Sums of squares in polynomial binary optimization

Greg Blekherman [1] João Gouveia [2] James Pfeiffer [3]

[1]Georgia Tech

[2]Universidade de Coimbra

[3]University of Washington

3rd July - EURO-INFORMS 2013 - Rome

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} = \min p(x)$ over all x such that

$$x \in \{x \mid p_i(x) = 0,\ i = 1, ..., t\}.$$

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} = \min p(x)$ over all x such that

$$x \in \mathcal{V}(\langle p_1, \cdots, p_t \rangle) = \mathcal{V}(I).$$

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} = \max \lambda$ such that

$$p(x) - \lambda \geq 0 \text{ for all } x \in \mathcal{V}(I).$$

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} \leq p_{\text{sos}} =$ max λ such that

$$p(x) - \lambda = \sum h_i^2 \bmod I.$$

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} \leq p_{\text{sos}} \leq p^k_{\text{sos}} = \max \lambda$ such that

$$p(x) - \lambda = \sum h_i^2 \bmod I, \quad \deg(h_i) \leq k.$$

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} \leq p_{\text{sos}} \leq p_{\text{sos}}^k = \max \lambda$ such that

$$p(x) - \lambda \in \Sigma_k[I].$$

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} \le p_{\text{sos}} \le p^k_{\text{sos}} = \max \lambda$ such that

$$p(x) - \lambda \in \Sigma_k[I].$$

However, there are other sum of squares certificates for nonnegativity.

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} \leq p_{\text{sos}} \leq p^k_{\text{sos}} = \max \lambda$ such that

$$p(x) - \lambda \in \Sigma_k[I].$$

However, there are other sum of squares certificates for nonnegativity.

Polynomial Optimization over Algebraic Varieties II

$$(p(x) - \lambda)g(x) \in \Sigma_k[I], \quad \text{for some positive } g(x)).$$

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} \leq p_{\text{sos}} \leq p^k_{\text{sos}} = \max\ \lambda$ such that

$$p(x) - \lambda \in \Sigma_k[I].$$

However, there are other sum of squares certificates for nonnegativity.

Polynomial Optimization over Algebraic Varieties II

$p_{\min} \leq p^{l,k}_{\text{sos}} = \max\ \lambda$ such that

$$(p(x) - \lambda)(1 + g(x)) \in \Sigma_k[I], \quad \text{for some } g(x) \in \Sigma_l([I]).$$

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} \leq p_{\text{sos}} \leq p_{\text{sos}}^{k} = \max \lambda$ such that

$$p(x) - \lambda \in \Sigma_k[I].$$

However, there are other sum of squares certificates for nonnegativity.

Polynomial Optimization over Algebraic Varieties II

$p_{\min} \leq p_{\text{sos}}^{l,k} = \max \lambda$ such that

$$(p(x) - \lambda)(1 + \Sigma_l[I]) \cap \Sigma_k[I] \neq \emptyset.$$

Sums of squares certificates

Polynomial Optimization over Algebraic Varieties

$p_{\min} \leq p_{\text{sos}} \leq p^k_{\text{sos}} = \max \lambda$ such that

$$p(x) - \lambda \in \Sigma_k[I].$$

However, there are other sum of squares certificates for nonnegativity.

Polynomial Optimization over Algebraic Varieties II

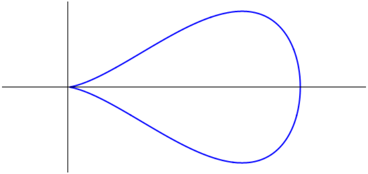
$p_{\min} \leq p^{l,k}_{\text{sos}} = \max \lambda$ such that

$$(p(x) - \lambda)(1 + \Sigma_l[I]) \cap \Sigma_k[I] \neq \emptyset.$$

This is not a linear SDP anymore, but is still doable.

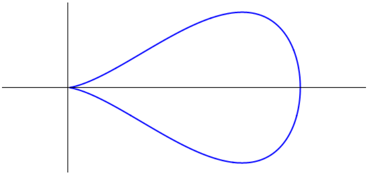
Example

Consider the teardrop curve given by $I = \langle x^4 - x^3 + y^2 \rangle$.



Example

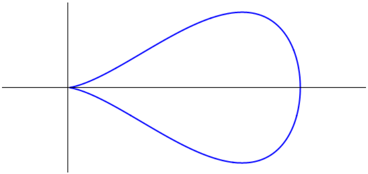
Consider the teardrop curve given by $I = \langle x^4 - x^3 + y^2 \rangle$.



Let $p(x) = x$

Example

Consider the teardrop curve given by $I = \langle x^4 - x^3 + y^2 \rangle$.

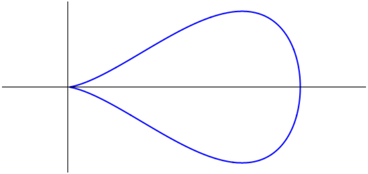


Let $p(x) = x$ then

$$p^2_{\text{sos}} = -0.1250,$$

Example

Consider the teardrop curve given by $I = \langle x^4 - x^3 + y^2 \rangle$.

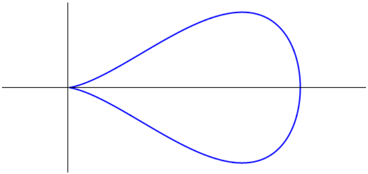


Let $p(x) = x$ then

$$p^2_{\text{sos}} = -0.1250, \quad p^3_{\text{sos}} = -0.0208,$$

Example

Consider the teardrop curve given by $I = \langle x^4 - x^3 + y^2 \rangle$.

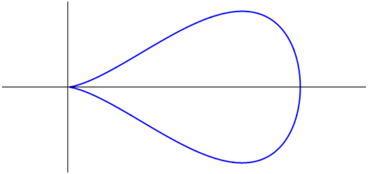


Let $p(x) = x$ then

$$p^2_{\text{sos}} = -0.1250, \quad p^3_{\text{sos}} = -0.0208, \quad p^4_{\text{sos}} = -0.0092, \quad \ldots$$

Example

Consider the teardrop curve given by $I = \langle x^4 - x^3 + y^2 \rangle$.



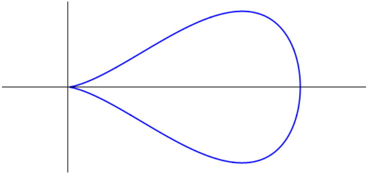
Let $p(x) = x$ then

$$p^2_{\text{sos}} = -0.1250, \quad p^3_{\text{sos}} = -0.0208, \quad p^4_{\text{sos}} = -0.0092, \quad \ldots$$

However $p^{1,2}_{\text{sos}} = p_{\text{min}} = 0$.

Example

Consider the teardrop curve given by $I = \langle x^4 - x^3 + y^2 \rangle$.



Let $p(x) = x$ then

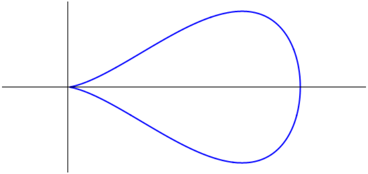
$$p_{\text{sos}}^2 = -0.1250, \quad p_{\text{sos}}^3 = -0.0208, \quad p_{\text{sos}}^4 = -0.0092, \quad \ldots$$

However $p_{\text{sos}}^{1,2} = p_{\min} = 0$. In fact

$$x^2 \cdot x = x^4 + y^2 \text{ modulo } I.$$

Example

Consider the teardrop curve given by $I = \langle x^4 - x^3 + y^2 \rangle$.



Let $p(x) = x$ then

$$p^2_{\text{sos}} = -0.1250, \quad p^3_{\text{sos}} = -0.0208, \quad p^4_{\text{sos}} = -0.0092, \quad \ldots$$

However $p^{1,2}_{\text{sos}} = p_{\min} = 0$. In fact

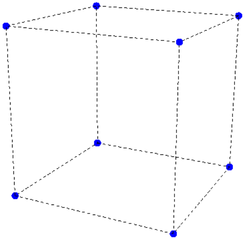
$$x^2 \cdot x = x^4 + y^2 \text{ modulo } I.$$

Multipliers make the relaxations less sensitive to singularities.

The n-cube

We are interested in the n-cube:

$$C_n = \{0,1\}^n = \{x \in \mathbb{R}^n \ : \ x_i^2 - x_i = 0, \ i = 1, \cdots, n\} = \mathcal{V}(I_n).$$

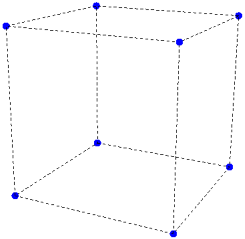


Cube C_3

The n-cube

We are interested in the n-cube:

$$C_n = \{0,1\}^n = \{x \in \mathbb{R}^n \ : \ x_i^2 - x_i = 0, \ i = 1, \cdots, n\} = \mathcal{V}(I_n).$$



Cube C_3

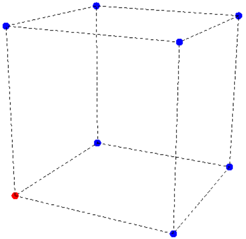
S_n acts on C_n by permuting coordinates, and if p is symmetric, it will be completely characterized by its evaluation at the levels T_k of the cube:

$$T_k = \{x \in C_n \ : \ \sum x_i = k\}.$$

The n-cube

We are interested in the n-cube:

$$C_n = \{0,1\}^n = \{x \in \mathbb{R}^n \ : \ x_i^2 - x_i = 0, \ i = 1, \cdots, n\} = \mathcal{V}(I_n).$$



Level T_0

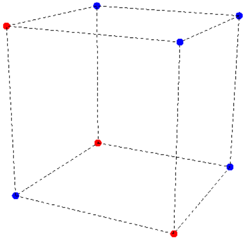
S_n acts on C_n by permuting coordinates, and if p is symmetric, it will be completely characterized by its evaluation at the levels T_k of the cube:

$$T_k = \{x \in C_n \ : \ \sum x_i = k\}.$$

The n-cube

We are interested in the n-cube:

$$C_n = \{0,1\}^n = \{x \in \mathbb{R}^n \ : \ x_i^2 - x_i = 0, \ i = 1, \cdots, n\} = \mathcal{V}(I_n).$$



Level T_1

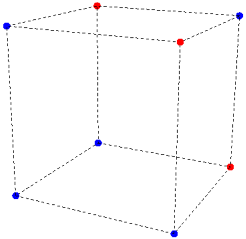
S_n acts on C_n by permuting coordinates, and if p is symmetric, it will be completely characterized by its evaluation at the levels T_k of the cube:

$$T_k = \{x \in C_n \ : \ \sum x_i = k\}.$$

The n-cube

We are interested in the n-cube:

$$C_n = \{0,1\}^n = \{x \in \mathbb{R}^n \ : \ x_i^2 - x_i = 0,\ i = 1, \cdots, n\} = \mathcal{V}(I_n).$$



Level T_2

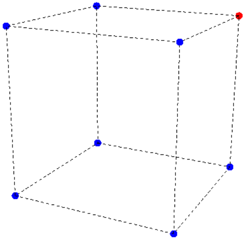
S_n acts on C_n by permuting coordinates, and if p is symmetric, it will be completely characterized by its evaluation at the levels T_k of the cube:

$$T_k = \{x \in C_n \ : \ \sum x_i = k\}.$$

The n-cube

We are interested in the n-cube:

$$C_n = \{0,1\}^n = \{x \in \mathbb{R}^n \ : \ x_i^2 - x_i = 0, \ i = 1, \cdots, n\} = \mathcal{V}(I_n).$$



Level T_3

S_n acts on C_n by permuting coordinates, and if p is symmetric, it will be completely characterized by its evaluation at the levels T_k of the cube:

$$T_k = \{x \in C_n \ : \ \sum x_i = k\}.$$

Main Result 1 - Bad news

Let p be a symmetric square-free polynomial attaining its minimum over C_n at level T_k, with $\deg p \leq k \leq n/2$.

Main Result 1 - Bad news

Let p be a symmetric square-free polynomial attaining its minimum over C_n at level T_k, with $\deg p \leq k \leq n/2$.

Theorem

If T_k is not a local extreme of p over $\mathbb{R}^n$ (seen as a polynomial in $\sum x_i$) then $p_{\min} > p_{\text{sos}}^{k-r,k}$, where $r = \lceil (\deg p)/2 \rceil$.

Main Result 1 - Bad news

Let p be a symmetric square-free polynomial attaining its minimum over C_n at level T_k, with $\deg p \leq k \leq n/2$.

Theorem

If T_k is not a local extreme of p over $\mathbb{R}^n$ (seen as a polynomial in $\sum x_i$) then $p_{\min} > p_{\text{sos}}^{k-r,k}$, where $r = \lceil (\deg p)/2 \rceil$.

This means that if the minimizer of p is "simple enough" and is close to the central levels of the cube, we need high level sos relaxations.

Main Result 1 - Bad news

Let p be a symmetric square-free polynomial attaining its minimum over C_n at level T_k, with $\deg p \leq k \leq n/2$.

Theorem

If T_k is not a local extreme of p over $\mathbb{R}^n$ (seen as a polynomial in $\sum x_i$) then $p_{\min} > p_{\mathrm{sos}}^{k-r,k}$, where $r = \lceil (\deg p)/2 \rceil$.

This means that if the minimizer of p is “simple enough” and is close to the central levels of the cube, we need high level sos relaxations.

The proof reduces to this lemma.

Lemma

If p has degree d and vanishes at T_k with $d \leq k \leq n - d$ then

$$p = (k - \sum x_i)q \mod I_n,$$

with $\deg q < \deg p$.

Sketch of Proof:

Consider the action of S_n in $\mathbb{R}[I]_k$.

Sketch of Proof:

Consider the action of S_n in $\mathbb{R}[I]_k$. It decomposes:

$$\mathbb{R}[I]_k = \mathbb{R}[I]_{=0} \oplus \mathbb{R}[I]_{=1} \oplus \mathbb{R}[I]_{=2} \oplus \cdots \oplus \mathbb{R}[I]_{=k}$$

Sketch of Proof:

Consider the action of S_n in $\mathbb{R}[I]_k$. It decomposes:

$$
\begin{array}{cccccccccccc}
\mathbb{R}[I]_k = & \mathbb{R}[I]_{=0} & \oplus & \mathbb{R}[I]_{=1} & \oplus & \mathbb{R}[I]_{=2} & \oplus & \cdots & \oplus & \mathbb{R}[I]_{=k} \\
& \wr\Vert & & \wr\Vert & & \wr\Vert & & & & \wr\Vert \\
& H_{n,0} & & H_{n,0} & & H_{n,0} & & \cdots & & H_{n,0} \\
& & & \oplus & & \oplus & & & & \oplus \\
& & & H_{n-1,1} & & H_{n-1,1} & & \cdots & & H_{n-1,1} \\
& & & & & \oplus & & & & \oplus \\
& & & & & H_{n-2,2} & & \cdots & & H_{n-2,2} \\
& & & & & & \ddots & & & \vdots \\
& & & & & & & & & H_{n-k,k}
\end{array}
$$

Sketch of Proof:

Consider the action of S_n in $\mathbb{R}[I]_k$. It decomposes:

$$\begin{array}{ccccccccccc}
\mathbb{R}[I]_k = & \mathbb{R}[I]_{=0} & \oplus & \mathbb{R}[I]_{=1} & \oplus & \mathbb{R}[I]_{=2} & \oplus & \cdots & \oplus & \mathbb{R}[I]_{=k} \\
 & \wr\| & & \wr\| & & \wr\| & & & & \wr\| \\
 & \color{red}{H_{n,0}} & & H_{n,0} & & H_{n,0} & & \cdots & & H_{n,0} \\
 & & & \oplus & & \oplus & & & & \oplus \\
 & & & \color{red}{H_{n-1,1}} & & H_{n-1,1} & & \cdots & & H_{n-1,1} \\
 & & & & & \oplus & & & & \oplus \\
 & & & & & \color{red}{H_{n-2,2}} & & \cdots & & H_{n-2,2} \\
 & & & & & & \ddots & & & \vdots \\
 & & & & & & & & & \color{red}{H_{n-k,k}}
\end{array}$$

Let M_j be the first copy of $H_{n-j,j}$ to appear,

Sketch of Proof:

Consider the action of S_n in $\mathbb{R}[I]_k$. It decomposes:

$$\begin{array}{cccccccccccc}
\mathbb{R}[I]_k = & \mathbb{R}[I]_{=0} & \oplus & \mathbb{R}[I]_{=1} & \oplus & \mathbb{R}[I]_{=2} & \oplus & \cdots & \oplus & \mathbb{R}[I]_{=k} \\
 & \wr\| & & \wr\| & & \wr\| & & & & \wr\| \\
 & H_{n,0} & & H_{n,0} & & H_{n,0} & & \cdots & & H_{n,0} \\
 & & & \oplus & & \oplus & & & & \oplus \\
 & & & H_{n-1,1} & & H_{n-1,1} & & \cdots & & H_{n-1,1} \\
 & & & & & \oplus & & & & \oplus \\
 & & & & & H_{n-2,2} & & \cdots & & H_{n-2,2} \\
 & & & & & & \ddots & & & \vdots \\
 & & & & & & & & & H_{n-k,k}
\end{array}$$

Let M_j be the first copy of $H_{n-j,j}$ to appear, then

$$\mathbb{R}[I]_k = \bigoplus_{j=0}^{k} M_j \oplus (k - \sum x_i) M_j \oplus \cdots \oplus (k - \sum x_i)^{k-j} M_j$$

Sketch of Proof:

Consider the action of S_n in $\mathbb{R}[I]_k$. It decomposes:

$$\begin{array}{ccccccccccc}
\mathbb{R}[I]_k = & \mathbb{R}[I]_{=0} & \oplus & \mathbb{R}[I]_{=1} & \oplus & \mathbb{R}[I]_{=2} & \oplus & \cdots & \oplus & \mathbb{R}[I]_{=k} \\
 & \wr\| & & \wr\| & & \wr\| & & & & \wr\| \\
 & H_{n,0} & & H_{n,0} & & H_{n,0} & & \cdots & & H_{n,0} \\
 & & & \oplus & & \oplus & & & & \oplus \\
 & & & H_{n-1,1} & & H_{n-1,1} & & \cdots & & H_{n-1,1} \\
 & & & & & \oplus & & & & \oplus \\
 & & & & & H_{n-2,2} & & \cdots & & H_{n-2,2} \\
 & & & & & & \ddots & & & \vdots \\
 & & & & & & & & & H_{n-k,k}
\end{array}$$

Let M_j be the first copy of $H_{n-j,j}$ to appear, then

$$\mathbb{R}[I]_k = \bigoplus_{j=0}^{k} M_j \oplus (k - \sum x_i) M_j \oplus \cdots \oplus (k - \sum x_i)^{k-j} M_j$$

and is enough to check that M_j does not vanish at T_k.

Application 1 - MaxCut

Recall that the maxcut problem over K_n can be reduced to

Binary polynomial formulation of MaxCut

$$\max p(x) = \sum_{i \neq j} (1 - x_i) x_j \text{ s.t. } x \in C_n$$

Application 1 - MaxCut

Recall that the maxcut problem over K_n can be reduced to

Binary polynomial formulation of MaxCut

$$\max p(x) = \sum_{i \neq j} (1 - x_i) x_j \text{ s.t. } x \in C_n$$

Laurent has proved that Lassere relaxations are of limited use.

Laurent

For $n = 2k + 1$, $p_{\text{sos}}^k > p_{\max}$.

Application 1 - MaxCut

Recall that the maxcut problem over K_n can be reduced to

Binary polynomial formulation of MaxCut

$$\max p(x) = \sum_{i \neq j} (1 - x_i) x_j \text{ s.t. } x \in C_n$$

Laurent has proved that Lassere relaxations are of limited use.

Laurent

For $n = 2k + 1$, $p_{\text{sos}}^{k} > p_{\max}$.

Note that p attains its maximum in C_n at T_k and T_{k+1}, which are not local maxima of p over $\mathbb{R}^n$.

First corollary of main result 1

For $n = 2k + 1$, $p_{\text{sos}}^{k-1,k} > p_{\max}$.

Application 2 - Global Optimization

Let p be any polynomial in $\mathbb{R}^n$.

Application 2 - Global Optimization

Let p be any polynomial in $\mathbb{R}^n$.

Artin (Hilbert's 17th Problem)

For some l, k, $p_{\text{sos}}^{l,k} = p_{\min}$.

Application 2 - Global Optimization

Let p be any polynomial in $\mathbb{R}^n$.

Artin (Hilbert's 17th Problem)

For some l, k, $p_{\text{sos}}^{l,k} = p_{\min}$.

We also know that these l, k can be very high. However there were no examples for such behavior.

Application 2 - Global Optimization

Let p be any polynomial in $\mathbb{R}^n$.

Artin (Hilbert's 17th Problem)

For some l, k, $p_{\text{sos}}^{l,k} = p_{\min}$.

We also know that these l, k can be very high. However there were no examples for such behavior.

Second corollary of main result 1

For any k there is a degree 4 polynomial in $\mathbb{R}^{2k+1}$ for which $p_{\min} < p_{\text{sos}}^{k-2,k}$.

Application 2 - Global Optimization

Let p be any polynomial in $\mathbb{R}^n$.

Artin (Hilbert's 17th Problem)

For some l, k, $p_{\text{sos}}^{l,k} = p_{\text{min}}$.

We also know that these l, k can be very high. However there were no examples for such behavior.

Second corollary of main result 1

For any k there is a degree 4 polynomial in $\mathbb{R}^{2k+1}$ for which $p_{\text{min}} < p_{\text{sos}}^{k-2,k}$.

This is proven by a perturbed extension of the polynomial on the previous example.

Main Result 2 - Not so bad news

We have showed lower bounds to the effectiveness of sos for binary polynomial programming. Luckily we also can show some upper bounds.

Main Result 2 - Not so bad news

We have showed lower bounds to the effectiveness of sos for binary polynomial programming. Luckily we also can show some upper bounds.

Theorem

Let p be a non constant quadratic polynomial in $\mathbb{R}^{2k+1}$, then $p_{\min} = p_{\text{sos}}^{k+1,k+2}$.

Main Result 2 - Not so bad news

We have showed lower bounds to the effectiveness of sos for binary polynomial programming. Luckily we also can show some upper bounds.

Theorem

Let p be a non constant quadratic polynomial in $\mathbb{R}^{2k+1}$, then $p_{\min} = p_{\text{sos}}^{k+1,k+2}$.

The proof is based in dimension counting.

Application - MaxCut revisited

Consider the weighted maxcut formulation.

Binary polynomial formulation of MaxCut

$$\max p_\omega(x) = \sum_{i \neq j} \omega_{ij}(1 - x_i)x_j \text{ s.t. } x \in C_n,$$

where ω_{ij} is the weight of edge $\{i, j\}$.

Application - MaxCut revisited

Consider the weighted maxcut formulation.

Binary polynomial formulation of MaxCut

$$\max p_\omega(x) = \sum_{i \neq j} \omega_{ij}(1 - x_i)x_j \text{ s.t. } x \in C_n,$$

where ω_{ij} is the weight of edge $\{i, j\}$.

The negative result proved by Laurent has an opposed positive conjecture.

Conjecture (Laurent)

If $n = 2k + 1$, $(p_\omega)_{\min} = (p_\omega)_{\text{sos}}^{k+1}$ for all weights.

Application - MaxCut revisited

Consider the weighted maxcut formulation.

Binary polynomial formulation of MaxCut

$$\max p_\omega(x) = \sum_{i \neq j} \omega_{ij}(1 - x_i)x_j \text{ s.t. } x \in C_n,$$

where ω_{ij} is the weight of edge $\{i, j\}$.

The negative result proved by Laurent has an opposed positive conjecture.

Conjecture (Laurent)

If $n = 2k + 1$, $(p_\omega)_{\min} = (p_\omega)_{\text{sos}}^{k+1}$ for all weights.

A weaker version can now be proved.

Corollary of main result 2

If $n = 2k + 1$, $(p_\omega)_{\min} = (p_\omega)_{\text{sos}}^{k+1,k+2}$ for all weights.

The End

Thank You