

Polynomial optimization on finite sets.

Mauricio Velasco
Universidad Católica del Uruguay (UCU)

RECO2023
Pontificia Universidad Católica

LECTURE 3:

Symmetry in Sums of squares.

Outline:

Assume n is an even integer and let $X := \{-1, 1\}^n$. In this lecture we will prove lower bounds on the degree needed to represent a nonnegative quadratic function in $\mathbb{R}[X]$ as a sum-of-squares.

Outline:

Assume n is an even integer and let $X := \{-1, 1\}^n$. In this lecture we will prove lower bounds on the degree needed to represent a nonnegative quadratic function in $\mathbb{R}[X]$ as a sum-of-squares.

Theorem. (Blekherman, Gouveia, Pfeiffer (2016))

The quadratic function

$$f := \left(-2 + \sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i \right)$$

*is nonnegative and **cannot be expressed** as a sum-of-squares of functions of degree strictly less than $n/2 + 1$ in $\mathbb{R}[X]$.*

Outline:

Assume n is an even integer and let $X := \{-1, 1\}^n$. In this lecture we will prove lower bounds on the degree needed to represent a nonnegative quadratic function in $\mathbb{R}[X]$ as a sum-of-squares.

Theorem. (Blekherman, Gouveia, Pfeiffer (2016))

The quadratic function

$$f := \left(-2 + \sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i \right)$$

*is nonnegative and **cannot be expressed** as a sum-of-squares of functions of degree strictly less than $n/2 + 1$ in $\mathbb{R}[X]$.*

It follows that $0 = f_{\min} > f_{(n/2)}$

Outline:

Assume n is an even integer and let $X := \{-1, 1\}^n$. In this lecture we will prove lower bounds on the degree needed to represent a nonnegative quadratic function in $\mathbb{R}[X]$ as a sum-of-squares.

Theorem. (Blekherman, Gouveia, Pfeiffer (2016))

The quadratic function

$$f := \left(-2 + \sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i \right)$$

*is nonnegative and **cannot be expressed** as a sum-of-squares of functions of degree strictly less than $n/2 + 1$ in $\mathbb{R}[X]$.*

It follows that $0 = f_{\min} > f_{(n/2)}$ so the smallest degree where equality holds for every quadratic f is at least $n/2 + 1$.

Theorem. (Blekherman, Gouveia, Pfieffer (2016))

The quadratic function

$$f := \left(-2 + \sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i \right)$$

*is nonnegative and **cannot be expressed** as a sum-of-squares of functions of degree strictly less than $n/2 + 1$ in $\mathbb{R}[X]$.*

Theorem. (Blekherman, Gouveia, Pfieffer (2016))

The quadratic function

$$f := \left(-2 + \sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i \right)$$

*is nonnegative and **cannot be expressed** as a sum-of-squares of functions of degree strictly less than $n/2 + 1$ in $\mathbb{R}[X]$.*

- The function f is **invariant under all permutations of the coordinates**.

Theorem. (Blekherman, Gouveia, Pfieffer (2016))

The quadratic function

$$f := \left(-2 + \sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i \right)$$

*is nonnegative and **cannot be expressed** as a sum-of-squares of functions of degree strictly less than $n/2 + 1$ in $\mathbb{R}[X]$.*

- The function f is **invariant under all permutations of the coordinates**.
- The proof [due to G.Blekherman (unpublished)] characterizes all possible **invariant sums-of-squares in $\mathbb{R}[X]$** of degree $\leq r$.

Theorem. (Blekherman, Gouveia, Pfieffer (2016))

The quadratic function

$$f := \left(-2 + \sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i \right)$$

*is nonnegative and **cannot be expressed** as a sum-of-squares of functions of degree strictly less than $n/2 + 1$ in $\mathbb{R}[X]$.*

- The function f is **invariant under all permutations of the coordinates**.
- The proof [due to G.Blekherman (unpublished)] characterizes all possible **invariant sums-of-squares in $\mathbb{R}[X]$** of degree $\leq r$.
- The key tool for this characterization will be the relationship between **representation theory** and sums-of-squares [Gatermann-Parrilo (2004)].

Plan for Lecture 3:

- 1 A primer on representation theory.
- 2 Invariant sums-of-squares.
- 3 Application: Invariant sums-of-squares in the hypercube.

Part 1:

A primer on representation theory.

Let G be a finite group

Definition.

A **representation of G** is a pair (V, ρ) where

- 1 V is a finite-dimensional vector space.
- 2 $\rho : G \rightarrow GL(V)$ is a group homomorphism.

Let G be a finite group

Definition.

A **representation of G** is a pair (V, ρ) where

- 1 V is a finite-dimensional vector space.
- 2 $\rho : G \rightarrow GL(V)$ is a group homomorphism.

A representation ρ makes each element g of the abstract group G into a concrete matrix $\rho(g)$.

Let G be a finite group

Definition.

A **representation** of G is a pair (V, ρ) where

- 1 V is a finite-dimensional vector space.
- 2 $\rho : G \rightarrow GL(V)$ is a group homomorphism.

A representation ρ makes each element g of the abstract group G into a concrete matrix $\rho(g)$.

Definition.

A **morphism** between two representations (V_1, ρ_1) and (V_2, ρ_2) of G is a linear map $T : V_1 \rightarrow V_2$ satisfying

$$T \circ \rho_1(g) = \rho_2(g) T \text{ for all } g \in G.$$

Example

Let $S_2 := \{id, \tau\}$ where $\tau = (12)$.

Example

Let $S_2 := \{id, \tau\}$ where $\tau = (12)$.

For $V = \mathbb{R}^2$ define $\rho : S_2 \rightarrow GL(V)$ by

$$\rho(id) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho(\tau) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

(V, ρ) is a representation of S_2 .

Example

Let $S_2 := \{id, \tau\}$ where $\tau = (12)$.

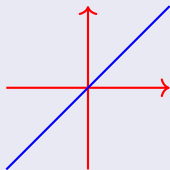
For $V = \mathbb{R}^2$ define $\rho : S_2 \rightarrow GL(V)$ by

$$\rho(id) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho(\tau) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

(V, ρ) is a representation of S_2 .

Remark.

Representations are the way in which abstract groups become concrete symmetries of a space.



$$\tau^2 = id \text{ so } \rho(\tau)\rho(\tau) = \rho(id)$$

Let (V, ρ) be a representation of G .

Definition.

A vector subspace $W \subseteq V$ is a **stable subspace** of V if

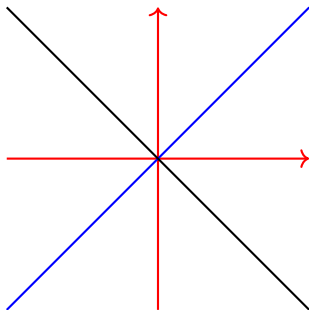
$$\forall w \in W \quad \forall g \in G \quad (\rho(g)(w) \in W).$$

Definition.

The representation (V, ρ) is **irreducible** if its only stable subspaces are $\{0\}$ and V .

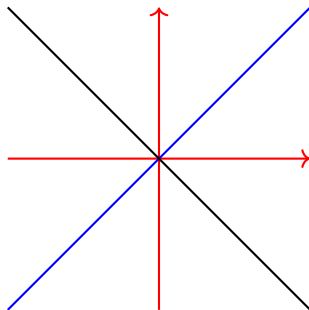
Example

The representation (V, ρ) of S_2 has two nontrivial stable subspaces



Example

The representation (V, ρ) of S_2 has two nontrivial stable subspaces



The representation (V, ρ) is NOT irreducible.

Why does this matter?

Representation Theory is important because of at least two reasons:

Why does this matter?

Representation Theory is important because of at least two reasons:

- 1 Representations appear **everywhere**.

Why does this matter?

Representation Theory is important because of at least two reasons:

- 1 Representations appear **everywhere**. For instance if (V_1, ρ_1) and (V_2, ρ_2) are two representations of G then $(V_1 \oplus V_2, \rho_1 \oplus \rho_2)$, $(V_1 \otimes V_2, \rho_1 \otimes \rho_2)$, $(\text{Hom}(V_1, V_2), \rho_1^* \otimes \rho_2)$ are representations of G .
- 2 There is an effective, complete and rather rigid **classification of the representations of G** , summarized in the following two slides...

Let G be a finite group. We have

Theorem. (Irreducible representations)

The following statements hold:

- 1 *There exists finitely many non-isomorphic irreducible representations of G . We denote them $V^{(1)}, V^{(2)}, \dots, V^{(c)}$ where c is the number of conjugacy classes of G .*
- 2 *The morphisms between irreducible representations satisfy*

$$\mathrm{Hom}_G(V^{(i)}, V^{(j)}) \cong \begin{cases} 0, & \text{if } i \neq j \\ \mathbb{C}, & \text{if } i = j. \end{cases}$$

Let G be a finite group. We have

Theorem. (Irreducible representations)

The following statements hold:

- 1 *There exists finitely many non-isomorphic irreducible representations of G . We denote them $V^{(1)}, V^{(2)}, \dots, V^{(c)}$ where c is the number of conjugacy classes of G .*
- 2 *The morphisms between irreducible representations satisfy*

$$\mathrm{Hom}_G(V^{(i)}, V^{(j)}) \cong \begin{cases} 0, & \text{if } i \neq j \\ \mathbb{C}, & \text{if } i = j. \end{cases}$$

The irreducible representations serve as building blocks of all other representations...

Let (W, ρ) be any representation of a finite group G and let \langle, \rangle be a G -invariant inner product on W ($\langle u, v \rangle = \langle \rho_W(g)u, \rho_W(g)v \rangle$).

Theorem.

There exist mutually orthogonal stable subspaces of W :

$$V_1^{(1)}, \dots, V_{m_1}^{(1)}, V_1^{(2)}, \dots, V_{m_2}^{(2)}, \dots, V_1^{(c)}, \dots, V_{m_c}^{(c)} \subseteq W$$

Such that:

① *The restriction of W to $V_j^{(i)}$ is isomorphic to $V^{(i)}$*

②

$$W = \left(\bigoplus_{i_1=1}^{m_1} V_{i_1}^{(1)} \right) \oplus \left(\bigoplus_{i_2=1}^{m_2} V_{i_2}^{(2)} \right) \oplus \dots \oplus \left(\bigoplus_{i_c=1}^{m_c} V_{i_c}^{(c)} \right)$$

③ *The integers m_1, \dots, m_c and the **isotypical components***

$$W_j := \left(\bigoplus_{i_j=1}^{m_j} V_{i_j}^{(j)} \right) \text{ for } j = 1, \dots, c$$

are uniquely determined.

The *takehome message* is:

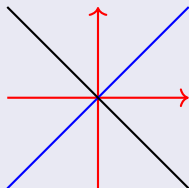
The *takehome message* is: On every problem involving symmetries there are **natural choices of coordinates**. Furthermore these coordinates can be computed effectively.

The *takehome message* is: On every problem involving symmetries there are **natural choices of coordinates**. Furthermore these coordinates can be computed effectively.

Example:

The *takehome message* is: On every problem involving symmetries there are **natural choices of coordinates**. Furthermore these coordinates can be computed effectively.

Example:

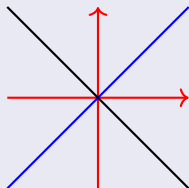


With respect to bases adapted to the stable subspaces we have

$$\rho(id) := \begin{matrix} & \text{triv} & \text{sgn} \\ \text{triv} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ \text{sgn} & \end{matrix} \quad \rho(\tau) := \begin{matrix} & \text{triv} & \text{sgn} \\ \text{triv} & \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ \text{sgn} & \end{matrix}$$

The *takehome message* is: On every problem involving symmetries there are **natural choices of coordinates**. Furthermore these coordinates can be computed effectively.

Example:



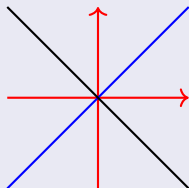
With respect to bases adapted to the stable subspaces we have

$$\rho(id) := \begin{matrix} & \text{triv} & \text{sgn} \\ \text{triv} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{matrix} \quad \rho(\tau) := \begin{matrix} & \text{triv} & \text{sgn} \\ \text{triv} & \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{matrix}$$

Both matrices became **simultaneously diagonal**.

The *takehome message* is: On every problem involving symmetries there are **natural choices of coordinates**. Furthermore these coordinates can be computed effectively.

Example:



With respect to bases adapted to the stable subspaces we have

$$\rho(id) := \begin{matrix} & \text{triv} & \text{sgn} \\ \text{triv} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ \text{sgn} & \end{matrix} \quad \rho(\tau) := \begin{matrix} & \text{triv} & \text{sgn} \\ \text{triv} & \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ \text{sgn} & \end{matrix}$$

Both matrices became **simultaneously diagonal**. In general the matrices will become **simultaneously block-diagonal**.

Part 2:

Invariant sums of squares.

Let $X \subseteq \mathbb{R}^n$ be a finite set and let G be a subgroup of permutations of X . The ring $\mathbb{R}[X]$ is naturally a representation of G .

Definition.

The **contragradient representation** $(\mathbb{R}[X], \rho^*)$ is the representation defined for $f \in \mathbb{R}[X]$ and $g \in G$ by the formula

$$[\rho^*(g)f](y) := f(g^{-1}(y))$$

Let $X \subseteq \mathbb{R}^n$ be a finite set and let G be a subgroup of permutations of X . The ring $\mathbb{R}[X]$ is naturally a representation of G .

Definition.

The **contragradient representation** $(\mathbb{R}[X], \rho^*)$ is the representation defined for $f \in \mathbb{R}[X]$ and $g \in G$ by the formula

$$[\rho^*(g)f](y) := f(g^{-1}(y))$$

This representation contains an important subring, the **ring of invariants** of X ,

$$\mathbb{R}[X]^G := \{f \in \mathbb{R}[X] : \forall g \in G (\rho^*(g)(f) = f)\}.$$

it is the isotypical component of the trivial representation.

Our main objective will be trying to understand the **invariant sums-of-squares** of some degree r , that is the sets

$$\Sigma_{\leq r}^G := \Sigma_{\leq r} \cap \mathbb{R}[X]^G.$$

Our main objective will be trying to understand the **invariant sums-of-squares** of some degree r , that is the sets

$$\Sigma_{\leq r}^G := \Sigma_{\leq r} \cap \mathbb{R}[X]^G.$$

*This set contains the **sums of invariant squares**, that is the elements of the form $s_1^2 + \cdots + s_m^2$ with $s_j \in \mathbb{R}[X]_{\leq r}^G$. Typically this inclusion is strict.*

Our main objective will be trying to understand the **invariant sums-of-squares** of some degree r , that is the sets

$$\Sigma_{\leq r}^G := \Sigma_{\leq r} \cap \mathbb{R}[X]^G.$$

*This set contains the **sums of invariant squares**, that is the elements of the form $s_1^2 + \cdots + s_m^2$ with $s_j \in \mathbb{R}[X]_{\leq r}^G$. Typically this inclusion is strict.*

Example:

Consider the action of S_2 in $\mathbb{R}[x, y]$ by coordinate permutations.

Our main objective will be trying to understand the **invariant sums-of-squares** of some degree r , that is the sets

$$\Sigma_{\leq r}^G := \Sigma_{\leq r} \cap \mathbb{R}[X]^G.$$

*This set contains the **sums of invariant squares**, that is the elements of the form $s_1^2 + \cdots + s_m^2$ with $s_j \in \mathbb{R}[X]_{\leq r}^G$. Typically this inclusion is strict.*

Example:

Consider the action of S_2 in $\mathbb{R}[x, y]$ by coordinate permutations. The polynomial $x^2 + y^2$ is invariant but x, y are not.

Our main objective will be trying to understand the **invariant sums-of-squares** of some degree r , that is the sets

$$\Sigma_{\leq r}^G := \Sigma_{\leq r} \cap \mathbb{R}[X]^G.$$

*This set contains the **sums of invariant squares**, that is the elements of the form $s_1^2 + \cdots + s_m^2$ with $s_j \in \mathbb{R}[X]_{\leq r}^G$. Typically this inclusion is strict.*

Example:

Consider the action of S_2 in $\mathbb{R}[x, y]$ by coordinate permutations. The polynomial $x^2 + y^2$ is invariant but x, y are not.

However

$$x^2 + y^2 = \left(\frac{x+y}{\sqrt{2}} \right)^2 + \left(\frac{x-y}{\sqrt{2}} \right)^2$$

The s_i are not invariant but do live in a fixed isotypical component.

Definition.

The **averaging operator** is the map $\mathcal{A} : \mathbb{R}[X] \rightarrow \mathbb{R}[X]$ given by

$$\mathcal{A}(f) := \frac{1}{|G|} \sum_{g \in G} \rho^*(g)(f)$$

The following properties hold:

- ① For every $f \in \mathbb{R}[X]$ the image $\mathcal{A}(f) \in \mathbb{R}[X]^G$.
- ② $\mathcal{A}(f) = f$ if and only if $f \in \mathbb{R}[X]^G$
- ③ It does not respect the product.

Definition.

The **averaging operator** is the map $\mathcal{A} : \mathbb{R}[X] \rightarrow \mathbb{R}[X]$ given by

$$\mathcal{A}(f) := \frac{1}{|G|} \sum_{g \in G} \rho^*(g)(f)$$

The following properties hold:

- 1 For every $f \in \mathbb{R}[X]$ the image $\mathcal{A}(f) \in \mathbb{R}[X]^G$.
- 2 $\mathcal{A}(f) = f$ if and only if $f \in \mathbb{R}[X]^G$
- 3 It does not respect the product.

Main result:

Assume $\mathbb{R}[X]_{\leq r}$ is a stable subspace of $\mathbb{R}[X]$ as a representation of the group G .

Theorem. (Invariant sums-of-squares)

If $\mathbb{R}[X]_{\leq r} = W_1 \oplus \cdots \oplus W_c$ is the isotypical decomposition of $\mathbb{R}[X]_{\leq r}$ then the following equality holds

$$\Sigma_{\leq r}^G = \mathcal{A}(\Sigma_{W_1}) + \cdots + \mathcal{A}(\Sigma_{W_c})$$

Proof

Suppose $f = s_1^2 + \cdots + s_N^2$ is an invariant sum of squares.

$$f = \mathcal{A}(f) = \sum_{i=1}^N \mathcal{A}(s_i^2)$$

Any summand $s = s_i \in \mathbb{R}[X]_{\leq r}$ has an isotypical decomposition

$$s = w_1 + \cdots + w_c$$

$$\mathcal{A}(s^2) = \mathcal{A} \left(w_1^2 + \cdots + w_c^2 + 2 \sum_{i < j} w_i w_j \right)$$

Since the W_i are distinct isotypical components $\mathcal{A}(w_i w_j) = 0$ for $i \neq j$ (**Exercise**) and we conclude

$$\mathcal{A}(s^2) = \mathcal{A}(w_1^2) + \cdots + \mathcal{A}(w_c^2)$$

Collecting similar terms we prove the statement.

Part 3:

Invariant sums of squares on the hypercube.

For n even let $X = \{-1, 1\}^n$, $G = S_n$ and $\ell(x_1, \dots, x_n) := \sum_{i=1}^n x_i$

Theorem. (Blekherman)

- 1 Every function $f \in \mathbb{R}[X]_{\leq n}^G$ can be written uniquely as a univariate polynomial in ℓ .
- 2 If $f \in \Sigma_{\leq n/2}^G$ then $f(\ell) = \sum_{i=0}^{n/2} p_i(\ell) s_i(\ell)$ where

$$p_i(\ell) = \prod_{j=1}^i ((2(n+1-j))^2 - \ell^2)$$

and s_i is a sum-of-squares of terms of degree $\leq n/2 - i$ in ℓ .

For n even let $X = \{-1, 1\}^n$, $G = S_n$ and $\ell(x_1, \dots, x_n) := \sum_{i=1}^n x_i$

Theorem. (Blekherman)

- 1 Every function $f \in \mathbb{R}[X]_{\leq n}^G$ can be written uniquely as a univariate polynomial in ℓ .
- 2 If $f \in \Sigma_{\leq n/2}^G$ then $f(\ell) = \sum_{i=0}^{n/2} p_i(\ell) s_i(\ell)$ where

$$p_i(\ell) = \prod_{j=1}^i ((2(n+1-j))^2 - \ell^2)$$

and s_i is a sum-of-squares of terms of degree $\leq n/2 - i$ in ℓ .

As a consequence:

- Every function $f(\ell) \in \Sigma_{\leq n/2}^G$ must be nonnegative in the **real interval** $[-2, 2]$ and thus

For n even let $X = \{-1, 1\}^n$, $G = S_n$ and $\ell(x_1, \dots, x_n) := \sum_{i=1}^n x_i$

Theorem. (Blekherman)

- 1 Every function $f \in \mathbb{R}[X]_{\leq n}^G$ can be written uniquely as a univariate polynomial in ℓ .
- 2 If $f \in \Sigma_{\leq n/2}^G$ then $f(\ell) = \sum_{i=0}^{n/2} p_i(\ell) s_i(\ell)$ where

$$p_i(\ell) = \prod_{j=1}^i ((2(n+1-j))^2 - \ell^2)$$

and s_i is a sum-of-squares of terms of degree $\leq n/2 - i$ in ℓ .

As a consequence:

- Every function $f(\ell) \in \Sigma_{\leq n/2}^G$ must be nonnegative in the **real interval** $[-2, 2]$ and thus
- The function $\ell(\ell - 2)$ is NOT in $\Sigma_{\leq n/2}^G$ proving...

The function $\ell(\ell - 2)$ is NOT in $\Sigma_{\leq n/2}^G$ proving...

Theorem. (Blekherman, Gouveia, Pfieffer (2016))

The nonnegative quadratic function

$$f := \left(-2 + \sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i \right)$$

cannot be expressed as a sum-of-squares of functions of degree strictly less than $n/2 + 1$ in $\mathbb{R}[X]$.

Decomposing the hypercube:

$\mathbb{R}[X]$ where $X = \{-1, 1\}^8$ as S_8 representation:

d	basis	$Isot$	$decomp$			
0	1	$S^{(8)}$				
1	x_1, \dots	$S^{(7,1)}$	$S^{(8)}$			
2	$x_1 x_2, \dots$	$S^{(6,2)}$	$S^{(7,1)}$	$S^{(8)}$		
3	$x_1 x_2 x_3, \dots$	$S^{(5,3)}$	$S^{(6,2)}$	$S^{(7,1)}$	$S^{(8)}$	
4	$x_1 x_2 x_3 x_4, \dots$	$S^{(4,4)}$	$S^{(5,3)}$	$S^{(6,2)}$	$S^{(7,1)}$	$S^{(8)}$
5	$x_1 x_2 x_3 x_4 x_5, \dots$	$S^{(5,3)}$	$S^{(6,2)}$	$S^{(7,1)}$	$S^{(8)}$	
6	$x_1 x_2 x_3 x_4 x_5 x_6, \dots$	$S^{(6,2)}$	$S^{(7,1)}$	$S^{(8)}$		
7	$x_1 x_2 x_3 x_4 x_5 x_6 x_7, \dots$	$S^{(7,1)}$	$S^{(8)}$			
8	$x_1 \dots x_8$	$S^{(8)}$				

Invariant sums-of-squares on the hypercube

Let $\ell(x_1, \dots, x_n) := \sum x_i$.

Lemma.

The following statements hold:

- 1 $\mathbb{R}[C]^G = \mathbb{R}[\ell] / \left(\ell \prod_{j=1}^n ((2j)^2 - \ell^2) \right).$
- 2 *The isotypical component of $\mathbb{R}[C]^G$ corresponding to the representation $S^{(n-k,k)}$ is given by*

$$W_{(n-k,k)} := \left\{ \sum_{j=0}^{n-2k} \ell^j f_j : f_j \in S^{(n-k,k)} \subseteq \mathbb{R}[C]_k \right\}$$

for $0 \leq k \leq n/2$.

Our Theorem on invariant sums-of-squares applied to the above isotypical decomposition implies Blekherman's characterization.