E-Companion to *"Sum of Squares Submodularity"*

Appendix A: Additional Material Related to Section 3

A.1. Background Material and Notation

The first two results we make use of are linked to multilinear polynomials and are straightforward to show.

LEMMA EC.1. *Let $F:\mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial. We have:*

$$\frac{\partial F(x)}{\partial x_i} = F(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n) - F(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_n), \forall i = 1,\ldots,n.$$

LEMMA EC.2. *Let $F:\mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial. We have:*

$$F(x_1,\ldots,x_i,\ldots,x_n) = x_i F(x_1,\ldots,1,\ldots,x_n) + (1-x_i)F(x_1,\ldots,0,\ldots,x_n),\ \forall i = 1,\ldots,n.$$

Some additional background material in algebraic geometry is required for some proofs, which readers familiar with the material may skip. We assume that some concepts, e.g., monomial orderings, radical ideals, and Gröbner bases, are known; see (Laurent 2009, Section 2) or (Blekherman et al. 2012, Appendix A.2) for definitions. Recall the definitions of the ideals given in (1) and (3).

PROPOSITION EC.1. *Assuming $s_n > \ldots > s_1 > y_n > \ldots > y_1 > x_n > \ldots > x_1$, the polynomials generating the ideals in (1) and (3) constitute reduced Gröbner bases of the corresponding ideals with respect to the graded lexicographic order.*

This is straightforward to show using, e.g., the definition of a reduced Gröbner basis. Given a polynomial $p \in \mathbb{R}[x]$ and an ideal with a reduced Gröbner basis, the *normal form* of p with respect to the Gröbner basis, denoted by $\overline{p}$, is the remainder of p when performing multivariate division on p by the Gröbner basis. This remainder is *unique*. Normal forms have the following properties.

LEMMA EC.3 **(see, e.g., Permenter and Parrilo (2012)).** *Let $p,q \in \mathbb{R}[x]$ and let I be an ideal with associated reduced Gröbner basis G. Further, let $\overline{p}$ and $\overline{q}$ be the normal forms of p and q with respect to G. We have:*

1. *Ideal membership:* $p \in I \Leftrightarrow \overline{p} = 0$
2. *Arithmetic identities:* $\overline{p(x)\cdot q(x)} = \overline{\overline{p(x)}\cdot\overline{q(x)}}$ *and* $\overline{p(x)+q(x)} = \overline{p(x)} + \overline{q(x)}$.
3. *Congruence modulo I:* $p \equiv q \mod I$ *if and only if* $\overline{p} = \overline{q}$.

Remark EC.1. To obtain the normal forms of a given polynomial p modulo any one of the three ideals under consideration, we simply use the equations that describe these ideals to sub out the higher degree variables/monomials (e.g., x_i^2 for $I_2[x]$) that appear in p by the lower degree variables/monomials (e.g., x_i for $I_2[x]$).

Remark EC.2. In the definition of t-sos modulo I (see Definition 5), if G is a reduced Gröbner basis of I, one can assume without loss of generality that q_r is in normal form with respect to G; see (Blekherman et al. 2012, Section 3.3.5). In particular, for all three of our ideals, we can assume that q_r is a multilinear polynomial in the appropriate variables.

Definition EC.1. Given $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, and for $i \in \{0, \ldots, n+1\}$, we let $[xy]^i$ to be the vector of variables of size n whose k^{th} entry is given by:

$$[xy]^i_k = \begin{cases} x_k & \text{if } k = 1, \ldots, i \\ y_k & \text{if } k = i+1, \ldots, n. \end{cases}$$

Furthermore, for a vector of variables x and $\alpha \in \mathbb{R}$ or a variable, we let $[x]^{k=\alpha}$ be the vector x with k^{th} entry set to α. We further let $[x]^{k=\alpha, j=\beta}$ be the vector x with k^{th} entry set to α and j^{th} entry set to β, where α, β are real numbers or variables.

A.2. Proofs of Results in Section 3.1.

Readers familiar with the sum of squares literature may find these arguments standard and may wish to skip this section as a consequence.

Proof of Proposition 3. Let $t \in \mathbb{N}$ and let f be t-sos-submodular with MLE F. Further, let $i, j \in \{1, \ldots, n\}, i \neq j$. We have:

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} \equiv \sum_r q_r^2(x_{-i,j}) \mod I_2[x_{-i,j}] \tag{EC.1}$$

for some $q_r \in \mathbb{R}[x_{-i,j}]$ such that $\deg(q_r) \leq t$. Recall from Section 2.4 that $\mathcal{V}_{\mathbb{R}}(I_2[x_{-i,j}]) = \{0,1\}^{n-2}$ and that (EC.1) implies

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} = \sum_r q_r^2(x_{-i,j}),\ \forall x_{-i,j} \in \mathcal{V}_{\mathbb{R}}(I_2[x_{-i,j}]).$$

As $q_r^2(x_{-i,j}) \geq 0$, $\forall x_{-i,j} \in \mathbb{R}^{n-2}$, we have $\sum_r q_r^2(x_{-i,j}) \geq 0$, $\forall x_{-i,j} \in \{0,1\}^{n-2}$, and $-\partial^2 F(x_{-i,j})/\partial x_i \partial x_j \geq 0$, $\forall x_{-i,j} \in \{0,1\}^{n-2}$. The result follows from Proposition 1, (iii'). □

We use the notation S^n to denote the set of symmetric matrices in $\mathbb{R}^{n \times n}$.

Proof of Proposition 4. Let $f : 2^\Omega \to \mathbb{R}$ be a set function and let F be its MLE. Following (Blekherman et al. 2012, Section 3.3.5), checking whether f is t-sos-submodular amounts to solving the following feasibility semidefinite program:

$$\begin{aligned} &\min_{Q_{ij}\in S^N, i,j=1,\ldots,n, i\neq j} \quad 0 \\ &\text{s.t. } -\overline{\frac{\partial^2 F(x)}{\partial x_i x_j}} = \operatorname{tr}\left(Q_{ij}\overline{z(x_{-i,j})z(x_{-i,j})^T}\right), \forall i,j=1,\ldots,n, i\neq j, \\ &Q_{ij} \succeq 0, \forall i,j=1,\ldots,n, i\neq j. \end{aligned} \tag{EC.2}$$

Here, the overline refers to the normal form with respect to $I_2[x_{-i,j}]$ (see Appendix A.1), $N = \sum_{k=0}^{t}\binom{n-2}{k}$, and $z(x_{-i,j})$ is the set of all multilinear monomials in $x_{-i,j}$ of degree less than or equal to t. As $x_{-i,j} \in \mathbb{R}^{n-2}$, this means that the size of $z(x_{-i,j})$ is N, which justifies the size of Q. □

To give the reader a clearer view of the SDP formulations of checking t-sos-submodularity for a given t, we derive the SDPs for testing 2 and 3-sos-submodularity on the function in Example 1. Code to solve these SDPs is provided here which uses the open source solver SCS.

Example EC.1. Recall the degree-5 set function f defined on a ground set of size $n = 5$ given in Example 1. Its multilinear extension F is given by

$$F(x_1,x_2,x_3,x_4,x_5) = 3x_1x_2x_3x_4x_5 - 4x_3x_1x_2 - 9x_4x_1x_2 - 12x_1x_2x_4x_5 - 4x_1x_2x_3x_5 - 4x_1x_2x_3x_4 + 2.$$

To determine whether f is 2-sos-submodular, we need to solve (EC.2) with $n = 5, N = 7$, and $z(x_{-i,j}) = [1, x_k, x_\ell, x_m, x_kx_\ell, x_kx_m, x_{\ell m}]^T$ for $\{k,\ell,m\} \in \{1,\ldots,5\}\backslash\{i,j\}$ and $i,j \in \{1,\ldots,5\}, i\neq j$. This optimization problem can be reformulated as an explicit SDP as we see now. We set $i = 1$ and $j = 2$ to illustrate. Other values of i and j are dealt with identically. Let $Q_{12} \in S^7$ and denote by $Q_{12}^{k\ell}$ the (k,ℓ)-entry of Q_{12}. The constraint

$$-\overline{\frac{\partial^2 F(x)}{\partial x_1 x_2}} = \operatorname{tr}\left(Q_{12}\overline{z(x_{-1,2})z(x_{-1,2})^T}\right)$$

requires us to derive the partial derivative of F with respect to x_1 and x_2 :

$$-\frac{\partial^2 F}{\partial x_1\,\partial x_2} = -3x_3x_4x_5 + 4x_3x_4 + 4x_3x_5 + 4x_3 + 12x_4x_5 + 9x_4.$$

With $z(x_{-1,2}) = [1, x_3, x_4, x_5, x_3x_4, x_3x_5, x_4x_5]^T$, the constraint then becomes:

$$\begin{aligned} &-3x_3x_4x_5 + 4x_3x_4 + 4x_3x_5 + 4x_3 + 12x_4x_5 + 9x_4 \\ &= \operatorname{tr}(Q_{12}\overline{z(x_{-1,2})z(x_{-1,2})^T}) \end{aligned}$$

$$
= \mathrm{tr}\left(Q_{12} \cdot \begin{pmatrix} 1 & x_3 & x_4 & x_5 & x_3x_4 & x_3x_5 & x_4x_5 \\ x_3 & x_3 & x_3x_4 & x_3x_5 & x_3x_4 & x_3x_5 & x_3x_4x_5 \\ x_4 & x_3x_4 & x_4 & x_4x_5 & x_3x_4 & x_3x_4x_5 & x_4x_5 \\ x_5 & x_3x_5 & x_4x_5 & x_5 & x_3x_4x_5 & x_3x_5 & x_4x_5 \\ x_3x_4 & x_3x_4 & x_3x_4 & x_3x_4x_5 & x_3x_4 & x_3x_4x_5 & x_3x_4x_5 \\ x_3x_5 & x_3x_5 & x_3x_4x_5 & x_3x_5 & x_3x_4x_5 & x_3x_5 & x_3x_4x_5 \\ x_4x_5 & x_3x_4x_5 & x_4x_5 & x_4x_5 & x_3x_4x_5 & x_3x_4x_5 & x_4x_5 \end{pmatrix} \right)
$$
$$
\begin{aligned}
&= Q_{12}^{11} + (2Q_{12}^{12} + Q_{12}^{22})x_3 + (2Q_{12}^{13} + Q_{12}^{33})x_4 + (2Q_{12}^{14} + Q_{12}^{44})x_5 + (2Q_{12}^{15} + 2Q_{12}^{23} + 2Q_{12}^{25} + 2Q_{12}^{35} + Q_{12}^{55})x_3x_4 \\
&\quad + (2Q_{12}^{16} + 2Q_{12}^{24} + 2Q_{12}^{26} + 2Q_{12}^{46} + Q_{12}^{66})x_3x_5 + (2Q_{12}^{17} + 2Q_{12}^{34} + 2Q_{12}^{37} + 2Q_{12}^{47} + Q_{12}^{77})x_4x_5 \\
&\quad + (2Q_{12}^{27} + 2Q_{12}^{36} + 2Q_{12}^{45} + 2Q_{12}^{56} + 2Q_{12}^{57} + 2Q_{12}^{67})x_3x_4x_5.
\end{aligned}
$$

By matching coefficients of the polynomial on the first line and the polynomial on the last three lines, we obtain the following linear equalities:

$$
\begin{aligned}
&Q_{12}^{11} = 0, \quad 2Q_{12}^{12} + Q_{12}^{22} = 4, \quad 2Q_{12}^{13} + Q_{12}^{33} = 9, \quad 2Q_{12}^{14} + Q_{12}^{44} = 0, \\
&2Q_{12}^{15} + 2Q_{12}^{23} + 2Q_{12}^{25} + 2Q_{12}^{35} + Q_{12}^{55} = 4, \quad 2Q_{12}^{16} + 2Q_{12}^{24} + 2Q_{12}^{26} + 2Q_{12}^{46} + Q_{12}^{66} = 4, \\
&2Q_{12}^{17} + 2Q_{12}^{34} + 2Q_{12}^{37} + 2Q_{12}^{47} + Q_{12}^{77} = 12, \quad 2Q_{12}^{27} + 2Q_{12}^{36} + 2Q_{12}^{45} + 2Q_{12}^{56} + 2Q_{12}^{57} + 2Q_{12}^{67} = -3,
\end{aligned}
$$

to be combined with the constraint $Q_{12} \succeq 0$. We can then proceed in an identical fashion for all $(i,j) \in \{1,\ldots,n\}, i \neq j$. We thus obtain the following SDP reformulation of (EC.2):

$$
\begin{aligned}
&\min_{Q_{12},Q_{13},Q_{14},Q_{15},Q_{23},Q_{24},Q_{25},Q_{34},Q_{35},Q_{45} \in S^7} \quad 0 \\
\text{s.t. } &Q_{12}^{11} = 0, \quad 2Q_{12}^{12} + Q_{12}^{22} = 4, \quad 2Q_{12}^{13} + Q_{12}^{33} = 9, \quad 2Q_{12}^{14} + Q_{12}^{44} = 0, \\
&2Q_{12}^{15} + 2Q_{12}^{23} + 2Q_{12}^{25} + 2Q_{12}^{35} + Q_{12}^{55} = 4, \quad 2Q_{12}^{16} + 2Q_{12}^{24} + 2Q_{12}^{26} + 2Q_{12}^{46} + Q_{12}^{66} = 4, \\
&2Q_{12}^{17} + 2Q_{12}^{34} + 2Q_{12}^{37} + 2Q_{12}^{47} + Q_{12}^{77} = 12, \quad 2Q_{12}^{27} + 2Q_{12}^{36} + 2Q_{12}^{45} + 2Q_{12}^{56} + 2Q_{12}^{57} + 2Q_{12}^{67} = -3 \\
&Q_{13}^{11} = 0, \quad 2Q_{13}^{12} + Q_{13}^{22} = 4, \quad 2Q_{13}^{13} + Q_{13}^{33} = 0, \quad 2Q_{13}^{14} + Q_{13}^{44} = 0, \\
&2Q_{13}^{15} + 2Q_{13}^{23} + 2Q_{13}^{25} + 2Q_{13}^{35} + Q_{13}^{55} = 4, \quad 2Q_{13}^{16} + 2Q_{13}^{24} + 2Q_{13}^{26} + 2Q_{13}^{46} + Q_{13}^{66} = 4, \\
&2Q_{13}^{17} + 2Q_{13}^{34} + 2Q_{13}^{37} + 2Q_{13}^{47} + Q_{13}^{77} = 0, \quad 2Q_{13}^{27} + 2Q_{13}^{36} + 2Q_{13}^{45} + 2Q_{13}^{56} + 2Q_{13}^{57} + 2Q_{13}^{67} = -3 \\
&\vdots \\
&Q_{45}^{11} = 0, \quad 2Q_{45}^{12} + Q_{45}^{22} = 0, \quad 2Q_{45}^{13} + Q_{45}^{33} = 0, \quad 2Q_{45}^{14} + Q_{45}^{44} = 0, \\
&2Q_{45}^{15} + 2Q_{45}^{23} + 2Q_{45}^{25} + 2Q_{45}^{35} + Q_{45}^{55} = 12, \quad 2Q_{45}^{16} + 2Q_{45}^{24} + 2Q_{45}^{26} + 2Q_{45}^{46} + Q_{45}^{66} = 0, \\
&2Q_{45}^{17} + 2Q_{45}^{34} + 2Q_{45}^{37} + 2Q_{45}^{47} + Q_{45}^{77} = 0, \quad 2Q_{45}^{27} + 2Q_{45}^{36} + 2Q_{45}^{45} + 2Q_{45}^{56} + 2Q_{45}^{57} + 2Q_{45}^{67} = -3, \\
&Q_{12}, Q_{13}, Q_{14}, Q_{15}, Q_{23}, Q_{24}, Q_{25}, Q_{34}, Q_{35}, Q_{45} \succeq 0.
\end{aligned} \tag{EC.3}
$$

The SDP (EC.3) can be solved numerically; see here for an implementation. In an echo of Example 1, we find that the problem is infeasible, implying that f is not 2-sos-submodular.

To test whether f is 3-sos-submodular, we change $z(x_{-i,j})$ to allow for multilinear monomials of degree up to 3, that is, we let $z(x_{-i,j}) = [1, x_k, x_\ell, x_m, x_k x_\ell, x_k x_m, x_{\ell m}, x_{k\ell m}]^T$ for $\{k,\ell,m\} \in$

$\{1,\ldots,5\}\backslash\{i,j\}$ and $i,j\in\{1,\ldots,5\}, i\neq j$. We then solve (EC.2) once again with decision variables $\{Q_{ij}\}_{i,j}$ that are of size 8×8. Proceeding in a similar way as above for its derivation, we obtain the slightly larger SDP:

$$
\begin{aligned}
&\min_{Q_{12},Q_{13},Q_{14},Q_{15},Q_{23},Q_{24},Q_{25},Q_{34},Q_{35},Q_{45}\in S^8} 0\\
\text{s.t. } & Q_{12}^{11}=0,\quad 2Q_{12}^{12}+Q_{12}^{22}=4,\quad 2Q_{12}^{13}+Q_{12}^{33}=9,\quad 2Q_{12}^{14}+Q_{12}^{44}=0,\\
& 2Q_{12}^{15}+2Q_{12}^{23}+2Q_{12}^{25}+2Q_{12}^{35}+Q_{12}^{55}=4,\quad 2Q_{12}^{16}+2Q_{12}^{24}+2Q_{12}^{26}+2Q_{12}^{46}+Q_{12}^{66}=4,\\
& 2Q_{12}^{17}+2Q_{12}^{34}+2Q_{12}^{37}+2Q_{12}^{47}+Q_{12}^{77}=12,\\
& 2Q_{12}^{18}+2Q_{12}^{27}+2Q_{12}^{28}+2Q_{12}^{36}+2Q_{12}^{38}+2Q_{12}^{45}+2Q_{12}^{48}+2Q_{12}^{56}+2Q_{12}^{57}+2Q_{12}^{58}+2Q_{12}^{67}+2Q_{12}^{68}+2Q_{12}^{78}+Q_{12}^{88}=-3\\
& \vdots\\
& Q_{45}^{11}=0,\quad 2Q_{45}^{12}+Q_{45}^{22}=0,\quad 2Q_{45}^{13}+Q_{45}^{33}=0,\quad 2Q_{45}^{14}+Q_{45}^{44}=0,\\
& 2Q_{45}^{15}+2Q_{45}^{23}+2Q_{45}^{25}+2Q_{45}^{35}+Q_{45}^{55}=12,\quad 2Q_{45}^{16}+2Q_{45}^{24}+2Q_{45}^{26}+2Q_{45}^{46}+Q_{45}^{66}=0,\\
& 2Q_{45}^{17}+2Q_{45}^{34}+2Q_{45}^{37}+2Q_{45}^{47}+Q_{45}^{77}=0,\\
& 2Q_{45}^{18}+2Q_{45}^{27}+2Q_{45}^{28}+2Q_{45}^{36}+2Q_{45}^{38}+2Q_{45}^{45}+2Q_{45}^{48}+2Q_{45}^{56}+2Q_{45}^{57}+2Q_{45}^{58}+2Q_{45}^{67}+2Q_{45}^{68}+2Q_{45}^{78}+Q_{45}^{88}=-3\\
& Q_{12},Q_{13},Q_{14},Q_{15},Q_{23},Q_{24},Q_{25},Q_{34},Q_{35},Q_{45}\succeq 0.
\end{aligned}
\tag{EC.4}
$$

The SDP (EC.4) is feasible, and a feasible solution can be obtained numerically; see here. This establishes that f is 3-sos-submodular, and hence submodular.

We now move onto the proof of Proposition 5, which makes use of the following result which is an immediate corollary of (Sakaue et al. 2016, Theorem 3.2).

Lemma EC.4 (**Sakaue et al. (2016)**). *Consider the binary polynomial optimization problem:*

$$
\begin{aligned}
&\min_z f(z)\\
&s.t.\ z\in\{-1,1\}^n,
\end{aligned}
\tag{EC.5}
$$

where f is a polynomial of degree d and in n variables. The semidefinite programming-based relaxation of order $\lceil\frac{n+d-1}{2}\rceil$ in Lasserre's hierarchy is an exact reformulation of (EC.5).

We can then combine Lemma EC.4, where we operate the affine change of variables $x_i=(z_i+1)/2$ for $i=1,\ldots,n$ to work on $\{0,1\}^n$ instead of $\{-1,1\}^n$, with known results in the sum of squares literature regarding the Lasserre hierarchy (see (Laurent 2009, Section 6.2)) to obtain the following corollary.

Corollary EC.1. *Let f be a degree-d, n-variate polynomial, nonnegative on $\{0,1\}^n$. For any $t\geq\lceil\frac{n+d-1}{2}\rceil$, f is t-sos modulo $I_2[x]$.*

Proof of Proposition 5. Let $f : 2^\Omega \to \mathbb{R}$ be a set function of degree d and let F be its MLE. Note that $-\frac{\partial^2 F(x)}{\partial x_i \partial x_j}$ is of degree at most $d-2$ for any $i, j \in \{1, \dots, n\}$. To obtain monomials of degree $d-2$ when expanding a sum of squares polynomial, it must be the case that this sum of squares polynomial has degree at least $\lceil \frac{d-2}{2} \rceil$.

Now suppose that f is of degree d and is submodular. Thus, from Proposition 1, it follows that $-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} \geq 0, \forall x \in \{0,1\}^n$ and $i, j \in \{1, \dots, n\}$. Fix $i, j \in \{1, \dots, n\}$ and note that $-\frac{\partial^2 F(x)}{\partial x_i \partial x_j}$ is of degree $d-2$ and in $n-2$ variables. The fact that $-\frac{\partial^2 F(x)}{\partial x_i \partial x_j}$ is t-sos mod $I_2[x_{-i,j}]$ for t at most $\lceil \frac{n+d-5}{2} \rceil$ follows Corollary EC.1. □

A.3. Proofs of Results in Section 3.2.

Recall the notation given in Definition EC.1 in Section A.1.

Lemma EC.5. *Let f be a set function and let F be its MLE. We have:*

$$\frac{\partial F}{\partial x_i}(x) - \frac{\partial F}{\partial x_i}(y) = \sum_{j=1, j\neq i}^{n} (y_j - x_j) \cdot \left(-\frac{\partial^2 F}{\partial x_i \partial x_j}([xy]^j) \right), \textit{ for } i = 1, \dots, n \tag{EC.6}$$

$$F(x) - F(y) = -\sum_{i=1}^{n} (y_i - x_i) \frac{\partial F}{\partial x_i}([xy]^i), \tag{EC.7}$$

$$F(x) - F(y) = -\sum_{i=1}^{n} (y_i - x_i) \frac{\partial F}{\partial x_i}([yx]^i) \tag{EC.8}$$

Proof. We first show (EC.6). Let $i \in \{1, \dots, n\}$. Using telescopic sums, we have:

$$\frac{\partial F}{\partial x_i}(x) - \frac{\partial F}{\partial x_i}(y) = \frac{\partial F}{\partial x_i}([xy]^n) - \frac{\partial F}{\partial x_i}([xy]^0) = \sum_{j=1, j\neq i}^{n} \left(\frac{\partial F}{\partial x_i}([xy]^j) - \frac{\partial F}{\partial x_i}([xy]^{j-1}) \right).$$

We observe that $[xy]^j$ and $[xy]^{j-1}$ only differ in entry j, that is, $[xy]^{j,j=1} = [xy]^{j-1,j=1}$ and $[xy]^{j,j=0} = [xy]^{j-1,j=0}$. Thus, from Lemmas EC.1 and EC.2,

$$\begin{aligned}
\frac{\partial F}{\partial x_i}([xy]^j) - \frac{\partial F}{\partial x_i}([xy]^{j-1}) =& x_j \frac{\partial F}{\partial x_i}([xy]^{j,j=1}) + (1-x_j) \frac{\partial F}{\partial x_i}([xy]^{j,j=0}) - y_j \frac{\partial F}{\partial x_i}([xy]^{j-1,j=1}) \\
&- (1-y_j) \frac{\partial F}{\partial x_i}([xy]^{j-1,j=0}) \\
&= (y_j - x_j) \cdot \left(\frac{\partial F}{\partial x_i}([xy]^{j,j=0}) - \frac{\partial F}{\partial x_i}([xy]^{j,j=1}) \right) \\
&= (y_j - x_j) \cdot \left(-\frac{\partial^2 F}{\partial x_i \partial x_j}([xy]^j) \right).
\end{aligned}$$

We now show (EC.7). Similarly to above, using telescopic sums, we have:

$$F(x) - F(y) = F([xy]^n) - F([xy]^0) = \sum_{j=1}^{n} \left(F([xy]^j) - F([xy]^{j-1}) \right).$$

As above, note that:

$$\begin{aligned}F([xy]^j) - F([xy]^{j-1}) &= x_j F([xy]^{j,j=1}) + (1-x_j)F([xy]^{j,j=0}) \\ &\quad - y_j F([xy]^{j-1,j=1}) - (1-y_j)F([xy]^{j-1,j=0}) \\ &= -(y_j - x_j)\frac{\partial F}{\partial x_j}([xy]^j_{-j}),\end{aligned}$$

which gives the conclusion. Finally, we show (EC.8). We use telescopic sums to obtain:

$$F(x) - F(y) = F([yx]^0) - F([yx]^n) = \sum_{j=1}^{n}\left(F([yx]^{j-1}) - F([yx]^j)\right).$$

Similarly to before, we have:

$$\begin{aligned}F([yx]^{j-1}) - F([yx]^j) &= x_j F([yx]^{j-1,j=1}) + (1-x_j)F([yx]^{j-1,j=0}) \\ &\quad - y_j F([yx]^{j,j=1}) - (1-y_j)F([yx]^{j,j=0}) \\ &= -(y_j - x_j)\frac{\partial F}{\partial x_j}([yx]^j),\end{aligned}$$

which leads to the conclusion. □

Proof of Theorem 1. Suppose that f is t-sos-submodular. For $i, j \in \{1, \ldots, n\}, i \neq j$, we have:

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} \equiv \sum_r q_{r,ij}^2(x_{-i,j}) \quad \text{mod } I_2[x_{-i,j}]$$

where $q_{r,ij} \in \mathbb{R}[x_{-i,j}]$ and $\deg(q_{r,ij}) \leq t$ for $r = 1, \ldots, m$. From Lemma EC.5, for fixed $i = 1, \ldots, n$, we then have:

$$\begin{aligned}\frac{\partial F}{\partial x_i}(x) - \frac{\partial F}{\partial x_i}(y) &= \sum_{j=1, j\neq i}^{n}(y_j - x_j)\left(-\frac{\partial^2 F([xy]^j)}{\partial x_i \partial x_j}\right) \\ &\equiv \sum_{j=1, j\neq i}^{n}\sum_r (y_j - x_j)q_{r,ij}^2([xy]^j_{-i,j}) \ \text{ mod } I_2[[xy]^j_{-i,j}].\end{aligned}$$

Now, for any $j \in \{1, \ldots, n\}, j \neq i$, note that $I_2[[xy]^j_{-i,j}] \subseteq I_1[x_{-i}, y_{-i}]$ and

$$(y_j - x_j)^2 \equiv y_j - x_j \quad \text{mod } I_1[x_{-i}, y_{-i}].$$

Thus,

$$\frac{\partial F(x)}{\partial x_i} - \frac{\partial F(y)}{\partial x_i} = \sum_{r=1}^{m}\sum_{j=1, j\neq i}^{n}\left((y_j - x_j)q_{r,ij}([xy]^j_{-i,j})\right)^2 \quad \text{mod } I_1[x_{-i}, y_{-i}].$$

As $\deg(q_{r,ij}) \leq t$ for $q_{r,ij} \in \mathbb{R}[x_{-i,j}]$, $\deg\left((y_j - x_j)q_{r,ij}([xy]^j_{-i,j})\right) \leq t+1$, which gives the result.

Now let $i \in \{1, \dots, n\}$. Suppose that we have:

$$\frac{\partial F(x)}{\partial x_i} - \frac{\partial F(y)}{\partial x_i} \equiv \sum_r q_r^i(x_{-i}, y_{-i})^2 \mod I_1[x_{-i}, y_{-i}] \tag{EC.9}$$

where $q_r^i \in \mathbb{R}[x_{-i}, y_{-i}]$ is multilinear (see Remark EC.2) and $\deg(q_r^i) \leq t+1$. Note that

$$\frac{\partial F(x)}{\partial x_i} - \frac{\partial F(x)}{\partial x_i} = 0 \equiv \sum_r q_r^i(x_{-i}, x_{-i})^2 \mod I_1[x_{-i}].$$

Thus, $q_r^i(x_{-i}, x_{-i}) \equiv 0 \mod I_1[x_{-i}]$ for any r. Now, let $j \in \{1, \dots, n\}$ such that $j \neq i$. We have:

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} = \frac{\partial F([x]^{j=0})}{\partial x_i} - \frac{\partial F([x]^{j=1})}{\partial x_i}. \tag{EC.10}$$

As q_r^i in (EC.9) is multilinear, we can write using Lemma EC.2,

$$\begin{aligned} q_r^i(x_{-i}, [x]_{-i}^{j=y_j}) &= x_j y_j q_r^i([x]_{-i}^{j=1}, [x]_{-i}^{j=1}) + x_j(1-y_j) q_r^i([x]_{-i}^{j=1}, [x]_{-i}^{j=0}) \\ &\quad + (1-x_j) y_j q_r^i([x]_{-i}^{j=0}, [x]_{-i}^{j=1}) + (1-x_j)(1-y_j) q_r^i([x]_{-i}^{j=0}, [x]_{-i}^{j=0}) \\ &\equiv (y_j - x_j) q_r^i([x]_{-i}^{j=0}, [x]_{-i}^{j=1}) \mod I_1[x_{-i}, [x]_{-i}^{j=y_j}], \end{aligned}$$

where we have used the fact that $q_r^i(x_{-i}, x_{-i}) \equiv 0 \mod I_1[x_{-i}]$ (and $I_1[x_{-i}] \subseteq I_1[x_{-i}, [x]_{-i}^{j=y_j}]$) and the fact that $x_j y_j \equiv x_j \mod I_1[x_{-i}, [x]_{-i}^{j=y_j}]$. As $\deg(q_r^i) \leq t+1$ for $q_r^i \in \mathbb{R}[x_{-i}, y_{-i}]$, it follows that $\deg(q_r^i) \leq t+1$ when $q_r^i \in \mathbb{R}[x_{-i}, [x]_{-i}^{j=y_j}]$. As the normal forms of $q_r^i(x_{-i}, [x]_{-i}^{j=y_j})$ and $q_r^i([x]_{-i}^{j=0}, [x]_{-i}^{j=1}) \mod I_1[x_{-i}, [x]_{-i}^{j=y_j}]$ are themselves, we can use Lemma EC.3 to conclude that $\deg(q_r^i([x]_{-i}^{j=0}, [x]_{-i}^{j=1})) \leq t$. Combining (EC.10) and (EC.9), we then have that:

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} \equiv \sum_r q_r^i([x]_{-i}^{j=0}, [x]_{-i}^{j=1})^2 \mod I_1[[x]_{-i}^{j=0}, [x]_{-i}^{j=1}]$$

As $I_1[[x]_{-i}^{j=0}, [x]_{-i}^{j=1}] = I_2[x_{-i,j}]$ and $\deg(q_r^i([x]_{-i}^{j=0}, [x]_{-i}^{j=1})) \leq t$, f is t-sos-submodular. □

Proof of Theorem 2. Suppose that f is t-sos-submodular. Using (EC.7) and (EC.8) and the notation in Definition (EC.1), we have:

$$\begin{aligned} G(x,y,s) &= F(x) + F(y) - F(x+y-s) - F(s) \\ &= -\sum_{i=1}^n (y_i - s_i) \frac{\partial F}{\partial x_i}([x(x+y-s)]^i) - \sum_{i=1}^n (s_i - y_i) \frac{\partial F}{\partial x_i}([sy]^i) \\ &= \sum_{i=1}^n (y_i - s_i) \left(\frac{\partial F}{\partial x_i}([sy]^i) - \frac{\partial F}{\partial x_i}([x(x+y-s)]^i) \right). \end{aligned} \tag{EC.11}$$

Let $i \in \{1, \ldots, n\}$. From Theorem 1, as f is t-sos-submodular, we have

$$\frac{\partial F}{\partial x_i}(x) - \frac{\partial F}{\partial x_i}(y) \equiv \sum_r q_{ri}^2(x_{-i}, y_{-i}) \mod I_1[x_{-i}, y_{-i}],$$

where $q_r(x_{-i}, y_{-i}) \in \mathbb{R}[x_{-i}, y_{-i}]$ and $\deg(q_{ri}(x_{-i}, y_{-i})) \leq t+1$. This implies that

$$\frac{\partial F}{\partial x_i}([sy]^i) - \frac{\partial F}{\partial x_i}([x(x+y-s)]^i) \equiv \sum_r q_{ri}^2([sy]_{-i}^i, [x(x+y-s)]_{-i}^i) \mod I_1[[sy]_{-i}^i, [x(x+y-s)]_{-i}^i].$$

Note that $\deg(q_{ri}([sy]_{-i}^i, [x(x+y-s)]_{-i}^i)) \leq t+1$. Substituting this in (EC.11), we get:

$$G(x, y, s) \equiv \sum_{i=1}^n \sum_r (y_i - s_i) q_{ri}^2([sy]_{-i}^i, [x(x+y-s)]_{-i}^i) \mod I_1[[sy]_{-i}^i, [x(x+y-s)]_{-i}^i].$$

It is straightforward to show that $I_1[[sy]_{-i}^i, [x(x+y-s)]_{-i}^i] \subseteq I_0[x, y, s]$. Furthermore, $y_i - s_i \equiv (y_i - s_i)^2 \mod I_0[x, y, s]$. Renaming $q_{ri}([sy]_{-i}^i, [x(x+y-s)]_{-i}^i)$ to be $\bar{q}_{ri}(x_{-i}, y_{-i}, s_{-i})$, we get that

$$G(x, y, s) \equiv \sum_{i=1}^n ((y_i - s_i) \cdot \bar{q}_{ri}(x_{-i}, y_{-i}, s_{-i}))^2 \mod I_0[x, y, s].$$

As $\deg(\bar{q}_{ri}(x_{-i}, y_{-i}, s_{-i})) \leq t+1$, the result follows.

For the converse, we have that:

$$G(x, y, s) \equiv \sum_r q_r(x, y, s)^2 \mod I_0[x, y, s] \tag{EC.12}$$

where $q_r \in \mathbb{R}[x, y, s]$ and $\deg(q_r) \leq t+2$. As mentioned in Remark EC.2, we can take q_r to be in normal form with respect to $I_0[x, y, s]$; in particular, we can assume that it is multilinear.

It is not hard to see that $G(x, y, x) = 0$ for any x, y and so $q_r(x, y, x) \equiv 0 \mod I_0[x, y, x]$ for any r. Let $i \in \{1, \ldots, n\}$. Using this and the fact that q_r is multilinear, we write, for fixed r:

$$\begin{aligned}
q_r(x, y, [x]^{i=s_i}) = {} & (1-x_i) y_i s_i q_r([x]^{i=0}, [y]^{i=1}, [x]^{i=1}) + x_i y_i (1-s_i) q_r([x]^{i=1}, [y]^{i=1}, [x]^{i=0}) \\
& + (1-x_i)(1-y_i) s_i q_r([x]^{i=0}, [y]^{i=0}, [x]^{i=1}) + x_i (1-y_i)(1-s_i) q_r([x]^{i=1}, [y]^{i=0}, [x]^{i=0}) \\
& \equiv (x_i - s_i) q_r([x]^{i=1}, [y]^{i=0}, [x]^{i=0}) \mod I_0[x, y, s].
\end{aligned}$$

We have that $\deg(q_r(x, y, [x]^{i=s_i})) \leq t+2$ as a polynomial in $\mathbb{R}[x, y, s]$. Following a similar argument as the previous proof regarding normal forms, it follows that $\deg(q_r([x]^{i=1}, [y]^{i=0}, [x]^{i=0})) \leq t+1$.

Now, from Lemma EC.1, note that

$$\frac{\partial F(x)}{\partial x_i} - \frac{\partial F(y)}{\partial x_i} = F([x]^{i=1}) + F([y]^{i=0}) - F([y]^{i=1}) - F([x]^{i=0}) = G([x]^{i=1}, [y]^{i=0}, [x]^{i=0}).$$

Plugging (EC.12) into this, we get:

$$\begin{aligned}\frac{\partial F(x)}{\partial x_i} - \frac{\partial F(y)}{\partial x_i} &\equiv G([x]^{i=1}, [y]^{i=0}, [x]^{i=0}) \\ &\equiv \sum_r q_r^2([x]^{i=1}, [y]^{i=0}, [x]^{i=0}) \mod I_0[[x]^{i=1}, [y]^{i=0}, [x]^{i=0}]\end{aligned}$$

It can easily be shown that $I_0[[x]^{i=1}, [y]^{i=0}, [x]^{i=0} \subseteq I_1[x_{-i}, y_{-i}]$. Thus,

$$\frac{\partial F(x)}{\partial x_i} - \frac{\partial F(y)}{\partial x_i} \equiv \sum_r q_r^2([x]^{i=1}, [y]^{i=0}, [x]^{i=0}) \mod I_1[x_{-i}, y_{-i}]$$

As $\deg(q_r([x]^{i=1}, [y]^{i=0}, [x]^{i=0})) \le t+1$, the result follows. $\square$

Proof of Theorem 3. Suppose that f is t-sos-submodular. Using Lemma EC.5, we get that

$$F(x) - F(y) + \sum_{i=1}^n (y_i - x_i)\frac{\partial F}{\partial x_i}(x) = \sum_{i=1}^n (y_i - x_i)\left(\frac{\partial F}{\partial x_i}(x) - \frac{\partial F}{\partial x_i}([xy]^i)\right).$$

From Theorem 1, as f is t-sos-submodular, we have, for any $i \in \{1, \ldots, n\}$,

$$\frac{\partial F}{\partial x_i}(x) - \frac{\partial F}{\partial x_i}([xy]^i) \equiv \sum_r q_{ri}^2(x_{-i}, [xy]^i_{-i}) \mod I_1[x_{-i}, [xy]^i_{-i}],$$

where $\deg(q_{ri}(x_{-i}, [xy]^i_{-i})) \le t+1$. Renaming $q_{ri}(x_{-i}, [xy]^i_{-i})$ to $\bar{q}_{ri}(x_{-i}, y_{-i})$ and noting that $\deg(\bar{q}_{ri}) \le t+1$ and $I_1[x_i, [xy]^i_{-i}] \subseteq I_1[x_{-i}, y_{-i}]$, we get that, for any $i \in \{1, \ldots, n\}$,

$$\frac{\partial F}{\partial x_i}(x) - \frac{\partial F}{\partial x_i}([xy]^i) \equiv \sum_r \bar{q}_{ri}^2(x_{-i}, y_{-i}) \mod I_1[x_{-i}, y_{-i}].$$

As $I_1[x_{-i}, y_{-i}] \subseteq I_1[x, y]$ and $y_i - x_i \equiv (y_i - x_i)^2 \mod I_1[x, y]$ for any i, it follows that

$$F(x) - F(y) + \sum_{i=1}^n (y_i - x_i)\frac{\partial F}{\partial x_i}(x) \equiv \sum_{i=1}^n \sum_r ((y_i - x_i) \cdot \bar{q}_{ri}(x_{-i}, y_{-i}))^2 \mod I_1[x, y]$$

and we obtain the result.

For the converse, suppose that

$$H_1(x, y) \equiv \sum_r q_r^2(x, y) \mod I_1[x, y] \tag{EC.13}$$

for $q_r \in \mathbb{R}[x, y]$ and $\deg(q_r) \le t+2$. As in Remark EC.2, we can assume that q_r is in normal form with respect to $I_1[x, y]$. In particular, q_r is multilinear. It is straightforward to see that $H(x, x) = 0$, thus $q_r(x, x) \equiv 0 \mod I_1[x, y]$ for any r. Likewise, $H([x]^{i=0}, [x]^{i=1}) = 0$, which implies that

$q_r([x]^{i=0},[x]^{i=1}) \equiv 0 \mod I_1[x,y]$. Now, fix $i, j \in \{1,\dots,n\}$ with $i \neq j$. Recalling the notation in Definition EC.1 and multilinearity of q_r and using the equivalences above, we have:

$$\begin{aligned} q_r(x,[x]^{i=y_i,j=y_j}) &\equiv (1-x_i)(1-x_j)y_iy_jq_r([x]^{i=0,j=0},[x]^{i=1,j=1}) \mod I_1[x,[x]^{i=y_i,j=y_j}] \\ &\equiv (y_i-x_i)(y_j-x_j)q_r([x]^{i=0,j=0},[x]^{i=1,j=1}) \mod I_1[x,[x]^{i=y_i,j=y_j}]. \end{aligned}$$

From this, we obtain, as done in previous proofs, that $\deg(q_r([x]^{i=0,j=0},[x]^{i=1,j=1})) \leq t$. Now, note that:

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} = H_1([x]^{i=0,j=0},[x]^{i=1,j=1}) \equiv \sum_r q_r^2([x]^{i=0,j=0},[x]^{i=1,j=1}) \mod I_1[[x]^{i=0,j=0},[x]^{i=1,j=1}],$$

where we have used Lemma EC.1 and the equivalence from (EC.13). As $I_1[[x]^{i=0,j=0},[x]^{i=1,j=1}] \subseteq I_2[x_{-i,j}]$, and we have $\deg(q_r([x]^{i=0,j=0},[x]^{i=1,j=1})) \leq t$, it follows that f is t-sos-submodular. □

Proof of Theorem 4. Suppose that f is t-sos-submodular. Similarly to the proof of Theorem 3, we can show that:

$$F(y)-F(x)-\sum_{i=1}^n (y_i-x_i)\frac{\partial F(y)}{\partial x_i} = \sum_{i=1}^n (y_i-x_i)\left(\frac{\partial F([yx]^i)}{\partial x_i} - \frac{\partial F(y)}{\partial x_i}\right).$$

The proof of the implication is then identical to that of Theorem 3.

For the converse, note that, once again,

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} = H_2([x]^{i=0,j=0},[x]^{i=1,j=1}).$$

The proof of the converse is then also identical to that of Theorem 3. □

Proof of Theorem 5. Suppose that f is t-sos-submodular and let $i \in \{1,\dots,n\}$. From Theorem 1, we have that:

$$\frac{\partial F(x)}{\partial x_i} - \frac{\partial F(y)}{\partial x_i} \equiv \sum_r q_r^i(x_{-i},y_{-i})^2 \mod I_1[x_{-i},y_{-i}] \tag{EC.14}$$

where $q_r^i \in \mathbb{R}[x_{-i},y_{-i}]$ and $\deg(q_r^i) \leq t+1$. Using Lemma EC.5, we have that:

$$F(x+y-s)-F(x) = \sum_{i=1}^n (y_i-s_i)\frac{\partial F([(x+y-s)x]^i)}{\partial x_i}. \tag{EC.15}$$

We also have:

$$F(x+y-s)-F(y) = \sum_{i=1}^n (x_i-s_i)\frac{\partial F([(x+-s)y]^i}{\partial x_i}. \tag{EC.16}$$

Now, using (EC.14), we get:

$$\frac{\partial F([(x+y-s)x]^i)}{\partial x_i} \equiv \frac{\partial F(x)}{\partial x_i} - \sum_r q_r^i(x_{-i}, [(x+y-s)x]_{-i}^i)^2 \mod I_1[x_{-i}, [(x+y-s)x]_{-i}^i],$$

where $\deg(q_r^i(x_{-i}, [(x+y-s)x]_{-i}^i)) \le t+1$. Likewise,

$$\frac{\partial F([(x+y-s)y]^i)}{\partial x_i} \equiv \frac{\partial F(x+y-s)}{\partial x_i} + \sum_r q_r^i([(x+y-s)y]_{-i}^i, (x+y-s)_{-i})^2$$
$$\mod I_1[[(x+y-s)y]_{-i}^i, (x+y-s)_{-i}],$$

where $\deg(q_r^i([(x+y-s)y]_{-i}^i, (x+y-s)_{-i})) \le t+1$. Plugging these equivalences back into (EC.15) and (EC.16) and noting that:

$$I_1[x_{-i}, [(x+y-s)x]_{-i}^i] \subseteq I_0[x,y,s] \text{ and } I_1[[(x+y-s)y]_{-i}^i, (x+y-s)_{-i}] \subseteq I_0[x,y,s],$$

we get the following equivalences, modulo $I_0[x,y,s]$,

$$F(x+y-s) - F(x) \equiv \sum_{i=1}^n (y_i - s_i)\left(\frac{\partial F(x)}{\partial x_i} - \sum_r q_r^i(x_{-i}, [(x+y-s)x]_{-i}^i)^2\right)$$
$$F(x+y-s) - F(y) \equiv \sum_{i=1}^n (x_i - s_i)\left(\frac{\partial F(x+y-s)}{\partial x_i} + \sum_r q_r^i([(x+y-s)y]_{-i}^i, (x+y-s)_{-i})^2\right)$$

Subtracting the first one from the second one, we obtain:

$$G_1(x,y,s) \equiv \sum_{i=1}^n (y_i - s_i)\sum_r q_r^i(x_{-i}, [(x+y-s)x]_{-i}^i)^2$$
$$+ \sum_{i=1}^n (x_i - s_i)\sum_r q_r^i([(x+y-s)y]_{-i}^i, (x+y-s)_{-i})^2 \mod I_0[x,y,s].$$

Noting that $(y_i - s_i) \equiv (y_i - s_i)^2 \mod I_0[x,y,s]$ and $(x_i - s_i) \equiv (x_i - s_i)^2 \mod I_0[x,y,s]$, we get that, modulo $I_0[x,y,s]$,

$$G_1(x,y,s) \equiv \sum_{i=1}^n \sum_r \left(\left((y_i - s_i) q_r^i(x_{-i}, [(x+y-s)x]_{-i}^i)\right)^2 + \left((x_i - s_i) q_r^i([(x+y-s)y]_{-i}^i, (x+y-s)_{-i})\right)^2\right).$$

The result follows as the degree of each term in the square is less than or equal to $t+2$.

For the converse, we have

$$G_1(x,y,s) \equiv \sum_r q_r^2(x,y,s) \mod I_0[x,y,s]$$

for some polynomials $q_r \in \mathbb{R}[x,y,s]$ with $\deg(q_r) \leq t+2$. We remark that $G_1(x,y,x) = H_1(x,y)$. Furthermore, $I_0[x,y,x] = I_1[x,y]$. Thus, setting $s = x$ and renaming $q_r(x,y,x)$ to $\bar{q}_r(x,y)$ in the above equation gives us:

$$H_1(x,y) = G_1(x,y,x) \equiv \sum_r \bar{q}_r^2(x,y) \mod I_1[x,y],$$

where $\deg(\bar{q}_r) \leq t+2$. This gives us the result from Theorem 3. □

Proof of Theorem 6. For the implication, using Lemma EC.5,

$$F(x) - F(s) = \sum_{i=1}^n (x_i - s_i) \frac{\partial F([xs]^i)}{\partial x_i} \text{ and } F(y) - F(s) = \sum_{i=1}^n (y_i - s_i) \frac{\partial F([ys]^i)}{\partial x_i}.$$

Then as f is t-sos-submodular, from Theorem 1, (EC.14) holds, which enables us to write

$$\frac{\partial F([xs]^i)}{\partial x_i} \equiv \frac{\partial F(s)}{\partial x_i} - \sum_r q_{ri}(s,[xs]^i)^2 \mod I_1[s_{-i},[xs]^i_{-i}]$$
$$\frac{\partial F([ys]^i)}{\partial x_i} \equiv \frac{\partial F(y)}{\partial x_i} + \sum_r q_{ri}(s,[ys]^i)^2 \mod I_1[s_{-i},[ys]^i_{-i}].$$

for any $i \in \{1,\dots,n\}$. The conclusion follows in an identical way to the proof of Theorem 5.

For the converse, set, once again $s = x$ and note that $G_2(x,y,x) = H_2(x,y)$. The conclusion follows in an identical way to the proof of Theorem 5. □

Proof of Proposition 6. It is easy to see that Theorem 1 involves n sos constraints as opposed to Theorems 2–6 which involve just one. To establish the size of the semidefinite programs under consideration, we consider the number of monomials in (x,y,s) (resp. (x,y)) of degree at most t corresponding to the Gröbner basis for $I_0[x,y,s]$ (resp. $I_1[x,y]$).

For the case of $I_1[x,y]$, consider any monomial of degree $k \leq t$. It contains at most one of either x_i or y_i for fixed $i \in \{1,\dots,n\}$. Indeed, $x_i y_i \equiv x_i \mod I_1[x,y]$ and thus any appearance of x_i and y_i simultaneously will reduce to x_i. As a consequence, one needs to pick k indexes out of the n possible and choose either x_i or y_i for each index. This corresponds to 2 choices. Thus, there are $\binom{n}{k}2^k$ possible monomials of degree k. As we are interested in the number of monomials of degree less than or equal to t, we get $\sum_{k=0}^t \binom{n}{k}2^k$. This explains the size of the semidefinite programs involved in Theorems 1, 3, and 4.

For the case of $I_0[x,y,s]$, consider once again any monomial of degree $k \leq t$. As before, it contains at most one of either x_i, y_i or s_i for fixed $i \in \{1,\dots,n\}$. Indeed, modulo $I_0[x,y,s]$, $x_i y_i \equiv s_i$, $x_i s_i \equiv s_i$, and $y_i s_i \equiv s_i$. As a consequence, one needs to pick k indexes out of n and choose either

x_i, y_i or s_i for each index. This corresponds to 3 choices. Thus, there are $\binom{n}{k}3^k$ possible monomials of degree k and a total of $\sum_{k=0}^{t}\binom{n}{k}3^k$ monomials of degree at most t. This explains the size of the semidefinite programs involved in Theorems 2, 5, and 6. □

A.4. Proofs of Results in Section 3.3.

Proof of Proposition 7. Nonnegative weighted sums is immediate from the fact that the set of t-sos polynomials are a cone and that differentiation is a linear operation. In the remainder, we let F be the MLE of f and G be the MLE of g.

For complementation, we have $G(x) = F(\mathbf{1} - x)$ for all $x \in \{0,1\}^n$, where $\mathbf{1}$ is the vector of size n of all ones. Let $i, j \in \{1, \ldots, n\}$, we have:

$$-\frac{\partial G_(x)}{\partial x_i \partial x_j} = -\frac{\partial F(\mathbf{1}-x)}{\partial x_i \partial x_j}.$$

As a t-sos polynomial remains t-sos when composed with an affine mapping, the result holds.

For restriction, setting $a = 1_A$, we have $G(x) = F(a \circ x)$. Thus, for $i, j \in \{1, \ldots, n\}$, we have:

$$-\frac{\partial G_(x)}{\partial x_i \partial x_j} = -a_i a_j \frac{\partial F(a \circ x)}{\partial x_i \partial x_j}.$$

Note that $a_i a_j \geq 0$ as $a_i, a_j \in \{0,1\}$. The result holds for the same reason as complementation.

For contraction, setting $a = 1_A$, we have $G(x) = F(a + x - a \circ x) - F(a)$. Thus, for $i, j \in \{1, \ldots, n\}$, we have:

$$-\frac{\partial G_(x)}{\partial x_i \partial x_j} = -(1-a_i)(1-a_j) \frac{\partial F(a \circ x)}{\partial x_i \partial x_j}.$$

Note that $(1-a_i)(1-a_j) \geq 0$ and the result follows similarly to restriction. □

The following lemmas are needed to show Proposition 8.

LEMMA EC.6. *Let $w_1 = w_2 = \ldots = w_n = 2$ and let $B = 2n-1$. Consider the set function $f(S) = \min\{B, \sum_{i\in S} w_i\}$. Its multilinear extension F is given by $F(x) = \sum_{i=1}^n w_i x_i - \prod_{i=1}^n x_i$.*

We use the notation from Section 2.2 in the proof. Namely, we write

$$F(x) = \sum_{T\subseteq\Omega} a(T) \prod_{i\in T} x_i, \text{ where } a(T) = \sum_{S\subseteq T} (-1)^{t-s} f(S), \forall T \subset \Omega.$$

Proof. Let $\tilde{B} = 2n$ and let $\tilde{f}(S) = \min\{\tilde{B}, \sum_{i\in S} w_i\} = \sum_{i\in S} w_i$. Let $\tilde{F}$ be the MLE of $\tilde{f}$ with coefficients $\tilde{a}_T$ for $T \subseteq \Omega$. It is quite straightforward to see that $\tilde{F}(x) = \sum_{i=1}^n w_i x_i$. Now, note that f and $\tilde{f}$ coincide on all sets S with $|S| \leq n-1$ as $\min\{B, \sum_{i\in S} w_i\} = \sum_{i\in S} w_i$ for any $|S| < n$. However, $f(\Omega) = 2n-1$ and $\tilde{f}(\Omega) = 2n$. It follows that

$$a(T) = \sum_{S\subseteq T} (-1)^{t-s} f(S) = \sum_{S\subseteq T} (-1)^{t-s} \tilde{f}(S) = \tilde{a}(T), \forall |T| < n.$$

For $T = \Omega$, we then have

$$\begin{aligned} a(\Omega) &= \sum_{T\subseteq\Omega} (-1)^{n-|T|} f(T) = \sum_{T\subseteq\Omega, |T|\le n-1} (-1)^{n-|T|} f(T) + (-1)^0 f(\Omega) \\ &= \sum_{T\subseteq\{1,\ldots,n\}, |T|\le n-1} (-1)^{n-|T|} \tilde{f}(T) + \tilde{f}(\Omega) - \tilde{f}(\Omega) + f(\Omega) \\ &= \tilde{a}(\Omega) - 2n + 2n - 1 = -1, \end{aligned}$$

which gives us the result. □

Lemma EC.7. *For $n > 1$, let $f : 2^\Omega \to \mathbb{R}$ be the set function with MLE:*

$$F(x) = 1 + \frac{n-1}{n} \sum_{i=1}^{n} x_i - \frac{2}{n-1} \sum_{1\le i<j\le n} x_i x_j.$$

We have that $F(0) = 1, F(1) = 0$, and $F(x) > 1$ for all $x \neq 0, 1$.

Proof. It is straightforward to see that $F(0) = 1$ and $F(1) = 1 + \frac{n-1}{n} \cdot n - \frac{2}{n-1} \cdot \frac{n(n-1)}{2} = 0$. Now, let $x \neq 0, 1$ and suppose that x has k ones and $n-k$ zeros. We have that $\sum_{i=1}^n x_i = k$ and $\sum_{1\le i<j\le n} x_i x_j = \binom{k}{2}$, thus:

$$F(x) = 1 + \frac{n-1}{n} \cdot k - \frac{2}{n-1} \frac{k(k-1)}{2} = 1 + k\left(\frac{n-1}{n} - \frac{k-1}{n-1}\right).$$

As $n-1 \ge k$, it can be shown that $\frac{n-1}{n} - \frac{k-1}{n-1} > 0$, and so $F(x) > 1$ for all $x \neq 0, 1$. □

Lemma EC.8. *For $n > 1$, let $f : 2^\Omega \to \mathbb{R}$ be the set function with MLE:*

$$F(x) = 2\sum_{i=1}^{n} x_i - \frac{4n-2}{n(n-1)} \sum_{1\le i<j\le n} x_i x_j.$$

We have that $F(0) = 0, F(1) = 1$, and $F(x) > 1$ for all $x \neq 0, 1$.

Proof. It is straightforward to see that $F(0) = 0$ and $F(1) = 2n - \frac{4n-2}{n(n-1)} \cdot \frac{n(n-1)}{2} = 1$. Now let $x \neq 0, 1$ and suppose that x has k ones as above. As before, we obtain

$$F(x) = 2k - \frac{4n-2}{n(n-1)} \frac{k(k-1)}{2}.$$

We define $h(k) = 2k - \frac{4n-2}{n(n-1)} \frac{k(k-1)}{2}$ for $1 \le k \le n-1$ and show that $h(k) > 1$ for all k. This implies that $F(x) > 1$ for all $x \neq 0, 1$. Note first that h is concave in k by, e.g., considering its second derivative. It follows that its minimum is at $k = 1$ or $k = n-1$. As $h(1) = 2 > 1$ and

$$h(n-1) = 2(n-1) - \frac{4n-2}{n(n-1)} \cdot \frac{(n-1)(n-2)}{2} = 3 - \frac{2}{n} > 1,$$

we obtain our result. □

LEMMA EC.9. *Let $F(x) = x_1x_2\ldots x_n$ for all $x \in \{0,1\}^n$. We have that F is n-sos modulo $I_2[x]$ but not $(n-1)$-sos modulo $I_2[x]$.*

Proof. To see that F is n-sos modulo $I_2[x]$, simply note that $F(x) \equiv (x_1x_2\ldots x_n)^2$ modulo $I_2[x]$. To show that F is not $(n-1)$-sos mod $I_2[x]$, we proceed by contradiction. Suppose that $F(x) \equiv \sum_r q_r(x)^2 \mod I_2[x]$ for some polynomials q_r with degree strictly less than n. As $F(x) = 0$ for any $x \neq 1$, it follows that $q_r(x) = 0$ for any $x \neq 1$ and all coefficients of q_r involving monomials containing less than or equal to $n-1$ variables are equal to zero; see, e.g., (Grabisch et al. 2000). As $q_r(x)$ does not contain the monomial $x_1x_2\ldots x_n$ (otherwise it would be of degree n), this implies that $q_r(x) = 0$ for all x. But this contradicts the fact that $F(x) = 1$ when $x = 1$. □

Proof of Proposition 8. We start first with partial minimization. For $n \in \mathbb{N}, B = 2n-1$, and $w_1 = \ldots = w_n = 2$, let $f : 2^{n+1} \to \mathbb{R}$ be the set function with MLE:

$$F : (x,z) \in \{0,1\}^n \times \{0,1\} \mapsto Bz + \left(\sum_{i=1}^n w_i x_i\right) \cdot (1-z)$$

and let $g : 2^n \to \mathbb{R}$ be the set function with MLE: $G : x \in \{0,1\}^n \mapsto \min_{z\in\{0,1\}} G(x,z)$. It is straightforward to check that f is 0-sos-submodular by noting that $-\partial F(x,z)/\partial x_i \partial x_j = 0$, for any $i \neq j$, and $-\partial F(x,z)/\partial x_i \partial z = w_i = \sqrt{2}^2$ for any i. Now, note that $g(x) = \min\{B, \sum_i w_i x_i\}$. Indeed if $B \leq \sum_i w_i x_i$ then $z = 1$ and $G(x) = B$ and conversely, if $B \geq \sum_i w_i x_i$ then $z = 0$ and $G(x) = \sum_i w_i x_i$. Thus following Lemma EC.6, we have that $G(x) = \sum_{i=1}^n w_i x_i - \prod_{i=1}^n x_i$ and $-\frac{\partial^2 G(x)}{\partial x_i \partial x_j} = \prod_{k\neq i,j} x_k$. The result follows from Lemma EC.9.

We now consider convolution. Let f be as defined in Lemma EC.7 and take $u = 0$. It is easy to see from Lemma EC.7 that $g(S) = \min_{T\subseteq S} f(T)$ is such that $g(S) = 1$ for any $S \neq \Omega$ and $g(\Omega) = 0$. Thus, its MLE is equal to

$$G(x) = \sum_{S\neq\Omega}\left(\prod_{i\in S} x_i \prod_{i\notin S}(1-x_i)\right) = \sum_{S}\left(\prod_{i\in S} x_i \prod_{i\notin S}(1-x_i)\right) - \prod_{i=1}^n x_i = 1 - \prod_{i=1}^n x_i.$$

Taking second-order derivatives, we obtain:

$$-\frac{\partial^2 G(x)}{\partial x_i \partial x_j} = \prod_{k=1,k\neq i,j}^n x_k.$$

The result follows from Lemma EC.9.

We finish with monotonization. Let f be a set function as defined in Lemma EC.8. It is easy to see that f is 0-sos-submodular as its second-order derivatives are negative numbers. Now define

$g(S) = \min_{T \supseteq S} f(T)$ for any $S \in \Omega$. It is easy to see from Lemma EC.8 that $g(\emptyset) = 0$ but that $g(S) = F(1) = 1$ for any set $S \neq \emptyset$. Thus, its MLE is equal to:

$$G(x) = \sum_{S \neq \emptyset} \left(\prod_{i \in S} x_i \prod_{i \notin S} (1 - x_i) \right) = \sum_{S} \left(\prod_{i \in S} x_i \prod_{i \notin S} (1 - x_i) \right) - \prod_{i=1}^{n} (1 - x_i) = 1 - \prod_{i=1}^{n} (1 - x_i).$$

Taking second-order derivatives, we obtain:

$$-\frac{\partial^2 G(x)}{\partial x_i \partial x_j} = \prod_{k=1, k \neq i,j}^{n} (1 - x_k).$$

Note that for any polynomial p over $\{0,1\}^n$, $p(x)$ is t-sos modulo $I_2[x]$ if and only if $p(1-x)$ is t-sos modulo $I_2[x]$. As $-\frac{\partial^2 G(1-x)}{\partial x_i \partial x_j} = \prod_{k \neq i,j} x_i$, the result follows from Lemma EC.9. □

A.5. Proofs of Results in Section 3.4.

Proof of Proposition 9. Let $f : 2^{\Omega} \to \mathbb{R}$ be a set function of degree d and let F be its MLE. When $d = 0$, F is a constant, and it is straightforward to check that, e.g., (i') in Proposition 1 holds. Thus, f is always submodular. Likewise, when $d = 1$, F is an affine function, that is $F(x) = a^T x + b$ where $a \in \mathbb{R}^n, b \in \mathbb{R}$. It is once again straightforward to check that (i') in Proposition 1 holds. Indeed, we have

$$F(x) + F(y) - F(x + y - x \circ y) - F(x \circ y) = a^T x + b + a^T y + b - a^T (x + y - x \circ y) - b - a^T (x \circ y) - b = 0.$$

Thus, f is also submodular (in fact, it is modular).

Now, let $d = 2$ and suppose that f is submodular. It follows from Proposition 1 (iii') that $F_2^{ij}(x_{-i,j}) := -\frac{\partial^2 F(x)}{\partial x_i \partial x_j} \geq 0$ for all $i, j \in \{1, \ldots, n\}$. As $\deg(f) = 2$, F_2^{ij} is in fact equal to a nonnegative constant C^{ij}. Hence, we can write

$$F_2^{ij}(x_{-i,j}) \equiv \left(\sqrt{C^{ij}} \right)^2 \mod I_2[x_{-i,j}],$$

which shows that f is 0-sos-submodular.

Finally, let $d = 3$. Billionnet and Minoux (1985) propose the following characterization of cubic submodular functions in Lemma 1: a cubic submodular set function f is submodular if and only if its MLE F can be written:

$$F(x) = \sum_{j \in J^-} c_j x_j - \sum_{j \in J^+} c_j x_j - \sum_{i,j \in IJ} c_{ij} x_i x_j - \sum_{(i,j,k) \in IJK^+} c_{ijk} x_i x_j x_+ \sum_{(i,j,k) \in IJK^-} c_{ijk} x_i x_j x_k + K,$$

where $c_j \geq 0, \forall j \in J^-$, $c_j \geq 0, \forall j \in J^+$, $c_{ij} \geq 0, \forall (i,j) \in IJ$, $c_{ijk} \geq 0, \forall (i,j,k) \in IJK^+ \cup IJK^-$, and $c_{ij} \geq \sum_{k \mid (i,j,k) \in IJK^-} c_{ijk}, \forall (i,j) \in \{1,\ldots,n\}^2, i<j$. Fix $i,j \in \{1,\ldots,n\}$, we have that:

$$
\begin{aligned}
-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} &= c_{ij} + \sum_{k \mid (i,j,k) \in IJK^+} c_{ijk}x_k - \sum_{k \mid (i,j,k) \in IJK^-} c_{ijk}x_k \\
&= c_{ij} - \sum_{k \mid (i,j,k) \in IJK^-} c_{ijk} + \sum_{k \mid (i,j,k) \in IJK^-} c_{ijk} + \sum_{k \mid (i,j,k) \in IJK^+} c_{ijk}x_k - \sum_{k \mid (i,j,k) \in IJK^-} c_{ijk}x_k \\
&\equiv \left(\sqrt{c_{ij} - \sum_{k \mid (i,j,k) \in IJK^-} c_{ijk}}\right)^2 + \sum_{k \mid (i,j,k) \in IJK^+} \left(\sqrt{c_{ijk}}x_k\right)^2 + \sum_{k \mid (i,j,k) \in IJK^-} \left(\sqrt{c_{ijk}}(1-x_k)\right)^2,
\end{aligned}
$$

where the last equivalence is mod $I_2[x_{-i,j}]$ and uses the assumptions on the coefficients of F given by (Billionnet and Minoux 1985). This shows that f is 1-sos-submodular. □

The next lemma is implied by the the results in (Laurent 2003, Fawzi et al. 2015, Sakaue et al. 2016). We restate it for clarity. It is needed for the proof of Proposition 10.

Lemma EC.10 (**Laurent (2003), Fawzi et al. (2015), Sakaue et al. (2016)**). *Let*

$$
p(x) = \begin{cases} 2\sum_{1\leq i<j\leq n} x_i x_j - (n-1)\sum_{i=1}^n x_i + \frac{n^2-1}{4} \text{ if } n \text{ odd} \\ 2\sum_{1\leq i<j\leq n} x_i x_j - (n-2)\sum_{i=1}^n x_i + \frac{n(n-2)}{4}, \text{ if } n \text{ even} \end{cases}, \quad \text{for } x \in \{0,1\}^n.
$$

We have that p is nonnegative (in fact $\lceil \frac{n}{2} \rceil$-sos when n is odd and $\lceil \frac{n+1}{2} \rceil$-sos when n is even) over $I_2[x]$ but not $(\lceil \frac{n}{2} \rceil - 1)$-sos over $I_2[x]$.

Proof. Suppose first that n is odd. Let $I[y] = \langle y_1^2 - 1, \ldots, y_n^2 - 1\rangle$ and for $y \in \{-1,1\}^n$, define

$$
q_{odd}(y) = \sum_{1\leq i<j\leq n} y_i y_j + \left\lfloor \frac{n}{2} \right\rfloor.
$$

Note that $\lfloor \frac{n}{2} \rfloor < \frac{n}{2}$ and thus $-\lfloor \frac{n}{2} \rfloor > -\frac{n}{2}$. From (Laurent 2003, Section 3), it follows that $q(y)$ is not $\lceil \frac{n}{2} \rceil - 1$-sos over $I[y]$. Further note that:

$$
q_{odd}(y) \equiv \frac{1}{2}\left(\sum_{i=1}^n y_i\right)^2 - \frac{n}{2} + \left\lfloor \frac{n}{2} \right\rfloor \quad \text{mod } I[y].
$$

As n is odd, $q_{odd}(y) = \frac{1}{2}\left(\sum_{i=1}^n y_i^2\right) - \frac{1}{2}$. As $\left(\sum_{i=1}^n y_i^2\right) \geq 1$ when $y \in \{-1,1\}^n$, it follows that $q_{odd}(y) \geq 0$ for all $y \in \{-1,1\}^n$. Thus, from (Fawzi et al. 2015), as q_{odd} is a nonnegative form over $\{-1,1\}^n$ and quadratic, $q_{odd}(y)$ is $\lceil \frac{n}{2} \rceil$-sos.

Suppose now that n is even and define

$$q_{even}(y) = \sum_{1\le i\le j\le n} y_i y_j + \sum_{i=1}^{n} y_i + \frac{n}{2}.$$

This polynomial is obtained from q_{odd} by considering the contraction of the complete graph over $n+1$ nodes to the complete graph over n nodes. Following (Laurent 2003, Proposition 3 (i)), we obtain that $q(y)$ is not $\lceil \frac{n}{2} \rceil - 1$-sos over $I[y]$. Further note that:

$$q_{even}(y) \equiv \frac{1}{2}\left(1 + \sum_{i=1}^{n} y_i\right)^2 - 1/2 \mod I[y].$$

For n even, as $\left(\sum_{i=1}^{n} 1 + y_i^2\right) \ge 1$ when $y \in \{-1,1\}^n$, it follows that $q_{even}(y) \ge 0$ for all $y \in \{-1,1\}^n$. Thus, from Sakaue et al. (2016), as q_{even} is a quadratic function, $q_{even}(y)$ is $\lceil \frac{n+1}{2} \rceil$-sos.

To get the result, note that if $x \in \{0,1\}^n$, then $2x - 1 \in \{-1,1\}^n$. Thus, as affine mappings maintain t-sos, the result holds. □

Proof of Proposition 10. We show the case when n is odd. The case where n is even is identical. We compute the second-order partial derivatives of F. For any $i, j \in \{1, \ldots, n-2\}, i \ne j$, we have

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j} = 2x_{n-1}x_n \equiv 2x_{n-1}^2 x_n^2 \mod I_2[x_{-i,j}].$$

For any $i \in \{1, \ldots, n\}$, we have:

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_{n-1}} = \left(-(n-3) + 2\sum_{j=1, j\ne i}^{n-2} x_j\right) x_n + (n-3) \equiv (n-3)(1-x_n)^2 + 2\sum_{j=1, j\ne i}^{n-2} x_j^2 x_n^2 \mod I_2[x_{-i,n-1}]$$

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_n} = \left(-(n-3) + 2\sum_{j=1, j\ne i}^{n-2} x_j\right) x_{n-1} + (n-3) \equiv (n-3)(1-x_{n-1})^2 + 2\sum_{j=1, j\ne i}^{n-2} x_j^2 x_{n-1}^2 \mod I_2[x_{-i,n}].$$

Finally, we have:

$$-\frac{\partial^2 F(x)}{\partial x_{n-1} \partial x_n} = 2\sum_{1\le i<j\le n-2} x_i x_j + \frac{n^2 - 4n + 3}{4} - (n-3)\sum_{i=1}^{n-2} x_i.$$

Note that $-\frac{\partial^2 F(x)}{\partial x_i \partial x_j}$ is 2-sos for any $i, j \in \{1, \ldots, n-2\}$, as is $-\frac{\partial^2 F(x)}{\partial x_i \partial x_{n-1}}$ and $-\frac{\partial^2 F(x)}{\partial x_i \partial x_n}$ for any $i \in \{1, \ldots, n\}$. However, following Lemma EC.10, $-\frac{\partial^2 F(x)}{\partial x_{n-1} \partial x_n}$ is $\lceil \frac{n-2}{2} \rceil$-sos over $I_2[x_{-n-1,n}]$ but not $\left(\lceil \frac{n-2}{2} \rceil - 1\right)$-sos over $I_2[x_{-n-1,n}]$. As $n \ge 4$ (given that $d \ge 4$), it follows that F is $\lceil \frac{n-2}{2} \rceil$-sos-submodular but not $\left(\lceil \frac{n-2}{2} \rceil - 1\right)$-sos-submodular. □

Proof of Proposition 11. Note that the function

$$F(x)=\sum_{e\in E} w_e\left[\left(1-\prod_{v\in e} x_v\right)-\prod_{v\in e}(1-x_v)\right]$$

is the multilinear extension of the weighted cut function on $\mathcal{H}$ and that if $\max_{e\in E}|e|=t$, then $\deg(F)=t$. Let $E_{ij}=\{e\in E : i,j\in e\}$. We have

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j}=\sum_{e\in E_{ij}} w_e\left[\prod_{v\in e, v\neq i,j} x_v+\prod_{v\in e, v\neq i,j}(1-x_v)\right].$$

Let $F_e^{ij}(x)=\prod_{v\in e, v\neq i,j} x_v+\prod_{v\in e, v\neq i,j}(1-x_v)$. We show that $F_e^{ij}(x)$ is $2\lfloor t/2-1\rfloor$-sos. If t is even, then it is easy to see that $\deg(F_e^{ij})\leq t-2$ and when t is odd, degree cancellations lead to $\deg(F_e^{ij})\leq t-3$. Furthermore, as F_e^{ij} takes on values 0 or 1 for all $x\in\{0,1\}^n$, we have that $F_e^{ij}(x)=\left(F_e^{ij}(x)\right)^2$ for all $x\in\{0,1\}^n$. As $w_e\geq 0, \forall e$, this shows that F is $(t-2)$-sos-submodular when t is even and $(t-3)$-sos-submodular when t is odd, i.e., $2\lfloor t/2-1\rfloor$-sos-submodular. □

Proof of Proposition 12. The multilinear extension of f can be written as

$$F(x)=\sum_{\ell=1}^{m}\left(1-\prod_{k:\ell\in A_k}(1-x_k)\right).$$

Note that the degree of F is equal to the maximum number of sets any element of Ω can belong to. Suppose $t=\max_{\ell=1,\ldots,m}|\{k:\ell\in A_k\}|$. Then we have

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j}=\sum_{\ell=1}^{m}\prod_{k:\ell\in A_i\cap A_j\cap A_k}(1-x_k),$$

which is a multilinear function of degree $t-2$. Since the above is equal to either 0 or 1 for $x\in\{0,1\}^n$, using arguments identical to above, the coverage function is $t-2$-sos-submodular. □

Proof of Proposition 13. Let F be the MLE of f. We have that:

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j}\equiv \phi(\sum_{k\neq i,j} x_k+1)+\phi(\sum_{k\neq i,j} x_k+1)-\phi(\sum_{k\neq i,j} x_k+2)-\phi(\sum_{k\neq i,j} x_k) \mod I_2[x_{-i,j}].$$

As ϕ is concave of degree $2d$ and univariate, this implies that $-\phi''\geq 0$, of degree $2d-2$, and univariate. It then follows from e.g. Blekherman et al. (2012) that $-\phi''$ is sos, and so $g(s,t)=\phi(\frac{1}{2}s+\frac{1}{2}t)-\frac{1}{2}\phi(s)-\frac{1}{2}\phi(t)=\sum_r(q_r(t,s))^2$, where q_r is a polynomial in $s,t\in\mathbb{R}^2$ of degree less than or equal to d (Ahmadi and Parrilo 2013). Taking $s=\sum_{k\neq i,j} x_k+2$ and $t=\sum_{j\neq i,j} x_k$ in the previous expression, we obtain that

$$-\frac{\partial^2 F(x)}{\partial x_i \partial x_j}\equiv\sum_r q_r^2\left(\sum_{k\neq i,j} x_k+2, \sum_{k\neq i,j} x_k\right) \mod I_2[x_{-i,j}]$$

and thus f is d-sos-submodular. □

Appendix B: Proof of Results in Section 4

B.1. Proof of Results in Section 4.2

We start with the proof of Proposition 15. To show this proof, we use the equivalent definition of the submodularity ratio $\gamma^*(f)$ as given in (Bian et al. 2017, Definition 1). As defined there, $\gamma^*(f)$ is the largest scalar γ such that:

$$\sum_{\omega \in L \backslash S} (f(S \cup \{\omega\}) - f(S)) \geq \gamma(f(L \cup S) - f(S)), \forall L, S \subseteq \{1, \ldots, n\}. \tag{EC.17}$$

Proof of Proposition 15. We show that $\gamma^*(f)$ is also the largest γ such that:

$$\sum_{i=1}^{n} (y_i - x_i) \frac{\partial F(x)}{\partial x_i} + \gamma(F(x) - F(y)) \geq 0, \forall x, y \in \{0,1\}^n, x \leq y. \tag{EC.18}$$

Let $x, y \in \{0,1\}^n$ such that $x \leq y$ and suppose that (EC.17) holds. We set S be the set such that $x = 1_S$. We further set L to be the set such that $y - x = 1_L$ (note that $x \leq y$ and so $y - x \in \{0,1\}^n$). We have that $L \cap S = \emptyset$ and that $y = 1_{L \cup S}$. Recalling that $F(1_S) = f(S)$ for any $S \in \{1, \ldots, n\}$, (EC.17) gives us:

$$\sum_{\omega \in L} (F(1_{S \cup \{\omega\}}) - F(x)) \geq \gamma(F(y) - F(x)).$$

Now, as $\omega \in L$ if and only if there exists $i \in \{1, \ldots, n\}$ such that $y_i - x_i = 1$, i.e., $y_i = 1$ and $x_i = 0$, we get:

$$\sum_{i=1}^{n} (y_i - x_i)(F(x + e_i) - F(x)) \geq \gamma(F(y) - F(x)),$$

which is equivalent from Lemma EC.1 to (EC.18).

Conversely, suppose now that (EC.18) holds for any $x, y \in \{0,1\}^n$ and $x \leq y$. We show that (EC.18) holds. Let $S, L \subseteq \{1, \ldots, n\}$. We let $y = 1_{L \cup S}$ and $x = 1_S$. Note that $(y_i - x_i) = 1$ for $i \in \{1, \ldots, n\}$ if and only if $i \in L \backslash S$. Thus (EC.18) gives us:

$$\sum_{i \in L \backslash S} (F(x + e_i) - F(x)) + \gamma(f(S) - f(S \cup L)) \geq 0$$

which is equivalent to $\sum_{i \in L \backslash S} (f(S \cup \{i\}) - f(S)) \geq \gamma(f(S \cup L) - f(S))$, which is none other than (EC.17). Our conclusion follows. □

Proof of Proposition 16. Note that we have: $F = F_k + T_k$, thus, by definition of m,

$$\frac{\partial F(x)}{\partial x_i} = \frac{\partial F_k(x)}{\partial x_i} + \frac{\partial T_k(x)}{\partial x_i} \geq \frac{\partial F_k(x)}{\partial x_i} + m, \forall x \in [0,1]^n.$$

Furthermore, by the triangle inequality,

$$|F(x) - F(y)| = |F_k(x) + T_k(x) - F_k(y) - T_k(y)| \le |F_k(x) - F_k(y)| + |T_k(x) - T_k(y)|.$$

Using the mean-value theorem and Hölder's theorem, it can be shown that $|T_k(x) - T_k(y)| \le M||x-y||_1, \forall x, y \in [0,1]^n$. Thus

$$|F(x) - F(y)| \le |F_k(x) - F_k(y)| + M||x-y||_1, \forall x, y \in [0,1]^n.$$

As $F(x) - F(y) \le 0$ for $x \le y$ and $\gamma \ge 0$, combining the bounds above, we have:

$$\begin{aligned}
&\sum_{i=1}^{n}(y_i - x_i)\frac{\partial F(x)}{\partial x_i} + \gamma(F(x) - F(y)) \\
&\ge \sum_{i=1}^{n}(y_i - x_i)\left(\frac{\partial F_k(x)}{\partial x_i} + m\right) + \gamma(F_k(x) - F_k(y)) - \gamma M||x-y||_1 \\
&= \sum_{i=1}^{n}(y_i - x_i)\left(\frac{\partial F_k(x)}{\partial x_i} + m - \gamma M\right) + \gamma(F_k(x) - F_k(y)), \forall x, y \in [0,1]^n, x \le y.
\end{aligned}$$

Thus, by solving (7), we obtain a lower bound on $\gamma^*(f)$. □

B.2. Proof of Results in Section 4.3

Proof of Proposition 17. We show the first claim of the proof, that is, for any irreducible submodular decomposition (g, h) of f, any other decomposition (g', h') that reduces to (g, h) results in a greater submodular approximation of f when using the same permutation π of Ω:

$$(g - \underline{\mathrm{h}}_\pi)(S) \le (g' - \underline{\mathrm{h}'_\pi})(S), \forall S \subseteq \Omega.$$

Since (g', h') can be reduced to (g, h), there exists a non-modular submodular p such that $g' = g + p$ and $h' = h + p$. For any $S \subseteq \Omega$ and any π such that $S^\pi_{|S|} = S$:

$$(g - \underline{\mathrm{h}}_\pi)(S) = (g' - p' - (\underline{\text{h' - p'}})_\pi)(S) = (g' - \underline{\mathrm{h}'_\pi})(S) + (\underline{\mathrm{p}}_\pi - p)(S) \le (g' - \underline{\mathrm{h}'_\pi})(S)$$

The second equality comes from linearity of the subgradient of any submodular function. The inequality stems from the fact that $\underline{\mathrm{p}}_\pi$ is a subgradient of π, and therefore $\underline{\mathrm{p}}_\pi \le p$.

We now show the second claim, which states that the above no longer holds when using two different permutations π_1 and π_2 to produce sub-gradients of h and h', via an example. Let $f = g - h$, where $g(S) = \sum_{i \in S} a_i$ and $h(S) = \sqrt{|S|}$, for all $S \subseteq \Omega$. Notice that it must be that (g, h) is irreducible, since g is modular. Now consider a reducible decomposition: $g' = g + \alpha h$ and $h' = (1+\alpha)h$. It is easy to see that $\underline{\mathrm{h}}_\pi(S) = \sum_i w_i^\pi$ with $w_i^\pi = \sqrt{\pi^{-1}(i)} - \sqrt{\pi^{-1}(i) - 1}$ and $\pi^{-1}(i)$ being equal to the position of element i in permutation π. Then:

$$(g - \underline{\mathrm{h}}_{\pi_1})(X) = \sum_{i \in S}(a_i - w_i^{\pi_1}), \text{ and } (g' - \underline{\mathrm{h}}'_{\pi_2})(X) = \sum_{i \in S}(a_i - (1+\alpha)w_i^{\pi_1}) + \alpha h(X).$$

Consider any permutation π_1 with element 1 in the last position, that is $\pi_1^{-1}(1) = n$ and π_2 with element 1 in its first position, $\pi_2^{-1}(1) = 1$. We have that

$$(g - \underline{\mathrm{h}}_{\pi_1})(\{1\}) = a_1 - \left(\sqrt{n} - \sqrt{n-1}\right), \text{ and } (g' - \underline{\mathrm{h}}'_{\pi_2})(\{1\}) = a_i - (1+\alpha)1 + \alpha = a_i - 1,$$

and so $(g - \underline{\mathrm{h}}_{\pi_1})(\{1\}) \geq (g' - \underline{\mathrm{h}}'_{\pi_2})(\{1\})\ \forall n \in \mathbb{N}$. □

Proof of Proposition 18. Suppose (G^*, H^*) is not an irreducible decomposition. Then there exists a non-modular submodular $P \in \mathbb{R}[x]$ such that $G := G^* - P$ and $H := H^* - P$, with $F = G - H$, and G and H submodular. By linearity of the second derivative, we have that:

$$\frac{\partial^2 H(x)}{\partial x_i \partial x_j} = \frac{\partial^2 H^*(x)}{\partial x_i \partial x_j} - \frac{\partial^2 P(x)}{\partial x_i \partial x_j}.$$

Further, due to submodularity, $\frac{\partial P(x)}{\partial x_i \partial x_j} \leq 0\ \forall i, j \in 1, \ldots, n$. In fact, there exists some $\bar{x} \in \{0,1\}^n, \bar{i}, \bar{j}$ such that $\frac{\partial P(\bar{x})}{\partial x_{\bar{i}} \partial x_{\bar{j}}} < 0$, since it is assumed that P is non-modular. Therefore

$$\sum_{x \in \{0,1\}^n} \sum_{i<j} \frac{\partial H(x)}{\partial x_i \partial x_j} = \sum_{x \in \{0,1\}^n} \sum_{i<j} \left(\frac{\partial H^*(x)}{\partial x_i \partial x_j} - \frac{\partial P(x)}{\partial x_i \partial x_j} \right) > \sum_{x \in \{0,1\}^n} \sum_{i<j} \frac{\partial H^*(x)}{\partial x_i \partial x_j},$$

which is a violation of our assumption that G^*, H^* is an optimal solution to (10). □

Proof of Proposition 19. We provide a reduction from $SUBMOD_4$ as defined in Proposition 2. Let q be a set function of degree 4 with MLE Q. We let $f = q$, or equivalently, $F = Q$ and take $k = 0$. We show that Q is submodular if and only if $F = G - H$ for submodular G and H and $\sum_{x \in \{0,1\}^n} \sum_{i<j} \frac{\partial^2 H(x)}{\partial x_i \partial x_j} \geq k = 0$.

If Q is submodular, then let $G = Q$ and $H = 0$ and the implication follows. Now suppose that there exist submodular G and H such that $Q = G - H$ with $\sum_{x \in \{0,1\}^n} \sum_{i<j} \frac{\partial^2 H(x)}{\partial x_i \partial x_j} \geq 0$. As H is submodular, we have that $\frac{\partial^2 H(x)}{\partial x_i \partial x_j} \leq 0, \forall x \in \{0,1\}^n, \forall i \neq j$, following Proposition 1 (iii'). Combining this with the assumption, we get that

$$\frac{\partial^2 H(x)}{\partial x_i \partial x_j} = 0, \forall x \in \{0,1\}^n, \forall i \neq j.$$

This implies that H is modular. Thus,

$$\frac{\partial^2 Q(x)}{\partial x_i \partial x_j} = \frac{\partial^2 G(x)}{\partial x_i \partial x_j} \geq 0, \forall x \in \{0,1\}^n, \forall i \neq j,$$

where the equality follows from the fact that $Q = G - H$ and the inequality follows from the fact that G is submodular. We obtain in consequence of Proposition 1 (iii') that Q is submodular and the converse follows. □

Appendix C: Details for Numerical Results

In this section, we provide additional details on the experimental set-up and numerical results for the experiments presented in Section 4. All experiments were executed on the ISyE High-Performance Computing cluster at Georgia Tech. We use Mosek (MOSEK ApS 2024) as our solver, JuMP (Dunning et al. 2017), and the SumOfSquares.jl library (Legat 2020). All code can be found here.

C.1. Details for Section 4.1

C.1.1. Details on regression methods. Given training data $\{x_i, y_i\}_{i=1}^m$, we solve the following problems to obtain a polynomial regressor, a t-sos-submodular regressor, and a polynomial regressor as in (Stobbe and Krause 2012), respectively:

$$F^k_{poly} = \underset{F\in\mathbb{R}[x],\deg(F)\leq k}{\arg\min} \sum_{i=1}^m (y_i - F(x_i))^2 + \lambda \sum_{T\subseteq\Omega,|T|\leq k} a(T)^2, \tag{EC.19}$$

$$\begin{aligned} F^k_{t-sos-submod} = \underset{F\in\mathbb{R}[x],\deg(F)\leq k}{\arg\min} & \sum_{i=1}^m (y_i - F(x_i))^2 + \lambda \sum_{T\subseteq\Omega,|T|\leq k} a(T)^2 \\ \text{s.t. } & F \text{ is } t\text{-sos submodular}\end{aligned} \tag{EC.20}$$

$$\begin{aligned} F^k_{necessary-submod} = \underset{F\in\mathbb{R}[x],\deg(F)\leq k}{\arg\min} & \sum_{i=1}^m (y_i - F(x_i))^2 + \lambda \sum_{T\subseteq\Omega,|T|\leq k} a(T)^2, \\ \text{s.t. } & a(\{i,j\}) + |a(T)| \leq 0,\ \forall \{i,j\} \subset T, \forall T \in \Omega, |T| \leq k,\end{aligned} \tag{EC.21}$$

The second term in the objective functions is equivalent to a ridge regression penalty on the coefficients of the polynomial fit to the data. We solve the training problem for $\lambda \in \{0.0, 10^{-4}, 10^{-3}, 10^{-2}\}$, and select the best performing λ using the validation set. This regularization is needed to avoid computational errors for both (EC.19) and (EC.21). We add it to (EC.20) too to ensure an apples-to-apples comparison. As explained in Section 4.1, the constraints appearing in (EC.21) are sufficient for submodularity when $k \in \{2, 3\}$ but are no longer sufficient when $k \geq 4$, in line with Proposition 2.

FlexSubNet fits a monotone submodular function $F_{FlexSubNet} = F^N(S;\theta)$ with the following recursive form: $F^n(S;\theta) = \phi_\theta(\alpha F^{n-1}(S;\theta) + (1-\alpha)m^n_\theta(S)), n = 1,\dots,N,$ and $F^0(S;\theta) = m_\theta(S)$, where m^n_θ and ϕ are modular and concave functions that are expressed by neural networks parameterized by θ. It learns θ via stochastic gradient descent on the problem $\min_\theta \sum_{i=1}^m (y_i - F^N(x_i;\theta))^2$.

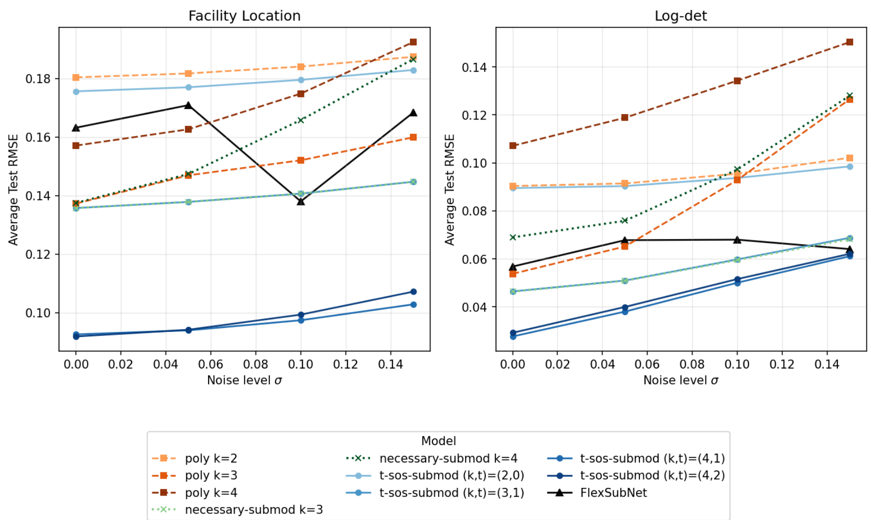

Figure EC.1 Comparison of root mean squared test error for various methods, fitted to noisy samples of facility location (right) and submodular log-determinant (left) functions.

We set the hyperparameters for this model as follows: batch size to three, number of hidden layers of the neural networks to four, learning rate to 0.005, and we train for 100 epochs. All remaining hyperparameters are set to their default as stated in the original paper and code (De and Chakrabarti 2022). Since the resulting model is sensitive to random initialization, we retrain with five different random initializations and select the best one using the validation set.

C.1.2. Additional experiments. We show submodular regression results for two additional synthetic set functions following (De and Chakrabarti 2022). We generate a log-determinantal function $F_{LD}(S) = \log\det(I + \sum_{s\in S} z_s z_s^\top)$, and a facility location function $F_{FL}(S) = \sum_{i=1}^n \max_{s\in S} \frac{z_i^\top z_s}{\|z_i\|\|z_s\|}$, with $z_i \sim \text{unif}[0,1]^{10}$, $i \in 1,\dots,n$. We set $n = 15$, and vary additive Gaussian noise as outlined in Section 4.1. The test performance of each method, averaged over five random problem instances, is presented in Figure EC.1 for the log-determinantal and the facility location functions.

C.1.3. Runtime. We report the average time to solve for F^k_{poly}, $F^k_{t-sos-submod}$, $F_{FlexSubNet}$ and $F^k_{necessary-submod}$ for $n = 15$ while varying k, t in Table EC.1.

Regression method	Polynomial (k)			t-sos-submodular (k,t)				FlexSubNet	Nec.-submod. (k)	
Parameters	2	3	4	$(2,0)$	$(3,1)$	$(4,1)$	$(4,2)$	–	3	4
Average runtimes (s)	5.56	5.90	7.30	18.26	20.07	32.68	5120.74	59.76	6.27	17.47

Table EC.1 Average training times for F^k_{poly}, $F^k_{t\text{-sos-submod}}$, $F_{FlexSubNet}$ and $F^k_{necessary-submod}$ on synthetic data.

C.2. Details for Section 4.2

We report the average time to solve for $\gamma_{trunc,sos}^{t,k}$ via (7) for $n = 10$ while varying k, t in Table EC.2.

(k,t)	(1,1)	(2,2)	(3,2)	(4,3)
Average runtime (s)	1.22	12.5	13.72	7996

Table EC.2 Average time to solve (7) for $n = 10$.

C.3. Details for Section 4.3

For ds decomposition, we seek to max $\sum_{x\in\{0,1\}^n}\sum_{i<k}\frac{\partial^2 H(x)}{\partial x_i \partial x_j}$, which we show how to write as a linear function in terms of the MLE coefficients $a(T)$:

$$\sum_{x\in\{0,1\}^n}\sum_{i<j}\frac{\partial^2 H(x)}{\partial x_i\partial x_j} = \sum_{x\in\{0,1\}^n}\sum_{i<j}\sum_{T\supseteq\{i,j\}} a(T)\prod_{k\in T\setminus\{i,j\}} x_k = \sum_{i<j}\sum_{T\supseteq\{i,j\}} a(T)\sum_{x\in\{0,1\}^n}\prod_{k\in T\setminus\{i,j\}} x_k$$
$$= \sum_{i<j}\sum_{T\supseteq\{i,j\}} a(T)2^{n-(|T|-2)} = \sum_{T\subseteq\Omega:|T|\geq 2} a(T)\binom{|T|}{2}2^{n-(|T|-2)}.$$

Table 3 shows the average optimality gap achieved over ten runs of the submodular-supermodular procedure for ten randomly generated set functions for each $n \in \{10, 15, 20\}$. Here we provide the disaggregated results, showing the average optimality gap attained for each individual randomly generated set function in Figure EC.2. These results show that not only does the t-sos-submodular-irreducible decomposition dominate on average, but also dominates each instance individually, showcasing the potential for drastic improvements. In Table EC.3 we report the average time it takes to solve (11) in order to obtain a t-sos-submodular irreducible decomposition.

	$n = 10$	$n = 15$	$n = 20$
Average runtime (s)	49.2	1,493	28,476

Table EC.3 Average time to solve (11) with $d = 4$, $t = 2$ for randomly generated set functions.

E-Companion References

Ahmadi AA, Parrilo PA (2013) A complete characterization of the gap between convexity and sos-convexity. *SIAM Journal on Optimization* 23(2):811–833.

Bian AA, Buhmann JM, Krause A, Tschiatschek S (2017) Guarantees for greedy maximization of non-submodular functions with applications. *International Conference on Machine Learning*, 498–507 (PMLR).

Billionnet A, Minoux M (1985) Maximizing a supermodular pseudoboolean function: A polynomial algorithm for supermodular cubic functions. *Discrete Applied Mathematics* 12(1):1–11.

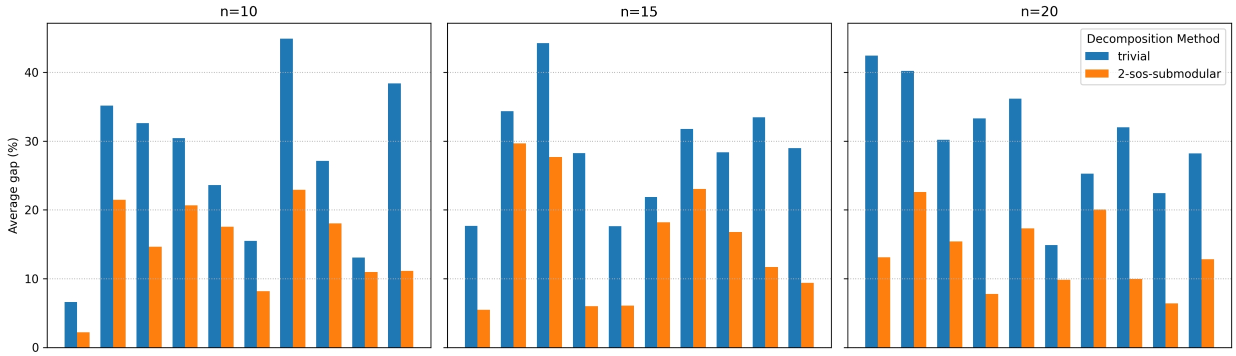

Figure EC.2 Comparison of average relative optimality gap obtained for ten runs of the submodular-supermodular procedure for the trivial decomposition and 2-sos-submodular-irreducible decomposition for each randomly generated set function in Section 4.3

Blekherman G, Parrilo PA, Thomas RR (2012) *Semidefinite Optimization and Convex Algebraic Geometry* (SIAM).

De A, Chakrabarti S (2022) Neural estimation of submodular functions with applications to differentiable subset selection. *Advances in Neural Information Processing Systems* 35:19537–19552.

Dunning I, Huchette J, Lubin M (2017) Jump: A modeling language for mathematical optimization. *SIAM Review* 59(2):295–320.

Fawzi H, Saunderson J, Parrilo PA (2015) Sparse sum-of-squares certificates on finite abelian groups. *2015 54th IEEE Conference on Decision and Control (CDC)*, 5909–5914 (IEEE).

Grabisch M, Marichal JL, Roubens M (2000) Equivalent representations of set functions. *Mathematics of Operations Research* 25(2):157–178.

Laurent M (2003) Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. *Mathematics of operations research* 28(4):871–883.

Laurent M (2009) Sums of squares, moment matrices and optimization over polynomials. *Emerging applications of algebraic geometry*, 157–270 (Springer).

Legat B (2020) Sumofsquares.jl: Sum-of-squares optimization in julia. `https://github.com/jump-dev/SumOfSquares.jl`.

MOSEK ApS (2024) The mosek optimization toolbox for matlab manual. version 10.0. `https://docs.mosek.com/`.

Permenter F, Parrilo PA (2012) Selecting a monomial basis for sums of squares programming over a quotient ring. *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 1871–1876 (IEEE).

Sakaue S, Takeda A, Kim S, Ito N (2016) Exact SDP relaxations with truncated moment matrix for binary polynomial optimization problems. *Proceedings of the 5th International Conference on Continuous Optimization ICCOPT*.

Stobbe P, Krause A (2012) Learning Fourier sparse set functions. *Artificial Intelligence and Statistics*, 1125–1133 (PMLR).