Fundamentos de teoria quântica

Introdução a otimização convexa

Rafael Rabelo – rabelo@ifi.unicamp.br

Departamento de Física da Matéria Condensada
Instituto de Física "Gleb Wataghin"
Universidade Estadual de Campinas

Reta e segmento de reta

Sejam dois pontos $x_1 \neq x_2$ de $\mathbb{R}^n$. Pontos da forma

$$y = \theta x_1 + (1 - \theta) x_2, \tag{1}$$

onde $\theta \in \mathbb{R}$, formam a *reta* que passam por x_1 e x_2.

Para $0 \leq \theta \leq 1$, tem-se o *segmento de reta* entre x_1 e x_2.

Conjunto afim

Um conjunto $C \subseteq \mathbb{R}^n$ é *afim* se a reta que passa entre quaisquer dois pontos em C está contida em C, ou seja, para $x_1, x_2 \in C$ e $\theta \in \mathbb{R}$, $\theta x_1 + (1-\theta)x_2 \in C$.

Combinação afim

Um ponto da forma

$$\theta_1 x_1 + \cdots + \theta_k x_k, \tag{2}$$

onde $\sum_{i=1}^{k} \theta_i = 1$ é uma *combinação afim* dos pontos $x_1, \ldots, x_k$.

Fecho afim

O conjunto de todas as combinações afins de $C \subseteq \mathbb{R}^n$ é chamado de *fecho afim de C*, denotado $\mathrm{aff}(C)$:

$$\mathrm{aff}(C) = \left\{ \theta_1 x_1 + \cdots + \theta_k x_k \,\middle|\, x_1, \ldots, x_k \in C; \sum_{i=1}^{k} \theta_i = 1 \right\}. \qquad (3)$$

O fecho afim de C é o menor conjunto afim que contém C.

Conjunto convexo

Um conjunto $C \subseteq \mathbb{R}^n$ é *convexo* se o segmento de reta que passa entre quaisquer dois pontos em C está contida em C, ou seja, para $x_1, x_2 \in C$ e $0 \leq \theta \leq 1$, $\theta x_1 + (1 - \theta) x_2 \in C$.

Combinação convexa

Um ponto da forma

$$\theta_1 x_1 + \cdots + \theta_k x_k, \tag{4}$$

onde $\theta_i \geq 0$, $\sum_{i=1}^k \theta_i = 1$, é uma *combinação convexa* dos pontos $x_1, \ldots, x_k$.

Ponto extremal

Seja C um conjunto convexo. Um ponto $x \in C$ que não pode ser escrito como combinação convexa de outros pontos de C é dito *extremal*.

Fecho convexo

O conjunto de todas as combinações convexas de $C \subseteq \mathbb{R}^n$ é chamado de *fecho convexo de* C, denotado $\mathrm{conv}(C)$:

$$\mathrm{conv}(C) = \left\{ \sum_{i=1}^{k} \theta_i x_i \,\middle|\, x_i \in C; \theta_i \geq 0, i = 1, \ldots, k; \sum_{i=1}^{k} \theta_i = 1 \right\}. \qquad (5)$$

O fecho convexo de C é o menor conjunto convexo que contém C.

Cone

Um conjunto $C \subseteq \mathbb{R}^n$ é um *cone* se para todo $x \in C$ e $\theta \geq 0$, $\theta x \in C$.

Um conjunto $C \subseteq \mathbb{R}^n$ é um *cone convexo* se é convexo e é um cone, ou seja, para todo par de pontos $x_1, x_2 \in C$ e $\theta_1, \theta_2 \geq 0$,

$$\theta_1 x_1 + \theta_2 x_2 \in C. \tag{6}$$

Combinação cônica

Um ponto da forma

$$\theta_1 x_1 + \cdots + \theta_k x_k, \tag{7}$$

onde $\theta_i \geq 0$, é uma *combinação cônica* dos pontos $x_1, \ldots, x_k$.

Fecho cônico

O conjunto de todas as combinações cônicas de $C \subseteq \mathbb{R}^n$ é chamado de *fecho cônico de* C, denotado $\text{cone}(C)$:

$$\text{cone}(C) = \left\{ \sum_{i=1}^{k} \theta_i x_i \middle| x_i \in C; \theta_i \geq 0, i = 1, \ldots, k \right\}. \tag{8}$$

O fecho cônico de C é o menor cone que contém C.

Hiperplano

Um *hiperplano* é um conjunto da forma

$$\left\{x | a^T x = b\right\}, \tag{9}$$

onde $a \in \mathbb{R}^n, a \neq 0$ e $b \in \mathbb{R}$.

Semi-espaços

Um hiperplano divide $\mathbb{R}^n$ em dois *semi-espaços*. Um semi-espaço (fechado) é um conjunto da forma

$$\left\{x \middle| a^T x \leq b\right\}, \tag{10}$$

onde $a \in \mathbb{R}^n, a \neq 0$ e $b \in \mathbb{R}$.

Bola euclidiana

Um *bola euclidiana* em $\mathbb{R}^n$, centrada em x_c e de raio r, é o conjunto

$$B(x_c, r) = \{x \| \|x - x_c\|_2 \leq r\}, \tag{11}$$

onde $r > 0$ e $\|.\|_2$ denota a norma euclidiana.

Bola da norma

Seja $\|.\|$ uma norma em $\mathbb{R}^n$. Uma *bola da norma* em $\mathbb{R}^n$, centrada em x_c e de raio r, é o conjunto

$$B(x_c, r) = \{x | \|x - x_c\| \leq r\}, \tag{12}$$

onde $r > 0$.

Cone da norma

Um *cone da norma* associado à norma $\|.\|$ é o conjunto

$$C = \{(x,t) | \|x\| \leq t\} \subseteq \mathbb{R}^{n+1}. \tag{13}$$

Politopo

Um *poliedro*, ou *politopo*, é o conjunto solução de um número finito de equações e desigualdades lineares:

$$P = \left\{ x \middle| a_i^T x \leq b_i, i = 1, \ldots, m; c_j^T x = d_j, j = 1, \ldots, p \right\}. \tag{14}$$

Em outros termos, é a interseção de um número finito de semi-espaços e hiperplanos.

É conveniente usar a forma compacta

$$P = \{x | Ax \preceq b, Cx = d\}, \tag{15}$$

onde

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}, \quad C = \begin{bmatrix} c_1^T \\ \vdots \\ c_p^T \end{bmatrix}, \tag{16}$$

e o símbolo $\preceq$ denota desigualdade elemento a elemento.

Politopo: representação de fecho convexo

Uma forma equivalente de descrever um politopo P é como o fecho convexo de um conjunto finito de pontos $C = \{x_1, \ldots, x_k\}$:

$$P = \mathrm{conv}(C) = \left\{ \theta_1 x_1 + \cdots + \theta_k x_k \middle| \theta \succeq 0, \mathbf{1}^T \theta = 1 \right\}. \qquad (17)$$

Matrizes simétricas

Denota-se por S^n o conjunto das *matrizes $n \times n$ simétricas*:

$$S^n = \left\{ X \in \mathbb{R}^{n \times n} \middle| X = X^T \right\}. \tag{18}$$

Este conjunto constitui um espaço vetorial real de dimensão $n(n+1)/2$.

Matrizes positivas semi-definidas

Denota-se por S^n_+ o conjunto das *matrizes $n \times n$ simétricas positivas semi-definidas*:

$$S^n_+ = \{X \in S^n | X \succeq 0\}, \tag{19}$$

onde $X \succeq 0$ denota que, para todo $y \in \mathbb{R}^n$, $y^T X y \geq 0$.

Cone das matrizes positivas semi-definidas

O conjunto S_+^n é um cone, ou seja, se $\theta_1, \theta_2 \geq 0$ e $A, B \in S_+^n$, então

$$\theta_1 A + \theta_2 B \in S_+^n. \tag{20}$$

Teorema do hiperplano separador

Sejam C e D dois conjuntos convexos que não se intersectam. Então existem $a \neq 0$ e b tais que $a^T x \leq b$ para todo $x \in C$ e $a^T y \geq b$ para todo $y \in D$.

O hiperplano $\left\{x \middle| a^T x = b\right\}$ é dito *hiperplano separador* para os conjuntos C e D.

Função convexa

Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é *convexa* se $\mathrm{dom}\, f$ é um conjunto convexo e se, para todo $x, y \in \mathrm{dom}\, f$, e $0 \leq \theta \leq 1$:

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y). \tag{21}$$

A função f é *estritamente convexa* se vale a desigualdade estrita para $x \neq y$ e $0 < \theta < 1$.

Função côncava

Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é (estritamente) *côncava* se $-f$ é (estritamente) convexa.

Exemplos

- Exponencial: e^{ax} é convexa em $\mathbb{R}$, para todo $a \in \mathbb{R}$.
- Potências: x^a é convexa em $\mathbb{R}_{++}$ quando $a \geq 1$ ou $a \leq 0$, e côncava quando $0 \leq a \leq 1$.
- Potências do valor absoluto: $|x|^p$, para $p \geq 1$, é convexa em $\mathbb{R}$.
- Logaritmo: $\log(x)$ é côncava em $\mathbb{R}_{++}$.
- Normas: toda norma em $\mathbb{R}^n$ é convexa.
- Função max: $f(x) = \max\{x_1, \ldots, x_n\}$ é convexa em R^n.
- Média geométrica: $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ é côncava em $\mathbb{R}^n_{++}$.
- Log-determinante: $f(X) = \log(\det(X))$ é côncava em S^n_{++}.

Problema de otimização

Um *problema de otimização* pode ser formulado como

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{sujeito a} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p. \end{aligned}$$

Definições

- $x \in \mathbb{R}^n$ é a *variável de otimização*;
- $f_0 : \mathbb{R}^n \to \mathbb{R}$ é a *função objetivo*;
- $f_i : \mathbb{R}^n \to \mathbb{R}$ são as *funções dos vínculos de desigualdade*;
- $h_i : \mathbb{R}^n \to \mathbb{R}$ são as *funções dos vínculos de igualdade*;

Definições

- o *domínio* do problema de otimização é

$$D = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i; \tag{22}$$

- um ponto $x \in D$ é dito *factível* se satisfaz todos os vínculos;
- o *valor ótimo* do problema é definido como

$$p^\star = \inf \{f_0(x) | f_i(x) \leq 0, i = 1, \ldots, m; h_i(x) = 0, i = 1, \ldots, p\}; \tag{23}$$

- um ponto $x^\star$ é dito *ótimo* se é factível e $f_0(x^\star) = p^\star$.

Problema de factibilidade

Um *problema de factibilidade* é formulado como

$$\begin{array}{ll} \text{encontre} & x \\ \text{sujeito a} & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p. \end{array}$$

Problema de otimização convexa

Um *problema de otimização convexa* é um problema de otimização da forma

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{sujeito a} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & a_i^T x = b_i, i = 1, \dots, p; \end{aligned}$$

onde $f_0, \dots, f_m$ são funções convexas.

Propriedade importante: todo ponto ótimo local é um ponto ótimo.

Programação linear

Um *programa linear (LP)* é um problema de otimização da forma

$$\begin{aligned} \text{minimize} \quad & c^T x + d \\ \text{sujeito a} \quad & Gx \preceq h \\ & Ax = b; \end{aligned}$$

onde $G \in \mathbb{R}^{m \times n}$ e $A \in \mathbb{R}^{p \times n}$.

Programas lineares são problemas de otimização convexa.

Programação semidefinida

Um *programa semi-definido (SDP)* é um problema de otimização da forma

$$\begin{aligned} \text{minimize} \quad & \operatorname{Tr}(CX) \\ \text{sujeito a} \quad & \operatorname{Tr}(A_i X) = b_i, \quad i = 1, \dots, p \\ & X \succeq 0; \end{aligned}$$

onde $X, C, A_1, \dots, A_p \in S^n$, e $\operatorname{Tr}(.)$ denota o traço de uma matriz quadrada, definido como a soma dos elementos de sua diagonal principal.

Programas semidefinidos também são problemas de otimização convexa.

Lagrangiano

Considere um problema de otimização

$$\begin{array}{rll} \text{minimize} & f_0(x) & \\ \text{sujeito a} & f_i(x) \leq 0, & i = 1, \dots, m \\ & h_i(x) = 0, & i = 1, \dots, p; \end{array}$$

onde $x \in \mathbb{R}^n$. Define-se o *Lagrangiano* $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associado ao problema como

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x). \tag{24}$$

As variáveis $\lambda \in \mathbb{R}^m$ e $\nu \in \mathbb{R}^p$ são chamadas *multiplicadores de Lagrange*.

Função dual de Lagrange

Dado um Lagrangiano L associado a um problema de otimização, sua *função dual (de Lagrange)* $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ é definida como

$$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu). \tag{25}$$

Para todo $\lambda \succeq 0$ e todo ν, $g(\lambda, \nu) \leq p^\star$.

O problema dual de Lagrange

O *problema dual de Lagrange*, dado um problema *primal*, é

$$\begin{array}{rl} \text{maximize} & g(\lambda, \nu) \\ \text{sujeito a} & \lambda \succeq 0. \end{array}$$

O problema dual é sempre um problema de otimização convexa, independentemente do problema primal.

Dualidade fraca

Sejam $p^\star$ o valor ótimo do problema primal e $d^\star$ o valor ótimo do problema dual. Então

$$d^\star \leq p^\star. \tag{26}$$

Dualidade fraca vale mesmo que o problema primal não seja convexo.

Dualidade forte

Em geral, quando o problema primal é convexo, vale a *dualidade forte*

$$d^\star = p^\star. \tag{27}$$

Assim, para problemas de otimização convexa, em geral é possível saber se o valor ótimo foi alcançado ou não resolvendo-se ambos os problemas primal e dual, e avaliando-se a diferença entre as soluções obtidas.