Limitantes de programação semidefinida para o número de contato

Fabrício Caluza Machado

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Mestre em Ciências

Programa: Ciência da Computação
Orientador: Prof. Dr. Fernando Mário de Oliveira Filho

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processos no. 2014/16058-0 e 2015/05648-4, e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

São Paulo, março de 2017

Limitantes de programação semidefinida para o número de contato

Esta versão da dissertação contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 21/02/2017. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Fernando Mário de Oliveira Filho (orientador) - IME-USP
- Profª. Drª. Sandra Augusta Santos - UNICAMP
- Prof. Dr. Frank Vallentin - Universität zu Köln

Resumo

MACHADO, F. C. **Limitantes de programação semidefinida para o número de contato**. 94 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. 2017.

O *número de contato* do $\mathbb{R}^n$ (em inglês, "kissing number") é o maior número de esferas de raio unitário e interiores dois-a-dois disjuntos que podem tocar simultaneamente uma esfera de raio unitário central. Nesta dissertação estudamos métodos que limitam o tamanho de tais configurações através de técnicas de otimização, como dualidade e programação semidefinida. O principal resultado obtido foi o cálculo de melhores limitantes para o número de contato nas dimensões 9 a 23; o que foi possível graças à exploração de simetrias dos polinômios presentes no limitante proposto por Bachoc e Vallentin (2008), levando à consideração de programas semidefinidos menores. Por fim, o limitante estudado é estendido para uma classe mais geral de problemas.

Palavras-chave: número de contato, códigos esféricos, empacotamento, programação semidefinida.

Abstract

MACHADO, F. C. **Semidefinite programming bounds for the kissing number**. 94 pp. Master thesis - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. 2017.

The *kissing number* of $\mathbb{R}^n$ is the maximum number of pairwise-nonoverlapping unit spheres that can simultaneously touch a central unit sphere. In this thesis we study methods to bound from above the size of such configurations using optimization techniques, like duality and semidefinite programming. The main result achieved is the computation of better bounds for the kissing number in dimensions 9 to 23; a result possible due to the exploitation of symmetries in the polynomials present in the bound proposed by Bachoc and Vallentin (2008), leading to the consideration of smaller semidefinite programs. Finally, the studied bound is extended to a bigger class of problems.

Keywords: kissing number, spherical codes, packing, semidefinite programming.

Sumário

Introdução

O *número de contato* do $\mathbb{R}^n$ (em inglês, "kissing number") denotado por τ_n, é o maior número de esferas de raio unitário e interiores dois-a-dois disjuntos que podem tocar simultaneamente uma esfera de raio unitário central.

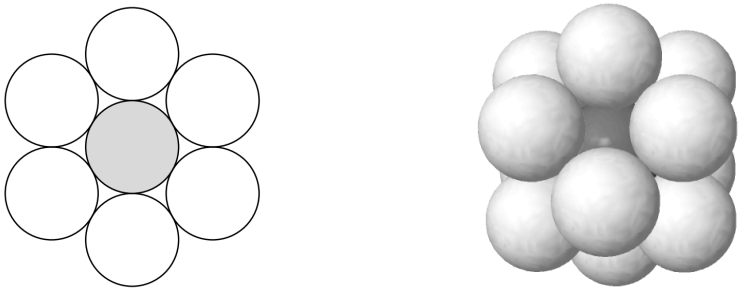

FIGURA 1. À esquerda, seis discos tocam um disco central no plano euclidiano. À direita, uma configuração com 12 esferas que mostra $\tau_3 \geq 12$.

Consideramos o espaço euclidiano $\mathbb{R}^n$ com seu produto interno $x \cdot y := \sum_{i=1}^{n} x_i y_i$ e denotamos a esfera unitária centrada na origem por $S^{n-1} := \{x \in \mathbb{R}^n : x \cdot x = 1\}$. A distância angular $d(x, y)$ entre dois pontos $x, y \in S^{n-1}$ é definida como $d(x, y) := \arccos(x \cdot y)$. Dada uma configuração de esferas que satisfaça as condições do problema, como os centros de três esferas que se tocam formam um triângulo equilátero, os pontos de contato dessas esferas com a esfera central possuem distância angular mínima de $\pi/3$ (Figura 2) e, reciprocamente, qualquer conjunto de pontos com distância angular mínima de $\pi/3$ fornece uma configuração de contato.

Conjuntos de pontos com distância angular mínima θ são chamados de *códigos esféricos*. Definindo o parâmetro (usamos $|C|$ para denotar a cardinalidade de um conjunto C)

$$A(n, \theta) := \max\{|C| : C \subset S^{n-1}, d(x, y) \geq \theta \text{ para } x, y \in C, x \neq y\},$$

que representa o maior número de pontos em S^{n-1} com distância angular mínima θ, temos que $\tau_n = A(n, \pi/3)$. O problema de determinar o maior código esférico para um certo valor de θ já foi considerado extensivamente e possui diversas aplicações, como no projeto de sinais de energia constante no chamado canal gaussiano em teoria das comunicações ou na limitação do número de soluções de certas equações diofantinas com formas quadráticas em teoria dos números (veja o artigo de Sloane [Slo81] ou os Capítulos 1 e 3 de Conway e Sloane [CS99] para mais referências). Também podemos enxergar um código esférico de distância angular mínima θ como um empacotamento de calotas esféricas[1] de raio $\theta/2$ em S^{n-1}.

Às vezes o problema de empacotamento de calotas esféricas em dimensão $n = 3$ é formulado de maneira diferente e pergunta-se qual o maior raio tal que existe

[1]Uma *calota esférica* de centro $e \in S^{n-1}$ e raio ϕ é o conjunto $\{x \in S^{n-1} : d(e, x) \leq \phi\}$.

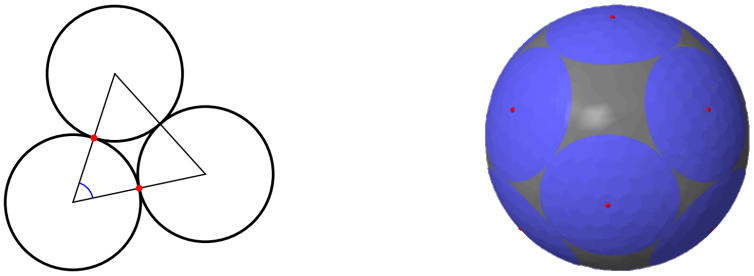

FIGURA 2. À esquerda, três esferas que se tocam simultaneamente e o ângulo de $\pi/3$ formado entre os pontos de contato. À direita, um empacotamento de calotas esféricas de raio $\pi/6$.

um empacotamento com um certo número N de calotas, posto dessa forma esse problema é conhecido como problema de Tammes (veja Coxeter [Cox62]).

Determinar $A(n, \theta)$ é um problema difícil que depende não apenas da construção de códigos esféricos grandes, mas também do desenvolvimento de técnicas que limitem o tamanho de qualquer código. Nesta dissertação estamos interessados em estudar métodos que limitem superiormente o tamanho de códigos esféricos, em especial do número de contato. Para isso será útil vermos $A(n, \theta)$ como o número de independência de um grafo: seja $G_{n,\theta}$ o grafo com conjunto de vértices S^{n-1} e no qual dois vértices distintos x, y são adjacentes se $d(x, y) < \theta$. Dessa forma, os códigos esféricos correspondem aos conjuntos independentes de $G_{n,\theta}$ e $\alpha(G_{n,\theta}) = A(n, \theta)$.

O número de contato só é conhecido exatamente nas dimensões 1, 2, 3, 4, 8 e 24. Enquanto é fácil ver que $\tau_2 = 6$, determinar τ_3 foi um problema muito mais difícil, que remonta a uma disputa em 1694 entre Isaac Newton, para o qual $\tau_3 = 12$, e David Gregory, que conjecturou $\tau_3 = 13$. Enquanto é possível exibir uma configuração com 12 bolas que tocam uma bola central (Figura 1), não é fácil mostrar que não é possível rearranjá-las de modo a encaixar uma 13ª bola no espaço que sobra (de fato, sobra tanto espaço que é possível alcançar qualquer permutação apenas girando as esferas em torno da esfera central, veja detalhes no final do Capítulo 1 de Conway e Sloane [CS99]; recentemente, Kusner et al. [KKLS16] fizeram um estudo das possíveis configurações de contato). Newton estava certo, mas uma prova de que $\tau_3 = 12$ só foi fornecida por Schütte e van der Waerden [SvdW53] em 1953.

A resposta para as dimensões 8 ($\tau_8 = 240$) e 24 ($\tau_{24} = 196560$) foi obtida em 1979 por Odlyzko e Sloane [OS79] e independentemente por Levenshtein [Lev79], baseados em um método proposto por Delsarte, Goethals e Seidel [DGS77] anos antes e conhecido como limitante de programação linear, que funciona especialmente bem quando a configuração ótima é simétrica e única, como ocorre nessas dimensões (as configurações relacionam-se respectivamente com os reticulados E_8 e de Leech; veja mais informações sobre esses reticulados no Capítulo 4 de Conway e Sloane [CS99]).

Em dimensão 4, o limitante de programação linear apenas mostra $\tau_4 \leq 25$. Ainda que por muito tempo conjecturado, o valor correto ($\tau_4 = 24$) só foi provado em 2003 por Musin [Mus08] com uma modificação desse método (veja também a exposição de Pfender e Ziegler [PZ04]).

Nas demais dimensões apenas são conhecidos limitantes. Nas dimensões menores do que 24, os melhores limitantes superiores conhecidos atualmente foram calculados por Machado e Oliveira [MdOF17] (o resultado principal desta dissertação), melhorando os resultados de Mittelman e Vallentin [MV10]; ambos os trabalhos usam um método proposto por Bachoc e Vallentin [BV08], conhecido como

limitante de programação semidefinida, que melhora o limitante de programação linear através da adição de restrições baseadas nas relações entre triplas de pontos.

Para completar o quadro, também são conhecidos resultados assintóticos. Kabatiansky e Levenshtein [KL78] mostram que $\tau_n \leq 2^{0.401n(1+o(1))}$, enquanto Wyner [Wyn65] mostra de forma não-construtiva que $\tau_n \geq 2^{(1-0.5\log_2 3)n(1+o(1))} = 2^{0.2075\ldots n(1+o(1))}$ (veja também as Seções 1.2.2 e 9.3.5 de Conway e Sloane [CS99]).

Contribuições e organização do trabalho

Esta dissertação organiza-se da seguinte forma. No Capítulo 1 são apresentadas introduções aos principais pré-requisitos do texto: teoria das representações (necessária para o aproveitamento das simetrias presentes no problema), kernels de Hilbert-Schmidt (necessários para a generalização dos limitantes inicialmente definidos para grafos finitos) e teoria de dualidade em programação cônica (necessária para a formulação dos programas de otimização que serão considerados). No Capítulo 2 é apresentado o limitante de programação linear e são demonstradas as propriedades dos polinômios harmônicos esféricos usadas pelo limitante. No Capítulo 3 é apresentado o limitante de programação semidefinida, o principal limitante considerado neste trabalho. No Capítulo 4 são apresentadas as técnicas de otimização polinomial necessárias para o cálculo do limitante, assim como os resultados obtidos. Por fim, no Capítulo 5 o limitante de programação semidefinida é estendido para uma classe mais geral de problemas, através dos "grafos topológicos de empacotamento".

O principal resultado obtido foi o cálculo de melhores limitantes para o número de contato nas dimensões 9 a 23; tal progresso foi possível graças à exploração de simetrias dos polinômios presentes no limitante de programação semidefinida, levando à resolução de problemas de otimização menores. Esse resultado é exibido na Seção 4.3 e já foi aceito para publicação:

[MdOF17] Fabrício C. Machado e Fernando M. de Oliveira Filho. Improving the semidefinite programming bound for the kissing number by exploiting polynomial symmetry. Aceito em *Experiment. Math.*, 2017. arXiv:1609.05167.

A apresentação dos limitantes de 3 pontos para grafos topológicos de empacotamento compactos feita na Seção 5.3 e a interpretação do limitante de programação semidefinida como um exemplo desses limitantes também é uma contribuição nova.

Notas sobre as notações

Nesta dissertação usamos a notação conhecida como *delta de Kronecker*, definida por

$$\delta_{i,j} = \begin{cases} 1 & \text{se } i = j, \\ 0 & \text{caso contrário.} \end{cases}$$

Se C é um conjunto finito, usamos a notação $\mathbb{R}^C$ para denotar o espaço vetorial das funções de C em $\mathbb{R}$. Tal como normalmente usado para o $\mathbb{R}^n$, vemos uma função $f \in \mathbb{R}^C$ como um vetor coluna indexado por C, vemos f^T como o vetor linha correspondente e uma função $A \in \mathbb{R}^{C\times C}$ como uma matriz, de modo que para $g \in \mathbb{R}^C$, definimos $g^T f \in \mathbb{R}$ por $g^T f := \sum_{x\in C} f(x)g(x)$ e $fg^T \in \mathbb{R}^{C\times C}$ por $(fg^T)(x,y) := f(x)g(y)$. Também definimos $Af \in \mathbb{R}^C$ por $(Af)(x) := \sum_{y\in C} A(x,y)f(y)$ e $g^T Af \in \mathbb{R}$ por $g^T Af := \sum_{x,y\in C} A(x,y)g(x)f(y)$.

Usamos $\langle , \rangle$ para denotar um produto interno em um espaço vetorial, quando o espaço for complexo convencionamos que ele é antilinear na segunda entrada,

de modo que $\langle u, \alpha v\rangle = \overline{\alpha}\,\langle u, v\rangle$. Também usamos a notação $\langle , \rangle$ para denotar o produto do traço entre matrizes, definido por $\langle A, B\rangle := \operatorname{tr}(B^T A) = \sum_{i,j} A_{i,j} B_{i,j}$ (ou $\operatorname{tr}(B^* A)$ no caso de matrizes complexas); observe que com as notações do parágrafo anterior, vale que $g^T A f = \left\langle A, f g^T \right\rangle$.

Usamos a notação $X \succeq 0$ para denotar que uma matriz é positivo-semidefinida, isto é, simétrica e tal que para qualquer vetor f vale $f^T X f \geq 0$ no caso real e hermitiana e tal que para qualquer vetor f vale $f^* X f \geq 0$ no caso complexo.

Nesta dissertação consideramos diversos problemas de otimização e usamos max ou min para denotar se eles são problemas de maximização ou minimização sem com isso estar afirmando que os valores ótimos desses problemas são atingidos (eles devem ser interpretados como sup e inf).

CAPÍTULO 1

Preliminares

1.1. Teoria das representações

Nesta seção vemos as definições e resultados básicos da teoria das representações. As principais referências são os dois primeiros capítulos de Fulton e Harris [FH91], os dois primeiros capítulos de Serre [Ser77] e o primeiro capítulo de Vilenkin [Vil68]. Essa teoria será útil pois em diversas partes desta dissertação (como no fim da Seção 2.2, na Seção 3.1 e na Seção 4.2) consideramos funções invariantes sob a ação de grupos e estamos interessados em simplificar suas descrições; isso é possível através do método de bloco-diagonalização descrito na Seção 1.1.3.

1.1.1. Definições. Neste trabalho consideramos grupos compactos, finitos ou infinitos. Um *grupo compacto* é um grupo que também é um espaço topológico compacto onde as operações do grupo (produto e inverso) são contínuas. Uma propriedade importante desses grupos é a existência e unicidade da *medida de Haar* (veja o Teorema 11.4 de Conway [Con90]), que é uma medida de Borel μ positiva, regular, que atribui valor positivo e não-nulo aos conjuntos abertos, normalizada de modo que $\mu(G) = 1$ e invariante (isto é, tal que para todo boreliano $E \subseteq G$ e $g \in G$, vale que $\mu(E) = \mu(gE) = \mu(Eg) = \mu(E^{-1})$, onde $gE = \{gx : x \in E\}$, $Eg = \{xg : x \in E\}$ e $E^{-1} = \{x^{-1} : x \in E\}$). Com a medida de Haar podemos construir uma integral invariante tal que

$$\int_G f(g)\,\mathrm{d}\mu(g) = \int_G f(hg)\,\mathrm{d}\mu(g) = \int_G f(gh)\,\mathrm{d}\mu(g) = \int_G f(g^{-1})\,\mathrm{d}\mu(g),$$

para todo $h \in G$ e f μ-integrável.

Seja V um espaço vetorial, nesta seção iremos sempre supor que os espaços vetoriais são complexos (outros corpos, como os reais, também podem ser considerados, porém a teoria de representações torna-se mais complicada nesses casos). O grupo formado pelos operadores lineares inversíveis de V é chamado de grupo linear e denotado $\mathrm{GL}(V)$. Uma *representação* de G é um homomorfismo contínuo ρ entre G e $\mathrm{GL}(V)$ (isto é, uma função contínua $\rho\colon G \to \mathrm{GL}(V)$ tal que $\rho(g)\rho(h) = \rho(gh)$ para todos $g, h \in G$). Às vezes o homomorfismo ρ é omitido, $\rho(g)$ é denotado diretamente por g e V é chamado de representação de G. Nesta dissertação consideramos representações em espaços de dimensão finita e mais geralmente em espaços de Hilbert (veja, e.g., o Capítulo II de Reed e Simon [RS72] para uma introdução aos espaços de Hilbert). Nesse segundo caso também supomos que os operadores $\rho(g)$ sejam contínuos em V.

Se $\langle , \rangle : V^2 \to \mathbb{C}$ é um produto interno de V, dizemos que a representação $\rho\colon G \to \mathrm{GL}(V)$ é *unitária* se

$$\langle \rho(g)v, \rho(g)w \rangle = \langle v, w \rangle$$

para todo $g \in G$ e $v, w \in V$.

Proposição 1.1. *Se G é um grupo compacto e $\rho\colon G \to \mathrm{GL}(V)$ é uma representação em um espaço V com um produto interno $\langle , \rangle_0$ qualquer, então existe um produto interno com relação ao qual a representação é unitária.*

Demonstração. Considerando inicialmente que G seja um grupo finito, um produto interno invariante pode ser construído por simetrização:

$$\langle u, v\rangle := \frac{1}{|G|}\sum_{g\in G}\langle \rho(g)u, \rho(g)v\rangle_0 .$$

A mesma coisa pode ser feita caso G seja infinito e compacto, nesse caso usamos sua medida de Haar μ:

$$\langle u, v\rangle := \int_G \langle \rho(g)u, \rho(g)v\rangle_0 \,\mathrm{d}\mu(g). \qquad \square$$

Apesar de a proposição anterior ter uma demonstração simples, ela permite-nos afirmar que toda representação de um grupo compacto em um espaço com produto interno pode ser considerada como uma representação unitária e assim a hipótese sobre a representação ser unitária, que aparece em diversas proposições, só é relevante quando houver uma afirmação sobre o espaço com um determinado produto interno. Às vezes supor que a representação é unitária também nos permite omitir a hipótese sobre o grupo ser finito ou compacto.

Exemplo 1.2 (Representação regular). Se G é um grupo finito, podemos considerar o espaço $\mathbb{C}^G$ das funções de G em $\mathbb{C}$ com o produto interno

$$\langle \phi, \psi\rangle := \frac{1}{|G|}\sum_{g\in G}\phi(g)\overline{\psi(g)}. \tag{1}$$

Nesse espaço pode-se definir uma representação unitária de G chamada *representação regular*, em que cada elemento $g \in G$ define um operador $\rho(g)$ que leva a função $\phi \in \mathbb{C}^G$ na função

$$\big(\rho(g)\phi\big)(s) := \phi(g^{-1}s).$$

O conjunto $\{\phi_g : g \in G\}$ onde $\phi_g(s) = 1$ se $g = s$ e $\phi_g(s) = 0$ caso contrário, é uma base de $\mathbb{C}^G$. Calculando a ação de $\rho(g)$ em um elemento ϕ_h dessa base, temos $\big(\rho(g)\phi_h\big)(s) = \phi_h(g^{-1}s)$ e logo $\rho(g)\phi_h = \phi_{gh}$.

No caso de G ser um grupo compacto, uma representação similar pode ser considerada sobre o espaço $L^2(G)$ com o produto interno

$$\langle \phi, \psi\rangle := \int_G \phi(g)\overline{\psi(g)}\,\mathrm{d}\mu(g),$$

mais detalhes podem ser encontrados na Seção 1.2.4 de Vilenkin [Vil68].

Duas representações $\rho\colon G \to \mathrm{GL}(V)$ e $\tau\colon G \to \mathrm{GL}(W)$ são ditas *equivalentes* se existe um isomorfismo $T\colon V \to W$ que comuta com elas, ou seja, tal que

$$T\rho(g) = \tau(g)T, \tag{2}$$

para todo $g \in G$. Nesse caso, τ pode ser obtida a partir de ρ por $\tau(g) = T\rho(g)T^{-1}$. Como o nome sugere, essa é uma relação de equivalência; o estudo das representações pode ser reduzido ao estudo das classes de equivalência dessa relação.

Denotamos o conjunto dos homomorfismos entre dois espaços vetoriais V e W por $\mathrm{Hom}(V, W)$. Se $T\colon V \to W$ é um homomorfismo que comuta com duas representações, como em (2), dizemos que T é um *G-homomorfismo*. O conjunto dos G-homomorfismos é um subespaço vetorial de $\mathrm{Hom}(V, W)$ que denotamos por $\mathrm{Hom}(V, W)^G$ (as representações ρ e τ ficam subentendidas no contexto). Usamos também os nomes *G-isomorfismo* para os isomorfismos que comutam com as representações (como no parágrafo anterior) e *G-endomorfismo* para quando $V = W$ e $\rho = \tau$; nesse último caso, são usadas também as notações $\mathrm{End}(V) := \mathrm{Hom}(V, V)$ e $\mathrm{End}(V)^G := \mathrm{Hom}(V, V)^G$.

Podemos ver os G-homomorfismos como homomorfismos que fazem o seguinte diagrama comutar para todo g em G:

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle \rho(g)}\downarrow & & \downarrow{\scriptstyle \tau(g)} \\ V & \xrightarrow{T} & W \end{array} \tag{3}$$

Se $\rho\colon G \to \mathrm{GL}(V)$ é uma representação e W é um subespaço de V tal que $\rho(g)w \in W$ para todo $w \in W$ e $g \in G$, dizemos que W é um *subespaço invariante*. Se W for um subespaço próprio, não-nulo e invariante, podemos restringir $\rho(g)$ à W e obter uma representação $\rho^W : G \to \mathrm{GL}(W)$. Dizemos que ρ^W é uma *subrepresentação* de ρ. Se V não possuir nenhum subespaço próprio e não-nulo invariante, dizemos que ρ é uma representação *irredutível*.

Vejamos agora formas de definir novas representações a partir de representações previamente conhecidas.

Soma direta. Dependendo do contexto, a soma direta $V \oplus W$ de dois espaços vetoriais é definida como soma direta interna ou soma direta externa. Caso V e W já sejam subespaços de um espaço maior e $V \cap W = \{0\}$, $V \oplus W$ (soma direta interna) é o espaço formado por todos os vetores da forma $v + w$, com $v \in V$ e $w \in W$; caso não valha $V \cap W = \{0\}$, falamos apenas na soma $V + W$. Caso V e W sejam espaços sem nenhuma relação predefinida, $V \oplus W$ (soma direta externa) é $V \times W$ com uma estrutura de espaço vetorial onde os elementos (v, w) são denotados por $v+w$; nesse caso $V \cap W = \{0\}$ é uma consequência da definição. No caso de espaços com produto interno, o produto interno na soma direta externa é definido de modo que os espaços V e W sejam ortogonais. Na soma direta interna os espaços V e W podem ou não ser ortogonais; quando é necessário ressaltar que a soma direta é ortogonal, usa-se a notação $V \perp W$.

Se $\rho\colon G \to \mathrm{GL}(V)$ e $\tau\colon G \to \mathrm{GL}(W)$ são duas representações de G, a *representação na soma direta* $\rho \oplus \tau = \omega\colon G \to \mathrm{GL}(V \oplus W)$ é definida por

$$\omega(g)(v + w) := \rho(g)(v) + \tau(g)(w),$$

para $g \in G$, $v \in V$ e $w \in W$. No caso de espaços de Hilbert de dimensão infinita e representações unitárias, a soma direta ortogonal enumerável também pode ser definida de forma semelhante.

Quociente. Se W é um subespaço de um espaço vetorial V, podemos definir o espaço quociente V/W como o espaço formado pelas classes de equivalência $v+W := \{v + w : w \in W\}$ e com as operações definidas de modo que a aplicação $v \mapsto v + W$ seja um homomorfismo.

Se $\rho\colon G \to \mathrm{GL}(V)$ é um representação de G e W é um subespaço invariante dessa representação, a *representação no espaço quociente* $\rho'\colon G \to \mathrm{GL}(V/W)$ é definida por

$$\rho'(g)(v + W) := \rho(g)v + W,$$

para $g \in G$ e $v \in V$. A representação está bem definida, isto é, temos que se $v + W = u + W$, então $\rho(g)v + W = \rho(g)u + W$ pois W é um subespaço invariante de ρ.

Proposição 1.3. *Se $\rho\colon G \to \mathrm{GL}(V)$ é um representação de G e V é soma direta de subespaços invariantes $V = W \oplus U$, então a subrepresentação ρ^U em U é equivalente à representação ρ' no espaço quociente V/W.*

Demonstração. Denotemos por A a transformação linear $A\colon U \to V/W$ dada por $Au = u + W$. Como V é soma direta entre W e U, temos que A é um isomorfismo entre U e V/W. O resultado segue assim que observarmos que A é um G-isomorfismo, de fato, para $u \in U$ e $g \in G$ temos:

$$A(\rho^U(g)u) = A(\rho(g)u) = \rho(g)u + W = \rho'(g)(u + W) = \rho'(g)(Au). \qquad \square$$

Dual. O espaço dual V^* de um espaço vetorial V é o espaço $\mathrm{Hom}(V, \mathbb{C})$ (no caso de espaços de Hilbert de dimensão infinita considera-se V', o dual topológico, constituído pela restrição de V^* aos homomorfismos contínuos). No caso de espaços de dimensão finita, se $\{v_1, \ldots, v_n\}$ é uma base de V, é conveniente considerarmos a base dual $\{v_1^*, \ldots, v_n^*\}$ correspondente em V^*, definida por $v_i^*(v_j) := \delta_{i,j}$.

Se $\rho\colon G \to \mathrm{GL}(V)$ é uma representação de G, a *representação no espaço dual* $\rho^* = \tau\colon G \to \mathrm{GL}(V^*)$ é definida por

$$(\tau(g)f)(v) := f(\rho(g^{-1})v),$$

para $g \in G$, $f \in V^*$ e $v \in V$. Com essa definição, temos que $(\tau(g)f)(\rho(g)v) = f(v)$ para $g \in G$, $f \in V^*$ e $v \in V$.

Produto tensorial. Apesar de o produto tensorial $V \otimes W$ poder ser definido para espaços de Hilbert (veja Reed e Simon [RS72] Capítulo II.4), aqui nos restringimos por simplicidade a espaços vetoriais de dimensão finita. Se V e W são dois espaços de dimensão finita, o produto tensorial $V \otimes W$ é definido como o espaço dual do espaço de todos os funcionais bilineares em $V \times W$. Para cada par de vetores $v \in V$ e $w \in W$, o produto tensorial $v \otimes w$ é o elemento de $V \otimes W$ definido por $(v \otimes w)(\varphi) := \varphi(v, w)$ para todo funcional bilinear $\varphi\colon V \times W \to \mathbb{C}$. Dessa forma, a aplicação $\otimes\colon V \times W \to V \otimes W$ é bilinear.

Temos também que se $\{v_1, \ldots, v_n\}$ é uma base de V e $\{w_1, \ldots, w_m\}$ é uma base de W, então $\{v_i \otimes w_j : i = 1, \ldots, n,\ j = 1, \ldots, m\}$ é uma base de $V \otimes W$; de fato, um funcional bilinear é determinado pelo seu valor nos pares (v_i, w_j), com $i = 1, \ldots, n$ e $j = 1, \ldots, m$, logo o conjunto de funcionais bilineares $\{\varphi_{p,q} : p = 1, \ldots, n, q = 1, \ldots, m\}$, onde $\varphi_{p,q}(v_i, w_j) := \delta_{p,i}\delta_{q,j}$, é uma base do espaço dos funcionais bilineares e o conjunto $\{v_i \otimes w_j : i = 1, \ldots, n, j = 1, \ldots, m\}$ pode ser identificado com a respectiva base dual, pois $(v_i \otimes w_j)(\varphi_{p,q}) = \varphi_{p,q}(v_i, w_j) = \delta_{p,i}\delta_{q,j} = \varphi_{i,j}^*(\varphi_{p,q})$ para todos $p, i = 1, \ldots, n$ e $q, j = 1, \ldots, m$.

No caso de espaços com produto interno, o produto interno em $V \otimes W$ é definido como $\langle e_{i_1} \otimes f_{j_1}, e_{i_2} \otimes f_{j_2} \rangle := \langle e_{i_1}, e_{i_2} \rangle \cdot \langle f_{j_1}, f_{j_2} \rangle$ e então é estendido por linearidade ao resto do espaço (assim a base com os vetores $e_i \otimes f_j$ é ortonormal, se as bases com os vetores e_i e f_j o forem em seus respectivos espaços). Fixando bases em cada espaço, não é difícil mostrar que a soma direta e o produto tensorial são associativos, comutativos e que vale a distributividade do produto sobre a soma (no sentido de isomorfismo entre espaços vetoriais).

Se $\rho\colon G \to \mathrm{GL}(V)$ e $\tau\colon G \to \mathrm{GL}(W)$ são duas representações de G, a *representação no produto tensorial* $\rho \otimes \tau = \psi\colon G \to \mathrm{GL}(V \otimes W)$ é definida por

$$\psi(g)(v \otimes w) := (\rho(g)v) \otimes (\tau(g)w)$$

para $g \in G$, $v \in V$ e $w \in W$ e estendida por linearidade. As propriedades associativa, comutativa e distributiva mencionadas para os espaços vetoriais também valem para as representações com essas definições.

Homomorfismos. Considerando V e W espaços de dimensão finita, temos que $V^* \otimes W$ pode ser identificado com $\mathrm{Hom}(V, W)$ definindo

$$(f \otimes w)(v) := f(v)w, \tag{4}$$

para $f \in V^*$ e $w \in W$. De fato, se $\{v_1, \ldots, v_n\}$ é uma base de V, $\{w_1, \ldots, w_m\}$ é uma base de W e $T \in \mathrm{Hom}(V, W)$ tal que $T(v_j) = \sum_{i=1}^m t_{ij} w_i$, então $\sum_{j=1}^n \sum_{i=1}^m t_{ij} v_j^* \otimes w_i = T$. Com essa identificação e as definições anteriores, se $\rho\colon G \to \mathrm{GL}(V)$ e $\tau\colon G \to \mathrm{GL}(W)$ são duas representações de G, obtemos uma representação $\rho^* \otimes \tau = \psi$ em $\mathrm{Hom}(V, W)$:

$$(\psi(g)T)(v) := \tau(g)T\rho(g^{-1})v.$$

O homomorfismo $\psi(g)T$ pode ser entendido como o homomorfismo que completa o seguinte diagrama

$$\begin{array}{ccc} V & \xrightarrow{\ T\ } & W \\ {\scriptstyle \rho(g)}\downarrow & & \downarrow{\scriptstyle \tau(g)} \\ V & \xrightarrow{\ \psi(g)T\ } & W \end{array} \tag{5}$$

Comparando com o diagrama (3), vemos que se T é um G-homomorfismo então

$$(\psi(g)T)v = \tau(g)T\rho(g)^{-1}v = T\rho(g)\rho(g)^{-1}v = Tv,$$

para todo $v \in V$ e $g \in G$ e portanto os G-homomorfismos são justamente os homomorfismos que ficam fixos sob essa representação. Já havíamos definido a notação $\mathrm{Hom}(V, W)^G$ para o espaço dos G-homomorfismos entre as representações V e W e a partir dessa observação podemos estender essa notação e usar V^G para o subespaço invariante de V que é fixado por uma certa representação, isto é:

$$V^G := \{v \in V : \rho(g)v = v \text{ para todo } g \in G\}. \tag{6}$$

1.1.2. Redutibilidade e Lema de Schur. Após termos visto operações para produzir novas representações, invertemos o ponto de vista e veremos que representações podem ser decompostas como soma direta de subrepresentações irredutíveis; tais representações são ditas *completamente redutíveis*. Restrigimos o estudo a espaços de dimensão finita, porém representações de grupos compactos em espaços de Hilbert também são completamente redutíveis (veja Folland [Fol16], Teorema 5.2).

Proposição 1.4. *Seja $\rho\colon G \to \mathrm{GL}(V)$ uma representação unitária em um espaço de dimensão finita e W um subespaço invariante. Então seu complemento ortogonal $W^\perp$ também é invariante e a representação em V é igual à soma direta de suas subrepresentações em W e $W^\perp$.*

Demonstração. Tome $x \in W^\perp$; desejamos mostrar que $\rho(g)x \in W^\perp$ para todo $g \in G$. Como a representação em V é unitária, temos que para qualquer $y \in W$,

$$\langle \rho(g)x, y\rangle = \langle x, \rho(g^{-1})y\rangle$$

e como W é um subespaço invariante, $\rho(g^{-1})y \in W$ e logo $\langle x, \rho(g^{-1})y\rangle = 0$. Portanto $W^\perp$ também é um subespaço invariante.

A afirmação sobre a representação em V ser igual à soma direta de suas subrepresentações segue diretamente da definição da representação na soma direta. □

Proposição 1.5. *Toda representação unitária em um espaço de dimensão finita pode ser escrita como soma direta ortogonal de subrepresentações irredutíveis.*

Demonstração. Seja $\rho\colon G \to \mathrm{GL}(V)$ uma representação unitária em um espaço de dimensão finita. Caso ρ seja irredutível, não há nada a provar, caso contrário, seja W_1 um subespaço próprio, não-nulo, invariante e irredutível. Pela Proposição 1.4, o complemento ortogonal de W_1 também é invariante, de modo que

obtemos $V = W_1 \perp U$ com W_1 irredutível. Caso U não seja irredutível, podemos repetir o argumento até obtermos a decomposição desejada. □

Se V é um espaço de representação decomposto como soma direta de subespaços invariantes e irredutíveis, podemos agrupar os subespaços cujas representações sejam equivalentes entre si. Obtemos:

$$\begin{aligned} V &= I_1 \perp \cdots \perp I_r, \\ I_k &= V_{k,1} \perp \cdots \perp V_{k,m_k}, \end{aligned} \tag{7}$$

onde os subespaços $V_{k,i}$ são invariantes e irredutíveis e tais que a subrepresentação em $V_{k,i}$ é equivalente à em $V_{l,j}$ se e somente se $k = l$. Os subespaços I_k são chamados de *componentes isotípicas* da representação.

Tal decomposição não é única em geral (considere por exemplo a representação $\rho\colon G \to \mathrm{GL}(V)$ em um espaço V de dimensão maior do que 1, tal que $\rho(g) = I$ para todo $g \in G$: nesse caso qualquer decomposição ortogonal em espaços de dimensão 1 serve), porém veremos no Teorema 1.15 que os números $m_1, \ldots, m_r$ são únicos e veremos no Teorema 1.17 que as componentes I_k também não dependem da decomposição escolhida.

A próxima proposição caracteriza os G-homomorfismos entre representações irredutíveis e possui diversas aplicações que serão vistas a seguir.

Proposição 1.6 (Lema de Schur). *Se $\rho\colon G \to \mathrm{GL}(V)$ e $\tau\colon G \to \mathrm{GL}(W)$ são duas representações irredutíveis de um grupo G em espaços de dimensão finita e $T \in \mathrm{Hom}(V, W)^G$, então:*

1. *T é um isomorfismo ou o homomorfismo nulo.*
2. *Se ρ e τ são equivalentes, então $\dim \mathrm{Hom}(V, W)^G = 1$. Em particular, se fixamos bases em V e W de modo que as matrizes de ρ e τ sejam iguais, então a matriz de T é igual à λI para algum $\lambda \in \mathbb{C}$.*

Demonstração. O primeiro item segue do fato do núcleo e da imagem de T, denotados por $\ker T$ e $\mathrm{Im}\, T$, serem subespaços invariantes de V e W, respectivamente. De fato, como T é um G-homomorfismo, temos que para todo $g \in G$,

$$T\rho(g) = \tau(g)T.$$

O núcleo de T é um subespaço invariante de V pois se $v \in \ker T$, então $T\rho(g)v = \tau(g)Tv = \tau(g)0 = 0$ e portanto $\rho(g)v \in \ker T$ para todo $g \in G$. A imagem de T é um subespaço invariante de W pois se $w \in \mathrm{Im}\, T$, então $w = Tv$ para algum $v \in V$ e $\tau(g)w = \tau(g)Tv = T\rho(g)v \in \mathrm{Im}\, T$ para todo $g \in G$. Como V e W são irredutíveis, ou $\ker T = V$ e $\mathrm{Im}\, T = \{0\}$ e T é o homomorfismo nulo, ou $\ker T = \{0\}$ e $\mathrm{Im}\, T = W$ e T é um isomorfismo.

O segundo item segue do fato de que se S e T são dois G-homomorfismos, então $S - \lambda T$ é um G-homomorfismo para todo $\lambda \in \mathbb{C}$. Como $\mathbb{C}$ é algebricamente fechado, temos que para algum λ o polinômio $\det(S - \lambda T) = 0$ e $S - \lambda T$ não pode ser um isomorfismo. Logo, pelo primeiro item, $S - \lambda T = 0$. □

Uma consequência interessante do Lema de Schur é a seguinte proposição, que diz que apesar de a dimensão do espaço de G-homomorfismos entre dois espaços equivalentes ser 1, se as representações forem unitárias, então existe um que preserva o produto interno entre os espaços. Tal fato é útil pois, se fixarmos uma base ortonormal em um dos espaços, podemos usar tal homomorfismo para determinar uma base ortonormal no outro espaço de modo que as matrizes das representações nos respectivos espaços sejam iguais. Na próxima seção usamos a existência de tais correspondências entre bases.

Proposição 1.7. *Sejam $\rho\colon G \to \mathrm{GL}(V)$ e $\tau\colon G \to \mathrm{GL}(W)$ duas representações unitárias, irredutíveis e equivalentes em espaços de dimensão finita. Então existe*

um G-isomorfismo $\phi\colon V \to W$ que preserva os produtos internos nos respectivos espaços de modo que $\langle u, v\rangle = \langle \phi u, \phi v\rangle$ para quaisquer $u, v \in V$.

Demonstração. Seja $\{e_1, \ldots, e_d\}$ uma base ortonormal de V e $T\colon V \to W$ um G-isomorfismo. Como pelo Lema de Schur, $\dim \mathrm{Hom}(V, W)^G = 1$, desejamos mostrar que

$$\langle Te_i, Te_j\rangle = \alpha\delta_{i,j},$$

para todos $i, j = 1, \ldots, d$ e algum $\alpha \in \mathbb{R}$, $\alpha > 0$; de modo que $\phi = \frac{1}{\sqrt{\alpha}}T$ seja o G-isomorfismo desejado.

Seja $A\colon V \to W$ a transformação linear definida para todo $j = 1, \ldots, d$ por

$$Ae_j := \sum_{i=1}^{d} \langle Te_j, Te_i\rangle\, Te_i.$$

Temos que A é um G-homomorfismo, pois para qualquer $i = 1, \ldots, d$ e $g \in G$:

$$\begin{aligned} A\rho(g)e_i &= \sum_{s=1}^{d} \rho_{s,i}(g) Ae_s = \sum_{s,t=1}^{d} \rho_{s,i}(g)\, \langle Te_s, Te_t\rangle\, Te_t = \sum_{t=1}^{d} \langle T\rho(g)e_i, Te_t\rangle\, Te_t \\ &= \sum_{t=1}^{d} \langle Te_i, T\rho(g^{-1})e_t\rangle\, Te_t = \sum_{s,t=1}^{d} \overline{\rho_{s,t}(g^{-1})}\, \langle Te_i, Te_s\rangle\, Te_t \\ &= \sum_{s=1}^{d} \langle Te_i, Te_s\rangle\, T\rho(g)e_s = \tau(g)Ae_i. \end{aligned}$$

Logo, pelo Lema de Schur segue que $A = \alpha T$ para algum $\alpha \in \mathbb{C}$. Temos então que $\langle Te_j, Te_i\rangle = \alpha\delta_{i,j}$ para $i, j = 1, \ldots, d$ e considerando $i = j$, como $\langle , \rangle$ é um produto interno, temos que $\alpha \in \mathbb{R}$ e $\alpha > 0$. □

1.1.3. Bloco-diagonalização de G-endomorfismos. A principal aplicação do Lema de Schur em que estamos interessados é uma descrição mais simples dos elementos de $\mathrm{End}(V)^G$; veremos que em uma certa base eles podem ser descritos através de matrizes bloco-diagonais.

Teorema 1.8. *Seja $\rho\colon G \to \mathrm{GL}(V)$ uma representação unitária em um espaço de dimensão finita. Então V possui uma base ortonormal onde a matriz de qualquer G-endomorfismo $T\colon V \to V$ é bloco-diagonal e composta por blocos $T_1, \ldots, T_r$, sendo que cada um desses blocos também é bloco-diagonal; cada bloco T_k é composto por d_k blocos iguais de tamanho $m_k \times m_k$. Os números $d_1, \ldots, d_r$ são as dimensões das r representações irredutíveis do grupo G e os números $m_1, \ldots, m_r$ são as correspondentes multiplicidades das componentes irredutíveis da decomposição de ρ em soma de subrepresentações irredutíveis.*

$$T = \begin{bmatrix} T_1 & & 0 \\ & \ddots & \\ 0 & & T_r \end{bmatrix},\ T_k = \begin{bmatrix} \Lambda_k & & 0 \\ & \ddots & \\ 0 & & \Lambda_k \end{bmatrix},\ \Lambda_k = \left[\lambda_{i,j}^k\right]_{i,j=1}^{m_k}.$$

Demonstração. Pela Proposição 1.5, podemos considerar V decomposto como soma direta ortogonal de subespaços irredutíveis:

$$\begin{aligned} V &= I_1 \perp \cdots \perp I_r, \\ I_k &= V_{k,1} \perp \cdots \perp V_{k,m_k}, \end{aligned}$$

com a subrepresentação de $V_{k,i}$ equivalente à de $V_{l,j}$ se e somente se $k = l$.

Usando a correspondência entre $\mathrm{End}(V)$ e $V^* \otimes V$, podemos ver um operador de V como um elemento de

$$\bigoplus_{k,l=1}^{r} \bigoplus_{i=1}^{m_k} \bigoplus_{j=1}^{m_l} V_{k,i}^* \otimes V_{l,j}.$$

Considerando a representação $\psi = \rho^* \otimes \rho$ nesse espaço, temos que cada um dos termos $V_{k,i}^* \otimes V_{l,j}$ é um subespaço invariante e portanto se $T \in (V^* \otimes V)^G$, então T é fixado por todo $\psi(g)$ e cada um dos componentes de T na soma anterior também é fixado. Pelo Lema de Schur, $\big(V_{k,i}^* \otimes V_{l,j}\big)^G$ é igual à $\{0\}$ se $k \neq l$ e possui dimensão 1 caso contrário.

Fixando bases ortonormais $\{e_{k,i,1}, \ldots, e_{k,i,d_k}\}$ em cada subespaço irredutível $V_{k,i}$ de modo que as matrizes das subrepresentações equivalentes sejam iguais, pelo segundo item do Lema de Schur temos que a matriz da componente de T em $\big(V_{k,i}^* \otimes V_{k,j}\big)^G$ é igual a $\lambda_{i,j}^k I$, com $\lambda_{i,j}^k \in \mathbb{C}$ para $i, j = 1, \ldots, m_k$ e $k = 1, \ldots, r$ (com I sendo a matriz identidade de ordem d_k).

Obtemos assim uma base de V em que ao ser ordenada em blocos $\{e_{k,1,s}, \ldots, e_{k,m_k,s}\}$ com $k = 1, \ldots, r$ e $s = 1, \ldots, d_k$, a matriz de T possui a estrutura descrita no enunciado. □

Esse teorema será usado na Seção 4.2, onde também será necessário um método para calcular a base descrita explicitamente, partindo das matrizes $\rho(g)$ da representação. Veremos como fazer isso no Teorema 1.21.

1.1.4. Caracteres e decomposição explícita de uma representação. A próxima aplicação do Lema de Schur será completar a caracterização da decomposição de uma representação em subrepresentações irredutíveis (Teoremas 1.15 e 1.17). Nesta seção voltamos também a considerar a representação regular (Exemplo 1.2) e vemos como os coeficientes das representações unitárias e irredutíveis de um grupo formam uma base ortonormal de $\mathbb{C}^G$ (Teorema 1.20). Por fim, vemos um método para calcular as bases dos subespaços invariantes a partir das matrizes da representação (Teorema 1.21), conforme motivado no fim da seção anterior.

Consideramos a seguir grupos finitos, porém a maior parte do que veremos estende-se para grupos compactos, conforme descrito no Capítulo 4 de Serre [Ser77], no Capítulo 5 de Folland [Fol16] ou na Seção 1.4 de Vilenkin [Vil68].

Seja $\rho\colon G \to \mathrm{GL}(V)$ uma representação de um grupo finito G em um espaço de dimensão finita. Ao fixarmos uma base, os coeficientes $\rho_{i,j}(g)$ das matrizes dessa representação são funções de G em $\mathbb{C}$. Vemos a seguir que os coeficientes das representações irredutíveis e unitárias são ortogonais em $\mathbb{C}^G$ (com o produto interno definido em (1)).

LEMA 1.9. *Sejam $\rho\colon G \to \mathrm{GL}(V)$ e $\tau\colon G \to \mathrm{GL}(W)$ duas representações irredutíveis de um grupo finito G em espaços de dimensões n e m respectivamente. Para $T_0 \in \mathrm{Hom}(V, W)$, considere*

$$T = \frac{1}{|G|} \sum_{g \in G} \tau(g^{-1}) T_0 \rho(g).$$

Então $T = 0$, se ρ e τ não forem equivalentes e $T = \big(\mathrm{tr}\, T_0/n\big)I$, caso $V = W$ e $\rho = \tau$.

DEMONSTRAÇÃO. Pode-se verificar que $T \in \mathrm{Hom}(V, W)^G$, logo pelo Lema de Schur, $T = 0$ se ρ e τ não forem equivalentes e $T = \lambda I$, caso $V = W$ e $\rho = \tau$. Podemos determinar λ calculando o traço de ambos os termos, $\mathrm{tr}\, T = \mathrm{tr}\, T_0$ e $\mathrm{tr}\,(\lambda I) = \lambda n$. □

PROPOSIÇÃO 1.10. *Sejam $\rho\colon G \to \mathrm{GL}(V)$, $\tau\colon G \to \mathrm{GL}(W)$ duas representações irredutíveis, não-equivalentes e unitárias de um grupo finito G em espaços de dimensões n e m respectivamente. Fixando bases ortonormais e denotando por $\rho_{i,j}(g)$ e $\tau_{i,j}(g)$ os coeficientes das matrizes dessas representações com relação a essas bases, temos:*

(1) *Para quaisquer $i_1, j_1 = 1, \ldots, n$ e $i_2, j_2 = 1, \ldots, m$,*
$$\langle \rho_{i_1,j_1}, \tau_{i_2,j_2} \rangle = 0.$$
(2) *Para quaisquer $i_1, j_1, i_2, j_2 = 1, \ldots, n$,*
$$\langle \rho_{i_1,j_1}, \rho_{i_2,j_2} \rangle = \delta_{i_1,i_2}\delta_{j_1,j_2}/n.$$

DEMONSTRAÇÃO. Sejam $\{e_1, \ldots, e_n\}$ e $\{f_1, \ldots, f_m\}$ as bases ortonormais de V e W. No caso do item (1), considere
$$T = \frac{1}{|G|} \sum_{g \in G} \tau(g^{-1})(e_{i_1}^* \otimes f_{i_2})\rho(g).$$
O resultado segue calculando o coeficiente t_{j_2,j_1} da matriz de T nas bases fixadas. Pelo Lema 1.9, $T = 0$ e logo $t_{j_2,j_1} = 0$. Por outro lado,
$$\begin{aligned} Te_{j_1} &= \frac{1}{|G|} \sum_{g \in G} \tau(g^{-1})(e_{i_1}^* \otimes f_{i_2})\rho(g)e_{j_1} \\ &= \frac{1}{|G|} \sum_{g \in G} \sum_{k=1}^{n} \rho_{k,j_1}(g)\tau(g^{-1})(e_{i_1}^* \otimes f_{i_2})e_k = \frac{1}{|G|} \sum_{g \in G} \rho_{i_1,j_1}(g)\tau(g^{-1})f_{i_2} \\ &= \frac{1}{|G|} \sum_{g \in G} \rho_{i_1,j_1}(g) \sum_{k=1}^{m} \tau_{k,i_2}(g^{-1})f_k = \sum_{k=1}^{m} \langle \rho_{i_1,j_1}, \tau_{i_2,k} \rangle f_k, \end{aligned}$$
e portanto,
$$0 = t_{j_2,j_1} = \langle \rho_{i_1,j_1}, \tau_{i_2,j_2} \rangle .$$
No caso do item (2), considere
$$T = \frac{1}{|G|} \sum_{g \in G} \rho(g^{-1})(e_{i_1}^* \otimes e_{i_2})\rho(g).$$
Novamente, o resultado segue calculando o coeficiente t_{j_2,j_1} da matriz de T na base fixada. Pelo Lema 1.9, $T = \big(\mathrm{tr}\,(e_{i_1}^* \otimes e_{i_2})/n\big)I = \big(\delta_{i_1,i_2}/n\big)I$ e logo $t_{j_2,j_1} = \delta_{i_1,i_2}\delta_{j_1,j_2}/n$. Por outro lado, de forma análoga ao cálculo anterior,
$$Te_{j_1} = \sum_{k=1}^{n} \langle \rho_{i_1,j_1}, \rho_{i_2,k} \rangle e_k,$$
e portanto,
$$\delta_{i_1,i_2}\delta_{j_1,j_2}/n = t_{j_2,j_1} = \langle \rho_{i_1,j_1}, \rho_{i_2,j_2} \rangle . \qquad \square$$

Introduzimos agora os caracteres das representações, que são ferramentas úteis para o estudo de representações em espaços de dimensão finita.

Seja $\rho : G \to \mathrm{GL}(V)$ uma representação de um grupo G em um espaço de dimensão finita V. O *caráter* da representação é definido como a função $\chi_\rho\colon G \to \mathbb{C}$,
$$\chi_\rho(g) := \mathrm{tr}\,\rho(g). \tag{8}$$

As primeiras propriedades do caráter seguem diretamente de propriedades da função traço. Em especial, o traço da matriz de um operador independe da base fixada no espaço (e com isso podemos falar no traço do operador, sem fazer menção a nenhuma base ou matriz), isso pode ser verificado através da identidade

$\operatorname{tr}(AB) = \operatorname{tr}(BA)$, satisfeita por quaisquer duas matrizes de dimensões compatíveis, e observando que se T e T' são as matrizes de um mesmo operador em bases distintas, então existe Q inversível tal que $T' = QTQ^{-1}$.

PROPOSIÇÃO 1.11. *Se χ_ρ é o caráter de uma representação $\rho\colon G \to \mathrm{GL}(V)$ de um grupo compacto em um espaço V de dimensão finita, então:*

(1) $\chi_\rho(e) = \dim V$ *(e é o elemento neutro de G),*
(2) $\chi_\rho(g^{-1}) = \overline{\chi_\rho(g)}$ *para todo $g \in G$,*
(3) $\chi_\rho(hgh^{-1}) = \chi_\rho(g)$ *para todos $g, h \in G$.*

DEMONSTRAÇÃO. Como V tem dimensão finita e como G é compacto, pela Proposição 1.1, podemos considerar V com um produto interno no qual a representação ρ é unitária. Consideremos as matrizes da representação relativas a uma base ortonormal.

Como a matriz de $\rho(e)$ é igual à matriz identidade em qualquer base, temos que $\chi_\rho(e) = \operatorname{tr} I = \dim V$. Também,

$$\chi_\rho(g^{-1}) = \sum_{i=1}^{n} \rho_{i,i}(g^{-1}) = \sum_{i=1}^{n} \overline{\rho_{i,i}(g)} = \overline{\chi_\rho(g)}.$$

E por fim,

$$\chi_\rho(hgh^{-1}) = \operatorname{tr}\big(\rho(h)\rho(g)\rho(h^{-1})\big) = \operatorname{tr}\rho(g) = \chi_\rho(g). \qquad \square$$

PROPOSIÇÃO 1.12. *Sejam $\rho\colon G \to \mathrm{GL}(V)$ e $\tau\colon G \to \mathrm{GL}(W)$ duas representações em espaços de dimensão finita:*

(1) *Se ρ e τ são equivalentes, então $\chi_\rho = \chi_\tau$.*
(2) *O caráter χ_ω da representação $\omega = \rho \oplus \tau$ em $V \oplus W$ é $\chi_\omega(g) = \chi_\rho(g) + \chi_\tau(g)$, para todo $g \in G$.*
(3) *O caráter χ_{ρ^*} da representação ρ^* em V^* é $\chi_{\rho^*}(g) = \overline{\chi_\rho(g)}$, para todo $g \in G$.*
(4) *O caráter χ_ψ da representação $\psi = \rho \otimes \tau$ em $V \otimes W$ é $\chi_\psi(g) = \chi_\rho(g)\chi_\tau(g)$, para todo $g \in G$.*

DEMONSTRAÇÃO. Se ρ e τ são equivalentes, então existe um G-isomorfismo T entre V e W e logo, para todo $g \in G$, temos

$$\chi_\rho(g) = \operatorname{tr}\rho(g) = \operatorname{tr}(T^{-1}\tau(g)T) = \operatorname{tr}\tau(g) = \chi_\tau(g).$$

Fixemos produtos internos em V e W de modo que as representações sejam unitárias e bases ortonormais $\{e_1, \ldots, e_n\}$ em V e $\{f_1, \ldots, f_m\}$ em W. No caso da soma direta, $\{e_1, \ldots, e_n\} \cup \{f_1, \ldots, f_m\}$ é uma base ortonormal de $V \oplus W$ e, para todo $g \in G$,

$$\begin{aligned}\chi_\omega(g) = \operatorname{tr}\omega(g) &= \sum_{i=1}^{n} \langle \omega(g)e_i, e_i\rangle + \sum_{f=1}^{m} \langle \omega(g)f_j, f_j\rangle \\ &= \sum_{i=1}^{n} \langle \rho(g)e_i, e_i\rangle + \sum_{f=1}^{m} \langle \tau(g)f_j, f_j\rangle = \operatorname{tr}\rho(g) + \operatorname{tr}\tau(g) = \chi_\rho(g) + \chi_\tau(g).\end{aligned}$$

No caso do espaço dual, a partir das definições de ρ^* e da base dual, temos para todos $i, j = 1, \ldots, n$ e $g \in G$,

$$\rho^*_{i,j}(g) = (\rho^*(g)e_j^*)(e_i) = e_j^*(\rho(g^{-1})e_i) = \rho_{j,i}(g^{-1}) = \overline{\rho_{i,j}(g)}$$

e logo para todo $g \in G$,

$$\chi_{\rho^*}(g) = \sum_{i=1}^{n} \rho^*_{i,i}(g) = \sum_{i=1}^{n} \overline{\rho_{i,i}(g)} = \overline{\chi_\rho(g)}.$$

No caso do produto tensorial, $\{e_i \otimes f_j : i = 1, \dots, n,\ j = 1, \dots, m\}$ é uma base ortonormal de $V \otimes W$ e, para todo $g \in G$,

$$\begin{aligned}\chi_\psi(g) = \operatorname{tr}\psi(g) &= \sum_{i=1}^{n}\sum_{f=1}^{m} \langle \psi(g)(e_i \otimes f_j), e_i \otimes f_j\rangle \\ &= \sum_{i=1}^{n}\sum_{f=1}^{m} \langle \rho(g)e_i \otimes \tau(g)f_j, e_i \otimes f_j\rangle = \sum_{i=1}^{n}\sum_{f=1}^{m} \langle \rho(g)e_i, e_i\rangle \langle \tau(g)f_j, f_j\rangle \\ &= \operatorname{tr}\rho(g)\operatorname{tr}\tau(g) = \chi_\rho(g)\chi_\tau(g). \quad \square\end{aligned}$$

O primeiro item da proposição anterior mostra que o caráter é uma função das classes de equivalência das representações. Veremos na Proposição 1.16 que essa função é injetiva, isto é, que se duas representações possuem o mesmo caráter, então são equivalentes. Portanto os caracteres caracterizam as classes de equivalência das representações.

Os caracteres das representações irredutíveis formam uma família ortonormal de funções:

Proposição 1.13. *Sejam $\rho\colon G \to \mathrm{GL}(V)$ e $\tau\colon G \to \mathrm{GL}(W)$ duas representações irredutíveis de um grupo finito G em espaços de dimensão finita.*

(1) *Se ρ e τ não são equivalentes, então $\langle \chi_\rho, \chi_\tau\rangle = 0$.*
(2) *$\langle \chi_\rho, \chi_\rho\rangle = 1$.*

Demonstração. Pela Proposição 1.1, podemos fixar produtos internos em V e W tais que ρ e τ sejam unitárias. Fixemos bases ortonormais e denotemos por $\rho_{i,j}(g)$ e $\tau_{i,j}(g)$ os coeficientes das matrizes das representações com relação a essas bases. Temos:

$$\begin{aligned}\langle \chi_\rho, \chi_\tau\rangle = \frac{1}{|G|}\sum_{g\in G}\chi_\rho(g)\overline{\chi_\tau(g)} &= \frac{1}{|G|}\sum_{g\in G}\sum_{i=1}^{n}\rho_{i,i}(g)\sum_{j=1}^{m}\overline{\tau_{j,j}(g)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{|G|}\sum_{g\in G}\rho_{i,i}(g)\overline{\tau_{j,j}(g)} = \sum_{i=1}^{n}\sum_{j=1}^{m}\langle \rho_{i,i}, \tau_{j,j}\rangle\end{aligned}$$

e então ambos os resultados seguem diretamente da Proposição 1.10. $\square$

Pelo terceiro item da Proposição 1.11, os caracteres são constantes dentro das classes de conjugação do grupo. Tais funções são chamadas *funções de classe* e formam um subespaço de $\mathbb{C}^G$.

Proposição 1.14. *Os caracteres das representações irredutíveis formam uma base ortonormal do espaço das funções de classe de um grupo finito G e portanto o número de representações irredutíveis não-equivalentes entre si é igual ao número de classes de conjugação de G.*

Demonstração. A Proposição 1.13 mostra que os caracteres formam um conjunto ortonormal; falta mostrar que eles geram o subespaço considerado.

Seja então $f \in \mathbb{C}^G$ uma função de classe tal que $\langle f, \chi_\rho\rangle = 0$ para qualquer representação ρ irredutível. Desejamos mostrar que $f = 0$.

Seja $\rho\colon G \to \mathrm{GL}(V)$ uma representação irredutível de G e considere

$$T = \sum_{g\in G}\overline{f(g)}\rho(g).$$

Temos que $T \in \mathrm{End}(V)^G$, pois para todo $h \in G$,

$$\rho(h)T = \sum_{g\in G}\overline{f(g)}\rho(hg) = \sum_{g\in G}\overline{f(h^{-1}gh)}\rho(gh) = \sum_{g\in G}\overline{f(g)}\rho(g)\rho(h) = T\rho(h).$$

Pelo Lema de Schur, $T = \lambda I$. Calculando o traço de T, obtemos

$$\operatorname{tr} T = \sum_{g \in G} \overline{f(g)} \chi_\rho(g) = |G| \langle \chi_\rho, f \rangle = 0,$$

e portanto $T = 0$. O mesmo argumento estende-se para qualquer representação em espaços de dimensão finita, pois pela Proposição 1.5 tais representações são completamente redutíveis e $T = 0$ em cada termo da decomposição.

Consideremos ρ sendo a representação regular de G e a base $\{\phi_g : g \in G\}$ de $G^{\mathbb{C}}$, definidas no Exemplo 1.2. Aplicando T em ϕ_e, obtemos

$$0 = T\phi_e = \sum_{g \in G} \overline{f(g)} \rho(g) \phi_e = \sum_{g \in G} \overline{f(g)} \phi_g.$$

Como $\{\phi_g : g \in G\}$ é linearmente independente em $\mathbb{C}^G$, concluimos que $f = 0$. □

Temos agora as ferramentas necessárias para a demonstração dos teoremas prometidos sobre a decomposição de uma representação como soma de subrepresentações irredutíveis.

Teorema 1.15. *Sejam G um grupo finito e $\rho^k \colon G \to \mathrm{GL}(V_k)$, com $k = 1, \ldots, r$, representantes das classes de representações irredutíveis do grupo. Denotemos por χ_k os caracteres dessas representações.*

Se $\rho \colon G \to \mathrm{GL}(V)$ é uma representação em um espaço de dimensão finita com decomposição em soma de subrepresentações irredutíveis:

$$V = I_1 \oplus \cdots \oplus I_r,$$

$$I_k = V_{k,1} \oplus \cdots \oplus V_{k,m_k},$$

de modo que a subrepresentação em $V_{k,i}$ é equivalente a ρ^k para todo $k = 1, \ldots, r$ e $i = 1, \ldots, m_k$, então $m_k = \langle \chi_\rho, \chi_k \rangle$ e os números m_k não dependem da decomposição particular.

Demonstração. Pelo segundo item da Proposição 1.12, podemos expressar χ_ρ como combinação linear dos caracteres das representações irredutíveis:

$$\chi_\rho = m_1 \chi_1 + \cdots + m_r \chi_r.$$

O resultado segue diretamente das relações de ortogonalidade vistas na Proposição 1.13. □

Proposição 1.16. *Se duas representações em espaços de dimensão finita possuem o mesmo caráter, então são equivalentes.*

Demonstração. Se duas representações têm o mesmo caráter, então pelo Teorema 1.15, elas podem ser decompostas em termos de representações irredutíveis de maneira semelhante e logo são equivalentes. □

Teorema 1.17. *Seja $\rho \colon G \to \mathrm{GL}(V)$ uma representação de um grupo finito G em um espaço de dimensão finita com decomposição em subrepresentações irredutíveis nos mesmos termos descritos no Teorema 1.15.*

Então a projeção p_k de V na componente isotípica I_k é dada pela fórmula

$$p_k = \frac{d_k}{|G|} \sum_{g \in G} \overline{\chi_k(g)} \rho(g),$$

com $d_k := \dim V_k$. Portanto as componentes isotípicas não dependem da decomposição em subrepresentações irredutíveis escolhida.

Demonstração. Como p_k é uma combinação linear dos operadores $\rho(g)$, temos que cada subespaço $V_{l,i}$ é um subespaço invariante de p_k.

Observe que $p_k \in \mathrm{End}(V)^G$, pois para todo $h \in G$,

$$\rho(h)p_k = \frac{d_k}{|G|}\sum_{g\in G}\overline{\chi_k(g)}\rho(hg) = \frac{d_k}{|G|}\sum_{g\in G}\overline{\chi_k(h^{-1}gh)}\rho(gh)$$
$$= \frac{d_k}{|G|}\sum_{g\in G}\overline{\chi_k(g)}\rho(g)\rho(h) = p_k\rho(h).$$

Fixando um subespaço invariante e irredutível $V_{l,i}$, pelo Lema de Schur, temos que a restrição de p_k é igual a λI. Comparando os traços, obtemos

$$\lambda d_l = \mathrm{tr}\,(\lambda I) = \mathrm{tr}\,p_k = \frac{d_k}{|G|}\sum_{g\in G}\overline{\chi_k(g)}\chi_l(g) = d_k\,\langle\chi_l,\chi_k\rangle$$

e pela Proposição 1.13, $p_k = \delta_{k,l}I$ em $V_{l,i}$.

Observando p_k em todos os termos da decomposição, vemos que p_k é a identidade em I_k e nulo em cada I_l com $l \neq k$. Portanto, p_k é a projeção de V em I_k. □

Proposição 1.18. *Se a representação $\rho\colon G \to \mathrm{GL}(V)$ de um grupo finito em um espaço de dimensão finita é unitária, então suas componentes isotípicas são ortogonais.*

Demonstração. Pela Proposição 1.5, o espaço V possui uma decomposição como soma direta ortogonal de subrepresentações irredutíveis. Usando essa decomposição ortogonal vemos que as componentes isotípicas também são ortogonais entre si. □

Calculando o caráter da representação regular podemos estabelecer mais resultados sobre as representações irredutíveis de um grupo finito.

Proposição 1.19. *Sejam χ_k, com $k = 1,\ldots,r$, os caracteres das representações irredutíveis de um grupo finito G e d_k as dimensões dos respectivos espaços de representação. Então:*

(1) *As dimensões d_k satisfazem $\sum_{k=1}^{r} d_k^2 = |G|$.*
(2) *Para todo $g \in G$, $g \neq e$, temos $\sum_{k=1}^{r} d_k\chi_k(g) = 0$.*

Demonstração. Sejam ρ a representação regular de G e $\{\phi_g : g \in G\}$ a base de $\mathbb{C}^G$, definidas no Exemplo 1.2. Para calcular χ_ρ, basta observar que a representação atua permutando os elementos da base e que se $g \neq e$, então $gh \neq h$ para todo $h \in G$. Assim, os termos diagonais das matrizes $\rho(g)$ nessa base são zero para todo $g \neq e$. Logo,

$$\chi_\rho(e) = \dim\mathbb{C}^G = |G| \quad \text{e} \quad \chi_\rho(g) = 0, \text{ para } g \in G\setminus\{e\}.$$

Pela Proposição 1.14, χ_ρ pode ser expresso como combinação linear dos caracteres χ_k. Como

$$\langle\chi_\rho,\chi_k\rangle = \frac{1}{|G|}\sum_{g\in G}\chi_\rho(g)\overline{\chi_k(g)} = \overline{\chi_k(e)} = d_k,$$

temos

$$\chi_\rho = d_1\chi_1 + \cdots + d_r\chi_r.$$

Aplicando essa fórmula em e, obtemos o primeiro item e aplicando em $g \in G$, $g \neq e$, obtemos o segundo. □

O próximo teorema mostra que os coeficientes das matrizes das representações unitárias e irredutíveis do grupo formam uma base ortonormal de $\mathbb{C}^G$. Como mencionado no início desta seção, a maior parte das proposições vistas podem ser estendidas para grupos compactos. No caso de $\mathbb{C}^G$, esse espaço é substituído pelo

espaço de Hilbert $L^2(G)$ e o teorema correspondente (conhecido como Teorema de Peter-Weyl — veja o Teorema 1 da Subeção 1.4.3 de Vilenkin [Vil68]) diz que essas funções formam um sistema ortonormal completo desse espaço.

TEOREMA 1.20. *Sejam $\rho^k\colon G \to \mathrm{GL}(V_k)$, com $k = 1, \dots, r$ as representações unitárias e irredutíveis de um grupo finito G e $\rho^k_{i,j}(g)$ os coeficientes das matrizes dessas representações com relação a bases ortonormais fixadas nesses espaços. As funções $\sqrt{d_k}\rho^k_{i,j}$, com $d_k := \dim V_k$, $k \in r$ e $i, j = 1, \dots, d_k$, formam uma base ortonormal de $\mathbb{C}^G$.*

DEMONSTRAÇÃO. Pela Proposição 1.10, essas funções formam um conjunto ortonormal. Pelo item 1 da Proposição 1.19, elas geram $\mathbb{C}^G$. □

As fórmulas de projeção dadas no Teorema 1.17 são úteis para construir explicitamente a decomposição de uma representação em componentes isotípicas. Não podemos esperar fórmulas semelhantes para a decomposição das componentes isotípicas em subrepresentações irredutíveis, visto que tal decomposição não é única, mas vemos a seguir um método que produz tal decomposição a partir da escolha de uma base para um certo subespaço da componente isotípica.

TEOREMA 1.21. *Seja $\rho\colon G \to \mathrm{GL}(V)$ uma representação unitária de um grupo finito G em um espaço de dimensão finita com decomposição em componentes isotípicas*

$$V = I_1 \perp \cdots \perp I_r.$$

Para $k = 1, \dots, r$, seja $\rho^k\colon G \to \mathrm{GL}(V_k)$ uma representação unitária e irredutível associada à k-ésima componente da decomposição anterior. Sejam d_k a dimensão de V_k e $\rho^k_{\alpha,\beta}(g)$ os coeficientes das matrizes dessa representação com relação a uma base ortonormal de V_k. Para $\alpha, \beta = 1, \dots, d_k$, considere o operador linear $p_{\alpha,\beta}\colon V \to V$:

$$p_{\alpha,\beta} := \frac{d_k}{|G|} \sum_{g \in G} \overline{\rho^k_{\alpha,\beta}(g)}\rho(g).$$

Então:

1. *O operador $p_{\alpha,\alpha}$ é uma projeção, sua imagem W_α está contida em I_k e $I_k = W_1 \perp \cdots \perp W_{d_k}$.*
2. *O mapa $p_{\alpha,\beta}$ é nulo em todo I_l com $l \neq k$ e em todo W_γ com $\gamma \neq \beta$; é um isomorfismo entre W_β e W_α.*
3. *Seja $\{x_1^{(1)}, \dots, x_1^{(m_k)}\}$ uma base ortonormal de W_1. Para $i = 1, \dots, m_k$ e $\alpha = 1, \dots, d_k$, seja $x_\alpha^{(i)} := p_{\alpha,1}x_1^{(i)}$. Então $\{x_1^{(i)}, \dots, x_{d_k}^{(i)}\}$ é uma base ortonormal de um subespaço $V_{k,i}$ tal que:*
   (a) *$V_{k,i}$ é um subespaço invariante da representação ρ e para todo $g \in G$, a matriz de $\rho^{V_{k,i}}(g)$ é igual a matriz de $\rho^k(g)$ nas suas respectivas bases já fixadas ($\rho^{V_{k,i}}$ é equivalente a ρ^k).*
   (b) *$I_k = V_{k,1} \perp \cdots \perp V_{k,m_k}$.*

DEMONSTRAÇÃO. Para $l \neq k$, consideremos uma decomposição da componente I_l em subespaços invariantes e irredutíveis

$$I_l = V_{l,1} \oplus \cdots \oplus V_{l,m_l}.$$

Como $p_{\alpha,\beta}$ é combinação linear dos operadores $\rho(g)$, os subespaços $V_{l,i}$ são invariantes para $p_{\alpha,\beta}$ e, fixando uma base ortonormal em cada subespaço $V_{l,i}$, vemos que os coeficientes da matriz de $p_{\alpha,\beta}|_{V_{l,i}}$ são nulos, pois pelo primeiro item da Proposição 1.10,

$$(p_{\alpha,\beta}|_{V_{l,i}})_{s,t} = d_k\langle \rho^k_{\alpha,\beta}, \rho^{V_{l,i}}_{s,t}\rangle = 0$$

e portanto $p_{\alpha,\beta}$ é nulo em todo I_l com $l \neq k$.

Para todos $\alpha, \beta, \gamma, \epsilon = 1, \ldots, d_k$, os operadores definidos no enunciado satisfazem

$$p_{\alpha,\beta} p_{\gamma,\epsilon} = \begin{cases} p_{\alpha,\epsilon} & \text{se } \beta = \gamma, \\ 0 & \text{caso contrário.} \end{cases}$$

De fato,

$$\begin{aligned} p_{\alpha,\beta} p_{\gamma,\epsilon} &= \frac{d_k^2}{|G|^2} \sum_{g,h \in G} \overline{\rho^k_{\alpha,\beta}(g) \rho^k_{\gamma,\epsilon}(h)} \rho(gh) = \frac{d_k^2}{|G|^2} \sum_{g,h \in G} \overline{\rho^k_{\alpha,\beta}(g) \rho^k_{\gamma,\epsilon}(g^{-1}h)} \rho(h) \\ &= \frac{d_k^2}{|G|^2} \sum_{g,h \in G} \overline{\rho^k_{\alpha,\beta}(g)} \rho^k_{\epsilon,\gamma}(h^{-1}g) \rho(h) \\ &= \frac{d_k^2}{|G|^2} \sum_{g,h \in G} \overline{\rho^k_{\alpha,\beta}(g)} \sum_{\zeta=1}^{d_k} \rho^k_{\epsilon,\zeta}(h^{-1}) \rho^k_{\zeta,\gamma}(g) \rho(h) \\ &= \frac{d_k^2}{|G|^2} \sum_{h \in G} \sum_{\zeta=1}^{d_k} \overline{\rho^k_{\zeta,\epsilon}(h)} \Big(\sum_{g \in G} \overline{\rho^k_{\alpha,\beta}(g)} \rho^k_{\zeta,\gamma}(g) \Big) \rho(h) \\ &= \frac{d_k^2}{|G|^2} \sum_{h \in G} \sum_{\zeta=1}^{d_k} \overline{\rho^k_{\zeta,\epsilon}(h)} \left(\frac{|G|}{d_k} \delta_{\alpha,\zeta} \delta_{\beta,\gamma} \right) \rho(h) = \delta_{\beta,\gamma} \frac{d_k}{|G|} \sum_{h \in G} \overline{\rho^k_{\alpha,\epsilon}(h)} \rho(h) \\ &= \delta_{\beta,\gamma} p_{\alpha,\epsilon}. \end{aligned}$$

Logo $p^2_{\alpha,\alpha} = p_{\alpha,\alpha}$ e $p_{\alpha,\alpha}$ é uma projeção com imagem contida em I_k. Pelo Teorema 1.17, $\sum_{\alpha=1}^{d_k} p_{\alpha,\alpha}$ é a projeção sobre I_k e então $I_k = W_1 + \cdots + W_{d_k}$. Para ver que a soma é direta e ortogonal, note que como os operadores $p_{\alpha,\beta}$ são definidos como uma combinação linear dos operadores $\rho(g)$, eles também preservam o produto interno de V e portanto se $\alpha \neq \beta$, $v \in W_\alpha$ e $u \in W_\beta$,

$$\langle v, u \rangle = \langle p_{\alpha,\alpha} v, p_{\beta,\beta} u \rangle = \langle p_{\beta,\beta} p_{\alpha,\alpha} v, p_{\beta,\beta} p_{\beta,\beta} u \rangle = 0,$$

pois $p_{\beta,\beta} p_{\alpha,\alpha} = 0$. Isso conclui a demonstração do item (1).

Como $p_{\alpha,\beta} p_{\gamma,\gamma} = 0$ se $\gamma \neq \beta$, o mapa $p_{\alpha,\beta}$ é nulo nos subespaços W_γ, com $\gamma \neq \beta$ e $\dim \operatorname{Im} p_{\alpha,\beta} \leq \dim W_\beta$. Como $p_{\alpha,\beta} p_{\beta,\alpha} = p_{\alpha,\alpha}$, temos que $W_\alpha \subseteq \operatorname{Im} p_{\alpha,\beta}$ e $\dim W_\alpha \leq \dim W_\beta$. Trocando α por β, temos $\dim W_\alpha = \dim W_\beta$ e portanto $\operatorname{Im} p_{\alpha,\beta} = W_\alpha$. Isso conclui a demonstração do item (2).

Que $\{x_1^{(i)}, \ldots, x_{d_k}^{(i)}\}$ é um conjunto ortonormal, segue do fato de I_k ser soma direta ortogonal dos subespaços W_α e $x_\alpha^{(i)} \in W_\alpha$ para todo $\alpha = 1, \ldots, d_k$.

A seguinte identidade é satisfeita para todo $h \in G$ e $\beta, \gamma = 1, \ldots, d_k$:

$$\rho(h) p_{\beta,\gamma} = \sum_{\alpha=1}^{d_k} \rho^k_{\alpha,\beta}(h) p_{\alpha,\gamma}.$$

De fato,

$$\begin{aligned} \rho(h) p_{\beta,\gamma} &= \frac{d_k}{|G|} \sum_{g \in G} \overline{\rho^k_{\beta,\gamma}(g)} \rho(hg) = \frac{d_k}{|G|} \sum_{g \in G} \overline{\rho^k_{\beta,\gamma}(h^{-1}g)} \rho(g) \\ &= \frac{d_k}{|G|} \sum_{g \in G} \rho^k_{\gamma,\beta}(g^{-1}h) \rho(g) = \frac{d_k}{|G|} \sum_{g \in G} \sum_{\alpha=1}^{d_k} \rho^k_{\gamma,\alpha}(g^{-1}) \rho^k_{\alpha,\beta}(h) \rho(g) \\ &= \sum_{\alpha=1}^{d_k} \rho^k_{\alpha,\beta}(h) \frac{d_k}{|G|} \sum_{g \in G} \overline{\rho^k_{\alpha,\gamma}(g)} \rho(g) = \sum_{\alpha=1}^{d_k} \rho^k_{\alpha,\beta}(h) p_{\alpha,\gamma} \end{aligned}$$

e disso segue o item (3) (a), já que para todo $g \in G$, $i = 1, \dots, m_k$ e $\beta = 1, \dots, d_k$,

$$\rho(g)x_\beta^{(i)} = \rho(g)p_{\beta,1}x_1^{(i)} = \sum_{\alpha=1}^{d_k} \rho_{\alpha,\beta}^k(g)p_{\alpha,1}x_1^{(i)} = \sum_{\alpha=1}^{d_k} \rho_{\alpha,\beta}^k(g)x_\alpha^{(i)}.$$

Para mostrar o item (3) (b), note que para $i, j = 1, \dots, m_k$, $i \neq j$, temos $V_{k,i} \perp V_{k,j}$. De fato, para quaisquer $\alpha, \beta = 1, \dots, d_k$, caso $\alpha \neq \beta$, temos $\langle x_\alpha^{(i)}, x_\beta^{(j)} \rangle = 0$ pois $x_\alpha^{(i)} \in W_\alpha$ e $x_\beta^{(i)} \in W_\beta$; caso $\alpha = \beta$,

$$\langle x_\alpha^{(i)}, x_\alpha^{(j)} \rangle = \langle p_{\alpha,1}x_1^{(i)}, p_{\alpha,1}x_1^{(j)} \rangle = \langle x_1^{(i)}, x_1^{(j)} \rangle = 0,$$

pois o conjunto $\{x_1^{(1)}, \dots, x_1^{(m_k)}\}$ foi escolhido ortonormal. A igualdade afirmada no item (3) (b) segue do fato de ambos os espaços terem dimensão $d_k m_k$. □

1.2. Kernels positivos, contínuos e invariantes

Uma matriz em $\mathbb{C}^{n\times n}$ pode ser vista tanto como uma função $\{1, \dots, n\} \times \{1, \dots, n\} \to \mathbb{C}$ quanto como um operador linear em $\mathbb{C}^n$. Nesta seção substituímos $\{1, \dots, n\}$ por um espaço topológico compacto de Hausdorff X. Consideramos funções complexas, pois na Seção 1.2.2 usamos a teoria de representações desenvolvida na última seção, entretanto todas as definições podem ser consideradas de forma análoga no caso real. Uma referência mais completa para kernels em espaços de Hilbert é o Capítulo VI de Reed e Simon [RS72]. Já a seção sobre kernels invariantes segue a Seção 3.3 de Bachoc, Gijswijt, Schrijver e Vallentin [BGSV12].

1.2.1. Kernels positivos e contínuos. Seja X um espaço topológico de Hausdorff compacto e ω uma medida de Radon (i.e., definida em todos os conjuntos de Borel, finita em compactos e internamente regular) estritamente positiva de X (i.e., todo conjunto aberto e não-vazio possui medida estritamente positiva), normalizada de modo que $\omega(X) = 1$. Consideramos o espaço de Hilbert $L^2(X)$ das funções complexas com quadrado integrável em X (na verdade, classes de equivalência, em que duas funções f e g são consideradas equivalentes se $\int_X \big(f(x)-g(x)\big)^2 \,\mathrm{d}\omega(x) = 0$) com o produto interno

$$\langle f, g \rangle = \int_X f(x)\overline{g(x)} \,\mathrm{d}\omega(x). \tag{9}$$

Um *kernel de Hilbert-Schmidt* é uma função $K \in L^2(X \times X)$. A um kernel associamos um *operador integral de Hilbert-Schmidt*, que é um operador da forma

$$T_K \colon L^2(X) \to L^2(X), \quad (T_K f)(x) = \int_X K(x, y)f(y) \,\mathrm{d}\omega(y).$$

Tal como uma matriz pode ser positivo-semidefinida, dizemos que um kernel K é *positivo* se seu operador associado for auto-adjunto e positivo, isto é, se para todo $f, g \in L^2(X)$, temos $\langle T_K f, g \rangle = \langle f, T_K g \rangle$ e $\langle T_K f, f \rangle \geq 0$.

Consideramos também funções contínuas $\mathcal{C}(X)$ e kernels contínuos $\mathcal{C}(X \times X)$, que de forma semelhante definem operadores integrais em $\mathcal{C}(X)$. Como estamos supondo X compacto, esses espaços são subespaços densos de $L^2(X)$ e $L^2(X \times X)$ (Teorema 3.14 de Rudin [Rud87]). Dizemos que um kernel contínuo é *hermitiano* se $K(x, y) = \overline{K(y, x)}$ para todo $x, y \in X$; o conjunto dos kernels contínuos e hermitianos é denotado por $\mathcal{C}(X \times X)_{\mathrm{her}}$ (no caso real trocamos o termo "hermitiano" por "simétrico" e usamos a notação $\mathcal{C}(X \times X)_{\mathrm{sym}}$), com a continuidade, a condição sobre o kernel ser hermitiano é equivalente a de seu operador associado ser auto-adjunto.

Denotamos o conjunto dos kernels contínuos e positivos por $\mathcal{C}(X \times X)_{\succeq 0}$. Esse conjunto possui uma caracterização muito útil que apresentamos a seguir como uma proposição, cuja demonstração pode ser encontrada em Bochner [Boc41] ou adaptada a partir da Proposição 3.35 de Folland [Fol16].

PROPOSIÇÃO 1.22. *Um kernel contínuo e hermitiano K é positivo se e somente se para qualquer subconjunto finito $\{x_1, ..., x_m\}$ de X, a matriz $\big(K(x_i, x_j)\big)_{i,j=1}^m$ é positivo-semidefinida.*

1.2.2. Kernels invariantes. Supomos agora que existe um grupo compacto G que age continuamente em X e que ω é G-invariante (i.e., $\omega(gA) = \omega(A)$ para todo $g \in G$ e A subconjunto mensurável de X). A ação de G em X induz uma representação unitária L em $L^2(X)$ definida por

$$(L(g)f)(x) := f(g^{-1}x). \tag{10}$$

Dizemos que um kernel K é *invariante* se seu operador associado T_K comuta com a representação de modo a ser um G-endomorfismo. Com a hipótese sobre a continuidade da ação, podemos restringir a representação à $\mathcal{C}(X)$ e se $K \in \mathcal{C}(X \times X)$, a condição de invariância é equivalente a dizer que $K(gx, gy) = K(x, y)$ para todos $x, y \in X$ e $g \in G$. De forma semelhante ao definido em (6), usamos as notações $L^2(X \times X)^G$, $\mathcal{C}(X \times X)^G_{\text{her}}$, $\mathcal{C}(X \times X)^G_{\succeq 0}$, etc, para os respectivos espaços de kernels invariantes.

Seja V um subespaço invariante de $\mathcal{C}(X)$ de dimensão finita e $V^{(2)}$ o subespaço de $\mathcal{C}(X \times X)$ gerado pelos kernels da forma $(x, y) \mapsto f_1(x)\overline{f_2(y)}$, com $f_1, f_2 \in V$. Note que se $\{v_1, \dots, v_n\}$ é uma base ortonormal de V, o operador associado ao kernel $(x, y) \to v_i(x)\overline{v_j(y)}$ corresponde ao operador $v_j^* \otimes v_i$ na notação da Seção 1.1.1. De fato, se $K(x, y) = v_i(x)\overline{v_j(y)}$ e $f \in V$, temos

$$(T_K f)(x) = \int_X v_i(x)\overline{v_j(y)} f(y)\, \mathrm{d}\omega(y) = \langle f, v_j \rangle\, v_i(x).$$

Logo os kernels de $V^{(2)}$ correspondem justamente aos operadores de $V^* \otimes V$.

Nos Capítulos 2 e 3 estamos interessados em descrever kernels contínuos, positivos e invariantes. Com a correspondência entre os kernels de $V^{(2)}$ e os operadores de $V^* \otimes V$, isso pode ser feito com o método de bloco-diagonalização de G-endomorfismos da Seção 1.1.3, conforme descrito a seguir.

PROPOSIÇÃO 1.23. *Considere V decomposto como soma ortogonal de subespaços irredutíveis*

$$V = I_1 \perp \cdots \perp I_r,$$
$$I_k = V_{k,1} \perp \cdots \perp V_{k,m_k},$$

com a subrepresentação de $V_{k,i}$ equivalente a de $V_{l,j}$ se e somente se $k = l$. Para cada $k = 1, \dots, r$ e $i = 1, \dots, m_k$, seja $\{e_{k,i,1}, \dots, e_{k,i,d_k}\}$ uma base ortonormal de $V_{k,i}$ tal que as matrizes das subrepresentações de G em subespaços equivalentes sejam iguais. Seja E_k a matriz de tamanho $m_k \times m_k$ cujas entradas são os kernels

$$\big(E_k\big)_{i,j}(x, y) = \sum_{s=1}^{d_k} e_{k,i,s}(x)\overline{e_{k,j,s}(y)}.$$

Um kernel $K \in V^{(2)}$ é positivo e invariante se e somente se existem matrizes positivo-semidefinidas $\Lambda_k = \big(\lambda^k_{i,j}\big)$ tais que

$$K = \sum_{k=1}^{r} \sum_{i,j=1}^{m_k} \lambda^k_{i,j} \big(E_k\big)_{i,j} = \sum_{k=1}^{r} \left\langle \Lambda_k, \overline{E_k} \right\rangle.$$

DEMONSTRAÇÃO. Segue do Lema de Schur e do método de bloco-diagonalização descrito no Teorema 1.8 que os kernels $\big(E_k\big)_{i,j}$ geram o espaço dos kernels invariantes. Basta observar que como, pela escolha das bases, a matriz de $L(g)$ em cada

subespaço $V_{k,i}$ é unitária e não depende de i, temos que $\big(E_k\big)_{i,j}(g^{-1}x, g^{-1}y) = \big(E_k\big)_{i,j}(x,y)$ e logo $T_{(E_k)_{i,j}} \in \big(V_{k,j}^* \otimes V_{k,i}\big)^G$.

A equivalência entre K ser um kernel positivo e as matrizes Λ_k serem positivo-semidefinidas segue de que para quaisquer $g, h \in \mathcal{C}(X)$,

$$\begin{aligned}\langle T_K g, h\rangle &= \int_X \int_X K(x,y) g(y) \overline{h(x)}\, \mathrm{d}\omega(x)\, \mathrm{d}\omega(y) \\ &= \sum_{k=1}^{r} \sum_{i,j=1}^{m_k} \lambda_{i,j}^k \sum_{s=1}^{d_k} \int_X \int_X e_{k,i,s}(x) \overline{e_{k,j,s}(y)} g(y) \overline{h(x)}\, \mathrm{d}\omega(x)\, \mathrm{d}\omega(y) \\ &= \sum_{k=1}^{r} \sum_{s=1}^{d_k} \sum_{i,j=1}^{m_k} \lambda_{i,j}^k \langle e_{k,i,s}, h\rangle \, \overline{\langle e_{k,j,s}, g\rangle},\end{aligned}$$

logo fazendo $g = h$,

$$\langle T_K g, g\rangle = \sum_{k=1}^{r} \sum_{s=1}^{d_k} \sum_{i,j=1}^{m_k} \lambda_{i,j}^k \langle e_{k,i,s}, g\rangle \, \overline{\langle e_{k,j,s}, g\rangle}.$$

Se as matrizes Λ_k forem positivo-semidefinidas, então para qualquer $g \in \mathcal{C}(X)$, o termo à direita na equação anterior é não-negativo e portanto K é um kernel positivo. Por outro lado, se K é um kernel positivo o termo à esquerda é não-negativo para todo $g \in \mathcal{C}(X)$ e para cada $k = 1, \ldots, r$ e $\alpha_1, \ldots, \alpha_{m_k} \in \mathbb{C}$, existe $g \in I_k$ que faz o termo à direita ser igual a $\sum_{i,j=1}^{m_k} \lambda_{i,j}^k \alpha_i \overline{\alpha_j}$, de modo que a matriz Λ_k é positivo-semidefinida. □

O teorema seguinte é muito semelhante à proposição anterior, porém possui hipóteses mais simples, o que facilita sua aplicação na Proposição 3.2 do Capítulo 3.

Teorema 1.24. *Considere V decomposto como soma direta de subespaços irredutíveis,*

$$\begin{aligned}V &= I_1 \perp \cdots \perp I_r, \\ I_k &= S_{k,1} \oplus \cdots \oplus S_{k,m_k},\end{aligned}$$

com a subrepresentação de $S_{k,i}$ equivalente a de $S_{l,j}$ se e somente se $k = l$. Para cada $k = 1, \ldots, r$ e $i = 1, \ldots, m_k$, seja $\{f_{k,i,1}, \ldots, f_{k,i,d_k}\}$ uma base ortogonal de $S_{k,i}$ com vetores cujas normas dependem apenas de k e i e tais que as matrizes da subrepresentação de G em subespaços equivalentes sejam iguais. Seja F_k a matriz de tamanho $m_k \times m_k$ cujas entradas são os kernels

$$\big(F_k\big)_{i,j}(x,y) := \sum_{s=1}^{d_k} f_{k,i,s}(x) \overline{f_{k,j,s}(y)}.$$

Um kernel $K \in V^{(2)}$ é positivo e invariante se e somente se existem matrizes positivo-semidefinidas $\Lambda_k = \big(\lambda_{i,j}^k\big)$ tais que

$$K = \sum_{k=1}^{r} \sum_{i,j=1}^{m_k} \lambda_{i,j}^k \big(F_k\big)_{i,j} = \sum_{k=1}^{r} \left\langle \Lambda_k, \overline{F_k} \right\rangle.$$

Demonstração. A demonstração deste teorema seguirá da Proposição 1.23, assim que estabelecermos uma relação entre as bases de V consideradas em cada enunciado. Fixe $k = 1, \ldots, r$ e considere uma decomposição ortogonal de I_k em subespaços invariantes e irredutíveis:

$$I_k = V_{k,1} \perp \cdots \perp V_{k,m_k}.$$

Para cada $i = 1, \ldots, m_k$, como $S_{k,i}$ é um subespaço invariante e irredutível da mesma componente isotípica, os subespaços são equivalentes entre si e existe um

G-isomorfismo entre $S_{k,i}$ e $V_{k,i}$. Pela Proposição 1.7, existe um G-isomorfismo que preserva o produto interno e como os vetores $f_{k,i,1}, \ldots, f_{k,i,d_k}$ são ortogonais e de mesmo tamanho, podemos escolher um G-isomorfismo $\phi_{k,i}\colon S_{k,i} \to V_{k,i}$ de modo que $\{e_{k,i,1}, \ldots, e_{k,i,d_k}\}$, com $e_{k,i,s} := \phi_{k,i}(f_{k,i,s})$, seja uma base ortonormal de $V_{k,i}$. Segue que as matrizes da representação em $V_{k,i}$ são iguais às matrizes da representação em $S_{k,i}$ com as suas respectivas bases já fixadas. Como V é soma direta dos subespaços invariantes $S_{k,i}$ e dos subespaços invariantes $V_{k,i}$, podemos definir um G-isomorfismo ϕ de V em V a partir das transformações $\phi_{k,i}$; para isso basta, para cada $v \in V$, escrever $v = \sum_{k=1}^{r}\sum_{i=1}^{m_k} s_{k,i}$, com $s_{k,i} \in S_{k,i}$ e então definir $\phi v := \sum_{k=1}^{r}\sum_{i=1}^{m_k} \phi_{k,i} s_{k,i}$.

Temos que ϕ^{-1} é um G-isomorfismo de V em V que leva os vetores $e_{k,i,s}$ nos vetores $f_{k,i,s}$ e, visto na base de vetores $e_{k,i,s}$, possui a mesma estrutura bloco-diagonal descrita na Seção 1.1.3. Isso implica que para cada $k = 1, \ldots, r$ existe uma matriz $A_k = (a^k_{i,j})_{i,j=1}^{m_k}$ inversível tal que

$$f_{k,i,s} = \sum_{j=1}^{m_k} a_{i,j} e_{k,j,s},$$

para todo $i = 1, \ldots, m_k$ e $s = 1, \ldots, d_k$. Escrevendo E_k como no enunciado da Proposição 1.23, temos

$$F_k = A_k E_k A_k^*.$$

Essa relação permite-nos obter o teorema a partir da proposição anterior. Pela Proposição 1.23, existem matrizes positivo-semidefinidas Λ'_k tais que $K = \sum_{k=1}^{r} \langle \Lambda'_k, \overline{E_k} \rangle$. Substituindo E_k por $A_k^{-1} F_k (A_k^{-1})^*$ e denotando $\overline{A_k}^{-1}$ por B_k, temos

$$\begin{aligned} K &= \sum_{k=1}^{r} \langle \Lambda'_k, \overline{E_k} \rangle = \sum_{k=1}^{r} \langle \Lambda'_k, B_k \overline{F_k} B_k^* \rangle = \sum_{k=1}^{r} \operatorname{tr}\left(B_k \overline{F_k}^* B_k^* \Lambda'_k\right) \\ &= \sum_{k=1}^{r} \operatorname{tr}\left(\overline{F_k}^* B_k^* \Lambda'_k B_k\right) = \sum_{k=1}^{r} \langle B_k^* \Lambda'_k B_k, \overline{F_k} \rangle \end{aligned}$$

e então fazendo $\Lambda_k = B_k^* \Lambda'_k B_k$, obtemos a expressão do teorema. □

Em certos casos, com o auxílio do Teorema de Stone-Weierstrass (Teorema IV.9 em Reed e Simon [RS72]), é possível considerar sequências de subespaços invariantes e de dimensão finita

$$V_0 \subset V_1 \subset \cdots \subset \mathcal{C}(X),$$

de modo que um kernel de $\mathcal{C}(X \times X)$ possa ser aproximado uniformemente por kernels de $V_d^{(2)}$ com d suficientemente grande.

Nas aplicações dos próximos capítulos, estamos interessados em kernels reais. Se G for um grupo cujas representações irredutíveis são realizáveis em espaços vetoriais reais, podemos ver o conjunto dos kernels reais como um subconjunto dos kernels complexos e encontrar bases para os subespaços irredutíveis e invariantes formadas apenas por funções reais, de modo que um kernel real possa ser representado por matrizes positivo-semidefinidas reais na expressão dada pelo Teorema 1.24.

1.3. Programação cônica

Programação cônica é uma grande classe de problemas de otimização definidos em termos de cones e que inclui os problemas de programação linear e semidefinida. Nesta seção vemos definições e resultados que incluem cones contidos em espaços vetoriais de dimensão infinita, o que será útil para a teoria desenvolvida no Capítulo 5, porém aqui, ao contrário das duas últimas seções, nos restringimos a espaços vetoriais reais. Vemos também uma teoria de dualidade que consiste no

estudo conjunto de um par de problemas de otimização que possuem uma relação de simetria entre si e que produzem limitantes para os valores das soluções viáveis um do outro. A principal referência desta seção são os Capítulos III e IV de Barvinok [Bar02], uma referência complementar para o estudo de cones em espaços vetoriais topológicos é o livro de Aliprantis e Tourky [AT07].

1.3.1. Teoremas de separação. Apresentamos primeiramente dois teoremas chamados "teoremas de separação" que são muito úteis adiante. Ambos mostram a existência de funcionais lineares que separam conjuntos convexos[1] disjuntos com certas propriedades topológicas.

Um *espaço vetorial topológico* é um espaço vetorial considerado junto com uma topologia onde as operações de adição e multiplicação por escalar são contínuas. O primeiro teorema supõe que um dos conjuntos possui interior não-vazio (Teorema III.3.2 de Barvinok [Bar02]):

Teorema 1.25. *Sejam V um espaço vetorial topológico e $A, B \subset V$ dois conjuntos convexos tais que $A \cap B = \emptyset$ e $\operatorname{int} A \neq \emptyset$. Então existe um hiperplano fechado que separa A e B, isto é, existe um funcional linear contínuo $f\colon V \to \mathbb{R}$ não-nulo tal que $f(x) \leq f(y)$ para todo $x \in A$ e $y \in B$.*

Um espaço vetorial topológico é *localmente convexo* se a sua topologia possuir uma base formada por conjuntos convexos; note que qualquer espaço normado é localmente convexo já que as bolas que geram a topologia desses espaços são convexas. Consideramos esses espaços pois assim temos o seguinte teorema (Teorema III.3.4 de Barvinok [Bar02]):

Teorema 1.26. *Sejam V um espaço vetorial topológico localmente convexo, $A \subset V$ um conjunto convexo fechado e $u \in V \setminus A$. Então existe um hiperplano fechado que separa estritamente u de A, isto é, existe um funcional linear contínuo $f\colon V \to \mathbb{R}$ tal que $f(x) < f(u)$ para todo $x \in A$.*

1.3.2. Dualidade entre espaços vetoriais. Sejam E e F dois espaços vetoriais reais. Se $\langle , \rangle : E \times F \to \mathbb{R}$ é uma forma bilinear *não-degenerada*, isto é, tal que $\langle e, f\rangle = 0$ para todo $e \in E$ implica $f = 0$ e $\langle e, f\rangle = 0$ para todo $f \in F$ implica $e = 0$, dizemos que $\langle , \rangle$ é uma *dualidade* entre E e F. Dizemos também que E e F formam um *par dual*. Um exemplo típico de par dual é quando E é um espaço vetorial topológico de Hausdorff localmente convexo, F é seu dual topológico (o espaço de todos os funcionais lineares contínuos) e $\langle e, f\rangle = f(e)$; pode-se mostrar que essa forma bilinear é não-degenerada com o Teorema 1.26 e o fato de $\{0\} \subset E$ ser fechado.

Note que a definição de par dual não pressupõe que os espaços E e F sejam espaços vetoriais topológicos. Podemos tornar os espaços E e F de uma dualidade em espaços vetoriais topológicos localmente convexos com a chamada *topologia fraca da dualidade* que em E é a topologia gerada pela base de conjuntos abertos da forma

$$\{e \in E : \alpha_i < \langle e, f_i\rangle < \beta_i,\ i = 1, \ldots, m\},$$

com $f_1, \ldots, f_m \in F$ e $\alpha_1, \beta_1, \ldots, \alpha_m, \beta_m \in \mathbb{R}$. A topologia fraca da dualidade em F é definida de maneira similar. Essas são as menores topologias com as quais os funcionais $\phi_f\colon E \to \mathbb{R},\ \phi_f(e) := \langle e, f\rangle$ e $\psi_e\colon F \to \mathbb{R},\ \psi_e(f) := \langle e, f\rangle$ são contínuos para todo $e \in E$ e $f \in F$. Com essas topologias temos a seguinte proposição (Teorema 4.2 de Barvinok [Bar02]):

Proposição 1.27. *Seja $\langle , \rangle : E \times F \to \mathbb{R}$ uma dualidade e considere E e F espaços topológicos com suas respectivas topologias fracas. Então, para cada $f \in F$,*

[1] Um subconjunto S de um espaço vetorial é *convexo* se para todo $u, v \in S$ e $\lambda \in [0, 1]$ temos $\lambda u + (1 - \lambda)v \in S$.

a função $\phi_f(e) = \langle e, f\rangle$ é um funcional linear contínuo em E, assim como para cada $e \in E$, a função $\psi_e(f) = \langle e, f\rangle$ é um funcional linear contínuo em F. Além do mais, todo funcional linear contínuo $\phi\colon E \to \mathbb{R}$ pode ser escrito como $\phi(e) = \langle e, f\rangle$ para um único $f \in F$ e todo funcional linear contínuo $\psi\colon F \to \mathbb{R}$ pode ser escrito como $\psi(f) = \langle e, f\rangle$ para um único $e \in E$.

Demonstração. Provemos primeiro que $\phi_f(e)$ é um funcional linear contínuo para todo $f \in F$. Seja $\epsilon > 0$ e $e_0 \in E$, para $\alpha := \phi_f(e_0)$, considere o aberto

$$U := \{e \in E : \alpha - \epsilon < \langle e, f\rangle < \alpha + \epsilon\}.$$

Temos que U é uma vizinhança de e_0 e para todo $e \in U$, temos $|\phi_f(e) - \phi_f(e_0)| \leq \epsilon$. Logo ϕ_f é contínuo em e_0 e, como e_0 é arbitrário, em todo E.

Provemos agora que qualquer funcional linear contínuo $\phi\colon E \to \mathbb{R}$ é igual à ϕ_f para algum $f \in F$ único. Como ϕ é contínuo em 0, existe uma vizinhança U de 0 tal que $|\phi(x)| < 1$ para todo $x \in U$. Temos que U pode ser escolhido dentro da base de abertos que define a topologia fraca e portanto é um conjunto da forma

$$U = \{e \in E : \alpha_i < \langle e, f_i\rangle < \beta_i,\ i = 1, \ldots, m\},$$

com $f_1, \ldots, f_m \in F$ e $\alpha_1, \beta_1, \ldots, \alpha_m, \beta_m \in \mathbb{R}$ (note que como $0 \in U$, então $\alpha_i < 0 < \beta_i$ para $i = 1, \ldots, m$). Para qualquer $\epsilon > 0$, temos também que se $\epsilon\alpha_i < \langle x, f_i\rangle < \epsilon\beta_i$ para $i = 1, \ldots, m$, então $(1/\epsilon)x \in U$ e, como ϕ é linear, $|\phi(x)| < \epsilon$. Logo, fazendo $\epsilon \to 0$, temos que se $\langle x, f_i\rangle = 0$ para $i = 1, \ldots, m$, então $\phi(x) = 0$, isto é, $\phi(x) = 0$ se $\phi_i(x) := \langle x, f_i\rangle = 0$ para $i = 1, \ldots, m$. Isso implica que[2] $\phi = \lambda_1\phi_1 + \cdots + \lambda_m\phi_m$, com $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ e portanto $\phi(x) = \langle x, f\rangle$ com $f := \lambda_1 f_1 + \cdots + \lambda_m f_m$. Para provar a unicidade, suponha que existam $f_1, f_2 \in F$ tais que $\phi(x) = \langle x, f_1\rangle = \langle x, f_2\rangle$ para todo $x \in E$. Então $\langle x, f_1 - f_2\rangle = 0$ para todo $x \in E$ e $f_1 - f_2 = 0$.

A demonstração para os funcionais $\psi\colon F \to \mathbb{R}$ é análoga. □

Daqui em diante consideramos implicitamente os pares duais com suas topologias. Também podemos considerar pares duais com topologias diferentes da fraca, mas nesse caso precisamos que as topologias sejam compatíveis com o par dual, isto é, o resultado da Proposição 1.27 torna-se uma hipótese adicional (também supomos que os espaços sejam localmente convexos). Novamente, o exemplo típico é E ser um espaço normado com a topologia da norma e F ser seu dual topológico com a topologia fraca (nesse caso também conhecida como topologia fraca-$*$).

Sejam $\langle , \rangle_1 : E_1 \times F_1 \to \mathbb{R}$ e $\langle , \rangle_2 : E_2 \times F_2 \to \mathbb{R}$ dualidades entre espaços vetoriais. Se $A\colon E_1 \to E_2$ e $A^*\colon F_2 \to F_1$ são transformações lineares tais que

$$\langle Ae, f\rangle_2 = \langle e, A^* f\rangle_1 \quad \text{para todos } e \in E_1,\ f \in F_2,$$

dizemos que A^* é a *transformação adjunta* de A. A seguinte proposição garante a existência e a unicidade da transformação adjunta:

Proposição 1.28. *Sejam $\langle , \rangle_1 : E_1 \times F_1 \to \mathbb{R}$ e $\langle , \rangle_2 : E_2 \times F_2 \to \mathbb{R}$ dualidades entre espaços vetoriais. Se $A\colon E_1 \to E_2$ é uma transformação linear contínua, então existe uma transformação adjunta $A^*\colon F_2 \to F_1$ e essa transformação é única.*

Demonstração. Para $f \in F_2$, seja $\phi\colon E_1 \to \mathbb{R}$ definido por $\phi(e) := \langle Ae, f\rangle_2$. Como A e $\langle \cdot, f\rangle_2$ são lineares e contínuos, temos que ϕ é um funcional linear contínuo e portanto, pela Proposição 1.27, existe um único $f' \in F_1$ tal que $\phi(e) = \langle e, f'\rangle_1$. Fazendo $A^* f := f'$, como $\langle , \rangle_1$ e $\langle , \rangle_2$ são bilineares, não é difícil ver que A^* é linear.

[2]Para ver isso, defina $\Phi\colon E \to \mathbb{R}^m$ por $\Phi(x) = (\phi_1(x), \ldots, \phi_m(x))$ e note que se $\Phi(x) = \Phi(y)$, então $\phi(x) = \phi(y)$, de modo que existe $\psi\colon \operatorname{Im}\Phi \to \mathbb{R}$ linear tal que $\phi = \psi \circ \Phi$ e, após estender ψ de $\operatorname{Im}\Phi$ para $\mathbb{R}^m$, temos que ψ pode ser escrita como uma combinação linear de suas coordenadas.

Para demonstrarmos a unicidade, se A_1^* e A_2^* são duas transformações tais que $\langle Ae, f\rangle_2 = \langle e, A_1^* f\rangle_1 = \langle e, A_2^* f\rangle_1$ para todo $e \in E_1$ e $f \in F_2$, então $\langle e, A_1^* f - A_2^* f\rangle_1 = 0$ para todo $e \in E_1$ e portanto $A_1^* f = A_2^* f$ para todo $f \in F_2$. □

1.3.3. Cones em espaços vetoriais topológicos. Um subconjunto K de um espaço vetorial E é chamado de *cone (convexo)* se para todos $\alpha, \beta \in \mathbb{R}$, $\alpha, \beta \geq 0$ e $x, y \in K$, temos $\alpha x + \beta y \in K$. Um cone é *pontudo* se $K \cap (-K) = \{0\}$.

Se $K \subset E$ é um cone, defimos a relação $\leq_K$ por

$$x \leq_K y \Leftrightarrow y - x \in K.$$

Não é difícil mostrar que $\leq_K$ é reflexiva e transitiva, e que se $x \leq_K y$, então $\alpha x + z \leq_K \alpha y + z$ para todos $z \in E$ e $\alpha \geq 0$. Entretanto, é necessário supor que K seja pontudo para que essa relação seja anti-simétrica e defina uma relação de ordem parcial em E.

Dada uma dualidade $\langle , \rangle : E \times F \to \mathbb{R}$ e um cone $K \subseteq E$, o *cone dual* $K^* \subseteq F$ é definido por

$$K^* := \{f \in F : \langle e, f\rangle \geq 0, \text{ para todo } e \in K\}.$$

Se K é um cone em F, definimos $K^* \subseteq E$ de maneira similar.

Proposição 1.29. *Seja $\langle , \rangle : E \times F \to \mathbb{R}$ uma dualidade e $K \subset E$ um cone. Se K é fechado, então $(K^*)^* = K$.*

Demonstração. Que $K \subseteq (K^*)^*$, segue diretamente da definição de cone dual. Para mostrarmos $(K^*)^* \subseteq K$, considere $x \notin K$. Como K é fechado, pelo Teorema 1.26, existe $f \in F$ tal que $\langle x, f\rangle < \langle u, f\rangle$ para todo $u \in K$. Como $0 \in K$, temos $\langle x, f\rangle < 0$. Ademais $f \in K^*$, pois se houvesse $y \in K$ tal que $\langle y, f\rangle < 0$, então tomando $\lambda = (\langle x, f\rangle / \langle y, f\rangle) + 1$, teríamos que $\langle \lambda y, f\rangle < \langle x, f\rangle$ com $\lambda y \in K$. Portanto $x \notin (K^*)^*$. □

Até aqui, todas as definições e resultados relativos a um par dual E, F guardam uma simetria entre os espaços, de modo que a ordem entre os espaços não seja relevante. A última proposição dá um passo a mais nessa direção ao mostrar que, supondo K fechado, $(K^*)^* = K$; isso permitirá que os programas primal e dual definidos na próxima seção sejam intercambiáveis. A hipótese sobre K ser fechado também possui essa simetria, já que K^* sempre é fechado (isso pode ser observado diretamente a partir da definição, que pode ser interpretada como uma intersecção de fechados).

Já observamos que a ordem induzida por um cone só é realmente uma relação de ordem parcial se o cone for pontudo. Infelizmente em geral não é verdade que o dual de um cone pontudo é pontudo, então precisamos considerar outras propriedades em conjunto para obtermos uma caracterização que preserve a simetria. A seguir, denotamos por $K - K$ o espaço vetorial gerado pelo cone K, isto é $K - K = \{x - y : x, y \in K\}$.

Proposição 1.30. *Seja $\langle , \rangle : E \times F \to \mathbb{R}$ uma dualidade e $K \subset E$ um cone.*

(1) *Se $K - K$ é denso em E, então K^* é pontudo.*
(2) *Se K é pontudo e fechado, então $K^* - K^*$ é denso em F.*

Demonstração. Para provarmos o item (1), seja $f \in K^* \cap (-K^*)$. Segue que $\langle e, f\rangle = 0$ para todo $e \in K - K$ e, pela continuidade de $\phi_f(x) = \langle x, f\rangle$ e densidade de $K - K$, temos que $\langle x, f\rangle = 0$ para todo $x \in E$. Logo $f = 0$ e K^* é pontudo.

Para provarmos o item (2), suponhamos por contradição que $K^* - K^*$ não seja denso em F. Então existe $f \in F$ não pertencente ao fecho de $K^* - K^*$ e pelo Teorema 1.26, existe $e \in E$ tal que $\langle e, f\rangle > 0$ e $\langle e, x\rangle = 0$ para todo $x \in K^* - K^*$ (aqui usamos que $K^* - K^*$ é um subespaço de F). Como K é fechado, pela Proposição 1.29, $(K^*)^* = K$ e logo $e \in K \cap (-K) = \{0\}$, uma contradição. □

A Proposição 1.30 motiva a seguinte definição: dada uma dualidade $\langle , \rangle : E \times F \to \mathbb{R}$, dizemos que o cone $K \subset E$ (ou F) é *próprio* se for fechado, pontudo e $K - K$ for denso em E (ou F).

PROPOSIÇÃO 1.31. *Se $\langle , \rangle : E \times F \to \mathbb{R}$ é uma dualidade e $K \subset E$ (ou F) é um cone próprio, então K^* também é próprio.* □

1.3.4. Par de programas primal e dual. A forma mais simétrica de apresentar o par de problemas de otimização cônica é a chamada forma canônica. Considere $\langle , \rangle_1 : E_1 \times F_1 \to \mathbb{R}$ e $\langle , \rangle_2 : E_2 \times F_2 \to \mathbb{R}$ duas dualidades, $K_1 \subset E_1$, $K_2 \subset E_2$ dois cones e $A\colon E_1 \to E_2$ uma transformação linear com adjunta. Fixados $c \in F_1$ e $b \in E_2$ temos respectivamente os *programas primal e dual na forma canônica*:

$$\begin{aligned} \max_{x \in E_1} \quad & \langle x, c \rangle_1 \\ & Ax \leq_{K_2} b, \qquad (11) \\ & x \geq_{K_1} 0. \end{aligned} \qquad\qquad \begin{aligned} \min_{y \in F_2} \quad & \langle b, y \rangle_2 \\ & A^* y \geq_{K_1^*} c, \qquad (12) \\ & y \geq_{K_2^*} 0. \end{aligned}$$

Trocando b, c e A por $-c, -b$ e $-A^*$, respectivamente, podemos reescrever o programa (12) na forma do programa (11) e, supondo que K_1 e K_2 sejam cones fechados de modo que valha a Proposição 1.29, o novo programa (12) torna-se equivalente ao antigo programa (11) e dessa forma os programas primal e dual são intercambiáveis.

O próximo teorema é conhecido como "teorema da dualidade fraca" e, apesar de ter uma demonstração simples, revela a principal característica do par de problemas de otimização: soluções viáveis de um produzem limitantes para o valor das soluções viáveis do outro.

TEOREMA 1.32. *Considere os problemas primal* (11) *e dual* (12) *na forma canônica.*

(1) *(Dualidade fraca) Para qualquer solução viável x do problema primal e qualquer solução viável y do dual, temos*
$$\langle x, c \rangle_1 \leq \langle b, y \rangle_2 .$$
Se $\langle x, c \rangle_1 = \langle b, y \rangle_2$, então x e y são soluções ótimas.

(2) *(Critério de otimalidade) Se x e y são soluções viáveis dos problemas primal e dual tais que*
$$\langle x, c - A^* y \rangle_1 = 0 \ \text{ e } \ \langle Ax - b, y \rangle_2 = 0,$$
então $\langle x, c \rangle_1 = \langle b, y \rangle_2$ e x e y são soluções ótimas.

(3) *(Folgas complementares) Se x e y são soluções ótimas dos problemas primal e dual tais que $\langle x, c \rangle_1 = \langle b, y \rangle_2$, então*
$$\langle x, c - A^* y \rangle_1 = 0 \ \text{ e } \ \langle Ax - b, y \rangle_2 = 0.$$

DEMONSTRAÇÃO. Sejam $x \in K_1$ e $y \in K_2^*$ tais que $Ax \leq_{K_2} b$ e $A^* y \geq_{K_1^*} c$. Todos os itens deste teorema seguem diretamente da seguinte cadeia de desigualdades:
$$\langle x, c \rangle_1 \leq \langle x, A^* y \rangle_1 = \langle Ax, y \rangle_2 \leq \langle b, y \rangle_2 .$$ □

Denotando o valor ótimo do problema primal por γ e o valor ótimo do problema dual por β, o teorema anterior nos mostra que sempre vale $\gamma \leq \beta$. Entretanto, em geral é possível que ocorra $\gamma < \beta$. A diferença $\beta - \gamma$ é chamada de *intervalo de dualidade* ("duality gap"). Existem teoremas que garantem $\beta = \gamma$ sob hipóteses adicionais, a forma mais fácil de mostrar-los é colocando os problemas na chamada forma padrão.

Considere $\langle , \rangle_1 : E_1 \times F_1 \to \mathbb{R}$ e $\langle , \rangle_2 : E_2 \times F_2 \to \mathbb{R}$ duas dualidades, $K \subset E_1$, um cone e $A\colon E_1 \to E_2$ uma transformação linear com adjunta. Fixados $c \in F_1$ e $b \in E_2$, temos respectivamente os *programas primal e dual na forma padrão*:

$$\begin{array}{ll} \max\limits_{x \in E_1} & \langle x, c\rangle_1 \\ & Ax = b, \\ & x \geq_K 0. \end{array} \qquad (13) \qquad\qquad \begin{array}{ll} \min\limits_{y \in F_2} & \langle b, y\rangle_2 \\ & A^* y \geq_{K^*} c. \end{array} \qquad (14)$$

A forma padrão pode ser vista como um caso particular da forma canônica, pois os programas (13) e (14) são iguais aos programas (11) e (12) quando $K_1 = K$ e $K_2 = \{0\}$. Entretanto, não há perda de generalidade ao considerarmos problemas nessa forma pois é possível reescrever problemas na forma canônica como problemas na forma padrão. Para ver como, usamos "variáveis de folga": considere $E = E_1 \oplus E_2$ e $F = F_1 \oplus F_2$, com a dualidade $\langle , \rangle : E \times F \to \mathbb{R}$ definida por $\langle (e_1, e_2), (f_1, f_2)\rangle = \langle e_1, f_1\rangle_1 + \langle e_2, f_2\rangle_2$, faça $K = K_1 \times K_2 \subset E$, $\hat{c} = (c, 0)$ e defina a transformação linear $\hat{A}\colon E \to E_2$ por $\hat{A}(x, v) = Ax + v$. Temos que $\hat{x} \geq_K 0$ implica que $\hat{x} = (x, v)$ com $x \in K_1$ e $v \in K_2$ e $\hat{A}\hat{x} = b$ implica que $Ax + v = b$, isto é $b \geq_{K_2} Ax$, de modo que escrevemos o programa (11) na forma do programa (13); o dual (14) deste problema também corresponde ao problema (12), pois $\hat{A}^* y = (A^* y, y)$ e $K^* = K_1^* \times K_2^*$.

Existem diversos teoremas que dão condições para a dualidade forte, isto é, $\gamma = \beta$. Vejamos a seguir dois deles, outros resultados semelhantes podem ser encontrados na Seção IV.7 de Barvinok [Bar02].

Teorema 1.33. *Considere os problemas primal* (13) *e dual* (14) *na forma padrão. Se o cone* $\hat{A}(K) := \{(Ax, \langle x, c\rangle_1) : x \in K\}$ *contido em* $E_2 \times \mathbb{R}$ *for fechado e existir uma solução viável para o problema primal, então* $\gamma = \beta$. *Se o valor ótimo do problema primal for limitado, então existe uma solução ótima para o primal.*

Demonstração. Se o problema primal não for limitado (isto é, $\gamma = \infty$), então pela dualidade fraca o dual não pode ser viável e $\gamma = \beta = \infty$. Podemos então supor que $\gamma < \infty$.

Considere a linha $L = \{(b, \tau) : \tau \in \mathbb{R}\}$, também contida em $E_2 \times \mathbb{R}$. A intersecção $L \cap \hat{A}(K)$ é um conjunto fechado de pontos da forma $(b, \langle x, c\rangle_1)$ com x viável para o primal. Como o problema primal é viável, temos que $L \cap \hat{A}(K) \neq \emptyset$ e como essa intersecção é fechada e estamos supondo que o primal é limitado, temos que existe uma solução ótima x_0 tal que $\langle x_0, c\rangle_1 = \gamma$.

Pela dualidade fraca, já sabemos que $\gamma \leq \beta$. Provemos agora que para qualquer $\epsilon > 0$, existe uma solução y viável para o dual tal que $\langle b, y\rangle_2 < \gamma + \epsilon$, de modo que $\beta \leq \gamma$. Como $(b, \gamma + \epsilon) \notin \hat{A}(K)$ e $\hat{A}(K)$ é convexo e fechado, pelo Teorema 1.26, existe $(y', \sigma) \in F_2 \times \mathbb{R}$ tal que

$$\langle b, y'\rangle_2 + \sigma(\gamma + \epsilon) < \langle Ax, y'\rangle_2 + \sigma \langle x, c\rangle_1 \qquad (15)$$

para todo $x \in K$. Como $0 \in K$, temos

$$\langle b, y'\rangle_2 + \sigma(\gamma + \epsilon) < 0.$$

Ademais, vale que

$$\langle Ax, y'\rangle_2 + \sigma \langle x, c\rangle_1 \geq 0$$

para todo $x \in K$, pois caso tivéssemos $\langle Ax, y'\rangle_2 + \sigma \langle x, c\rangle_1 < 0$ para algum $x \in K$, como K é um cone, a condição (15) seria violada para algum $\lambda x \in K$ com $\lambda > 0$. Substituindo em (15) a solução ótima $x_0 \in K$ do primal, como $Ax_0 = b$, $\langle x_0, c\rangle = \gamma$ e $\epsilon > 0$, obtemos $\sigma < 0$. Fazendo $y := -(1/\sigma)y'$, temos $\langle b, y\rangle_2 < \gamma + \epsilon$ e também $\langle Ax, y\rangle_2 - \langle x, c\rangle_1 \geq 0$ para todo $x \in K$, o que implica $\langle x, A^* y - c\rangle_1 \geq 0$, de modo que $A^* y \geq_{K^*} c$. Portanto y é viável para o dual e $\beta = \gamma$. □

A seguir um outro teorema semelhante que também garante a dualidade forte. Dessa vez a hipótese é sobre a existência de um ponto no interior de um conjunto e em vez de usarmos o Teorema 1.26 para a separação, usamos o Teorema 1.25.

TEOREMA 1.34. *Considere os problemas primal* (13) *e dual* (14) *na forma padrão. Seja* $\mathcal{A}(K) \subset E_2 \times \mathbb{R}$ *o conjunto* $\mathcal{A}(K) := \{(Ax, t) : x \in K, t \leq \langle x, c\rangle_1\}$. *Se o primal possuir uma solução viável* $\tilde{x}$ *tal que* $(A\tilde{x}, \langle \tilde{x}, c\rangle_1) \in \operatorname{int} \mathcal{A}(K)$, *então* $\gamma = \beta$. *Se o valor ótimo do problema primal for limitado, então existe uma solução ótima para o dual.*

DEMONSTRAÇÃO. Se o problema primal não for limitado ($\gamma = \infty$), então pela dualidade fraca o dual não pode ser viável e $\gamma = \beta = \infty$. Podemos então supor que $\gamma < \infty$.

Considere o conjunto $L = \{(b, s) : s > \gamma\}$, também contido em $E_2 \times \mathbb{R}$. A intersecção entre L e $\mathcal{A}(K)$ é vazia pois caso $(b, t) \in \mathcal{A}(K) \cap L$, existiria $x \in K$ viável para o primal com $\langle x, c\rangle_1 > \gamma$, o que contradiz o fato de γ ser o valor ótimo do primal. Logo, pelo Teorema 1.25, existe $(y', \sigma) \in F_2 \times \mathbb{R} \setminus \{(0, 0)\}$ tal que

$$\langle b, y'\rangle_2 + \sigma s \leq \langle Ax, y'\rangle_2 + \sigma t \tag{16}$$

para todo $s > \gamma$, $x \in K$ e $t \leq \langle x, c\rangle_1$.

Como $0 \in K$, para $t = \langle 0, c\rangle_1 = 0$, temos $\langle b, y'\rangle_2 + \sigma s \leq 0$ para todo $s > \gamma$ e logo também vale

$$\langle b, y'\rangle_2 + \sigma\gamma \leq 0. \tag{17}$$

Ademais, temos

$$\langle Ax, y'\rangle_2 + \sigma \langle x, c\rangle_1 \geq 0 \tag{18}$$

para todo $x \in K$, pois caso tivéssemos $\langle Ax, y'\rangle_2 + \sigma \langle x, c\rangle_1 < 0$ para algum $x \in K$, como K é um cone, a condição (16) seria violada para algum $\lambda x \in K$ com $\lambda > 0$.

Provemos agora que $\sigma < 0$. Caso $\sigma > 0$, fixando $s > \gamma$ e $x \in K$, a condição (16) seria violada para $t \to -\infty$. Também não é possível que $\sigma = 0$, pois caso isso ocorra teríamos que $y' \neq 0$ e, de (17) e (18), $\langle A\tilde{x}, y'\rangle_2 = \langle b, y'\rangle = 0$. Como $(A\tilde{x}, \langle \tilde{x}, c\rangle_1) \in \operatorname{int} \mathcal{A}(K)$, existe uma vizinhança U de b tal que para todo $u \in U$, existe $x \in K$ com $Ax = u$. Como $y' \neq 0$ e $\langle , \rangle_2$ é não-degenerada, existe $v \in E_2$ tal que $\langle v, y'\rangle_2 > 0$. Como a função $\varphi\colon \mathbb{R} \to E_2$ definida por $\varphi(\lambda) := \lambda v + b$ é contínua, $\varphi^{-1}(U)$ é aberto em $\mathbb{R}$ e como $0 \in \varphi^{-1}(U)$, existe $\epsilon > 0$ tal que se $|\lambda| < \epsilon$, então $\lambda v + b \in U$. Logo existe $x' \in K$ tal que $-(\epsilon/2)v + b = Ax'$ e então $\langle Ax', y'\rangle_2 < 0$, o que entra em contradição com (18).

Definindo $y := -(1/\sigma)y'$, a partir de (18) temos $\langle x, A^*y - c\rangle_1 \geq 0$ para todo $x \in K$ e portanto $A^*y \geq_{K^*} c$, isto é, y é solução viável para o dual. Por outro lado, a relação (17) mostra que $\langle b, y\rangle_2 \leq \gamma$ e como, pela dualidade fraca, vale $\langle b, y\rangle_2 \geq \gamma$, concluímos que $\langle b, y\rangle_2 = \gamma$ e portanto y é uma solução ótima do dual. □

Em problemas de dimensão infinita é difícil aplicar o Teorema 1.34 pois muitos cones de interesse possuem interior vazio. Em espaços de dimensão finita, como em problemas de programação semidefinida, esse teorema leva a uma condição muito útil para a dualidade forte, às vezes chamada de condição de Slater:

PROPOSIÇÃO 1.35. *Considere os problemas primal* (13) *e dual* (14) *na forma padrão e todos os espaços sendo de dimensão finita. Se existir* $\tilde{x} \in \operatorname{int} K$ *viável para o primal e o valor ótimo do primal for limitado, então* $\gamma = \beta$ *e existe uma solução ótima para o dual.*

DEMONSTRAÇÃO. Podemos supor que a transformação A é sobrejetora (como já estamos suponto que o primal é viável, a substituição de E_2 pela imagem de A não altera o programa (13) e estamos encontrando uma solução ótima para o programa (14), cujo espaço de variáveis é apenas reduzido). Como $\tilde{x} \in \operatorname{int} K$, existe

uma bola em torno de $\tilde{x}$ contida em K e como A é sobrejetora e os espaços possuem dimensão finita, a imagem dessa bola por A também contém uma bola em E_2. Temos então que $(A\tilde{x}, \langle \tilde{x}, c \rangle_1) \in \operatorname{int} \mathcal{A}(K)$ e o resultado segue do Teorema 1.34. □

CAPÍTULO 2

O limitante de programação linear

Neste capítulo obtemos o assim chamado limitante de programação linear, criado por Delsarte, Goethals e Seidel [DGS77] para limitar o tamanho dos códigos esféricos. Fazemos isso a partir do número teta de Lovász, conforme apresentado por Bachoc, Nebe, Oliveira e Vallentin [BNdOFV09]. Após vermos o número teta de Lovász em grafos finitos e sua extensão com kernels, a forma final do limitante é obtida com o auxílio de certas propriedades dos polinômios harmônicos esféricos, que estudamos em detalhe ao final deste capítulo.

2.1. O número teta de Lovász para grafos finitos

Lovász apresentou em 1979 [Lov79] o parâmetro $\vartheta(G)$, que limita superiormente o número de independência de um grafo finito $G = (V, E)$. Existem diversas formas equivalentes de apresentá-lo (veja o artigo original ou o artigo de Knuth [Knu94]), uma delas é como o valor ótimo do seguinte programa de otimização semidefinida:

$$\begin{aligned} \max \quad & \sum_{u,v \in V} X(u,v) \\ & X \in \mathbb{R}^{V \times V},\ X \succeq 0, \\ & \sum_{u \in V} X(u,u) = 1, \\ & X(u,v) = 0 \quad \text{se } \{u,v\} \in E. \end{aligned} \tag{19}$$

Não é difícil provar que $\vartheta(G)$ limita o número de independência $\alpha(G)$:

Proposição 2.1. *Para todo grafo finito $G = (V, E)$, temos $\vartheta(G) \geq \alpha(G)$.*

Demonstração. Para qualquer subconjunto $C \subseteq V$ não-vazio e independente, seja $\chi_C \colon V \to \{0, 1\}$ sua função característica (tal que $\chi_C(u) = 1$ se $u \in C$ e $\chi_C(u) = 0$, caso contrário). Defina a matriz $X \in \mathbb{R}^{V \times V}$ por

$$X(u,v) = \frac{1}{|C|} \chi_C(u) \chi_C(v).$$

Temos que X é uma solução viável para o Programa (19) cujo valor é igual a $\sum_{u,v \in V} X(u,v) = |C|$. □

Observando as soluções viáveis que representam os conjuntos independentes no Programa (19), podemos obter um limitante melhor adicionando restrições de não-negatividade nas entradas de X. Esse limitante será mais adequado para a nossa aplicação e será denotado por $\vartheta'(G)$, seguindo Schrijver [Sch79] ($X \geq 0$ significa

que todas as entradas de X devem ser não-negativas):

$$\begin{aligned} \max \quad & \sum_{u,v \in V} X(u,v) \\ & X \in \mathbb{R}^{V \times V},\ X \geq 0,\ X \succeq 0, \\ & \sum_{u \in V} X(u,u) = 1, \\ & X(u,v) = 0 \quad \text{se } \{u,v\} \in E. \end{aligned} \tag{20}$$

Voltaremos a esse programa na Seção 5.2, quando vemos formas de fortalecê-lo. Veremos agora um programa em que não apenas o valor ótimo, mas qualquer solução viável produz um limitante para o número de independência. Este é o programa dual, conforme a teoria de dualidade de programação cônica vista na Seção 1.3.

$$\begin{aligned} \min \quad & \lambda \\ & \lambda \in \mathbb{R},\ Z \in \mathbb{R}^{V \times V},\ Z \succeq 0, \\ & Z(u,u) = \lambda - 1 \text{ para todo } u \in V, \\ & Z(u,v) \leq -1 \quad \text{se } u,v \in V,\ u \neq v,\ \{u,v\} \notin E. \end{aligned} \tag{21}$$

Como $X = \frac{1}{|V|} I$ é uma solução estritamente viável para o Programa (20), vale a condição de Slater (Proposição 1.35) para a dualidade forte e assim o valor ótimo do Programa (21) também é $\vartheta'(G)$. Provamos a seguir diretamente a afirmação de que soluções viáveis para o Programa (21) produzem limitantes superiores para o número de independência.

Proposição 2.2. *Se λ, Z satisfazem as restrições do Programa* (21)*, então $\alpha(G) \leq \lambda$.*

Demonstração. Se $C \subseteq V$ é um conjunto independente não-vazio qualquer, como $Z \succeq 0$, temos

$$0 \leq \sum_{u,v \in C} Z(u,v) \leq |C|(\lambda - 1) + (|C|^2 - |C|)(-1),$$

o que implica $|C| \leq \lambda$. □

2.2. Extensão do número teta linha de Lovász para códigos esféricos

Voltamos agora ao problema central desta dissertação, em que desejamos limitar o tamanho máximo de um código esférico $C \subset S^{n-1}$ com distância angular mínima θ, valor denotado por $A(n,\theta)$. Como descrito na introdução, esse problema pode ser formulado como o número de independência do grafo $G_{n,\theta}$ que possui conjunto de vértices S^{n-1} e arestas entre vértices com distância angular menor que θ.

Notemos aqui que S^{n-1} com sua medida de superfície ω satisfaz as hipóteses impostas sobre X na Seção 1.2 e assim podemos estender o número teta linha de Lovász para o grafo $G_{n,\theta}$ substituindo as matrizes no Programa (21) por kernels contínuos (note que estamos usando kernels reais):

$$\begin{aligned} \min \quad & \lambda \\ & \lambda \in \mathbb{R},\ Z \in \mathcal{C}(S^{n-1} \times S^{n-1})_{\succeq 0}, \\ & Z(u,u) = \lambda - 1 \text{ para todo } u \in S^{n-1}, \\ & Z(u,v) \leq -1 \quad \text{se } u \cdot v \in [-1, \cos\theta]. \end{aligned} \tag{22}$$

A hipótese sobre a continuidade dos kernels permite-nos usar a caracterização dos kernels contínuos e positivos dada na Proposição 1.22, que faz com que o mesmo argumento usado na demonstração da Proposição 2.2 se aplique a esse limitante.

PROPOSIÇÃO 2.3. *Se λ, Z satisfazem as restrições do Programa* (22)*, então $A(n,\theta) \leq \lambda$.* □

Também seria natural tentar considerar o Programa (20) com kernels, entretanto o resultado é um programa que não se aplica ao grafo $G_{n,\theta}$ que estamos considerando. Conforme visto na Seção 1.3, para escrevermos o programa dual de (22) é necessário considerar o cone dual de $\mathcal{C}(S^{n-1} \times S^{n-1})_{\succeq 0}$, que está contido em um espaço de medidas. Veremos mais detalhes sobre essa relação de dualidade no Capítulo 5.

Até aqui temos um método para limitar o tamanho de códigos esféricos, porém ainda não temos técnicas para efetivamente construir kernels que sejam soluções viáveis do Programa (22). O próximo passo será observar que podemos nos restringir a kernels invariantes, semelhantes aos vistos na Seção 1.2.2.

O grupo ortogonal

$$\mathrm{O}(\mathbb{R}^n) := \{T \in \mathbb{R}^{n\times n} \ : \ T^T T = I\}$$

age sobre os kernels de $\mathcal{C}(S^{n-1} \times S^{n-1})$:

$$(T \circ Z)(x,y) := Z(T^{-1}x, T^{-1}y).$$

Se λ, Z satisfazem as restrições do Programa (22) e $T \in \mathrm{O}(\mathbb{R}^n)$, então λ, $T \circ Z$ também satisfazem. De fato, para qualquer $u \in S^{n-1}$, $(T \circ Z)(u,u) = Z(T^{-1}u, T^{-1}u) = \lambda - 1$ e se $u \cdot v \in [-1, \cos\theta]$, então $(T \circ Z)(u,v) = Z(T^{-1}u, T^{-1}v) \leq -1$ (pois $(T^{-1}u) \cdot (T^{-1}v) = u \cdot v$). Além disso, $T \circ Z$ é um kernel contínuo (T é um operador linear e portanto estamos apenas compondo operações contínuas) e positivo (substitua $\{x_1, \ldots, x_n\} \subset S^{n-1}$ por $\{T^{-1}x_1, \ldots, T^{-1}x_n\}$ ao aplicar a Proposição 1.22).

O grupo ortogonal é um grupo topológico com sua topologia herdada de $\mathbb{R}^{n\times n}$ e, conforme mencionado na Seção 1.1.1, possui uma medida de Haar normalizada que denotamos por μ. Temos que se (λ, Z) é uma solução viável do problema, então $(\lambda, \bar{Z})$ com

$$\bar{Z}(x,y) := \int_{\mathrm{O}(\mathbb{R}^n)} (T \circ Z)(x,y)\, \mathrm{d}\mu(T)$$

é uma solução invariante com o mesmo valor. Portanto podemos restringir o Programa (22) a kernels invariantes sem enfraquecer o limitante:

$$\begin{aligned} \min \quad & \lambda \\ & \lambda \in \mathbb{R},\ Z \in \mathcal{C}(S^{n-1} \times S^{n-1})_{\succeq 0}^{\mathrm{O}(\mathbb{R}^n)}, \\ & Z(u,u) = \lambda - 1 \text{ para todo } u \in S^{n-1}, \\ & Z(u,v) \leq -1 \quad \text{se } u \cdot v \in [-1, \cos\theta]. \end{aligned} \tag{23}$$

A vantagem da restrição a kernels invariantes é que eles podem ser caracterizados de uma forma simples. A *órbita* de um par de pontos $(u,v) \in S^{n-1} \times S^{n-1}$ sob a ação de $\mathrm{O}(\mathbb{R}^n)$ é o conjunto

$$O(u,v) := \{(Tu, Tv) \ : \ T \in \mathrm{O}(\mathbb{R}^n)\}.$$

Tal órbita é determinada pelo produto interno $u \cdot v$. Como um kernel invariante deve ser constante entre pontos de uma mesma órbita, um kernel contínuo e invariante pode ser representado por uma função contínua $f\colon [-1,1] \to \mathbb{R}$:

$$Z(u,v) = f(u \cdot v).$$

Para expressar um kernel invariante e positivo consideramos os resultados sobre kernels invariantes exibidos na Seção 1.2.2. É necessário estudar a representação de $\mathrm{O}(\mathbb{R}^n)$ em $\mathcal{C}(S^{n-1})$, o que será feito em detalhe na Seção 2.4. A conclusão será o seguinte teorema [Sch42]:

TEOREMA 2.4 (Teorema de Schoenberg). *Um kernel $K\colon S^{n-1} \times S^{n-1} \to \mathbb{R}$ é contínuo, positivo e $\mathrm{O}(\mathbb{R}^n)$-invariante se e somente se*

$$K(x, y) = \sum_{k=0}^{\infty} a_k P_k^n(x \cdot y),$$

com $a_k \geq 0$, tais que $\sum_{k=0}^{\infty} a_k$ converge. Nesse caso, a série acima converge absolutamente e uniformemente em $S^{n-1} \times S^{n-1}$.

Acima, P_k^n é o polinômio de Gegenbauer de grau k, normalizado de modo que $P_k^n(1) = 1$. Esses polinômios podem ser obtidos a partir dos polinômios $1, t, t^2 \ldots$ através do processo de Gram-Schmidt com o produto interno

$$\langle f, g \rangle = \int_{-1}^{1} f(t)g(t)(1 - t^2)^{(n-3)/2} \, \mathrm{d}t.$$

Os polinômios P_k^n serão estudados na Seção 2.4.4.

2.3. O limitante de programação linear

Usando o Teorema de Schoenberg para reescrever o Programa (23) e truncando a série no d-ésimo termo, obtemos o chamado limitante de programação linear, devido a Delsarte, Goethals e Seidel [DGS77]:

$$\begin{aligned} \min \quad & 1 + \sum_{k=1}^{d} a_k \\ & a_k \in \mathbb{R},\ a_k \geq 0 \qquad \text{para } k = 1, \ldots, d, \\ & \sum_{k=1}^{d} a_k P_k^n(t) \leq -1 \qquad \text{para todo } t \in [-1, \cos\theta]. \end{aligned} \tag{24}$$

Substituímos a variável λ por $1 + \sum_{k=1}^{d} a_k$ e fixamos $a_0 = 0$, já que como $P_0^n(t) = 1$ para todo $t \in \mathbb{R}$, qualquer solução viável do programa com $a_0 > 0$ pode ser transformada em uma solução viável de valor menor fazendo $a_0 = 0$.

Apesar de a validade desse limitante resultar do mesmo argumento usado na demonstração da Proposição 2.2, a seguir é fornecida uma outra prova de que soluções viáveis desse programa limitam o tamanho de um código esférico. Essa prova se baseia em uma propriedade dos polinômios P_k^n que será demonstrada na Seção 2.4.4.

PROPOSIÇÃO 2.5. *Se $C \subset S^{n-1}$ é finito e $w\colon C \to \mathbb{C}$, então*

$$\sum_{c, c' \in C} w(c)\overline{w(c')} P_k^n(c \cdot c') \geq 0.$$

TEOREMA 2.6. *Se $a_1, \ldots, a_d$ satisfazem as restrições do Programa (24), então $A(n, \theta) \leq \left\lfloor 1 + \sum_{k=1}^{d} a_k \right\rfloor$.*

DEMONSTRAÇÃO. Seja $C \subset S^{n-1}$, $C \neq \emptyset$, um código esférico de distância angular mínima θ. Considere a quantidade

$$S(C) := \sum_{u, v \in C} \left(1 + \sum_{k=1}^{d} a_k P_k^n(u \cdot v) \right).$$

Por um lado, a Proposição 2.5 implica que

$$S(C) = |C|^2 + \sum_{k=1}^{d} a_k \sum_{u,v \in C} P_k^n(u \cdot v) \geq |C|^2. \tag{25}$$

Por outro lado,

$$\begin{aligned} S(C) &= \sum_{u \in C} \left(1 + \sum_{k=1}^{d} a_k P_k^n(u \cdot u)\right) + \sum_{\substack{u,v \in C, \\ u \neq v}} \left(1 + \sum_{k=1}^{d} a_k P_k^n(u \cdot v)\right) \\ &\leq |C| \left(1 + \sum_{k=1}^{d} a_k\right), \end{aligned} \tag{26}$$

pois se $u, v \in C$ e $u \neq v$, então $u \cdot v \in [-1, \cos\theta]$. Juntando as duas desigualdades e considerando que $A(n, \theta)$ é um número inteiro, obtemos o resultado desejado. □

O limitante de programação linear é assim chamado pois o Programa (24) pode ser visto como um programa linear com infinitas restrições (uma para cada $t \in [-1, \cos\theta]$). Uma forma de lidar com essas restrições é usando *amostragem*: escolha um subconjunto finito $A \subset [-1, \cos\theta]$ de amostras e resolva o programa linear apenas com essas restrições. Se A for uma amostra grande e suficientemente boa, o polinômio $f(t) = 1 + \sum_{k=1}^{d} a_k P_k^n(t)$ será quase viável para o programa original e poderá ser modificado de modo a produzir um limitante válido (seja $M = \max_{t \in [-1, \cos\theta]} f(t)$ e caso $0 < M < 1$, substitua f por $\frac{f-M}{1-M}$ para obter uma solução viável). Outro método para se calcular uma solução do Programa (24) é usando programação semidefinida e a representação de polinômios não-negativos por somas de quadrados, esse método será visto na Seção 4.1.

Na Tabela 2.1 vemos os valores dos limitantes superiores calculados por Odlyzko e Sloane [OS79] para o número de contato ($\theta = \pi/3$) usando a técnica de amostragem.

n	**lim. inferior**	**lim. superior**
3	12	13
4	24	25
5	40	46
6	72	82
7	126	140
8	240	240
9	306	380
10	500	595
11	582	915
12	840	1416
13	1154	2233

n	**lim. inferior**	**lim. superior**
14	1606	3492
15	2564	5431
16	4320	8313
17	5346	12215
18	7398	17877
19	10668	25901
20	17400	37974
21	27720	56852
22	49896	86537
23	93150	128096
24	196560	196560

TABELA 2.1. Limitantes superiores para o número de contato calculados por Odlyzko e Sloane [OS79]. Os valores dos limitantes inferiores são obtidos através de construções explícitas e podem ser encontrados no livro de Conway e Sloane [CS99], à parte das dimensões 13 e 14, nas quais os limitantes foram obtidos por Zinoviev e Ericson [ZE99].

É interessante observar que o limitante de programação linear não mostra que $\tau_3 = 12$ (resultado devido a Schütte e van der Waerden [SvdW53]) e nem que $\tau_4 = 24$ (resultado conjecturado por muito tempo, mas que só foi provado em 2003, por Musin [Mus08]). Mas, talvez surpreendentemente, o limitante de programação linear mostra que $\tau_8 = 240$ e $\tau_{24} = 196560$. O valor do limitante obtido nessas

dimensões é justo (não faz uso da função piso) e isso nos dá informações sobre os códigos esféricos que atingem esses valores. Por exemplo, em dimensão 8 o polinômio $f(t) = 1 + \sum_{k=1}^{d} a_k P_k^n(t)$ que produz o limitante é:

$$\begin{aligned} f(t) &= 1 + 8P_1^8(t) + 25P_2^8(t) + 52P_3^8(t) + \frac{133}{2}P_4^8(t) + 60P_5^8(t) + \frac{55}{2}P_6^8(t) \\ &= \frac{320}{3}(t+1)(t+1/2)^2t^2(t-1/2). \end{aligned}$$

Como o limitante é justo, todas as desigualdades na demonstração do Teorema 2.6 devem valer com igualdade para um código esférico $C \subset S^7$ de distância angular mínima $\pi/3$ e tamanho 240. Da igualdade na inequação (26) resulta que se $x, y \in C$, então $x \cdot y \in \{-1, -1/2, 0, 1/2, 1\}$. Também, definindo

$$A_t := \frac{1}{|C|}|\{(x, y) \in C \times C : x \cdot y = t\}|,$$

para $t \in \{-1, -1/2, 0, 1/2, 1\}$, obtemos a partir da igualdade na inequação (25) que

$$A_{-1}P_k^8(-1) + A_{-1/2}P_k^8(-1/2) + A_0P_k^8(0) + A_{1/2}P_k^8(1/2) = -1$$

para $k = 0, \ldots, 6$, o que implica que $A_{-1} = 1, A_{-1/2} = A_{1/2} = 56$ e $A_0 = 126$. Usando essas propriedades, Bannai e Sloane [BS81] mostraram que as configurações com 240 pontos em dimensão 8 são iguais a menos de isometrias e correspondem à obtida pelo reticulado E_8. De forma semelhante, em dimensão 24 as configurações com 196560 pontos correspondem a menos de isometrias à obtida pelo reticulado de Leech (para informações sobre esses reticulados, veja o Capítulo 4 de Conway e Sloane [CS99]).

Para a dimensão 17, o Programa (24) apenas produz o limitante $\tau_{17} \leq 12218$. Para a determinação do limitante $\tau_{17} \leq 12215$, Odlyzko e Sloane [OS79] usaram uma modificação do Programa (24) que se baseia na observação de que em uma calota esférica de raio $\phi < \pi/6$ cabe no máximo um ponto de um código esférico e em uma calota esférica de raio $\phi < \arccos\left(\sqrt{2/3}\right)$ cabem no máximo dois pontos de um código esférico ($\arccos\left(\sqrt{2/3}\right)$ é o ângulo entre o centro e os vértices de um triângulo esférico de lados com comprimento angular $\pi/3$), o que leva ao seguinte programa:

$$\begin{aligned} \min \quad & 1 + \sum_{k=1}^{d} a_k + b_1 + 2b_2 \\ & a_k \in \mathbb{R},\ a_k \geq 0 && \text{para } k = 1, \ldots, d, \\ & b_1, b_2 \in \mathbb{R},\ b_1, b_2 \geq 0, \\ & \sum_{k=1}^{d} a_k P_k^n(t) \leq -1 + b_1 + b_2 && \text{para todo } t \in [-1, -\sqrt{3}/2], \\ & \sum_{k=1}^{d} a_k P_k^n(t) \leq -1 + b_2 && \text{para todo } t \in [-\sqrt{3}/2, -\sqrt{2/3}], \\ & \sum_{k=1}^{d} a_k P_k^n(t) \leq -1 && \text{para todo } t \in [-\sqrt{2/3}, 1/2]. \end{aligned} \tag{27}$$

Esse programa é válido para qualquer dimensão n e a princípio é capaz de produzir limitantes melhores, visto que fixando $b_1 = b_2 = 0$ voltamos à formulação anterior. Porém apenas em dimensão 17, Odlyzko e Sloane [OS79] foram capazes de

produzir um limitante melhor com essa modificação (experimentos com a técnica de somas de quadrados que veremos na Seção 4.1 mostram que também é possível obter $\tau_{16} \leq 8312$, $\tau_{19} \leq 25900$, $\tau_{21} \leq 56851$ e $\tau_{23} \leq 128095$ com o Programa (27)).

A prova de que uma solução viável para o Programa (27) produz um limitante para o número de contato é semelhante à prova do Teorema 2.6, só que com a separação em três casos para se limitar o termo $\sum_{\substack{u,v \in C, \\ u \neq v}} f(u \cdot v)$ e o uso das observações sobre os códigos em calotas. A prova de que $\tau_4 = 24$, de Musin [Mus08], também modifica o Programa (24) usando restrições sobre códigos dentro de calotas, porém com argumentos mais elaborados e específicos para a dimensão 4.

2.4. Polinômios harmônicos esféricos

Nesta seção obtemos uma caracterização dos kernels $\mathrm{O}(\mathbb{R}^n)$-invariantes em termos dos polinômios de Gegenbauer, uma família de polinômios ortogonais que surgem a partir do estudo dos espaços de polinômios harmônicos esféricos, que vemos corresponder aos subespaços invariantes e irredutíveis da representação de $\mathrm{O}(\mathbb{R}^n)$ em $\mathcal{C}(S^{n-1})$. Grande parte desta seção é baseada na Seção IX.2 do livro de Vilenkin [Vil68], entretanto o Capítulo 9 do livro de Andrews, Askey e Roy [AAR99] e o Capítulo 5 do livro de Axler, Bourdon e Ramey [ABR01] também são úteis (o primeiro livro foca na teoria das representações, o segundo nas propriedades dos polinômios que surgem e o terceiro nas funções harmônicas). No restante desta seção consideramos polinômios com n variáveis reais e usamos as notações $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ e $r^2 = x_1^2 + \ldots + x_n^2$.

2.4.1. Espaços de polinômios esféricos, homogêneos e harmônicos. Consideramos espaços vetoriais complexos de polinômios com n variáveis reais. Chamamos de *polinômio esférico* a imagem de um polinômio sob a aplicação $p \mapsto p|_{S^{n-1}}$, que restringe um polinômio à função que este define na esfera. Note que essa aplicação é linear, porém não é injetiva: se f e g são dois polinômios e $h(x) = f(x) + (1 - r^2)g(x)$, então $f(\xi) = h(\xi)$ para qualquer $\xi \in S^{n-1}$.

Denotemos por $\mathrm{Pol}(S^{n-1})_{\leq d}$ o espaço vetorial complexo formado pela imagem dos polinômios de grau no máximo d sob essa aplicação. Esse espaço é um subespaço de $L^2(S^{n-1})$ e, tal como definido na Seção 1.2, podemos considerá-lo com o produto interno

$$\langle f, g \rangle = \int_{S^{n-1}} f(\xi)\overline{g(\xi)}\,\mathrm{d}\omega(\xi)$$

e com a representação unitária L do grupo ortogonal

$$(L(T)f)(\xi) = f(T^{-1}\xi),$$

para $f \in \mathrm{Pol}(S^{n-1})_{\leq d}$ e $T \in \mathrm{O}(\mathbb{R}^n)$ (observe que $\mathrm{Pol}(S^{n-1})_{\leq d}$ é um subespaço invariante com relação à representação definida em $L^2(S^{n-1})$).

Denotemos por P_m^n o espaço dos polinômios homogêneos de n variáveis e grau m. Se $f \in P_m^n$ e $x = r\xi$ com $\xi \in S^{n-1}$, então

$$f(r\xi) = r^m f(\xi).$$

Logo f é determinado por sua restrição à esfera e a aplicação entre P_m^n e $L^2(S^{n-1})$ é injetora. Para cada $m \leq d$, podemos então identificar P_m^n com sua imagem em $\mathrm{Pol}(S^{n-1})_{\leq d}$.

Dizemos que um polinômio f é *harmônico* se $\Delta f(x) = 0$, onde

$$\Delta := \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}$$

é conhecido como operador de Laplace. Denotemos por H_m^n o espaço dos polinômios harmônicos e homogêneos de grau m. Da mesma forma que com P_m^n, podemos

identificar H_m^n com o espaço de polinômios esféricos correspondente. Às vezes esse espaço também é chamado de espaço dos polinômios harmônicos esféricos de grau m (omitindo o termo "homogêneos").

2.4.2. Projeção harmônica. Um mesmo polinômio esférico pode corresponder à restrição à esfera de diversos polinômios. Vemos nesta seção que todo polinômio esférico coincide de maneira única com um polinômio harmônico. O seguinte resultado estabelece a chamada projeção harmônica de um polinômio:

PROPOSIÇÃO 2.7. *O espaço P_m^n é soma direta do subespaço H_m^n com o subespaço $r^2 P_{m-2}^n$ dos polinômios da forma $r^2 f$, com $f \in P_{m-2}^n$. Em símbolos:*

$$P_m^n = H_m^n \oplus r^2 P_{m-2}^n.$$

DEMONSTRAÇÃO. Provemos primeiro que

$$H_m^n \cap r^2 P_{m-2}^n = \{0\}.$$

Seja $f \in r^2 P_{m-2}^n$ com $f \neq 0$, mostraremos que $\Delta f(x) \neq 0$. Para isso, escreva $f(x) = r^{2k} g(x)$ com $g \in P_{m-2k}^n$, $0 < k \leq \lfloor m/2 \rfloor$ e g não divisível por r^2.

Temos

$$\begin{aligned}
\Delta f(x) &= \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}\Big(r^{2k} g(x)\Big) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\Big(2k r^{2k-2} x_i g(x) + r^{2k} \frac{\partial}{\partial x_i} g(x)\Big) \\
&= \sum_{i=1}^{n} \Big(4k(k-1) r^{2k-4} x_i^2 g(x) + 2k r^{2k-2} g(x) + 4k r^{2k-2} x_i \frac{\partial}{\partial x_i} g(x) + r^{2k} \frac{\partial^2}{\partial x_i^2} g(x)\Big) \\
&= 4k(k-1) r^{2k-2} g(x) + 2kn r^{2k-2} g(x) + 4k r^{2k-2} \Big(\sum_{i=1}^{n} x_i \frac{\partial}{\partial x_i} g(x)\Big) + r^{2k} \Delta g(x) \\
&= 2k\big(2(k-1) + n + 2(m-2k)\big) r^{2k-2} g(x) + r^{2k} \Delta g(x) \\
&= 2k(n + 2m - 2k - 2) r^{2k-2} g(x) + r^{2k} \Delta g(x).
\end{aligned}$$

Suponha por absurdo que $\Delta f(x) = 0$. A partir de $k \geq 1$ e $m \geq 2k$, obtemos $(n + 2m - 2k - 2) > 0$. Cancelando r^{2k-2}, vemos que g é divisível por r^2, o que é uma contradição. Portanto $\Delta f(x) \neq 0$.

Provemos agora que

$$H_m^n + r^2 P_{m-2}^n = P_m^n.$$

Para isso, calculemos as dimensões desses espaços. Temos que a dimensão de P_m^n é igual ao número de n-uplas de inteiros não-negativos que somam m, que é igual a $\binom{n+m-1}{m}$. Como a dimensão de $r^2 P_{m-2}^n$ é igual à dimensão de P_{m-2}^n e $H_m^n \cap r^2 P_{m-2}^n = \{0\}$, temos que

$$\dim H_m^n \leq \dim P_m^n - \dim P_{m-2}^n,$$

por outro lado, se $f \in P_m^n$, então $\Delta f(x) \in P_{m-2}^n$ e a condição $\Delta f(x) = 0$ impõe não mais do que $\dim(P_{m-2}^n)$ condições lineares nos coeficientes de $f(x)$. Logo,

$$\dim H_m^n \geq \dim P_m^n - \dim P_{m-2}^n$$

e portanto

$$\dim H_m^n = \dim P_m^n - \dim P_{m-2}^n = \binom{n+m-1}{m} - \binom{n+m-3}{m-2}. \tag{28}$$

Como a dimensão de $H_m^n \cap r^2 P_{m-2}^n$ é 0, temos que a dimensão de $H_m^n + r^2 P_{m-2}^n$ é igual a de P_m^n e portanto a soma é igual ao espaço todo. □

Aplicando a proposição anterior a P_{m-2}^n e repetindo o processo, obtemos:

TEOREMA 2.8. *Se f é um polinômio homogêneo de grau m, existem e são únicos polinômios $h_{m-2k} \in H^n_{m-2k}$, com $0 \leq k \leq \lfloor m/2 \rfloor$, tais que*

$$f(x) = \sum_{k=0}^{\lfloor m/2 \rfloor} r^{2k} h_{m-2k}(x).$$

Em particular, para $\xi \in S^{n-1}$ vale

$$f(\xi) = \sum_{k=0}^{\lfloor m/2 \rfloor} h_{m-2k}(\xi). \qquad \square$$

Como todo polinômio é soma de polinômios homogêneos, podemos concluir que todo polinômio de grau d coincide na esfera com um polinômio harmônico (não homogêneo) de grau menor ou igual a d e logo

$$\mathrm{Pol}(S^{n-1})_{\leq d} = H^n_0 + \cdots + H^n_d.$$

Ainda falta determinar se a soma anterior é direta, isto é, se $H^n_k \cap H^n_l = \{0\}$, quando vistos como subespaços de $L^2(S^{n-1})$. A próxima proposição mostra não apenas isso, mas também que esses espaços são ortogonais.

PROPOSIÇÃO 2.9. *Os espaços H^n_k e H^n_l de polinômios harmônicos esféricos de graus k e l, com $k \neq l$, são ortogonais entre si.*

DEMONSTRAÇÃO. Sejam $f_k \in H^n_k$ e $f_l \in H^n_l$. Pelo teorema de Green (veja §1.11 de Arfken e Weber [AW01]), temos

$$\int_{\|x\| \leq 1} f_k(x) \Delta \overline{f_l(x)} - \overline{f_l(x)} \Delta f_k(x) \, \mathrm{d}x = \int_{S^{n-1}} f_k(\xi) \nabla \overline{f_l(\xi)} \cdot \xi - \overline{f_l(\xi)} \nabla f_k(\xi) \cdot \xi \, \mathrm{d}\omega(\xi).$$

Como f_k e f_l são funções harmônicas, o lado esquerdo é nulo. Como f_k é um polinômio homogêneo de grau k e f_l é um polinômio homogêneo de grau l, temos que $\nabla f_k(x) \cdot x = k f_k(x)$ e $\nabla \overline{f_l(x)} \cdot x = l \overline{f_l(x)}$, logo o lado direito é igual a

$$(l-k) \int_{S^{n-1}} f_k(\xi) \overline{f_l(\xi)} \, \mathrm{d}\omega(\xi)$$

e como $k \neq l$, obtemos o resultado desejado. $\square$

2.4.3. Invariância e irredutibilidade de H^n_m. Retornando à representação L de $\mathrm{O}(\mathbb{R}^n)$ sobre $L^2(S^{n-1})$ (definida em (10)), vemos agora que os espaços de polinômios harmônicos e esféricos H^n_m são invariantes.

PROPOSIÇÃO 2.10. *O espaço H^n_m é um subespaço invariante da representação L de $\mathrm{O}(\mathbb{R}^n)$ em $L^2(S^{n-1})$.*

DEMONSTRAÇÃO. Sejam $f \in H^n_m$ e $T = (t_{ij})^n_{i,j=1} \in \mathrm{O}(\mathbb{R}^n)$, desejamos mostrar que $L(T)f \in H^n_m$. Que $L(T)f$ seja um polinômio homogêneo de grau m, segue do fato de T^{-1} ser um operador linear inversível. Que $L(T)f$ seja harmônico, segue do fato de a ação do operador de Laplace comutar com a ação do grupo ortogonal e portanto $\Delta\big(L(T)f(x)\big) = L(T)\big(\Delta f(x)\big) = 0$. Verifiquemos essa afirmação (usamos a seguir que $T^{-1} = T^T$ e portanto que $\sum_{l=1}^n t_{li} t_{lj}$ é igual a 1, se $i = j$, e igual a 0,

caso contrário):

$$\begin{aligned}\Delta\big(L(T)f(x)\big) &= \sum_{l=1}^{n}\frac{\partial^2}{\partial x_l^2}\big(f(T^{-1}x)\big) = \sum_{l=1}^{n}\frac{\partial}{\partial x_l}\Big(\sum_{i=1}^{n}t_{li}\frac{\partial}{\partial x_i}f(T^{-1}x)\Big)\\ &= \sum_{l=1}^{n}\sum_{i=1}^{n}t_{li}\sum_{j=1}^{n}t_{lj}\frac{\partial^2}{\partial x_j\partial x_i}f(T^{-1}x)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\Big(\sum_{l=1}^{n}t_{li}t_{lj}\Big)\frac{\partial^2}{\partial x_j\partial x_i}f(T^{-1}x)\\ &= \sum_{i=1}^{n}\frac{\partial^2}{\partial x_i^2}f(T^{-1}x) = \Delta f(T^{-1}x) = L(T)\big(\Delta f(x)\big). \qquad \square\end{aligned}$$

Agora que sabemos que os espaços H_m^n são invariantes, podemos também observar que as representações de $\mathrm{O}(\mathbb{R}^n)$ nesses espaços não podem ser equivalentes, visto que já calculamos em (28) a dimensão desses espaços e elas diferem entre si.

Para provar a irredutibilidade dos espaços H_m^n é preciso considerar funções zonais esféricas. Dizemos que uma função $f\colon S^{n-1}\to\mathbb{C}$ é *zonal esférica com polo* $e\in S^{n-1}$ se f é invariante sob a ação do subgrupo de $\mathrm{O}(\mathbb{R}^n)$ que estabiliza e, isto é,

$$\{T\in\mathrm{O}(\mathbb{R}^n) : Te = e\}.$$

O valor de uma função zonal esférica em um ponto $\xi\in S^{n-1}$ depende apenas do produto interno $e\cdot\xi$, logo as funções zonais esféricas podem ser associadas a funções no intervalo $[-1,1]$.

O próximo lema mostra que as funções zonais esféricas formam um subespaço de H_m^n de dimensão 1. A irredutibilidade de H_m^n é provada na proposição seguinte como uma consequência disso.

Lema 2.11. *A menos de um fator multiplicativo, existe exatamente um polinômio harmônico e homogêneo de grau m que é invariante sob todas as transformações ortogonais que fixam um ponto $e\in S^{n-1}$. Esse polinômio é dado pela expressão*

$$f(x) = \sum_{k=0}^{m} c_k (x\cdot e)^{m-k}(r^2-(x\cdot e)^2)^{k/2},$$

onde $c_k = 0$ se k é ímpar e

$$c_{k+2} = -c_k\frac{(m-k)(m-k-1)}{(k+2)(n+k-1)}, \qquad \text{para } 0\le k\le m-2.$$

Demonstração. Podemos supor que $e=(1,0,\dots,0)$, pois para qualquer outro $e'\in S^{n-1}$, existe $T\in\mathrm{O}(\mathbb{R}^n)$ tal que $Te'=e$ e se f é uma função zonal esférica com polo e', pode-se verificar que $L(T)f$ é uma função zonal esférica com polo e.

Provemos primeiro que existe no máximo um polinômio como no enunciado. Seja $f\in H_m^n$ invariante sob a ação das transformações ortogonais que fixam e. Logo $f(x_1,\dots,x_n) = f(x_1,\sqrt{r^2-x_1^2},0,\dots,0)$ (estamos denotando $\sum_{i=2}^n x_i^2$ por $r^2-x_1^2$) e podemos olhar apenas para os monômios de f que contêm a primeira e a segunda variáveis. Temos

$$f(x) = \sum_{k=0}^{m} c_k x_1^{m-k}(r^2-x_1^2)^{k/2}$$

e como também vale que $f(x_1,\dots,x_n) = f(x_1,-\sqrt{r^2-x_1^2},0,\dots,0)$, temos também

$$f(x) = \sum_{k=0}^{m} c_k(-1)^k x_1^{m-k}(r^2-x_1^2)^{k/2},$$

logo $c_k=0$ para todo k ímpar.

A partir de $\Delta f(x) = 0$, podemos obter uma relação entre os coeficientes c_k:

$$\begin{aligned}
\Delta f(x) &= \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}\Big(\sum_{k=0}^{m} c_k x_1^{m-k}(r^2 - x_1^2)^{k/2}\Big) \\
&= \sum_{k=0}^{m-2} c_k(m-k)(m-k-1)x_1^{m-k-2}(r^2-x_1^2)^{k/2} \\
&\qquad + \sum_{i=2}^{n}\sum_{k=2}^{m} c_k x_1^{m-k} k\Big((r^2-x_1^2)^{\frac{k-2}{2}} + (k-2)x_i^2(r^2-x_1^2)^{\frac{k-2}{2}-1}\Big) \\
&= \sum_{k=0}^{m-2} c_k(m-k)(m-k-1)x_1^{m-k-2}(r^2-x_1^2)^{k/2} \\
&\qquad + \sum_{k=2}^{m} c_k k(n+k-3)x_1^{m-k}(r^2-x_1^2)^{\frac{k-2}{2}} \\
&= \sum_{k=0}^{m-2} \Big(c_k(m-k)(m-k-1) + c_{k+2}(k+2)(n+k-1)\Big)x_1^{m-k-2}(r^2-x_1^2)^{k/2},
\end{aligned}$$

e portanto,

$$c_{k+2} = -c_k \frac{(m-k)(m-k-1)}{(k+2)(n+k-1)}$$

para todo k entre 0 e $m-2$.

Essa recorrência determina todos os coeficientes c_k como múltiplos de c_0, assim provamos que existe no máximo um polinômio, como queríamos.

Reciprocamente, seja f como no enunciado. Como seus coeficientes satisfazem a relação de recorrência, temos $\Delta f = 0$. Vemos diretamente de sua expressão que f é um polinômio homogêneo de grau m e vemos também que determina uma função zonal esférica de polo e, pois sua expressão depende apenas de $x \cdot e$ e r^2. □

Proposição 2.12. *O espaço H_m^n é um subespaço irredutível da representação L de $\mathrm{O}(\mathbb{R}^n)$ em $L^2(S^{n-1})$.*

Demonstração. Seja F um subespaço invariante e não-nulo de H_m^n. Desejamos mostrar que F não pode ser um subespaço próprio.

Consideremos $\{R_1, \ldots, R_h\}$ uma base ortonormal de F e seja $R\colon S^{n-1} \to \mathbb{C}^h$ o vetor

$$R(\xi) := \begin{pmatrix} R_1(\xi) \\ \vdots \\ R_h(\xi) \end{pmatrix}.$$

Podemos definir o kernel $K \in F^{(2)}$,

$$K(\xi,\eta) := \sum_{i=1}^{h} \overline{R_i(\xi)} R_i(\eta) = R(\xi)^* R(\eta).$$

Como a representação de $\mathrm{O}(\mathbb{R}^n)$ é unitária e F é um subespaço invariante, a ação de uma matriz $T \in \mathrm{O}(\mathbb{R}^n)$ na base $\{R_1(\xi), \ldots, R_h(\xi)\}$ produz outra base ortonormal de F, logo existe uma matriz unitária U tal que $R(T\xi) = UR(\xi)$ e

$$K(T\xi, T\eta) = R(\xi)^* U^* U R(\eta) = R(\xi)^* R(\eta) = K(\xi,\eta).$$

Portanto $K(\xi,\eta)$ é invariante sob a ação de $\mathrm{O}(\mathbb{R}^n)$. Fixando ξ e fazendo $f(\eta) := K(\xi,\eta)$, temos que $f \in F$, que f é não-nulo (por ser uma combinação de funções linearmente independentes) e que f é invariante sob a ação das transformações ortogonais que fixam ξ.

Caso F seja um subespaço próprio, então o complemento ortogonal G de F em H^n_m é não-nulo e pelo mesmo argumento podemos definir $g \in G$ invariante sob a ação das transformações ortogonais que fixam ξ. Como F e G são ortogonais, as funções f e g são linearmente independentes, o que contradiz o Lema 2.11.

Concluímos que H^n_m não pode possuir subespaços invariantes próprios e não-nulos sob a ação de $\mathrm{O}(\mathbb{R}^n)$ e portanto é irredutível. □

Observe que considerando a Proposição 1.18, os resultados desta seção dão-nos uma demonstração alternativa da Proposição 2.9.

Os resultados provados até aqui podem ser agrupados no seguinte teorema:

Teorema 2.13. *O espaço* $\mathrm{Pol}(S^{n-1})_{\leq d}$ *dos polinômios esféricos de grau no máximo* d*, sob a ação do grupo ortogonal* $\mathrm{O}(\mathbb{R}^n)$*, decompõe-se como soma ortogonal de subespaços invariantes, irredutíveis e não-equivalentes entre si:*

$$\mathrm{Pol}(S^{n-1})_{\leq d} = H^n_0 \perp \cdots \perp H^n_d. \qquad \square$$

2.4.4. Polinômios de Gegenbauer e kernels $\mathrm{O}(\mathbb{R}^n)$-invariantes. Vimos no Lema 2.11 que o valor em $\xi \in S^{n-1}$ de um polinômio harmônico, homogêneo de grau m e zonal esférico de polo e é dado por um polinômio de grau m em $\xi \cdot e$. Vemos agora que esses polinômios são múltiplos dos polinômios de Gegenbauer, uma conhecida família de polinômios ortogonais, e que podem ser usados para definir kernels positivos e $\mathrm{O}(\mathbb{R}^n)$-invariantes.

Comecemos definindo esses polinômios. A partir do Lema 2.11 e com alguns ajustes nos parâmetros, definimos para todo n inteiro positivo e m inteiro não-negativo o polinômio:

$$P^n_m(u) := \sum_{k=0}^{\lfloor m/2 \rfloor} c_k u^{m-2k}(1-u^2)^k, \tag{29}$$

com $c_0 = 1$ e coeficientes c_k definidos pela recorrência

$$c_k = -c_{k-1}\frac{(m-2k+2)(m-2k+1)}{2k(n+2k-3)}, \qquad \text{para } 1 \leq k \leq \lfloor m/2 \rfloor. \tag{30}$$

A primeira propriedade que estabeleceremos é a chamada "fórmula de adição", que ressalta a ligação desses polinômios com H^n_m.

Proposição 2.14 (Fórmula de adição). *Se* $\{R_1, \ldots, R_{h^n_m}\}$ *é uma base ortonormal de* H^n_m*, então para quaisquer* ξ *e* η *em* S^{n-1}*, vale:*

$$P^n_m(\xi \cdot \eta) = \frac{1}{h^n_m}\sum_{i=1}^{h^n_m} \overline{R_i(\xi)} R_i(\eta). \tag{31}$$

Demonstração. A exemplo do que foi feito na demonstração da Proposição 2.12, seja $K \in (H^n_m)^{(2)}$,

$$K(\xi,\eta) := \sum_{i=1}^{h^n_m} \overline{R_i(\xi)} R_i(\eta).$$

Temos que $K(\xi,\eta)$ é $\mathrm{O}(\mathbb{R}^n)$-invariante. Fixando ξ, temos que $p(\eta) := K(\xi,\eta)$ pertence a H^n_m e é invariante sob a ação de transformações ortogonais que fixam ξ, logo pelo Lema 2.11 e pela forma como os polinômios P^n_m foram definidos, temos

$$K(\xi,\eta) = cP^n_m(\xi \cdot \eta)$$

para alguma constante c.

Falta determinar que $c = h_m^n$. Para isso observe que, pela expressão (29), $P_m^n(1) = 1$ e, lembrando que $\int_{S^{n-1}} \mathrm{d}\omega(x) = 1$,

$$K(\xi,\xi) = \int_{S^{n-1}} K(\xi,\xi)\,\mathrm{d}\omega(\xi) = \sum_{i=1}^{h_m^n} \int_{S^{n-1}} \overline{R_i(\xi)} R_i(\xi)\,\mathrm{d}\omega(\xi) = \sum_{i=1}^{h_m^n} 1 = h_m^n. \quad \square$$

A partir da fórmula de adição segue facilmente a Proposição 2.5, usada na prova do limitante de programação linear e que, em conjunto com a Proposição 1.22, estabelece a positividade dos kernels $(x,y) \mapsto P_k^n(x \cdot y)$.

Proposição 2.5. *Para todo C, subconjunto finito de S^{n-1} e $w\colon C \to \mathbb{C}$, vale:*

$$\sum_{c,c' \in C} w(c)\overline{w(c')} P_k^n(c \cdot c') \geq 0.$$

Demonstração. Pela fórmula de adição (Proposição 2.14),

$$\begin{aligned}\sum_{c,c' \in C} w(c)\overline{w(c')} P_k^n(c \cdot c') &= \sum_{c,c' \in C} w(c)\overline{w(c')} \frac{1}{h_k^n} \sum_{i=1}^{h_k^n} R_i(c)\overline{R_i(c')} \\ &= \frac{1}{h_k^n} \sum_{i=1}^{h_k^n} \Big| \sum_{c \in C} w(c) R_i(c) \Big|^2 \geq 0. \quad \square\end{aligned}$$

Para $\lambda > -1/2$, os *polinômios de Gegenbauer* $C_k^\lambda(u)$ (também conhecidos como polinômios ultraesféricos — veja Andrews, Askey e Roy [AAR99], Seção 6.4 ou Szegö [Sze39], Seção 4.7) são uma família de polinômios ortogonais com relação ao peso $(1-u^2)^{\lambda-1/2}$ no intervalo $[-1,1]$. Isso significa que eles são dois-a-dois ortogonais com relação ao produto interno

$$(f,g)_\lambda = \int_{-1}^{1} f(u)g(u)(1-u^2)^{\lambda-1/2}\,\mathrm{d}u$$

do espaço $L^2([-1,1])$. Essa condição, junto com a informação de que o polinômio $C_k^\lambda(u)$ possui grau k, determina esses polinômios a menos de um fator multiplicativo; na literatura eles são usualmente encontrados com a normalização $C_k^\lambda(1) = \binom{k+2\lambda-1}{k}$. A relação de ortogonalidade também permite identificá-los com os chamados polinômios de Jacobi (veja e.g., Szegö [Sze39], Capítulo 4) de parâmetros $\alpha = \beta = \lambda - 1/2$.

Proposição 2.15. *Os polinômios $P_k^n(u)$ definidos em (29) e (30) são múltiplos dos polinômios de Gegenbauer com parâmetro $\lambda = (n-2)/2$. Temos*

$$P_k^n(u) = C_k^{(n-2)/2}(u)/C_k^{(n-2)/2}(1).$$

Demonstração. Precisamos mostrar que os polinômios $P_k^n(u)$ são dois-a-dois ortogonais com relação ao peso $(1-u^2)^{(n-3)/2}$ no intervalo $[-1,1]$.

Fixando $k \neq l$ e $\eta \in S^{n-1}$, pela Proposição 2.14 temos que $f(\xi) := P_k^n(\xi \cdot \eta) \in H_k^n$ e $g(\xi) := P_l^n(\xi \cdot \eta) \in H_l^n$. Como pela Proposição 2.9 esses espaços são ortogonais, vale

$$\int_{S^{n-1}} P_k^n(\xi \cdot \eta) P_l^n(\xi \cdot \eta)\,\mathrm{d}\omega(\xi) = 0. \tag{32}$$

A seguir precisamos distinguir as medidas superficiais das esferas S^{n-1} e S^{n-2}, portanto denotamos por ω_n a medida normalizada de S^{n-1} (que até aqui, como na equação anterior, denotamos por ω). Escrevendo $\xi = u\eta + \sqrt{1-u^2}\zeta$ com $u \in [-1,1]$, $\zeta \in S^{n-2} \subset (\mathbb{R}\eta)^\perp$ e usando que (Corolário A.5 de Axler, Bourdon e

Ramey [ABR01])

$$\mathrm{d}\omega_n(\xi) = \frac{(n-1)\Gamma(n/2+1)}{n\pi^{1/2}\Gamma(n/2+1/2)}(1-u^2)^{(n-3)/2}\,\mathrm{d}u\,\mathrm{d}\omega_{n-1}(\zeta), \tag{33}$$

a integral em (32) pode ser reescrita como

$$\begin{aligned}\frac{(n-1)\Gamma(n/2+1)}{n\pi^{1/2}\Gamma(n/2+1/2)}\int_{S^{n-2}}\int_{-1}^{1} P_k^n(u)P_l^n(u)(1-u^2)^{(n-3)/2}\,\mathrm{d}u\,\mathrm{d}\omega_{n-1}(\zeta)\\ = \frac{(n-1)\Gamma(n/2+1)}{n\pi^{1/2}\Gamma(n/2+1/2)}\int_{-1}^{1} P_k^n(u)P_l^n(u)(1-u^2)^{(n-3)/2}\,\mathrm{d}u,\end{aligned}$$

o que mostra a ortogonalidade mencionada. □

Diversas propriedades dos polinômios de Gegenbauer (que como já mencionado, são um caso particular dos polinômios de Jacobi) podem ser encontradas na literatura; veja por exemplo Szegö [Sze39], em especial a fórmula (4.5.1), que fornece outro método para calcular os polinômios P_k^n explicitamente.

A principal aplicação dos polinômios de Gegenbauer é a descrição dos kernels positivos, contínuos e $\mathrm{O}(\mathbb{R}^n)$-invariantes. A partir do Teorema 1.24, considerando a decomposição de $\mathrm{Pol}(S^{n-1})_{\leq d}$ vista no Teorema 2.13 (no caso, todos os subespaços irredutíveis não são equivalentes entre si e as componentes isotípicas possuem multiplicidade 1) e a Proposição 2.14, temos que um kernel positivo e invariante de $\mathrm{Pol}(S^{n-1})^{(2)}_{\leq d}$ é dado por uma combinação positiva dos kernels $(\xi,\eta)\mapsto P_m^n(\xi\cdot\eta)$, com $m=0,\ldots,d$. Note que como as componentes isotípicas possuem multiplicidade 1, a condição de positividade de um kernel invariante implica na sua imagem ser real.

Segue do Teorema de Stone-Weierstrass (Teorema IV.9 em Reed e Simon [RS72]) que $\mathrm{Pol}(S^{n-1})^{(2)}$ é denso em $\mathcal{C}(S^{n-1}\times S^{n-1})$, logo um kernel contínuo, positivo e invariante pode ser aproximado com qualquer precisão por combinações positivas dos kernels $(\xi,\eta)\mapsto P_m^n(\xi\cdot\eta)$, com $m=0,1,\ldots$. Terminemos esta seção demonstrando o Teorema de Schoenberg [Sch42], enunciado na Seção 2.2, que estabelece esse resultado, acrescentando a convergência absoluta e uniforme dessas aproximações.

Teorema 2.4 (Teorema de Schoenberg). *Um kernel $K\colon S^{n-1}\times S^{n-1}\to\mathbb{R}$ é contínuo, positivo e $\mathrm{O}(\mathbb{R}^n)$-invariante se e somente se:*

$$K(x,y)=\sum_{k=0}^{\infty} a_k P_k^n(x\cdot y), \tag{34}$$

com $a_k\geq 0$, para todo k e tais que $\sum_{k=0}^{\infty}a_k$ converge. Nesse caso, a série acima converge absolutamente e uniformemente em $S^{n-1}\times S^{n-1}$.

Demonstração. Provemos primeiro que se $a_0,a_1,\ldots$ são não-negativos e tais que $\sum_{k=0}^{\infty}a_k$ converge, então o kernel K definido em (34) é contínuo, positivo e $\mathrm{O}(\mathbb{R}^n)$-invariante.

Segue da Proposição 2.5 que para todo $u\in[-1,1]$ e k inteiro não-negativo, $|P_k^n(u)|\leq P_k^n(1)=1$ e portanto a série $\sum_{k=0}^{\infty}a_kP_k^n(x\cdot y)$ converge absolutamente para todos $x,y\in S^{n-1}$.

Como $|\sum_{k=m}^{\infty}a_kP_k^n(u)|\leq\sum_{k=m}^{\infty}a_k$, a série $\sum_{k=0}^{\infty}a_kP_k^n(x\cdot y)$ converge uniformemente. Como o limite uniforme de funções contínuas é contínuo, o kernel K dado pela equação (34) é contínuo.

O kernel K é positivo pois pelas Proposições 2.5 e 1.22, os kernels $(x,y)\mapsto P_k^n(x\cdot y)$ são positivos e os coeficientes a_k são todos maiores ou iguais a zero. Por fim, o kernel K é $\mathrm{O}(\mathbb{R}^n)$-invariante, pois sua expressão depende apenas de $x\cdot y$.

Provemos agora que se $K\colon S^{n-1}\times S^{n-1}\to\mathbb{R}$ é um kernel contínuo, positivo e $\mathrm{O}(\mathbb{R}^n)$-invariante, então existem números $a_0, a_1, \dots$ não-negativos tais que $\sum_{k=0}^{\infty} a_k$ converge e K satisfaz a expressão (34).

Seja $f\colon [-1,1]\to\mathbb{R}$ a função contínua tal que $K(x,y)=f(x\cdot y)$. Como os polinômios P_k^n formam um sistema ortogonal completo em $L^2([-1,1])$ com o produto interno

$$(f,g)_n := \int_{-1}^{1} f(u)g(u)(1-u^2)^{(n-3)/2}\,\mathrm{d}u,$$

existem números $a_0, a_1, \dots$ tais que

$$f=\sum_{k=0}^{\infty} a_k P_k^n,$$

com convergência em $L^2([-1,1])$.

Provemos primeiro que $a_l\geq 0$ para todo l inteiro não-negativo. Por um lado, $(f,P_l^n)_n=\sum_{k=0}^{n} a_k(P_k^n,P_l^n)_n = a_l(P_l^n,P_l^n)_n$. Por outro lado, usando c para denotar $\frac{(n-1)\Gamma(n/2+1)}{n\pi^{1/2}\Gamma(n/2+1/2)}$ e a equação (33), temos

$$(f,P_l^n)_n=\int_{-1}^{1} f(u)P_l^n(u)(1-u^2)^{(n-3)/2}\,\mathrm{d}u = c\int_{S^{n-1}} f(\xi\cdot\eta)P_l^n(\xi\cdot\eta)\,\mathrm{d}\omega(\xi),$$

para qualquer $\eta\in S^{n-1}$ e então se $\{R_1,\dots,R_{h_l^n}\}$ for uma base ortonormal de H_l^n, usando a Proposição 2.14,

$$(f,P_l^n)_n = c\int_{S^{n-1}}\int_{S^{n-1}} f(\xi\cdot\eta)P_l^n(\xi\cdot\eta)\,\mathrm{d}\omega(\xi)\,\mathrm{d}\omega(\eta) = \frac{c}{h_l^n}\sum_{i=1}^{h_l^n}\langle T_K R_i, R_i\rangle \geq 0,$$

pois K é um kernel positivo. Disso resulta $a_l\geq 0$.

Vejamos agora que a série $\sum_{k=0}^{\infty} a_k$ converge. Para cada $m=0,1,\dots$, considere a função

$$f_m(u) := f(u) - \sum_{k=0}^{m} a_k P_k^n(u).$$

Essa função é contínua e além disso, $f_m=\sum_{k=m+1}^{\infty} a_k P_k^n$ com convergência em $L^2([-1,1])$ e como, pelas Proposições 2.5 e 1.22, os kernels $(x,y)\mapsto P_k^n(x\cdot y)$ são positivos, temos que $(x,y)\to f_m(x\cdot y)$ também é positivo e logo $f_m(1)\geq 0$ para qualquer m. Disso resulta que

$$f(1)-\sum_{k=0}^{m} a_k = f(1)-\sum_{k=0}^{m} a_k P_k^n(1) = f_m(1)\geq 0,$$

e portanto a série $\sum_{k=0}^{\infty} a_k$ é monótona e limitada superiormente por $f(1)$, logo convergente. □

CAPÍTULO 3

O limitante de programação semidefinida

Considerando kernels invariantes sob a ação de apenas um subgrupo de $\mathrm{O}(\mathbb{R}^n)$, obtemos um aperfeiçoamento do limitante apresentado no capítulo anterior, devido a Bachoc e Vallentin [BV08]. Na Proposição 3.4 obtemos matrizes cujos coeficientes satisfazem uma propriedade semelhante à vista na Proposição 2.5 e, seguindo o artigo de Bachoc e Vallentin [BV08], derivamos o limitante diretamente a partir dessas propriedades. As demonstrações de algumas propriedades sobre os polinômios harmônicos são apresentadas em uma seção no fim do capítulo. Na Seção 5.5 veremos uma outra forma de obter o limitante como um aperfeiçoamento do número teta de Lovász.

3.1. A ação do subgrupo estabilizador de um ponto

Fixemos um ponto $e \in S^{n-1}$ e consideremos o subgrupo H que estabiliza e sob a ação de $\mathrm{O}(\mathbb{R}^n)$, isto é,

$$H := \{T \in \mathrm{O}(\mathbb{R}^n) : Te = e\}.$$

Escrevendo $\mathbb{R}^n$ como soma direta entre o subespaço gerado por e e seu complemento ortogonal, vemos que H é isomorfo a $\mathrm{O}(\mathbb{R}^{n-1})$.

Quando consideramos a ação de H em $L^2(S^{n-1})$, como H é um subgrupo de $\mathrm{O}(\mathbb{R}^n)$, os espaços de polinômios harmônicos e homogêneos H_m^n ainda são invariantes mas já não são mais irredutíveis. Cada um desses espaços pode ser decomposto em subespaços invariantes e irredutíveis menores que, conforme indicado pelo isomorfismo entre H e $\mathrm{O}(\mathbb{R}^{n-1})$, são equivalentes a espaços H_i^{n-1}. Na Seção 3.3 demonstramos o seguinte teorema:

Teorema 3.1. *O espaço H_m^n sob a ação de H (o subgrupo de $\mathrm{O}(\mathbb{R}^n)$ que estabiliza $e \in \mathbb{R}^n$, fixado) decompõe-se como soma ortogonal de subespaços invariantes, irredutíveis e não equivalentes entre si:*

$$H_m^n = H_{0,m}^{n-1} \perp H_{1,m}^{n-1} \perp \cdots \perp H_{m,m}^{n-1},$$

com a representação de H em $H_{i,m}^{n-1}$ equivalente à representação de $\mathrm{O}(\mathbb{R}^{n-1})$ em H_i^{n-1}.

Escrevendo essa decomposição em conjunto com a que tínhamos no Teorema 2.13, obtemos uma decomposição de $\mathrm{Pol}(S^{n-1})_{\leq d}$ em subespaços invariantes e irredutíveis sob a ação de H:

$$\begin{array}{rcccccccc} \mathrm{Pol}(S^{n-1})_{\leq d} & = & H_0^n & \perp & H_1^n & \perp & \cdots & \perp & H_d^n \\ & = & H_{0,0}^{n-1} & \perp & H_{0,1}^{n-1} & \perp & \cdots & \perp & H_{0,d}^{n-1} \\ & & & \perp & H_{1,1}^{n-1} & \perp & \cdots & \perp & H_{1,d}^{n-1} \\ & & & & & & \cdots\cdots\cdots & & \\ & & & & & & & \perp & H_{d,d}^{n-1}. \end{array}$$

Agrupando os subespaços irredutíveis e equivalentes entre si, obtemos

$$\begin{aligned} \mathrm{Pol}(S^{n-1})_{\leq d} &= I_0 \perp \cdots \perp I_d, \\ I_k &= H_{k,k}^{n-1} \perp \cdots \perp H_{k,d}^{n-1} \simeq (d-k+1)H_k^{n-1}. \end{aligned} \tag{35}$$

Podemos usar essa decomposição em conjunto com o Teorema 1.24 para descrever os kernels positivos e H-invariantes de $\mathrm{Pol}(S^{n-1})^{(2)}_{\leq d}$. Conforme descrito pelo teorema, para cada $k = 0, \ldots, d$, precisamos construir bases ortogonais para os subespaços invariantes e irredutíveis de I_k com vetores de mesma norma e tais que as matrizes das subrepresentações irredutíveis e equivalentes sejam iguais; isso será feito na demonstração da Proposição 3.2, porém considerando uma decomposição alternativa de I_k como soma direta de subespaços invariantes e irredutíveis

$$I_k = S_{k,0} \oplus \cdots \oplus S_{k,d-k}. \tag{36}$$

Para cada $i = 0, \ldots, d-k$, seja $\{f_{k,i,1}, \ldots, f_{k,i,h_k^{n-1}}\}$ a base de $S_{k,i}$ com as características descritas e seja F_k^n a matriz de tamanho $(d-k+1) \times (d-k+1)$ cujas entradas são os kernels

$$\big(F_k^n\big)_{i,j}(x,y) = \sum_{s=1}^{h_k^{n-1}} f_{k,i,s}(x)\overline{f_{k,j,s}(y)},$$

com $i, j = 0, \ldots, d-k$.

Conforme descrito no Teorema 1.24, um kernel $K \in \mathrm{Pol}(S^{n-1})^{(2)}_{\leq d}$ positivo e H-invariante pode ser descrito através de matrizes positivo-semidefinidas Λ_k de tamanho $(d-k+1) \times (d-k+1)$:

$$K = \sum_{k=0}^{d} \sum_{i,j=0}^{d-k} \lambda_{i,j}^k \big(F_k^n\big)_{i,j} = \sum_{k=0}^{d} \left\langle \Lambda_k, \overline{F_k^n} \right\rangle.$$

Como os kernels $\big(F_k^n\big)_{i,j}$ são H-invariantes, o valor de $\big(F_k^n\big)_{i,j}(x,y)$ depende apenas da órbita de (x,y) sob a ação de H. Tal órbita é determinada pelos produtos internos $u = e \cdot x$, $v = e \cdot y$ e $t = x \cdot y$, logo podemos reescrever a expressão anterior através de funções nas variáveis u, v e t definindo matrizes Y_k^n cujas entradas são tais que

$$\big(Y_k^n\big)_{i,j}(e \cdot x, e \cdot y, x \cdot y) := \big(F_k^n\big)_{i,j}(x,y)$$

para todo $i, j = 0, \ldots, d-k$.

A seguir, calculamos os coeficientes das matrizes Y_k^n explicitamente.

Proposição 3.2. *Para $k = 0, \ldots, d$ e $i, j = 0, \ldots, d-k$, as entradas das matrizes Y_k^n definidas anteriormente são:*

$$(Y_k^n)_{i,j}(u,v,t) = u^i v^j Q_k^{n-1}(u,v,t), \tag{37}$$

com

$$Q_k^{n-1}(u,v,t) := \big((1-u^2)(1-v^2)\big)^{k/2} P_k^{n-1}\left(\frac{t-uv}{\sqrt{(1-u^2)(1-v^2)}}\right),$$

com as funções P_k^n sendo os polinômios ortogonais de Gegenbauer, introduzidos na Seção 2.4.4.

Demonstração. Podemos escrever $x \in S^{n-1}$ como soma entre sua componente na direção de e e seu complemento ortogonal

$$x = ue + \sqrt{1-u^2}\zeta,$$

com $u = x \cdot e$ e $\zeta \in S^{n-2} \subset (\mathbb{R}e)^\perp$. Para cada $i = 0, \ldots, d-k$, definimos $\varphi_i \colon H_k^{n-1} \to \mathrm{Pol}(S^{n-1})_{\leq d}$ de modo que para $f \in H_k^{n-1}$,

$$\varphi_i(f)(x) := \big(1/\sqrt{h_k^{n-1}}\big) u^i (1-u^2)^{k/2} f(\zeta).$$

Note que f é um polinômio homogêneo de grau k em $n-1$ variáveis e a multiplicação por $(1-u^2)^{k/2}$ garante que $\varphi_i(f)$ seja uma função polinomial de grau $k+i$ nas coordenadas de x. A transformação φ_i comuta com as ações de $\mathrm{O}(\mathbb{R}^{n-1})$ em

H_k^{n-1} e de H em $\mathrm{Pol}(S^{n-1})_{\leq d}$ e assim $\varphi_i(H_k^{n-1})$ é um subespaço de $\mathrm{Pol}(S^{n-1})_{\leq d}$ equivalente à H_k^{n-1}. Obtemos então $d-k+1$ subespaços independentes entre si e equivalentes à H_k^{n-1}, tal como na decomposição (36). Fixando uma base ortonormal $\{g_{k,1},\dots,g_{k,h_k^{n-1}}\}$ em H_k^{n-1}, usamos suas imagens $f_{k,i,s} := \varphi_i(g_{k,s})$ para determinar bases em $\varphi_i(H_k^{n-1})$ e calcular $(Y_k^n)_{i,j}(u,v,t)$.

Sejam $x,y \in S^{n-1}$ tais que $x = ue + \sqrt{1-u^2}\zeta$, $y = ve + \sqrt{1-v^2}\xi$ e $x\cdot y = t$, então:

$$\begin{aligned}
(Y_k^n)_{i,j}(u,v,t) &= \sum_{s=1}^{h_k^{n-1}} f_{k,i,s}(x)\overline{f_{k,j,s}(y)} \\
&= \frac{1}{h_k^{n-1}} \sum_{s=1}^{h_k^{n-1}} u^i(1-u^2)^{k/2} g_s(\zeta) v^j (1-v^2)^{k/2} \overline{g_s(\xi)} \\
&= u^i v^j \big((1-u^2)(1-v^2)\big)^{k/2} \frac{1}{h_k^{n-1}} \sum_{s=1}^{h_k^{n-1}} g_s(\zeta)\overline{g_s(\xi)} \\
&= u^i v^j \big((1-u^2)(1-v^2)\big)^{k/2} P_k^{n-1}(\zeta\cdot\xi) \\
&= u^i v^j \big((1-u^2)(1-v^2)\big)^{k/2} P_k^{n-1}\left(\frac{t-uv}{\sqrt{(1-u^2)(1-v^2)}}\right),
\end{aligned}$$

sendo que na penúltima passagem usamos a Proposição 2.14 e na última usamos que $t = (ue+\sqrt{1-u^2}\zeta)\cdot(ve+\sqrt{1-v^2}\xi)$. □

Os resultados discutidos até aqui podem ser resumidos pelo seguinte teorema.

Teorema 3.3. *Um kernel $K \in \mathrm{Pol}(S^{n-1})^{(2)}_{\leq d}$ é positivo e H-invariante se e somente se existem matrizes positivo-semidefinidas Λ_k de tamanho $(d-k+1)\times(d-k+1)$, para $k=0,\dots,d$ tais que*

$$K(x,y) = \sum_{k=0}^{d} \left\langle \Lambda_k, Y_k^n(e\cdot x, e\cdot y, x\cdot y)\right\rangle,$$

com as matrizes Y_k^n definidas na Proposição 3.2. Ademais, se K é um kernel real, seus coeficientes Λ_k também são reais. □

Algumas observações sobre a última proposição e teorema:

(1) Existem diversas formas de decompor I_k como soma de subespaços invariantes e irredutíveis, levando a diferentes construções de Y_k^n. Porém, conforme descrito na demonstração do Teorema 1.24, essas opções levam a transformações do tipo $Y_k^n \mapsto AY_k^nA^*$, com A sendo uma matriz inversível, o que por sua vez leva a transformações semelhantes nas matrizes Λ_k.

(2) Note que as entradas de Y_k^n são polinômios. Isso segue do fato de $P_k^{n-1}(u)$ ser um polinômio par ou ímpar, dependendo de k (isso pode ser observado diretamente na expressão (29)). Também pode-se observar que esses polinômios possuem grau menor ou igual a $2d$, $Y_0^n(1,1,1)$ é igual a J_{d+1} (a matriz em $\mathbb{R}^{(d+1)\times(d+1)}$ com 1 em todas as suas entradas) e $Y_k^n(1,1,1)$ é a matriz nula se $k>0$.

(3) A afirmação sobre um kernel com imagem real poder ser expresso por matrizes Λ_k reais segue do fato de todas as entradas de Y_k^n serem polinômios linearmente independentes e de imagem real.

A relação entre as matrizes Y_k^n e os kernels positivos nos permite mostrar o seguinte resultado (compare-o com a Proposição 2.5):

PROPOSIÇÃO 3.4. *Para todo subconjunto finito $C \subset S^{n-1}$, $k = 0, \dots, d-k$ e $\mu\colon C \to \mathbb{C}$, temos*

$$\sum_{x,y\in C} \mu(x)\overline{\mu(y)}Y_k^n(e\cdot x, e\cdot y, x\cdot y) \succeq 0.$$

DEMONSTRAÇÃO. Dado $\omega \in \mathbb{C}^{d-k+1}$, seja $\Lambda_k = \omega\omega^*$. Pelo Teorema 3.3, o kernel $K(x,y) = \langle \Lambda_k, Y_k^n(e\cdot x, e\cdot y, x\cdot y)\rangle$ é positivo e pela Proposição 1.22, a restrição de K a $C\times C$ é uma matriz positivo-semidefinida, logo:

$$\begin{aligned} 0 &\le \sum_{x,y\in C} \mu(x)\overline{\mu(y)}K(x,y) = \sum_{x,y\in C} \mu(x)\overline{\mu(y)} \sum_{i,j=0}^{d-k} \omega_i\overline{\omega_j}(Y_k^n)_{i,j}(e\cdot x, e\cdot y, x\cdot y) \\ &= \sum_{i,j=0}^{d-k} \omega_i\overline{\omega_j} \sum_{x,y\in C} \mu(x)\overline{\mu(y)}(Y_k^n)_{i,j}(e\cdot x, e\cdot y, x\cdot y). \end{aligned}$$

□

No início desse capítulo fixamos um ponto $e \in S^{n-1}$ e então tudo que se seguiu, incluindo a definição das matrizes Y_k^n, foram dependentes desse ponto. Essa escolha é artificial, visto que nenhum ponto da esfera é especial na definição dos códigos esféricos. Uma forma de levar isso em conta é definindo a matriz S_k^n para $k = 0, \dots, d-k$,

$$S_k^n(u,v,t) := \frac{1}{6}\sum_{\sigma\in\mathcal{S}_3} Y_k^n(\sigma(u,v,t)), \tag{38}$$

com $\mathcal{S}_3$ sendo o grupo de permutações das variáveis u, v e t. As matrizes S_k^n satisfazem uma relação semelhante à Proposição 3.4, que será usada na próxima seção.

PROPOSIÇÃO 3.5. *Para todo subconjunto finito $C \subset S^{n-1}$ e $k = 0, \dots, d-k$, temos:*

$$\sum_{x,y,z\in C} S_k^n(z\cdot x, z\cdot y, x\cdot y) \succeq 0.$$

DEMONSTRAÇÃO. Observemos inicialmente que a ação de $\mathrm{O}(\mathbb{R}^n)$ em S^{n-1} é transitiva, isto é, para todo $z \in S^{n-1}$, existe $\psi_z \in \mathrm{O}(\mathbb{R}^n)$ tal que $\psi_z e = z$. A ação de $\mathrm{O}(\mathbb{R}^n)$ também preserva o produto interno, isto é, $(\psi x)\cdot(\psi y) = x\cdot y$ para todos $\psi \in \mathrm{O}(\mathbb{R}^n)$ e $x, y \in S^{n-1}$. A seguir, dados $C \subset S^{n-1}$ e $\psi \in \mathrm{O}(\mathbb{R}^n)$, denotemos por $\psi(C)$ o conjunto $\{\psi x : x \in C\}$.

Para qualquer subconjunto finito $C \in S^{n-1}$, temos:

$$\begin{aligned} \sum_{x,y,z\in C} S_k^n(z\cdot x, z\cdot y, x\cdot y) &= \frac{1}{6}\sum_{x,y,z\in C}\sum_{\sigma\in\mathcal{S}_3} Y_k^n\big(\sigma(z\cdot x, z\cdot y, x\cdot y)\big) \\ &= \sum_{x,y,z\in C} Y_k^n(z\cdot x, z\cdot y, x\cdot y) \\ &= \sum_{z\in C}\sum_{x,y\in C} Y_k^n\big(e\cdot(\psi_z^{-1}x), e\cdot(\psi_z^{-1}y), (\psi_z^{-1}x)\cdot(\psi_z^{-1}y)\big) \\ &= \sum_{z\in C}\sum_{x,y\in\psi_z^{-1}(C)} Y_k^n(e\cdot x, e\cdot y, x\cdot y) \succeq 0. \end{aligned}$$

Sendo que na segunda igualdade trocamos a ordem entre o somatório em $\mathcal{S}_3$ e o somatório triplo em C e, para cada $\sigma \in \mathcal{S}_3$, trocamos os nomes de x, y e z de modo a obter o mesmo resultado. Cada parcela do último termo é positiva semidefinida pela Proposição 3.4 aplicada em $\psi_z^{-1}(C)$, com $\mu(x) = 1$ para todo $x \in \psi_z^{-1}(C)$. □

3.2. O limitante de programação semidefinida

As relações satisfeitas pelas funções determinadas na seção anterior podem ser usadas para a construção de um novo limitante para o tamanho dos códigos

esféricos e consequentemente do número de contato. Contruímos inicialmente um problema de otimização cuja solução ótima produz um limitante e em seguida, por um argumento de dualização, obtemos um problema em que qualquer solução viável produz um limitante.

O primeiro passo é, para cada código esférico $C \subset S^{n-1}$ de distância angular mínima θ, definir uma função que mede a distribuição dos produtos internos entre triplas de pontos do código, isto é:

$$x_C(u,v,t) := \frac{1}{|C|}\big|\{(c,c',c'') \in C^3 : c\cdot c' = u,\ c\cdot c'' = v,\ c'\cdot c'' = t\}\big|.$$

Como estamos considerando códigos esféricos com distância angular mínima θ, definimos os domínios $I := [-1, \cos\theta]$ e

$$\Delta := \{(u,v,t) \in I^3 : \text{existem } c,c',c'' \in S^{n-1},\ c\cdot c' = u,\ c\cdot c'' = v,\ c'\cdot c'' = t\}, \quad (39)$$

que corresponde às triplas de produtos internos entre pontos distintos de um código com distância angular mínima θ.

Construímos o limitante a partir de propriedades satisfeitas por essas funções. Temos que $x_C(u,v,t) \geq 0$ para todo $u,v,t \in [-1,1]$ e $x_C(u,v,t) = 0$ para quase todos os pontos, exceto um número finito de triplas de $(I \cup \{1\})^3$ (e por isso os somatórios a seguir estão bem definidos), também:

$$\begin{aligned}
&x_C(1,1,1) = 1,\\
&x_C(\sigma(u,v,t)) = x_C(u,v,t) \text{ para todo } \sigma \in \mathcal{S}_3,\\
&|C|^2 = \sum_{u,v,t\in[-1,1]} x_C(u,v,t) = 1 + 3\sum_{u\in I} x_C(u,u,1) + \sum_{(u,v,t)\in\Delta} x_C(u,v,t),\\
&|C| = \sum_{u\in[-1,1]} x_C(u,u,1) = 1 + \sum_{u\in I} x_C(u,u,1).
\end{aligned}$$

Comparando as duas últimas identidades, obtemos

$$1 + 3\sum_{u\in I} x_C(u,u,1) + \sum_{(u,v,t)\in\Delta} x_C(u,v,t) = \Big(1 + \sum_{u\in I} x_C(u,u,1)\Big)^2,$$

o que implica

$$\sum_{u\in I} x_C(u,u,1) + \sum_{(u,v,t)\in\Delta} x_C(u,v,t) - \Big(\sum_{u\in I} x_C(u,u,1)\Big)^2 \geq 0,$$

equivalente a

$$\begin{pmatrix} 1 & \sum_{u\in I} x_C(u,u,1) \\ \sum_{u\in I} x_C(u,u,1) & \sum_{u\in I} x_C(u,u,1) + \sum_{(u,v,t)\in\Delta} x_C(u,v,t) \end{pmatrix} \succeq 0. \quad (40)$$

A Proposição 2.5 (com $w(c) = 1$ para todo $c \in C$), para $k = 1,\ldots,d$, implica na seguinte relação satisfeita pelas funções x_C:

$$\sum_{u\in[-1,1]} x_C(u,u,1)P_k^n(u) = 1 + \sum_{u\in I} x_C(u,u,1)P_k^n(u) \geq 0. \quad (41)$$

E a Proposição 3.5, para $k = 0,\ldots,d$, implica em:

$$\begin{aligned}
\sum_{u,v,t\in[-1,1]} x_C(u,v,t)S_k^n(u,v,t) = S_k^n(1,1,1) + 3\sum_{u\in I} x_C(u,u,1)S_k^n(u,u,1)\\
+ \sum_{(u,v,t)\in\Delta} x_C(u,v,t)S_k^n(u,v,t) \succeq 0. \quad (42)
\end{aligned}$$

Considerando as relações (40), (41) e (42) satisfeitas pelas funções x_C, podemos escrever o seguinte problema de otimização:

$$
\begin{aligned}
\max \quad & 1+\sum_{u\in I} x(u,u,1) \\
& x\colon (I\cup\{1\})^3 \to \mathbb{R}_{\geq 0}, \\
& x(u,v,t)=0 \text{ exceto para um número finito de triplas em } (I\cup\{1\})^3, \\
& \begin{pmatrix}1&0\\0&0\end{pmatrix} + \sum_{u\in I} x(u,u,1)\begin{pmatrix}0&1\\1&1\end{pmatrix} + \sum_{(u,v,t)\in\Delta} x(u,v,t)\begin{pmatrix}0&0\\0&1\end{pmatrix} \succeq 0, \\
& 1+\sum_{u\in I} x(u,u,1)P_k^n(u) \geq 0 \quad \text{para } k=1,\dots,d, \\
& S_k^n(1,1,1) + 3\sum_{u\in I} x(u,u,1)S_k^n(u,u,1) \\
& \qquad\qquad + \sum_{(u,v,t)\in\Delta} x(u,v,t)S_k^n(u,v,t) \succeq 0 \quad \text{para } k=0,\dots,d.
\end{aligned}
\tag{43}
$$

Como as funções x_C construídas a partir dos códigos esféricos são soluções viáveis para esse problema, o valor ótimo de (43) produz um limitante para $A(n,\theta)$. Entretanto, nessa forma não é claro como podemos limitar superiomente o valor de uma solução ótima. Isso pode ser feito construindo-se um problema dual com técnicas semelhantes às explicadas na Seção 1.3. O resultado é o seguinte programa em que qualquer solução viável produz um limitante para $A(n,\theta)$:

$$
\begin{aligned}
\min \quad & 1+\sum_{k=1}^{d} a_k + b_{11} + \langle F_0, J_{d+1}\rangle \\
& a_k\in\mathbb{R},\ a_k\geq 0 \qquad \text{para } k=1,\dots,d, \\
& B=\begin{pmatrix}b_{11}&b_{12}\\b_{12}&b_{22}\end{pmatrix}\succeq 0, \\
& F_k\in\mathbb{R}^{(d-k+1)\times(d-k+1)}, \quad F_k\succeq 0 \qquad \text{para } k=0,\dots,d, \\
& \text{(a)}\ \ \sum_{k=1}^{d} a_k P_k^n(u) + 2b_{12} + b_{22} \\
& \qquad\qquad +3\sum_{k=0}^{d}\langle F_k, S_k^n(u,u,1)\rangle \leq -1 \quad \text{para } u\in I, \\
& \text{(b)}\ \ b_{22} + \sum_{k=0}^{d}\langle F_k, S_k^n(u,v,t)\rangle \leq 0 \qquad \text{para } (u,v,t)\in\Delta.
\end{aligned}
\tag{44}
$$

Esse é o limitante de programação semidefinida, conforme proposto por Bachoc e Vallentin [BV08]. O limitante é assim chamado pois pode ser visto como um programa de otimização semidefinida com infinitas restrições, indexadas pelos conjuntos I e Δ. Veremos no próximo capítulo como encontrar soluções viáveis através de um programa de otimização semidefinida finito.

Em vez de ver os detalhes sobre o processo de dualização, apresentamos a seguir uma prova direta de que uma solução viável do Programa (44) fornece um limitante superior para $A(n,\theta)$ (repare como a demonstração se baseia apenas na Proposição 2.5 e na Proposição 3.5):

TEOREMA 3.6. *Se $a_1, \ldots, a_d, B, F_0, \ldots, F_d$ satisfazem as restrições do Programa (44), então $A(n,\theta) \le \lfloor 1 + \sum_{k=1}^{d} a_k + b_{11} + \langle F_0, J_{d+1}\rangle \rfloor$.*

DEMONSTRAÇÃO. Seja $C \subset S^{n-1}$ um conjunto não-vazio de pontos com distância angular mínima θ. Considere a quantidade

$$S(C) := \sum_{x,y\in C} \Big(1 + \sum_{k=1}^{d} a_k P_k^n(x\cdot y) + 2b_{12} - 2b_{22}\Big) + \sum_{x,y,z\in C} \Big(b_{22} + \sum_{k=0}^{d} \langle F_k, S_k^n(z\cdot x, z\cdot y, x\cdot y)\rangle\Big).$$

Como $a_k \ge 0$ para $k = 1, \ldots, d$, pela Proposição 2.5, temos

$$\sum_{x,y\in C} \Big(1 + \sum_{k=1}^{d} a_k P_k^n(x\cdot y) + 2b_{12} - 2b_{22}\Big) \ge |C|^2(1 + 2b_{12} - 2b_{22}).$$

Como $F_k \succeq 0$ para $k = 0, \ldots, d$ e o produto $\langle A, B\rangle$ de duas matrizes positivo-semidefinidas é sempre não-negativo, pela Proposição 3.5 temos

$$\sum_{x,y,z\in C} \Big(b_{22} + \sum_{k=0}^{d} \langle F_k, S_k^n(z\cdot x, z\cdot y, x\cdot y)\rangle\Big) \ge |C|^3 b_{22}$$

e portanto

$$S(C) \ge |C|^2(1 + 2b_{12} - 2b_{22}) + |C|^3 b_{22}.$$

Por outro lado,

$$\begin{aligned}
&\sum_{x,y\in C} \Big(1 + \sum_{k=1}^{d} a_k P_k^n(x\cdot y) + 2b_{12} - 2b_{22}\Big) \\
&= |C|\Big(1 + \sum_{k=1}^{d} a_k + 2b_{12} - 2b_{22}\Big) + \sum_{\substack{x,y\in C:\\ x\ne y}} \Big(1 + \sum_{k=1}^{d} a_k P_k^n(x\cdot y) + 2b_{12} - 2b_{22}\Big)
\end{aligned}$$

e (usando a condição (b) do programa)

$$\begin{aligned}
&\sum_{x,y,z\in C} \Big(b_{22} + \sum_{k=0}^{d} \langle F_k, S_k^n(z\cdot x, z\cdot y, x\cdot y)\rangle\Big) \\
&\qquad = \sum_{s=1}^{3} \sum_{\substack{x,y,z\in C:\\ |\{x,y,z\}|=s}} \Big(b_{22} + \sum_{k=0}^{d} \langle F_k, S_k^n(z\cdot x, z\cdot y, x\cdot y)\rangle\Big) \\
&\qquad \le |C|\Big(b_{22} + \langle J_{d+1}, F_0\rangle\Big) + 3 \sum_{\substack{x,y\in C:\\ x\ne y}} \Big(b_{22} + \sum_{k=0}^{d} \langle F_k, S_k^n(x\cdot y, x\cdot y, 1)\rangle\Big),
\end{aligned}$$

logo (usamos agora a condição (a) do programa),

$$S(C) \le |C|\Big(1 + \sum_{k=1}^{d} a_k + 2b_{12} - b_{22} + \langle F_0, J_{d+1}\rangle\Big).$$

Juntando as duas desigualdades para $S(C)$ e cancelando $|C|$, obtemos

$$|C|(1 + 2b_{12} - 2b_{22}) + |C|^2 b_{22} \le 1 + \sum_{k=1}^{d} a_k + 2b_{12} - b_{22} + \langle F_0, J_{d+1}\rangle,$$

o que implica

$$|C| + \Big((b_{11} - 2b_{12} + b_{22}) + |C|(2b_{12} - 2b_{22}) + |C|^2 b_{22}\Big) \leq 1 + \sum_{k=1}^{d} a_k + b_{11} + \langle F_0, J_{d+1}\rangle$$

e finalmente

$$|C| \leq 1 + \sum_{k=1}^{d} a_k + b_{11} + \langle F_0, J_{d+1}\rangle\,,$$

já que $B \succeq 0$ implica que a expressão entre parênteses é não-negativa, qualquer que seja o valor de $|C|$.

Como isso vale para qualquer conjunto C de pontos com distância angular mínima θ, e considerando que $A(n,\theta)$ é um número inteiro, obtemos o resultado desejado. □

3.3. A decomposição de H_m^n sob a ação de H

Nesta seção demonstramos o Teorema 3.1 e vemos os detalhes de como H_m^n se decompõe como soma de subespaços irredutíveis sob a ação de H, o subgrupo de $\mathrm{O}(\mathbb{R}^n)$ que estabiliza um certo ponto $e \in S^{n-1}$. Tal como na Seção 2.4, usamos $r^2 = x_1^2 + \cdots + x_n^2$ e por simplicidade supomos que $e = (1, 0, \ldots, 0)$.

Para $k = 0, \ldots, m$, denotemos por $x_1^{m-k} H_k^{n-1}$ o espaço dos polinômios da forma $f(x) = x_1^{m-k} h_k(x')$, com $h_k \in H_k^{n-1}$ e $x' = (x_2, \ldots, x_n) \in \mathbb{R}^{n-1}$. Os polinômios desse espaço são homogêneos de grau m e $x_1^{m-k} H_k^{n-1}$ é um subespaço invariante de P_m^n (o espaço dos polinômios homogêneos de n variáveis e grau m) sob a ação de H, cuja representação é equivalente à representação de $\mathrm{O}(\mathbb{R}^{n-1})$ em H_k^{n-1} e logo é irredutível.

Na próxima proposição, vemos uma outra decomposição do espaço P_m^n semelhante à que vimos na Proposição 2.7.

Proposição 3.7. *Seja $r^2 P_{m-2}^n$ o espaço de polinômios da forma $r^2 f(x)$, com $f \in P_{m-2}^n$ e para $k = 0, \ldots, m$, seja $x_1^{m-k} H_k^{n-1}$ o espaço dos polinômios da forma $x_1^{m-k} h_k(x')$, com $h_k \in H_k^{n-1}$ e $x' = (x_2, \ldots, x_n) \in \mathbb{R}^{n-1}$. Então,*

$$P_m^n = \bigoplus_{k=0}^{m} x_1^{m-k} H_k^{n-1} \oplus r^2 P_{m-2}^n.$$

Demonstração. Primeiramente, mostremos que esses subespaços geram todo P_m^n. Seja $f \in P_m^n$, dividindo f por r^2, substituindo x_1^2 por $r^2 - (x_2^2 + \cdots + x_n^2)$ no resto e colocando x_1 em evidência, podemos escrever f na forma

$$f(x) = r^2 F(x) + x_1 \varphi_1(x') + \varphi_2(x'),$$

com $F \in P_{m-2}^n$, $\varphi_1 \in P_{m-1}^{n-1}$ e $\varphi_2 \in P_m^{n-1}$. Aplicando a decomposição vista no Teorema 2.8 em φ_1 e φ_2, obtemos

$$f(x) = r^2 F(x) + \sum_{k=0}^{\lfloor (m-1)/2 \rfloor} x_1 (r^2 - x_1^2)^k \hat{h}_{m-2k-1}(x') + \sum_{k=0}^{\lfloor m/2 \rfloor} (r^2 - x_1^2)^k \hat{h}_{m-2k}(x'),$$

com $\hat{h}_s \in H_s^{n-1}$. Expandindo os binômios e agrupando os termos com r^2, obtemos

$$f(x) = r^2 f_1(x) + \sum_{k=0}^{m} x_1^{m-k} h_k(x')$$

com $f_1 \in P_{m-2}^n$ e $h_k \in H_k^{n-1}$, conforme desejado.

Para mostrar que a soma é direta basta calcular a dimensão dos espaços. Usando que $\dim(P_m^n) = \binom{n+m-1}{m}$ e $\dim(H_m^n) = \binom{n+m-1}{m} - \binom{n+m-3}{m-2}$ (conforme a equação (28)), temos que a soma das dimensões dos espaços é

$$\begin{aligned}
&\sum_{k=0}^{m} \dim(H_k^{n-1}) + \dim(P_{m-2}^n) \\
&\qquad = \sum_{k=0}^{m} \left(\binom{n+k-2}{k} - \binom{n+k-4}{k-2} \right) + \binom{n+m-3}{m-2} \\
&\qquad = \sum_{k=0}^{m} \left(\binom{n+k-2}{n-2} - \binom{n+k-4}{n-2} \right) + \binom{n+m-3}{n-1} \\
&\qquad = \binom{n+m-2}{n-2} - \binom{n+m-3}{n-2} + \binom{n+m-3}{n-1} \\
&\qquad = \binom{n+m-1}{n-1} = \dim(P_m^n). \qquad \square
\end{aligned}$$

A seguir demonstramos o Teorema 3.1, apresentado no início deste capítulo.

TEOREMA 3.1. *O espaço H_m^n sob a ação de H (o subgrupo de $\mathrm{O}(\mathbb{R}^n)$ que estabiliza $e \in \mathbb{R}^n$, fixado) decompõe-se como soma ortogonal de subespaços invariantes, irredutíveis e não equivalentes entre si:*

$$H_m^n = H_{0,m}^{n-1} \perp H_{1,m}^{n-1} \perp \cdots \perp H_{m,m}^{n-1},$$

com a representação de H em $H_{i,m}^{n-1}$ equivalente à representação de $\mathrm{O}(\mathbb{R}^{n-1})$ em H_i^{n-1}.

DEMONSTRAÇÃO. Como os subespaços presentes na decomposição da Proposição 3.7 são subespaços invariantes com relação à representação do grupo H, temos pela Proposição 1.3 que a representação de H em $\bigoplus_{k=0}^{m} x_1^{m-k} H_k^{n-1}$ é equivalente à representação de H no espaço quociente $P_m^n / r^2 P_{m-2}^n$ que, pela decomposição de P_m^n vista na Proposição 2.7, é equivalente à representação de H em H_m^n. Disto resulta a decomposição indicada no enunciado. A ortogonalidade entre os subespaços segue da Proposição 1.18, sendo que o fato da representação de H em H_m^n ser unitária segue do fato da representação de $\mathrm{O}(\mathbb{R}^n)$ em $L^2(S^{n-1})$ o ser. $\square$

CAPÍTULO 4

Cálculo do limitante de programação semidefinida

Neste capítulo vemos como encontrar soluções viáveis para o limitante de programação semidefinida dado no Programa (44) e então limitantes para o problema do número de contato através do Teorema 3.6.

As restrições (a) e (b) do programa formam um conjunto infinito de restrições e para considerá-las é necessário aplicar alguma técnica, como a de amostragem vista na Seção 2.3. Neste capítulo vemos outra, baseada na representação de polinômios por somas de quadrados via programação semidefinida. No fim também vemos uma forma de verificar rigorosamente os resultados obtidos.

4.1. Otimização polinomial e soma de quadrados

Um polinômio de n variáveis $p \in \mathbb{R}[x]$ (denotemos $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$) é dito *soma-de-quadrados* se puder ser escrito como

$$p = q_1^2 + \cdots + q_m^2,$$

com $q_i \in \mathbb{R}[x]$. Expressar um polinômio como uma soma de quadrados é uma forma de certificar que um polinômio é não-negativo para todo $x \in \mathbb{R}^n$, entretanto nem todo polinômio não-negativo pode ser expresso dessa forma. Um exemplo é o polinômio de Motzkin:

$$p(x_1, x_2) = x_1^2 x_2^2 (x_1^2 + x_2^2 - 3) + 1.$$

Esse polinômio é não-negativo, o que é claro se $x_1^2+x_2^2-3 \geq 0$ e, caso $x_1^2+x_2^2-3 < 0$, fazendo $y^2 = 3-x_1^2-x_2^2$, temos $p(x_1, x_2) = -x_1^2x_2^2y^2+1 \geq -\big((x_1^2+x_2^2+y^2)/3\big)^3+1 = 0$ (pela desigualdade entre as médias aritmética e geométrica). Temos também que esse polinômio não pode ser expresso como uma soma de quadrados. Para ver isso considere que $p = \sum_{i=1}^m r_i^2$, com polinômios r_i de grau no máximo 3; como os coeficientes de x_1^6 e x_2^6 em p são iguais a zero, vemos que os coeficientes de x_1^3 e x_2^3 nos polinômios r_i devem ser zero. Analogamente, como os coeficientes de x_1^4, x_2^4, x_1^2 e x_2^2 em p são nulos, os coeficientes de x_1^2, x_2^2, x_1 e x_2 nos polinômios r_i também devem ser iguais a zero; podemos então escrever $r_i = a_i x_1 x_2^2 + b_i x_1^2 x_2 + c_i x_1 x_2 + d_i$ e analisando o coeficiente de $x_1^2x_2^2$ em p, temos $-3 = \sum_{i=1}^m c_i^2$, o que gera uma contradição. Hilbert [Hil88] mostrou que essas duas propriedades (ser globalmente não-negativo e ser soma-de-quadrados) são equivalentes apenas para polinômios de uma variável, polinômios de grau dois e polinômios de duas variáveis e grau quatro.

Apesar de soma-de-quadrados ser uma propriedade mais restritiva do que ser não-negativo, vemos a seguir que podemos determinar se um polinômio pode ser escrito como soma de quadrados usando programação semidefinida, o que na prática pode ser resolvido por computadores.

Proposição 4.1. *Seja $p \in \mathbb{R}[x]$ um polinômio de grau $2d$ e B uma base de $\mathbb{R}[x]_{\leq d}$, o espaço dos polinômios de grau no máximo d. Seja $v_B \colon B \to \mathbb{R}[x]$ tal que $v_B(q) = q$ para todo $q \in B$. Então p é soma-de-quadrados se e somente se existe uma matriz positivo-semidefinida $Q \colon B \times B \to \mathbb{R}$ tal que $p = v_B^T Q v_B = \langle Q, v_B v_B^T \rangle$.*

Demonstração. Digamos que exista uma matriz Q positivo-semidefinida tal que $p = v_B^T Q v_B$. Ela pode ser fatorada como $Q = RR^T$ e assim $p = v_B^T RR^T v_B = \sum_{i=1}^m \left(r_i^T v_B\right)^2$, com os vetores r_i denotando as colunas de R de modo que $r_i^T v_B$ seja um polinômio.

Reciprocamente, se p puder ser escrito como $p = \sum_{i=1}^m q_i^2$, com $q_i \in \mathbb{R}[x]_{\leq d}$, como B é uma base desse espaço, existe $r_i \in \mathbb{R}^B$ tal que $r_i^T v_B = q_i$ e podemos definir uma matriz R com as colunas formadas pelos vetores r_i, de modo que $p = v_B^T RR^T v_B$. □

A proposição anterior permite reduzir a questão da representatividade de um polinômio como soma de quadrados pela da viabilidade de um programa semidefinido. Note que a identidade $p = \langle Q, v_B v_B^T \rangle$ é uma igualdade entre polinômios e deve ser transformada em um conjunto de restrições lineares nas entradas de Q. Isso pode ser feito considerando uma base $B_=$ de $\mathbb{R}[x]_{\leq 2d}$ e comparando os coeficientes de ambos os lados nessa base; usualmente escolhemos B e $B_=$ como as bases monomiais de seus respectivos espaços, entretanto na próxima seção veremos um caso em que será conveniente escolher bases diferentes.

Vimos até aqui uma forma de certificar que um polinômio é não-negativo em todo $\mathbb{R}^n$, consideremos agora polinômios não-negativos em *conjuntos semi-algébricos fechados*, isto é, conjuntos da forma

$$\{x \in \mathbb{R}^n : g_1(x) \geq 0,\, g_2(x) \geq 0,\, \ldots,\, g_s(x) \geq 0\}, \tag{45}$$

com $g_1, \ldots, g_s \in \mathbb{R}[x]$. Uma condição suficiente para um polinômio $p \in \mathbb{R}[x]$ ser não-negativo em tal conjunto é a existência de polinômios $q_0, \ldots, q_s$ somas-de-quadrados tais que

$$p = q_0 + g_1 q_1 + \cdots + g_s q_s. \tag{46}$$

Tal como no caso global, nem todo polinômio nâo-negativo em um conjunto semi-algébrico pode ser expresso dessa forma, entretanto existem resultados que, com algumas hipóteses adicionais, garantem a existência de tal representação. Um desses teoremas é devido a Putinar [Put93]:

Teorema 4.2 (Positvstellensatz de Putinar). *Seja $K = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots,\, g_s(x) \geq 0\}$ um conjunto semi-algébrico compacto. Suponha que exista um polinômio P na forma $P = r_0 + g_1 r_1 + \cdots + g_s r_s$, com $r_0, \ldots, r_s$ somas-de-quadrados, tal que o conjunto $\{x \in \mathbb{R}^n : P(x) \geq 0\}$ seja compacto. Então, todo polinômio p positivo em K pode ser escrito como $p = q_0 + g_1 q_1 + \cdots + g_s q_s$, com $q_0, \ldots, q_s$ somas-de-quadrados.*

O teorema não diz nada sobre o grau dos polinômios $q_0, \ldots, q_s$ que aparecem na expressão (ao contrário do caso global, isso não é claro, pois podem haver cancelamentos entre os termos dos diversos polinômios presentes na fórmula). Na prática, na busca por uma representação com programação semidefinida é necessário limitar o grau desses polinômios, o que acarreta em alguma perda de generalidade.

Usando essa técnica podemos reescrever as condições (a) e (b) do Programa (44). Seja B uma base de $\mathbb{R}[u]_{\leq d}$ e $v_{B,u} \colon B \to \mathbb{R}$ o vetor obtido de v_B quando avaliamos cada entrada em u, de modo que $v_{B,u}(p) = p(u)$. Para a condição (a), usamos $g(u) := (u+1)(\cos\theta - u)$, de modo que $I = \{u \in \mathbb{R} : g(u) \geq 0\}$ e substituímos a restrição por

$$\sum_{k=1}^{d} a_k P_k^n(u) + 2b_{12} + b_{22} + 3\sum_{k=0}^{d} \langle F_k, S_k^n(u,u,1)\rangle + \langle Q_0, V_{0,u}\rangle + \langle Q_1, g(u) V_{1,u}\rangle = -1, \tag{47}$$

com Q_0, $Q_1 \succeq 0$ sendo duas variáveis novas para o programa, $V_{0,u} = v_{B_0,u} v_{B_0,u}^T$ com $B_0 = \{1, u, u^2, \ldots, u^d\}$ e $V_{1,u} = v_{B_1,u} v_{B_1,u}^T$ com $B_1 = \{1, u, u^2, \ldots, u^{d-1}\}$. Os graus são escolhidos de modo que o maior grau em todos os polinômios que aparecem no lado esquerdo de (47) seja $2d$.

Para a condição (b), também podemos reescrever Δ como um conjunto semi-algébrico. Para isso, observamos que a matriz

$$\begin{pmatrix} 1 & u & v \\ u & 1 & t \\ v & t & 1 \end{pmatrix}$$

é a matriz de Gram de três pontos distintos $c, c', c'' \in S^{n-1}$ com distância angular mínima θ se e somente se for positivo-semidefinida e $u, v, t \in I$. Isso é equivalente a seu determinante ser não-negativo e $g(u), g(v), g(t) \geq 0$, isto é:

$$\Delta = \{\, (u, v, t) \in \mathbb{R}^3 : g_i(u, v, t) \geq 0 \text{ para } i = 1, \ldots, 4 \,\},$$

com

$$\begin{aligned} g_1(u, v, t) &:= g(u), \quad g_2(u, v, t) := g(v), \\ g_3(u, v, t) &:= g(t), \quad g_4(u, v, t) := 1 + 2uvt - u^2 - v^2 - t^2. \end{aligned} \tag{48}$$

Assim a restrição (b) pode ser substituída por

$$b_{22} + \sum_{k=0}^{d} \langle F_k, S_k^n \rangle + \langle R_0, V_d \rangle + \langle R_1, g_1 V_{d-1} \rangle + \langle R_2, g_2 V_{d-1} \rangle + \langle R_3, g_3 V_{d-1} \rangle + \langle R_4, g_4 V_{d-2} \rangle = 0, \tag{49}$$

com $R_0, \ldots, R_4 \succeq 0$ sendo variáveis novas para o programa e $V_{d'} = v_{B_{d'}} v_{B_{d'}}^T$, com $B_{d'}$ sendo a base monomial de $\mathbb{R}[u, v, t]_{\leq d'}$ para cada d' que aparece na fórmula. Novamente, os graus são escolhidos de modo que o maior grau em todos os polinômios que aparecem no lado esquerdo de (49) seja $2d$.

Note que os tamanhos das matrizes presentes na restrição (49) crescem rapidamente com d. O número de monômios com 3 variáveis e grau até d é $\binom{d+3}{3}$, o que com $d = 14$ por exemplo, leva à consideração de matrizes de tamanho 680×680. Mittelman e Vallentin [MV10] resolveram o Programa (44) com $\theta = \pi/3$, as restrições (a) e (b) substituídas por (47) e (49) e $d \leq 14$, obtendo limitantes para o número de contato nas dimensões $n = 3, \ldots, 24$. Em todos os casos, esses limitantes foram os melhores conhecidos até então, sendo os resultados novos na maior parte dos casos ainda em aberto. Os valores serão apresentados na Seção 4.3, junto com os novos limitantes obtidos.

Devido ao tamanho do problema de programação semidefinida a ser resolvido e dificuldades numéricas, Mittelman e Vallentin [MV10] usaram *solvers* com precisão numérica elevada e mais lentos, como o SDPA-GMP [Nak10], e por isso só foi possível considerar problemas com $d \leq 14$ usando essa formulação. A situação é diferente da que ocorre com o limitante de programação linear, onde é possível realizar cálculos com polinômios de grau crescente, até o ponto em que não se observam mais melhorias significativas no limitante fornercido [OS79]; nas tabelas presentes no artigo de Mittelman e Vallentin [MV10] observa-se que o valor do limitante obtido ainda pode ser melhorado com o incremento de d.

A partir da forma como as matrizes S_k^n foram definidas em (38), temos que os polinômios presentes em suas entradas são simétricos com relação às permutações entre as variáveis u, v e t; o domínio Δ da condição (b) do Programa (44) também possui essa simetria. Vemos na próxima seção uma forma de usar essa propriedade para reduzir o tamanho das matrizes usadas na formulação com somas de quadrados.

4.2. Somas de quadrados com polinômios invariantes

Seja G um grupo finito que age linearmente em $\mathbb{R}^n$, o que induz uma ação no espaço de polinômios em n variáveis $\mathbb{R}[x]$ definida de modo que para todo $g \in G$, $p \in \mathbb{R}[x]$ e $x \in \mathbb{R}^n$,

$$gp(x) := p(g^{-1}x).$$

Dizemos que um polinômio p é *invariante* quando $gp = p$ para todo $g \in G$.

Como um exemplo inicial, consideremos o polinômio de duas variáveis

$$p(x_1, x_2) = x_1^2 + x_2^2.$$

Esse polinômio é soma-de-quadrados e é invariante com relação à permutação de suas variáveis; ele também pode ser escrito como

$$p(x_1, x_2) = \frac{1}{2}(x_1 + x_2)^2 + \frac{1}{2}(x_1 - x_2)^2,$$

a soma do quadrado de um polinômio simétrico com o quadrado de um polinômio anti-simétrico (isto é, tal que $p(x_1, x_2) = -p(x_2, x_1)$). Vemos a seguir uma técnica desenvolvida por Gatermann e Parrilo [GP04] (veja também o texto de Bachoc, Gijswijt, Schrijver e Vallentin [BGSV12]) que mostra como podemos reduzir o tamanho dos programas semidefinidos relacionados à decomposição em soma de quadrados quando o polinômio é invariante.

Supomos por simplicidade que G seja um grupo cujas representações irredutíveis sejam todas realizáveis em espaços vetoriais reais, o que será suficiente para a aplicação no limitante para o número de contato. (Caso contrário, passa a ser necessário o uso de programação semidefinida com matrizes complexas e polinômios com coeficientes complexos na decomposição em soma de quadrados.) A técnica pode ser resumida pelo seguinte teorema, que será demonstrado na Seção 4.2.2, onde também é descrito como calcular os coeficientes das matrizes V_d^i explicitamente.

Teorema 4.3. *Seja G um grupo finito que possui todas as suas representações irredutíveis realizáveis em espaços vetoriais reais e $d_1, \ldots, d_r$ as dimensões dessas representações. Então existem matrizes V_d^i de tamanho $m_i \times m_i$ para $i = 1, \ldots, r$ com polinômios invariantes e de grau até $2d$ em suas entradas tais que um polinômio $p \in \mathbb{R}[x]$ de grau $2d$ invariante sob a ação de G é soma-de-quadrados se e somente se existem matrizes positivo-semidefinidas $Q^1, \ldots, Q^r$ tais que*

$$p = \left\langle Q^1, V_d^1 \right\rangle + \cdots + \left\langle Q^r, V_d^r \right\rangle.$$

Os números $m_1, \ldots, m_r$ são as multiplicidades das componentes irredutíveis da representação de G em $\mathbb{R}[x]_{\leq d}$, de modo que $m_1 d_1 + \cdots + m_r d_r = \binom{n+d}{d}$.

4.2.1. Aplicação ao limitante. Voltemos a considerar a restrição (b) do Programa (44). Anteriormente a substituímos pela condição

$$b_{22} + \sum_{k=0}^{d} \langle F_k, S_k^n \rangle + r_0 + g_1 r_1 + \cdots + g_4 r_4 = 0,$$

com os polinômios $g_1, \ldots, g_4$ definidos em (48) e os polinômios $r_0, \ldots, r_4$ somas-de-quadrados, que em (49) já foram substituídos pela expressão dada pela Proposição 4.1. Apesar de os polinômios presentes nas matrizes S_k^n serem invariantes sob a ação do grupo $\mathcal{S}_3$ que age permutando suas variáveis e de o domínio Δ possuir propriedade semelhante, não podemos supor sem prejuízo ao limitante que os polinômios $r_0, \ldots, r_4$ sejam invariantes. Isso ocorre porque os polinômios $g_1, \ldots, g_4$ não o são, caso estivéssemos usando polinômios $s_1, \ldots, s_4$ invariantes e tivéssemos

$$b_{22} + \sum_{k=0}^{d} \langle F_k, S_k^n \rangle + r_0 + s_1 r_1 + \cdots + s_4 r_4 = 0, \tag{50}$$

com polinômios $r_1, \dots r_4$ somas-de-quadrados não necessariamente invariantes, poderíamos simetrizar toda a equação:

$$\frac{1}{|\mathcal{S}_3|}\sum_{\sigma\in\mathcal{S}_3}\sigma\Big(b_{22}+\sum_{k=0}^{d}\langle F_k, S_k^n\rangle + r_0 + s_1r_1 + \cdots + s_4r_4\Big) = 0$$

e, como os outros polinômios presentes são invariantes, obter

$$b_{22}+\sum_{k=0}^{d}\langle F_k, S_k^n\rangle+\Big(\frac{1}{|\mathcal{S}_3|}\sum_{\sigma\in\mathcal{S}_3}\sigma r_0\Big)+s_1\Big(\frac{1}{|\mathcal{S}_3|}\sum_{\sigma\in\mathcal{S}_3}\sigma r_1\Big)+\cdots+s_4\Big(\frac{1}{|\mathcal{S}_3|}\sum_{\sigma\in\mathcal{S}_3}\sigma r_4\Big) = 0,$$

uma expressão com polinômios somas-de-quadrados e invariantes.

Vemos no próximo lema que de fato podemos representar o domínio Δ com polinômios invariantes:

LEMA 4.4. *Considere os polinômios*

$$\begin{aligned} &s_1 := g_1 + g_2 + g_3, \quad s_2 := g_1g_2 + g_1g_3 + g_2g_3,\\ &s_3 := g_1g_2g_3, \ \text{e}\ s_4 := g_4, \end{aligned} \tag{51}$$

com os polinômios g_i definidos em (48)*. Então*

$$\Delta = \{(u,v,t)\in\mathbb{R}^3 \ : \ s_i(u,v,t)\geq 0, \ i = 1,\dots,4\}.$$

DEMONSTRAÇÃO. Como $s_1, \dots, s_4$ são combinações positivas de produtos de $g_1, \dots, g_4$, temos que $g_i(u,v,t)\geq 0$ para $i = 1, \dots, 4$ implica em $s_i(u,v,t)\geq 0$ para $i = 1, \dots, 4$.

Para a direção contrária, podemos supor que $g_1(u,v,t) < 0$ (g_2 e g_3 são análogos e se $g_4(u,v,t) < 0$, então $s_4(u,v,t) = g_4(u,v,t) < 0$). Suponha que $s_2(u,v,t)\geq 0$ e $s_3(u,v,t)\geq 0$. Então $(g_1g_2g_3)(u,v,t)\geq 0$ e assim $(g_2g_3)(u,v,t)\leq 0$. Também, $(g_1g_2 + g_1g_3 + g_2g_3)(u,v,t)\geq 0$ implica que

$$(g_1(g_2+g_3))(u,v,t)\geq -(g_2g_3)(u,v,t)\geq 0,$$

e então $(g_2+g_3)(u,v,t)\leq 0$. Portanto, $s_1(u,v,t) = (g_1+g_2+g_3)(u,v,t) < 0$. □

Assim, a equação (50) com os polinômios $s_1, \dots, s_4$ dados em (51) e os polinômios $r_0, \dots, r_4$ somas-de-quadrados e invariantes é uma condição suficiente para a restrição (b) do Programa (44) ser satisfeita. Como feito na seção anterior, escolhemos os graus dos polinômios r_i de modo que o maior grau em todos os polinômios que aparecem no lado esquerdo de (50) seja $2d$; essa escolha é útil pois evita que o problema fique sobredeterminado pelas restrições lineares, o que afeta a estabilidade numérica do programa e a verificação dos resultados que será descrita na Seção 4.4. Usando o Teorema 4.3 no lugar da Proposição 4.1 para representar os polinômios r_i, a condição (b) é substituída por:

$$\begin{aligned} b_{22} + \sum_{k=0}^{d}&\langle F_k, S_k^n\rangle + \langle R_0^1, V_d^1\rangle + \langle R_0^2, V_d^2\rangle + \langle R_0^3, V_d^3\rangle\\ &+ \langle R_1^1, s_1V_{d-1}^1\rangle + \langle R_1^2, s_1V_{d-1}^2\rangle + \langle R_1^3, s_1V_{d-1}^3\rangle\\ &+ \langle R_2^1, s_2V_{d-2}^1\rangle + \langle R_2^2, s_2V_{d-2}^2\rangle + \langle R_2^3, s_2V_{d-2}^3\rangle\\ &+ \langle R_3^1, s_3V_{d-3}^1\rangle + \langle R_3^2, s_3V_{d-3}^2\rangle + \langle R_3^3, s_3V_{d-3}^3\rangle\\ &+ \langle R_4^1, s_4V_{d-2}^1\rangle + \langle R_4^2, s_4V_{d-2}^2\rangle + \langle R_4^3, s_4V_{d-2}^3\rangle = 0, \end{aligned} \tag{52}$$

com as matrizes $R_i^j \succeq 0$ sendo variáveis novas para o programa. O teorema foi aplicado com $r = 3$ pois sabe-se que o grupo $\mathcal{S}_3$ de permutações de três elementos possui três representações irredutíveis distintas, conhecidas como *trival*, *alternante*, ambas de dimensão um, e *padrão*, de dimensão dois: todas realizáveis em espaços vetoriais reais.

4.2.2. Demonstração do Teorema 4.3. Vejamos $\mathbb{R}[x]$ como um subconjunto do espaço de polinômios com coeficientes complexos $\mathbb{C}[x]$ e estendamos a ação de G para $\mathbb{C}[x]$ de modo a termos um espaço vetorial complexo e assim possamos usar a teoria de representações vista na Seção 1.1. Em seguida nos restrinjamos ao espaço de dimensão finita $\mathbb{C}[x]_{\leq d}$, com polinômios de grau no máximo d, e fixemos um produto interno $\langle , \rangle$ de modo que pela Proposição 1.1 a representação de G seja unitária. Esse produto interno pode ser escolhido de modo que $\langle p, q\rangle \in \mathbb{R}$ para todos $p, q \in \mathbb{R}[x]$; para ver isso considere como produto interno inicial o produto interno $\langle , \rangle_0$ que faz a base monomial ser ortonormal (que possui essa propriedade) e considere a simetrização efetuada na demonstração da Proposição 1.1 (a simetrização preserva a propriedade, já que $gp \in \mathbb{R}[x]$, se $p \in \mathbb{R}[x]$ e $g \in G$).

Seja B uma base ortonormal de $\mathbb{C}[x]_{\leq d}$ escolhida de modo que $B \subset \mathbb{R}[x]_{\leq d}$; isso pode ser feito através do processo de Gram-Schmidt aplicado à base monomial, já que $\langle p, q\rangle \in \mathbb{R}$ para todos $p, q \in \mathbb{R}[x]$. Seja $v_B \colon B \to \mathbb{C}[x]$ um vetor definido como na Proposição 4.1, de modo que $v_B(q) = q$ para todo $q \in B$. Para $x \in \mathbb{R}^n$, denotemos por $v_{B,x}$ o vetor obtido de v_B quando avaliamos cada entrada em x.

Para cada $g \in G$, seja $P_g \colon B \times B \to \mathbb{R}$ a matriz na base B da representação em $\mathbb{C}[x]_{\leq d}$; as matrizes P_g são reais pois $B \subset \mathbb{R}[x]$ e a representação é induzida pela ação linear de G em $\mathbb{R}^n$. Pode-se verificar que para cada $g \in G$ e $x \in \mathbb{R}^n$,

$$v_{B,g^{-1}x} = P_g^T v_{B,x}.$$

De fato, para $q \in B$,

$$v_{B,g^{-1}x}(q) \,=\, gq(x) \,=\, \sum_{p\in B} P_g(p,q)p(x) \,=\, \sum_{p\in B} P_g(p,q)v_{B,x}(p) \,=\, (P_g^T v_{B,x})(q).$$

Se $p \in \mathbb{R}[x]_{\leq 2d}$ é soma-de-quadrados e invariante sob a ação de G, temos pela Proposição 4.1 que existe $Q \colon B \times B \to \mathbb{R}$ positivo-semidefinida tal que $p = v_B^T Q v_B$ e assim para cada $x \in \mathbb{R}^n$,

$$\begin{aligned} p(x) = \frac{1}{|G|}\sum_{g\in G} p(g^{-1}x) &= \frac{1}{|G|}\sum_{g\in G} v_{B,g^{-1}x}^T Q v_{B,g^{-1}x} \\ &= \frac{1}{|G|}\sum_{g\in G} \big(v_{B,x}^T P_g\big) Q \big(P_g^T v_{B,x}\big) = v_{B,x}^T \Big(\frac{1}{|G|}\sum_{g\in G} P_g Q P_g^T\Big) v_{B,x}. \end{aligned}$$

Logo podemos definir a matriz

$$\overline{Q} := \frac{1}{|G|}\sum_{g\in G} P_g Q P_g^T,$$

que também é positivo-semidefinida e satisfaz

$$P_g \overline{Q} = \overline{Q} P_g$$

para todo $g \in G$; temos que $\overline{Q}$ é a matriz de um G-endomorfismo e, conforme descrito no Teorema 1.8, o espaço $\mathbb{C}[x]_{\leq d}$ possui uma base ortonormal W na qual a transformação dada por $\overline{Q}$ é bloco-diagonal.

Seguindo a notação usada na demonstração do Teorema 1.8, consideremos $\mathbb{C}[x]_{\leq d}$ decomposto como soma direta ortogonal de subespaços irredutíveis:

$$\begin{aligned} \mathbb{C}[x]_{\leq d} &= I_1 \perp \cdots \perp I_r, \\ I_k &= V_{k,1} \perp \cdots \perp V_{k,m_k}, \end{aligned}$$

com a subrepresentação de $V_{k,i}$ equivalente à de $V_{l,j}$ se e somente se $k = l$ e fixemos bases ortonormais $\{e_{k,i,1}, \ldots, e_{k,i,d_k}\}$ em cada subespaço irredutível $V_{k,i}$ de modo que as matrizes das subrepresentações equivalentes sejam iguais e tenhamos a base $W = \{e_{k,i,s} : k = 1, \ldots, r, i = 1, \ldots, m_k, s = 1, \ldots, d_k\}$.

Os polinômios $e_{k,i,s}$ podem ser determinados pelo processo descrito no Teorema 1.21, que usa as matrizes P_g da representação e matrizes das representações irredutíveis de G, que estamos supondo serem reais. Assim $W \subset \mathbb{R}[x]$ e a matriz $U\colon B \times W \to \mathbb{R}$ de mudança de base entre B e W é real. Como B e W são ortonormais, $U^{-1} = U^T$ e a matriz da transformação definida por $\overline{Q}$ na base W é dada por $U^T\overline{Q}U$, que além de bloco-diagonal é também positivo-semidefinida.

Podemos escrever p como soma de quadrados usando $U^T\overline{Q}U$:

$$\begin{aligned} p = v_B^T\overline{Q}v_B = v_B^T(UU^T)\overline{Q}(UU^T)v_B &= (v_B^TU)(U^T\overline{Q}U)(U^Tv_B) \\ &= v_W^T(U^T\overline{Q}U)v_W = \left\langle U^T\overline{Q}U, v_Wv_W^T \right\rangle, \end{aligned}$$

com $v_W = U^Tv_B$. De acordo com o Teorema 1.8, a matriz $U^T\overline{Q}U$ é composta por r blocos distintos $Q^1, \ldots, Q^r$, com Q^k sendo de tamanho $m_k \times m_k$ e repetindo-se d_k vezes. Assim podemos reescrever p como a soma de r termos:

$$p = \left\langle Q^1, V_d^1 \right\rangle + \cdots + \left\langle Q^r, V_d^r \right\rangle,$$

com V_d^k sendo o bloco correspondente à Q^k em $v_Wv_W^T$ e cujas entradas são:

$$(V_d^k)_{i,j} = \sum_{s=1}^{d_k} e_{k,i,s}e_{k,j,s}.$$

Note que com essa expressão também fica claro que as entradas das matrizes V_d^k são polinômios de grau até $2d$, já que os polinômios $e_{k,i,s}$ pertencem à $\mathbb{C}[x]_{\leq d}$ e que são polinômios invariantes, pois as matrizes das representações de G nos subespaços $V_{k,i}$ com as bases $\{e_{k,i,s}, \ldots, e_{k,i,d_k}\}$ são unitárias e não dependem de i. □

4.3. Resultados

O Programa (44) foi resolvido com $\theta = \pi/3$ e com as restrições (a) e (b) substituídas por (47) e (52), respectivamente. Essas restrições são identidades polinomiais que devem ser reescritas como restrições lineares nas variáveis do problema, isso pode ser feito escolhendo bases para $\mathbb{R}[u]_{\leq 2d}$ e $\mathbb{R}[u,v,t]_{\leq 2d}$ e expandindo todos os polinômios nessas bases. Para a restrição (47), simplesmente escolhemos a base monomial de $\mathbb{R}[u]_{\leq 2d}$. Para a restrição (52), notemos que todos os polinômios que aparecem são invariantes, então podemos considerar menos restrições se nos restringirmos a uma base do subespaço de polinômios invariantes de $\mathbb{R}[u,v,t]_{\leq 2d}$. Uma forma de escolher tal base é considerar todas as triplas (a,b,c) de inteiros não-negativos tais que $a + 2b + 3c \leq 2d$ e para cada uma delas pegar o polinômio $(u+v+t)^a(u^2+v^2+t^2)^b(u^3+v^3+t^3)^c$. Pela Proposição 1.1.2 de Sturmfels [Stu08], esses polinômios geram o subespaço de polinômios invariantes de grau no máximo $2d$ e pelo Teorema 1.1.1 do mesmo livro junto com um argumento de dimensão, vê-se que esses polinômios formam uma base do subespaço.

Obtemos assim um programa semidefinido menor do que o considerado na Seção 4.1, o que na prática gera uma grande diferença. Por exemplo, experimentos com o *solver* SDPA-GMP [Nak10] em um processador de 2.4 GHz demoraram 9 dias para resolver o programa apresentado na Seção 4.1 no caso $d = 11$ e $n = 12$, que envolve matrizes de tamanho $\binom{3+11}{3} = 364$. Após a simplificação, a matriz é decomposta em três blocos de tamanhos 83, 41 e 120 e o programa resultante pôde ser resolvido em menos de 12 horas sob condições similares.

Dessa forma foi possível calcular o valor do limitante para o número de contato com d até 16 em um tempo inferior a 6 semanas e melhorar os resultados obtidos por Mittelmann e Vallentin [MV10] nas dimensões 9 a 23.

Os resultados são exibidos na Tabela 4.1. Seguindo Mittelmann e Vallentin, a tabela inclui diversos valores de d e dígitos decimais, já que a sequência de valores

n	*lim. inf.*	d	*lim. sup. anterior [MV10]*	*lim. sup. novo*	n	*lim. inf.*	d	*lim. sup. anterior [MV10]*	*lim. sup. novo*
3	12	14	12.38180947	12.381921	14	1606	14	3183.133169	3183.348148
		15		12.374682			15		3180.112464
		16		12.368591			16		3177.917052
4	24	14	24.06628391	24.066298	15	2564	14	4866.245659	4866.795537
		15		24.062758			15		4862.382161
		16		24.056903			16		4858.505436
5	40	14	44.99899685	44.999047	16	4320	14	7355.809036	7356.238006
		15		44.987727			15		7341.324655
		16		44.981067			16		7332.776399
6	72	14	78.24061272	78.240781	17	5346	14	11072.37543	11073.844334
		15		78.212731			15		11030.170254
		16		78.187761			16		11014.183845
7	126	14	134.4488169	134.456246	18	7398	14	16572.26478	16575.934858
		15		134.330898			15		16489.848647
		16		134.270201			16		16469.090329
9	306	14	364.0919287	364.104934	19	10668	14	24812.30254	24819.810569
		15		363.888016			15		24654.968481
		16		363.675154			16		24575.871259
10	500	14	554.5075418	554.522392	20	17400	14	36764.40138	36761.630730
		15		554.225840			15		36522.436885
		16		553.827497			16		36402.675795
11	582	14	870.8831157	870.908146	21	27720	14	54584.76757	54579.036297
		15		869.874183			15		54069.067238
		16		869.244985			16		53878.722941
12	840	14	1357.889300	1357.934329	22	49896	14	82340.08003	82338.035075
		15		1357.118955			15		81688.317095
		16		1356.603728			16		81376.459564
13	1154	14	2069.587585	2069.675634	23	93150	14	124416.9796	124509.320059
		15		2067.388613			15		123756.492951
		16		2066.405173			16		123328.397290

TABELA 4.1. Limitantes inferiores e superiores para o número de contato nas dimensões 3, ..., 24. As dimensões 8 e 24 foram omitidas já que o limitante de programação linear já é justo. Todos os limitantes inferiores podem ser encontrados no livro de Conway e Sloane [CS99], exceto nas dimensões 13 e 14, obtidas por Zinoviev e Ericson [ZE99]. As melhorias sobre os limitantes previamente conhecidos estão sublinhadas. Todos os limitantes foram verificados rigorosamente com o método descrito na Seção 4.4.

nos indica o quão forte o limitante de programação semidefinida de Bachoc e Vallentin [BV08] pode ser se polinômios de grau mais alto forem usados. Os dígitos decimais na dimensão 4 também são interessantes, já que um limitante justo poderia dar-nos informação sobre as configurações ótimas (ainda é um problema em aberto se a configuração de 24 pontos em dimensão 4 dada pelos vetores de menor tamanho do reticulado D_4 é única; como descrito na Seção 2.3, para as dimensões 8 e 24 a unicidade foi provada por Bannai e Sloane [BS81] usando a cota justa dada pelo limitante de programação linear).

Observamos também que muitos dos novos valores para $d = 14$ são na verdade maiores que os valores correspondentes obtidos por Mittelmann e Vallentin [MV10]. Isso ocorre pois os problemas resolvidos não são exatamente iguais: um programa usa a restrição (49) enquanto o outro usa a restrição (52).

4.4. Verificação rigorosa dos resultados

Os dados usados como entrada para o *solver* não são exatos, já que é necessário usar aritmética de ponto flutuante para calcular as matrizes V_d^i do Teorema 4.3 e o *solver* também usa aritmética de ponto flutuante para resolver o problema de otimização; portanto a solução obtida no fim é apenas uma aproximação numérica e o limitante obtido pode não ser rigoroso. Entretanto, caso a solução seja composta por matrizes positivo-definidas e próximas o suficiente de uma solução viável, é possível argumentar que ela pode ser transformada em uma solução viável sem alterar muito o resultado, produzindo assim um limitante rigoroso.

A técnica usada para verificar a solução é uma adaptação do processo usado por Dostert, Guzmán, Oliveira e Vallentin [DGdOFV17]. O primeiro passo é encontrar uma solução com matrizes positivo-definidas junto com um limitante inferior para o seu menor autovalor. Em seguida devemos verificar o quanto a solução viola as restrições do problema; se o limitante inferior do menor autovalor for grande comparado com a violação das restrições (em um sentido que será descrito precisamente a seguir), concluímos que é possível transformar a solução em uma solução exata sem alterar o valor do limitante. Essa comparação não pode ser feita com aritmética de ponto flutuante, em vez disso, a verificação se baseia em uma biblioteca de aritmética intervalar, como a MPFI [RR05], que representa números como intervalos e faz todos os arredondamentos para fora.

É conveniente usar uma restrição semelhante à restrição (49) no lugar da restrição (52):

$$b_{22} + \sum_{k=0}^{d} \langle F_k, S_k^n \rangle + \langle R_0, V_d \rangle + \langle R_1, s_1 V_{d-1} \rangle + \langle R_2, s_2 V_{d-2} \rangle \\ + \langle R_3, s_3 V_{d-3} \rangle + \langle R_4, s_4 V_{d-2} \rangle = 0, \quad (53)$$

com $R_0, \ldots, R_4 \succeq 0$ e $V_{d'} = v_{B_{d'}} v_{B_{d'}}^T$, com $B_{d'}$ sendo a base monomial de $\mathbb{R}[u, v, t]_{\leq d'}$ para cada d' que aparece na equação, pois (53) usa matrizes baseadas na base monomial de $\mathbb{R}[u, v, t]_{\leq d'}$, que podem ser calculadas de forma exata e assim conhecemos precisamente toda a entrada do problema. Dada uma solução obtida pelo *solver* para o problema com a restrição (52), podemos convertê-la em uma solução para o problema com a restrição (53). Para isso, note que no processo descrito na demonstração do Teorema 4.3, a matriz $\overline{Q}$ torna-se bloco-diagonal após a mudança entre a base monomial e a base dada pelos Teoremas 1.8 e 1.21, portanto a conversão se resume em desfazer essa mudança de base. O tamanho do programa aumenta, já que estamos perdendo a estrutura bloco-diagonal, mas isso não é um problema, já que o programa já foi resolvido pelo *solver* e a conversão não é uma operação cara.

Para obter uma solução com matrizes positivo-definidas, o problema é resolvido com uma pequena mudança de variáveis, em que cada variável X é substituída por $X' + \lambda_{\min}$, com a restrição $X' \succeq 0$. Isso faz a restrição semidefinida ser mais rígida, o que penaliza o valor do limitante obtido, mas nos dá uma solução com menor autovalor aproximadamente $\lambda_{\min}$. Esse parâmetro auxiliar deve ser escolhido pequeno o suficiente para que a perda no limitante seja pequena, mas grande o suficiente para que a comparação ao final do processo funcione. Após alguns testes, valores em torno de 10^{-8} e 10^{-10} funcionaram bem.

Para tornar o limitante rigoroso, para cada variável X da solução obtida, procuramos um $\lambda_X > 0$ tal que $X - \lambda_X I$ tenha uma decomposição de Cholesky $L_X L_X^T$; normalmente esse λ_X é um pouco menor do que o $\lambda_{\min}$ usado na etapa anterior. Todos esses cálculos ainda são feitos com aritmética de ponto flutuante, mas agora

podemos expressar X como

$$X = L_X L_X^T + \lambda_X I$$

e interpretar o lado direito como a solução aproximada do problema de forma que tenhamos um limitante para o seu menor autovalor; a partir deste ponto todos os cálculos são feitos com aritmética intervalar.

O próximo passo é checar as restrições (47) e (53); como ambos os casos são similares, consideremos a restrição (53). Como dito, os coeficientes da restrição são conhecidos exatamente e assim podemos usar aritmética intervalar para calcular o polinômio r que aparece no lado esquerdo de (53) quando todas as variáveis são substituídas pelas correspondentes expressões $L_X L_X^T + \lambda_X I$. Temos que r provavelmente não será 0, porém observamos que $V_d = v_{B_d} v_{B_d}^T$ possui a base monomial dos polinômios de grau no máximo $2d$ em suas entradas e assim podemos encontrar A tal que $r = \langle A, V_d \rangle$ (isso pode ser feito com A sendo simétrica, basta colocar os coeficientes de r nos lugares correspondentes de A e então substituir A por $(A+A^T)/2$). Substituindo R_0 por $L_{R_0} L_{R_0}^T + \lambda_{R_0} - A$ em (53), temos uma solução exata; mas agora temos que checar se $L_{R_0} L_{R_0}^T + \lambda_{R_0} - A \succeq 0$. Isso será satisfeito se $\|A\| \leq \lambda_{R_0}$ (onde $\|A\| = \langle A, A \rangle^{1/2}$ é a norma de Frobenius de A); de fato, tomando qualquer vetor $w \in \mathbb{R}^{B_d}$, temos

$$\begin{aligned}
w^T(L_{R_0} L_{R_0}^T + \lambda_{R_0} - A)w &= w^T(L_{R_0} L_{R_0}^T + \lambda_{R_0})w - \sum_{p,q \in B_d} A(p,q)w(p)w(q) \\
&\geq \lambda_{R_0} w^T w - \Big(\sum_{q \in B_d} w(q)^2 \Big)^{1/2} \Big(\sum_{q \in B_d} \Big(\sum_{p \in B_d} A(p,q)w(p) \Big)^2 \Big)^{1/2} \\
&\geq \lambda_{R_0} w^T w - \Big(\sum_{q \in B_d} w(q)^2 \Big)^{1/2} \Big(\sum_{q \in B_d} \big(\sum_{p \in B_d} A(p,q)^2 \big) \big(\sum_{p \in B_d} w(p)^2 \big) \Big)^{1/2} \\
&= \lambda_{R_0} w^T w - \Big(\sum_{q \in B_d} w(q)^2 \Big) \Big(\sum_{p,q \in B_d} A(p,q)^2 \Big)^{1/2} \\
&= (\lambda_{R_0} - \|A\|) w^T w \geq 0.
\end{aligned}$$

A condição $\lambda_{R_0} \geq \|A\|$ pode ser verificada com a aritmética intervalar se os valores de $\|A\|$ e λ_{R_0} forem distintos o suficiente.

Para este processo funcionar, devemos escolher no primeiro passo $\lambda_{\min}$ grande o suficiente comparado com o resíduo $\|A\|$, que por sua vez depende do tamanho do problema e da precisão numérica do *solver*. Isso apenas pode ser feito sem uma grande perda no valor do limitante se o problema for resolvido com um *solver* de alta precisão; aqui todos os programas semidefinidos foram resolvidos com o SDPA-GMP [Nak10] configurado com 200 bits de precisão (equivalente a cerca de 60 casas decimais). Todos os resultados descritos na última coluna da Tabela 4.1 passaram por essa verificação.

CAPÍTULO 5

Extensão para grafos topológicos de empacotamento

Vemos neste capítulo como estender o limitante de programação semidefinida, até aqui definido para limitar o tamanho de códigos esféricos, para uma classe maior de problemas. Consideramos o número de independência de grafos com o conjunto de vértices possivelmente infinito, porém com certas propriedades topológicas adicionais que surgem em problemas de empacotamento e que serão descritas na Seção 5.1.

Grande parte deste capítulo segue o trabalho de de Laat e Vallentin [dLV15], onde é discutida a noção de "limitantes de k pontos", útil para comparar a complexidade de diversos limitantes para o número de independência; veremos que o limitante de programação linear é um limitante de 2 pontos, enquanto o limitante de programação semidefinida é um limitante de 3 pontos.

5.1. Grafos topológicos de empacotamento

Estamos interessados em grafos cujos vértices que sejam "próximos" sejam adjacentes e nos quais vértices adjacentes continuem adjacentes após sofrer "pequenas perturbações". Mais precisamente, um grafo cujo conjunto de vértices é um espaço topológico de Hausdorff é chamado de *grafo topológico de empacotamento* se cada clique finito estiver contido em um clique aberto.

Note que a condição presente na definição só precisa ser verificada para cliques de tamanho um e dois. De fato, se $\{x_1, \dots, x_t\}$ é um clique de tamanho $t > 2$, podemos escolher cliques abertos $U_{\{i,j\}}$ que contenham cada par $\{x_i, x_j\}$ com $i \neq j$, definir $U_i = \bigcap_{j \neq i} U_{\{i,j\}}$ e então tomar $U = \bigcup_{i=1}^{t} U_i$ como um clique aberto que contenha $\{x_1, \dots, x_t\}$. Qualquer grafo considerado com a topologia discreta (por exemplo, um grafo finito) é um grafo topológico de empacotamento, entretanto topologias mais fracas levam a condições mais fortes: nesta dissertação consideramos grafos com conjunto de vértices compacto, nesse caso a condição de grafo topológico de empacotamento implica naturalmente que os conjuntos independentes sejam finitos, pois o conjunto de vértices possui uma cobertura finita por cliques.

Exemplo 5.1 (Grafos de distância). Um *grafo de distância* $G = (V, E)$ é um grafo no qual $V = (V, d)$ é um espaço métrico e em que existe $D \subseteq (0, \infty)$ tal que x, y são adjacentes precisamente quando $d(x, y) \in D$. Se D for aberto e contiver o intervalo $(0, \delta)$ para algum $\delta > 0$, então pode-se mostrar que G é um grafo topológico de empacotamento. O grafo $G_{n,\theta}$ apresentado na introdução para modelar o problema dos códigos esféricos é um grafo de distância.

Exemplo 5.2 (Empacotamento de corpos). Seja (V, d) um espaço métrico, K um subconjunto de V com interior não-vazio e Γ um grupo topológico que age continuamente em V preservando distâncias. Um *empacotamento* de K em V é uma coleção de imagens $\{gK : g \in S\}$ com $S \subseteq \Gamma$ tal que se $f, g \in S$, então $\operatorname{int}(fK) \cap \operatorname{int}(gK) = \emptyset$.

Podemos definir um grafo G com conjunto de vértices Γ e cujos conjuntos independentes correspondem aos empacotamentos. Para isso, note que $\operatorname{int}(fK) \cap \operatorname{int}(gK) \neq \emptyset$ se e somente se $\operatorname{int} K \cap \operatorname{int}(f^{-1}gK) \neq \emptyset$, de modo que definindo

$$X := \{g \in \Gamma \setminus \{e\} : \operatorname{int} K \cap \operatorname{int}(gK) \neq \emptyset\},$$

tomamos como arestas os conjuntos $\{f, g\} \subset \Gamma$ tais que $f^{-1}g \in X$. Esse grafo é um grafo de Cayley[1] e X é o seu conjunto de conexão.

Provemos que G é um grafo topológico de empacotamento. Seja $\{f, g\} \subset \Gamma$ um clique de tamanho um ou dois. Como $\operatorname{int}(fK) \cap \operatorname{int}(gK) \neq \emptyset$, existe uma bola aberta de centro x_0 e raio r, que denotaremos por $B(x_0, r)$, contida em $\operatorname{int}(fK) \cap \operatorname{int}(gK)$. Como a ação de Γ preserva distâncias, isso implica que $B(f^{-1}x_0, r) \cup B(g^{-1}x_0, r) \subset K$. Seja

$$U := \{h \in \Gamma : hf^{-1}x_0 \in B(x_0, r/2) \text{ ou } hg^{-1}x_0 \in B(x_0, r/2)\},$$

temos que U é aberto, pois é a união das imagens inversas de $B(x_0, r/2)$ sob duas funções contínuas, $\{f, g\} \subseteq U$ e U é um clique, pois para qualquer $h \in U$, o aberto $B(x_0, r/2)$ está contido em hK; de fato, se $h \in U$, temos que $d(hf^{-1}x_0, x_0) < r/2$ (o argumento é análogo no caso com g) e para qualquer $x \in B(x_0, r/2)$,

$$\begin{aligned} d(h^{-1}x, f^{-1}x_0) &\leq d(h^{-1}x, h^{-1}x_0) + d(h^{-1}x_0, f^{-1}x_0) \\ &= d(x, x_0) + d(hf^{-1}x_0, x_0) \\ &< r, \end{aligned}$$

logo $h^{-1}x \in B(f^{-1}x_0, r) \subset K$ e $x \in hK$.

Este exemplo engloba diversos problemas considerados na literatura, entre eles o problema dos códigos binários corretores de erro, que correspondem ao empacotamento de bolas no cubo de Hamming $\{0, 1\}^n$ (um domínio finito), o problema do empacotamento de calotas esféricas em S^{n-1} (o problema central desta dissertação, um domínio compacto) e o problema do empacotamento de corpos convexos no espaço euclideano (um domínio localmente compacto).

5.1.1. Topologia da família de conjuntos independentes. Uma das principais consequências de um grafo ser um grafo topológico de empacotamento é a estrutura topológica da sua família de conjuntos independentes. Vejamos primeiro como definir uma topologia para essa família.

Seja $G = (V, E)$ um grafo topológico de empacotamento. Denotemos por $\operatorname{sub}_t(V)$ a família de subconjuntos com até t vértices, por I_t a família de subconjuntos independentes com até t vértices, por $I_{=t}$ a família de subconjuntos independentes com exatamente t vértices e também $I'_t := I_t \setminus \{\emptyset\}$ e $\operatorname{sub}'_t(V) := \operatorname{sub}_t(V) \setminus \{\emptyset\}$. Desejamos atribuir uma topologia à[2] I'_t. Para isso, vemos I'_t como um subconjunto de $\operatorname{sub}'_t(V)$ e, conforme feito na Seção 2 de Handel [Han00], consideramos V^t com a topologia produto e em seguida a imagem de V^t sob o mapa

$$q\colon V^t \to \operatorname{sub}'_t(V), \quad q(v_1, ..., v_t) = \{v_1, ..., v_t\}$$

com a topologia quociente, isto é, tal que $S \subseteq \operatorname{sub}'_t(V)$ é aberto se $q^{-1}(S)$ for aberto em V^t (essa topologia também é conhecida como a maior topologia que faz q ser contínua).

---

[1] Em geral, se Γ é um grupo, $X \subset \Gamma$ é tal que $e \notin \Gamma$ e $X = X^{-1}$, o *grafo de Cayley* $G = \operatorname{Cay}(\Gamma, X)$ com *conjunto de conexão* X é o grafo com conjunto de vértices $V(G) = \Gamma$ e arestas $E(G) = \{\{f, g\} \subset \Gamma : f^{-1}g \in X\}$. As condições sobre X garantem que o grafo seja simples.

[2] Em de Laat e Vallentin [dLV15] é considerado I_t, aqui o conjunto vazio não será importante, porém a distinção na notação será mantida.

É conveniente construir uma base de abertos para $\mathrm{sub}'_t(V)$. Para $r = 1, \ldots, t$ e $U_1, \ldots, U_r \subseteq V$, seja

$$(U_1, \ldots, U_r)_t := \{S \in \mathrm{sub}'_t(V) : S \subseteq U_1 \cup \cdots \cup U_r,\ S \cap U_i \neq \emptyset \text{ para } i = 1, \ldots, r\}.$$

Conforme observado por Handel [Han00],

$$q^{-1}((U_1, \ldots U_r)_t) = \bigcup_{\substack{\tau\colon \{1,\ldots,t\} \to \{1,\ldots,r\} \\ \tau \text{ é sobrejetora}}} U_{\tau(1)} \times \cdots \times U_{\tau(t)}$$

e portanto se $U_1, \ldots, U_r$ forem abertos em V, então $(U_1, \ldots, U_r)_t$ é aberto em $\mathrm{sub}'_t(V)$. Ademais, conforme a Proposição 2.11 de Handel [Han00], se $\mathcal{B}$ é uma base de abertos de V, então

$$\mathcal{B}_t := \{(U_1, \ldots, U_r)_t : 1 \leq r \leq t, U_1, \ldots, U_r \in \mathcal{B}\}$$

é uma base de abertos de $\mathrm{sub}'_t(V)$ e se $\{u_1, \ldots, u_r\}$ é um elemento de um conjunto aberto U de $\mathrm{sub}'_t(V)$, então existem vizinhanças U_i de cada u_i tais que $(U_1, \ldots, U_r)_t$ seja uma vizinhança de $\{u_1, \ldots, u_r\}$ contida em U.

Como a topologia de $\mathrm{sub}'_t(V)$ foi definida a partir das operações de produto e quociente, se V for compacto, então segue imediatamente da compacidade de V que $\mathrm{sub}'_t(V)$ é compacto. Também pode-se mostrar que $\mathrm{sub}'_t(V)$ é Hausdorff (veja Handel [Han00], Proposição 2.7).

Finalmente, temos uma topologia em I'_t e em $I_{=t}$ vendo-os como subconjuntos de $\mathrm{sub}'_t(V)$. Observe que $I_{=1}$ é homeomorfo a V, por isso em alguns momentos trocamos um pelo outro no texto.

Vejamos agora duas proposições nas quais a hipótese do grafo ser um grafo topológico de empacotamento assume papel central.

Proposição 5.3. *Seja $G = (V, E)$ um grafo topológico de empacotamento com conjunto de vértices compacto. Então I'_t é compacto para todo $t \in \mathbb{N}$.*

Demonstração. Mostremos que I'_t é um subconjunto fechado do espaço compacto $\mathrm{sub}'_t(V)$. Tome $\{x_1, \ldots, x_r\} \in \mathrm{sub}'_t(V) \setminus I'_t$; sem perda de generalidade podemos supor que x_1 e x_2 são adjacentes. Pela hipótese, existe um clique aberto $U \subseteq V$ que contém $\{x_1, x_2\}$ e, como V é Hausdorff, existem abertos disjuntos $U_1, U_2 \subseteq U$ tais que $x_1 \in U_1$ e $x_2 \in U_2$. Considere $(U_1, U_2, V, \ldots, V)_t$, onde V aparece $r - 2$ vezes. Temos que $(U_1, U_2, V, \ldots, V)_t$ é uma vizinhança de $\{x_1, \ldots, x_r\}$ contida em $\mathrm{sub}'_t(V) \setminus I'_t$, pois se $S \in (U_1, U_2, V, \ldots, V)_t$, então S contém $s_1 \in U_1$ e $s_2 \in U_2$, que são adjacentes. □

Proposição 5.4. *Seja $G = (V, E)$ um grafo topológico de empacotamento. Então o mapa $c\colon I'_t \to \mathbb{N}$, $c(S) = |S|$ é contínuo para todo $t \in \mathbb{N}$. Em particular, $I_{=r}$ é aberto e fechado em I'_t para $r = 1, \ldots, t$.*

Demonstração. Seja $\{x_1, \ldots, x_r\} \in c^{-1}(r)$. Pela hipótese sobre G, existem cliques abertos e disjuntos $U_1, \ldots, U_r$ tais que $x_i \in U_i$ para $i = 1, \ldots, r$. Logo o conjunto $(U_1, \ldots, U_r)_t \cap I'_t$ é aberto em I'_t e contém $\{x_1, \ldots, x_r\}$. Temos que $(U_1, \ldots, U_r)_t \cap I'_t \subseteq c^{-1}(r)$, pois se $S \in (U_1, \ldots, U_r)_t \cap I'_t$, então $|S| \geq r$ já que os conjuntos U_i são disjuntos e $|S| \leq r$ já que os conjuntos U_i são cliques. □

5.2. Matriz de momentos e limitantes de k pontos para grafos finitos

Calcular o número de independência de um grafo finito é um problema NP-difícil que atraiu a atenção de muitos pesquisadores para o desenvolvimento de técnicas que encontrem soluções aproximadas ou limitantes através do uso de programação linear e semidefinida (tal como o número teta de Lovász, apresentado na Seção 2.1).

Essas técnicas aplicam-se mais geralmente ao caso em que desejamos otimizar uma função linear sobre um politopo $P := \text{conv}(K \cap \{0,1\}^n)$, com K um conjunto semi-algébrico (como em (45)), e envolvem técnicas de extensão e projeção ("lift-and-project methods") em que consideramos projeções de objetos em espaços de dimensão maior. Exemplos são a hierarquia de Lovász-Schrijver [LS91] e a de Lasserre [Las02], que constroem uma sequência de relaxações $K \supseteq K^1 \supseteq K^2 \supseteq \cdots \supseteq K^n = P$ que convergem para P em n passos e em que, sob certas condições sobre K, pode-se otimizar uma função linear em K^t em tempo polinomial para qualquer t fixo.

Laurent [Lau03] apresenta esses métodos em um quadro comum usando matrizes de momentos, que definimos a seguir. Para um conjunto finito V, $t \in \{1, \ldots, |V|\}$ e $y\colon \text{sub}_{2t}(V) \to \mathbb{R}$ (usamos por simplicidade as notações $y_S = y(S), y_i = y_{\{i\}}$ e $y_{ij} = y_{\{i,j\}}$), a *matriz truncada de momentos* $M_t(y)$ é a matriz

$$M_t(y)\colon \text{sub}_t(V) \times \text{sub}_t(V) \to \mathbb{R}, \quad M_t(y)_{S,T} := y_{S \cup T}. \tag{54}$$

Note que quando $x \in \{0,1\}^V$ e y é dado por $y_S = \prod_{i \in S} x_i$ para todo $S \in \text{sub}_{2t}(V)$, temos que $M_t(y) = yy^T \succeq 0$.

No caso do problema do número de independência de um grafo $G = (V, E)$, consideramos $K = \{x \in \mathbb{R}^V : 1 - x_u - x_v \geq 0 \text{ para todo } \{u, v\} \in E\}$. A hierarquia de Lasserre possui uma descrição simples nesse caso, que conforme descrito por Laurent [Lau03] (Lema 13, sendo que aqui estamos trocando $t+1$ por t de modo a começar a contagem em $t = 1$), no passo t produz um limitante $\text{las}_t(G)$ para o número de independência, definido como o valor ótimo do seguinte problema de otimização:

$$\begin{aligned} \max \quad & \sum_{u \in V} y_u \\ & y\colon \text{sub}_{2t}(V) \to \mathbb{R}, \\ & y_S = 0 \text{ para todo } S \in \text{sub}_{2t}(V) \setminus I_{2t}, \\ & y_\emptyset = 1, \\ & M_t(y) \succeq 0. \end{aligned} \tag{55}$$

Não é difícil mostrar diretamente que $\alpha(G) \leq \text{las}_t(G)$. Para isso, basta verificar que para qualquer $I \subseteq V$ independente, y definido por $y_S = 1$ se $S \subseteq I$ e $y_S = 0$ caso contrário é uma solução viável de valor $|I|$. Laurent [Lau03] também mostra (Proposição 14) que $\text{las}_{\alpha(G)}(G) = \alpha(G)$.

Como explicado por Bachoc, Gijswijt, Schrijver e Vallentin [BGSV12] (Seção 4.3), uma variação dessa hierarquia é obtida restringindo y a conjuntos com pelo menos um elemento, removendo a linha e a coluna correspondentes ao conjunto vazio de $M_t(y)$ (denotemos a matriz resultante por $M'_t(y)$) e adicionando restrições de não-negatividade. A condição $y_\emptyset = 1$ é substituída por $\sum_{u \in V} y_u = 1$ e o objetivo por $\sum_{u,v \in V} y_{u,v}$. Como $M_1(y) \succeq 0$ implica que $(\sum_{u \in V} y_u)^2 \leq y_\emptyset \sum_{u,v \in V} y_{u,v}$, obtemos uma relaxação. A variação descrita corresponde ao valor ótimo do seguinte problema de otimização, que denominamos número teta-linha de Lovász estendido

$\vartheta'_t(G)$, visto que para $t = 1$ obtemos o Programa (21):

$$\begin{aligned} \max \quad & \sum_{u,v\in V} y_{u,v} \\ & y\colon \mathrm{sub}'_{2t}(V) \to \mathbb{R}_{\geq 0}, \\ & y_S = 0 \text{ para todo } S \in \mathrm{sub}'_{2t}(V) \setminus I'_{2t}, \\ & \sum_{u\in V} y_u = 1, \\ & M'_t(y) \succeq 0. \end{aligned} \tag{56}$$

Diversos limitantes para $\alpha(G)$ podem ser obtidos considerando diferentes submatrizes principais de $M_{|V|}(y)$ e restrições extras. A denominação *limitante de k pontos* refere-se ao tamanho k do maior subconjunto de V considerado no domínio de y e serve como parâmetro de medição da complexidade de um limitante (em termos do número de variáveis usadas) assim como de seu potencial (espera-se que um limitante que considere relações entre subconjuntos maiores de vértices produza resultados melhores). Nesse sentido, os limitantes $\mathrm{las}_t(G)$ e $\vartheta'_t(G)$ são limitantes de $2t$ pontos.

Como o cálculo de um limitante de 4 pontos já é muitas vezes inviável na prática, é interessante a formulação de um limitante de 3 pontos que produza alguma melhoria sobre o número teta-linha de Lovász. Podemos fazer isso considerando submatrizes principais de $M'_2(y)$ escolhidas de modo que só apareçam conjuntos com até três vértices. Sejam $M'_w(y)$ as matrizes definidas para cada $w \in V$ por:

$$M'_w(y)\colon V \times V \to \mathbb{R}, \quad M'_w(y)_{u,v} := y_{\{u,v,w\}}. \tag{57}$$

Partindo de $\vartheta'_1(G)$ e adicionando a condição de que as matrizes $M'_w(y)$ sejam positivo-semidefinidas, obtemos o seguinte problema de otimização cujo valor ótimo denotamos por $\mathrm{3PB}(G)$:

$$\begin{aligned} \max \quad & \sum_{u,v\in V} y_{u,v} \\ & y\colon \mathrm{sub}'_3(V) \to \mathbb{R}_{\geq 0}, \\ & y_S = 0 \text{ para todo } S \in \mathrm{sub}'_3(V) \setminus I'_3, \\ & \sum_{u\in V} y_u = 1, \\ & M'_1(y) \succeq 0, \\ & M'_w(y) \succeq 0 \text{ para todo } w \in V. \end{aligned} \tag{58}$$

5.3. Limitantes de 3 pontos para grafos topológicos de empacotamento

Nesta seção generalizamos o Programa (58) para grafos topológicos de empacotamento compactos. Antes, precisamos discutir os espaços que serão considerados a fim de termos pares duais e podermos usar a teoria de dualidade vista na Seção 1.3.

Tendo em vista a condição $y_S = 0$ para todo $S \in \mathrm{sub}'_3(V)\setminus I'_3$ no Programa (58), substituímos o domínio de y por I'_3, que pela Proposição 5.3 é compacto. Seja $\mathcal{C}(I'_3)$ o conjunto das funções reais e contínuas em I'_3 considerado inicialmente com a topologia induzida pela norma do supremo. Pelo Teorema de Representação de Riesz (veja por exemplo Berg, Christensen e Ressel [BCR84], Teorema 2.4), seu dual topológico pode ser identificado com o espaço $\mathcal{M}(I'_3)$ das medidas de Radon com

sinal (i.e., definidas na σ-álgebra de Borel, finitas em compactos e internamente regulares) e, conforme visto na Seção 1.3.2, esses espaços formam um par dual com a dualidade $\langle f, \mu\rangle := \int f \,\mathrm{d}\mu$ para todo $f \in \mathcal{C}(I'_3)$ e $\mu \in \mathcal{M}(I'_3)$. Nesse caso, as topologias fracas da dualidade são conhecidas como topologia fraca de $\mathcal{C}(I'_3)$ e fraca-$*$ de $\mathcal{M}(I'_3)$. Consideramos o cone $\mathcal{C}(I'_3)_{\geq 0}$ das funções não-negativas e denotamos seu cone dual $(\mathcal{C}(I'_3)_{\geq 0})^*$ por $\mathcal{M}(I'_3)_{\geq 0}$. A variável y torna-se $\lambda \in \mathcal{M}(I'_3)_{\geq 0}$.

Para representar a condição $M'_1(y) \succeq 0$ de (58), consideramos o espaço dos kernels contínuos e simétricos $\mathcal{C}(V \times V)_{\mathrm{sym}}$ e o cone $\mathcal{C}(V \times V)_{\succeq 0}$ dos kernels contínuos e positivos (aqui, em vez de usar a definição de positividade introduzida na Seção 1.2.1, que depende de uma medida previamente definida em V, usamos a caracterização da Proposição 1.22, isto é, dizemos que um kernel contínuo e simétrico é *positivo* se a sua restrição a qualquer subconjunto finito de V é uma matriz positivo-semidefinida). Denotamos por $\mathcal{M}(V \times V)_{\mathrm{sym}}$ o dual topológico de $\mathcal{C}(V \times V)_{\mathrm{sym}}$ (que também pode ser identificado com o espaço das medidas de Radon com sinal simétricas, isto é, tais que $\mu(E, E') = \mu(E', E)$ para todos os borelianos $E, E' \subset V$) e denotamos por $\mathcal{M}(V \times V)_{\succeq 0}$ o cone dual de $\mathcal{C}(V \times V)_{\succeq 0}$. Usamos a mesma técnica usada por de Laat e Vallentin [dLV15] e generalizamos o operador $M'_1 \colon \mathbb{R}^{\mathrm{sub}'_3(V)} \to \mathbb{R}^{V \times V}$ através de seu operador adjunto. Como para quaisquer $y \in \mathbb{R}^{\mathrm{sub}'_3(V)}$ com $y_S = 0$ para todo $S \in \mathrm{sub}'_3(V) \setminus I'_3$ e $Y \in \mathbb{R}^{V \times V}$ vale

$$\langle M'_1(y), Y\rangle = \sum_{u,v \in V} y_{\{u,v\}} Y(u,v) = \sum_{S \in I'_3} y_S \sum_{\substack{u,v \in V: \\ \{u,v\}=S}} Y(u,v),$$

definimos o operador $A_1 \colon \mathcal{C}(V \times V)_{\mathrm{sym}} \to \mathcal{C}(I'_3)$ por:

$$A_1 K(S) := \sum_{\substack{u,v \in V: \\ \{u,v\}=S}} K(u,v). \tag{59}$$

Note que $A_1 K(S)$ é dado por expressões diferentes dependendo do tamanho de S:

$$\begin{aligned} A_1 K(\{u\}) &= K(u,u) && \text{se } \{u\} \in I_{=1}, \\ A_1 K(\{u,v\}) &= K(u,v) + K(v,u) && \text{se } \{u,v\} \in I_{=2}, \\ A_1 K(\{u,v,w\}) &= 0 && \text{se } \{u,v,w\} \in I_{=3}. \end{aligned}$$

Porém ainda assim $A_1 K$ é uma função contínua pois, pela Proposição 5.4, I'_3 pode ser escrito como a união disjunta dos abertos $I_{=1}$, $I_{=2}$ e $I_{=3}$ e $A_1 K$ é contínua em cada uma dessas partes. Como $\|A_1 K\|_\infty \leq 2\|K\|_\infty$, A_1 é um operador contínuo quando $\mathcal{C}(V \times V)_{\mathrm{sym}}$ e $\mathcal{C}(I'_3)$ são considerados com suas topologias induzidas pelas respectivas normas do supremo e portanto A_1 também é contínuo quando o domínio e o contradomínio são considerados com suas topologias fracas (veja Megginson [Meg98], Teorema 2.5.11), logo pela Proposição 1.28 existe um operador adjunto $A_1^* \colon \mathcal{M}(I'_3) \to \mathcal{M}(V \times V)_{\mathrm{sym}}$. A condição $M'_1(y) \succeq 0$ torna-se $A_1^* \lambda \in \mathcal{M}(V \times V)_{\succeq 0}$, com $\lambda \in \mathcal{M}(I'_3)$.

Para representar as condições $M'_w(y) \succeq 0$ para todo $w \in V$ do Programa (58), definimos $\mathcal{C}(V \times V \times V)_{\mathrm{sym}}$ como o espaço das funções contínuas $T \colon V \times V \times V \to \mathbb{R}$ simétricas nas primeira e segunda coordenadas, isto é, tais que $T(u,v,w) = T(v,u,w)$ para todos $u, v, w \in V$. Definimos $\mathcal{C}(V \times V \times V)_{\succeq 0}$ como o cone das funções $T \in \mathcal{C}(V \times V \times V)_{\mathrm{sym}}$ para as quais as matrizes $\big(T(x_i, x_j, w)\big)_{i,j=1}^n$ são positivo-semidefinidas para todo $n \in \mathbb{N}$ e $w, x_1, \ldots, x_n \in V$. Denotamos por $\mathcal{M}(V \times V \times V)_{\mathrm{sym}}$ o dual topológico de $\mathcal{C}(V \times V \times V)_{\mathrm{sym}}$ (que também pode ser identificado com o espaço das medidas de Radon com sinal simétricas, isto é, tais que $\mu(E, E', E'') = \mu(E', E, E'')$ para todos os borelianos $E, E', E'' \subset V$) e por $\mathcal{M}(V \times V \times V)_{\succeq 0}$ o cone dual de $\mathcal{C}(V \times V \times V)_{\succeq 0}$. Como para quaisquer $y \in \mathbb{R}^{\mathrm{sub}'_3(V)}$ com $y_S = 0$ para

todo $S \in \mathrm{sub}_3'(V) \setminus I_3'$ e $\{Y_w\}_{w\in V} \subset \mathbb{R}^{V\times V}$ vale

$$\sum_{w\in V} \langle M_w'(y), Y_w\rangle = \sum_{u,v,w\in V} y_{\{u,v,w\}} Y_w(u,v) = \sum_{S\in I_3'} y_S \sum_{\substack{u,v,w\in V:\\ \{u,v,w\}=S}} Y_w(u,v),$$

definimos o operador $A_{3\mathrm{PB}}\colon \mathcal{C}(V\times V\times V)_{\mathrm{sym}} \to \mathcal{C}(I_3')$ por:

$$A_{3\mathrm{PB}}T(S) := \sum_{\substack{u,v,w\in V:\\ \{u,v,w\}=S}} T(u,v,w). \tag{60}$$

Note que $A_{3\mathrm{PB}}T(S)$ é dado por expressões diferentes dependendo do tamanho de S:

$$\begin{aligned}
A_{3\mathrm{PB}}T(\{u\}) &= T(u,u,u) && \text{se } \{u\}\in I_{=1},\\
A_{3\mathrm{PB}}T(\{u,v\}) &= T(u,v,u)+T(v,u,u)+T(v,v,u)\\
&\quad + T(u,v,v)+T(v,u,v)+T(u,u,v) && \text{se } \{u,v\}\in I_{=2},\\
A_{3\mathrm{PB}}T(\{u,v,w\}) &= T(u,v,w)+T(v,u,w)+T(u,w,v)\\
&\quad + T(w,u,v)+T(v,w,u)+T(w,v,u) && \text{se } \{u,v,w\}\in I_{=3}.
\end{aligned}$$

Novamente, é uma consequência da Proposição 5.4 que $A_{3\mathrm{PB}}T$ seja uma função contínua, pois I_3' pode ser escrito como a união disjunta dos abertos $I_{=1}$, $I_{=2}$ e $I_{=3}$ e $A_{3\mathrm{PB}}T$ é contínua em cada uma dessas partes. Como $\|A_{3\mathrm{PB}}T\|_\infty \le 6\|T\|_\infty$, temos que $A_{3\mathrm{PB}}$ é um operador contínuo com relação às topologias das normas e também com relação às topologias fracas e portanto existe o operador adjunto $A_{3\mathrm{PB}}^*\colon \mathcal{M}(I_3') \to \mathcal{M}(V\times V\times V)_{\mathrm{sym}}$. As condições $M_w'(y)\succeq 0$ tornam-se $A_{3\mathrm{PB}}^*\lambda \in \mathcal{M}(V\times V\times V)_{\succeq 0}$, com $\lambda\in\mathcal{M}(I_3')$.

Por fim, a condição $\sum_{u\in V} y_u = 1$ torna-se simplesmente $\lambda(I_{=1})=1$ e a função objetivo $\sum_{u,v\in V} y_{u,v}$ torna-se $\lambda(I_{=1})+2\lambda(I_{=2})$. Com todas essas substituições, definimos para um grafo topológico de empacotamento compacto $G=(V,E)$ o parâmetro $3\mathrm{PB}(G)$ como o valor ótimo do seguinte programa de otimização:

$$\begin{aligned}
\max\quad & \lambda(I_{=1})+2\lambda(I_{=2})\\
& \lambda\in\mathcal{M}(I_3')_{\succeq 0},\\
& \lambda(I_{=1})=1,\\
& A_1^*\lambda\in\mathcal{M}(V\times V)_{\succeq 0},\\
& A_{3\mathrm{PB}}^*\lambda\in\mathcal{M}(V\times V\times V)_{\succeq 0}.
\end{aligned}\tag{61}$$

Esse programa é uma generalização do Programa (58) pois segue da forma como ele foi construído que se G for um grafo com conjunto de vértices finito e considerado com a topologia discreta, os Programas (58) e (61) são iguais.

Após alguma manipulação, podemos construir o dual do Programa (61) usando a teoria vista na Seção 1.3.4. Obtemos o seguinte programa:

$$\begin{aligned}
\min\quad & a\\
& a\in\mathbb{R},\\
& K\in\mathcal{C}(V\times V)_{\succeq 0},\\
& T\in\mathcal{C}(V\times V\times V)_{\succeq 0},\\
& A_1K + A_{3\mathrm{PB}}T \le_{\mathcal{C}(I_3')_{\succeq 0}} (a-1)1_{I_{=1}} - 21_{I_{=2}}.
\end{aligned}\tag{62}$$

A função $1_X\in\mathcal{C}(I_3')$ é a função indicadora que assume valor $1_X(S)=1$ se $S\in X$ e $1_X(S)=0$ caso contrário; para $X=I_{=1}$ e $X=I_{=2}$, essa função é

contínua devido à Proposição 5.4. A notação $\leq_{\mathcal{C}(I_3')_{\geq 0}}$ foi introduzida na Seção 1.3.3 e significa que a função à esquerda deve ter valor menor ou igual do que a função à direita para todo $S \in I_3'$.

É uma consequência da forma como foi construído e do processo de dualização que soluções viáveis para o Programa (62) produzam limitantes para $\alpha(G)$. Vejamos uma prova direta desse fato:

TEOREMA 5.5. *Se $G = (V, E)$ é um grafo topológico de empacotamento com conjunto de vértices compacto e a, K, T satisfazem as restrições do Programa* (62)*, então $\alpha(G) \leq a$.*

DEMONSTRAÇÃO. Seja $C \subset V$ um conjunto não-vazio e independente qualquer. Considere a quantidade:

$$\sum_{\substack{S \subseteq C: \\ S \in I_3'}} A_1 K(S) + A_{3\text{PB}} T(S).$$

Por um lado, temos

$$\begin{aligned}\sum_{\substack{S \subseteq C: \\ S \in I_3'}} A_1 K(S) + A_{3\text{PB}} T(S) &\leq \binom{|C|}{1}(a-1) + \binom{|C|}{2}(-2) + \binom{|C|}{3}0 \\ &= |C|(a - |C|).\end{aligned}$$

Pelo outro lado,

$$\begin{aligned}\sum_{\substack{S \subseteq C: \\ S \in I_3'}} A_1 K(S) + A_{3\text{PB}} T(S) &= \sum_{\substack{S \subseteq C: \\ S \in I_3'}} \left(\sum_{\substack{u,v \in V: \\ \{u,v\}=S}} K(u,v) + \sum_{\substack{u,v,w \in V: \\ \{u,v,w\}=S}} T(u,v,w) \right) \\ &= \sum_{u,v \in C} K(u,v) + \sum_{u,v,w \in C} T(u,v,w) \geq 0.\end{aligned}$$

Juntando as duas desigualdades, obtemos $|C| \leq a$. □

5.4. Simetrização para grafos homogêneos

Seja $G = (V, E)$ um grafo topológico de empacotamento com conjunto de vértices compacto. Dizemos que G é um *grafo homogêneo* se existe um grupo compacto Γ, subgrupo do grupo de automorfismos de G, que age continuamente em V e tal que a ação de Γ em V é transitiva.

A ação de Γ pode ser estendida para $\mathcal{C}(V \times V)_{\text{sym}}$ e para $\mathcal{C}(V \times V \times V)_{\text{sym}}$ componente a componente:

$$(\psi \circ K)(u,v) := K(\psi^{-1}u, \psi^{-1}v), \quad (\psi \circ T)(u,v,w) := T(\psi^{-1}u, \psi^{-1}v, \psi^{-1}w),$$

para $K \in \mathcal{C}(V \times V)_{\text{sym}}$, $T \in \mathcal{C}(V \times V \times V)_{\text{sym}}$, $\psi \in \Gamma$ e $u, v, w \in V$.

Se a, K e T satisfazem as restrições do Programa (62) e $\psi \in \Gamma$, então a, $\psi \circ K$ e $\psi \circ T$ também as satisfazem. Para ver isso, note que

$$A_1(\psi \circ K)(S) = \sum_{\substack{u,v \in V: \\ \{u,v\}=S}} K(\psi^{-1}u, \psi^{-1}v) = A_1 K(\psi^{-1}S)$$

e

$$A_{3\text{PB}}(\psi \circ T)(S) = \sum_{\substack{u,v,w \in V: \\ \{u,v,w\}=S}} T(\psi^{-1}u, \psi^{-1}v, \psi^{-1}w) = A_{3\text{PB}} T(\psi^{-1}S),$$

logo, como $\psi^{-1}S \in I_3'$ se e somente se $S \in I_3'$ e $|\psi^{-1}S| = |S|$, também vale

$$A_1(\psi \circ K) + A_{3\text{PB}}(\psi \circ T) \leq_{\mathcal{C}(I_3')_{\geq 0}} (a-1)1_{I_{=1}} - 21_{I_{=2}}.$$

Ademais, temos que $\psi \circ K \in \mathcal{C}(V \times V)_{\succeq 0}$ e $\psi \circ T \in \mathcal{C}(V \times V \times V)_{\succeq 0}$; a continuidade segue do fato de estarmos compondo funções contínuas e a positividade segue diretamente da definição, basta notar que restringir $\psi \circ K$ a um subconjunto finito $\{x_1, \ldots, x_n\}$ de V é o mesmo que restringir K a $\{\psi^{-1}x_1, \ldots, \psi^{-1}x_n\}$ e o mesmo argumento vale para $\psi \circ T$.

Como Γ é um grupo compacto e possui uma medida de Haar, podemos simetrizar uma solução a, K, T do Programa (62) e obter uma solução de mesmo valor com K e T invariantes sob a ação de Γ. Restringindo-nos a soluções invariantes, podemos simplificar o cálculo do limitante. Vemos na próxima proposição como uma função Γ-invariante de $\mathcal{C}(V \times V \times V)_{\succeq 0}$ pode ser expressa de um jeito mais simples:

Proposição 5.6. *Seja V um espaço topológico de Hausdorff compacto e Γ um grupo compacto que age continuamente em V de forma transitiva. Fixe $e \in V$ e seja H o subgrupo estabilizador de e sob a ação de Γ. Então existe uma bijeção Φ entre as funções Γ-invariantes de $\mathcal{C}(V \times V \times V)_{\succeq 0}$ e os kernels H-invariantes de $\mathcal{C}(V \times V)_{\succeq 0}$. A cada $T \in \mathcal{C}(V \times V \times V)^{\Gamma}_{\succeq 0}$, associamos $\Phi(T) \in \mathcal{C}(V \times V)^{H}_{\succeq 0}$ dado por*

$$\Phi(T)(x, y) := T(x, y, e).$$

Reciprocamente, a função inversa Φ^{-1} associa a cada $R \in \mathcal{C}(V \times V)^{H}_{\succeq 0}$ a função $\Phi^{-1}(R) \in \mathcal{C}(V \times V \times V)^{\Gamma}_{\succeq 0}$ dada por

$$\Phi^{-1}(R)(x, y, z) := R(\psi_z^{-1}x, \psi_z^{-1}y)$$

onde, para cada $z \in V$, $\psi_z \in \Gamma$ é tal que $\psi_z e = z$.

Demonstração. Para ver que $\Phi(T)$ é H-invariante, note que para $\psi \in H$ temos

$$\begin{aligned}\Phi(T)(\psi^{-1}x, \psi^{-1}y) &= T(\psi^{-1}x, \psi^{-1}y, e) \\ &= T(\psi^{-1}x, \psi^{-1}y, \psi^{-1}e) = T(x, y, e) = \Phi(T)(x, y).\end{aligned}$$

A positividade de $\Phi(T)$ segue diretamente da definição de $\mathcal{C}(V \times V \times V)_{\succeq 0}$ aplicada a $u = e$.

A função Φ^{-1} está bem definida, isto é, não depende da escolha de ψ_z, pois se escolhermos outro $\psi'_z \in \Gamma$ tal que $\psi'_z e = z$, então $\psi_z'^{-1}\psi_z \in H$ e

$$R(\psi_z^{-1}x, \psi_z^{-1}y) = R\big((\psi_z'^{-1}\psi_z)\psi_z^{-1}x, (\psi_z'^{-1}\psi_z)\psi_z^{-1}y\big) = R(\psi_z'^{-1}x, \psi_z'^{-1}y).$$

Temos que $\Phi^{-1}(R)$ é Γ-invariante, já que para $\gamma \in \Gamma$,

$$\Phi^{-1}(R)(\gamma x, \gamma y, \gamma z) = R(\psi_{\gamma z}^{-1}\gamma x, \psi_{\gamma z}^{-1}\gamma y)$$

e como R é H-invariante e $\psi_z^{-1}\gamma^{-1}\psi_{\gamma z} \in H$, temos

$$R(\psi_{\gamma z}^{-1}\gamma x, \psi_{\gamma z}^{-1}\gamma y) = R(\psi_z^{-1}x, \psi_z^{-1}y) = \Phi^{-1}(R)(x, y, z).$$

Provemos que $\Phi^{-1}(R)$ é contínua. Pela definição de continuidade, precisamos mostrar que para quaisquer pontos $x, y, z \in V$ e $\epsilon > 0$, existem vizinhanças $V_x, V_y, V_z \subset V$ de x, y e z tais que se $(a, b, c) \in V_x \times V_y \times V_z$, então

$$|\Phi^{-1}(R)(a, b, c) - \Phi^{-1}(R)(x, y, z)| < \epsilon.$$

Como R é contínua, existem vizinhanças $V_1, V_2 \subset V$ de $\psi_z^{-1}x$ e $\psi_z^{-1}y$ (fixemos algum $\psi_z \in \Gamma$ tal que $\psi_z e = z$) tais que se $(u, v) \in V_1 \times V_2$, então

$$|R(u, v) - R(\psi_z^{-1}x, \psi_z^{-1}y)| < \epsilon.$$

Como R é H-invariante, podemos substituir $V_1 \times V_2$ pelo aberto

$$U := \bigcup_{h \in H} hV_1 \times hV_2$$

de $V \times V$ que é H-invariante e também satisfaz que se $(u, v) \in U$, então $|R(u, v) - R(\psi_z^{-1}x, \psi_z^{-1}y)| < \epsilon$. Como a ação de Γ em V é contínua, existem vizinhanças $A \subset \Gamma$ de ψ_z e $V_x, V_y \subset V$ de x e y tais que se $\psi \in A$, $a \in V_x$ e $b \in V_y$, então $(\psi^{-1}a, \psi^{-1}b) \in U$; como U é H-invariante, podemos trocar A pelo também aberto $AH := \{\psi h : \psi \in A, h \in H\} \subset \Gamma$ sem alterar a conclusão anterior. Definindo $V_z := \{\psi e : \psi \in AH\} \subset V$, temos que $z \in V_z$ (pois $\psi_z \in A$) e que se $c \in V_z$, então $\psi_c \in AH$ (para qualquer $\psi_c \in \Gamma$ tal que $\psi_c e = c$). Logo obtemos que se $a \in V_x$, $b \in V_y$ e $c \in V_z$, então $(\psi_c^{-1}a, \psi_c^{-1}b) \in U$, o que por sua vez implica que $|R(\psi_c^{-1}a, \psi_c^{-1}b) - R(\psi_z^{-1}x, \psi_z^{-1}y)| < \epsilon$ e portanto $|\Phi^{-1}(R)(a, b, c) - \Phi^{-1}(R)(x, y, z)| < \epsilon$, conforme desejávamos. Para concluirmos que $\Phi^{-1}(R)$ é contínua, resta mostrar que V_z é um conjunto aberto de V, para isso, considere a aplicação $f\colon \Gamma/H \to V$ dada por

$$f(\psi H) = \psi e.$$

Não é difícil ver que f está bem definida, que é contínua e que é uma bijeção entre os dois espaços, ademais, como Γ é compacto, Γ/H também o é e como V é um espaço de Hausdorff, segue que f é um homeomorfismo entre os dois espaços (conforme o Teorema 5.E de Simmons [Sim63]). Finalmente, como V_z é a imagem de $[AH] := \{\psi H \in \Gamma/H : \psi \in AH\}$ por f e como AH ser aberto em Γ implica que $[AH]$ é aberto em Γ/H, concluímos que V_z é aberto em V.

Para ver que $\Phi^{-1}(R)$ pertence ao cone $\mathcal{C}(V \times V \times V)_{\succeq 0}$, note que se $n \in \mathbb{N}$ e $w, x_1, \ldots, x_n \in V$, então a matriz $\left(\Phi^{-1}(R)(x_i, x_j, w)\right)_{i,j=1}^{n}$ é igual à matriz $\left(R(\psi_w^{-1}x_i, \psi_w^{-1}x_j)\right)_{i,j=1}^{n}$ que é positivo-semidefinida pois $R \in \mathcal{C}(V \times V)_{\succeq 0}$.

Por fim, note que Φ^{-1} é de fato a função inversa de Φ, pois para $T \in \mathcal{C}(V \times V \times V)_{\succeq 0}^{\Gamma}$,

$$\Phi^{-1}(\Phi(T))(x, y, z) = \Phi(T)(\psi_z^{-1}x, \psi_z^{-1}y) = T(\psi_z^{-1}x, \psi_z^{-1}y, e) = T(x, y, z)$$

e para $R \in \mathcal{C}(V \times V)_{\succeq 0}^{H}$,

$$\Phi(\Phi^{-1}(R))(x, y) = \Phi^{-1}(R)(x, y, e) = R(\psi_e^{-1}x, \psi_e^{-1}y) = R(x, y). \qquad \square$$

5.5. O limitante de programação semidefinida revisto

Voltemos ao problema dos códigos esféricos. O grafo $G_{n,\theta}$ que modela o problema é um grafo topológico de empacotamento com conjunto de vértices $V = S^{n-1}$ e é homogêneo sob a ação de $\Gamma = \mathrm{O}(\mathbb{R}^n)$. Portanto, podemos usar o Programa (62) em conjunto com as observações feitas na última seção para obter um limitante para $A(n, \theta)$, que veremos ser muito semelhante ao limitante apresentado no Capítulo 3.

Já vimos como representar um kernel contínuo, positivo e $\mathrm{O}(\mathbb{R}^n)$-invariante $K \in \mathcal{C}(S^{n-1} \times S^{n-1})_{\succeq 0}^{\mathrm{O}(\mathbb{R}^n)}$ com o Teorema 2.4,

$$K(x, y) = \sum_{k=0}^{d} a_k P_k^n(x \cdot y), \tag{63}$$

com $a_k \geq 0$, sendo que aqui a série é apresentada truncada no d-ésimo termo, de modo a termos um número finito de variáveis, e usamos os polinômios de Gegenbauer P_k^n estudados na Seção 2.4.4.

De acordo com a Proposição 5.6, podemos representar uma função $T \in \mathcal{C}(S^{n-1} \times S^{n-1} \times S^{n-1})_{\succeq 0}^{\mathrm{O}(\mathbb{R}^n)}$ através de um kernel $\Phi(T) \in \mathcal{C}(S^{n-1} \times S^{n-1})_{\succeq 0}^{H}$, com H sendo o subgrupo estabilizador de um ponto $e \in S^{n-1}$ sob a ação de $\mathrm{O}(\mathbb{R}^n)$. Vimos na Seção 3.1 como representar um kernel positivo e H-invariante de $\mathrm{Pol}_{\leq d}(S^{n-1})^{(2)}$, que consiste no espaço gerado pelos kernels da forma $(x, y) \mapsto f_1(x)f_2(y)$, com $f_1, f_2 \in \mathrm{Pol}(S^{n-1})_{\leq d}$; segue do Teorema de Stone-Weierstrass (Teorema IV.9 em Reed e Simon [RS72]) que $\mathrm{Pol}(S^{n-1})^{(2)}$ é denso em $\mathcal{C}(S^{n-1} \times S^{n-1})$ e logo kernels de $\mathcal{C}(S^{n-1} \times S^{n-1})$ podem ser uniformemente aproximados por kernels de

$\mathrm{Pol}_{\leq d}(S^{n-1})^{(2)}$ com d suficientemente grande. Conforme descrito pelo Teorema 3.3, um kernel $\Phi(T) \in \mathrm{Pol}(S^{n-1})^{(2)}_{\leq d}$ é positivo e H-invariante se e somente se existem matrizes positivo-semidefinidas F_k de tamanho $(d-k+1) \times (d-k+1)$ para $k = 0, \ldots, d$ tais que

$$\Phi(T)(x,y) = \sum_{k=0}^{d} \langle F_k, Y_k^n(e \cdot x, e \cdot y, x \cdot y) \rangle \,, \tag{64}$$

com as matrizes Y_k^n dadas pela Proposição 3.2.

Para aplicar as expressões (63) e (64) ao Programa (62), devemos distinguir três casos, de acordo com o tamanho de $S \in I_3'$:

Caso $|S| = 1$ (denotemos $S = \{x\}$), temos:

$$A_1 K(S) = K(x,x) = \sum_{k=0}^{d} a_k P_k^n(1) = \sum_{k=0}^{d} a_k$$

e

$$\begin{aligned} A_{3\mathrm{PB}} T(S) = T(x,x,x) = \Phi(T)(\psi_x^{-1}x, \psi_x^{-1}x) &= \sum_{k=0}^{d} \langle F_k, Y_k^n(1,1,1) \rangle \\ &= \langle F_0, J_{d+1} \rangle \end{aligned}$$

(conforme observado após o Teorema 3.3).

Caso $|S| = 2$ (denotemos $S = \{x, y\}$, com $x \cdot y = u$), temos:

$$\begin{aligned} A_1 K(S) = K(x,y) + K(y,x) &= \sum_{k=0}^{d} a_k \big(P_k^n(x \cdot y) + P_k^n(y \cdot x) \big) \\ &= 2 \sum_{k=0}^{d} a_k P_k^n(u) \end{aligned}$$

e

$$\begin{aligned} A_{3\mathrm{PB}} T(S) &= T(x,y,x) + T(y,x,x) + T(y,y,x) \\ &\quad + T(x,y,y) + T(y,x,y) + T(x,x,y) \\ &= \Phi(T)(\psi_x^{-1}x, \psi_x^{-1}y) + \Phi(T)(\psi_x^{-1}y, \psi_x^{-1}x) + \Phi(T)(\psi_x^{-1}y, \psi_x^{-1}y) \\ &\quad + \Phi(T)(\psi_y^{-1}x, \psi_y^{-1}y) + \Phi(T)(\psi_y^{-1}y, \psi_y^{-1}x) + \Phi(T)(\psi_y^{-1}x, \psi_y^{-1}x) \\ &= \sum_{k=0}^{d} \Big\langle F_k, \big(Y_k^n(1,u,u) + Y_k^n(u,1,u) + Y_k^n(u,u,1) \\ &\qquad + Y_k^n(u,1,u) + Y_k^n(1,u,u) + Y_k^n(u,u,1) \big) \Big\rangle \\ &= 6 \sum_{k=0}^{d} \big\langle F_k, S_k^n(u,u,1) \big\rangle. \end{aligned}$$

Caso $|S| = 3$ (denotemos $S = \{x, y, z\}$, com $z \cdot x = u$, $z \cdot y = v$ e $x \cdot y = t$), temos:

$$A_1 K(S) = 0$$

e

$$\begin{aligned}
A_{3\mathrm{PB}}T(S) &= T(x,y,z) + T(y,x,z) + T(x,z,y) \\
&\quad + T(z,x,y) + T(y,z,x) + T(z,y,x) \\
&= \Phi(T)(\psi_z^{-1}x, \psi_z^{-1}y) + \Phi(T)(\psi_z^{-1}y, \psi_z^{-1}x) + \Phi(T)(\psi_y^{-1}x, \psi_y^{-1}z) \\
&\quad + \Phi(T)(\psi_y^{-1}z, \psi_y^{-1}x) + \Phi(T)(\psi_x^{-1}y, \psi_x^{-1}z) + \Phi(T)(\psi_x^{-1}z, \psi_x^{-1}y) \\
&= \sum_{k=0}^{d} \Big\langle F_k, \big(Y_k^n(u,v,t) + Y_k^n(v,u,t) + Y_k^n(t,v,u) \\
&\qquad + Y_k^n(v,t,u) + Y_k^n(t,u,v) + Y_k^n(u,t,v)\big) \Big\rangle \\
&= 6 \sum_{k=0}^{d} \big\langle F_k, S_k^n(u,v,t) \big\rangle.
\end{aligned}$$

Substituindo essas expressões no Programa (62) e substituindo $a-1$ pela expressão que aparece quando $|S| = 1$, obtemos quase exatamente o Programa (44), como formulado por Bachoc e Vallentin [BV08] (a única variável faltando é a variável B, que vem de uma restrição extra no programa primal).

Conclusão

Neste trabalho são calculados novos limitantes para o número de contato nas dimensões 9 a 23. Para isso é usado o limitante de programação semidefinida proposto por Bachoc e Vallentin [BV08] e as melhorias são possíveis através da aplicação de técnicas de otimização polinomial com polinômios invariantes introduzidas por Gatermann e Parrilo [GP04], que permitem a redução do tamanho dos programas semidefinidos considerados.

Também é estabelecida uma relação entre o limitante de programação semidefinida para códigos esféricos e um limitante para o número de independência de grafos finitos que se baseia em submatrizes principais da matriz de momentos com entradas relativas a conjuntos de até 3 elementos. Essa conexão é possível através de uma adaptação de uma técnica usada por de Laat e Vallentin [dLV15] que permite estender tais limitantes para grafos topológicos de empacotamento compactos.

Essa extensão é interessante não só por proporcionar uma melhor compreensão do limitante estudado neste trabalho, como também por sugerir formas de fortalecê-lo. Diversos limitantes podem ser formulados em termos da matriz de momentos, em de Laat e Vallentin [dLV15] a hierarquia de Lasserre [Las02, Lau03] é estendida para grafos topológicos de empacotamento compactos. O principal desafio para o cálculo desses limitantes é a resolução dos problemas de otimização polinomial resultantes, que passam a envolver polinômios com mais variáveis e resultar em programas semidefinidos muito grandes.

Referências Bibliográficas

[AAR99] George E. Andrews, Richard Askey e Ranjan Roy. *Special functions*, volume 71 de *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1999. 37, 43

[ABR01] Sheldon Axler, Paul Bourdon e Wade Ramey. *Harmonic function theory*, volume 137 de *Graduate Texts in Mathematics*. Springer-Verlag, New York, segunda edição, 2001. 37, 44

[AT07] Charalambos D. Aliprantis e Rabee Tourky. *Cones and duality*, volume 84 de *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2007. 24

[AW01] George B. Arfken e Hans J. Weber. *Mathematical methods for physicists*. Harcourt/Academic Press, Burlington, MA, quinta edição, 2001. 39

[Bar02] Alexander Barvinok. *A course in convexity*, volume 54 de *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002. 24, 28

[BCR84] Christian Berg, Jens P. R. Christensen e Paul Ressel. *Harmonic analysis on semigroups*, volume 100 de *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1984. Theory of positive definite and related functions. 71

[BGSV12] Christine Bachoc, Dion C. Gijswijt, Alexander Schrijver e Frank Vallentin. Invariant semidefinite programs. Em *Handbook on semidefinite, conic and polynomial optimization*, volume 166 de *Internat. Ser. Oper. Res. Management Sci.*, páginas 219–269. Springer, New York, 2012. 20, 60, 70

[BNdOFV09] Christine Bachoc, Gabriele Nebe, Fernando M. de Oliveira Filho e Frank Vallentin. Lower bounds for measurable chromatic numbers. *Geom. Funct. Anal.*, 19(3):645–661, 2009. 31

[Boc41] Salomon Bochner. Hilbert distances and positive definite functions. *Ann. of Math. (2)*, 42:647–656, 1941. 20

[BS81] Eiichi Bannai e Neil J. A. Sloane. Uniqueness of certain spherical codes. *Canad. J. Math.*, 33(2):437–449, 1981. 36, 64

[BV08] Christine Bachoc e Frank Vallentin. New upper bounds for kissing numbers from semidefinite programming. *J. Amer. Math. Soc.*, 21(3):909–924, 2008. 2, 47, 52, 64, 78, 79

[Con90] John B. Conway. *A course in functional analysis*, volume 96 de *Graduate Texts in Mathematics*. Springer-Verlag, New York, segunda edição, 1990. 5

[Cox62] Harold S. M. Coxeter. The problem of packing a number of equal nonoverlapping circles on a sphere. *Transactions of the New York Academy of Sciences*, 24(3 Series II):320–331, 1962. 2

[CS99] John H. Conway e Neil J. A. Sloane. *Sphere packings, lattices and groups*, volume 290 de *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*.

Springer-Verlag, New York, terceira edição, 1999. Com contribuições adicionais de E. Bannai, R. E. Borcherds, J. Leech, S. P. Norton, A. M. Odlyzko, R. A. Parker, L. Queen e B. B. Venkov. 1, 2, 3, 35, 36, 64

[DGdOFV17] Maria Dostert, Cristóbal Guzmán, Fernando M. de Oliveira Filho e Frank Vallentin. New upper bounds for the density of translative packings of three-dimensional convex bodies with tetrahedral symmetry. Aceito em *Discrete Comput. Geom.*, 2017. arXiv:1510.02331. 65

[DGS77] Philippe Delsarte, Jean-Marie Goethals e Johan J. Seidel. Spherical codes and designs. *Geometriae Dedicata*, 6(3):363–388, 1977. 2, 31, 34

[dLV15] David de Laat e Frank Vallentin. A semidefinite programming hierarchy for packing problems in discrete geometry. *Math. Program.*, 151(2, Ser. B):529–553, 2015. 67, 68, 72, 79

[FH91] William Fulton e Joe Harris. *Representation theory*, volume 129 de *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991. A first course, Readings in Mathematics. 5

[Fol16] Gerald B. Folland. *A course in abstract harmonic analysis*. Textbooks in Mathematics. CRC Press, Boca Raton, FL, segunda edição, 2016. 9, 12, 20

[GP04] Karin Gatermann e Pablo A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *J. Pure Appl. Algebra*, 192(1-3):95–128, 2004. 60, 79

[Han00] David Handel. Some homotopy properties of spaces of finite subsets of topological spaces. *Houston J. Math.*, 26(4):747–764, 2000. 68, 69

[Hil88] David Hilbert. Über die Darstellung definiter Formen als Summe von Formenquadraten. *Math. Ann.*, 32(3):342–350, 1888. 57

[KKLS16] Rob Kusner, Wöden Kusner, Jeffrey C. Lagarias e Senya Shlosman. The twelve spheres problem. *arXiv preprint*, 2016. arXiv:1611.10297. 2

[KL78] Grigorii A. Kabatiansky e Vladimir I. Levenshtein. Bounds for packings on the sphere and in space. *Problemy Peredachi Informacii*, 14(1):3–25, 1978. 3

[Knu94] Donald E. Knuth. The sandwich theorem. *Electron. J. Combin.*, 1:Article 1, approx. 48 pp. (electronic), 1994. 31

[Las02] Jean B. Lasserre. An explicit equivalent positive semidefinite program for nonlinear 0-1 programs. *SIAM J. Optim.*, 12(3):756–769, 2002. 70, 79

[Lau03] Monique Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.*, 28(3):470–496, 2003. 70, 79

[Lev79] Vladimir I. Levenshtein. Boundaries for packings in n-dimensional Euclidean space. *Dokl. Akad. Nauk SSSR*, 245(6):1299–1303, 1979. 2

[Lov79] László Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979. 31

[LS91] László Lovász e Alexander Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. Optim.*, 1(2):166–190, 1991. 70

[MdOF17] Fabrício C. Machado e Fernando M. de Oliveira Filho. Improving the semidefinite programming bound for the kissing number by exploiting polynomial symmetry. Aceito em *Experiment. Math.*, 2017. arXiv:1609.05167. 2, 3

[Meg98] Robert E. Megginson. *An introduction to Banach space theory*, volume 183 de *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998. 72

[Mus08] Oleg R. Musin. The kissing number in four dimensions. *Ann. of Math. (2)*, 168(1):1–32, 2008. 2, 35, 37

[MV10] Hans D. Mittelmann e Frank Vallentin. High-accuracy semidefinite programming bounds for kissing numbers. *Experiment. Math.*, 19(2):175–179, 2010. 2, 59, 63, 64

[Nak10] Maho Nakata. A numerical evaluation of highly accurate multiple-precision arithmetic version of semidefinite programming solver: SDPA-GMP,-QD and-DD. Em *Computer-Aided Control System Design (CACSD), 2010 IEEE International Symposium on*, páginas 29–34. IEEE, 2010. 59, 63, 66

[OS79] Andrew M. Odlyzko e Neil J. A. Sloane. New bounds on the number of unit spheres that can touch a unit sphere in n dimensions. *J. Combin. Theory Ser. A*, 26(2):210–214, 1979. 2, 35, 36, 59

[Put93] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.*, 42(3):969–984, 1993. 58

[PZ04] Florian Pfender e Günter M. Ziegler. Kissing numbers, sphere packings, and some unexpected proofs. *Notices Amer. Math. Soc.*, 51(8):873–883, 2004. 2

[RR05] Nathalie Revol e Fabrice Rouillier. Motivations for an arbitrary precision interval arithmetic and the MPFI library. *Reliab. Comput.*, 11(4):275–290, 2005. 65

[RS72] Michael Reed e Barry Simon. *Methods of modern mathematical physics. I. Functional analysis.* Academic Press, New York-London, 1972. 5, 8, 20, 23, 44, 76

[Rud87] Walter Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, terceira edição, 1987. 20

[Sch42] Isaac J. Schoenberg. Positive definite functions on spheres. *Duke Math. J.*, 9:96–108, 1942. 34, 44

[Sch79] Alexander Schrijver. A comparison of the Delsarte and Lovász bounds. *IEEE Trans. Inform. Theory*, 25(4):425–429, 1979. 31

[Ser77] Jean-Pierre Serre. *Linear representations of finite groups.* Springer-Verlag, New York-Heidelberg, 1977. Traduzido da segunda edição francesa por Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42. 5, 12

[Sim63] George F. Simmons. *Introduction to topology and modern analysis.* McGraw-Hill Book Co., Inc., New York-San Francisco, Calif.-Toronto-London, 1963. 76

[Slo81] Neil J. A. Sloane. Tables of sphere packings and spherical codes. *IEEE Trans. Inform. Theory*, 27(3):327–338, 1981. 1

[Stu08] Bernd Sturmfels. *Algorithms in invariant theory.* Texts and Monographs in Symbolic Computation. Springer Wien New York, Vienna, segunda edição, 2008. 63

[SvdW53] Kurt Schütte e Bartel L. van der Waerden. Das Problem der dreizehn Kugeln. *Math. Ann.*, 125:325–334, 1953. 2, 35

[Sze39] Gabor Szegö. *Orthogonal Polynomials*. American Mathematical Society, New York, 1939. American Mathematical Society Colloquium Publications, v. 23. 43, 44

[Vil68] Neil Ja. Vilenkin. *Special functions and the theory of group representations*. Traduzido do russo por V. N. Singh. Translations of Mathematical Monographs, Vol. 22. American Mathematical Society, Providence, R. I., 1968. 5, 6, 12, 18, 37

[Wyn65] Aaron D. Wyner. Capabilities of bounded discrepancy decoding. *Bell Systems Tech. J.*, 44:1061–1122, 1965. 3

[ZE99] Victor A. Zinov'ev e Thomas Ericson. New lower bounds for contact numbers in small dimensions. *Problemy Peredachi Informatsii*, 35(4):3–11, 1999. 35, 64