$ depends only on the cardinality $|S|$. Thus we may also assume without loss of generality that $\widetilde{\mathbb{E}} \boldsymbol{x}^S = f(|S|)$ for some $f : \mathbb{N} \to \mathbb{R}$ with $f(k) = 0$ for all odd k.

Finally, suppose $|S| \leq n-2$. We note that by the SOS Cauchy-Schwarz inequality, splitting $S = S_1 \sqcup S_2$ with $|S_1| = \lfloor |S|/2 \rfloor \leq (n-3)/2$ and $|S_2| = \lceil |S|/2 \rceil \leq (n-1)/2$, we have

$$\left| \widetilde{\mathbb{E}} \left[\left(\sum_{i=1}^n x_i \right) \boldsymbol{x}^S \right] \right| = \left| \widetilde{\mathbb{E}} \left[\left(\sum_{i=1}^n x_i \right) \boldsymbol{x}^{S_1} \cdot \boldsymbol{x}^{S_2} \right] \right| \leq \left(\widetilde{\mathbb{E}} \left[\left(\sum_{i=1}^n x_i \right)^2 \right] \right)^{1/2} = 0 \tag{9.6}$$

by our assumption. On the other hand, we may expand by the previous observation

$$0 = \widetilde{\mathbb{E}} \left[\left(\sum_{i=1}^n x_i \right) \boldsymbol{x}^S \right] = |S| f(|S|-1) + (n - |S|) f(|S|+1), \tag{9.7}$$

and rearranging and reindexing this gives the recurrence

$$f(k+2) = -\frac{k+1}{n-k-1} f(k). \tag{9.8}$$

Combined with $\widetilde{\mathbb{E}}[1] = f(0) = 1$, this specifies all values of $f(k)$, and thus our entire pseudoexpectation! Remarkably, the many symmetries of this problem imply that there is only one plausible pseudoexpectation that we need to consider for our lower bound.

We may also give a closed form for these values,

$$\widetilde{\mathbb{E}} \boldsymbol{x}^S = f(|S|) = \mathbb{1}\{|S| \text{ even}\} (-1)^{|S|/2} \prod_{i=0}^{|S|/2-1} \frac{2i+1}{n-2i-1}, \tag{9.9}$$

with the first few non-zero values given by

$$\widetilde{\mathbb{E}}[1] = 1, \quad \widetilde{\mathbb{E}}[x_i x_j] = -\frac{1}{n-1}, \quad \widetilde{\mathbb{E}}[x_i x_j x_k x_\ell] = \frac{3}{(n-1)(n-3)}. \tag{9.10}$$

We may also check that the condition on the objective function is indeed satisfied:

$$\widetilde{\mathbb{E}} \left(\sum_{i=1}^n x_i \right)^2 = n f(0) + n(n-1) f(2) = n - n(n-1) \frac{1}{n-1} = 0. \tag{9.11}$$

All other conditions on $\widetilde{\mathbb{E}}$ except for positivity are automatically satisfied by our construction. Thus, defining the pseudomoment matrix $\boldsymbol{Y} \in \mathbb{R}^{\binom{[n]}{\leq (n-1)/2} \times \binom{[n]}{\leq (n-1)/2}}$ to have entries

$$Y_{S,T} := \widetilde{\mathbb{E}}[\boldsymbol{x}^S \cdot \boldsymbol{x}^T] = \widetilde{\mathbb{E}}[\boldsymbol{x}^{S \triangle T}] = f(|S \triangle T|), \tag{9.12}$$

the following will occupy our attention for the rest of the chapter.

Lemma 9.3. $\boldsymbol{Y} \succeq \boldsymbol{0}$.

Clearly, Theorem 9.1 follows immediately from our construction and Lemma 9.3.

9.2 Degree 4 Lower Bound

To introduce the main ideas of the analysis, let us consider just the first non-trivial submatrix of $\boldsymbol{Y}$, that indexed by sets of size 0, 1, and 2. Abusing notation for this section, let us also call this matrix $\boldsymbol{Y}$. Divided into blocks according to sets of these three sizes, this submatrix looks like

$$\boldsymbol{Y} = \begin{bmatrix} 1 & \mathbf{0} & -\frac{1}{n-1}\mathbf{1}^\top \\ \mathbf{0} & \boldsymbol{Y}^{(2)} & \mathbf{0} \\ -\frac{1}{n-1}\mathbf{1} & \mathbf{0} & \boldsymbol{Y}^{(4)} \end{bmatrix}, \tag{9.13}$$

where $\boldsymbol{Y}^{(2)} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{Y}^{(4)} \in \mathbb{R}^{\binom{[n]}{2}\times\binom{[n]}{2}}$ have entries

$$Y^{(2)}_{ij} = \left\{ \begin{array}{ll} 1 & \text{if } i = j, \\ -\frac{1}{n-1} & \text{if } i \neq j \end{array} \right\}, \tag{9.14}$$

$$Y^{(4)}_{ST} = f(|S \triangle T|) = \left\{ \begin{array}{ll} 1 & \text{if } |S\triangle T| = 0, \\ -\frac{1}{n-1} & \text{if } |S\triangle T| = 2, \\ \frac{3}{(n-1)(n-3)} & \text{if } |S\triangle T| = 4 \end{array} \right\}. \tag{9.15}$$

We observe that, after permuting the rows and columns, $\boldsymbol{Y}$ is the direct sum of $\boldsymbol{Y}^{(2)}$ with the matrix indexed by the subsets of size 0 and 2. We have

$$\boldsymbol{Y}^{(2)} = \left(1 + \frac{1}{n-1}\right)\boldsymbol{I}_n - \frac{1}{n-1}\mathbf{1}\mathbf{1}^\top = \frac{n}{n-1}\boldsymbol{I}_n - \frac{n}{n-1}\hat{\mathbf{1}}\hat{\mathbf{1}}^\top \succeq \mathbf{0}, \tag{9.16}$$

where $\hat{\mathbf{1}} = \mathbf{1}/\sqrt{n}$, a unit vector. Thus it suffices to consider the remaining direct summand,

$$\begin{bmatrix} 1 & -\frac{1}{n-1}\mathbf{1}^\top \\ -\frac{1}{n-1}\mathbf{1} & \boldsymbol{Y}^{(4)} \end{bmatrix} \overset{?}{\succeq} \mathbf{0}. \tag{9.17}$$

By the Schur complement reduction, it furthermore suffices to consider the matrix

$$\boldsymbol{Y}^{(4)} - \frac{1}{(n-1)^2}\mathbf{1}\mathbf{1}^\top \overset{?}{\succeq} \mathbf{0}. \tag{9.18}$$

To this end, we will try to understand the spectrum of $\boldsymbol{Y}^{(4)}$. More generally, we will see that our reasoning will also describe the matrix of any $\boldsymbol{Z} \in \mathbb{R}^{\binom{n}{2}\times\binom{n}{2}}$ whose entries are given by

$$Z_{S,T} = g(|S\triangle T|), \tag{9.19}$$

as is the case both for $\boldsymbol{Y}^{(4)}$ and for the whole left-hand side of (9.18).

We will first take a very hands-on approach, and then will discuss how our calculations generalize in several fruitful directions. We are interested in finding eigenvectors $\boldsymbol{v} \in \mathbb{R}^{\binom{[n]}{2}}$ of $\boldsymbol{Z}$ as above. We compute

$$(\boldsymbol{Z}\boldsymbol{v})_{\{i,j\}} = g(0)v_{\{i,j\}} + g(2)\sum_{k\notin\{i,j\}}(v_{\{i,k\}} + v_{\{j,k\}}) + g(4)\sum_{\{k,\ell\}\cap\{i,j\}=\varnothing} v_{\{k,\ell\}} \overset{?}{=} \lambda v_{\{i,j\}}. \tag{9.20}$$

We first observe that if $\boldsymbol{v} = \mathbf{1}$ then this eigenvector equation indeed holds, with eigenvalue given by counting the nubmer of terms in each sum above,

$$\lambda_0 := g(0) + (n-2)g(2) + \binom{n-2}{2}g(4). \tag{9.21}$$

Now, suppose $\langle \boldsymbol{v}, \mathbf{1} \rangle = 0$ (as must be the case for every other eigenvector). Using this, we may reduce the eigenvector equation, removing the last sum:

$$(\boldsymbol{Z}\boldsymbol{v})_{\{i,j\}} = (g(0) - g(4))v_{\{i,j\}} + (g(2) - g(4)) \sum_{k \notin \{i,j\}} (v_{\{i,k\}} + v_{\{j,k\}}) + g(4) \underbrace{\sum_{\{k,\ell\}} v_{\{k,\ell\}}}_{= \langle \boldsymbol{v}, \mathbf{1} \rangle = 0}. \tag{9.22}$$

The next step is not so obvious, but we propose considering $v_{\{i,j\}} = u_i + u_j$ for some $\boldsymbol{u} \in \mathbb{R}^n$. In order to have $\langle \boldsymbol{v}, \mathbf{1} \rangle = 0$, we must have $\langle \boldsymbol{u}, \mathbf{1} \rangle = 0$. For such $\boldsymbol{v}$, we have

$$\begin{aligned}
(\boldsymbol{Z}\boldsymbol{v})_{\{i,j\}} &= (g(0) - g(4))(u_i + u_j) + (g(2) - g(4)) \sum_{k \notin \{i,j\}} (u_i + u_j + 2u_k) \\
&= (g(0) + (n-2)g(2) - (n-1)g(4))(u_i + u_j) + 2(g(2) - g(4)) \sum_{k \notin \{i,j\}} u_k \\
&= (g(0) + (n-2)g(2) - (n-1)g(4))(u_i + u_j) - 2(g(2) - g(4))(u_i + u_j) \\
&= (g(0) + (n-4)g(2) - (n-3)g(4))v_{\{i,j\}},
\end{aligned} \tag{9.23}$$

so indeed any such $\boldsymbol{v}$ is an eigenvector with eigenvalue

$$\lambda_1 := g(0) + (n-4)g(2) - (n-3)g(4). \tag{9.24}$$

Let us write V_1 for the span of these vectors, which has dimension $\dim(V_1) = n - 1$ (due to the constraint that $\langle \boldsymbol{u}, \mathbf{1} \rangle = 0$).

Finally, suppose $\boldsymbol{v} \in (\mathbf{1} \oplus V_1)^\perp = \mathbf{1}^\perp \cap V_1^\perp$. We further simplify the eigenvector equation using that $\boldsymbol{v} \in V_1^\perp$. Namely, let $\boldsymbol{w} \in \mathbb{R}^{\binom{n}{2}}$ have $w_{\{i,j\}} = u_i + u_j$ for $\boldsymbol{u} = \boldsymbol{e}_i + \boldsymbol{e}_j$. We then have $w_{\{k,\ell\}} = |\{i,j\} \cap \{k,\ell\}|$. We have

$$0 = \langle \boldsymbol{w}, \boldsymbol{v} \rangle = 2v_{\{i,j\}} + \sum_{k \notin \{i,j\}} (v_{\{i,k\}} + v_{\{j,k\}}). \tag{9.25}$$

Applying this in the eigenvector equation (starting with our previous reduction), we have

$$(\boldsymbol{Z}\boldsymbol{v})_{\{i,j\}} = (g(0) - g(4))v_{\{i,j\}} - 2(g(2) - g(4))v_{\{i,j\}} = (g(0) - 2g(2) + g(4))v_{\{i,j\}}, \tag{9.26}$$

so that any such vector is an eigenvector with eigenvalue

$$\lambda_2 := g(0) - 2g(2) + g(4). \tag{9.27}$$

Let us write V_2 for the span of these vectors, which has dimension $\dim(V_2) = \dim((\mathbf{1} \oplus V_1)^\perp) = \binom{n}{2} - n$.

We thus arrive at the following general linear-algebraic fact.

Lemma 9.4. *Suppose that $\boldsymbol{Z} \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}}$ with entries given by $Z_{S,T} = g(|S \triangle T|)$. Then, $\boldsymbol{Z}$ has at most three distinct eigenspaces $V_0 = \mathsf{span}(\mathbf{1}), V_1, V_2$ as described above, with associated eigenvalues*

$$\lambda_0 = g(0) + (n-2)g(2) + \binom{n-2}{2} g(4), \tag{9.28}$$

$$\lambda_1 = g(0) + (n-4)g(2) - (n-3)g(4), \tag{9.29}$$

$$\lambda_2 = g(0) - 2g(2) + g(4). \tag{9.30}$$

Lemma 9.3 follows from Lemma 9.4 by a straightforward calculation that we leave as an exercise. In particular, applying Lemma 9.4 to $\boldsymbol{Y}^{(4)}$ itself (where $g = f$) and using that $\mathbf{1}$ is an eigenvector of $\boldsymbol{Y}^{(4)}$, it suffices to check that, with $\lambda_0, \lambda_1, \lambda_2$ computed with $g = f$, we have

$$\lambda_0 \geq \frac{\binom{n}{2}}{(n-1)^2}, \tag{9.31}$$

$$\lambda_1 \geq 0, \tag{9.32}$$

$$\lambda_2 \geq 0. \tag{9.33}$$

9.3 Spectra of Matrices with Entrywise Symmetry

It likely seems rather magical that we were able to obtain a result like Lemma 9.4. To somewhat demystify our calculations, let us indicate two ways in which our approach can be considerably generalized.

9.3.1 Representation Theory

Suppose that G is a finite group acting on a finite set X. Then, G also acts on $\mathbb{R}^X$ by permutation matrices: the action of G on X may be identified with a group homomorphism $\phi : G \to S_{|X|}$, and likewise $\mathbb{R}^X$ may be identified with $\mathbb{R}^{|X|}$ (in both cases fixing an ordering of X), in which case $g \in G$ acts as the permutation matrix of $\phi(g)$, which we denote $\boldsymbol{\Pi}_{\phi(g)}$.

Suppose also that we have a matrix $\boldsymbol{Z} \in \mathbb{R}^{X \times X}_{\mathsf{sym}}$ that *commutes* with this action, which may be written in equivalent ways as having, for all $g \in G$,

$$\boldsymbol{\Pi}_{\phi(g)} \boldsymbol{Z} = \boldsymbol{Z} \boldsymbol{\Pi}_{\phi(g)} \quad \Leftrightarrow \quad \boldsymbol{\Pi}_{\phi(g)} \boldsymbol{Z} \boldsymbol{\Pi}_{\phi(g)}^{\top} = \boldsymbol{Z} \tag{9.34}$$

$$\Leftrightarrow \quad Z_{\phi(g)(x),\phi(g)(y)} = Z_{x,y} \text{ for all } x, y \in X. \tag{9.35}$$

Our setting was a special case of this setting. We had $G = S_n$, so that the homomorphism ϕ was just the identity, and had $X = \binom{[n]}{2}$, where for $S = \{i, j\} \in X$ and $g \in S_n$ we had the action $g(S) := \{g(i), g(j)\}$. Any $\boldsymbol{Z}$ with $Z_{S,T} = g(|S \triangle T|)$ then satisfies the invariance (9.35) (indeed, one may show that the invariance (9.35) is *equivalent* to $\boldsymbol{Z}$ having this form for some f).

The representation theory of finite groups provides us with a structure theory of such matrices $\boldsymbol{Z}$ (in addition to treating many further generalizations of this setup). In particular, it shows that the eigenspaces of such $\boldsymbol{Z}$ are *invariant subspaces* of the action of G: these subspaces $V \subset \mathbb{R}^X$ must have $gv \in V$ for every $g \in G$ and $v \in V$. (One may check that the subspaces V_0, V_1, V_2 from our example satisfy this property for the action of S_n on $\mathbb{R}^{\binom{[n]}{2}}$.)

Representation theory also describes the possible decompositions of $\mathbb{R}^X$ into invariant subspaces; in the language of representation theory, the above setup makes $\mathbb{R}^X$ (equipped with the S_n action) a *representation* of G, and any invariant $V \subset \mathbb{R}^X$ is a *subrepresentation* (which, by invariance, is itself a representation), while a minimal invariant subspace is an *irreducible* subrepresentation. There is a rich combinatorial theory characterizing

the irreducible representations of S_n in particular, allowing analogs of the decomposition $\mathbb{R}^{\binom{[n]}{2}} = V_0 \oplus V_1 \oplus V_2$ to be computed in many cases.

Some standard mathematical references are [Ful97, FH04], while [Dia88] gives a more down-to-earth treatment with a particular focus on $G = S_n$.

9.3.2 Association Schemes

There is another approach to understanding matrices like $\boldsymbol{Z}$, which generalizes to a different, usually narrower, class of matrices, but has the advantage of sometimes allowing us to understand the *eigenvalues* of $\boldsymbol{Z}$ without a detailed understanding of the *eigenspaces*.

Definition 9.5. *An* association scheme *is a set of matrices $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_m \in \{0,1\}^{N \times N}_{\mathsf{sym}}$ satisfying the following properties:*

1. $\boldsymbol{A}_1 = \boldsymbol{I}_N$,
2. $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{1}\boldsymbol{1}^\top$,
3. $\boldsymbol{A}_i \boldsymbol{A}_j = \sum_{k=1}^m c_{ijk} \boldsymbol{A}_k$ *for some* $c_{ijk} \in \mathbb{R}$ *and all* $i, j \in [m]$.

We may view each $\boldsymbol{A}_i$ as a *relation* on $[N] \times [N]$, where for every $a, b \in [N]$, (a, b) belongs to exactly one relation among those specified by $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$. Often it is a useful intuition to view the unique i for which $(\boldsymbol{A}_i)_{ab} = 1$ to be a "distance" between a and b specified by the association scheme.

The case we saw earlier was the association scheme with $N = \binom{n}{2}$, where $(\boldsymbol{A}_{i/2+1})_{S,T} = \mathbb{1}\{|S \triangle T| = i\}$ for $i = 0, 2, 4$ (and thus $m = 3$). One may check that Condition 3 above indeed holds for this example. The $\boldsymbol{Z}$ we considered had

$$\boldsymbol{Z} = g(0)\boldsymbol{A}_1 + g(2)\boldsymbol{A}_1 + g(4)\boldsymbol{A}_2. \tag{9.36}$$

This, and its generalizations to larger subsets of $[n]$, are called *Johnson schemes* and are among the best-studied examples of association schemes.

The most restrictive property of association schemes is Condition 3 above, which, along with Condition 1, may be seen as imposing that the $\boldsymbol{A}_i$ generate a *subalgebra* of $\mathbb{R}^{N \times N}_{\mathsf{sym}}$. One consequence is that the $\boldsymbol{A}_i$ commute:

$$\boldsymbol{A}_i \boldsymbol{A}_j = (\boldsymbol{A}_j \boldsymbol{A}_i)^\top = \left(\sum_{k=1}^m c_{jik} \boldsymbol{A}_k \right)^\top = \sum_{k=1}^m c_{jik} \boldsymbol{A}_k = \boldsymbol{A}_j \boldsymbol{A}_i. \tag{9.37}$$

Thus the $\boldsymbol{A}_i$ are simultaneously diagonalizable, explaining why we could find a simultaneous collection of eigenspaces for any $\boldsymbol{Z}$ in the Johnson scheme, and explaining why the eigenvalues λ_i were *linear* functions of the entry values $g(k)$.

It turns out that, knowing the c_{ijk} (called the *intersection numbers* of the scheme) it is often possible to describe these functions giving the eigenvalues without explicitly describing the eigenspaces. A basic version of this observation is tricky but elementary to derive; we present it in Exercise 9.2.

Some references on further aspects of this theory are [Del73, BCN89, Sei91, GS06], and some uses of association schemes in simplifying semidefinite programs and SOS arguments are discussed in [GR99, MW13], the latter of which is part of a long line of work we will study in Chapter 11.

9.4 Full Proof Strategy for Lemma 9.3

We have only shown, already with some difficulty, the part of Lemma 9.3 concerning a small submatrix of our original $\boldsymbol{Y}$. Let us outline the ideas behind the proof of the full result.

First, note that we may always decompose $\boldsymbol{Y}$ into a direct sum of two principal submatrices, those with set indices S having even and odd size, respectively. For the sake of simplicity, let us just consider the case of $|S|$ even; similar reasoning applies to the other submatrix as well.

Based on our strategy for handling the submatrix of $\boldsymbol{Y}$ indexed by $|S| \in \{0, 2\}$, the following strategy seems appealing: let $\boldsymbol{Y}^{(0)} = [1], \boldsymbol{Y}^{(4)}, \boldsymbol{Y}^{(8)}, \dots$ be the diagonal blocks of $\boldsymbol{Y}$, with $\boldsymbol{Y}^{(4k)} \in \mathbb{R}^{\binom{[n]}{4k} \times \binom{[n]}{4k}}$. Consider repeatedly taking the Schur complement in $\boldsymbol{Y}$ with respect to the smallest diagonal block remaining, and let $\widetilde{\boldsymbol{Y}}^{(4k)}$ be the submatrix indexed by $\binom{[n]}{4k} \times \binom{[n]}{4k}$ after having taken k such Schur complements. For example, from our argument above, we have

$$\widetilde{\boldsymbol{Y}}^{(0)} = [1], \tag{9.38}$$

$$\widetilde{\boldsymbol{Y}}^{(4)} = \boldsymbol{Y}^{(4)} - \frac{1}{(n-1)^2} \mathbf{1}\mathbf{1}^\top. \tag{9.39}$$

By the Schur complement characterization of positive semidefiniteness, it suffices to show that each $\widetilde{\boldsymbol{Y}}^{(4k)} \succeq \mathbf{0}$. Moreover, the computation of iterated Schur complements only requires the inversion of each $\widetilde{\boldsymbol{Y}}^{(4k)}$, along with some (substantial) bookkeeping. Laurent in [Lau03] mentions that a natural proof strategy could be based on proving the following observation.

Lemma 9.6. *For all $k \geq 0$, $\widetilde{\boldsymbol{Y}}^{(4k)}$ belongs to the Johnson scheme on $\mathbb{R}^{\binom{[n]}{2k} \times \binom{[n]}{2k}}$ (and likewise for the analogous statement for the submatrix of $\boldsymbol{Y}$ with $|S|$ odd).*

This would at least imply that there is a relatively straightforward algebraic way to perform the necessary inversion and to complete this inductive proof. Unfortunately, Laurent was not able to prove Lemma 9.6; it was only proved more recently by [KM22] as part of a complete description of the spectrum of $\boldsymbol{Y}$.

Instead, Laurent used the following trickier approach. Note that, since as we showed above $\widetilde{\mathbb{E}}[(\sum_{i=1}^n x_i) p(\boldsymbol{x})] = 0$ for any p, $\boldsymbol{Y} \in \mathbb{R}^{\binom{[n]}{\leq (n-1)/2} \times \binom{[n]}{\leq (n-1)/2}}$ has a kernel of dimension at least $\binom{n}{\leq (n-1)/2 - 1}$. Thus, by Cauchy's interlacing theorem, it suffices to identify a strictly positive definite principal submatrix of $\boldsymbol{Y}$ of dimension $\binom{n}{\leq (n-1)/2} - \binom{n}{\leq (n-1)/2 - 1} = \binom{n}{(n-1)/2}$.

Laurent identifies such a submatrix as the one indexed by $\binom{[n-1]}{(n-1)/2} \cup \binom{[n-1]}{(n-1)/2 - 1}$. A few observations are in order. First, this submatrix indeed has the correct dimension, since a basic binomial coefficient identity gives

$$\binom{n-1}{(n-1)/2} + \binom{n-1}{(n-1)/2 - 1} = \binom{n}{(n-1)/2}. \tag{9.40}$$

Second, this submatrix itself decomposes as the direct sum of the two submatrices indexed by $\binom{[n-1]}{(n-1)/2}$ and $\binom{[n-1]}{(n-1)/2-1}$, as $(n-1)/2$ and $(n-1)/2 - 1$ have opposite parity. Third, it is reasonable to expect these submatrices to be strictly positive definite, since they correspond

to evaluations of $\widetilde{\mathbb{E}}$ on the restricted space of polynomials $\mathbb{R}[x_1, \ldots, x_{n-1}]$, where we omit x_n, and thus the "problematic" linear form $\sum_{i=1}^n x_i$ that generates elements of the kernel of $\boldsymbol{Y}$ does not belong to this space. Finally, each of these two submatrices belong to a suitable Johnson scheme since they involve set indices of fixed size, so their eigenvalues can be computed explicitly. Laurent's proof of Lemma 9.3 in [Lau03] (and likewise Grigoriev's proof of a similar statement in [Gri01a]) computes these eigenvalues, and uses some rather elaborate manipulations of sums of binomial coefficients to show that they are positive.

Exercises

Exercise 9.1. *Follow the steps below to give an alternate motivation for the Grigoriev-Laurent pseudoexpectation* $\widetilde{\mathbb{E}}$.[1]

1. *Show the binomial coefficient identity*

$$\sum_{\ell=0}^{2k} \binom{m}{\ell} \binom{m}{2k-\ell} (-1)^\ell = \binom{m}{k} (-1)^k. \tag{9.41}$$

2. *Suppose* n *is even, and let* $\mu = \mathsf{Unif}(\{\boldsymbol{x} \in \{\pm 1\}^n : \sum_{i=1}^n x_i = 0\})$. *Let* $S \subseteq [n]$. *Compute* $\mathbb{E}_{\boldsymbol{x} \sim \mu}[\boldsymbol{x}^S]$, *simplified using the result of Part 1 to a closed form in terms of binomial coefficients.*

3. *Now, suppose* n *is odd. Describe a formal extension of binomial coefficients to fractional inputs, in particular making sense of coefficients of the form* $\binom{n/2}{k}$ *when* n *is odd, so evaluating the result of Part 2 with odd* n *and this formal extension recovers the Grigoriev-Laurent values of* $\widetilde{\mathbb{E}}[\boldsymbol{x}^S]$ *as in* (9.9).

Exercise 9.2. *Recall the definition of an association scheme generated by* $\boldsymbol{I}_N = \boldsymbol{A}_1, \ldots, \boldsymbol{A}_m \in \{0,1\}^{N \times N}_{\mathsf{sym}}$ *from Definition 9.5 The definition implies that the* $\boldsymbol{A}_i$ *commute, and so are simultaneously diagonalizable. That is, there exist projection matrices* $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_d$ *to mutually orthogonal subspaces of* $\mathbb{R}^N$ *and some* λ_{ij} *such that*

$$\boldsymbol{A}_i = \sum_{j=1}^d \lambda_{ij} \boldsymbol{P}_j, \tag{9.42}$$

so that $\lambda_{i1}, \ldots, \lambda_{id}$ *are the eigenvalues of* $\boldsymbol{A}_i$. *In this problem we will show how one can find the* λ_{ij} *from the* c_{ijk}.

1. *Show that the* $\boldsymbol{A}_i$ *are linearly independent.*

2. *Show that if* $\boldsymbol{A}$ *is a symmetric matrix and* $\boldsymbol{P}$ *is the orthogonal projection to the eigenspace of an eigenvalue* λ, *then* $\boldsymbol{P}$ *is a polynomial in* $\boldsymbol{A}$. *Conclude that we may take* $d = m$ *above (i.e., the number of distinct eigenspaces of each* $\boldsymbol{A}_i$ *is at most the total number of* $\boldsymbol{A}_i$ *in the scheme, as we saw in the Johnson scheme example in class), and that in this case the* $\boldsymbol{P}_j$ *are a basis for the span of the* $\boldsymbol{A}_i$.

[1] I learned of this argument from Cristopher Moore, who credits it to Robert Kleinberg.

3. *Show that* $\lambda_{ik}\lambda_{jk} = \sum_{\ell=1}^{d} c_{ij\ell}\lambda_{\ell k}$.

4. *Let* $\boldsymbol{E} \in \mathbb{R}^{m\times m}$ *have* $E_{ij} = \lambda_{ji}$. *Let* $\boldsymbol{L}_i \in \mathbb{R}^{m\times m}$ *have* $(\boldsymbol{L}_i)_{kj} = c_{ijk}$. *Show that* $\boldsymbol{E}$ *is non-singular, and that*

$$\boldsymbol{E}\boldsymbol{L}_i\boldsymbol{E}^{-1} = \mathsf{diag}(\lambda_{i1},\ldots,\lambda_{im}). \tag{9.43}$$

That is, the distinct eigenvalues of $\boldsymbol{A}_i \in \mathbb{R}^{N\times N}_{\mathsf{sym}}$ *are the eigenvalues of* $\boldsymbol{L}_i \in \mathbb{R}^{m\times m}$, *usually a much smaller matrix.*

NOTES

OTHER SOURCES We have mostly followed Laurent's paper [Lau03], as well as some further elaboration mentioned in [KM22]. Our explicit treatment of spectrum of $\boldsymbol{Y}$ in the degree 4 case is mentioned in [DM15] (for a different problem).

OTHER PROOFS Several other proofs of the Grigoriev-Laurent lower bound have appeared in the literature. The approaches of [KLM16, Pot17] give techniques for simplifying the proof of positivity for the "natural" pseudomoments for problems having sufficient degrees of symmetry. For the specific case of optimization over the hypercube, [BGP16] gives a powerful approach that also allows the treatment of *rational* SOS proofs. This approach, using representation theory of the action of S_n on polynomials over $\{\pm 1\}^n$, was also adapted in [KM22] to give a detailed analysis of the matrix $\boldsymbol{Y}$ from the original proofs that we worked with in this chapter.

10 | Case Study 5: Constraint Satisfaction Problems

We now proceed to a more complicated lower bound, where we will use *random* instances of a problem rather than deterministic ones. We will study *constraint satisfaction problems (CSPs)*; we first recall some basic notions used in formulating such problems.

10.1 Background on Constraint Satisfaction Problems

CSPs are formulated over Boolean variables $x_1, \dots, x_n \in \{0,1\}$, where we equate 0 with "False" and 1 with "True." We call a choice of such values an *assignment*. Under these interpretations, we write $\neg x_i$ for the negation of x, $\wedge$ for the Boolean AND operation, $\vee$ for the Boolean OR operation, and $\oplus$ for the Boolean XOR operation.

We will consider the following CSPs:

1. *3-SAT*: For $x_{ab} \in \{x_1, \dots, x_n\}$ and $\tilde{x}_{ab} \in \{x_{ab}, \neg x_{ab}\}$, find an assignment $\boldsymbol{x}$ that makes the following formula true:

$$(\tilde{x}_{11} \vee \tilde{x}_{12} \vee \tilde{x}_{13}) \wedge \cdots \wedge (\tilde{x}_{m1} \vee \tilde{x}_{m2} \vee \tilde{x}_{m3}). \tag{10.1}$$

2. *MAX-3-SAT*: In the above setting, find an assigment $\boldsymbol{x}$ that makes as many of the clauses $\tilde{x}_{i1} \vee \tilde{x}_{12} \vee \tilde{x}_{13}$ true as possible.

3. *3-XORSAT*: For $x_{ab} \in \{x_1, \dots, x_n\}$ and $b_1, \dots, b_m \in \{0,1\}$, find an assignment $\boldsymbol{x}$ that makes the following formula true:

$$(x_{11} \oplus x_{12} \oplus x_{13} \oplus b_1) \wedge \cdots \wedge (x_{m1} \oplus x_{m2} \oplus x_{m3} \oplus b_m). \tag{10.2}$$

4. *MAX-3-XORSAT*: In the above setting, find an assignment $\boldsymbol{x}$ that makes as many of the clauses $x_{i1} \oplus x_{i2} \oplus x_{i3} \oplus b_i$ true as possible.

3-SAT and MAX-3-SAT are well-known to be NP-hard; the same is known for MAX-3-XORSAT as well. It turns out that 3-XORSAT is actually easy to solve. The key insight is that the XOR operation is the same as addition on $\{0,1\}$ modulo 2, i.e., in the finite field $\mathbb{F}_2$ (that is why we use the $\oplus$ notation). In particular, the general instance in (10.2) is the same as the system of linear equations over $\mathbb{F}_2$ given in matrix form by $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{x} \in \mathbb{F}_2^n$ is an indeterminate, $\boldsymbol{b} \in \mathbb{F}_2^m$ consists of the b_i, and $\boldsymbol{A} \in \mathbb{F}_2^{m \times n}$ has entries $A_{ij} = 1$ if x_j

participates in clause i and $A_{ij} = 0$ otherwise. Whether such a system has a solution or not can be determined by Gaussian elimination (as in the maybe more familiar case from real- or complex-valued linear algebra), but this algorithm is "fragile" in the sense that it can only determine exact satisfiability of a 3-XORSAT formula, but not, e.g., the satisfiability of 99% of the clauses.

10.2 Polynomial Encoding and Main Theorem

We will study whether, for unsatisfiable CSPs, low-degree SOS can produce a proof of unsatisfiability (also called *refuting* satisfiability). To do this, we need to formulate CSPs as polynomial systems. Let us work over the $x \in \{\pm 1\}^n$ hypercube; we may encode this with the constraints

$$x_i^2 - 1 = 0 \text{ for all } i \in [n]. \tag{10.3}$$

For 3-SAT, letting $s_{ij} = 1$ if $\widetilde{x}_{ij} = x_{ij}$ and $s_{ij} = -1$ if $\widetilde{x}_{ij} = x_{ij}$, we have the equivalence

$$\widetilde{x}_{i1} \vee \widetilde{x}_{i2} \vee \widetilde{x}_{i3} \quad \Leftrightarrow \quad s_{i1}x_{i1} + s_{i2}x_{i2} + s_{i3}x_{i3} \geq -1. \tag{10.4}$$

For 3-XORSAT, letting $c_i := (-1)^{b_i} \in \{\pm 1\}$, we have the equivalence

$$x_{i1} \oplus x_{i2} \oplus x_{i3} \oplus b_i \quad \Leftrightarrow \quad x_{i1}x_{i2}x_{i3} = c_i. \tag{10.5}$$

Theorem 10.1 ([Gri01b, Sch08])**.** *For arbitrarily large n, there exist unsatisfiable 3-SAT (respectively, 3-XORSAT) formulas on n variables and degree $\Omega(n)$ pseudoexpectations that satisfy the constraints $\{x_i^2 = 1\}_{i=1}^n$ and the constraints of* (10.4) *(respectively,* (10.5)*) for each clause of the formula.*

Moreover, as we will see, these instances can be taken to be "very unsatisfiable": the optimal value of MAX-3-SAT or MAX-3-XORSAT on them is some δm for $\delta \in (0,1)$ nearly as small as possible, while the value of an SOS relaxation of these optimization problems is m until degree $\Omega(n)$ of the SOS hierarchy. Thus, Theorem 10.1 also gives quantitative integrality gaps for the MAX versions of these CSPs.

10.3 Random Instances

We will construct the "hard" instances for Theorem 10.1 by simply taking *random* instances of the appropriate CSP, having $m = \alpha n$ clauses for some large constant α.

Definition 10.2 (Random formulas)**.** *We define* canonical distributions *for 3-SAT and 3-XORSAT formulas. For 3-SAT, we sample from this distribution by taking the support of each clause $\{x_{i1}, x_{i2}, x_{i3}\}$ to be a uniformly random subset of $\{x_1, \dots, x_n\}$ of size three. We then draw $t_{ij} \sim \mathsf{Unif}(\{0,1\})$ and take $\widetilde{x}_{ij} := t_{ij} \oplus x_{ij}$; the s_{ij} used in writing the instance as a polynomial problem are then $s_{ij} = (-1)^{t_{ij}}$. For 3-XORSAT, we draw the x_{ij} in the same way, and then draw $b_j \sim \mathsf{Unif}(\{0,1\})$.*

Proposition 10.3. *For either 3-SAT or 3-XORSAT, there is an $\alpha_0 \in \mathbb{R}$ such that, whenever $\alpha \geq \alpha_0$, with high probability (as $n \to \infty$) a random instance drawn from the appropriate canonical distribution is unsatisfiable.*

Proof. Let F be a random formula. Consider a fixed $\boldsymbol{x} \in \{0,1\}^n$. Let γ be the probability that $\boldsymbol{x}$ satisfies the first clause of the formula. This is easily computed to be

$$\gamma = \begin{cases} 1/2 & \text{for 3-XORSAT,} \\ 7/8 & \text{for 3-SAT.} \end{cases} \tag{10.6}$$

Since the clauses are drawn independently, $\mathbb{P}[\boldsymbol{x} \text{ satisfies } F] = \gamma^m$. Thus, by a union bound,

$$\mathbb{P}[F \text{ is satisfiable}] \leq 2^n \cdot \gamma^m = (2\gamma^\alpha)^n, \tag{10.7}$$

and, since $\gamma < 1$, for sufficiently large α the quantity being raised to the nth power is smaller than 1. □

Indeed, a similar argument using a Hoeffding inequality shows that, with high probability, the value of MAX-3-SAT or MAX-3-XORSAT is with high probability at most $(\gamma + o_{\alpha\to\infty}(1))m$, as we have suggested above.

We next show that it suffices to consider only 3-XORSAT; the result from 3-SAT will follow automatically.

Proposition 10.4. *There is a distribution over pairs of formulas (F, F'), each on n variables and m clauses, where F is a 3-XORSAT formula, F' is a 3-SAT formula, the marginal distribution of each is the appropriate canonical distribution, and $F \Rightarrow F'$ (i.e., whenever $\boldsymbol{x}$ satisfies F, $\boldsymbol{x}$ also satisfies F').*

Proof. Let F' be built as in Definition 10.2, where the t_{ij} are the indicators of whether x_{ij} is negated. Then, let F have clauses of the form

$$x_{i1} \oplus x_{i2} \oplus x_{i3} \oplus (\underbrace{1 \oplus t_{i1} \oplus t_{i2} \oplus t_{i3}}_{=:b_i}). \tag{10.8}$$

Note that, since $t_{ij} \sim \mathrm{Unif}(\{0,1\})$, the law of b_i is $\mathrm{Unif}(\{0,1\})$ as well, so the law of F is the canonical 3-XORSAT distribution.

On the other hand, recalling that in constructing F' we have $\tilde{x}_{ij} = t_{ij} \oplus x_{ij}$, note that the above formula is satisfied if and only if

$$(t_{i1} \oplus x_{i1}) \oplus (t_{i2} \oplus x_{i2}) \oplus (t_{i3} \oplus x_{i3}) = \tilde{x}_{i1} \oplus \tilde{x}_{i2} \oplus \tilde{x}_{i3} = 1. \tag{10.9}$$

But, whenever this holds, at least one of the $\tilde{x}_{ij}$ must be true; that is, the above clause implies $\tilde{x}_{i1} \vee \tilde{x}_{i2} \vee \tilde{x}_{i3}$. Thus, whenever $\boldsymbol{x}$ satisfies F it also satisfies F', as desired. □

Corollary 10.5. *Suppose that Theorem 10.1 holds with high probability for random 3-XORSAT formulas with sufficiently large α. Then, it also holds with high probability for random 3-SAT formulas with sufficiently large α.*

Proof. Let (F, F') be drawn as in Proposition 10.4. Suppose α is large enough that F' is unsatisfiable with high probability and that Theorem 10.1 applies. Since $F \Rightarrow F'$ under this joint distribution, F is also unsatisfiable with high probability.

Suppose $\widetilde{\mathbb{E}}$ is a degree $\Omega(n)$ pseudoexpectation satisfying the constraints generated by F (a 3-XORSAT formula). We then claim that $\widetilde{\mathbb{E}}$ also satisfies the constraints generated by F'. It suffices to show the implication clause by clause, which amounts to showing that there is an SOS proof of $x + y + z \geq -1$ (the expression of a 3-SAT clause being satisfied) under the constraints $xyz - 1 = x^2 - 1 = y^2 - 1 = z^2 - 1 = 0$ (the expression of the corresponding 3-XORSAT clause being satisfied, plus the hypercube constraints).

There is in fact a degree 8 SOS proof. Recall from Exercise 2.6 on the triangle inequalities that there exist $q_1, q_2, q_3 \in \mathbb{R}[x, y, z]$ and $s \in \mathsf{SOS}$ such that $\deg(q_i), \deg(s) \leq 2$

$$xy + yz + xz + 1 = (x^2 - 1)q_1 + (y^2 - 1)q_2 + (z^2 - 1)q_3 + s. \tag{10.10}$$

Multiplying either side by xyz and applying the constraint $xyz - 1 = 0$ then leaves $x + y + z + 1$ on the left-hand side, and a term generated by the constraints plus s on the right-hand side. □

Justified by this result, we will only look at the case of 3-XORSAT from now on. Let us revise our notation to make this slightly simpler. Let $S_1, \dots, S_m \sim \mathsf{Unif}(\binom{[n]}{3})$, and $c_1, \dots, c_m \sim \mathsf{Unif}(\{\pm 1\})$, all independently. Then, to prove Theorem 10.1, it suffices to show that there exists a pseudoexpectation $\widetilde{\mathbb{E}}$ of degree $D \geq \epsilon n$ for some $\epsilon > 0$ respecting the constraints

$$x_i^2 - 1 = 0 \quad \text{for all } i \in [n], \tag{10.11}$$

$$\boldsymbol{x}^{S_j} - c_j = 0 \quad \text{for all } j \in [m]. \tag{10.12}$$

(Recall the notation $\boldsymbol{x}^{S_j} = \prod_{a \in S_j} x_a$.)

10.4 Pseudoexpectation Construction

Unlike the situation Chapter 9, we do not have enough constraints and symmetry to force us to pick a single highly symmetric $\widetilde{\mathbb{E}}$ satisfying these constraints. Instead, we will make the "simplest possible" choice after making sure that all immediate consequences of the constraints are satisfied. Note that, as before, it suffices to specify the multilinear pseudomoments $\widetilde{\mathbb{E}}\boldsymbol{x}^S$ for $S \in \binom{[n]}{\leq D}$. We do this by the following iterative process.

1. Define $\widetilde{\mathbb{E}}\boldsymbol{x}^{\emptyset} = \widetilde{\mathbb{E}}1 := 1$.
2. Define $\widetilde{\mathbb{E}}\boldsymbol{x}^{S_j} := c_j$ for all $j \in [m]$.
3. While it is possible to do so, make an arbitrary choice of $S, T \in \binom{[n]}{\leq D}$ such that $\widetilde{\mathbb{E}}\boldsymbol{x}^S$ and $\widetilde{\mathbb{E}}\boldsymbol{x}^T$ have both been defined, but $\widetilde{\mathbb{E}}\boldsymbol{x}^{S \triangle T}$ has not yet been defined and has $|S \triangle T| \leq D$. Then, set $\widetilde{\mathbb{E}}\boldsymbol{x}^{S \triangle T} := (\widetilde{\mathbb{E}}\boldsymbol{x}^S)(\widetilde{\mathbb{E}}\boldsymbol{x}^T) \in \{\pm 1\}$.
4. For all remaining $S \in \binom{[n]}{\leq D}$ not yet defined, set $\widetilde{\mathbb{E}}\boldsymbol{x}^S := 0$.

This is a natural choice since all of the definitions made in Step 3 have SOS proofs based on multiplying some of the constrained $\boldsymbol{x}^{S_j}$ together and then removing repeated powers using the hypercube constraints $x_i^2 = 1$ repeatedly. (It turns out that we do not need to worry about whether these SOS proofs actually have degree smaller than D; it will not be a problem for us to enforce more of these constraints than we strictly need to.)

What could go wrong with this construction? The main issue is that the procedure in Step 3 does not *a priori* produce a unique $\widetilde{\mathbb{E}}$: different orderings of processing the assignments in Step 3 could yield different $\widetilde{\mathbb{E}}$. Indeed, if it is possible to derive different values of some $\widetilde{\mathbb{E}}\boldsymbol{x}^S$ by short sequences of this procedure, then SOS proves a *contradiction* derived from the constraints, in which case SOS *would* refute satisfiability, since there could not exist $\widetilde{\mathbb{E}}$ satisfying the constraints! Thus to show our lower bound it is important that we exclude this possibility.

To do this, we will reason in more logical terms, describing how a value of $\widetilde{\mathbb{E}}\boldsymbol{x}^S$ is "derived" from the constraints in Step 3 above. Instead of keeping track of the sets A for which $\widetilde{\mathbb{E}}\boldsymbol{x}^A$ gets defined in this process, we keep track of the subsets of *constraints* that are involved in each step of the derivation.

Definition 10.6. *A* D-derivation *of* S *is a sequence* $T_0, T_1, \ldots, T_t \subseteq [m]$ *satisfying the following properties:*

1. $T_0 = \emptyset$*;*
2. *for all* $i \geq 1$*, either* $|T_i| = 1$ *or* $T_i = T_j \triangle T_k$ *for some* $0 \leq j, k \leq i-1$*;*
3. $|\triangle_{a \in T_i} S_a| \leq D$ *for all* $0 \leq i \leq t$*; and*
4. $\triangle_{a \in T_t} S_a = S$*.*

The following is the key result making $\widetilde{\mathbb{E}}$ well-defined.

Lemma 10.7. *With high probability, all* $S \subseteq [n]$ *with* $|S| \leq D$ *having any* D*-derivation have the same final constraint set* T_t *in all* D*-derivations of* S*.*

If this is true then, clearly, $\widetilde{\mathbb{E}}\boldsymbol{x}^S$ is defined in Step 3 if and only if there exists a D-derivation of S, and if $T_0, \ldots, T_t$ is such a derivation, then the unique value assigned is $\widetilde{\mathbb{E}}\boldsymbol{x}^S := \prod_{a \in T_t} c_a$.

We introduce the following bookkeeping mechanism that will be useful for proving Lemma 10.7.

Definition 10.8. *The* variable-constraint graph $G = (V, E)$ *of our formula is the bipartite graph with* n *vertices on the left side of the bipartition denoted* $\mathcal{L} \subset V$ *and identified with* $[n]$*,* m *vertices on the right side denoted* $\mathcal{R} \subset V$ *and identified with* $[m]$*, and an edge* $\{i, j\}$ *connecting* $i \in \mathcal{L} = [n]$ *to* $j \in \mathcal{R} = [m]$ *if and only if* $i \in S_j$ *(i.e., if and only if* x_i *participates in the* j*th clause of the 3-XORSAT formula).*

In our setup, G is a random bipartite graph that is "right-3-regular," i.e., having $\deg(v) = 3$ for all $v \in \mathcal{R}$.

We note that Lemma 10.7 has nothing to do with the "right-hand sides" c_j of the 3-XORSAT clauses; it only makes a statement about G. It turns out that its statement is closely related to *expansion* in G.

Definition 10.9. *We say a bipartite graph G as in Definition 10.8 is a* (β, γ)-expander *if, for all $T \subseteq \mathcal{R}$ with $|T| \leq \beta m$, we have $|\partial T| \geq \gamma |T|$, where $\partial T \subseteq \mathcal{L}$ is the set of neighbors of T.*

Lemma 10.10. *For all $\alpha > 0$ and $0 < \gamma < 2$, there exists $\beta = \beta(\alpha, \delta) > 0$ such that the random bipartite graph G associated to the random 3-XORSAT formula on n variables and $m = \alpha n$ clauses is with high probability a (β, γ)-expander.*

We note that, by considering $T = \partial x$ for a "typical" $x \in \mathcal{L}$, we have $|\partial T| \lesssim 2|T|$ since all vertices of T have a common neighbor in x and each has at most two other neighbors, so $\gamma < 2$ is a necessary restriction.

Proof of Lemma 10.10. Let $\widetilde{\beta} := \alpha\beta$, so that $\beta m = \widetilde{\beta} n$ We rewrite

$$\begin{aligned}
&\mathbb{P}[G \text{ is not a } (\beta, \gamma)\text{-expander}] \\
&\quad = \mathbb{P}[\text{there exist } S \subseteq \mathcal{L}, T \subseteq \mathcal{R} \text{ with } |S| \leq \widetilde{\beta}\gamma n, |T| \leq \widetilde{\beta} n, S = \partial T]
\end{aligned}$$

and, bounding this by the number of ways to choose S, multiplied by the number of ways to choose $\widetilde{\beta} n$ neighborhoods of size 3 inside S, multiplied by the probability that some $T \subseteq \mathcal{R}$ has exactly those neighborhoods gives

$$\leq \binom{n}{\widetilde{\beta}\gamma n} \cdot \frac{\binom{\widetilde{\beta}\gamma n}{3}^{\widetilde{\beta} n}}{(\widetilde{\beta} n)!} \cdot \left(\frac{m}{\binom{n}{3}} \right)^{\widetilde{\beta} n}$$

Using that $k! \geq (k/e)^k$ and $(n/k)^k \leq \binom{n}{k} \leq (en/k)^k$, we further bound

$$\begin{aligned}
&\leq \left(\frac{en}{\widetilde{\beta}\gamma n} \right)^{\widetilde{\beta}\gamma n} \left(\frac{e\widetilde{\beta}^3 \gamma^3 n^3}{\widetilde{\beta} n} \right)^{\widetilde{\beta} n} \left(\frac{27\alpha}{n^2} \right)^{\widetilde{\beta} n} \\
&= \left(\frac{27 e^{\gamma+1} \gamma^3 \alpha}{\gamma^\gamma} \widetilde{\beta}^{2-\gamma} \right)^{\widetilde{\beta} n}.
\end{aligned}$$

Since $\gamma < 2$, for sufficiently small $\widetilde{\beta} > 0$ (achieved by choosing sufficiently small $\beta > 0$) the quantity in parentheses is smaller than 1, giving the result. □

Proof of Lemma 10.7. Suppose that there are two different D-derivations $\emptyset, T_1^{(1)}, \ldots, T_{t_1}^{(1)}$ and $\emptyset, T_1^{(2)}, \ldots, T_{t_2}^{(2)}$ of the same set S, and, for the sake of contradiction, suppose $T_{t_1}^{(1)} \neq T_{t_2}^{(2)}$.

On the other hand, we must have $\triangle_{a \in T_{t_1}^{(1)}} S_a = S = \triangle_{a \in T_{t_2}^{(2)}} S_a$. Consider then the concatenation of these two derivations, with a final step added at the end:

$$(\emptyset, T_1, \ldots, T_t) := (\emptyset, T_1^{(1)}, \ldots, T_{t_1}^{(1)}, T_1^{(2)}, \ldots, T_{t_2}^{(2)}, T_{t_1}^{(1)} \triangle T_{t_2}^{(2)}), \tag{10.13}$$

so that $t = t_1 + t_2 + 1$. By the above observations, we must have $T_t \neq \emptyset$, and $\triangle_{a \in T_t} S_a = \emptyset$. Thus, this is a non-trivial D-derivation of $\emptyset$.

The general plan of the rest of the proof is to derive a contradiction as follows. First, we show that T_t must be large. Note that, since $\triangle_{a \in T_t} S_a = \emptyset$, every vertex in ∂T_t must be incident to at least two elements of T_t. Since there are $3|T_t|$ edges leaving T_t, we find that

$|\partial T_t| \le \frac{3}{2}|T_t|$. Now, let β be small enough that G is with high probability a $(\beta/\alpha, \frac{7}{4})$ expander by Lemma 10.10. Choose $\epsilon = \frac{1}{100}\beta$ and set $D = \epsilon n = \frac{1}{100}\beta n$. Since T_t has small expansion, we must have

$$|T_t| \ge \frac{\beta}{\alpha} m = \beta n = 100D. \tag{10.14}$$

Next, we note that since either $|T_i| = 1$ or $|T_i| \le |T_j| + |T_k|$ for some $j, k < i$, the maximum size of the $|T_i|$ encountered thus far at most doubles at each step in the derivation. At an intuitive level, the T_i must grow fairly slowly. In particular, suppose i is the smallest index for which $|T_i| \ge 10D$. Then, in fact T_i must be of "intermediate" size, having

$$10D \le |T_i| \le 20D. \tag{10.15}$$

From the right-hand inequality, we have $|T_i| \le \frac{1}{5}\beta n$, so the expansion property applies to T_i, and we have $|\partial T_i| \ge \frac{7}{4}|T_i|$.

Finally, we will show that this large expansion means that $|\triangle_{a \in T_i} S_a|$ must be large. Let k be the number of vertices of $\partial T_i \subseteq \mathcal{L}$ that have exactly one neighbor in T_i. Since each corresponding variable occurs only once in the symmetric difference, $|\triangle_{a \in T_i} S_a| \ge k$. Thus to derive a contradiction it suffices to show that $k > D$.

Since the total number of edges leaving T_i is $3|T_i|$, we have $k + 2(|\partial T_i| - k) \le 3|T_i|$, or $k \ge 2|\partial T_i| - 3|T_i| \ge 2 \cdot \frac{7}{4}|\partial T_i| - 3|T_i| = \frac{1}{2}|T_i| \ge 5D$, as desired. □

10.5 Proof of Theorem 10.1

We are now ready to proceed to our proof of the main lower bound.

Proof of Theorem 10.1. Choose α sufficiently large that with high probability a random 3-XORSAT formula on n variables and $m = \alpha n$ clauses is unsatisfiable, and choose $D = \epsilon n$ sufficiently small as in the proof of Lemma 10.7. We will work on the event that the random formula is unsatisfiable, and that the variable-clause graph G is an expander as needed in the proof of Lemma 10.7.

On this high probability event, consider $\widetilde{\mathbb{E}}$ as constructed above. This $\widetilde{\mathbb{E}}$ respects the hypercube constraints $x_i^2 - 1 = 0$ by construction. Consider one of the other constraints $\boldsymbol{x}^{S_j} - c_j = 0$. To show that $\widetilde{\mathbb{E}}$ respects this constraint, it suffices to show that $\widetilde{\mathbb{E}}[(\boldsymbol{x}^{S_j} - c_j)\boldsymbol{x}^T] = 0$ for all $|T| \le D - 3$, which is equivalent to showing

$$\widetilde{\mathbb{E}}\boldsymbol{x}^{S_j \triangle T} \stackrel{?}{=} c_j \widetilde{\mathbb{E}}\boldsymbol{x}^T = (\widetilde{\mathbb{E}}\boldsymbol{x}^{S_j})(\widetilde{\mathbb{E}}\boldsymbol{x}^T) \tag{10.16}$$

If $\widetilde{\mathbb{E}}\boldsymbol{x}^{S_j \triangle T}$ is set to a non-zero value in our construction, then $\widetilde{\mathbb{E}}\boldsymbol{x}^{(S_j \triangle T) \triangle S_j} = \widetilde{\mathbb{E}}\boldsymbol{x}^T$ must also be set to $\widetilde{\mathbb{E}}\boldsymbol{x}^T = (\widetilde{\mathbb{E}}\boldsymbol{x}^{S_j \triangle T})(\widetilde{\mathbb{E}}\boldsymbol{x}^{S_j})$. Conversely, if $\widetilde{\mathbb{E}}\boldsymbol{x}^T$ is set to a non-zero value, then we must also set $\widetilde{\mathbb{E}}\boldsymbol{x}^{S_j \triangle T} = (\widetilde{\mathbb{E}}\boldsymbol{x}^{S_j})(\widetilde{\mathbb{E}}\boldsymbol{x}^T)$. Thus, $\widetilde{\mathbb{E}}\boldsymbol{x}^{S_j \triangle T} \ne 0$ if and only if $\widetilde{\mathbb{E}}\boldsymbol{x}^T \ne 0$, and if both hold then (10.16) also holds. On the other hand, if neither hold, then both sides of (10.16) are zero, so it still holds. Thus $\widetilde{\mathbb{E}}$ respects the clause constraints $\boldsymbol{x}^{S_j} - c_j = 0$ as well.

It remains to show positivity. Consider the pseudomoment matrix $\boldsymbol{Y} \in \mathbb{R}^{\binom{[n]}{\le D/2} \times \binom{[n]}{\le D/2}}$. It is tempting to argue that, whenever $Y_{S,T} = \widetilde{\mathbb{E}}\boldsymbol{x}^S \boldsymbol{x}^T = \widetilde{\mathbb{E}}\boldsymbol{x}^{S \triangle T} \ne 0$, then $Y_{S,T} = (\widetilde{\mathbb{E}}\boldsymbol{x}^S)(\widetilde{\mathbb{E}}\boldsymbol{x}^T)$, so that $\boldsymbol{Y}$ is rank one and positive semidefinite. But this is not quite correct—indeed, it cannot

be correct, since if $\boldsymbol{Y}$ were rank one then it would be an integral point, corresponding to an actual satisfying assignment for the 3-XORSAT formula, which does not exist!

The issue is that we may have $\widetilde{\mathbb{E}}\boldsymbol{x}^S = 0$ or $\widetilde{\mathbb{E}}\boldsymbol{x}^T = 0$ while still having $Y_{S,T} = \widetilde{\mathbb{E}}\boldsymbol{x}^{S\triangle T} \neq 0$, since $S\triangle T$ can be derived in other ways than deriving S and T and then taking their symmetric difference. We thus make the following more nuanced argument. Define the relation $S \sim T$ if $Y_{S,T} = \widetilde{\mathbb{E}}\boldsymbol{x}^{S\triangle T} \neq 0$. It is straightforward to check that this is an equivalence relation, with transitivity following from the symmetric difference identity $(R\triangle S)\triangle(S\triangle T) = R\triangle T$. Let $C_1 \sqcup \cdots \sqcup C_k = \binom{[n]}{\leq D/2}$ be the decomposition into equivalence classes under this relation. Note that then $\boldsymbol{Y}$ decomposes as a direct sum of the principal submatrices indexed by each C_i; call this matrix $\boldsymbol{Y}^{(i)}$.

We will show that *each* $\boldsymbol{Y}^{(i)}$ is a rank one positive semidefinite matrix. Note that all entries of this submatrix are non-zero. Let us fix some $A_i \in C_i$, and suppose $S, T \in C_i$. Then, we have $S\triangle T = (S\triangle A_i)\triangle(T\triangle A_i)$. By the definition of the equivalence relation, both $\widetilde{\mathbb{E}}\boldsymbol{x}^{S\triangle A_i} \neq 0$ and $\widetilde{\mathbb{E}}\boldsymbol{x}^{T\triangle A_i} \neq 0$. So, by Lemma 10.7, we must have $\widetilde{\mathbb{E}}\boldsymbol{x}^{S\triangle T} = (\widetilde{\mathbb{E}}\boldsymbol{x}^{S\triangle A_i})(\widetilde{\mathbb{E}}\boldsymbol{x}^{T\triangle A_i})$, completing the proof. □

Notes

Derandomizing It is interesting to ask whether a similar result to Theorem 10.1 can be achieved without relying on random instances. Randomness plays two roles in the construction: first, the randomness of the variable-clause graph G is used to ensure expansion, and second, the randomness of the right-hand side values c_j is used to ensure unsatisfiability. One might hope to find explicit graphs G satisfying the expansion property in Lemma 10.10; however, as far as I know, deterministic such expander graphs with the amount of flexibility our argument needs have not been constructed. It is, however, possible to produce deterministic unsatisfiable 3-XORSAT instances that admit similar SOS lower bounds using *Tseitin tautologies* on (non-bipartite) expander graphs, as [Gri01b] also showed. However, these are not "very unsatisfiable" in the way that our random instances here are; these results do not show large integrality gaps for SOS relaxations of MAX-3-XORSAT. More recently, achieving the best of both worlds, [DFHT20] gave a construction of a deterministic instance based on *high-dimensional expanders* that does achieve a constant MAX-3-XORSAT integrality gap. Their work leaves an interesting open problems: they only treat SOS degree $D = O(\sqrt{\log n})$, so it remains to improve these results to polynomial or linear degrees.

Open Problem 10.1 (Deterministic MAX-3-XORSAT lower bounds)**.** *Find explicit sequences of 3-XORSAT formulas on n variables and m clauses (with $m, n \to \infty$) such that at most $(\frac{1}{2} + \epsilon)m$ clauses are satisfiable by any assignment, while SOS of degree $D = \Omega(n)$ fails to refute satisfiability.*

11 | Case Study 6: Large Cliques in Random Graphs

To conclude our study of SOS lower bounds, we will look at a more elaborate example from this literature, which introduced the important technique of *pseudocalibration* that has since been used for many other lower bounds. Pseudocalibration was developed to prove SOS lower bounds on relaxations of the *clique number* of random graphs. We recall the basic definitions below.

We write $G(n,p)$ for the *Erdős-Rényi random graph* distribution on graphs with n vertices, where each edge is present independently with probability p. For the sake of brevity, we write $G := G(n, \frac{1}{2})$, as this is the only case we study in detail here. A *clique* in a graph is a complete subgraph. The *clique number*, denoted $\omega(G)$, is the size of the largest clique. We will be interested in the behavior of $\omega(G)$ when $G \sim G$.

Proposition 11.1. *For any $\epsilon > 0$, for $G \sim G$, with high probability $\omega(G) \in [(2-\epsilon)\log_2 n, (2+\epsilon)\log_2 n$.*

Proof. We only give the proof of the upper bound using the first moment method; the lower bound follows by a similar but more involved calculation using the second moment method. By the union bound followed by standard estimates, we have

$$\mathbb{P}[\omega(G) \geq k] \leq \binom{n}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}} \leq n^k 2^{-k^2/2} = 2^{k\log_2 n - k^2/2}. \tag{11.1}$$

When $k \geq (2+\epsilon)\log_2 n$, then this is at most $2^{-\Omega(k)} = n^{-\Omega(1)}$ and thus goes to zero as $n \to \infty$. □

11.1 Planted Clique Model and Information-Theoretic Threshold

While the problem we ultimately study will only involve bounding $\omega(G)$ when $G \sim G$, the study of this problem is motivated by and turns out to be related to the following other random graph distribution.

Definition 11.2 (Planted clique model)**.** *We write $P_k = P_k(n)$ for the random graph distribution where $G \sim P_k$ is sampled by sampling $H \sim G$, choosing $C^\star \sim \mathsf{Unif}(\binom{[n]}{k})$, and letting G equal the union of H with the clique on the vertices of $C^\star$. This special clique on $C^\star$ in P_k is called the* planted *clique.*

The following statistical question was posed by [Jer92] and [Kuč95]:

> "For what k can we *detect* or *recover* the planted clique in $\mathcal{P}_k$?"

To *detect* means, given either $G \sim \mathcal{G}$ or $G \sim \mathcal{P}_k$, to determine which distribution G was sampled from; in our case, this just means to tell whether a clique has been planted in $G \sim \mathcal{G}$ or not. To *recover* means, given $G \sim \mathcal{P}_k$, to return $\widehat{C} = \widehat{C}(G) \in \binom{[n]}{k}$ with $\widehat{C} \approx C^\star$. (In more traditional statistics language, these are *hypothesis testing* and *estimation*, respectively.)

Without computational constraints on these procedures, it is natural to expect that the *threshold* or *critical* value of k around which the possibility of detection and recovery changes is $k_{\mathsf{stat}} \approx 2\log_2 n$, the typical value of $\omega(G)$ under $G \sim \mathcal{G}$. (The subscript indicates that this is the *statistical* threshold of this problem, sometimes also called an *information-theoretic* threshold, as opposed to the *computational* threshold discussed below.) Indeed, this is the case as the following results show.

Proposition 11.3 ([CX16])**.** *If $k \le (2-\epsilon)\log_2 n$, then both detection and recovery are impossible (there is no $f : G \mapsto \{0,1\}$ so that $f(G) = 0$ with high probability under $G \in \mathcal{G}$ and $f(G) = 1$ with high probability under $G \sim \mathcal{P}_k$, and there is no $\widehat{C}(G)$ such that $\widehat{C} = C^\star$ with high probability under $G \sim \mathcal{P}_k$).*

This result was likely folklore before the reference given; we will see an argument for the negative part later.

Proposition 11.4. *If $k \ge (2+\epsilon)\log_2 n$, then the estimator $\widehat{C}$ outputting the largest clique in G equals $C^\star$ with high probability under $G \sim \mathcal{P}_k$, and a hypothesis test thresholding $\omega(G)$ (say, at $2+\epsilon/2$) distinguishes $\mathcal{G}$ and $\mathcal{P}_k$ with high probability.*

This second result is just a direct corollary of Proposition 11.1 and the fact that $\omega(G) \ge k$ with probability 1 when $G \sim \mathcal{P}_k$.

However, computing $\omega(G)$ or finding its maximizer (a maximum clique in G) are NP-hard problems. So, the existence of this test and estimator does not address the more relevant algorithmic question: what is the threshold k_{comp} so that, when $k \gtrsim k_{\mathsf{comp}}$, then an *efficient* (say, polynomial time) algorithm can detect or recover a planted clique? In fact, we will see that there is much evidence that, for this problem, $k_{\mathsf{comp}} \gg k_{\mathsf{stat}}$, a phenomenon called a *statistical-to-computational* or *information-computation gap*.

11.2 Basic Algorithms for Recovering Planted Cliques

To predict how we expect k_{comp} to behave, let us consider two relatively simple classes of algorithms, focusing here on recovery of a planted clique.

11.2.1 Degree Thresholding

The first, quite simple, algorithmic idea, observed by [Kuč95], is that adding the clique $C^\star$ to a random graph increases the degree of the vertices involved in the clique. Thus we may try to estimate $C^\star$ by just finding the highest-degree vertices in the graph and outputting those. The following establishes when this works.

Theorem 11.5 ([Kuč95]). *There is a constant $C > 0$ such that, if $k \geq C\sqrt{n \log n}$, then, letting $\widehat{C}(G)$ output the set of the k vertices of highest degree in G, with high probability under $G \sim \mathcal{P}_k$, $\widehat{C}(G) = C^\star$.*

Proof Sketch. For any fixed $i \in [n]$, the law of $\deg(i)$ when $H \sim \mathcal{G}$ is $\mathsf{Bin}(n-1, \frac{1}{2})$. In particular, these have mean $\approx \frac{n}{2}$ and variance $\approx \frac{\sqrt{n}}{2}$ and are $O(n)$-subgaussian. Thus the maximum degree in $H \sim \mathcal{G}$ is, with high probability, $O(\sqrt{n \log n})$. The same then applies for the maximum degree of all $i \notin C^\star$ under $G \sim \mathcal{P}_k$, and the result follows. □

11.2.2 Spectral Algorithms

It is then natural to ask if this simple idea is optimal or not. A different, more sophisticated approach shows that we may in fact remove the $\sqrt{\log n}$ factor that the degree thresholding algorithm requires.

The idea of this improved algorithm is to reason not in terms of the graph structure but in terms of the spectral properties of the adjacency matrix. Let us derive an approximate description of $\boldsymbol{A} \in \mathbb{R}^{n \times n}_{\mathsf{sym}}$, the adjacency matrix of $G \sim \mathcal{P}_k$. First, consider $\boldsymbol{A}^{(0)}$ the adjacency matrix of $H \sim \mathcal{G}$. Separating it into the mean and the centered fluctuations and ignoring the diagonal, we have

$$\boldsymbol{A}^{(0)} = \mathbb{E}\boldsymbol{A}^{(0)} + (\underbrace{\boldsymbol{A}^{(0)} - \mathbb{E}\boldsymbol{A}^{(0)}}_{=:\boldsymbol{\Delta}}) \approx \frac{1}{2}\mathbf{1}\mathbf{1}^\top + \boldsymbol{\Delta}, \tag{11.2}$$

where $\boldsymbol{\Delta}$ has i.i.d. entries above the diagonal distributed as $\mathsf{Unif}(\{\pm\frac{1}{2}\})$.

Now, we may view the addition of $C^\star$ to H as adding (ignoring both the diagonal and repeated edges) an all-ones submatrix to $\boldsymbol{A}^{(0)}$ indexed by $C^\star$. Letting $\boldsymbol{x}^\star$ have $x^\star_i = \mathbb{1}\{i \in C^\star\}$, we then have

$$\boldsymbol{A} \approx \boldsymbol{A}^{(0)} + \boldsymbol{x}^\star \boldsymbol{x}^{\star\top} \approx \frac{1}{2}\mathbf{1}\mathbf{1}^\top + \boldsymbol{x}^\star \boldsymbol{x}^{\star\top} + \boldsymbol{\Delta}. \tag{11.3}$$

We may then try to estimate $\boldsymbol{x}^\star$ by computing the top eigenvector (that with largest eigenvalue) of $\boldsymbol{A} - \frac{1}{2}\mathbf{1}\mathbf{1}^\top \approx \boldsymbol{x}^\star \boldsymbol{x}^{\star\top} + \boldsymbol{\Delta}$. By the eigenvector perturbation bound of Proposition 5.6, this should give us a good estimate when $\|\boldsymbol{x}^\star \boldsymbol{x}^{\star\top}\| = k \gg \|\boldsymbol{\Delta}\|$. Random matrix theory gives us tools for understanding this remaining norm.

Theorem 11.6 ([Gem80, AGZ10]). *Suppose $\boldsymbol{\Delta} \in \mathbb{R}^{n \times n}_{\mathsf{sym}}$ has $\Delta_{ii} = 0$ and $\Delta_{ij} \sim \rho$ independently for $i < j$, where $\mathbb{E}_{x \sim \rho}[x] = 0$ and $\mathbb{E}_{x \sim \rho}[e^{tx}] < \infty$ for some $t > 0$. Let $\sigma^2 := \mathbb{E}_{x \sim \rho}[x^2]$. Then, $\|\boldsymbol{\Delta}\|/\sqrt{n} \to 2\sigma$ in probability. In particular, with high probability $\|\boldsymbol{\Delta}\| = O(\sqrt{n})$.*

Reasoning very loosely, one may guess this scaling as follows: we have $\mathbb{E}\|\boldsymbol{\Delta}\|_F^2 = O(n^2) = \mathbb{E}\sum_{i=1}^n \lambda_i(\boldsymbol{\Delta})^2$. Thus, we might expect the typical eigenvalue to have $\lambda_i(\boldsymbol{\Delta})^2 = O(n)$, and if the extreme eigenvalues also have this scaling then indeed $\|\boldsymbol{\Delta}\| = O(\sqrt{n})$.

Thus we see that this estimator should succeed when $k \gg \sqrt{n}$. In fact, even when $k \sim \sqrt{n}$, it is possible to start with this spectral estimator and then refine it with a combinatorial procedure to recover $C^\star$, as the following result showed.

Theorem 11.7 ([AKS98]). *There is a constant $C > 0$ and a polynomial-time algorithm such that, if $k \geq C\sqrt{n}$ and $(C^\star, G) \sim \mathcal{P}_k$, then the algorithm returns $C^\star$ with high probability.*

11.2.3 The Planted Clique Hypothesis

The best algorithms we have seen, recovering the planted clique when $k \gtrsim \sqrt{n}$, are still very far from the statistical threshold $k_{\mathsf{stat}} \sim \log n$. However, it is believed that the behavior of the spectral algorithm above is essentially optimal.

Conjecture 11.8. *There is no polynomial-time algorithm that, when $k \ll \sqrt{n}$, either achieves detection between $\mathcal{G}$ and $\mathcal{P}_k$ or recovery under $\mathcal{P}_k$.*

Already some evidence for this conjecture was given by Jerrum in [Jer92], who showed that a natural algorithm based on Markov Chain Monte Carlo does not succeed in quickly recovering the planted clique in this regime. However, for a long time after his work this was the only evidence available for Conjecture 11.8, so the claim was plausible but not overwhelmingly convincing—it seemed just as likely that more powerful algorithms might improve on the performance of spectral algorithms. In the rest of this chapter, we will see that constant-degree SOS relaxations cannot improve on spectral algorithms, giving much stronger evidence for the Conjecture.

11.3 Sum-of-Squares Relaxations: Introduction and Degree 2

When discussing SOS relaxations of the planted clique problem, we will focus on using them for *detection* or distinguishing between $\mathcal{G}$ and $\mathcal{P}_k$. First, notice that $\omega(G)$ may be written as the polynomial optimization problem

$$\omega(G) = \left\{ \begin{array}{ll} \text{maximize} & \sum_{i=1}^n x_i \\ \text{subject to} & x_i^2 - x_i = 0 \text{ for all } i \in [n], \\ & x_i x_j = 0 \text{ for all } i \not\sim_G j \end{array} \right\}, \tag{11.4}$$

where $\sim_G$ is the adjacency relation in G. We note that $i \not\sim_G j$ is equivalent to $A_{ij} = 0$, which will be a useful way to interpret this constraint later. This is a correct formulation because the first constraint imposes $x_i \in \{0,1\}$, while the second imposes that those i for which $x_i = 1$ must form a clique in G. Let us write $\mathsf{SOS}_D(G)$ for the value of the degree D relaxation of this problem.

Since $\mathsf{SOS}_D(G)$ is indeed a relaxation, when $G \sim \mathcal{P}_k$, we must have $\mathsf{SOS}_D(G) \geq k$. Thus, we have the following simple observation.

Proposition 11.9. *Suppose $k = k(n) \in \mathbb{N}$ is increasing and $\epsilon > 0$. If when $G \sim \mathcal{G}$ we have $\mathsf{SOS}_D(G) \leq (1-\epsilon)k$ with high probability as $n \to \infty$, then an algorithm computing $\mathsf{SOS}_D(G)$ and thresholding its value can distinguish $\mathcal{G}$ and $\mathcal{P}_k$ with high probability.*

When we talk about "lower bounds against SOS for the detection problem," we are really talking about showing that this kind of algorithm does not succeed. Thus, it will suffice to show that $\mathsf{SOS}_D(G) \gtrsim \sqrt{n}$, say for any fixed D and with high probability as $n \to \infty$.

Let us consider the first of these relaxations, the case $D = 2$. You may verify that, in the Lasserre form, this program may be written in the form

$$\mathsf{SOS}_2(G) = \left\{ \begin{array}{ll} \text{maximize} & \sum_{i=1}^n x_i \\ \text{subject to} & \boldsymbol{X} := \begin{bmatrix} 1 & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{M} \end{bmatrix} \succeq \boldsymbol{0}, \\ & \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{M} \in \mathbb{R}^{n \times n}_{\mathsf{sym}}, \\ & M_{ii} = x_i \text{ for all } i \in [n], \\ & M_{ij} = 0 \text{ for all } i \not\sim_G j \end{array} \right\}. \tag{11.5}$$

The final constraint may be viewed as stating that $\boldsymbol{M}$ has the same "sparsity pattern" as $\boldsymbol{A}$: M_{ij} is allowed to be non-zero only when A_{ij} is non-zero. In Exercise 11.1, you will show that this SDP is the same as the Lovász ϑ function bounding $\omega(G)$ that you may have encountered before.

Theorem 11.10 ([FK00])**.** *When $G \sim \mathcal{G}$, with high probability $\mathsf{SOS}_2(G) \gtrsim \sqrt{n}$.*

Proof. We build our pseudomoment matrix $\boldsymbol{X}$ to be as naive as possible, satisfying the linear constraints but taking all degree 1 pseudomoments equal, and all non-zero degree 2 pseudomoments equal. Such matrices are, for some $a, b \in \mathbb{R}$,

$$\boldsymbol{X} = \begin{bmatrix} 1 & a\boldsymbol{1}^\top \\ a\boldsymbol{1} & a\boldsymbol{I}_n + b\boldsymbol{A} \end{bmatrix}. \tag{11.6}$$

The objective value of such a matrix is an, so we want to show that it is possible to have $\boldsymbol{X} \succeq \boldsymbol{0}$ with $a \gtrsim 1/\sqrt{n}$.

By the Schur complement condition, $\boldsymbol{X} \succeq \boldsymbol{0}$ if and only if

$$a\boldsymbol{I}_n - a^2\boldsymbol{1}\boldsymbol{1}^\top + b\boldsymbol{A} \succeq \boldsymbol{0}. \tag{11.7}$$

Moreover, using our decomposition for $\boldsymbol{A}$ into the mean and centered fluctuation and being slightly more careful, we have

$$\boldsymbol{A} = \mathbb{E}\boldsymbol{A} + (\underbrace{\boldsymbol{A} - \mathbb{E}\boldsymbol{A}}_{=:\boldsymbol{\Delta}}) = \frac{1}{2}\boldsymbol{1}\boldsymbol{1}^\top - \frac{1}{2}\boldsymbol{I}_n + \boldsymbol{\Delta}. \tag{11.8}$$

Substituting and rearranging, it suffices to show that

$$\left(\frac{b}{2} - a^2\right)\boldsymbol{1}\boldsymbol{1}^\top + \left(a - \frac{b}{2}\right)\boldsymbol{I}_n + b\boldsymbol{\Delta} \succeq \boldsymbol{0}. \tag{11.9}$$

Using that $\boldsymbol{\Delta} \succeq -\|\boldsymbol{\Delta}\|\boldsymbol{I}_n = -C\sqrt{n}\boldsymbol{I}_n$ for some $C > 0$ by Theorem 11.6 and grouping the last two terms together, it suffices to have

$$\frac{1}{2}b \geq a^2, \tag{11.10}$$

$$a \geq \frac{3}{2}bC\sqrt{n}. \tag{11.11}$$

We may achieve this taking $a \sim 1/\sqrt{n}$ and $b \sim 1/n$ with suitable constants, completing the proof. □

11.4 Feige-Krauthgamer Pseudomoments and Kelner's Polynomial

It is natural to try to continue in the same spirit as our degree 2 lower bound for larger degrees D, defining

$$\widetilde{\mathbb{E}}\boldsymbol{x}^S := \mathbb{1}\{S \text{ is a clique in } G\} f(|S|) \tag{11.12}$$

for some $f : \mathbb{N} \to \mathbb{R}$. To derive the "right" values of f, it is helpful to suppose that we augment the underlying polynomial problem with the constraint $\sum_{i=1}^n x_i = k$ and search for *any* pseudoexpectation satisfying this additional constraint, rather than maximizing $\widetilde{\mathbb{E}}[\sum_{i=1}^n x_i]$. (This is a stronger version of the SOS relaxation that many papers in this literature worked with.) In this case, we may calculate roughly

$$k^a = \widetilde{\mathbb{E}} \left(\sum_{i=1}^n x_i \right)^a \tag{11.13}$$

which we suppose is dominated by a-cliques,

$$\approx \#\{a\text{-cliques in } G\} \cdot a! \cdot f(a), \tag{11.14}$$

and since, by the calculation in the proof of Proposition 11.1, we expect the number of a-cliques to be roughly $\binom{n}{a} 2^{-\binom{a}{2}} \sim n^a$ for small a, we have

$$\sim n^a f(a). \tag{11.15}$$

Thus we expect the scaling

$$\widetilde{\mathbb{E}}\boldsymbol{x}^S \asymp \mathbb{1}\{S \text{ is a clique in } G\} \left(\frac{k}{n}\right)^{|S|}. \tag{11.16}$$

Indeed, this is the scaling that we saw in the proof of Theorem 11.10, where $k \sim \sqrt{n}$ and $\widetilde{\mathbb{E}} x_i \sim n^{-1/2}$ and $\widetilde{\mathbb{E}} x_i x_j \sim n^{-1}$ for all $i \sim j$. This choice of pseudomoments (with some small adjustments) was studied by Feige and Krauthgamer in [FK03], who showed that it gives tight lower bounds in the weaker *Lovász-Schrijver* hierarchy of semidefinite programs. For this reason we call these the *FK pseudomoments*, and write $\widetilde{\mathbb{E}} = \widetilde{\mathbb{E}}^{\mathsf{FK}}$ for the pseudoexpectation defined by (11.16).

Another interpretation of this choice is that it "pretends" to behave like an expectation under $G \sim \mathcal{P}_k$. Indeed, we have

$$\left(\frac{k}{n}\right)^{|S|} \approx \mathop{\mathbb{P}}_{(C^\star, G) \sim \mathcal{P}_k} [S \subseteq C^\star] = \mathop{\mathbb{E}}_{(\boldsymbol{x}^\star, G) \sim \mathcal{P}_k} [(\boldsymbol{x}^\star)^S]. \tag{11.17}$$

Starting from this and "brutely" applying the clique constraints by multiplying by the indicator in (11.16) is another way to build the FK pseudomoments.

However, it turns out that there is a problem with using the FK pseudomoments for higher-degree SOS relaxations, whose discovery is attributed to Kelner (see, e.g., discussion

in [Hop18b]). Let us write $\widetilde{\mathbb{E}}_G^{\mathsf{FK}}$ for the FK pseudomoments built from a given graph G. Note that the above observations may be written in the form

$$\mathop{\mathbb{E}}_{G\sim\mathcal{G}} \widetilde{\mathbb{E}}_G^{\mathsf{FK}}[\boldsymbol{x}^S] \approx \mathop{\mathbb{E}}_{(\boldsymbol{x}^\star,G)\sim\mathcal{P}_k}[(\boldsymbol{x}^\star)^S], \tag{11.18}$$

since in taking the expectation over G on the left-hand side we only introduce a further factor of $\mathbb{P}[S \text{ is a clique in } G] = 2^{-\binom{|S|}{2}}$, which is of constant order for low-degree SOS.

We will to conditions of the form (11.18), which play a crucial role in the *pseudocalibration* construction to come, but for now we just notice first that, if (11.18) holds for all $\boldsymbol{x}^S$ with $|S| \le D$, then by linearity it also holds for all $p(\boldsymbol{x})$ with $\deg(p) \le D$. But moreover, if we want $\widetilde{\mathbb{E}}_G^{\mathsf{FK}}$ to "pretend" to take an expectation over $\mathcal{P}_k$, we might expect the same to hold when p depends on G as well. Let us view G as encoded in its ± 1 adjacency matrix $\boldsymbol{G} \in \{\pm 1\}^{n\times n}_{\mathsf{sym}}$ with $G_{ii} := 1$ by convention. Then, if we have some $p(\boldsymbol{x}, \boldsymbol{G})$ with $\deg_{\boldsymbol{x}}(p) \le D$ and $\deg_{\boldsymbol{G}}(p)$ small as well, then we might expect to have

$$\mathop{\mathbb{E}}_{\boldsymbol{G}\sim\mathcal{G}} \widetilde{\mathbb{E}}_G^{\mathsf{FK}}[p(\boldsymbol{x},\boldsymbol{G})] \approx \mathop{\mathbb{E}}_{(\boldsymbol{x}^\star,\boldsymbol{G})\sim\mathcal{P}_k}[p(\boldsymbol{x}^\star,\boldsymbol{G})]. \tag{11.19}$$

We give further reasoning for why this is desirable in the next section, but let us take it for granted for the moment to identify a "bad" $p(\boldsymbol{x}, \boldsymbol{G})$.

We consider *Kelner's polynomial*

$$p(\boldsymbol{x},\boldsymbol{G}) := \sum_{i=1}^n \left(\sum_{j=1}^n G_{ij}x_j\right)^4 = \|\boldsymbol{G}\boldsymbol{x}\|_4^4 = \sum_{i,j,k,\ell,m=1}^n G_{ij}G_{ik}G_{i\ell}G_{im}x_jx_kx_\ell x_m \tag{11.20}$$

Consider (11.19) evaluated with this polynomial. On the right-hand side, we have by conditioning on $C^\star$ first

$$\mathop{\mathbb{E}}_{(\boldsymbol{x}^\star,\boldsymbol{G})\sim\mathcal{P}_k}[G_{ij}G_{ik}G_{i\ell}G_{im}x_j^\star x_k^\star x_\ell^\star x_m^\star] = \mathop{\mathbb{E}}_{C^\star} \mathbb{1}\{j,k,\ell,m \in C^\star\} \mathop{\mathbb{E}}_{\boldsymbol{G}\sim\mathcal{P}_k}[G_{ij}G_{ik}G_{i\ell}G_{im} \mid C^\star]$$

and, since conditional on $C^\star$ the G_{ij} are independent and are either 1 with probability 1 if $i,j \in C^\star$ or $i = j$, or distributed as $\mathsf{Unif}(\{\pm 1\})$ otherwise, we find

$$\begin{aligned} &= \mathop{\mathbb{E}}_{C^\star} \mathbb{1}\{j,k,\ell,m \in C^\star\}\mathbb{1}\{i \in C^\star\} \\ &= \mathop{\mathbb{E}}_{C^\star} \mathbb{1}\{i,j,k,\ell,m \in C^\star\}, \end{aligned} \tag{11.21}$$

and thus, summing over all $i,j,k,\ell,m \in [n]$, since $|C^\star| = k$ we have

$$\mathop{\mathbb{E}}_{(\boldsymbol{x}^\star,\boldsymbol{G})\sim\mathcal{P}_k}[p(\boldsymbol{x}^\star,\boldsymbol{G})] = k^5. \tag{11.22}$$

On the other hand, on the left-hand side upon expanding we have

$$\begin{aligned} &\mathop{\mathbb{E}}_{G\sim\mathcal{G}} \widetilde{\mathbb{E}}_G^{\mathsf{FK}}[G_{ij}G_{ik}G_{i\ell}G_{im}x_jx_kx_\ell x_m] \\ &\quad\approx \left(\frac{k}{n}\right)^{|\{j,k,\ell,m\}|} \mathop{\mathbb{E}}_{G\sim\mathcal{G}}\Big[G_{ij}G_{ik}G_{i\ell}G_{im}\mathbb{1}\{\{j,k,\ell,m\} \text{ are a clique in } G\}\Big] \end{aligned}$$

where the indicator in the expectation is independent of the remaining terms, so we may factor out $\mathbb{P}[\{j,k,\ell,m\} \text{ are a clique in } G]$, which is only of constant order, so we ignore it in our approximation, finding

$$\approx \left(\frac{k}{n}\right)^{|\{j,k,\ell,m\}|} \mathop{\mathbb{E}}_{G\sim\mathcal{G}}[G_{ij}G_{ik}G_{i\ell}G_{im}]. \tag{11.23}$$

If $i \notin \{j,k,\ell,m\}$, then the remaining expectation is 1 if each index among j,k,ℓ,m occurs an even number of times, and 0 otherwise. If $i \in \{j,k,\ell,m\}$, say $i = j$, then $G_{ij}G_{ik}G_{i\ell}G_{im} = G_{jk}G_{j\ell}G_{jm}$, so again the expectation is 1 if one of k,ℓ,m equals j and the other two of k,ℓ,m are equal, and zero otherwise. Thus in either case the expectation can only be non-zero if each index among j,k,ℓ,m occurs an even number of times.

Putting things together, the left-hand side scales as

$$\mathop{\mathbb{E}}_{G\sim\mathcal{G}} \widetilde{\mathbb{E}}_G^{\mathsf{FK}}[p(\boldsymbol{x},\boldsymbol{G})] \lesssim n\left(n^2\left(\frac{k}{n}\right)^2 + n\left(\frac{k}{n}\right)\right) = O(k^2 n), \tag{11.24}$$

where the outer factor of n corresponds to the summation over i and the inner terms count the contributions from $|\{j,k,\ell,m\}| = 2$ and $|\{j,k,\ell,m\}| = 1$, respectively.

Finally, comparing the left- and right-hand sides, we see that (11.19) is violated once $k^2 n \ll k^5$, or

$$k \ll n^{1/3}. \tag{11.25}$$

That is, we might expect the FK pseudomoments to *fail* to be feasible for some low degree of SOS for clique sizes smaller than $k = n^{1/3} \gg n^{1/2}$, and in particular the FK pseudomoments should *not* prove the tight lower bound we are interested in.

The following shows that, using Kelner's polynomial, we may in fact show that the FK pseudomoments fail to satisfy a concrete inequality of pseudoexpectations admitting a low-degree SOS proof.

Lemma 11.11. *For any $\widetilde{\mathbb{E}}$ of degree at least 6 satisfying the clique constraints on a graph G, $\widetilde{\mathbb{E}}\|\boldsymbol{G}\boldsymbol{x}\|_4^4 \geq \widetilde{\mathbb{E}}(\sum_{i=1}^n x_i)^5$.*

Proof. Note that we have the identity

$$1 - x_i = (1-x_i)^2 - (x_i^2 - x_i), \tag{11.26}$$

and thus for any $s(\boldsymbol{x}) \in \mathsf{SOS}$ with $\deg(s) \leq 4$, we have

$$0 \leq \widetilde{\mathbb{E}}[(1-x_i)s(\boldsymbol{x})] = \widetilde{\mathbb{E}}[s(\boldsymbol{x})] - \widetilde{\mathbb{E}}[x_i s(\boldsymbol{x})], \tag{11.27}$$

or

$$\widetilde{\mathbb{E}}[x_i s(\boldsymbol{x})] \leq \widetilde{\mathbb{E}}[s(\boldsymbol{x})]. \tag{11.28}$$

Now, expanding the pseudoexpectation we are trying to bound and applying this, we have

$$\begin{aligned}\widetilde{\mathbb{E}}\|\boldsymbol{G}\boldsymbol{x}\|_4^4 &= \sum_{i=1}^n \widetilde{\mathbb{E}}\left(\sum_{j=1}^n G_{ij}x_j\right)^4 \\ &\geq \sum_{i=1}^n \widetilde{\mathbb{E}}\left[x_i\left(\sum_{j=1}^n G_{ij}x_j\right)^4\right]\end{aligned}$$

and, expanding fully now,

$$= \sum_{i,j,k,\ell,m=1}^{n} G_{ij}G_{ik}G_{i\ell}G_{im}\widetilde{\mathbb{E}}[x_i x_j x_k x_\ell x_m]$$

where the pseudoexpectation is non-zero only if $\{i,j,k,\ell,m\}$ form a clique in G, in which case $G_{ij} = G_{ik} = G_{i\ell} = G_{im} = 1$, so we have

$$= \sum_{i,j,k,\ell,m=1}^{n} \widetilde{\mathbb{E}}[x_i x_j x_k x_\ell x_m]$$
$$= \widetilde{\mathbb{E}}\left(\sum_{i=1}^{n} x_i\right)^5,$$

completing the proof. □

With some more technicalities, degree 6 can also be lowered to degree 4. With a slightly more careful analysis, one may show the following.

Corollary 11.12. *For $k = k(n) \ll n^{1/3}$, any pseudomoments $\widetilde{\mathbb{E}}_G^{\mathsf{FK}}$ satisfying the scaling (11.16) are with high probability (as $n \to \infty$ and under $G \sim \mathcal{G}$) not feasible for degree 4 SOS.*

These pseudomoments *do* show that $\mathsf{SOS}_4(G) \geq n^{1/3}/\mathsf{polylog}(n)$ with high probability, as was first shown by [DM15]. Moreover, the best bound lower bound on degree D SOS that can be proved with the FK pseudomoments scales as $\mathsf{SOS}_D(G) \geq n^{1/(D/2+1)}/\mathsf{polylog}(n)$ (for D constant as $n \to \infty$), as shown by [HKP15]. On the other hand, it is possible to improve on the FK pseudomoments at least for degree 4—[HKP15] found a somewhat *ad hoc* adjustment of these pseudomoments to deal with Kelner's polynomial and similar obstructions, and showed that their adjusted pseudomoments do give the essentially optimal bound $\mathsf{SOS}_4(G) \geq n^{1/2}/\mathsf{polylog}(n)$. Next, we will discuss the seminal improved pseudomoment construction that allowed this bound to be extended to all constant D.

11.5 Pseudocalibration

Let us revisit the condition (11.19) that we proposed above, which we repeat below: for a pseudoexpectation $\widetilde{\mathbb{E}}_G$ constructed from a graph G to achieve a strong lower bound over $G \sim \mathcal{G}$, we claimed that we might expect to have

$$\mathop{\mathbb{E}}_{G\sim\mathcal{G}} \widetilde{\mathbb{E}}_G[p(\boldsymbol{x},\boldsymbol{G})] \approx \mathop{\mathbb{E}}_{(\boldsymbol{x}^\star,\boldsymbol{G})\sim\mathcal{P}_k}[p(\boldsymbol{x}^\star,\boldsymbol{G})]. \tag{11.29}$$

Why and for what p should we expect such a "calibration" property to hold?

11.5.1 Motivating Argument

Let us step back and remember the broader context: we are interested in whether solving the $\mathsf{SOS}_D(G)$ program can distinguish $G \sim \mathcal{G}$ and $G \sim \mathcal{P}_k$; Conjecture 11.8 tells us that, for $k \ll n^{1/2}$ and D constant, we expect this to be impossible.

Suppose $\widetilde{\mathbb{E}}_G$ is the optimizer of $\mathsf{SOS}_D(G)$. (We should expect our lower bound construction to be close to the optimizer if we are to prove a tight lower bound!) Previously we restricted our attention to using $\mathsf{SOS}_D(G)$ to solve this detection problem in the following *specific* way: we would compute the *value* of the relaxation, i.e., compute $\widetilde{\mathbb{E}}_G[\sum_{i=1}^n x_i]$, and threshold this quantity. For $G \sim \mathcal{P}_k$ it would be at least k, so if for $G \sim \mathcal{G}$ this quantity is, say, at most $(1-\epsilon)k$ with high probability, then this detection procedure would succeed (that was the content of our Proposition 11.9).

The first key idea is that there is no reason to use the specific polynomial $\sum_{i=1}^n x_i$: if we could replace that with some other polynomial $p(\boldsymbol{x}, \boldsymbol{G})$, we could attempt the same thresholding procedure. Thus, if Conjecture 11.8 holds, we should at least expect to have

$$\mathop{\mathbb{E}}_{G \sim \mathcal{G}} \widetilde{\mathbb{E}}_G[p(\boldsymbol{x}, \boldsymbol{G})] \approx \mathop{\mathbb{E}}_{G \sim \mathcal{P}_k} \widetilde{\mathbb{E}}_G[p(\boldsymbol{x}, \boldsymbol{G})]. \tag{11.30}$$

Which polynomials p do we expect this to hold for? For the evaluation of the pseudoexpectation to make sense, we need $\deg_{\boldsymbol{x}}(p) \le D$. But also, for $p(\boldsymbol{x}, \boldsymbol{G})$ to be computable efficiently the degree in the $\boldsymbol{G}$ variables must be bounded, say by some other constant $\deg_{\boldsymbol{G}}(p) \le \Delta$. (If we did not include this constraint, then we could take, e.g., $p(\boldsymbol{x}, \boldsymbol{G}) = \omega(G)$ which, being a function of the Boolean matrix $\boldsymbol{G}$ is some polynomial in $\boldsymbol{G}$, in which case (11.30) would fail even once $k \gg \log n$.)

The second idea is that, when k is close to $n^{1/2}$ (as large as possible for Conjecture 11.8 to still apply), then when $(\boldsymbol{x}^\star, G) \sim \mathcal{P}_k$, we expect to have

$$\widetilde{\mathbb{E}}_G \approx \mathbb{E}_{\delta_{\boldsymbol{x}^\star}}. \tag{11.31}$$

That is, as we approach the threshold of detection becoming possible, we expect the SOS optimizer under the planted distribution $\mathcal{P}_k$ to behave like just an evaluation at the indicator $\boldsymbol{x}^\star$ of the planted clique. (Another way to think about this is that around this threshold we expect SOS to successfully solve the problem of recovering the planted clique.)

If we believe this, then combining (11.30) and (11.31) we find

$$\mathop{\mathbb{E}}_{G \sim \mathcal{G}} \widetilde{\mathbb{E}}_G[p(\boldsymbol{x}, \boldsymbol{G})] \approx \mathop{\mathbb{E}}_{G \sim \mathcal{P}_k} \widetilde{\mathbb{E}}_G[p(\boldsymbol{x}, \boldsymbol{G})] \approx \mathop{\mathbb{E}}_{(\boldsymbol{x}^\star, \boldsymbol{G}) \sim \mathcal{P}_k} p(\boldsymbol{x}^\star, \boldsymbol{G}), \tag{11.32}$$

the same relation we proposed earlier. The requirement that (11.32) hold for all p with $\deg_{\boldsymbol{x}}(p) \le D$ and $\deg_{\boldsymbol{G}}(p) \le \Delta$ is called *pseudocalibration* of the pseudoexpectation $\widetilde{\mathbb{E}}_G$.

11.5.2 Deriving Pseudocalibrated Pseudomoments

We next show how (11.32) constrains $\widetilde{\mathbb{E}}_G$ sufficiently that there is essentially only one pseudocalibrated pseudoexpectation, and pseudocalibration can in fact be used to derive an explicit description of this $\widetilde{\mathbb{E}}_G$.

Consider what happens when we plug into (11.32) a monomial $p(\boldsymbol{x}, \boldsymbol{G}) = \boldsymbol{x}^T \boldsymbol{G}^S$. Note that $\boldsymbol{x} \in \mathbb{R}^n$, so $T \subseteq [n]$, while the degrees of freedom of the indices of $\boldsymbol{G}$ may be identified with $\binom{[n]}{2}$ (corresponding to the possible edges in G), so $S \subseteq \binom{[n]}{2}$. Pseudocalibration then gives us, upon using linearity of $\widetilde{\mathbb{E}}_G$,

$$\mathop{\mathbb{E}}_{\boldsymbol{G} \sim \mathcal{G}} \left[\boldsymbol{G}^S \, \widetilde{\mathbb{E}}_G[\boldsymbol{x}^T] \right] = \mathop{\mathbb{E}}_{(\boldsymbol{x}^\star, \boldsymbol{G}) \sim \mathcal{P}_k} \left[(\boldsymbol{x}^\star)^T \boldsymbol{G}^S \right]. \tag{11.33}$$

Consider the function $f_T(\boldsymbol{G}) := \widetilde{\mathbb{E}}_G[\boldsymbol{x}^T]$ (where we recall that we identify the graph G with the matrix $\boldsymbol{G}$). The left-hand side above is then none other than the *Boolean Fourier transform* of f_T, evaluated at the index S:

$$\widehat{f_T}(S) = \mathop{\mathbb{E}}_{(\boldsymbol{x}^\star, \boldsymbol{G}) \sim \mathcal{P}_k} \left[(\boldsymbol{x}^\star)^T \boldsymbol{G}^S \right]. \tag{11.34}$$

We know that the full Fourier transform, by Fourier inversion, would fully determine f_T, and doing this for all T would fully determine $\widetilde{\mathbb{E}}_G$:

$$\widetilde{\mathbb{E}}_G[\boldsymbol{x}^T] = f_T(\boldsymbol{G}) = \sum_{S \subseteq \binom{[n]}{2}} \widehat{f_T}(S) \boldsymbol{G}^S. \tag{11.35}$$

Unfortunately, we only expect pseudocalibration to ensure the above for sufficiently small S, having $|S| \leq \Delta$. To fully determine $\widetilde{\mathbb{E}}_G$, we must then make some choice of what $\widehat{f_T}(S)$ should be for large S. The choice that we make is, as a last assumption, to set these to zero:

$$\textbf{assume } \widehat{f_T}(S) = 0 \text{ for } |S| > \Delta. \tag{11.36}$$

One way to justify this is to note that, as an evalution of the solution of the degree D SOS relaxation, $\widetilde{\mathbb{E}}_G$ is efficiently computable from $\boldsymbol{G}$. On the other hand, if the polynomial expansion (11.35) has non-zero terms for many large S, then $\widetilde{\mathbb{E}}_G$ would not be efficiently computable *as a polynomial*, in the usual brute force way of evaluating polynomials. Of course, some special polynomials can be computed faster than by term-by-term summation, but this heuristic gives some justification for the assumption (11.36).

Under this assumption, we have enough information to fully determine $\widetilde{\mathbb{E}}_G$, and we arrive at the first version of our pseudocalibrated pseudoexpectation:

$$\widetilde{\mathbb{E}}_G^{(0)}[\boldsymbol{x}^T] := \sum_{\substack{S \subseteq \binom{[n]}{2} \\ |S| \leq \Delta}} \left(\mathop{\mathbb{E}}_{(\boldsymbol{x}^\star, \boldsymbol{G}) \sim \mathcal{P}_k} \left[(\boldsymbol{x}^\star)^T \boldsymbol{G}^S \right] \right) \boldsymbol{G}^S. \tag{11.37}$$

We will see that to satisfy the actual SOS constraints we will need to make some further adjustments to this construction, but this is very close to the pseudoexpectation that is used to prove our lower bound.

11.5.3 Computing Fourier Coefficients

Our description in (11.37) may not seem fully satisfying, since the Fourier coefficients are still given as vague-looking expectations over $\mathcal{P}_k$. But actually, these are not hard to compute

in closed form, and we give the calculation below. The following notion will be useful; the notation encourages us to think of S as a set of "potential edges" in G (or actual edges in the complete graph K_n), and defines the set of vertices that these edges touch:

$$\mathsf{vert}(S) := \bigcup_{\{i,j\} \in S} \{i,j\} = \{i \in [n] : i \in A \text{ for some } A \in S\}. \tag{11.38}$$

Lemma 11.13 (Pseudocalibration Fourier coefficients)**.** *For any $T \subseteq [n]$ and $S \subseteq \binom{[n]}{2}$, we have*

$$\mathop{\mathbb{E}}_{(\boldsymbol{x}^\star, \boldsymbol{G}) \sim \mathcal{P}_k} \left[(\boldsymbol{x}^\star)^T \boldsymbol{G}^S \right] = \mathop{\mathbb{P}}_{C^\star \sim \mathsf{Unif}(\binom{[n]}{k})} [T \cup \mathsf{vert}(S) \subseteq C^\star] \approx \left(\frac{k}{n}\right)^{|T \cup \mathsf{vert}(S)|}. \tag{11.39}$$

Proof. Expanding the definitions and then conditioning on $C^\star$, we may rewrite:

$$\begin{aligned}
\mathop{\mathbb{E}}_{(\boldsymbol{x}^\star, \boldsymbol{G}) \sim \mathcal{P}_k} (\boldsymbol{x}^\star)^T \boldsymbol{G}^S &= \mathop{\mathbb{E}}_{(C^\star, \boldsymbol{G}) \sim \mathcal{P}_k} \mathbb{1}\{T \subseteq C^\star\} \prod_{\{i,j\} \in S} G_{ij} \\
&= \mathop{\mathbb{E}}_{C^\star \sim \mathsf{Unif}(\binom{[n]}{k})} \mathbb{1}\{T \subseteq C^\star\} \mathop{\mathbb{E}}_{G \sim \mathcal{P}_k} \left[\prod_{\{i,j\} \in S} G_{ij} \,\middle|\, C^\star \right]
\end{aligned}$$

Here, the inner expectation is zero if any of the G_{ij} in the product are not forced to equal 1 by the conditioning. Thus, we have

$$\begin{aligned}
&= \mathop{\mathbb{E}}_{C^\star \sim \mathsf{Unif}(\binom{[n]}{k})} \mathbb{1}\{T \subseteq C^\star\} \mathbb{1}\{\mathsf{vert}(S) \subseteq C^\star\} \\
&= \mathop{\mathbb{E}}_{C^\star \sim \mathsf{Unif}(\binom{[n]}{k})} \mathbb{1}\{T \cup \mathsf{vert}(S) \subseteq C^\star\} \\
&= \mathop{\mathbb{P}}_{C^\star \sim \mathsf{Unif}(\binom{[n]}{k})} [T \cup \mathsf{vert}(S) \subseteq C^\star],
\end{aligned}$$

and the remaining approximation for small $|T \cup \mathsf{vert}(S)|$ is straightforward. □

Actually, it will be convenient for us to assume the approximation above holds exactly, so we will make a first adjustment to our construction by plugging it in directly, obtaining more explicit, negligibly different pseudomoments

$$\widetilde{\mathbb{E}}^{(1)}_G [\boldsymbol{x}^T] := \sum_{\substack{S \subseteq \binom{[n]}{2} \\ |S| \le \Delta}} \left(\frac{k}{n}\right)^{|T \cup \mathsf{vert}(S)|} \boldsymbol{G}^S. \tag{11.40}$$

11.6 Adjustments to Satisfy Relaxation Constraints

We will have to further adjust our construction of $\widetilde{\mathbb{E}}_G$ in order to satisfy the SOS constraints. Recall that we wrote $\omega(G)$ as a polynomial optimization problem in (11.4), and based on this we may write the feasibility conditions on $\widetilde{\mathbb{E}}_G$ as follows:

1. $\widetilde{\mathbb{E}}_G$ is linear,

2. $\widetilde{\mathbb{E}}_G[(x_i^2 - x_i)p(\boldsymbol{x})] = 0$ whenever $\deg(p) \le D-2$,

3. $\widetilde{\mathbb{E}}_G[x_i x_j p(\boldsymbol{x})] = 0$ whenever $i \not\sim_G j$ and $\deg(p) \le D-2$,

4. $\widetilde{\mathbb{E}}_G[1] = 1$, and

5. $\widetilde{\mathbb{E}}_G[p(\boldsymbol{x})^2] \ge 0$ whenever $\deg(p) \le D/2$.

Since our general plan is to specify $\widetilde{\mathbb{E}}_G[\boldsymbol{x}^T]$ for sets T and extend first to multisets and then to arbitrary polynomials by linearity, Conditions 1 and 2 will always be satisfied automatically so long as we stick to this class of constructions.

11.6.1 Clique Constraints

Let us next consider Condition 3. By expanding $p(\boldsymbol{x})$ as a sum of monomials and applying Condition 2 as needed, we see that it suffices to show this equivalent condition:

3′. $\widetilde{\mathbb{E}}_G[\boldsymbol{x}^T] = 0$ whenever $T \in \binom{[n]}{\le D}$ is not a clique in G.

We observe that this condition *almost* holds: grouping the terms in (11.40) by the set $R := T \cup \mathsf{vert}(S)$, we have

$$\widetilde{\mathbb{E}}_G^{(1)}[\boldsymbol{x}^T] = \sum_{\substack{R \subseteq [n] \\ T \subseteq R}} \left(\frac{k}{n}\right)^{|R|} \sum_{\substack{S \subseteq \binom{[n]}{2} \\ R \setminus T \subseteq \mathsf{vert}(S) \subseteq R \\ |S| \le \Delta}} \boldsymbol{G}^S$$

and, expanding the definition of $\boldsymbol{G}^S$, we have

$$= \sum_{\substack{R \subseteq [n] \\ T \subseteq R}} \left(\frac{k}{n}\right)^{|R|} \sum_{\substack{S \subseteq \binom{[n]}{2} \\ R \setminus T \subseteq \mathsf{vert}(S) \subseteq R \\ |S| \le \Delta}} (-1)^{\#\{\{i,j\} \in S : i \not\sim_G j\}}. \tag{11.41}$$

Note first that for R large enough there are no terms in the inner sum since it becomes impossible to have both $R \setminus T \subseteq \mathsf{vert}(S)$ and $|S| \le \Delta$, so there are really only finitely many terms in the outer sum. Note also that for R small enough, namely for $\binom{|R|}{2} \le \Delta$, we may discard the size constraint on S, $|S| \le \Delta$. Suppose that this is the case, and that T is not a clique in G. That means that are some $i_0, j_0 \in T$ with $i_0 \not\sim_G j_0$. We may then reorganize the inner sum as

$$\begin{aligned} \sum_{\substack{S \subseteq \binom{[n]}{2} \\ R \setminus T \subseteq \mathsf{vert}(S) \subseteq R}} (-1)^{\#\{\{i,j\} \in S : i \not\sim_G j\}} &= \sum_{\substack{S \subseteq \binom{[n]}{2} \\ R \setminus T \subseteq \mathsf{vert}(S) \subseteq R \\ \{i_0, j_0\} \notin S}} \left((-1)^{\#\{\{i,j\} \in S : i \not\sim_G j\}} + (-1)^{\#\{\{i,j\} \in S \cup \{\{i_0,j_0\}\} : i \not\sim_G j\}} \right) \\ &= 0, \end{aligned} \tag{11.42}$$

since each pair of grouped terms now are opposite signs.

Thus, the only reason that Condition 3′ does not hold for $\widetilde{\mathbb{E}}_G^{(1)}$ is that there are some R for which the inner sum is not empty, but we *cannot* disregard the condition $|S| \le \Delta$; this

Construction	Formula (on x^T)	Source	Issue
$\widetilde{\mathbb{E}}_G^{(0)}$	$\sum_{\substack{S \subseteq \binom{[n]}{2} \\ \lvert S\rvert \le \Delta}} \left(\mathbb{E}_{\mathcal{P}_k}\left[(x^\star)^T G^S\right]\right) G^S$	Pseudocalibration and low-degree truncation	Inconvenient combinatorial coefficients
$\widetilde{\mathbb{E}}_G^{(1)}$	$\sum_{\substack{S \subseteq \binom{[n]}{2} \\ \lvert S\rvert \le \Delta}} \left(\frac{k}{n}\right)^{\lvert T \cup \mathsf{vert}(S)\rvert} G^S$	Approximate evaluation of Fourier coefficients	Does not satisfy clique constraints
$\widetilde{\mathbb{E}}_G^{(2)}$	$\sum_{\substack{S \subseteq \binom{[n]}{2} \\ \lvert T \cup \mathsf{vert}(S)\rvert \le \Delta}} \left(\frac{k}{n}\right)^{\lvert T \cup \mathsf{vert}(S)\rvert} G^S$	Different application of truncation	Does not satisfy normalization $\widetilde{\mathbb{E}}[1] = 1$
$\widetilde{\mathbb{E}}_G^{(3)} = \widetilde{\mathbb{E}}_G$	$\widetilde{\mathbb{E}}_G^{(2)}[x^T] / \widetilde{\mathbb{E}}_G^{(2)}[1]$	Normalization	None!

Table 11.1: The sequence of constructions eventually leading to a valid pseudoexpectation for the planted clique problem.

condition actually imposes a non-trivial truncation of the inner sum. In that case, the above grouping fails because it can be that $\{i_0, j_0\} \notin S$ with $|S| = \Delta$, so that $S \cup \{\{i_0, j_0\}\}$ is not included in the sum.

To deal with this, we adjust our construction again, changing the way that we perform the truncation on the size of S to make sure the above issue never occurs. Namely, we take:

$$\widetilde{\mathbb{E}}_G^{(2)}[x^T] := \sum_{\substack{S \subseteq \binom{[n]}{2} \\ \lvert T \cup \mathsf{vert}(S)\rvert \le \Delta}} \left(\frac{k}{n}\right)^{\lvert T \cup \mathsf{vert}(S)\rvert} G^S. \tag{11.43}$$

Following the above argument again, one may then check the following.

Proposition 11.14. *For any graph G and any T not a clique in G, $\widetilde{\mathbb{E}}_G^{(2)}[x^T] = 0$.*

11.6.2 Normalization Constraint

We are now ready to state the final pseudoexpectation that we will work with to prove our main lower bound:

$$\widetilde{\mathbb{E}}_G[x^T] = \widetilde{\mathbb{E}}_G^{(3)}[x^T] := \frac{\widetilde{\mathbb{E}}_G^{(2)}[x^T]}{\widetilde{\mathbb{E}}_G^{(2)}[1]}. \tag{11.44}$$

This will automatically satisfy Conditions 1 through 4 above, so it will remain to prove positivity.

However, it will be useful for us to know at a quantitative level that this renormalization does not change the values of the pseudoexpectation too much, i.e., that $\widetilde{\mathbb{E}}_G^{(2)}[1] \approx 1$. (At the very least, we would like $\widetilde{\mathbb{E}}_G^{(2)}[1] > 0$ so that positive semidefiniteness of $\widetilde{\mathbb{E}}_G^{(3)}$ is equivalent to that of $\widetilde{\mathbb{E}}_G^{(2)}$.)

Lemma 11.15 (Approximate normalization of $\widetilde{\mathbb{E}}_G^{(2)}$)**.** *Suppose $k \le \sqrt{n}$. The random variable $Y = Y(G) := \widetilde{\mathbb{E}}_G^{(2)}[1]$ satisfies the following properties:*

$$\mathbb{E}[Y] = 1, \tag{11.45}$$

$$\mathsf{Var}[Y] = O_\Delta\left(\left(\frac{k}{\sqrt{n}}\right)^4\right). \tag{11.46}$$

In particular, if $k = n^{1/2-\epsilon}$ and Δ is constant, then $|Y - 1| = O(n^{-\Omega(\epsilon)})$ with high probability.

We note that, remarkably, the $k \sim \sqrt{n}$ threshold appears independently of any of our previous algorithmic reasoning in this calculation! We will return to this phenomenon later when we study low-degree polynomial algorithms and their connection to pseudocalibration.

Proof. From the definition of $\widetilde{\mathbb{E}}_G^{(2)}$ evaluated with $T = \emptyset$, we have

$$Y = \sum_{\substack{S \subseteq \binom{[n]}{2} \\ |\mathsf{vert}(S)| \le \Delta}} \left(\frac{k}{n}\right)^{|\mathsf{vert}(S)|} \boldsymbol{G}^S. \tag{11.47}$$

We note that $\mathbb{E}[\boldsymbol{G}^S] = 0$ for any $S \neq \emptyset$, since $G_{ij} \sim \mathsf{Unif}(\{\pm 1\})$ independently. Thus the only contribution to $\mathbb{E}[Y]$ is from $S = \emptyset$, and we indeed find $\mathbb{E}[Y] = 1$.

For computing the second moment $\mathbb{E}[Y^2]$, we note that $\mathbb{E}[\boldsymbol{G}^S \boldsymbol{G}^{S'}] = \mathbb{E}[\boldsymbol{G}^{S \triangle S'}]$ is zero unless $S = S'$, in which case it is 1 (indeed, this statement is just the orthonormality of the Boolean Fourier basis $\boldsymbol{G}^S$). Thus we have

$$\mathbb{E}[Y^2] = \sum_{\substack{S \subseteq \binom{[n]}{2} \\ |\mathsf{vert}(S)| \le \Delta}} \left(\frac{k}{n}\right)^{2|\mathsf{vert}(S)|} = 1 + \underbrace{\sum_{\substack{S \subseteq \binom{[n]}{2} \\ 2 \le |\mathsf{vert}(S)| \le \Delta}} \left(\frac{k}{n}\right)^{2|\mathsf{vert}(S)|}}_{= \mathsf{Var}[Y]}, \tag{11.48}$$

where we isolate the contribution of $S = \emptyset$. Grouping according to $t = |\mathsf{vert}(S)|$ and making some elementary estimates, we have

$$\begin{aligned}
\mathsf{Var}[Y] &\le \sum_{t=2}^{\Delta} \left(\frac{k}{n}\right)^{2t} \underbrace{\binom{n}{t}}_{\le n^t} \underbrace{\#\{\text{distinct graphs on } t \text{ labelled vertices}\}}_{\le 2^{t^2}} \\
&\le 2^{\Delta^2} \sum_{t=2}^{\Delta} \left(\frac{k}{\sqrt{n}}\right)^{2t} \\
&= O_\Delta\left(\left(\frac{k}{\sqrt{n}}\right)^4\right),
\end{aligned}$$

where we use that, since $k/\sqrt{n} \le 1$ by assumption, up to terms depending on Δ, the value of the sum is proportional to the first term. □

The last thing that we will need for our lower bound is to verify that the objective function of the SOS relaxation, $\widetilde{\mathbb{E}}[\sum_{i=1}^n x_i]$, is still roughly k even after all of our adjustments. This result is quite similar to Lemma 11.15, so we only sketch the proof to highlight the main differences.

Lemma 11.16. *Suppose $k \le \sqrt{n}$. The random variable $Z = Z(G) := \widetilde{\mathbb{E}}_G^{(2)}[\sum_{i=1}^n x_i]$ satisfies the following properties:*

$$\mathbb{E}[Z] = k, \tag{11.49}$$

$$\mathsf{Var}[Y] = O_\Delta\left(k^2 \left(\frac{k}{\sqrt{n}}\right)^4\right). \tag{11.50}$$

In particular, if $k = n^{1/2-\epsilon}$ and Δ is constant, then $Z \ge n^{1/2-\epsilon}/2$ with high probability.

Proof Sketch. Expanding as in Lemma 11.15, we have

$$Z = \sum_{i=1}^{n} \sum_{\substack{S \subseteq \binom{[n]}{2} \\ |\mathsf{vert}(S) \cup \{i\}| \le \Delta}} \left(\frac{k}{n}\right)^{|\mathsf{vert}(S) \cup \{i\}|} \boldsymbol{G}^S. \tag{11.51}$$

In each inner sum, the only contribution to $\mathbb{E}[Z]$ is again from $S = \emptyset$, for which the contribution is k/n. Thus we have $\mathbb{E}[Z] = (k/n) \cdot n = k$.

Switching the order of summation and grouping according to the coefficient of $\boldsymbol{G}^S$, neglecting some minor subtleties with the truncation for large S, we essentially have

$$\begin{aligned} Z &\approx \sum_{\substack{S \subseteq \binom{[n]}{2} \\ |\mathsf{vert}(S)| \le \Delta}} \left(|S| \left(\frac{k}{n}\right)^{|\mathsf{vert}(S)|} + (n - |S|) \left(\frac{k}{n}\right)^{|\mathsf{vert}(S)|+1} \right) \boldsymbol{G}^S \\ &= \sum_{\substack{S \subseteq \binom{[n]}{2} \\ |\mathsf{vert}(S)| \le \Delta}} \left(|S| + \frac{n - |S|}{n} \cdot k \right) \left(\frac{k}{n}\right)^{|\mathsf{vert}(S)|} \boldsymbol{G}^S \end{aligned}$$

and, since we have $|S| \ll k \ll \sqrt{n}$, we further have

$$\begin{aligned} &\approx k \sum_{\substack{S \subseteq \binom{[n]}{2} \\ |\mathsf{vert}(S)| \le \Delta}} \left(\frac{k}{n}\right)^{|\mathsf{vert}(S)|} \boldsymbol{G}^S \\ &= kY, \end{aligned}$$

where Y is as in Lemma 11.15. The rest of the proof then goes through as before, though to be fully precise we must account for the approximations above carefully. □

11.6.3 Formulating Main Lower Bound

Finally, with the above results established, we may formulate the remaining positivity result that we will need to prove, and use this to formulate and prove our main theorem (which, assuming that result, will be nothing but a combination of the various preliminary results). We do this here before moving on to the complexities of proving positivity.

Lemma 11.17 (Positivity of $\widetilde{\mathbb{E}}_G^{(2)}$)**.** *Suppose D is constant not depending on n, $k = n^{1/2-\epsilon}$ for a fixed $\epsilon > 0$ also not depending on n, and $\Delta = C \cdot D/\epsilon$ for a universal constant $C > 0$. Then, with high probability, for all p with $\deg(p) \le D/2$, $\widetilde{\mathbb{E}}_G^{(2)}[p(\boldsymbol{x})^2] \ge 0$.*

Theorem 11.18 (Planted clique lower bound; special case of [BHK+19])**.** *Suppose D is constant not depending on n, $k = n^{1/2-\epsilon}$ for a fixed $\epsilon > 0$ also not depending on n, and $\Delta = C \cdot D/\epsilon$ for a universal constant $C > 0$. Let $\widetilde{\mathbb{E}}_G = \widetilde{\mathbb{E}}_G^{(3)}$ be as in (11.44) or Table 11.6.2 (this pseudoexpectation is a function of G, k, and Δ). Then, with high probability, $\widetilde{\mathbb{E}}_G$ is feasible for the degree D SOS relaxation of $\omega(G)$, and has objective value $\widetilde{\mathbb{E}}_G[\sum_{i=1}^n x_i] \ge n^{1/2-\epsilon}/3$.*

We note that the original reference [BHK+19] also allowed D to grow modestly with n, which is an interesting extension of this lower bound, but for the sake of simplicity in our exposition we do not allow this.

Proof. $\widetilde{\mathbb{E}}_G$ satisfies linearity, respects the Boolean constraints $x_i^2 - x_i = 0$, and has $\widetilde{\mathbb{E}}_G[1] = 1$ by construction. It respects the clique constraints $x_i x_j = 0$ when $i \not\sim_G j$ by Proposition 11.14.

We have $\widetilde{\mathbb{E}}_G = \widetilde{\mathbb{E}}_G^{(2)} / \widetilde{\mathbb{E}}_G^{(2)}[1]$. By Lemma 11.17, $\widetilde{\mathbb{E}}_G^{(2)}$ satisfies the positivity constraint with high probability, and by Lemma 11.15 we have $\widetilde{\mathbb{E}}_G^{(2)}[1] > 0$ with high probability. Thus, also with high probability, $\widetilde{\mathbb{E}}_G$ satisfies the positivity constraint.

Finally, by Lemma 11.16 we have $\widetilde{\mathbb{E}}_G^{(2)}[\sum_{i=1}^n x_i] \geq n^{1/2-\epsilon}/2$ with high probability, and by Lemma 11.15 we have $\widetilde{\mathbb{E}}_G^{(2)}[1] \leq \frac{3}{2}$ with high probability. Thus, also with high probability, $\widetilde{\mathbb{E}}_G[\sum_{i=1}^n x_i] = \widetilde{\mathbb{E}}_G^{(2)}[\sum_{i=1}^n x_i] / \widetilde{\mathbb{E}}_G^{(2)}[1] \geq n^{1/2-\epsilon}/3$. □

11.7 Proof of Positivity

11.7.1 Graphical Matrices

Notes

Exercises

Exercise 11.1 (Lovász ϑ function)**.** *Let G be a graph on vertex set $[n]$. An* orthonormal representation *of G is a set of unit vectors $v_1, \ldots, v_n \in \mathbb{S}^{d-1}$ so that $\langle v_i, v_j \rangle = 0$ whenever $i \not\sim_G j$. The* Lovász ϑ function *of G is the quantity*

$$\vartheta(G) := \min_{\substack{c \in \mathbb{S}^{d-1} \\ (v_1,\ldots,v_n) \text{ an orthonormal} \\ \text{representation of } G}} \max_{i \in [n]} \frac{1}{\langle c, v_i \rangle^2}. \tag{11.52}$$

This is a bound on the size of the maximum independent set in G, and geometrically may be viewed as finding the narrowest cone (centered on c) on which an orthonormal representation of G may lie. Let $\overline{G}$ be the graph complement of G. Show that

$$\vartheta(\overline{G}) = \mathsf{SOS}_2(G). \tag{11.53}$$

Part IV

Miscellaneous Background

A | Linear Algebra

A.1 Symmetric Matrices

Definition A.1 (Frobenius inner product)**.** $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle := \sum_{i,j} X_{ij} Y_{ij} = \mathsf{Tr}(\boldsymbol{X}^\top \boldsymbol{Y})$.

Definition A.2 (Eigenvalues)**.** *For $\boldsymbol{X} \in \mathbb{R}^{n\times n}_{\mathsf{sym}}$, we write $\lambda_1(\boldsymbol{X}) \geq \cdots \geq \lambda_n(\boldsymbol{X})$ for the ordered eigenvalues. We also sometimes write $\lambda_{\mathsf{max}}(\boldsymbol{X}) = \lambda_1(\boldsymbol{X})$ and $\lambda_{\mathsf{min}}(\boldsymbol{X}) = \lambda_n(\boldsymbol{X})$.*

A.2 Positive Semidefinite Matrices

Definition A.3. *Let $\boldsymbol{X} \in \mathbb{R}^{n\times n}_{\mathsf{sym}}$. We say that $\boldsymbol{X}$ is* positive definite *if $\boldsymbol{v}^\top \boldsymbol{X} \boldsymbol{v} > 0$ for all $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$, written $\boldsymbol{X} \succ \boldsymbol{0}$, and* positive semidefinite *if $\boldsymbol{v}^\top \boldsymbol{X} \boldsymbol{v} \geq 0$ for all $\boldsymbol{v} \in \mathbb{R}^n$, written $\boldsymbol{X} \succeq \boldsymbol{0}$. For another $\boldsymbol{Y} \in \mathbb{R}^{n\times n}_{\mathsf{sym}}$ we write $\boldsymbol{X} \succ \boldsymbol{Y}$ if $\boldsymbol{X} - \boldsymbol{Y} \succ \boldsymbol{0}$ and $\boldsymbol{X} \succeq \boldsymbol{Y}$ if $\boldsymbol{X} - \boldsymbol{Y} \succeq \boldsymbol{0}$.*

Proposition A.4. *Let $\boldsymbol{X} \in \mathbb{R}^{n\times n}_{\mathsf{sym}}$. Then, the following are equivalent:*

1. *$\boldsymbol{X} \succeq \boldsymbol{0}$, i.e., $\boldsymbol{v}^\top \boldsymbol{X} \boldsymbol{v} \geq 0$ for all $\boldsymbol{v} \in \mathbb{R}^n$.*
2. *$\lambda_i(\boldsymbol{X}) \geq 0$ for all $i \in [n]$.*
3. *There exist $\boldsymbol{v}_i \in \mathbb{R}^N$ for some $N \geq 1$ with $X_{ij} = \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$.*
4. *There exist $\boldsymbol{v}_i \in \mathbb{R}^n$ with $X_{ij} = \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$.*
5. *There exist $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$ for some $N \geq 1$ with $\boldsymbol{X} = \sum_{i=1}^N \boldsymbol{v}_i \boldsymbol{v}_i^\top$.*
6. *There exist $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ with $\boldsymbol{X} = \sum_{i=1}^n \boldsymbol{v}_i \boldsymbol{v}_i^\top$.*
7. *There exists $\boldsymbol{A} \in \mathbb{R}^{n\times N}$ for some $N \geq 1$ with $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{A}^\top$.*
8. *There exists $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{A}^\top$.*
9. *There exists $\boldsymbol{A} \in \mathbb{R}^{n\times n}_{\mathsf{sym}}$ with $\boldsymbol{X} = \boldsymbol{A}^2$.*

B | Convex Optimization

B.1 Lagrangian Duality

To come.

Bibliography

[AGH+14] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of machine learning research*, 15:2773–2832, 2014.

[AGJ14] Animashree Anandkumar, Rong Ge, and Majid Janzamin. Guaranteed non-orthogonal tensor decomposition via alternating rank-1 updates. *arXiv preprint arXiv:1402.5180*, 2014.

[AGJ15] Animashree Anandkumar, Rong Ge, and Majid Janzamin. Learning overcomplete latent variable models through tensor methods. In *Conference on Learning Theory*, pages 36–112. PMLR, 2015.

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Cambridge University Press, 2010.

[AH19] Amir Ali Ahmadi and Georgina Hall. On the construction of converging hierarchies for polynomial optimization based on certificates of global positivity. *Mathematics of Operations Research*, 44(4):1192–1207, 2019.

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998.

[ALPTJ11] Radosław Adamczak, Alexander E Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann. Sharp bounds on the rate of convergence of the empirical covariance matrix. *Comptes Rendus Mathematique*, 349(3-4):195–200, 2011.

[AN04] Noga Alon and Assaf Naor. Approximating the cut-norm via Grothendieck's inequality. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80, 2004.

[Art27] Emil Artin. Über die zerlegung definiter funktionen in quadrate. In *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, volume 5, pages 100–115, 1927.

[Ban15] Afonso S Bandeira. Ten lectures and forty-two open problems in the mathematics of data science, 2015.

[BBH+12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *44th Annual ACM Symposium on Theory of Computing (STOC 2012)*, pages 307–326. ACM, 2012.

[BCMV14] Aditya Bhaskara, Moses Charikar, Ankur Moitra, and Aravindan Vijayaraghavan. Smoothed analysis of tensor decompositions. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 594–603, 2014.

[BCN89] Andries E Brouwer, Arjeh M Cohen, and Arnold Neumaier. Association schemes. In *Distance-Regular Graphs*, pages 43–78. Springer, 1989.

[BCO14] Cristiano Bocci, Luca Chiantini, and Giorgio Ottaviani. Refined methods for the identifiability of tensors. *Annali di Matematica Pura ed Applicata (1923-)*, 193(6):1691–1702, 2014.

[BDER16] Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

[Ben17] Olivier Benoist. Writing positive polynomials as sums of (few) squares. *EMS Newsletter*, (105):8–13, 2017.

[BGJR88] Francisco Barahona, Martin Grötschel, Michael Jünger, and Gerhard Reinelt. An application of combinatorial optimization to statistical physics and circuit layout design. *Operations Research*, 36(3):493–513, 1988.

[BGP16] Grigoriy Blekherman, João Gouveia, and James Pfeiffer. Sums of squares on the hypercube. *Mathematische Zeitschrift*, 284(1-2):41–54, 2016.

[BHK+19] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

[BIK+96] Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, and Pavel Pudlák. Lower bounds on Hilbert's Nullstellensatz and propositional proofs. *Proceedings of the London Mathematical Society*, 3(1):1–26, 1996.

[BKM19] Jess Banks, Robert Kleinberg, and Cristopher Moore. The Lovász theta function for random regular graphs and community detection in the hard regime. *SIAM Journal on Computing*, 48(3):1098–1119, 2019.

[BKS14] Boaz Barak, Jonathan A Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 31–40. ACM, 2014.

[BKS15] Boaz Barak, Jonathan A Kelner, and David Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 143–151, 2015.

[BM86] Francisco Barahona and Ali Ridha Mahjoub. On the cut polytope. *Mathematical Programming*, 36(2):157–173, 1986.

[BPR06] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in Real Algebraic Geometry*. 2006.

[Bro87] W Dale Brownawell. Bounds for the degrees in the Nullstellensatz. *Annals of Mathematics*, 126(3):577–591, 1987.

[BS14] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014.

[BS16] Boaz Barak and David Steurer. Proofs, beliefs, and algorithms through the lens of sum-of-squares. http://www.sumofsquares.org/public/index.html, 2016.

[Buc06] Bruno Buchberger. Bruno buchberger's phd thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal. *Journal of symbolic computation*, 41(3-4):475–511, 2006.

[Cat12] Olivier Catoni. Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 1148–1185, 2012.

[CEP71] John WS Cassels, William J Ellison, and Albrecht Pfister. On sums of squares and on elliptic curves over function fields. *Journal of Number Theory*, 3(2):125–149, 1971.

[CLOS94] David Cox, John Little, Donal O'Shea, and Moss Sweedler. Ideals, varieties, and algorithms. *American Mathematical Monthly*, 101(6):582–586, 1994.

[Col75] George E Collins. Quantifier elimination for real closed fields by cylindrical algebraic decompostion. In *Automata theory and formal languages*, pages 134–183. Springer, 1975.

[Con] Keith Conrad. Pfister's theorem on sums of squares. https://kconrad.math.uconn.edu/blurbs/linmultialg/pfister.pdf.

[CR79] Stephen A Cook and Robert A Reckhow. The relative efficiency of propositional proof systems. *The Journal of Symbolic Logic*, 44(1):36–50, 1979.

[CST11] Kevin P Costello, Asaf Shapira, and Prasad Tetali. Randomized greedy: new variants of some classic approximation algorithms. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 647–655. SIAM, 2011.

[CW04] Moses Charikar and Anthony Wirth. Maximizing quadratic programs: extending Grothendieck's inequality. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 54–60. IEEE, 2004.

[CX16] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research*, 17(1):882–938, 2016.

[DdL+22] Jingqiu Ding, Tommaso d'Orsi, Chih-Hung Liu, Stefan Tiegel, and David Steurer. Fast algorithm for overcomplete order-3 tensor decomposition. *arXiv preprint arXiv:2202.06442*, 2022.

[Del73] Philippe Delsarte. An algebraic approach to the association schemes of coding theory. *Philips Res. Rep. Suppl.*, 10:vi+–97, 1973.

[DFHT20] Irit Dinur, Yuval Filmus, Prahladh Harsha, and Madhur Tulsiani. Explicit SoS lower bounds from high-dimensional expanders. *arXiv preprint arXiv:2009.05218*, 2020.

[Dia88] Persi Diaconis. Group representations in probability and statistics. *Lecture Notes Monograph Series*, 11, 1988.

[DK70] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[DKP20] Ilias Diakonikolas, Daniel M Kane, and Ankit Pensia. Outlier robust mean estimation with subgaussian rates via stability. *Advances in Neural Information Processing Systems*, 33:1830–1840, 2020.

[DL09] Michel Marie Deza and Monique Laurent. *Geometry of cuts and metrics*. Springer, 2009.

[DL22] Jules Depersin and Guillaume Lecué. Robust sub-gaussian estimation of a mean vector in nearly linear time. *The Annals of Statistics*, 50(1):511–536, 2022.

[DLCC07] Lieven De Lathauwer, Josphine Castaing, and Jean-Franois Cardoso. Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Transactions on Signal Processing*, 55(6):2965–2973, 2007.

[DLV04] Nikhil R Devanur, Richard J Lipton, and Nisheeth K Vishnoi. On the complexity of Hilbert's 17th problem. In *International Conference on Foundations of Software Technology and Theoretical Computer Science*, pages 237–249. Springer, 2004.

[dlVKM07] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 53–61, 2007.

[DM15] Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *28th Annual Conference on Learning Theory (COLT 2015)*, pages 523–562, 2015.

[DW12] Andrew C Doherty and Stephanie Wehner. Convergence of SDP hierarchies for polynomial optimization on the hypersphere. *arXiv preprint arXiv:1210.5048*, 2012.

[Erd67] Paul Erdős. On bipartite subgraphs of a graph. *Matematika Lapok*, 18:283–288, 1967.

[FF20] Kun Fang and Hamza Fawzi. The sum-of-squares hierarchy on the sphere and applications in quantum information theory. *Mathematical Programming*, pages 1–30, 2020.

[FH04] William Fulton and Joe Harris. *Representation theory: a first course*. Springer Science & Business Media, 2004.

[FH14] Péter Frenkel and Péter Horváth. Minkowski's inequality and sums of squares. *Open Mathematics*, 12(3):510–516, 2014.

[FK00] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures & Algorithms*, 16(2):195–208, 2000.

[FK03] Uriel Feige and Robert Krauthgamer. The probable value of the Lovász-Schrijver relaxations for maximum independent set. *SIAM Journal on Computing*, 32(2):345–370, 2003.

[FKP19] Noah Fleming, Pravesh Kothari, and Toniann Pitassi. *Semialgebraic Proofs and Efficient Algorithm Design*. Now the Essence of Knowledge, 2019.

[FL06] Uriel Feige and Michael Langberg. The RPR2 rounding technique for semidefinite programs. *Journal of Algorithms*, 60(1):1–23, 2006.

[FS02] Uriel Feige and Gideon Schechtman. On the optimality of the random hyperplane rounding technique for max cut. *Random Structures & Algorithms*, 20(3):403–440, 2002.

[FSP16] Hamza Fawzi, James Saunderson, and Pablo A Parrilo. Sparse sums of squares on finite abelian groups and improved semidefinite lifts. *Mathematical Programming*, 160(1-2):149–191, 2016.

[Ful97] William Fulton. *Young tableaux: with applications to representation theory and geometry*. Cambridge University Press, 1997.

[Gem80] Stuart Geman. A limit theorem for the norm of random matrices. *The Annals of Probability*, pages 252–261, 1980.

[GKD18] Fred Glover, Gary Kochenberger, and Yu Du. A tutorial on formulating and using QUBO models. *arXiv preprint arXiv:1811.11538*, 2018.

[GM15] Rong Ge and Tengyu Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. *arXiv preprint arXiv:1504.05287*, 2015.

[GM17] Rong Ge and Tengyu Ma. On the optimization landscape of tensor decompositions. In *Advances in Neural Information Processing Systems*, pages 3653–3663, 2017.

[GR99] Michel X Goemans and Franz Rendl. Semidefinite programs and association schemes. *Computing*, 63(4):331–340, 1999.

[Gri01a] Dima Grigoriev. Complexity of Positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001.

[Gri01b] Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613–622, 2001.

[Gro56] Alexandre Grothendieck. *Résumé de la théorie métrique des produits tensoriels topologiques*. Soc. de Matemática de São Paulo, 1956.

[GS06] Christopher D Godsil and Sung Y Song. Association schemes. In *Handbook of Combinatorial Designs*, pages 351–355. Chapman and Hall/CRC, 2006.

[GV01] Dima Grigoriev and Nicolai Vorobjov. Complexity of Null- and Positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1-3):153–160, 2001.

[GW95] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

[Hab39] Walter Habicht. Über die zerlegung strikte definiter formen in quadrate. *Commentarii Mathematici Helvetici*, 12(1):317–322, 1939.

[Har70] Richard A Harshman. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis. 1970.

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001.

[Her98] Grete Hermann. The question of finitely many steps in polynomial ideal theory. *ACM SIGSAM Bulletin*, 32(3):8–30, 1998.

[Hil88] David Hilbert. Über die darstellung definiter formen als summe von formenquadraten. *Mathematische Annalen*, 32(3):342–350, 1888.

[Hil90] David Hilbert. Ueber die theorie der algebraischen formen. *Mathematische annalen*, 36(4):473–534, 1890.

[Hil93a] David Hilbert. Über die vollen invariantensysteme. *Mathematische Annalen*, 42(3):313–373, 1893.

[Hil93b] David Hilbert. Über ternäre definite formen. *Acta Mathematica*, 17(1):169, 1893.

[HKM21] Samuel B Hopkins, Gautam Kamath, and Mahbod Majid. Efficient mean estimation with pure differential privacy via a sum-of-squares exponential mechanism. *arXiv preprint arXiv:2111.12981*, 2021.

[HKP15] Samuel B Hopkins, Pravesh K Kothari, and Aaron Potechin. Sos and planted clique: Tight analysis of mpw moments at all degrees and an optimal lower bound at degree four. *arXiv preprint arXiv:1507.05230*, 2015.

[HL96] Thomas Hofmeister and Hanno Lefmann. A combinatorial design approach to MAXCUT. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 439–452. Springer, 1996.

[HL03] Didier Henrion and Jean-Bernard Lasserre. GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi. *ACM Transactions on Mathematical Software (TOMS)*, 29(2):165–194, 2003.

[HL05] Didier Henrion and Jean-Bernard Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. In *Positive polynomials in control*, pages 293–310. Springer, 2005.

[HLZ20] Sam Hopkins, Jerry Li, and Fred Zhang. Robust and heavy-tailed mean estimation made simple, via regret minimization. *Advances in Neural Information Processing Systems*, 33:11902–11912, 2020.

[Hop18a] Sam Hopkins. Clustering and sum of squares proofs: Six blog posts on unsupervised learning. 2018.

[Hop18b] Samuel Hopkins. *Statistical inference and the sum of squares method*. PhD thesis, Cornell University, 2018.

[Hop18c] Samuel B Hopkins. Mean estimation with sub-gaussian rates in polynomial time. *arXiv preprint arXiv:1809.07425*, 2018.

[HSSS16] Samuel B Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *48th Annual ACM Symposium on Theory of Computing (STOC 2016)*, pages 178–191, 2016.

[HST19] Samuel B Hopkins, Tselil Schramm, and Luca Trevisan. Subexponential LPs approximate max-cut. *arXiv preprint arXiv:1911.10304*, 2019.

[Hur98] Adolf Hurwitz. Über die komposition der quadratischen formen von beliebig vielen variabeln. 1898.

[HV91] David J. Haglin and Shankar M. Venkatesan. Approximation and intractability results for the maximum cut problem and its variants. *IEEE Transactions on Computers*, 40(01):110–113, 1991.

[HW79] Godfrey Harold Hardy and Edward Wright. *An introduction to the theory of numbers*. Oxford university press, 1979.

[Jer92] Mark Jerrum. Large cliques elude the Metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992.

[JH16] Cédric Josz and Didier Henrion. Strong duality in Lasserre's hierarchy for polynomial optimization. *Optimization Letters*, 10(1):3–10, 2016.

[Kar72] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

[Kar99] Howard Karloff. How good is the Goemans-Williamson MAX CUT algorithm? *SIAM Journal on Computing*, 29(1):336–350, 1999.

[KHG+14] Gary Kochenberger, Jin-Kao Hao, Fred Glover, Mark Lewis, Zhipeng Lü, Haibo Wang, and Yang Wang. The unconstrained binary quadratic programming problem: a survey. *Journal of combinatorial optimization*, 28(1):58–81, 2014.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, 2002.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[KLM16] Adam Kurpisz, Samuli Leppänen, and Monaldo Mastrolilli. Sum-of-squares hierarchy lower bounds for symmetric formulations. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 362–374. Springer, 2016.

[KLYZ12] Erich L Kaltofen, Bin Li, Zhengfeng Yang, and Lihong Zhi. Exact certification in global polynomial optimization via sums-of-squares of rational functions with rational coefficients. *Journal of Symbolic Computation*, 47(1):1–15, 2012.

[KM22] Dmitriy Kunisky and Cristopher Moore. The spectrum of the Grigoriev-Laurent pseudomoments. *arXiv preprint arXiv:2203.05693*, 2022.

[KMV21] Pravesh K Kothari, Pasin Manurangsi, and Ameya Velingker. Private robust estimation by stabilizing convex relaxations. *arXiv preprint arXiv:2112.03548*, 2021.

[KO06] Subhash Khot and Ryan O'Donnell. SDP gaps and UGC-hardness for MAXCUT-GAIN. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 217–226. IEEE, 2006.

[Kol88] János Kollár. Sharp effective Nullstellensatz. *Journal of the American Mathematical Society*, pages 963–975, 1988.

[KOTZ14] Manuel Kauers, Ryan O'Donnell, Li-Yang Tan, and Yuan Zhou. Hypercontractive inequalities via SOS, and the Frankl-Rödl graph. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1644–1658. SIAM, 2014.

[KR08] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2- ε. *Journal of Computer and System Sciences*, 74(3):335–349, 2008.

[Kri64] Jean-Louis Krivine. Anneaux préordonnés. *Journal d'Analyse Mathématique*, 12(1):307–326, 1964.

[Kuč95] Luděk Kučera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2-3):193–212, 1995.

[KV05] Subhash Khot and Nisheeth K Vishnoi. On the unique games conjecture. In *46th Annual Symposium on Foundations of Computer Science (FOCS 2005)*, 2005.

[KV15] Subhash A Khot and Nisheeth K Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative-type metrics into ℓ_1. *Journal of the ACM (JACM)*, 62(1):1–39, 2015.

[Las01] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[Lau03] Monique Laurent. Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. *Mathematics of Operations Research*, 28(4):871–883, 2003.

[Lau09] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009.

[LL78] Anneli Lax and Peter D Lax. On sums of squares. *Linear algebra and its applications*, 20(1):71–75, 1978.

[LM19] Gábor Lugosi and Shahar Mendelson. Sub-gaussian estimators of the mean of a random vector. *The annals of statistics*, 47(2):783–794, 2019.

[Lov79] László Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information theory*, 25(1):1–7, 1979.

[LP68] Joram Lindenstrauss and Aleksander Pełczyński. Absolutely summing operators in l_p-spaces and their applications. *Studia Mathematica*, 29(3):275–326, 1968.

[LPR14] Henri Lombardi, Daniel Perrucci, and Marie-Françoise Roy. An elementary recursive bound for effective Positivstellensatz and Hilbert 17th problem. *arXiv preprint arXiv:1404.2338*, 2014.

[LRA93] Sue E Leurgans, Robert T Ross, and Rebecca B Abel. A decomposition for three-way arrays. *SIAM Journal on Matrix Analysis and Applications*, 14(4):1064–1083, 1993.

[Mag15] Victor Magron. Error bounds for polynomial optimization over the hypercube using Putinar type representations. *Optimization Letters*, 9(5):887–895, 2015.

[Mar08] Murray Marshall. *Positive polynomials and sums of squares*. Number 146. American Mathematical Soc., 2008.

[Mik20] Dan Mikulincer. A CLT in Stein's distance for generalized Wishart matrices and higher order tensors. *arXiv preprint arXiv:2002.10846*, 2020.

[Min22] Stanislav Minsker. U-statistics of growing order and sub-gaussian mean estimators with sharp constants. *arXiv preprint arXiv:2202.11842*, 2022.

[MM82] Ernst W Mayr and Albert R Meyer. The complexity of the word problems for commutative semigroups and polynomial ideals. *Advances in Mathematics*, 46(3):305–329, 1982.

[Moi20] Ankur Moitra. Sum of squares in theoretical computer science. In *Sum of Squares: Theory and Applications*, volume 77 of *Proceedings of Symposia in Applied Mathematics*. American Mathematical Society, 2020.

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 21–30. IEEE, 2005.

[Mot67] Theodore Samuel Motzkin. The arithmetic-geometric inequality. *Inequalities (Proc. Sympos. Wright-Patterson Air Force Base)*, pages 205–224, 1967.

[MR95] Sanjeev Mahajan and Hariharan Ramesh. Derandomizing semidefinite programming based approximation algorithms. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 162–169. IEEE, 1995.

[MS08] Claire Mathieu and Warren Schudy. Yet another algorithm for dense max cut: go greedy. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 176–182. Citeseer, 2008.

[MSS16] Tengyu Ma, Jonathan Shi, and David Steurer. Polynomial-time tensor decompositions with sum-of-squares. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 438–446. IEEE, 2016.

[MW13] Raghu Meka and Avi Wigderson. Association schemes, non-commutative polynomial concentration, and sum-of-squares lower bounds for planted clique. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 20, page 10, 2013.

[Nag69a] Aleksandr V Nagaev. Integral limit theorems taking large deviations into account when cramérâĂŹs condition does not hold. i. *Theory of Probability & Its Applications*, 14(1):51–64, 1969.

[Nag69b] Aleksandr V Nagaev. Integral limit theorems taking large deviations into account when cramérâĂŹs condition does not hold. ii. *Theory of Probability & Its Applications*, 14(2):193–208, 1969.

[Nes98] Yurii Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization Methods and Software*, 9(1-3):141–160, 1998.

[Nes00] Yurii Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, pages 405–440. Springer, 2000.

[Nie14] Jiawang Nie. Optimality conditions and finite convergence of Lasserre's hierarchy. *Mathematical programming*, 146(1):97–121, 2014.

[NZ21] Ivan Nourdin and Guangqu Zheng. Asymptotic behavior of large Gaussian correlated Wishart matrices. *Journal of Theoretical Probability*, pages 1–30, 2021.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[OS18] Ryan O'Donnell and Tselil Schramm. Sherali-Adams strikes back. *arXiv preprint arXiv:1812.09967*, 2018.

[OZ13] Ryan O'Donnell and Yuan Zhou. Approximability and proof complexity. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1537–1556. Society for Industrial and Applied Mathematics, 2013.

[Par00] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

[Par03] Pablo A Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, 2003.

[PD13] Alexander Prestel and Charles Delzell. *Positive polynomials: from Hilbert's 17th problem to real algebra*. Springer Science & Business Media, 2013.

[Pfi67] Albrecht Pfister. Zur darstellung definiter funktionen als summe von quadraten. *Inventiones mathematicae*, 4(4):229–237, 1967.

[Pfi95] Albrecht Pfister. *Quadratic forms with applications to algebraic geometry and topology*, volume 217. Cambridge University Press, 1995.

[Pól28] George Pólya. Über positive darstellung von polynomen. *Vierteljschr. Naturforsch. Ges. Zürich*, 73(141-145):2, 1928.

[Pot17] Aaron Potechin. Sum of squares lower bounds from symmetry and a good story. *arXiv preprint arXiv:1711.11469*, 2017.

[Pow11a] Victoria Powers. Positive polynomials and sums of squares: Theory and practice. *Real Algebraic Geometry*, 1:78–149, 2011.

[Pow11b] Victoria Powers. Rational certificates of positivity on compact semialgebraic sets. *Pacific journal of mathematics*, 251(2):385–391, 2011.

[Pow21] Victoria Powers. Certificates of positivity for real polynomials. 2021.

[PP08] Helfried Peyrl and Pablo A Parrilo. Computing sum of squares decompositions with rational coefficients. *Theoretical Computer Science*, 409(2):269–281, 2008.

[PR01] Victoria Powers and Bruce Reznick. A new bound for Pólya's theorem with applications to polynomials positive on polyhedra. *Journal of pure and applied algebra*, 164(1-2):221–229, 2001.

[PT82] Svatopluk Poljak and Daniel Turzik. A polynomial algorithm for constructing a large bipartite subgraph, with an application to a satisfiability problem. *Canadian Journal of Mathematics*, 34(3):519–524, 1982.

[PT94] Svatopluk Poljak and Zsolt Tuza. The expected relative error of the polyhedral approximation of the max-cut problem. *Operations Research Letters*, 16(4):191–198, 1994.

[Put93] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *40th Annual ACM Symposium on Theory of Computing (STOC 2008)*, pages 245–254. ACM, 2008.

[Rez95] Bruce Reznick. Uniform denominators in Hilbert's seventeenth problem. *Mathematische Zeitschrift*, 220(1):75–97, 1995.

[Rez00] Bruce Reznick. Some concrete aspects of Hilbert's 17th problem. *Contemporary Mathematics*, 253:251–272, 2000.

[Rez07] Bruce Reznick. On Hilbert's construction of positive polynomials. *arXiv preprint arXiv:0707.2156*, 2007.

[Sch91] Konrad Schmüdgen. The k-moment problem for compact semi-algebraic sets. *Mathematische Annalen*, 289(1):203–206, 1991.

[Sch08] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 593–602. IEEE, 2008.

[Sch12] Konrad Schmüdgen. Around Hilbert's 17th problem. *Documenta Mathematica, Optimization stories, extra volume ISMP*, pages 433–438, 2012.

[Sch16] Claus Scheiderer. Sums of squares of polynomials with rational coefficients. *Journal of the European Mathematical Society*, 18(7):1495–1513, 2016.

[Sei91] JJ Seidel. Introduction to association schemes. *Séminaire Lotharingien de Combinatoire [electronic only]*, 26:B26g–17, 1991.

[SG74] Sartaj Sahni and Teofilo Gonzales. P-complete problems and approximate solutions. In *15th Annual Symposium on Switching and Automata Theory (swat 1974)*, pages 28–32. IEEE, 1974.

[Shi19] Jonathan Shi. *Tensor rank decompositions via the pseudo-moment method*. PhD thesis, Cornell University, 2019.

[Sho87] Naum Zuselevich Shor. An approach to obtaining global extremums in polynomial mathematical programming problems. *Cybernetics*, 23(5):695–700, 1987.

[SL22] Lucas Slot and Monique Laurent. Sum-of-squares hierarchies for binary polynomial optimization. *Mathematical Programming*, pages 1–40, 2022.

[Ste74] Gilbert Stengle. A Nullstellensatz and a Positivstellensatz in semialgebraic geometry. *Mathematische Annalen*, 207(2):87–97, 1974.

[STT07] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovász-Schrijver LP relaxations of vertex cover and max cut. In *Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing*, pages 302–310, 2007.

[Stu02] Bernd Sturmfels. *Solving systems of polynomial equations*. Number 97. American Mathematical Society, 2002.

[Tau70] Olga Taussky. Sums of squares. *The American Mathematical Monthly*, 77(8):805–830, 1970.

[Tre12] Luca Trevisan. Max cut and the smallest eigenvalue. *SIAM Journal on Computing*, 41(6):1769–1786, 2012.

[Tro15] Joel A Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[Vit81] Paul MB Vitányi. How well can a graph be n-colored? *Discrete Mathematics*, 34(1):69–80, 1981.

[Wed72] Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972.

[YWS15] Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.