

Sumário

Sumário	1
12/3/2013	2
1 Programação Linear	2
1.1 Introdução	2
1.2 Formulação Geral	3
1.3 Programação inteira	7
12/3/2013	8
2/4/2013	8
1.4 Forma equacional	8
1.5 Solução Viável Básica	9
1.6 Convexidade e poliedros convexos	10
9/4/2013	10
2 Programação Semidefinida	10
2.1 Revisão de Álgebra Linear	10
9/4/2013	13
2.2 Dualidade em programação linear	14
2.3 Polítopos de conjuntos estáveis	15
31/4/2013	15
2.4 Caracterização da equivalência	16
7/5/2013	16
2.5 Programação Semidefinida	16
21/5/2013	17
2.6 Capacidade de Shannon	17
28/5/2013	20
4/6/2013	21
11/6/2013	22
2.7 Polítopo $\text{STAB}(G)$	23
18/6/2013	24
3 Conjectura de Erdős–Simonovits–Sidorenko	24
3.1 Espaços mensuráveis	25
3.2 Desigualdade de Jensen	26
3.3 Desigualdade log-soma	26
Referências Bibliográficas	27

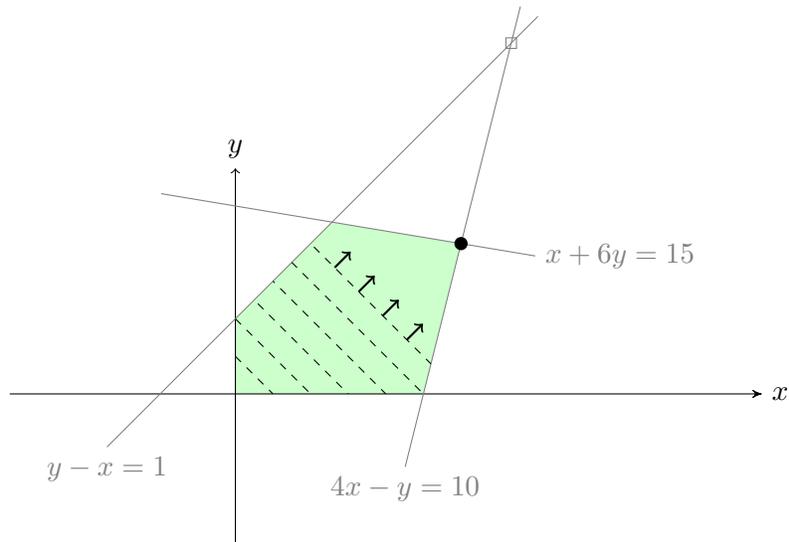


Figura 1: Representação gráfica de um PL com duas variáveis. A região colorida representa o conjunto de *soluções viáveis*, as linhas tracejadas representam (cada uma) um conjunto de pontos para o qual a *função objetivo* assume valor igual a uma constante $x + y = c$ (“cotas” ou “curvas de nível”). O ponto marca a *solução ótima*. O quadrado assinala a solução ótima se a restrição $x + 6y \leq 15$ é removida.

2013/3/12

1 Programação Linear

1.1 Introdução

Introduzimos a seguir o assunto “programação linear”. Uma referência é [MG07].

Programação linear (PL) consiste no estudo da otimização de uma função linear, sujeita a restrições lineares. Uma função nas variáveis x_1, x_2, \dots, x_n é *linear* se pode ser escrita como $\sum_{i=1}^n \alpha_i x_i$, com $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. Uma inequação nas variáveis x_1, x_2, \dots, x_n é linear se pode ser escrita como $\sum_{i=1}^n \beta_i x_i \leq c$, com $\beta_1, \dots, \beta_n \in \mathbb{R}$.

Exemplo (*Formulação de um PL*)

O seguinte PL, com variáveis x e y , está representado graficamente na figura 1.

$$\begin{aligned} \max \quad & x + y \\ \text{s.a.} \quad & -x + y \leq 1 \\ & x + 6y \leq 15 \\ & 4x - y \leq 10 \\ & x, y \geq 0 \end{aligned}$$

Note que, se removemos a restrição $4x - y \geq 10$, então, (exercício!) para todo ponto (x, y) satisfazendo as restrições do PL, existe um ponto (x', y') que também satisfaz as restrições, e tal que $x' + y' \geq x + y$.

	cenoura	repolho	pepino	mínimo a incluir (mg)
Vitamina A (mg)	35	0.5	0.5	0.5
Vitamina C (mg)	60	300	10	15
Fibras (mg)	30	20	10	4
Preço (\$/Kg)	0.75	0.5	0.15	

Tabela 1: Restrições de consumo para uma dieta “boa”: consumir, entre outras coisas, ao menos 4 miligramas de fibras, e 300 miligramas de vitamina C. Os preços afetam a função-objetivo que queremos minimizar: o custo.

1.2 Formulação Geral

Para $a, b \in \mathbb{R}^n$, com $a = (a_1, \dots, a_n)$ e $b = (b_1, \dots, b_n)$, convencionamos que $a \leq b$ quando $a_i \leq b_i$, para $i = 1, \dots, n$. Assim, podemos (exercício!) escrever qualquer PL como

$$\begin{aligned} \max \quad & c^T x \\ \text{s.a.} \quad & Ax \leq b \end{aligned}$$

Onde $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, o vetor c^T representa c transposto, e A é uma matriz real $m \times n$.

Nessa formulação, chamamos $c^T x$ de **função objetivo**, e o sistema $Ax \leq b$ de **restrições**. Um vetor $x' \in \mathbb{R}^n$ é chamado **solução viável** se $Ax' \leq b$. Uma solução viável $x^* \in \mathbb{R}^n$ é **solução ótima** se $c^T x^*$ é máximo. Um PL pode não ter, ter exatamente uma, ou ter infinitas soluções ótimas (exercício!).

Observação 1.1 As soluções de um PL formam um poliedro convexo.

Nem sempre é simples identificar que um problema tem formulação como PL. Veremos exemplos.

Exemplo (Dieta Ótima)

Desejamos encontrar uma dieta de custo mínimo, que seja capaz de satisfazer certas necessidades básicas do consumo diário das vitaminas A e C, e de fibras. O preço dos itens disponíveis para composição da dieta, assim como seus custos, estão descritos na tabela 1.

Se representamos por x_{cen} , x_{rep} e x_{pep} , respectivamente, a quantidade a consumir de cenoura, repolho e pepino, obtemos a seguinte formulação.

$$\begin{aligned} \min \quad & 0.75x_{\text{cen}} + 0.5x_{\text{rep}} + 0.15x_{\text{pep}} \\ \text{s.a.} \quad & 35x_{\text{cen}} + 0.5x_{\text{rep}} + 0.5x_{\text{pep}} \geq 0.5 \\ & 60x_{\text{cen}} + 300x_{\text{rep}} + 10x_{\text{pep}} \geq 15 \\ & 30x_{\text{cen}} + 20x_{\text{rep}} + 10x_{\text{pep}} \geq 4 \end{aligned}$$

Exemplo (Fluxo em Rede)

Temos uma rede, digamos, de computadores, conectados entre si como mostra a figura 2. As conexões têm limitações: cada uma possui uma taxa máxima de transmissão de dados. Queremos saber qual a máxima taxa possível de O a N.

Restrições: as conexões podem ser usadas em qualquer sentido, mas não em ambos os sentidos simultaneamente. Ademais, a quantidade de dados que chega a qualquer dos nós A,B,C,D,E deve ser a mesma que dele sai. Observe (exercício!) que isso significa que a quantidade que sai de O é *igual* à que chega a N.

Note que a taxa máxima não excede a soma das taxas das conexões com extremo em O — donde é menor ou igual a 6. Ademais, a quantidade que passa pelo nó A é no máximo 2, pelo modo como as capacidades das conexões O–A, A–B e A–D estão distribuídas. Observe (exercício!) que é de fato possível transmitir informação à taxa “5”; como mais não é possível, essa taxa é máxima.

Observe ainda que, em geral, se consideramos uma partição A, B do conjunto $\{O, A, B, C, D, E, N\}$ tal que O e N estejam em partições diferentes, então a soma da taxa de transmissão das arestas que conectam A a B (mais precisamente, arestas que conectam $a \in A$ a $b \in B$) é um limitante superior para a taxa máxima de transmissão de O a N. É um teorema (teorema fluxo máximo–corte mínimo) que o menor valor atingido por esse limitante é de fato a taxa máxima que pode ser transmitida.

Indicando por x_{uv} a taxa de transmissão de u a v , para algum esquema de transmissão (de modo que $x_{uv} = -x_{vu}$), temos a seguinte formulação.

$$\begin{aligned}
 \max \quad & x_{oa} + x_{ob} + x_{oc} \\
 \text{s.a.} \quad & x_{oa} - x_{ab} - x_{ad} = 0 \\
 & x_{ab} + x_{ob} - x_{be} = 0 \\
 & x_{oc} + x_{ec} - x_{cd} = 0 \\
 & x_{ad} + x_{cd} - x_{dn} = 0 \\
 & x_{be} - x_{ec} - x_{en} = 0 \\
 & -3 \leq x_{oa} \leq 3 \\
 & -1 \leq x_{ob} \leq 1 \\
 & -2 \leq x_{oc} \leq 2 \\
 & -1 \leq x_{ab} \leq 1 \\
 & -1 \leq x_{ad} \leq 1 \\
 & -2 \leq x_{be} \leq 2 \\
 & -4 \leq x_{cd} \leq 4 \\
 & -4 \leq x_{dn} \leq 4 \\
 & -4 \leq x_{ec} \leq 4 \\
 & -1 \leq x_{en} \leq 1
 \end{aligned}$$

Exemplo (Planejamento da sorveteria)

Uma fábrica de sorvete deseja planejar sua produção mensal de sorvete, de modo a maximizar seus ganhos. Basearemos o planejamento em um estudo de mercado prévio, que nos dá estimativas da demanda mês a mês (em toneladas). A produção em excesso de um mês fica estocada para venda nos meses subsequentes. Ademais, há os custos de operação da fábrica: um custo de 50 \$/(tonelada · mês) pago com base na mudança na produção: se no mês i a fábrica produzia x_i toneladas de sorvete, e no mês $i + 1$ passará a produzir x_{i+1} , o custo associado à alteração é $50|x_{i+1} - x_i|$; e um custo de 20 \$/(tonelada · mês) que é pago pelo sorvete que é estocado.

Aqui denotamos por x_i a quantidade de sorvete que foi produzida no mês i e por s_i a quantidade deixada de estoque ao fim do mês i , com $i = 1, 2, \dots, 12$. Definimos $x_0 = s_0 = 0$, para escrever a formulação de PL a seguir.

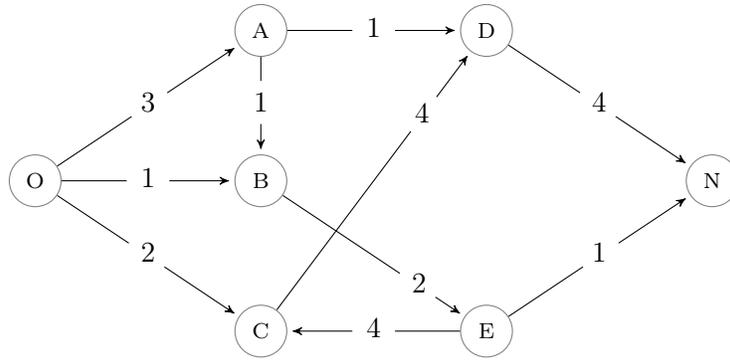


Figura 2: Rede com capacidade nas conexões. Conexões entre nós (O, A, por exemplo) são indicadas por linhas. O fluxo máximo que pode passar em uma ligação é indicado pelos números. Cada conexão pode transmitir em qualquer sentido, mas em apenas um. Qual a maior taxa de envio possível de O para A?

$$\begin{aligned}
 \min \quad & \langle \text{mudança de produção} \rangle + \langle \text{estoque} \rangle \\
 = \min \quad & 50 \sum_{i=1}^{12} |x_i - x_{i-1}| + s_i \\
 \text{s.a.} \quad & x_0 = s_0 = s_{12} = 0 \\
 & x_i + s_{i-1} - d_i = s_i, \quad i = 1, 2, \dots, 12 \\
 & x_i, s_i \geq 0
 \end{aligned}$$

A presença do módulo na função objetivo torna a formulação não-linear. Podemos, contudo, contornar esse empecilho usando uma variável que mede a *variação da produção*. Definimos, assim, as variáveis

$$y_i = \begin{cases} x_i - x_{i-1} & \text{se } x_i - x_{i-1} \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

$$z_i = \begin{cases} -(x_i - x_{i-1}) & \text{se } x_i - x_{i-1} < 0 \\ 0 & \text{caso contrário} \end{cases}$$

Neste caso, o PL é

$$\begin{aligned}
 \min \quad & 50 \sum_{i=1}^{12} (y_i + z_i) + 20 \sum_{i=1}^{12} s_i \\
 \text{s.a.} \quad & x_0 = s_0 = s_{12} = 0, \\
 & x_i + s_{i-1} - d_i = s_i, \\
 & x_i - x_{i-1} = y_i - z_i, \\
 & s_i, x_i, y_i, z_i \geq 0 \quad i = 1, 2, \dots, 12
 \end{aligned}$$

Note que se existe solução viável x em que ambos y_i e z_i são positivos, então há solução viável x' em que todas as restrições são atendidas, e ademais a função objetivo assume valor $c^T x' < c^T x$ (exercício!).

Exemplo (Aproximação por Reta)

Suponha que desejamos encontrar a reta que melhor aproxima um conjunto de pontos $P = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$. A resposta depende do modo pelo qual avaliemos “quão boa” é uma reta candidata a aproximação. Para manter a função objetivo linear, usaremos o seguinte critério: queremos a reta $r = ax + b$ que minimize a soma $\sum_i^N |ax_i + b - y_i| = |e_i|$ das distâncias (“erros”) verticais entre os pontos e r .

Usamos outra estratégia, que dispensa o uso do módulo “ $|\cdot|$ ”, observando que minimizar $|z|$ é minimizar o máximo $\max(z, -z)$. Assim, chegamos à formulação abaixo.

$$\begin{aligned} \min \quad & \sum_{i=1}^N e_i \\ \text{s.a.} \quad & ax_i + b - y_i \leq e_i \\ & -(ax_i + b - y_i) \leq e_i. \end{aligned}$$

Exemplo (Separar Pontos)

Neste exemplo, temos um conjunto de pontos “pretos” e “brancos”, que queremos separar usando uma reta r — de modo que pontos de mesma cor estejam todos de um mesmo lado de r .

Sejam p_1, \dots, p_n os pontos brancos e q_1, \dots, q_n os pontos pretos. Gostaríamos de descrever a reta como uma função linear $r = ax + b$.

Dividimos a solução em dois casos:

1. a reta r é vertical ($a = \infty$);
2. a reta r não é vertical ($a \in \mathbb{R}$), os pontos brancos estão “acima” e os pontos pretos “abaixo” da reta; e
3. a reta r não é vertical ($a \in \mathbb{R}$), os pontos pretos estão “acima” e os pontos brancos “abaixo” da reta.

É simples, dada a coleção de pontos, identificar qual o caso que se aplica. No primeiro caso, toda reta vertical entre os pontos horizontalmente mais próximos das coleções é uma reta divisora.

A seguir, indicamos por $x(P)$ a coordenada x do ponto P .

Tratamos agora o caso dois (o caso 3 segue por analogia). Primeiro, escolhemos a medida segundo a qual comparamos duas retas divisoras, a função objetivo que queremos otimizar. Tomamos o somatório da distância vertical entre cada ponto e a reta divisória

$$\begin{aligned} \min \quad & \sum_{i=1}^n (y(p_i) - a \cdot x(p_i) - b) + (a \cdot x(q_i) + b - y(q_i)). \\ \text{s.a.} \quad & a \cdot x(p_i) + b < y(p_i) \\ & a \cdot x(p_i) + b > y(q_i) \end{aligned}$$

(Uma alternativa seria minimizar o somatório das distâncias entre cada ponto e r , mas isso leva a uma função objetivo quadrática. A técnica usada nesse caso é *método dos mínimos quadrados*.)

Aqui encontramos uma dificuldade. É preciso eliminar desigualdades estritas, caso contrário (exercício!), dada qualquer solução viável, podemos nos aproximar arbitrariamente da solução ótima sem, contudo, nunca alcançá-la.

Para contornar o problema, adicionamos uma distância δ até a reta, e a formulação torna-se

$$\begin{aligned} \max \quad & \delta \\ \text{s.a.} \quad & y(p_i) \geq a \cdot x(p_i) + b \\ & y(q_i) \leq a \cdot x(q_i) + b \end{aligned}$$

comprimento do rolo	quantidade
135 cm	97
108 cm	61
93 cm	31
92 cm	21

Tabela 2: Pedido de faixas de papel.

corte	perda	corte	perda	corte	perda
108 + 93 + 93	6	93 + 93 + 93	31	135 + 108	57
108 + 93 + 92	7	93 + 93 + 92	32	135 + 93	72
108 + 92 + 92	8	93 + 92 + 92	33	135 + 92	73
135 + 135	30	92 + 92 + 92	34	108 + 108	84

Tabela 3: Modos de corte possíveis para um rolo com 300 cm. Na tabela, por exemplo, 135 + 93 significa que após o corte temos uma faixa de 135 cm e outra de 93 cm.

e, se a solução ótima apresenta $\delta > 0$, temos a reta desejada.

Exemplo (Rolos de Papel)

Uma indústria possui uma máquina que produz rolos de papel de comprimento 3 m cada. Ela recebe o pedido descrito na tabela 2, em que diversos rolos (de comprimento menor do que 3 metros cada um) foram pedidos. A pergunta é como cortar os rolos de 3 metros de modo a minimizar a quantidade de rolos necessários para atender ao pedido?

Para resolver o problema, consideramos variáveis associadas aos *modos de corte* possíveis de um rolo de 3 m. Cada modo de corte produz uma certa quantidade de pedaços de cada tipo. Assim, por exemplo, um dos modos de corte possíveis produz 3 pedaços de 93 cm (e desperdiça um trecho de 31 cm); outro produz 2 pedaços: um de 135 cm e outro de 108 cm (com um desperdício de 57 cm).

Se contamos o número de vezes que usamos cada modo de corte, sabemos a quantidade de rolos consumida, e, ademais, a quantidade de pedaços de cada tipo que produzimos (por exemplo definimos as variáveis $x_{93,93,93}$ como o número de vezes que dividimos um rolo em três pedaços de 93 cm; uma outra variável poderia ser $x_{135,108}$).

As soluções deste último problema têm uma restrição a mais: o número de vezes que usamos cada corte deve ser um número inteiro. Note que, a princípio, um PL não tem necessariamente solução ótima inteira, donde é possível que essa restrição force um resultado diferente. Contudo, a solução ótima do problema “inteiro” não pode ser melhor (em termos de minimizar ou maximizar) a função objetivo que a versão *relaxada* (em que as variáveis podem assumir valores reais) (exercício!).

É em geral muito mais complicado computacionalmente resolver um problema com restrição aos inteiros. Esses problemas são estudo da *programação inteira*, ou PI.

Assim, dado um problema inteiro, podemos tentar aproximar sua solução usando sua versão relaxada. Veremos que, essa pode por vezes ser uma boa ideia, mas nem sempre.

1.3 Programação inteira

A solução inteira não pode ser melhor do que o PL geral, mas é possível que o seja aceitável aproximar o PL inteiro por sua versão real (também chamada de *relaxação* do problema).

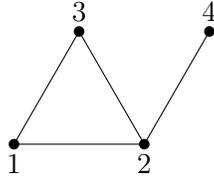


Figura 3: Exemplo de grafo. Os pontos representam os vértices e as linhas conectando os pontos são arestas.

No exemplo sobre rolos de papel, a perda por arredondamento é de menos de 0.25%.

2013/3/12

Veremos exemplos dos variados tipos de comportamento que observamos ao aproximar um PL por sua relaxação.

Grafos

Definição 1.2 (Grafo). Um **grafo** é um par ordenado $G = (V, E)$, onde V é dito conjunto de **vértices** de G , e $E \subseteq \{\{u, v\} : u, v \in V\}$ contém as **arestas** de G . Se $\{u, v\} \in E$, dizemos que u e v estão **ligados**, **conectados** ou que são **vizinhos**

Exemplo (Grafo)

É comum representar grafos por meio de desenhos, onde vértices são pontos ligados por linhas — as arestas que os conectam. Um exemplo de tal representação pode ser visto na figura 3, que exhibe o grafo $G = (V, E)$ em que $V = \{1, 2, 3, 4\}$ e $E = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{2, 4\}\}$.

Definição 1.3 (Grafo bipartido). Um grafo $G = (A \cup B, E)$ é dito A, B -bipartido se o conjunto de vértices pode ser particionado em classes A e B , com $A \cap B = \emptyset$, de modo que toda aresta $e \in E$ tenha exatamente um vértice em cada classe.

Definição 1.4 (Emparelhamento). Um **emparelhamento** M é um subconjunto das arestas de um grafo tal que dois elementos disjuntos de M tem interseção vazia. Um emparelhamento é perfeito se todo vértice do grafo pertence a alguma aresta do emparelhamento.

Exemplo (Emparelhamento perfeito de peso máximo)

Para grafos bipartidos (ou seja sem circuitos de comprimento ímpar), a solução ótima do PI, quando existe, coincide com a do PL. Isso não acontece quando existem circuitos de comprimento ímpar.

Exemplo (Cobertura por vértices mínima)

Neste caso obtemos uma 2-aproximação.

Exemplo (Conjunto independente máximo)

A solução ótima encontrada na relaxação pode ser pior do que qualquer constante.

2013/4/2

1.4 Forma equacional

A **forma equacional** ou **forma padrão** de um PL é a seguinte.

$$\begin{aligned} \max \quad & c^T x \\ \text{s.a.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

onde $A \in M(m \times n)$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ e $0 \in \mathbb{R}^n$. A restrição $x \geq 0$ é chamada condição de **não-negatividade**. Podemos sempre formular um PL na forma equacional.

Exemplo (Rescrever PL na forma equacional)

$$\begin{aligned} \max \quad & 3x_1 - 2x_2 \\ \text{s.a.} \quad & 2x_1 - x_2 \leq 4 \\ & x_1 + 3x_2 \geq 5 \\ & x_2 \geq 0 \end{aligned}$$

Primeiro eliminamos a desigualdade “ $2x_1 - x_2 \leq 4$ ”, acrescentando a variável de folga x_3 .

$$\begin{aligned} 2x_1 - x_2 + x_3 &= 4 \\ x_3 &\geq 0 \end{aligned}$$

Aplicamos a mesma técnica à restrição “ $x_1 + 3x_2 \geq 5$ ”.

$$\begin{aligned} -x_1 - 3x_2 + x_4 &= -5 \\ x_4 &\geq 0 \end{aligned}$$

Finalmente, forçamos a condição de não-negatividade, substituindo a variável x_1 pela diferença entre suas partes positiva e negativa.

$$\begin{aligned} x_1 &= y_1 - z_1 \\ y_1, z_1 &\geq 0 \end{aligned}$$

Partindo de um PL com m equações em n incógnitas, e aplicando o método exibido acima, terminamos com uma formulação que tem até $2n + m$ variáveis e m equações (sem contar as condições de não-negatividade). Nesse procedimento, assumimos que

1. $Ax = b$ tem ao menos uma solução, e
2. as linhas de A são linearmente independentes (l.i.).

1.5 Solução Viável Básica

Seja $A \in M(m \times n)$, com $n \geq m$ uma matriz de posto m . (Posto é o número de linhas l.i. da matriz.) Seja ainda $B \subseteq \{1, \dots, n\}$, e A_B a matriz das colunas de A com índice em B .

Exemplo (Notação A_B)

$$A = \begin{pmatrix} 1 & 5 & 3 \\ 0 & 1 & 3 \end{pmatrix}, \quad B = \{1, 3\} \quad \implies \quad A_B = \begin{pmatrix} 1 & 3 \\ 0 & 3 \end{pmatrix}$$

Uma solução viável $x \in \mathbb{R}^n$ de um PL na forma equacional é dita **solução viável básica** se existe $B \subseteq \{1, \dots, n\}$ satisfazendo $|B| = m$ (número de restrições) e tal que

- A_B é não-singular (*i.e.*, as colunas são l.i.), e
- $x_j = 0$ se $j \notin B$.

Exemplo (Uma solução viável básica)

$$A = \begin{pmatrix} 1 & 5 & 3 & 4 & 6 \\ 0 & 1 & 3 & 5 & 6 \end{pmatrix}, \quad b = \{14, 7\}, \quad Ax = b, \quad x \geq 0, \quad B = \{2, 4\}.$$

Temos o sistema $A_B x_B = b$,

$$5x_2 + 4x_4 = 14, \quad x_2 + 5x_4 = 7,$$

de solução $x_4 = 1, x_2 = 2$. Assim, $x = (0, 2, 0, 1, 0)$ é básica (e viável, pois suas entradas são não-negativas).

Definição 1.5 (Variável básica). Se $x = (x_i)$ é solução viável básica (SVB) então x_j tal que $j \in B$ é **variável básica**, e x_k tal que $k \notin B$ é **variável não-básica**.

Lema 1.6 Seja x SVB de um PL na forma equacional. Então x é básica se e somente se as colunas de A_K são l.i., onde $k = \{j \in \{1, \dots, n\} : x_j \geq 0\}$.

1.6 Convexidade e poliedros convexos

Falamos da interpretação geométrica das restrições de um PL; de poliedros, fecho convexo e vértices.

2013/4/9

2 Programação Semidefinida

Programação semidefinida estuda a otimização de uma função objetivo linear sujeita a restrições que podem ser escritas por meio de matrizes positivas semidefinidas; uma referência para o assunto é [Lov95].

2.1 Revisão de Álgebra Linear

Tópicos: definição de autovalor e traço; propriedades de matrizes simétricas; definição de menor simétrico; teorema do entrelaçamento de autovalores; definição de matriz semidefinida positiva. Definição de produto interno e cone; enunciado de uma versão do teorema de separação de Hahn-Banach.

Dada uma matriz A $n \times n$, definimos um **autovetor** como o vetor $v \neq 0 \in \mathbb{R}^n$ tal que existe $\lambda \in \mathbb{C}$ com $Av = \lambda v$. O valor λ é dito **autovalor** de A .

Proposição 2.1 O valor λ é autovalor de $A \iff \det(A - \lambda I) = 0$.

O **traço** da matriz $A_{n \times n} = (A_{ij})_{1 \leq i, j \leq n}$ é dado por $\text{tr}(A) = \sum_{i=1}^n A_{ii}$, e é igual à soma dos autovalores da matriz (levando em conta a multiplicidade de autovalores).

Fato 2.2 Se A é simétrica com entradas reais, *i.e.*, $A = A^T$, então todos os autovalores de A são reais.

Demonstração. Seja A uma matriz com entradas complexas e que é igual a sua transposta conjugada: $A = A^* = \overline{A^T}$. Ou seja, se $A = (a_{jk})_{n \times n} = (x_{jk} + iy_{jk})_{n \times n}$, onde i é a unidade imaginária, vale

$$a_{jk} = x_{kj} + iy_{kj} = \overline{a_{kj}} = x_{kj} - iy_{kj},$$

para $1 \leq i, j \leq n$. (Nesse caso A é chamada matriz **hermitiana**.) Fixe um autovalor λ de A , e v autovetor de A . Seja $\langle x, y \rangle = x^T \overline{y}$ o produto interno de $x, y \in \mathbb{C}^n$. Então

$$\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle Av, v \rangle = \langle v, A^* v \rangle = \langle v, Av \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \langle v, v \rangle.$$

E como $\langle v, v \rangle \neq 0$, temos $\lambda = \overline{\lambda}$, e, portanto, $\lambda \in \mathbb{R}$. □

Fato 2.3 Toda matriz simétrica é diagonalizável, *i.e.*, $A = U^{-1}DU$, onde D é uma matriz diagonal e $U^{-1} = U^T$. Ademais, $A = \sum_{i=1}^n \lambda_i U_i U_i^T$.

Um **menor simétrico** de $A_{n \times n}$ é uma submatriz B obtida removendo-se algumas linhas de A e as respectivas colunas. Mais formalmente, B é menor simétrico de A se existe um conjunto de linhas $R \subset [n]$ tal que $B = (A_{ij})_{i \in R, j \in R}$.

Exemplo (*Menor simétrico*)

$$A = \begin{pmatrix} 1 & 4 & -1 & 7 \\ 4 & 2 & 3 & 0 \\ -1 & 3 & 5 & 4 \\ 7 & 0 & 4 & 3 \end{pmatrix} \quad B = \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix}$$

Teorema 2.4 (Entrelaçamento de autovalores, Cauchy). Seja $A_{n \times n}$ uma matriz simétrica com autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, e seja B um menor simétrico $(n-k) \times (n-k)$ de A com autovalores $\mu_1 \geq \mu_2 \geq \dots \geq \mu_{n-k}$. Vale $\lambda_i \leq \mu_i \leq \lambda_{i+k}$, para todo $1 \leq i \leq n-k$.

Uma matriz simétrica A é dita **positiva semidefinida** (PSD) se todos seus autovalores são não-negativos. Tal propriedade é denotada por $A \succeq 0$. A matriz é **positiva definida** se todos os seus autovalores são *positivos*.

Proposição 2.5 Dada uma matriz $n \times n$ real simétrica A , as seguintes afirmações são equivalentes.

1. a matriz A é PSD,
2. a forma quadrática $x^T A x$ é não-negativa, para todo $x \in \mathbb{R}^n$,
3. podemos escrever A como uma matriz de Gram de vetores $u_1, \dots, u_n \in \mathbb{R}^m$, para algum m . Isso significa que $A_{ij} = u_i^T u_j$, ou, equivalentemente, $A = U^T U$ para alguma matriz U ,
4. a matriz A é uma combinação linear não-negativa de matrizes da forma $x x^T$, para algum $x \in \mathbb{R}^n$,
5. o determinante de todo menor simétrico de A é não-negativo.

Observação 2.6 Para toda matriz A PSD, temos $\text{tr}(A) = 0$ se e somente se $A = 0$.

Demonstração. Temos que $A = U^T D U$, onde U é tal que $U^T = U^{-1}$. Assim, uma vez que $\text{tr}(XY) = \text{tr}(YX)$ (exercício!), temos $\text{tr}(A) = \text{tr}(U^T D U) = \text{tr}(D U U^T) = \text{tr}(D I) = \text{tr}(D)$. \square

A soma de semidefinidas é semidefinida.

A matriz PSD mais simples é da forma aa^T , onde $a \in \mathbb{R}^n$.

1. aa^T é PSD: $x^T(aa^T)x = (x^T a)(a^T x) = (a^T x)(a^T x) \geq 0$
2. seja $a = (a_1, \dots, a_n)$. Sem perda de generalidade, tome uma linha de aa^T que tem um elemento não-nulo. Essa linha (exercício!) gera alguma outra.

Proposição 2.7 Se A e B são PSD então $\text{tr}(AB) \geq 0$.

Demonstração. Pela propriedade 2.5.3, existem U, V tais que $A = U^T U$, e $B = V^T V$. Logo, $\text{tr}(AB) = \text{tr}(U^T U V^T V) = \text{tr}(V U^T U V^T) = \text{tr}((U V^T)^T (U V^T)) = \text{tr}(C^T C)$. Ora, mas para $C = (c_{ij})_{n \times n}$ vale

$$(C^T C)_{ii} = \sum_{j=1}^n c_{jj}^2 \geq 0. \quad \square$$

Em particular, se $\text{tr}(AB) = 0$ temos $AB = 0$ (pela observação 2.6), e assim $C = 0 = U V^T$, e $AB = U^T U V^T V = U^T 0 V = 0$.

Definição 2.8 (Produto interno). O **produto interno** usual de $M_n(\mathbb{R})$ é expresso por

$$\langle A, B \rangle = \sum_{i=1}^n \sum_{j=1}^n A_{ij} B_{ij} = \text{tr}(A^T B).$$

Definição 2.9 (Cone). Um conjunto C é dito **cone** se para todo vetor $x \in C$ e todo escalar positivo λ temos $\lambda x \in C$.

Definição 2.10 (Cone convexo). Seja C um subconjunto de um espaço vetorial. Dizemos que C é um **cone convexo** se C é cone e para todo par de vetores $x, y \in C$ temos $x + y \in C$.

Note que um cone C é convexo se e somente se é um *cone convexo* como na definição acima. Isto pois para $\lambda = 1/2$ e $x, y \in C$ temos $2(x/2 + y/2) \in C$ e, para $0 < \lambda < 1$ e $x, y \in C$ temos $\lambda x, (1 - \lambda)y \in C$ e também $\lambda x + (1 - \lambda)y \in C$.

Sistemas de inequações lineares homogêneas

$$\begin{aligned} a_1^T x &\geq 0 \\ &\vdots \\ a_m^T x &\geq 0 \end{aligned}$$

podem definir cones convexos. Quando os sistemas são finitos, temos os chamados **cone convexos poliedrais**.

Dado um cone convexo C , definimos o **cone polar** C^* :

$$C^* = \{x \in \mathbb{R}^n : x^T y \geq 0, \text{ para todo } y \in C\}.$$

Um **ponto de acumulação** p de $C \subseteq \mathbb{R}^n$ é um ponto $x \in \mathbb{R}^n$ tal que, para todo $\delta \geq 0$, existe $c \in C$ tal que a distância $d(p, c)$ é menor do que δ . O **fecho** de um conjunto é o conjunto de seus pontos de acumulação. Intuitivamente, podemos pensar no fecho de $C \subseteq \mathbb{R}^n$ como a união de C com sua fronteira.

Fato 2.11 O cone polar C^*

1. é convexo,
2. é fechado ($\overline{C^*} = C^*$), pois é interseção de semi-espacos
3. é tal que $(C^*)^* = \overline{C}$, onde \overline{C} é o fecho de C .

As demonstrações dos teoremas 2.12 e 2.13 a seguir foram enviadas por Eric Osami Endo.

Teorema 2.12 Seja C um conjunto convexo e fechado em \mathbb{R}^n e $x_0 \notin C$. Existe elemento $y \in C$ tal que $\langle y, x_0 \rangle < \langle y, x \rangle$ para todo $x \in C$.

Demonstração. Como C é fechado, existe $v \in C$ tal que

$$\|x_0 - v\| = \inf\{\|x_0 - x\| : x \in C\}.$$

Sabemos, para todo $x \in C$,

$$\begin{aligned} \|x_0 - x\|^2 &= \|x_0 - v + v - x\|^2 \\ &= \|x_0 - v\|^2 + 2\langle x_0 - v, v - x \rangle + \|v - x\|^2. \end{aligned}$$

Como $\|x_0 - v\| \leq \|x_0 - x\|$, temos

$$\|v - x\|^2 + 2\langle x_0 - v, v - x \rangle \geq 0.$$

Seja $\bar{x} = (1 - \lambda)v + \lambda x \in C$ com $0 < \lambda \leq 1$. Como $\bar{x} \neq v$, vale

$$\|v - \bar{x}\|^2 + 2\langle x_0 - v, v - \bar{x} \rangle \geq 0.$$

Como $v - \bar{x} = \lambda(v - x)$, temos

$$\lambda^2\|v - x\|^2 + 2\lambda\langle x_0 - v, v - x \rangle \geq 0.$$

E, como $\lambda \neq 0$, obtemos

$$\lambda\|v - x\|^2 + 2\langle x_0 - v, v - x \rangle \geq 0.$$

Tomando o limite com $\lambda \rightarrow 0$, concluímos $\langle x_0 - v, v - x \rangle \geq 0$ ou seja, $\langle v - x_0, x - v \rangle \geq 0$. Seja $y = v - x_0$. Temos $\langle y, x - v \rangle \geq 0$, o que implica $\langle y, v \rangle \leq \langle y, x \rangle$, para todo $x \in C$.

Ademais, como $x_0 \neq v$, temos $\|y\| > 0$, assim $\langle y, v - x_0 \rangle > 0$, e concluímos que $\langle y, x_0 \rangle < \langle y, v \rangle$.

Portanto $\langle y, x_0 \rangle < \langle y, x \rangle$ para todo $x \in C$, como queríamos. \square

Teorema 2.13 Sejam C e D dois conjuntos convexos disjuntos não-vazios em \mathbb{R}^n , com C fechado e D compacto. Então existe $y \in \mathbb{R}^n$ tal que $\langle y, d \rangle < \langle y, c \rangle$ para todo $c \in C$ e $d \in D$.

Demonstração. Seja $C - D = \{c - d : c \in C, d \in D\}$. Como C e D são convexos, então $C - D$ é convexo (exercício!). Mostremos que $C - D$ é fechado.

Seja $x_n = c_n - d_n$ uma sequência convergente de elementos de $C - D$, e x o limite. Mostremos que $x \in C - D$.

Como D é compacto, existe d_{n_k} subsequência convergente a $d \in D$. Como $c_{n_k} = x_{n_k} + d_{n_k}$ e ambas são convergentes, segue que c_{n_k} é convergente a $x + d$. Como C é fechado, temos $x + d \in C$. Logo $x = (x + d) - d$ pertence a $C - D$. Portanto $C - D$ é fechado.

Não é difícil de ver que $0 \notin C - D$, pois C e D são disjuntos. Portanto, pelo teorema 2.12, temos que existe $y \in \mathbb{R}^n$ tal que

$$\langle y, 0 \rangle < \langle y, c - d \rangle,$$

para todo $c - d \in C - D$. Como $\langle y, 0 \rangle = 0$, obtemos

$$\langle y, d \rangle < \langle y, c \rangle$$

para todo $c \in C$ e $d \in D$, como queríamos. \square

Observações: Um conjunto de \mathbb{R}^n é dito *compacto* se é fechado e limitado. Aqui, usamos a seguinte equivalência dos conjuntos compactos:

Um conjunto K é compacto se e somente se toda sequência em K possui uma subsequência convergente.

2013/4/9

Teorema 2.14 Se C é cone fechado, então $(C^*)^* = C$.

Demonstração. Demonstramos apenas que $C \supseteq (C^*)^*$.

Primeiro, mostramos que $x \notin C$ implica $x \notin (C^*)^*$. Por Hahn–Banach, existe $y \in \mathbb{R}^n$ tal que $x^T y < \inf\{z^T y : z \in C\} = M$. Como $0 \in C$, sabemos que $M \leq 0^T y = 0$. Vamos mostrar que $M = 0$. Suponha, por absurdo, que $M < 0$. Então existe $z \in C$ tal que $z^T y < 0$. Defina $\alpha = 2x^T y / z^T y > 0$. Como C é um cone, temos $\alpha z \in C$. De modo que $(\alpha z)^T y = \alpha z^T y = 2x^T y < x^T y$, absurdo. Portanto $M = 0$. Assim, $z^T y \geq 0$, para todo $z \in C$, o que implica $y \in C^*$. \square

2.2 Dualidade em programação linear

Lema 2.15 (Farkas). Sejam $a_1, \dots, a_m, x \in \mathbb{R}^n$, e $b_1, \dots, b_m \in \mathbb{R}$. Não existe solução para o sistema

$$\begin{aligned} a_1^T x &\leq b_1 \\ a_2^T x &\leq b_2 \\ &\vdots \\ a_m^T x &\leq b_m \end{aligned}$$

se e somente se existem $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ tais que

- $\sum_{i=1}^m \lambda_i a_i = 0$, e
- $\sum_{i=1}^m \lambda_i b_i = -1$.

Lema 2.16 (Farkas, versão de inferência). Sejam $a_1, \dots, a_m, c \in \mathbb{R}^n$, e $b_1, \dots, b_m, d \in \mathbb{R}$. Suponhamos que o sistema

$$\begin{aligned} a_1^T x &\leq b_1 \\ a_2^T x &\leq b_1 \\ &\vdots \\ a_m^T x &\leq b_m \end{aligned} \tag{1}$$

possui solução. Então as soluções de 1 satisfazem $c^T x \leq d$ se e somente se existem $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ tais que

- $c = \sum_{i=1}^m \lambda_i a_i = \lambda^T A$, e
- $d = \sum_{i=1}^m \lambda_i b_i = \lambda^T b$.

O problema **dual** de um problema de programação linear

$$\begin{aligned} \max \quad & c^T x \\ \text{s.a.} \quad & a_1^T x \leq b_1 \\ & \vdots \\ & a_m^T x \leq b_m \end{aligned} \tag{2}$$

é o PL definido pelo sistema

$$\begin{aligned} \min \quad & b^T y \\ \text{s.a.} \quad & A^T y = c \\ & y \geq 0 \end{aligned} \tag{3}$$

onde $b^T = (b_1, \dots, b_m)$ é um vetor coluna e $A^T = (a_1^T, \dots, a_m^T)$ é uma matriz cujas linhas são formadas pela transposição dos a_i , para $i = 1, \dots, m$.

Teorema 2.17 (Dualidade fraca). Se x é solução viável do PL 2 e y é solução viável do PL 3, então $b^T y \geq c^T x$.

Demonstração. Note que $b^T y \geq (Ax)^T y = x^T A^T y = x^T (A^T y) = x^T c = c^T x$. □

Teorema 2.18 (Dualidade forte). Se o PL 2 possui solução ótima, então o seu dual também possui solução ótima, e ambos os problemas possuem mesmo valor ótimo.

Teorema 2.19 (Folgas complementares, ou “complementary slackness”). Seja x um solução viável do PL 2 e y uma solução viável do PL 3. Então x é solução ótima do PL 2 e y uma solução ótima do PL 3 se e somente se as seguintes condições são satisfeitas

$$\left(b_i - \sum_{j=1}^n a_{ij}x_j \right) y_i = 0, \quad i = 1, \dots, n.$$

Note (exercício!) que a versão de inferência do teorema 1 (de Farkas) implica o teorema 2.18 (dualidade forte).

2.3 Polítopos de conjuntos estáveis

Seja $G = (V, E)$ um grafo sem vértices isolados, isto é, um grafo em que todo vértice está ligado a ao menos um outro. Um conjunto de vértices $S \subseteq V$ é dito **independente**, ou **estável** se não existe aresta no grafo que liga dois vértices em S (isto é, se $u, v \in S$, então $\{u, v\} \notin E$). O número de estabilidade de um grafo, denotado por $\alpha(G)$, representa o tamanho do maior conjunto estável de G . Formalmente, $\alpha(G)$ é o maior inteiro não-negativo tal que não existe conjunto $S \subseteq V$ independente com $|S| > \alpha(G)$.

2013/4/31

O problema da estabilidade é calcular

$$\alpha(G) = \max\{r \geq 1 : \text{existem } r \text{ vértices 2-a-2 não adjacentes em } G\}.$$

Fixado $S \subseteq V$, definimos o vetor de incidência χ^S dado por

$$\chi^S = \begin{cases} 1 & \text{se } i \in S \\ 0 & \text{caso contrário.} \end{cases}$$

Ademais, definimos $\text{STAB}(G)$ como o fecho convexo

$$\text{conv}\{\chi^S : S \subseteq V \text{ é um conj. independente}\}.$$

Nosso problema então admite a formulação

$$\alpha(G) = \max\{e^T x : x \in \text{STAB}(G)\}$$

onde $e = (1, \dots, 1)^T$ é um vetor coluna.

Veremos agora uma série de restrições que podemos impor a $x = (x_i)_{i \in V}$ sem perder soluções ótimas do problema.

Restrição da não-negatividade Sabemos que $x_i \geq 0$, para todo $i \in V$. Como cada coordenada satisfaz a restrição, o fecho convexo também a satisfaz.

Restrição da adjacência Para todo $ij \in E$, temos $x_i + x_j \leq 1$.

Restrição da clique Para toda clique B de G , temos $\sum_{i \in B} x_i \leq 1$.

2.4 Caracterização da equivalência

Definimos grafos **número cromático**, **subgrafo induzido**, e grafo **perfeito**.

Seja $G = (V, E)$ um grafo. Uma atribuição de cores aos vértices de G é uma **coloração própria** de G se aresta alguma teve os extremos coloridos com a mesma cor. O número cromático $\chi(G)$ é o mínimo número de cores em uma coloração própria de G . Verifique que, para qualquer grafo $G = (V, E)$, vale $\omega(G) \leq \chi(G) \leq |V|$ (exercício!).

Um grafo $H = (V', E')$ é **subgrafo** de $G = (V, E)$ se $V' \subseteq V$ e $E' \subseteq E$. Dizemos que $S = (V'', E'')$ é **subgrafo induzido** de G se S é subgrafo de G , e toda aresta de G entre vértices $u, v \in V''$ está em E'' . Em símbolos, $S = (V'', E'')$ é subgrafo induzido de $G = (V, E)$

$$V'' \subseteq V \quad \text{e} \quad E'' = \{\{u, v\} : u, v \in V'', \{u, v\} \in E\}.$$

Definição 2.20 (Grafo perfeito). Um grafo G é **perfeito** se todo subgrafo H induzido de G satisfaz $\chi(H) = \omega(H)$.

Teorema 2.21 (Fulkerson–Chvátal). Seja G um grafo. Temos $\text{STAB}(G) = \text{QSTAB}(G)$ se e somente se G é um grafo perfeito.

Teorema 2.22 (Conjectura fraca dos grafos perfeitos — Lóvasz, 1972). Um grafo G é perfeito se e somente se seu complemento \overline{G} é perfeito.

Teorema 2.23 (Conjectura forte dos grafos perfeitos — Chudnovsky, Robertson, Seymour, 2002). Um grafo G é perfeito se e somente se nem G nem \overline{G} contém circuito ímpar induzido com pelo menos 5 arestas.

2013/5/7

2.5 Programação Semidefinida

Definição 2.24 Um problema de programação semidefinida (PSD) é um problema de otimização da forma

$$\begin{aligned} \min \quad & c^T x = \sum_{i=1}^n c_i x_i \\ \text{s.a.} \quad & x_1 A_1 + \cdots + x_n A_n - B \succeq 0 \end{aligned}$$

onde A_1, \dots, A_n, B são matrizes simétricas $m \times m$ e $c \in \mathbb{R}^n$ é dado.

Definimos $\mathcal{F}_1 = \{x \in \mathbb{R}^n : x_1 A_1 + \cdots + x_n A_n - B \succeq 0\}$ e dizemos que $x \in \mathbb{R}^n$ é uma **solução viável** do problema PSD se $x \in \mathcal{F}_1$. Ademais, se $x_1 A_1 + \cdots + x_n A_n - B \succ 0$ dizemos que $x \in \mathbb{R}^n$ é uma **solução estritamente viável** do PSD, e definimos $v_{\text{primal}} = \inf_x \{c^T x : x \in \mathcal{F}_1\}$. Note que $v_{\text{primal}} \in \mathbb{R} \cup \{\pm\infty\}$, onde $v_{\text{primal}} = +\infty$ se $\mathcal{F}_1 = \emptyset$; e também $v_{\text{primal}} = -\infty$ se $\mathcal{F}_1 \neq \emptyset$ e, para todo $x \in \mathcal{F}_1$, existe $y \in \mathcal{F}_1$ tal que $c^T y < c^T x$.

Exemplo (Um problema PSD)

Considere o problema PSD a seguir.

$$\begin{aligned} \min \quad & x \\ \text{s.a.} \quad & \begin{pmatrix} x & 1 \\ 1 & 0 \end{pmatrix} \succeq 0. \end{aligned}$$

Podemos escrever a condição acima como

$$\begin{pmatrix} x & 1 \\ 1 & 0 \end{pmatrix} = x \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}}_{A_1} - \underbrace{\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}}_B \succeq 0, \quad \text{e ainda} \quad c = 1$$

Como $\det \begin{pmatrix} x & 1 \\ 1 & 0 \end{pmatrix} = -1 < 0$, segue que $\mathcal{F}_1 = \emptyset$, e portanto $v_{\text{primal}} = +\infty$.

Propriedades de PSD

Lema 2.25 (versão homogênea). Sejam A_1, \dots, A_n, B matrizes simétricas $m \times m$. Então o sistema $x_1 A_1 + \dots + x_n A_n - B \succ 0$ não tem solução se e somente se existe $Y \in \mathbb{R}^{m \times m}$, onde $Y \neq 0$, e $Y \succeq 0$ é tal que $A_i \cdot Y = 0$, para todo $1 \leq i \leq n$.

Proposição 2.26 Se $A \succ 0$, então $A \cdot B > 0$, para todo $B \neq 0$ tal que $B \succeq 0$.

Note que, em geral, $A \succ 0$ não é sinônimo de $A \neq 0$ e $A \succeq 0$.

Lema 2.27 (versão não-homogênea). Sejam A_1, \dots, A_n, B matrizes simétricas $m \times m$. Então o sistema $x_1 A_1 + \dots + x_n A_n - B \succ 0$ não tem solução se e somente se existe $Y \in \mathbb{R}^{m \times m}$, onde $Y \neq 0$, e $Y \succeq 0$ é tal que

$$\begin{cases} A_i \cdot Y = 0 & \text{para todo } 1 \leq i \leq n, \text{ e} \\ B \cdot Y \geq 0. \end{cases}$$

Definição 2.28 O dual do PSD

$$\begin{aligned} \min \quad & c^T x \\ \text{s.a.} \quad & x_1 A_1 + \dots + x_n A_n - B \succeq 0 \end{aligned}$$

é um problema de otimização da forma

$$\begin{aligned} \max \quad & B \cdot Y \\ \text{s.a.} \quad & A_1 \cdot Y = c_1 \\ & \vdots \\ & A_n \cdot Y = c_n \\ & Y \succeq 0 \end{aligned}$$

E definimos $v_{\text{dual}} = \sup\{B \cdot Y : A_i \cdot Y = c_i, 1 \leq i \leq n \text{ e } Y \succeq 0\}$.

Teorema 2.29 Suponhamos que os problemas PSD primal e dual possuem soluções viáveis. Então

- (a) $v_{\text{primal}} \geq v_{\text{dual}}$
- (a) se o problema PSD primal (respectivamente dual) tem uma solução estritamente viável, então existe solução ótima do dual (resp. primal) e $v_{\text{primal}} = v_{\text{dual}}$.

2013/5/21

2.6 Capacidade de Shannon

Ruído na comunicação

Considere a seguinte situação: desejamos enviar uma mensagem por um canal ruidoso. Se a mensagem é uma sequência de k símbolos escolhidos de um conjunto de n símbolos possíveis, temos ao total n^k mensagens possíveis. Vamos imaginar que possam ocorrer problemas na transmissão dos símbolos, e alguns deles podem ser confundidos.

É natural perguntar: qual o número de mensagens *diferentes* que podemos enviar, sem perigo de confusão? Vamos tentar responder a essa pergunta, supondo que conhecemos de antemão os pares de símbolos que podem ser confundidos. Nesta seção estamos interessados em resolver o seguinte problema a seguir.

Exemplo (*Mensagens compostas por duas cores*)

Suponha que as mensagens são sequências de duas cores, escolhidas entre as seguintes: vermelho, bege, verde, azul e roxo. Suponha, ademais que as possíveis confusões sejam

mensagens	primeira cor	segunda cor
1	vermelho	vermelho
2	bege	verde
3	roxo	azul
4	azul	bege
5	verde	roxo

Tabela 4: Algumas mensagens que não podem ser confundidas

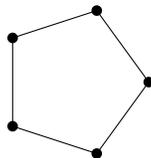


Figura 4: O circuito C_5 .

- vermelho e bege;
- bege e verde;
- verde e azul;
- azul e roxo; e
- roxo e vermelho.

Quantas mensagens distintas podemos enviar, sem correr o risco de haver confusão na mensagem enviada? A tabela 4 lista algumas mensagens possíveis. Note que não é possível confundir quaisquer das mensagens da tabela 4 (exercício!). Existe algum esquema de mensagens que permite enviar mais mensagens?

Formulação da capacidade de Shannon para grafos

A apresentação acima segue a apresentação feita na tese de bacharelado de Femke [Bek11], sobre [Lov79]. Seja V um conjunto finito (de símbolos) tal que alguns pares de V podem ser confundidos. Definimos o grafo $G = (V, E)$ com arestas $\{uv : u \text{ e } v \text{ podem ser confundidos}\}$. A figura 4 contém o grafo correspondente à situação do exemplo acima.

Sejam $a_i, b_j \in V$ para $1 \leq i, j \leq k$. Dizemos que as mensagens

$$a_1 a_2 \cdots a_k \quad \text{e} \quad b_1 b_2 \cdots b_k$$

se confundem se $a_i = b_i$ ou $a_i b_i \in E$ para todo $i \in V$. Denotamos por $\alpha_k(G)$ a quantidade máxima de mensagens de tamanho k em G . Note que $\alpha_1(G) = \alpha(G)$. Ademais $\alpha_2(C_5) \geq 5$ e, em geral $\alpha_k(C_5) \geq 5^{k/2}$, se k é par (exercício!).

Definição 2.30 A *capacidade de Shannon* de um grafo G é

$$\Theta(G) = \sup_{k \in \mathbb{Z}_+^*} \sqrt[k]{\alpha_k(G)}.$$

Para nosso problema, temos $\Theta(G) \geq \sqrt{5}$. Buscamos agora construir um grafo G^k tal que $\alpha(G^k) = \alpha_k(G)$. Temos

$$V(G^k) = V(G)^k = \underbrace{V(G) \times \cdots \times V(G)}_k = \{(v_1, v_2, \dots, v_k) ; v_i \in V(G), \text{ para } 1 \leq i \leq k\},$$

e a aresta $(a_1 \cdots a_k)(b_1 \cdots b_k) \in E(G^k)$ se as mensagens $a_1 \cdots a_k$ e $b_1 \cdots b_k$ se confundem. O grafo G^k é chamado **produto forte** de k cópias de G .

Definição 2.31 O **produto forte** dos grafos G e H é o grafo $G \cdot H$ com vértices $V(G \cdot H) = V(G) \times V(H)$ e arestas $(u_1, v_1)(u_2, v_2)$ se

- $u_1 = u_2$ ou $u_1 u_2 \in E(G)$, e ainda
- $v_1 = v_2$ ou $v_1 v_2 \in E(G)$.

Teorema 2.32 Vale que $\alpha(G^k) \geq \alpha(G)^k$.

Demonstração. Seja U um conjunto estável máximo em G . Vamos mostrar que U^k é estável em G^k .

Sejam $u_1 \dots u_k \in V(U^k)$ e $v_1 \dots v_k \in V(U^k)$ distintos. Então existe um índice $1 \leq i \leq k$ tal que $u_i \neq v_i$. Note que $u_i v_i \notin E(G)$. Assim $(u_1 \dots u_k)(v_1 \dots v_k) \notin E(G^k)$, e $\alpha(G^k) \geq \alpha(G)^k$. \square

Teorema 2.33 É verdade que $\Theta(G) \geq \alpha(G)$.

Demonstração. Por definição,

$$\Theta(G) = \sup_{k \in \mathbb{Z}_+^*} \sqrt[k]{\alpha_k(G)} \geq \sup_{k \in \mathbb{Z}_+^*} \sqrt[k]{\alpha(G)^k} \geq \alpha(G).$$

\square

Note que a recíproca não é verdade (exercício!).

Teorema 2.34 Para todo $p, q \in \mathbb{Z}_+^*$ vale que $\alpha(G^{p+q}) \geq \alpha(G^p)\alpha(G^q)$.

A prova é um exercício.

Teorema 2.35 É verdade que

$$\Theta(G) = \sup_{k \in \mathbb{Z}_+^*} \sqrt[k]{\alpha(G^k)} = \lim_{k \rightarrow \infty} \sqrt[k]{\alpha(G^k)}.$$

A demonstração do teorema feita na reunião usa o lema de Fekete abaixo.

Lema 2.36 (Lema de Fekete). Seja $(a_n)_{n \geq 0}$ uma sequência de reais superlinear, isto é, que satisfaz $a_{n+m} \geq a_n + a_m$. Então

$$\sup_{k \in \mathbb{Z}_+^*} \frac{a_k}{k} = \lim_{k \rightarrow \infty} \frac{a_k}{k}.$$

A demonstração a seguir foi enviada por Eric Ossami Endo.

Demonstração. Para todo $i, j, k \geq 1$, temos $a_{jk+i} \geq a_{jk} + a_i \geq ja_k + a_i$. Logo, fixados quaisquer $i, k \geq 1$, temos

$$\begin{aligned} \liminf_{j \rightarrow \infty} \left(\frac{a_{jk+i}}{jk+i} \right) &\geq \liminf_{j \rightarrow \infty} \left(\frac{ja_k + a_i}{jk+i} \right) \\ &= \liminf_{j \rightarrow \infty} \left(\frac{ja_k}{jk+i} + \frac{a_i}{jk+i} \right) \\ &= \liminf_{j \rightarrow \infty} \left(\frac{a_k}{k} \frac{jk}{jk+i} + \frac{a_i}{jk+i} \right) \\ &= \frac{a_k}{k}. \end{aligned}$$

A última igualdade vale pois $\lim_{j \rightarrow \infty} jk/(jk+i) = 1$ e $\lim_{j \rightarrow \infty} a_i/(jk+i) = 0$. Como a desigualdade acima vale para todo i , temos, para todo $k \geq 1$ fixo,

$$\begin{aligned} \liminf \left(\frac{a_n}{n} \right) &= \inf_{1 \leq i \leq k} \liminf_{j \rightarrow \infty} \left(\frac{a_{jk+i}}{jk+i} \right) \\ &\geq \inf_{1 \leq i \leq k} \left(\frac{a_k}{k} \right) \\ &= \frac{a_k}{k}. \end{aligned}$$

Logo, $\liminf \left(\frac{a_n}{n} \right) \geq \sup_k \frac{a_k}{k}$, e concluímos que

$$\lim_{n \rightarrow \infty} \frac{a_n}{n} \geq \sup_{n \in \mathbb{Z}_+^*} \frac{a_n}{n}.$$

Como o outro lado da desigualdade sempre vale, obtemos a igualdade desejada. \square

Representação ortonormal para grafos

Uma **representação ortonormal** de um grafo G é uma função $u : V \rightarrow \mathbb{R}^n$, para algum n , tal que

1. $u(i) = u_i$ é um vetor unitário para todo $i \in V$, isto é, $\|u_i\|_2 = 1$;
2. se $ij \notin E(G)$, então $u_i^T u_j = 0$, ou seja u_i e u_j são ortonormais.

Note que todo grafo possui uma representação ortonormal.

Definimos o **produto tensorial** de $v = (v_1, \dots, v_m) \in \mathbb{R}^m$ e $w = (w_1, \dots, w_n) \in \mathbb{R}^n$ como

$$v \circ w = (v_1 w_1, v_1 w_2, \dots, v_1 w_n, v_2 w_1, \dots, v_m w_n) \in \mathbb{R}^{mn}.$$

Teorema 2.37 Seja $u : V(G) \rightarrow \mathbb{R}^m$ uma representação ortonormal de G , e $v : V(H) \rightarrow \mathbb{R}^n$ uma representação ortonormal de H . Então a função $w : V(G \cdot H) \rightarrow \mathbb{R}^{mn}$ que mapeia $(i, j) \rightarrow w(i, j) = u_i \circ v_j$ é uma representação ortonormal de $G \cdot H$.

Teorema 2.38 (Lóvasz, 1979). É verdade que $\Theta(C_5) = \sqrt{5}$.

Teorema 2.39 (Lóvasz). É verdade que, para todo n ímpar

$$\Theta(C_n) \leq \frac{n \cos \frac{\pi}{n}}{1 + \cos \frac{\pi}{n}}.$$

2013/5/28

Função ϑ de Lovász

Uma representação ortonormal $u : V(G) \rightarrow \mathbb{R}^k$ de um grafo possui *valor*

$$\min_c \max_{1 \leq i \leq n} \frac{1}{(c^T u_i)^2},$$

onde c percorre todos os vetores unitários. Definimos a **função ϑ de Lovász** por

$$\vartheta(G) = \min_{(u_1, \dots, u_n)} \min_c \max_{1 \leq i \leq n} \frac{1}{(c^T u_i)^2}.$$

A representação ortonormal $u = (u_1, \dots, u_n)$ é *ótima* se atinge o mínimo em $\vartheta(G)$ e o menor c é dito **cabo**. Vimos os seguintes exemplos.

Exemplo (Função de Lovász para o grafo completo K_n)

Como todas as arestas possíveis estão em K_n , a condição 2 nunca é aplicável, e podemos tomar a representação com $k = 1$ e $u(1) = 1$, donde $\vartheta(K_n) \leq 1$. Como sabemos em geral que $\vartheta(G) \geq 1$ (exercício!), segue que $\vartheta(K_n) = 1$.

Exemplo (Função de Lovász para o grafo vazio $\overline{K_n}$)
 $\vartheta(\overline{K_n}) = n$.

Lema 2.40 Para grafos G e H como acima, vale $\vartheta(G \cdot H) = \vartheta(G)\vartheta(H)$.

Lema 2.41 Para todo grafo G , vale $\alpha(G) \leq \vartheta(G)$.

Teorema 2.42 $\Theta(G) \leq \vartheta(G)$.

Veremos a seguir outros modos de calcular a função ϑ .

Lema 2.43 Seja A uma matriz simétrica e $\lambda \in \mathbb{R}$. Então $\lambda I - A$ é PSD se e somente se $\lambda_i \leq \lambda$ para todo λ_i autovalor de A .

Teorema 2.44 Seja G um grafo com $|V(G)| = n$. Então $\vartheta(G)$ é o menor valor dentre todos os autovalores das matrizes $A = (a_{ij})_{n \times n}$ que satisfazem

$$a_{ij} = 1 \quad \text{se } i = j \text{ ou } ij \notin E(G).$$

2013/6/4

Teorema 2.45 Todo grafo G tem uma representação ortonormal ótima (u_1, \dots, u_n) com cabo c tal que

$$\vartheta(G) = \frac{1}{(c^T u_i)^2}, \quad 1 \leq i \leq n.$$

Demonstração. Seja (v_1, \dots, v_n) uma representação ortonormal ótima de G e seja d um vetor unitário qualquer. Considere a matriz simétrica $A = (a_{ij})_{1 \leq i, j \leq n}$ tal que

$$a_{ij} = 1 \quad \text{se } i = j \text{ ou } ij \notin E(G), \quad (4)$$

e, se λ é o maior autovalor de A , então $\vartheta(G) \geq \lambda$. Como A satisfaz 4, existe uma representação ortonormal (u_1, \dots, u_n) e um vetor unitário c tais que

$$\lambda = \frac{1}{(c^T u_i)^2} \quad \text{para } 1 \leq i \leq n \quad \text{e} \quad \vartheta(G) \leq \lambda.$$

Logo $\vartheta(G) = \lambda = (c^T u_i)^{-2}$, para $1 \leq i \leq n$, e (u_1, \dots, u_n) é uma representação ortonormal ótima com cabo c . \square

Teorema 2.46 Seja $G = (V, E)$ um grafo com $V = \{1, \dots, n\}$. O valor $\vartheta(G)$ é solução viável ótima do seguinte problema PSD.

$$\begin{aligned} \max \quad & \text{tr}(BJ) \\ \text{s.a.} \quad & B \succeq 0, \quad B = (b_{ij})_{1 \leq i, j \leq n} \in M_{n \times n} \\ & b_{ij} = 0 \text{ se } ij \in E \\ & \text{tr}(B) = 1. \end{aligned}$$

onde $J = (J_{ij})_{1 \leq i, j \leq n}$ é dada por $J_{ij} = 1$, para todo $1 \leq i, j \leq n$.

Demonstração. Considere o conjunto \mathcal{F} das matrizes que satisfazem às restrições do PSD no teorema. Formalmente,

$$\mathcal{F} = \{B \in M_{n \times n} : B \succeq 0, b_{ij} = 0 \text{ se } ij \in E(G), \text{tr}(B) = 1\}.$$

Nossa demonstração segue em partes. *i)* Demonstramos que $\vartheta(G) \geq \max_{B \in \mathcal{F}} \text{tr}(BJ)$. Em seguida, construímos uma matriz $B \in M_{n \times n}$ tal que $B \in \mathcal{F}$ e $\vartheta(G) = \text{tr}(BJ)$.

Sejam $(i_1, j_1), \dots, (i_m, j_m)$ arestas de G com $m = |E|$. Para cada $h = (h_1, \dots, h_m)^T \in S^{n-1}$, definimos

$$\tilde{h} = \left(h_{i_1} h_{j_1}, \dots, h_{i_m} h_{j_m}, \left(\sum_{i=1}^n h_i \right)^2 \right)^T \in \mathbb{R}^{m+1}.$$

Seja $\tilde{H} = \{\tilde{h} : h \in S^{n-1}\}$ e $Z = (0, 0, \dots, \vartheta(G))^T \in \mathbb{R}^{m+1}$. □

Afirmação 2.47 O vetor Z pertence à interseção de todos os convexos que contêm \tilde{H} . Ou seja, $Z \in \text{conv}(\tilde{H})$.

Afirmção 2.48 O maior autovalor λ de uma matriz simétrica A é

$$\lambda = \max\{h^t A h; h \in S^{n-1}\},$$

onde S^{n-1} é a esfera unitária (conjunto dos vetores unitários).

Lema 2.49 Seja (u_1, \dots, u_n) uma representação ortonormal de $G = (V, E)$, seja (v_1, \dots, v_n) uma representação ortonormal de \overline{G} e ainda $c, d \in \mathbb{R}^n$. Temos a seguinte desigualdade.

$$\sum_{i=1}^n (u_i^T c)^2 (v_i^T d)^2 \leq \|c\|^2 \|d\|^2.$$

Corolário 2.50 Se (v_1, \dots, v_n) é representação ortonormal de \overline{G} e $d \in S^{n-1}$, então

$$\vartheta(G) \geq \sum_{i=1}^n (v_i^T d)^2.$$

Teorema 2.51 Seja \mathcal{A} o conjunto das representações ortonormais de \overline{G} . Vale a seguinte igualdade.

$$\vartheta(G) = \max_{(v_1, \dots, v_n) \in \mathcal{A}} \max_{d \in S^{n-1}} \sum_{i=1}^n (v_i^T d)^2.$$

2013/6/11

Teorema 2.52 Vale a seguinte igualdade.

$$\vartheta(G \cdot H) = \vartheta(G) \vartheta(H).$$

A prova a seguir baseia-se no seguinte fato, cuja prova é deixada de exercício para o leitor.

Afirmção 2.53 Sejam G e H grafos. Vale que $\overline{G} \cdot \overline{H} \subseteq \overline{G \cdot H}$.

Demonstração. Sabemos que

$$\vartheta(G) = \max_{\substack{(v_1, \dots, v_n) \\ \text{rep. ort. de } G}} \max_{d \in S^{n-1}} \sum_{i=1}^n (v_i^T d)^2.$$

Sejam (v_1, \dots, v_n) e (w_1, \dots, w_n) , respectivamente, representações ortonormais de \overline{G} e de \overline{H} , e sejam $c, d \in S^{n-1}$ tais que

$$\vartheta(G) = \sum_{i=1}^n (v_i^T c)^2 \quad \text{e} \quad \vartheta(H) = \sum_{i=1}^n (w_i^T d)^2.$$

Pela contrapositiva do fato 2.53, temos que se $e \notin \overline{G \cdot H}$, então $e \notin \overline{G} \cdot \overline{H}$.

Assim, (exercício!) $(v_i \circ w_j)_{i=1, j=1}^{n, m}$ é uma representação ortonormal de $\overline{G \cdot H}$. Portanto,

$$\begin{aligned} \vartheta(G \cdot H) &\geq \sum_{i=1}^n \sum_{j=1}^m ((v_i \circ w_j)^2 (c \circ d)^2) \\ &= \sum_{i=1}^n \sum_{j=1}^m (v_i^T c)^2 (w_j^T d)^2 \\ &= \vartheta(G) \vartheta(H). \end{aligned} \quad \square$$

2.7 Politopo STAB(G)

Vimos, para $G = (V, E)$, com $V = \{1, \dots, n\}$ que

- $\text{STAB}(G) = \text{conv}\{\chi_S : S \text{ conjunto estável em } G\}$ e
- $\text{QSTAB}(G) = \{x \in \mathbb{R}^n : x \geq 0, \sum_{i \in B} x_i \leq 1 \text{ para toda clique } B \text{ em } G\}$.

Definimos $\text{TSTAB}(G) = \{((v_i^T d)^2)_{i=1}^n \in \mathbb{R}^n : (v_1, \dots, v_n) \text{ rep.ort. de } \overline{G} \text{ e } d \in S^{n-1}\}$.

Seja u_1, \dots, u_n uma representação ortonormal de G e seja $c \in S^{n-1}$. A restrição de $(x_1, \dots, x_n) \in \mathbb{R}^n$

$$\sum_{i=1}^n (u_i^T c)^2 x_i \leq 1$$

é chamada **restrição da ortogonalidade**. Temos o seguinte teorema, que demonstraremos parcialmente.

Teorema 2.54 Vale a seguinte equivalência:

$$\text{TSTAB}(G) = \{x \in \mathbb{R}^n : x \geq 0 \text{ e } x \text{ satisfaz a restrição da ortogonalidade}\}.$$

Prova parcial. Fixe uma representação ortonormal v_1, \dots, v_n de \overline{G} e $d \in S^{n-1}$. Para toda representação ortonormal u_1, \dots, u_n de G e $c \in S^{n-1}$, vale que

$$\sum_{i=1}^n (u_i^T c)^2 (v_i^T d)^2 \leq \|c\|^2 \|d\|^2 = 1. \quad \square$$

Teorema 2.55 Para todo grafo G , temos $\text{STAB}(G) \subseteq \text{TSTAB}(G) \subseteq \text{QSTAB}(G)$

Demonstração. Primeiro demonstramos que $\text{STAB}(G) \subseteq \text{TSTAB}(G)$. É suficiente argumentar que $\text{TSTAB}(G)$ é convexo e que todos os conjuntos estáveis de G pertencem a $\text{TSTAB}(G)$. A convexidade decorre de $\text{TSTAB}(G)$ ser interseção de semiespaços. Para a outra condição, fixe um conjunto estável S em G . Sabemos que χ_S é formado de zeros e uns e como são ortonormais, podemos completar seu conjunto para formar uma base ortonormal de \mathbb{R}^n . Assim,

$$\sum_{i=1}^n (u_i^T c)^2 \chi_S^i = \sum_{i \in S} (u_i^T c)^2 \leq \|c\|^2 = 1.$$

Demonstramos agora que $\text{TSTAB}(G) \subseteq \text{QSTAB}(G)$. Fixe $x \in \text{TSTAB}(G)$. Seja B uma clique de G qualquer. Tome $c = e_1$ e

$$v_i = \begin{cases} e_1 & \text{se } i \in B \\ e_{i+1} & \text{caso contrário} \end{cases}.$$

Então os vetores são unitários e se ij não é aresta de G ...

Note que (u_1, \dots, u_n) é uma representação ortonormal em G . Portanto

$$\sum_{i=1}^n (u_i^T c)^2 x_i = \sum_{i \notin B} (u_i^T c)^2 x_i + \sum_{i \in B} (u_i^T c)^2 x_i = \sum x_i. \quad \square$$

Teorema 2.56 (Lovász, 1979). Para todo grafo G , vale que $\omega(G) \leq \vartheta(G) \leq \chi(G)$.

Demonstração. Sabemos que $\alpha(G) \leq \vartheta(G)$, o que implica que $\alpha(\overline{G}) \leq \vartheta(\overline{G})$, e assim $\omega(G) \leq \vartheta(G)$. Seja $k(G) = \max\{e^T x : x \in \text{QSTAB}(G)\}$, onde e é um vetor com uma coordenada de valor 1 e as demais 0. Pelo teorema anterior, vale $\vartheta(G) \leq k(G)$.

Lembramos que uma *clique* de um grafo é um subgrafo completo. Definimos uma **cobertura de G por cliques** como um conjunto de cliques tal que todo vértice de G está em alguma delas. Definimos ainda $\overline{\chi}(G)$ como o tamanho da menor cobertura de G por cliques.

Vamos mostrar que $k(G) \leq \overline{\chi}(G)$. Sejam $\{C_1, \dots, C_p\}$ uma cobertura mínima de G por cliques. Sem perda de generalidade (exercício!), podemos supor que as cliques são vértice-disjuntas. Assim

$$\sum_{i=1}^n x_i = \sum_{j=1}^p \underbrace{\sum_{i \in V(C_j)} x_i}_{\leq 1} \leq p = \overline{\chi}(G).$$

Portanto $e^T x \leq \overline{\chi}(G)$ para todo $x \in \text{QSTAB}(G)$ e $k(G) \leq \overline{\chi}(G)$. Assim, temos

$$\vartheta(\overline{G}) \leq k(\overline{G}) \leq \overline{\chi}(\overline{G}) = \chi(G). \quad \square$$

Teorema 2.57 O conjunto $\text{TSTAB}(G)$ é um politopo se e somente se G é perfeito.

Nesse caso, o teorema 2.55 implica que $\text{STAB}(G) = \text{TSTAB}(G) = \text{QSTAB}(G)$. Enunciamos ainda outro teorema, sem demonstração.

Teorema 2.58 Toda função objetivo *linear* pode ser maximizada sobre $\text{TSTAB}(G)$, com erro arbitrariamente pequeno e em tempo polinomial.

2013/6/18

3 Conjectura de Erdős–Simonovits–Sidorenko

Tópicos: homomorfismos de grafos, densidade de homomorfismos; espaços mensuráveis, desigualdade de Jensen (esperança de função convexa), acoplamentos independentes.

Sejam G, H grafos. Uma função $f : V(H) \rightarrow V(G)$ é dita **homomorfismo** se para toda aresta $xy \in E(H)$ vale $f(x)f(y) \in E(G)$. O conjunto de todos os homomorfismos de H para G é denotado por

$$\text{hom}(H, G) = \{f : V(H) \rightarrow V(G), \text{ tal que } f \text{ é homomorfismo}\}.$$

É fácil ver que todo grafo G bipartido satisfaz $\text{hom}(G, K_2) \neq \emptyset$. Definimos a **densidade de homomorfismos** de H para G por

$$t(H, G) = \frac{|\text{hom}(H, G)|}{|V(G)|^{|V(H)|}}.$$

Podemos interpretar a densidade de homomorfismos como uma probabilidade sobre todos os mapas possíveis de $V(H)$ para $V(G)$: a probabilidade de um mapa escolhido uniformemente ao acaso ser um homomorfismo (isto é, de preservar as arestas de H).

Em particular, $t(K_2, G) = 2|E(G)|/|V(G)|^2$ é “essencialmente” a densidade de arestas de G . Dizemos “essencialmente” pois a densidade de arestas é $|E(G)|/(\binom{|V(G)|}{2})$.

Gostaríamos de estimar $t(H, G)$ de algum modo. Uma pergunta natural é se vale a aproximação $t(H, G) \approx t(K_2, G)^{|E(H)|}$. Para H bipartido, temos a conjectura a seguir [Sid93].

Conjectura 3.1 (Conjectura de Erdős–Simonovits–Sidorenko). Se H é um grafo bipartido, então

$$t(H, G) \geq t(K_2, G)^{|E(H)|}.$$

Um grafo H bipartido é dito **de Sidorenko** se para todo G vale a conjectura 3.1. Note que ocorre igualdade quando G é um grafo vazio. A conjectura só foi demonstrada para algumas classes de grafos. É um fato que, se H não é bipartido, então existe G tal que a conjectura é falsa.

3.1 Espaços mensuráveis

Dizemos que X é um **espaço mensurável discreto** se X é enumerável e está munido da σ -álgebra $\mathcal{P}(X)$. Lembramos que $\mathcal{P}(X)$ é a família de todos os $2^{|X|}$ subconjuntos de X .

Informalmente, dizemos que $\mu : \mathcal{P}(X) \rightarrow [0, 1]$ é uma **probabilidade** sobre X se dá a probabilidade dos subconjuntos de X . Definimos $\hat{\mu} : X \rightarrow [0, 1]$ como a função que mapeia $x \in X$ na probabilidade do evento unitário $\mu(\{x\})$. O **suporte** de uma medida de probabilidade sobre um espaço mensurável discreto (X, μ) é o “menor” subconjunto de X com medida positiva:

$$\bigcap_{\substack{S \in \mathcal{P}(X) \\ \mu(S) > 0}} S = \bigcup_{\substack{x \in X \\ \mu(\{x\}) > 0}} x.$$

Seja X um espaço mensurável discreto, e μ, ν duas probabilidades sobre X . Dizemos que μ é **absolutamente contínua** em relação a ν se para todo subconjunto A de X vale que $\mu(A) = 0$ sempre que $\nu(A) = 0$. Escrevemos “ $\mu \ll \nu$ ” para indicar que μ é absolutamente contínua em relação a ν .

Para espaços discretos isso equivale a exigir, para todo $x \in X$, que $\hat{\mu}(x) = 0$ quando $\hat{\nu}(x) = 0$. Em outras palavras, o suporte $\text{supp}(\hat{\mu})$ de ν é subconjunto do suporte $\text{supp}(\hat{\nu})$ de μ . Nesse caso, a **entropia relativa** $D(\mu \parallel \nu)$ como

$$D(\mu \parallel \nu) = \sum_{x \in \text{supp}(\hat{\mu})} \hat{\mu}(x) \log \frac{\hat{\mu}(x)}{\hat{\nu}(x)}.$$

(Para espaços que não são discretos $D(\mu \parallel \nu) = \int_X \ln \partial\mu/\partial\nu \partial\mu$.)

Re-enunciamos agora a conjectura 3.1 explicitamente em termos de probabilidades. Para isso, fixados grafos H e G , definimos $\nu(H, G)$ como a probabilidade uniforme sobre o conjunto

$$V(G)^{V(H)} = \{f : V(H) \rightarrow V(G)\},$$

onde “uniforme” significa que cada elemento do conjunto ocorre com a mesma probabilidade. Ademais, definimos a probabilidade $\mu(H, G)$, sobre o mesmo conjunto, como

$$\mu(H, G)(A) = \frac{|A \cap \text{hom}(H, G)|}{|\text{hom}(H, G)|}.$$

Note que μ é absolutamente contínua em relação a ν , e podemos calcular

$$\begin{aligned}
-\ln(t(H, G)) &= \ln \frac{1/|\text{hom}(H, G)|}{1/|V(G)^{V(H)}|} \\
&= \sum_{f \in \text{hom}(H, G)} \ln \frac{\hat{\mu}(H, G)(f)}{\hat{\nu}(H, G)(f)} \\
&= \sum_{f \in \text{hom}(H, G)} \hat{\mu}(H, G)(f) \ln \frac{\hat{\mu}(H, G)(f)}{\hat{\nu}(H, G)(f)}
\end{aligned}$$

Conjectura 3.2 (Conjectura 3.1, versão entropia). Se H é um grafo bipartido, então

$$D(\mu(H, G) \parallel \nu(H, G)) \leq |E(H)|D(\mu(K_2, G) \parallel \nu(K_2, G)).$$

3.2 Desigualdade de Jensen

Seja J um intervalo, e $f : J \rightarrow \mathbb{R}$ uma função de J com imagem \mathbb{R} . A função f é **convexa** se e somente se para todo $t \in [0, 1]$ e quaisquer $x, y \in J$, temos

$$tf(x) + (1-t)f(y) \geq f(tx + (1-t)y).$$

Teorema 3.3 (Desigualdade de Jensen). Seja J um intervalo, f uma função convexa, X um espaço de probabilidade e Y uma variável aleatória que assume valores em J e tem esperança $\mathbb{E}(Y) \in J$. Vale que $\mathbb{E}(f(Y)) \geq f(\mathbb{E}(Y))$.

Teorema 3.4 Se $\mu \ll \nu$, então a entropia relativa satisfaz $D(\mu \parallel \nu) \geq 0$.

Demonstração. Aplicamos a desigualdade de Jensen à função $f(y) = -\ln(y)$.

$$\begin{aligned}
D(\mu \parallel \nu) &= - \sum_{x \in \text{supp } \hat{\mu}} -\hat{\mu}(x) \ln \frac{\hat{\mu}(x)}{\hat{\nu}(x)} \\
&= \sum_{x \in \text{supp } \hat{\mu}} -\hat{\mu}(x) \ln \frac{\hat{\nu}(x)}{\hat{\mu}(x)} \\
&\geq -\ln \sum_{x \in \text{supp } \hat{\mu}} \frac{\hat{\nu}(x)}{\hat{\mu}(x)} \hat{\mu}(x) \\
&\geq -\ln \sum_{x \in \text{supp } \hat{\mu}} \hat{\nu}(x) \\
&\geq 0
\end{aligned}$$

□

3.3 Desigualdade log-soma

Teorema 3.5 (Desigualdade log-soma). Se $(a_i)_{i \in I}$ e $(b_i)_{i \in I}$ são seqüências enumeráveis de reais positivos, com somas finitas, isto é

$$a = \sum_{i \in I} a_i < +\infty \quad \text{e} \quad b = \sum_{i \in I} b_i < +\infty,$$

então

$$\sum_{i \in I} a_i \ln \frac{a_i}{b_i} \leq a \ln \frac{a}{b}.$$

Demonstração. Seja $f :]0, +\infty[\rightarrow \mathbb{R}$ a função $f(x) = x \ln x$. Suas derivadas

$$f'(x) = \ln x + 1 \quad \text{e} \quad f''(x) = \frac{1}{x}$$

somente assumem valores positivos, o que implica que f é convexa. Assim podemos aplicar Jensen

$$\sum_{i \in I} b_i \frac{a_i}{b_i} \ln \frac{a_i}{b_i} = b \sum_{i \in I} \frac{b_i}{b} f\left(\frac{a_i}{b_i}\right) \geq b f\left(\sum_{i \in I} \frac{b_i a_i}{b_i b}\right) = b f\left(\frac{a}{b}\right) = a \ln \frac{a}{b}. \quad \square$$

Seja X um espaço mensurável discreto, e μ, ν medidas de probabilidade sobre X . Seja ainda $M = \nu(\text{supp}(\hat{\mu}))$. Definimos $\alpha : \mathcal{P}(X) \rightarrow [0, 1]$ como $\alpha(A) = \nu(A \cap \text{supp}(\hat{\mu}))/M$. Note que α é uma medida de probabilidade. Se α é absolutamente contínua em relação a ν , então μ também é, e $D(\alpha \parallel \nu) \leq D(\mu \parallel \nu)$.

Demonstração. Note que

$$D(\mu \parallel \nu) = \sum_{x \in \text{supp} \hat{\mu}} \hat{\mu}(x) \ln \frac{\hat{\mu}(x)}{\hat{\nu}(x)} \geq 1 \ln \frac{1}{M}.$$

Definindo $\hat{\alpha}(x) = \alpha(\{x\})$, sabemos que $\hat{\alpha}(x) = \nu(\{x\})/M$, para todo $x \in \text{supp}(\hat{\alpha})$. Então $\text{supp} \hat{\alpha} = \text{supp} \hat{\mu}$, e

$$D(\alpha \parallel \nu) = \sum_{x \in \text{supp} \hat{\mu}} \hat{\alpha}(x) \ln \frac{\hat{\alpha}(x)}{\hat{\nu}(x)} = \sum_{x \in \text{supp} \hat{\mu}} \frac{\hat{\nu}(x)}{M} \ln \frac{1}{M} = \ln \frac{1}{M} \leq D(\mu \parallel \nu). \quad \square$$

Referências Bibliográficas

- [Bek11] Femke Bekius. The shannon capacity of a graph. Bachelor's thesis, supervised by prof. dr. Alexander Schrijver, KdV Instituut voor wiskunde Faculteit der Natuurwetenschappen, Wiskunde en Informatica Universiteit van Amsterdam, 07 2011.
- [Lov79] L. Lovász. On the shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25(1):1–7, 1979.
- [Lov95] László Lovász. Semidefinite programs and combinatorial optimization, 1995.
- [MG07] J. Matoušek and Gärtner. *Understanding and Using Linear Programming*. Universitext (En ligne). Springer, 2007.
- [Sid93] Alexander Sidorenko. A correlation inequality for bipartite graphs. *Graphs Combin.*, 9(2):201–204, 1993.