

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

JOSÉ CARLOS RUBIANES SILVA

**Método Subgradiente Condicional com
Sequência Ergódica**

Goiânia
2011

JOSÉ CARLOS RUBIANES SILVA

Método Subgradiente Condicional com Sequência Ergódica

Dissertação apresentada ao Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de concentração: Matemática Aplicada.

Orientador: Prof. Dr. Jefferson D. Gonçalves de Melo

Goiânia
2011

Dados Internacionais de Catalogação na Publicação (CIP)
GPT/BC/UFG

S586m	<p>Silva, José Carlos Rubianes. Silva, José Carlos Rubianes. Método Subgradiente Condicional com Sequência Ergódica [manuscrito] / José Carlos Rubianes Silva. - 2011. 92 f. : il.</p> <p>Orientador: Prof. Dr. Jefferson D. Gonçalves de Melo. Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística, 2011. Bibliografia.</p> <p>1. Programação Convexa. 2. Métodos Subgradientes. 3. Subgradiente Condicional. 4. Relaxação Lagrangiana. 5. Dualidade Lagrangiana. 6. Convergência Ergódica. I. Título.</p> <p style="text-align: right;">CDU: 531.314</p>
-------	--

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

José Carlos Rubianes Silva

Graduou-se em Matemática na UNAC - Universidad Nacional del Callao - Lima - Perú. Atualmente é aluno de Doutorado da Unicamp - Universidade Estadual de Campinas.

Aos meus pais José Carlos e Magda e meu irmão Ivan.

Agradecimentos

Meus sinceros agradecimentos:

Ao Professor Jefferson Melo, pela orientação, disponibilidade ao esclarecimento de todas as dúvidas e sugestões, bem como à paciência, ao incentivo, à confiança, à dedicação e à amizade demonstrada.

Aos Professores José Yunier Bello Cruz, Luis Román e Orizon Pereira Ferreira, pelo apoio e as sugestões dadas durante as apresentações nos seminários.

Agradeço os membros da banca, especialmente, Professor Maicon Marques Alves pelos comentários e sugestões.

A meus amigos e colegas do programa, com os quais formamos um bom grupo de estudos ao longo destes anos, resolvendo as listas, discutindo os problemas e acrescentando juntos nossos conhecimentos.

A meu amigo Milton Gabriel, com quem fizemos algumas matérias do mestrado juntos, e quem é um dos melhores amigos que eu tenho aqui no Brasil, e com quem discuti muitos problemas de otimização.

À Universidade federal de Goiás, em particular, ao programa de Pós-Graduação em Matemática, pela oportunidade que me concederam para participar do Mestrado, pela disponibilidade do local de estudo e de suas instalações.

Agradeço à Capes pelo apoio financeiro.

"Se as pessoas não acreditam que a Matemática é simples, é porque elas não perceberam o quão complicado é a vida".

Jonh Von Neumann.
Matemático

Resumo

Silva, José Carlos Rubianes. **Método Subgradiente Condicional com Sequência Ergódica**. Goiânia, 2011. 92p. Dissertação de Mestrado. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Nesta dissertação consideramos um problema de otimização convexo e estudamos variações do método subgradiente aplicado ao problema dual obtido via uma função Lagrangiana. Estudamos o método subgradiente condicional desenvolvido por Larsson et al, o qual é uma simples variação do método subgradiente “usual”. A principal diferença é que os subgradientes são condicionados a um conjunto restrição, mais especificamente, o comportamento da função fora do conjunto restrição não é levado em conta. Uma motivação para estudar tais métodos consiste principalmente na sua simplicidade, em especial, estes métodos são bastante usados em problemas de grande porte. O método subgradiente, quando aplicado a um problema dual, é relativamente eficaz para obter boas aproximações de soluções duais e do valor ótimo, no entanto, não possui a mesma eficiência para obter soluções primais. Analisamos uma estratégia para obter boas aproximações de soluções primais via método subgradiente condicional, com pouco custo computacional adicional. Esta estratégia consiste em construir uma sequência ergódica das soluções obtidas durante a resolução dos subproblemas Lagrangianos. Mostraremos que os pontos limites desta sequência ergódica são soluções primais. Consideramos diferentes regras para o tamanho do passo, em particular, seguindo as idéias de Nedic e Ozdaglar, apresentamos estimativas da sequência ergódica com o conjunto de soluções primais e/ou o conjunto viável quando usamos a regra de passos constantes.

Palavras-chave

Programação convexa, Métodos subgradientes, Subgradiente condicional, Relaxação Lagrangiana, Dualidade Lagrangiana, Convergência ergódica.

Abstract

Silva, José Carlos Rubianes. **Método Subgradiente Condicional com Sequência Ergódica**. Goiânia, 2011. 92p. MSc. Dissertation. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

In this dissertation we consider a primal convex optimization problem and we study variants of subgradient method applied to the dual problem obtained via a Lagrangian function. We analyze the conditional subgradient method developed by Larsson et al, which is a variant of the “usual” subgradient method. In this variant, the subgradients are conditioned to a constraint set, more specifically, the behavior of the objective function outside of the constraint set is not taken into account. One motivation for studying such methods is primarily its simplicity, in particular, these methods are widely used in large-scale problems. The subgradient method, when applied to a dual problem, is relatively effective to obtain a good approximation of a dual solution and the optimal value, but it is not efficient to obtain primal solutions. We study a strategy to obtain good approximations of primal solutions via conditional subgradient method, under suitable additional computational costs. This strategy consists of constructing an ergodic sequence of solutions of the Lagrangian subproblems. We show that the limit points of this ergodic sequence are primal solutions. We consider different step sizes rule, in particular, following the ideas of Nedic and Ozdaglar, using the constant step size rule, we present estimates of the ergodic sequence and primal solutions and / or the feasible set.

Keywords

Convex programming, Subgradient methods, Conditional subgradient, Lagrangean relaxation, Lagrangean duality, Ergodic convergence.

Sumário

1	Introdução	3
2	Convexidade e dualidade Lagrangiana	5
2.1	Análise convexa	5
2.1.1	Conjuntos convexos	5
2.1.2	Funções convexas	10
2.2	Relaxação Lagrangiana	17
2.3	Dualidade Lagrangiana	18
2.3.1	Relaxação Lagrangiana e o problema dual	18
2.3.2	Condições de otimalidade global	22
2.3.3	Dualidade forte para programas convexos	22
2.4	Exemplos	25
2.5	Propriedades de diferenciabilidade da função dual	27
2.5.1	Subdiferenciabilidade de funções convexas	27
2.5.2	Diferenciabilidade da função dual Lagrangiana	30
3	Método Subgradiente Condicional	33
3.1	Subgradiente Condicional	33
3.2	Algoritmo Subgradiente Condicional	37
3.3	Método subgradiente com projeção	47
3.3.1	Análise do método	49
4	Método subgradiente condicional com sequência ergódica	54
4.1	Método subgradiente dual	54
4.1.1	Dualidade Lagrangiana	55
4.2	Convergência do método subgradiente dual	59
4.2.1	Convergência ergódica primal	63
5	Soluções primais duais aproximadas	69
5.1	Problemas primal e dual	69
5.1.1	Qualificação de Slater e conjunto de multiplicadores limitado	70
5.2	Método Subgradiente	72
5.2.1	Multiplicadores limitados	73
5.3	Soluções primais aproximadas	76
5.3.1	Propriedades básicas da sequência ergódica primal	76
5.3.2	Propriedades da sequência ergódica primal sob a qualificação de Slater	79
5.4	Método subgradiente modificado sob a qualificação de Slater	81
6	Conclusões	90

Referências Bibliográficas

91

Introdução

Muitos problemas de otimização provenientes de diversas aplicações consistem em minimizar uma função convexa não diferenciável. Tais problemas podem ser encontrados, por exemplo, em redes neurais, problemas de alocação, entre outros. Para resolvê-los temos que empregar métodos especiais que funcionam na ausência de diferenciabilidade e conseguem explorar a convexidade e possivelmente outras estruturas especiais que o problema possa ter, por exemplo, estrutura separável das variáveis.

Em diferentes áreas da matemática, costuma-se associar a um determinado problema (problema primal) um segundo problema (o problema dual), o qual é, em geral, menos complicado de resolver que o problema primal. Tal procedimento constitui uma forma interessante e eficiente, tanto do ponto de vista teórico quanto computacional, para obter aproximações da solução do problema original. Claramente, na prática tal formulação dual é eficaz somente quando resolver o problema dual é extremamente mais simples que resolver o problema primal. Por meio de uma função Lagrangiana podemos associar a um programa de otimização convexo um problema dual. A função Lagrangiana é convexa (côncava) em relação a suas variáveis primais (duais), obtendo portanto um problema dual que consiste em maximizar a função dual (côncava) com restrições convexas simples (\mathbb{R}_+^m). As soluções para este par de programas primal-dual são pontos de sela da função Lagrangiana. As soluções primais podem ser aproximadas por meio dos subproblemas, que são geralmente mais simples de resolver que o problema original.

Recentemente, muitas ferramentas computacionais eficientes tem sido exploradas para obter soluções de problemas de programação não linear. Em particular, para resolver problemas de otimização convexas de grande porte com estrutura, a relaxação Lagrangiana juntamente com métodos do tipo subgradiente são utilizados de forma eficiente e são bastante usados na prática. A relaxação Lagrangiana engloba as restrições na função objetivo, tornando problemas restritos em problemas irrestritos, mantendo a estrutura de convexidade, além de originar problemas com estruturas separáveis na maioria dos problemas práticos de grande porte. Os métodos subgradientes são usados principalmente devido sua simplicidade. Tais métodos não são de descida, porém possuem

outra característica especial, a distância do iterado atual ao conjunto solução não é maior que a distância do conjunto solução ao iterado anterior. Outro método bastante usado na prática para resolver problemas convexos não diferenciáveis é o método de feixes, o qual é bastante eficiente porém possui uma complexidade computacional alta comparado com métodos subgradientes. Não estudaremos este método aqui, veja [5] para um estudo aprofundado do método de feixe.

Apesar de sua popularidade, o método subgradiente dual, em geral, não obtém solução para o problema primal, o qual é o nosso principal interesse. A principal razão está ligada ao fato que a função dual é tipicamente não suave em uma solução ótima dual e portanto uma solução ótima primal é em geral uma combinação convexa das imagens das funções restrições aplicadas nas soluções do subproblema Lagrangiano. Além do mais, as soluções do subproblema dual, em geral, não são viáveis para o problema original. Uma estratégia interessante para obter solução primal via métodos do tipo subgradiente aplicados ao problema dual é usar uma sequência ergódica primal. Tal estratégia foi proposta para problemas lineares por Shor [23], e estendida para programas convexos, não necessariamente lineares, por Larsson et al. [18]. Nedic e Ozdaglar [20], estudaram o método subgradiente dual com sequência ergódica no caso em que a regra do comprimento de passo constante é empregada, estimativas entre a sequência ergódica e o conjunto de soluções primais e o conjunto viável são apresentadas. Os principais resultados apresentados nesta dissertação foram obtidos por Larsson et al. [18] e Nedic e Ozdaglar [20].

A organização da dissertação é a seguinte: No Capítulo 2, apresentamos alguns resultados básicos de análise convexa e dualidade Lagrangiana. No Capítulo 3, apresentamos o método subgradiente condicional e suas principais propriedades de convergência. Consideramos uma escolha particular de subgradientes condicionais, obtidos por projeções relativamente simples, obtendo um esquema de fácil implementação. No Capítulo 4, estudamos o método subgradiente condicional aplicado ao problema dual Lagrangiano de um problema convexo. Construímos uma sequência ergódica de soluções do subproblema Lagrangiano e apresentamos suas principais propriedades de convergência. No Capítulo 5, estudamos o método subgradiente com sequência ergódica, onde usamos a regra de comprimento de passo constante. Primeiro, mostramos que a sequência dual gerada pelo método subgradiente é limitada sob a condição de regularidade de Slater. Apresentaremos estimativas (por-iteração) para a violação das restrições, e cotas superior e inferior da função objetivo sobre a sequência ergódica primal. Por último, apresentamos nossa conclusão e possíveis trabalhos futuros.

Convexidade e dualidade Lagrangiana

Este capítulo reúne alguns resultados básicos sobre convexidade e dualidade Lagrangiana, em especial os resultados relacionados aos programas convexos com ausência de brecha de dualidade.

2.1 Análise convexa

2.1.1 Conjuntos convexos

Definição 2.1 (*Conjunto convexo*) Seja $\Omega \subset \mathbb{R}^n$. O conjunto Ω é convexo se para quaisquer $x, y \in \Omega$ e $\alpha \in [0, 1]$, tem-se

$$\alpha x + (1 - \alpha)y \in \Omega.$$

O ponto $\alpha x + (1 - \alpha)y \in \Omega$, onde $\alpha \in [0, 1]$, se chama a *combinação convexa* de x e y (com parâmetro α).

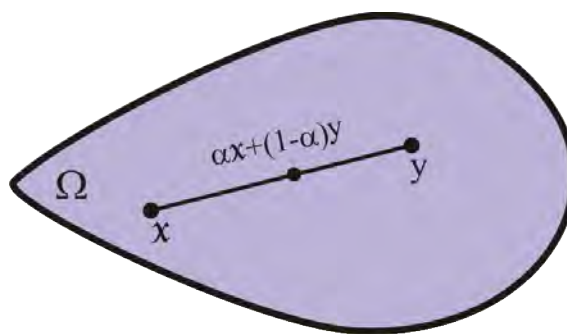


Figura 2.1: Ilustração da definição de um conjunto convexo.

Exemplo 2.2 (*Conjuntos convexos e não convexos*) Da definição de um conjunto convexo, podemos estabelecer que

- (a) O conjunto vazio é um conjunto convexo.
- (b) O conjunto $\{x \in \mathbb{R}^n / \|x\| \leq r\}$ é convexo para cada valor de $r \in \mathbb{R}$.

- (c) Fixado $r > 0$, o conjunto $\{x \in \mathbb{R}^n / \|x\| = r\}$ é não convexo.
 (d) O conjunto $\{0, 1, 2\} \subset \mathbb{R}$ não é convexo.

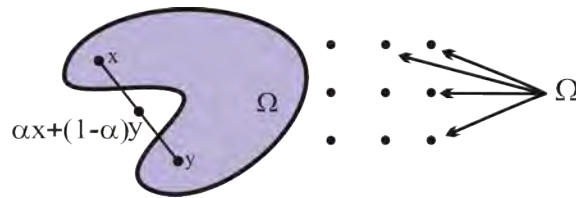


Figura 2.2: Dois conjuntos não convexos.

Proposição 2.3 (Interseção convexa) Suponha que Ω_λ , $\lambda \in I$, é qualquer coleção de conjuntos convexos. Então a interseção,

$$\Omega = \bigcap_{\lambda \in I} \Omega_\lambda$$

é um conjunto convexo.

Demonstração. Ver [12, Proposição 3.2.1]. □

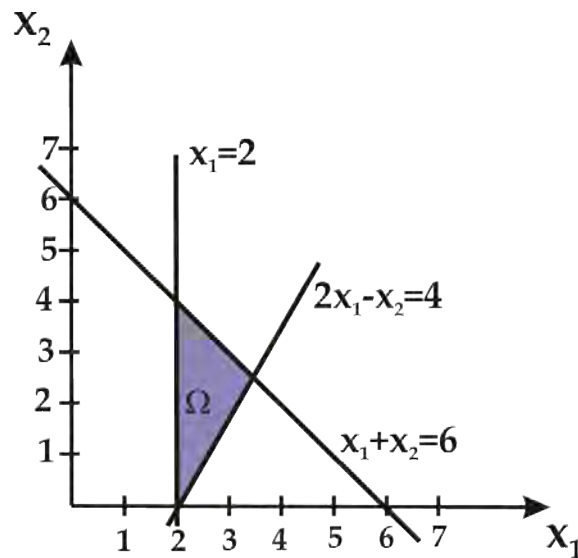


Figura 2.3: Ilustração do poliedro limitado P .

Definição 2.4 (Conjunto poliedral) Um conjunto $P \subset \mathbb{R}^n$ é um poliedro se existe uma matriz $A \in \mathbb{R}^{m \times n}$ e um vetor $b \in \mathbb{R}^m$ tal que

$$P = \{x \in \mathbb{R}^n / Ax \leq b\}.$$

A importância de poliedros é óbvia, uma vez que o conjunto de soluções viáveis de todo problema de programação linear é um poliedro.

Exemplo 2.5 (Poliedro)

(a) A Figura 2.3 mostra o poliedro limitado

$$P = \{ x \in \mathbb{R}^2 / x_1 \geq 2; x_1 + x_2 \leq 6; 2x_1 - x_2 \leq 4 \}.$$

(b) O poliedro ilimitado

$$P = \{ x \in \mathbb{R}^2 / x_1 \geq 2; x_1 + x_2 \leq 2; 3x_1 - x_2 \geq 0 \} \text{ é mostrado na Figura 2.4.}$$

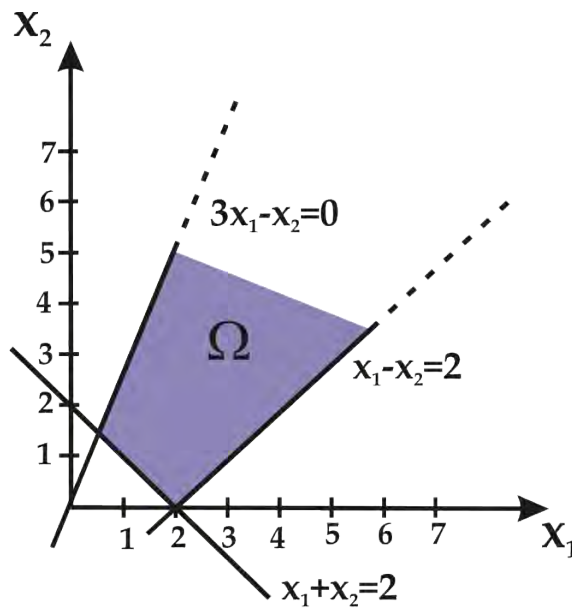


Figura 2.4: Ilustração do poliedro ilimitado P .

Definição 2.6 (Cone) Um conjunto $K \subset \mathbb{R}^n$ é um cone quando

$$x \in K \implies tx \in K, \quad \forall t \in \mathbb{R}_+.$$

Em adição, se K é convexo então K é chamado cone convexo.

Observação 2.7 Pela Definição (2.6), se K é um cone não vazio então necessariamente $0 \in K$. Intuitivamente, K é um conjunto de direções.

Exemplo 2.8 Alguns exemplos de cone são: o espaço \mathbb{R}^n , qualquer subespaço vetorial de \mathbb{R}^n , o octante não negativo \mathbb{R}_+^n .

Exemplo 2.9 (Cone)

(a) O conjunto $\{ x \in \mathbb{R}^n / Ax \leq 0 \}$, onde $A \in \mathbb{R}^{m \times n}$, é um cone. Como este conjunto é um poliedro, este tipo de cone em geral é chamado de cone poliedral.

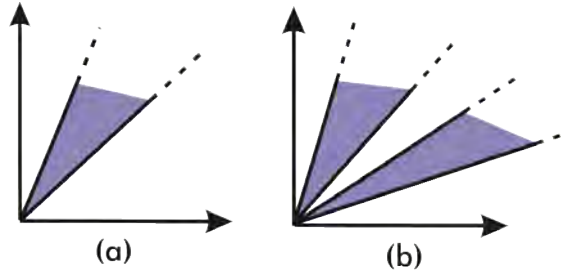


Figura 2.5: (a) Um cone convexo em \mathbb{R}^2 . (b) Um cone não convexo em \mathbb{R}^2 .

(b) A Figura 2.5(a) ilustra um cone convexo e Figura 2.5(b) ilustra um cone não convexo

Definição 2.10 (*Direção viável*) Dizemos que $d \in \mathbb{R}^n$ é uma direção viável em relação ao conjunto Ω no ponto $\bar{x} \in \Omega$, quando existe $\varepsilon > 0$ tal que

$$\bar{x} + td \in \Omega \quad \forall t \in [0, \varepsilon].$$

Definimos a distância entre um ponto $x \in \mathbb{R}^n$ e o conjunto $\Omega \subset \mathbb{R}^n$ como

$$\text{dist}(x, \Omega) = \inf_{y \in \Omega} \|y - x\|.$$

Definição 2.11 (*Direção tangente*) Dizemos que $d \in \mathbb{R}^n$ é uma direção tangente em relação ao conjunto Ω no ponto $\bar{x} \in \Omega$ quando

$$\text{dist}(\bar{x} + td, \Omega) = o(t), \quad t \in \mathbb{R}_+,$$

onde $o(t)$ possui a propriedade de que $o(t)/t \rightarrow 0$ quando $t \rightarrow 0^+$.

O conjunto de todas as direções tangentes é chamado o cone tangente a Ω no ponto $\bar{x} \in \Omega$, denotado por $T_\Omega(\bar{x})$.

Definição 2.12 (*Cone dual*) O cone dual (ou cone polar) de um cone $K \subset \mathbb{R}^n$ é definido por

$$K^* = \{ y \in \mathbb{R}^n / \langle y, d \rangle \leq 0, \quad \forall d \in K \}.$$

Definição 2.13 (*Cone normal*) sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e $\bar{x} \in \Omega$. O cone normal no ponto \bar{x} em relação ao conjunto Ω é dado por

$$N_\Omega(\bar{x}) = \{ d \in \mathbb{R}^n / \langle d, x - \bar{x} \rangle \leq 0, \quad \forall x \in \Omega \}.$$

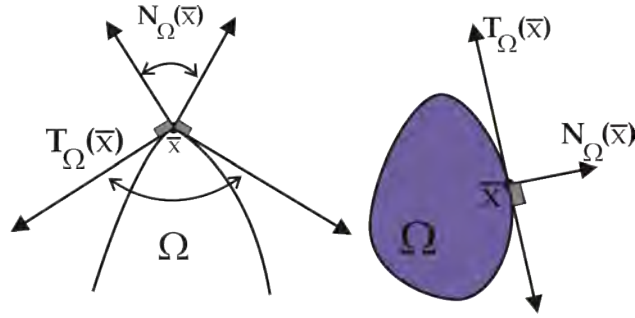


Figura 2.6: Exemplos do cone normal de um conjunto convexo.

A seguir, mostramos que no caso convexo, o dual do cone tangente é exatamente o cone normal definido acima. Veja Figura (2.6).

Dizemos que Ω é regular no ponto \bar{x} quando:

$$N_{\Omega}(\bar{x}) = T_{\Omega}(\bar{x})^*.$$

Proposição 2.14 *Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo não vazio. Então para todo $\bar{x} \in \Omega$, temos*

$$x \in T_{\Omega}(\bar{x})^* \text{ se, e somente se, } x^t(\bar{x} - x) \geq 0, \quad \forall \bar{x} \in \Omega.$$

Além disso, Ω é regular para todo $\bar{x} \in \Omega$. Em particular, temos

$$T_{\Omega}(\bar{x})^* = N_{\Omega}(\bar{x}), \quad T_{\Omega}(\bar{x}) = N_{\Omega}(\bar{x})^*.$$

Demonstração. Ver [4, Proposição 4.6.3]. □

Teorema 2.15 *(Teorema de separação) Sejam $\Omega_1 \subset \mathbb{R}^n$ e $\Omega_2 \subset \mathbb{R}^n$ conjuntos convexos não vazios tais que $\Omega_1 \cap \Omega_2 = \emptyset$. Então existem $a \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tais que*

$$\langle a, x_1 \rangle \leq c \leq \langle a, x_2 \rangle \quad \forall x_1 \in \Omega_1, \quad \forall x_2 \in \Omega_2.$$

Demonstração. Ver [12, Teorema 3.2.6] □

Teorema 2.16 *(Teorema de separação estrita) Sejam $\Omega_1 \subset \mathbb{R}^n$ e $\Omega_2 \subset \mathbb{R}^n$ conjuntos convexos, fechados e não vazios. Suponhamos que um deles também seja limitado.*

Então $\Omega_1 \cap \Omega_2 = \emptyset$ se, e somente se, existem $a \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tais que

$$\langle a, x_1 \rangle < c < \langle a, x_2 \rangle \quad \forall x_1 \in \Omega_1, \quad \forall x_2 \in \Omega_2.$$

Demonstração. Ver [12, Teorema 3.2.7] □

Exemplo 2.17 (ilustração do teorema de separação) Considere o conjunto fechado e convexo $\Omega = \{x \in \mathbb{R}^2 / \|x\| \leq 1\}$ e o ponto $y = (1.5, 1.5)$. Como $y \notin \Omega$ pelo Teorema de separação temos que existe uma reta em \mathbb{R}^2 que separa y de Ω . Esta reta no entanto não é única. Ver figura 2.7.

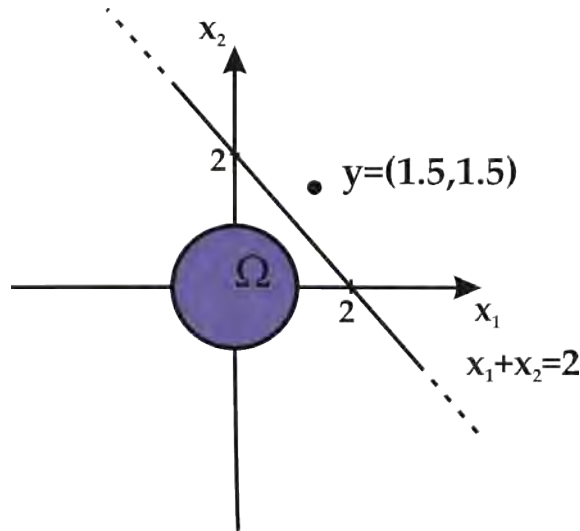


Figura 2.7: Ilustração do Teorema de separação.

2.1.2 Funções convexas

Definição 2.18 (Função convexa)

Se $\Omega \subset \mathbb{R}^n$ é um conjunto convexo, diz-se que a função $f : \Omega \rightarrow \mathbb{R}$ é convexa em Ω quando para quaisquer $x \in \Omega$, $y \in \Omega$ e $\alpha \in [0, 1]$, tem-se

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

A função diz-se estritamente convexa quando a desigualdade acima é estrita para todos os $x \neq y$ e $\alpha \in (0, 1)$.

Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é côncava (resp. estritamente côncava) quando $-f$ é convexa (resp. estritamente convexa).

Exemplo 2.19 (Funções convexas) Usando a definição podemos estabelecer que:

- (a) A função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por $f(x) = \|x\|$ é convexa em \mathbb{R}^n .
- (b) Seja $c \in \mathbb{R}^n$, $a \in \mathbb{R}$. A função afim $f(x) = c^t x + a = \sum_{j=1}^n c_j x_j + a$ é convexa e côncava sobre \mathbb{R}^n .

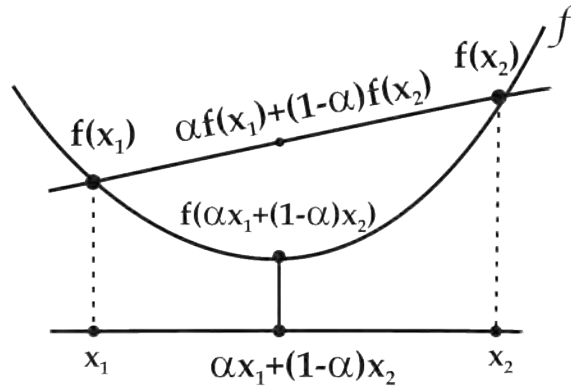


Figura 2.8: Uma função convexa.

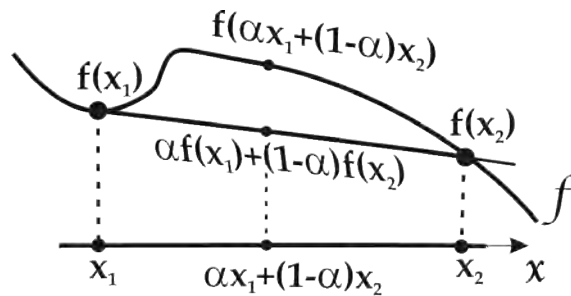


Figura 2.9: Uma função não convexa.

Proposição 2.20 (*Soma de funções convexas*) Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e $f_i : \Omega \rightarrow \mathbb{R}$, $i = 1, \dots, p$, funções convexas em Ω . Então para quaisquer $\alpha_i \in \mathbb{R}_+$, $i = 1, \dots, p$, a função

$$f : \Omega \rightarrow \mathbb{R}, \quad f(x) = \sum_{i=1}^p \alpha_i f_i(x)$$

é convexa em Ω .

Demonstração. Ver [12, Proposição 3.4.1]. □

Proposição 2.21 (*Convexidade da composição de funções*) Sejam $g : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa, $h : \mathbb{R} \rightarrow \mathbb{R}$ uma função convexa e não decrescente. Então a função

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad f(x) = h(g(x))$$

é convexa.

Demonstração. Ver [12, Proposição 3.4.3] □

Exemplo 2.22 *Suponha que a função g é convexa.*

(a) A função $x \rightarrow -\log(-g(x))$ é convexa sobre o conjunto $\{x \in \mathbb{R}^n / g(x) < 0\}$.

(b) A função $x \rightarrow -1/g(x)$ é convexa sobre o conjunto $\{x \in \mathbb{R}^n / g(x) < 0\}$.

(c) A função $x \rightarrow 1/\log(-g(x))$ é convexa sobre o conjunto $\{x \in \mathbb{R}^n / g(x) < -1\}$.

Definição 2.23 Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo. Dizemos que $g : \Omega \rightarrow \mathbb{R}^m$ é convexa em Ω se todas as funções $g_j : \Omega \rightarrow \mathbb{R}$, $j = 1, \dots, m$, são convexas em Ω .

Teorema 2.24 (Continuidade de funções convexas) Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : \Omega \rightarrow \mathbb{R}$ uma função convexa em Ω . Então f é localmente Lipschitz-contínua em Ω . Em particular, f é contínua em Ω .

Demonstração. Ver [12, Teorema 3.4.2]. □

Em particular, se $f : \Omega \rightarrow \mathbb{R}$ é convexa no conjunto convexo $\Omega \subset \mathbb{R}^n$, então f é contínua no interior de Ω .

Exemplo 2.25 Sejam $\Omega = [0, 1]$ e $f : \Omega \rightarrow \mathbb{R}$,

$$f(x) = \begin{cases} 2, & \text{se } x = -1 \text{ ou } 1, \\ x^2, & \text{se } -1 < x < 1. \end{cases}$$

Note que f é convexa em Ω , mas f não é contínua nos pontos $x = 1$ e $x = -1$ (na fronteira do seu domínio); veja a Figura 2.10.

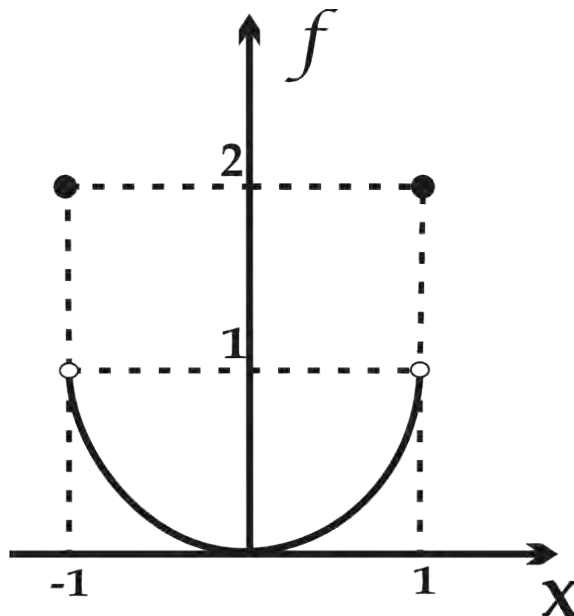


Figura 2.10: Ilustração do Exemplo 2.25.

Definição 2.26 Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é denominada *semi-continua inferior* (resp. *superior*) em um ponto $x \in \mathbb{R}^n$ se, para todo $\{x^k\} \subset \mathbb{R}^n$ com $x^k \rightarrow x$ tem-se

$$f(x) \leq \liminf_{k \rightarrow \infty} f(x^k) \quad \left(\text{resp. } f(x) \geq \limsup_{k \rightarrow \infty} f(x^k) \right).$$

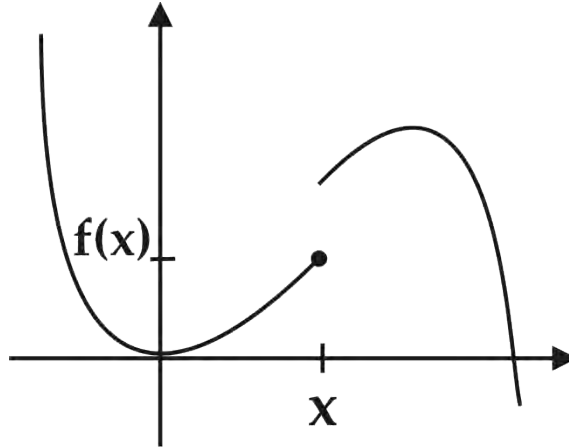


Figura 2.11: Função semi-continua inferior.

Definição 2.27 Considere o conjunto $\Omega \subset \mathbb{R}^n$ não vazio. A função $\delta_\Omega : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ definida por

$$\delta_\Omega(x) = \begin{cases} 0, & x \in \Omega \\ +\infty, & x \notin \Omega \end{cases} \quad (2-1)$$

é denominada *função indicadora de Ω* .

Note que Ω é um conjunto convexo e fechado se, e somente se, δ_Ω é uma função convexa e semi-continua inferior.

Observação 2.28 Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo. Uma função convexa $f : \Omega \rightarrow \mathbb{R}$, pode ser estendida a uma função convexa $\hat{f} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ da seguinte forma:

$$\hat{f}(x) = \begin{cases} f(x), & x \in \Omega \\ +\infty, & x \notin \Omega \end{cases} \quad (2-2)$$

Definição 2.29 Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é dita *convexa própria* quando $f(x) < +\infty$ para algum $x \in \mathbb{R}^n$.

Se f é uma função convexa finita, e Ω é um conjunto convexo, fechado e não vazio, então

$$f(x) + \delta_\Omega(x) = \begin{cases} f(x), & x \in \Omega \\ +\infty, & x \notin \Omega \end{cases} \quad (2-3)$$

é semi-contínua inferior, própria e convexa.

Considere o problema

$$\begin{aligned} \min f(x) \\ \text{s.a } x \in \Omega, \end{aligned} \quad (2-4)$$

Note que o problema (2-4) é equivalente a:

$$\begin{aligned} \min f(x) + \delta_{\Omega}(x) \\ \text{s.a } x \in \mathbb{R}^n. \end{aligned} \quad (2-5)$$

- Definição 2.30** (a) Uma combinação afim de elementos x_1, \dots, x_k de \mathbb{R}^n é um elemento da forma $\sum_{i=1}^k \alpha_i x_i$, onde $\alpha_i \geq 0$, $i = 1, \dots, k$.
- (b) Uma variedade afim em \mathbb{R}^n é um conjunto contendo todas suas combinações afins. É fácil verificar que a interseção de variedades afins é ainda uma variedade afim.
- (c) Seja $S \subset \mathbb{R}^n$ um conjunto não vazio. O fecho afim de S , denotado por $\text{aff}(S)$, é a variedade afim gerada por S , isto é, a interseção de todas as variedades afins que contém S .

Observação 2.31 O fecho afim de S é a menor variedade afim que contém S e pode ser construída diretamente de S colecionando todas as combinações afins de elementos de S .

Definição 2.32 Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo. O interior relativo de Ω , denotado por $\text{rint}(\Omega)$ é o interior de Ω com relação a topologia relativa do fecho afim de Ω ; isto é, $x \in \text{rint}(\Omega)$ se, e somente se, $x \in \text{aff}(\Omega)$ e existe $\delta > 0$ tal que $\text{aff}(\Omega) \cap B(x, \delta) \subset \Omega$.

- Exemplo 2.33** (a) Se $\Omega = \{x\}$, com $x \in \mathbb{R}^n$, então $\text{aff}(\Omega) = \{x\}$ e $\text{rint}(\Omega) = \{x\}$.
- (b) Se $\Omega = B(x_0, \delta)$, com $x_0 \in \mathbb{R}^n$ e $\delta > 0$, então $\text{aff}(\Omega) = \mathbb{R}^n$ e $\text{rint}(\Omega) = B(x_0, \delta)$.

Considere agora um conjunto $\Omega \subset \mathbb{R}^n$, um ponto $x \in \mathbb{R}^n$ e o problema de encontrar um ponto de Ω mais próximo de x . Este problema pode não ter solução e quando tem, não garantimos unicidade. No entanto, conforme provaremos a seguir, se Ω é fechado, então existe solução. Se além de fechado, for convexo, a solução é única e será chamada de projeção de x sobre Ω , denotada por $P_{\Omega}(x)$. Veja ilustração na Figura (2.12).

Definição 2.34 Definiremos a projeção de $x \in \mathbb{R}^n$ sobre Ω , do seguinte modo:

$$P_{\Omega}(x) := \arg \min \{ \|y - x\| : y \in \Omega \}. \quad (2-6)$$

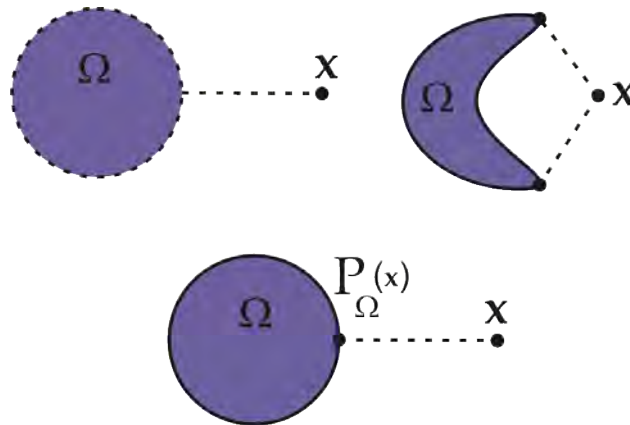


Figura 2.12: *Projeção sobre um conjunto.*

Teorema 2.35 (*Teorema da Projeção*)

- (a) *Seja $\Omega \subset \mathbb{R}^n$ um conjunto fechado e não-vazio. Então para todo ponto $x \in \mathbb{R}^n$, existe uma projeção de x sobre Ω .*
- (b) *Se, além de ser fechado, o conjunto Ω for convexo, então para todo ponto $x \in \mathbb{R}^n$, existe uma única projeção de x sobre Ω .*
- (c) *Um ponto $\bar{x} \in \Omega$ é a projeção de $P_\Omega(x)$ se, e somente se,*

$$\langle x - \bar{x}, y - \bar{x} \rangle \leq 0, \quad \text{para todo } y \in \Omega.$$

Demonstração. Veja [12, Teorema 3.2.4]. □

A seguir, mostramos que o operador de projeção é não expansivo (i.e., Lipschitz-contínuo com módulo um), veja a Figura (2.13).

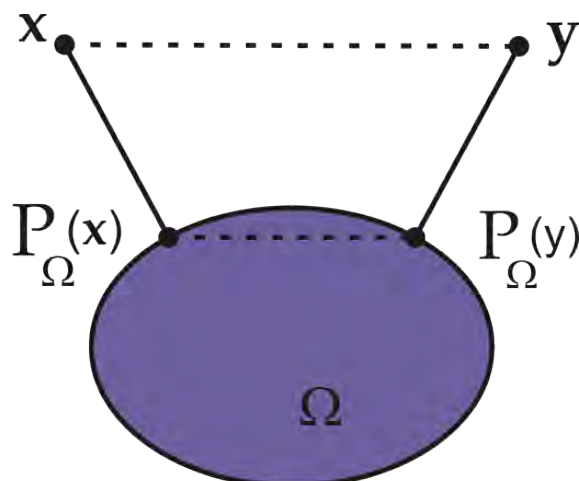


Figura 2.13: *O operador de projeção é não expansivo.*

Proposição 2.36 (*Operador de projeção é não expansivo*) *Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo e fechado. Então para $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^n$ quaisquer,*

$$\|P_{\Omega}(x) - P_{\Omega}(y)\| \leq \|x - y\|.$$

Demonstração. Ver [12, Corolário 3.2.3]. □

A seguir apresentamos algumas propriedades do operador de projeção sobre um cone convexo e fechado.

Proposição 2.37 *Seja K um cone convexo e fechado. Então $y_x = P_K(x)$ se, e somente se,*

$$y_x \in K, \quad x - y_x \in K^*, \quad \langle x - y_x, y_x \rangle = 0.$$

Demonstração. Ver [11, Proposição 3.2.1]. □

Algumas propriedades que são consequência da Proposição (2.37) para todo $x \in \mathbb{R}^n$ são,

$$P_{\Omega}(x) = 0 \quad \text{se, e somente se,} \quad x \in \Omega^*;$$

$$P_{\Omega}(\alpha x) = \alpha P_{\Omega}(x) \quad \text{para todo} \quad \alpha \geq 0;$$

$$P_{\Omega}(-x) = -P_{-\Omega}(x).$$

Teorema 2.38 *(Decomposição de Moreau) Seja K um cone convexo e fechado. Para $x, x_1, x_2 \in \mathbb{R}^n$, as seguintes propriedades são equivalentes:*

- (a) $x = x_1 + x_2$ com $x_1 \in K$, $x_2 \in K^*$ e $\langle x_1, x_2 \rangle = 0$;
- (b) $x_1 = P_K(x)$ e $x_2 = P_{K^*}(x)$.

Demonstração. Ver [11, Teorema 3.2.5]. □

2.2 Relaxação Lagrangiana

Consideremos o problema

$$f^* = \min f(x) \quad (2-7a)$$

$$s.a \ x \in \Omega \quad (2-7b)$$

onde $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função dada e $\Omega \subset \mathbb{R}^n$, definimos a relaxação de (2-7) como sendo o problema

$$f_R^* = \min f_R(x) \quad (2-8a)$$

$$s.a \ x \in \Omega_R \quad (2-8b)$$

onde $f_R: \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função com a propriedade que $f_R \leq f$ sobre Ω , e onde $\Omega \subset \Omega_R$.

Para estes dois problemas temos o seguinte resultado básico.

Teorema 2.39 (*Teorema de Relaxação*)

- (a) [relaxação] $f_R^* \leq f^*$.
- (b) [inviabilidade] Se (2-8) é não viável, então também o é (2-7).
- (c) [relaxação ótima] Se o problema (2-8) tem solução x_R^* , o qual satisfaz que

$$x_R^* \in \Omega \quad e \quad f_R(x_R^*) = f(x_R^*), \quad (2-9)$$

então x_R^* é uma solução para (2-7) também.

Demonstração. O resultado em (a) é óbvio, como toda solução viável em (2-7) é também viável em (2-8) e tem valor funcional inferior no último problema. O resultado em (b) segue por razões semelhantes. Para obter o resultado em (c), note que

$$f(x_R^*) = f_R(x_R^*) \leq f_R(x) \leq f(x), \quad x \in \Omega$$

a partir do qual o resultado segue. □

2.3 Dualidade Lagrangiana

Nesta seção, vamos formular o problema dual Lagrangiano e estabelecer a sua convexidade. O Teorema da Dualidade Fraca também é estabelecido, e discutiremos um pouco sobre multiplicadores de Lagrange e "brecha de dualidade".

2.3.1 Relaxação Lagrangiana e o problema dual

Considere o problema de otimização

$$f^* = \min f(x) \quad (2-10a)$$

$$s.a \ x \in \Omega \quad (2-10b)$$

$$g_j(x) \leq 0, \quad j = 1, 2, \dots, m, \quad (2-10c)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ ($j = 1, 2, \dots, m$) são funções dadas, e $\Omega \subset \mathbb{R}^n$. Para este problema, assumimos que

$$-\infty < f^* < \infty, \quad (2-11)$$

i.e., f é limitado inferiormente e que o problema tem ao menos uma solução viável.

Para um vetor arbitrário $\mu \in \mathbb{R}^m$, definimos a função Lagrangiana

$$L(x, \mu) = f(x) + \sum_{j=1}^m \mu_j g_j(x) = f(x) + \mu^t g(x). \quad (2-12)$$

O vetor $\mu^* \in \mathbb{R}^m$ é dito um multiplicador de Lagrange se for não negativo e satisfaz $f^* = \inf_{x \in \Omega} L(x, \mu^*)$.

Teorema 2.40 (*Multiplicadores de Lagrange*) *Seja μ^* um multiplicador de Lagrange. Então x^* é uma solução ótima para (2-10) se, e somente se, x^* é viável em (2-10) e*

$$x^* \in \arg \min_{x \in \Omega} L(x, \mu^*), \quad e \quad \mu_j^* g_j(x^*) = 0, \quad j = 1, 2, \dots, m. \quad (2-13)$$

Demonstração. Se x^* é uma solução ótima para (2-10), então em particular é viável, e

$$f^* = f(x^*) \geq L(x^*, \mu^*) \geq \inf_{x \in \Omega} L(x, \mu^*) = f^*,$$

onde as desigualdades decorrem da viabilidade de x^* e a definição de multiplicador de Lagrange. Assim, (2-13) se verifica.

Reciprocamente, se x^* é viável e vale (2-13), então usando a definição de um multiplicador de Lagrange,

$$f(x^*) = L(x^*, \mu^*) = \inf_{x \in \Omega} L(x, \mu^*) = f^*,$$

obtemos que x^* é uma solução ótima. \square

Seja

$$q(\mu) = \inf_{x \in \Omega} L(x, \mu) \quad (2-14)$$

a função dual Lagrangiana, definida pelo valor ínfimo da função Lagrangiana sobre Ω ; o problema dual Lagrangiano é

$$\max q(\mu), \quad (2-15a)$$

$$s.a. \mu \geq 0. \quad (2-15b)$$

Observe que é possível que $q(\mu) = -\infty$ para algum μ . Se isto é verdade para todos os $\mu \geq 0$, então

$$\sup q(\mu),$$

$$s.a. \mu \geq 0.$$

é igual a $-\infty$.

O domínio efetivo de q é

$$D_q = \{ \mu \in \mathbb{R}^m / q(\mu) > -\infty \}.$$

Teorema 2.41 (*Problema dual convexo*) O domínio efetivo D_q de q é convexo, e q é côncavo sobre D_q .

Demonstração. Sejam $x \in \mathbb{R}^n$, $\mu, \bar{\mu} \in \mathbb{R}^m$, e $\alpha \in [0, 1]$. temos que

$$L(x, \alpha\mu + (1 - \alpha)\bar{\mu}) = \alpha L(x, \mu) + (1 - \alpha)L(x, \bar{\mu}).$$

Tome o ínfimo sobre $x \in \Omega$ em ambos lados; logo,

$$\begin{aligned} \inf_{x \in \Omega} L(x, \alpha\mu + (1 - \alpha)\bar{\mu}) &= \inf_{x \in \Omega} \{ \alpha L(x, \mu) + (1 - \alpha)L(x, \bar{\mu}) \} \\ &\geq \inf_{x \in \Omega} \alpha L(x, \mu) + \inf_{x \in \Omega} (1 - \alpha)L(x, \bar{\mu}) \\ &= \alpha \inf_{x \in \Omega} L(x, \mu) + (1 - \alpha) \inf_{x \in \Omega} L(x, \bar{\mu}), \end{aligned}$$

devido ao fato de que $\alpha \in [0, 1]$, e que a soma dos valores ínfimo pode ser menor que o ínfimo da soma. Por isso,

$$q(\alpha\mu + (1 - \alpha)\bar{\mu}) \geq \alpha q(\mu) + (1 - \alpha)q(\bar{\mu}).$$

Esta desigualdade tem duas implicações: se μ e $\bar{\mu}$ pertencem a D_q , então $\alpha\mu + (1 - \alpha)\bar{\mu}$, também pertence a D_q , e ainda, q é côncava sobre D_q . \square

Como o problema dual Lagrangiano é sempre convexo (maximização de uma função côncava) com restrições convexas simples, ele pode ser resolvido de forma eficiente. No capítulo seguinte veremos como uma solução ótima dual pode ser usada para obter uma solução ótima primal.

Teorema 2.42 (*Teorema da Dualidade Fraca*) *Sejam x e μ viáveis em (2-10) e (2-15), respectivamente. Então,*

$$q(\mu) \leq f(x).$$

Em particular vale,

$$q^* \leq f^*$$

Se $q(\mu) = f(x)$, então o par (x, μ) é ótimo no seu respectivo problema.

Demonstração. Para todo $\mu \geq 0$ e $x \in \Omega$ com $g(x) \leq 0$,

$$q(\mu) = \inf_{x \in \Omega} L(x, \mu) \leq f(x) + \mu^t g(x) \leq f(x),$$

assim

$$q^* = \sup_{\mu \geq 0} q(\mu) \leq \inf_{x \in \Omega: g(x) \leq 0} f(x) = f^*.$$

\square

Dualidade fraca é também uma consequência do Teorema de Relaxação: Para qualquer $\mu \in \mathbb{R}_+^m$, seja

$$S = \Omega \cap \{ x \in \mathbb{R}^n / g(x) \leq 0 \}, \quad (2-17a)$$

$$S_R = \Omega, \quad (2-17b)$$

$$f_R = L(., \mu). \quad (2-17c)$$

Então, o Teorema de dualidade fraca é o resultado da parte (a) do Teorema 2.39, de onde relaxação Lagrangiana é uma relaxação em termos da definição na Seção 2.3.

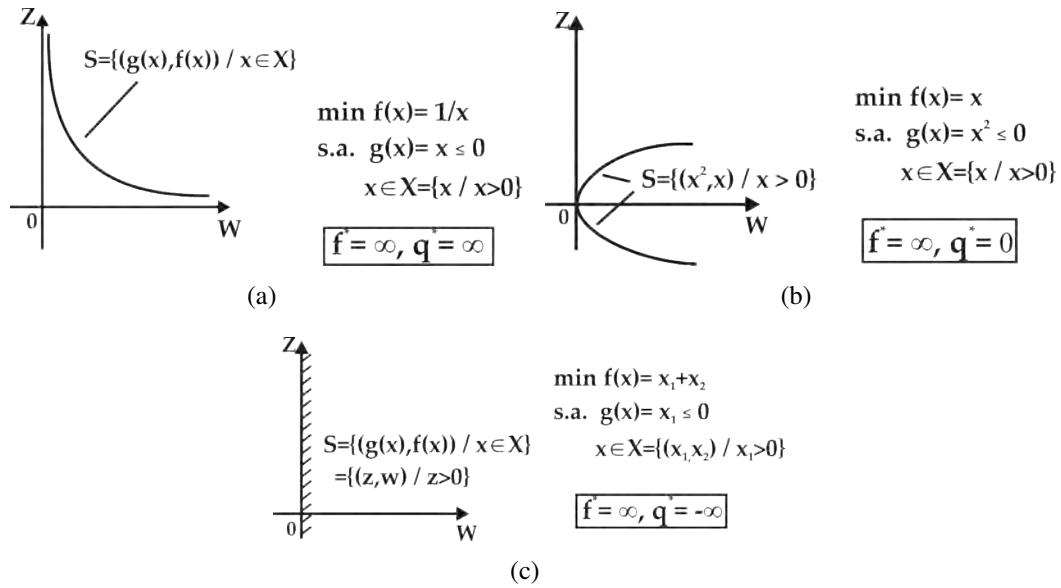


Figura 2.14: Exemplos onde Ω é não vazio mas o problema primal é não viável.

Se nossa hipótese inicial (2-11), é falsa, então o que implica a dualidade fraca? Suponha que $f^* = -\infty$. Então dualidade fraca implica que $q(\mu) = -\infty$ para todo $\mu \in \mathbb{R}_+^m$, i.e., o problema dual é não viável. Suponhamos então que $\Omega \neq \emptyset$ mas que $\Omega \cap \{x \in \mathbb{R}^n / g(x) \leq 0\}$ é vazio. Então, $f^* = \infty$, por convenção. A função dual satisfaz $q(\mu) < \infty$ para todo $\mu \in \mathbb{R}_+^m$, mas é possível que $q^* = -\infty$, $-\infty < q^* < \infty$, ou $q^* = \infty$. Ver Figuras 2.14(a), 2.14(b) e 2.14(c).

Se $q^* = f^*$ dizemos que não há brecha de dualidade. Se existe um vetor de multiplicadores de Lagrange, pelo teorema da dualidade fraca, isto implica que não há brecha de dualidade. O recíproco não é verdadeiro em geral: pode haver casos em que não existe multiplicador de Lagrange, mesmo quando não há brecha de dualidade, nesse caso, porém, o problema dual Lagrangiano não pode ter uma solução ótima, como mostra a seguinte proposição.

Proposição 2.43 (Brecha de dualidade e a existência de multiplicadores de Lagrange)

- (a) Se não houver brecha de dualidade, então o conjunto de multiplicadores de Lagrange é igual ao conjunto de soluções ótimas duais (que no entanto pode ser vazia).
- (b) Se houver brecha de dualidade, então não existem multiplicadores de Lagrange.

Demonstração. Pela definição, um vetor $\mu^* \in \mathbb{R}_+^m$ é um multiplicador de Lagrange se, e somente se, $f^* = q(\mu^*) \leq q^*$, a igualdade segue da definição de $q(\mu^*)$ e a desigualdade da definição de q^* sobre \mathbb{R}_+^m . Pela dualidade fraca, esta relação é válida se, e somente se, não há brecha de dualidade e μ^* é uma solução ótima dual. □

2.3.2 Condições de otimalidade global

O resultado seguinte caracteriza cada solução ótima primal e dual. É aplicável somente na presença de multiplicadores de Lagrange; em outras palavras, o sistema (2-18) é consistente se, e somente se, existe um multiplicador de Lagrange e não há brecha de dualidade.

Teorema 2.44 (*Condições de otimalidade global na ausência de brecha de dualidade*)
O vetor (x^*, μ^*) é um par de solução ótima primal e um multiplicador de Lagrange se, e somente se,

$$\mu^* \geq 0 \quad , \quad (\text{Viabilidade dual}) \quad (2-18a)$$

$$x^* \in \arg \min_{x \in \Omega} L(x, \mu^*) \quad , \quad (\text{Otimalidade Lagrangiana}) \quad (2-18b)$$

$$x^* \in \Omega, \quad g(x^*) \leq 0 \quad , \quad (\text{Viabilidade primal}) \quad (2-18c)$$

$$\mu_j^* g_j(x^*) = 0, \quad j = 1, 2, \dots, m \quad . \quad (\text{Condição de complementaridade}) \quad (2-18d)$$

Demonstração. Suponhamos que vale (2-18). Aplicamos o Teorema de relaxação (2.39) como segue, consideremos a relação (2-17), com $\mu = \mu^* \in \mathbb{R}_+^m$ e $x_R = x^*$. Em seguida, observe as seguintes equivalências:

1. A solução relaxada x_R , é uma solução ótima Lagrangiana. (*Otimalidade Lagrangiana é satisfeita*).
2. Que $x_R \in \Omega$, significa que ele é viável para o problema primal. (*Viabilidade Primal é satisfeita*).
3. Que $f^* = f_R^*$, significa que $(\mu^*)^t g(x^*) = 0$. (*Condição de complementaridade é satisfeita*).

Para concluir, se as condições em (2-18) são satisfeitas, então o Teorema 2.39 implica que o vetor (x^*, μ^*) é um par de solução ótima primal e um multiplicador de Lagrange.

Reciprocamente, se (x^*, μ^*) é um par de solução ótima primal e um multiplicador de Lagrange, então eles são obviamente primal viável e dual viável, respectivamente. As duas últimas equações em (2-18) seguem do Teorema 2.40. \square

2.3.3 Dualidade forte para programas convexos

Vimos como a dualidade fraca permite formar um problema de otimização convexa que proporciona uma cota inferior para o problema primal original, mesmo quando este não é convexo.

Dizemos que o problema satisfaz a dualidade forte, se não há brecha de dualidade. Mais importante, porém, é que a dualidade forte está profundamente associada com a convexidade do problema original, e é em especial no âmbito da convexidade que as soluções ótimas primais e duais estão ligadas através das condições de otimalidade global proporcionadas na seção anterior. Começamos por analisar o caso de restrições de desigualdade, provando este resultado com mais detalhe.

Consideremos o problema convexo (2-10), onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ são funções convexas e $\Omega \subset \mathbb{R}^n$ é um conjunto convexo. Para este problema, temos que definir uma qualificação das restrições.

Definição 2.45 (*Qualificação de Slater*) Dizemos que o problema (2-10) com Ω convexo, f e g_j convexas sobre Ω , satisfaz a condição de qualificação de Slater quando existir $x \in \Omega$ tal que

$$g_j(x) < 0 \quad \text{para todo } j = 1, 2, \dots, m. \quad (2-19)$$

Teorema 2.46 (*Dualidade Forte*) Suponha que valem (2-11) e a qualificação de Slater (2-19) para o problema convexo (2-10). Então

- (a) Não há brecha de dualidade entre os problemas primal e dual. Além disso, existe pelo menos um multiplicador de Lagrange μ^* .
- (b) Se o mínimo em (2-10) é atingido para algum x^* , então o par (x^*, μ^*) satisfaz as condições de otimalidade global (2-18).

Demonstração. (a) Começamos por estabelecer a existência de um multiplicador de Lagrange (e ausência de brecha de dualidade).

Primeiro, consideramos o seguinte subconjunto de \mathbb{R}^{m+1} :

$$A = \{ (z_1, \dots, z_m, w)^t / \exists x \in \Omega \text{ com } g_j(x) \leq z_j, \quad j = 1, \dots, m; \quad f(x) \leq w \}.$$

É fácil verificar que A é convexo. Observe que $(0^t, f^*)^t$, $0 \in \mathbb{R}^n$ não é um ponto interior de A ; caso contrário, para algum $\varepsilon > 0$ vale que o ponto $(0^t, f^* - \varepsilon)^t \in A$ o que é uma contradição pela definição de f^* . Portanto, pelo Teorema de separação 2.15, podemos encontrar um hiperplano passando por $(0^t, f^*)^t$ tal que A está em um dos dois semi espaços. Em particular, existe um vetor $(\mu^t, \beta)^t \neq (0^t, 0)^t$ tal que

$$\beta f^* \leq \beta w + \mu^t z, \quad (z^t, w)^t \in A. \quad (2-20)$$

Isso implica que

$$\beta \geq 0, \quad \mu \geq 0, \quad (2-21)$$

como para cada $(z^t, w)^t \in A$ temos que

$$(z^t, w + \gamma)^t \in A \text{ e } (z_1, \dots, z_{j-1}, z_j + \gamma, z_{j+1}, \dots, z_m, w)^t \in A \quad \forall \gamma > 0 \text{ e } j = 1, \dots, m.$$

Afirmamos que $\beta > 0$. De fato, se não fosse o caso, então $\beta = 0$ e (2-20) implicaria que $\mu^t z \geq 0$ para todo par $(z^t, w)^t \in A$. Mas como $(g(\bar{x}^t), 0)^t \in A$, onde \bar{x} é tal que satisfaz a qualificação de Slater, obtemos que

$$0 \leq \sum_{j=1}^m \mu_j g_j(\bar{x}),$$

como por (2-21) $\mu \geq 0$ e por hipótese \bar{x} satisfaz a qualificação de Slater, isto implica que $\mu = 0$. Isto significa, no entanto, que

$$(\mu^t, \beta)^t = (0^t, 0)^t,$$

chegando a uma contradição. Portanto, podemos afirmar que $\beta > 0$. Nós também, sem perda de generalidade, podemos supor que $\beta = 1$.

Assim, como $(g(x)^t, f(x))^t \in A$ para cada $x \in \Omega$, de (2-20) temos

$$f^* \leq f(x) + \mu^t g(x), \quad x \in \Omega.$$

Tomando o ínfimo sobre $x \in \Omega$ e usando o fato que $\mu \geq 0$ obtemos

$$f^* \leq \inf_{x \in \Omega} \{f(x) + \mu^t g(x)\} = q(\mu) \leq \sup_{\mu \geq 0} q(\mu) = q^*.$$

Usando o Teorema de Dualidade Fraca (2.42) segue que μ é um vetor multiplicador de Lagrange, e não existe brecha de dualidade.

(b) O resultado segue do Teorema 2.44. □

2.4 Exemplos

Exemplo 2.47 (Um problema dual diferenciável explícito)

Consideremos o problema

$$\begin{aligned} \min f(x) &= x_1^2 + x_2^2, \\ \text{s.a. } x_1 + x_2 &\geq 4, \\ x_j &\geq 0, \quad j = 1, 2. \end{aligned}$$

Definimos $g(x) = -x_1 - x_2 + 4$ e $\Omega = \{ (x_1, x_2) / x_j \geq 0, j = 1, 2 \}$.

Então a função dual Lagrangiana é

$$\begin{aligned} q(\mu) &= \min_{x \in \Omega} L(x, \mu) = f(x) - \mu(x_1 + x_2 - 4) \\ &= 4\mu + \min_{x \in \Omega} \{x_1^2 + x_2^2 - \mu x_1 - \mu x_2\} \\ &= 4\mu + \min_{x_1 \geq 0} \{x_1^2 - \mu x_1\} + \min_{x_2 \geq 0} \{x_2^2 - \mu x_2\}, \quad \mu \geq 0. \end{aligned}$$

Para um $\mu \geq 0$ fixo, o mínimo é atingido em $x_1(\mu) = \frac{\mu}{2}$, $x_2(\mu) = \frac{\mu}{2}$.

Substituindo esta expressão em $q(\mu)$, obtemos que

$$q(\mu) = f(x(\mu)) - \mu(x_1(\mu) + x_2(\mu) - 4) = 4\mu - \frac{\mu^2}{2}.$$

Note que q é estritamente côncava e diferenciável, (devido ao fato de que f e g são diferenciáveis e $x(\mu)$ é único), pelo item (d) do Teorema de Danskin 2.52.

Temos então que

$$q'(\mu) = 4 - \mu = 0 \iff \mu = 4.$$

Como $\mu \geq 4$, este é o ótimo no problema dual, logo

$$\mu^* = 4 \quad \text{e} \quad x^* = (x_1(\mu^*), x_2(\mu^*))^t = (2, 2)^t,$$

e também $f(x^*) = q(\mu^*) = 8$.

Este é um exemplo onde a função dual é diferenciável, neste caso o ótimo x^* também é único.

Exemplo 2.48 (Um problema dual implícito não diferenciável) Consideremos o problema linear

$$\begin{aligned} \min f(x) &= -x_1 - x_2, \\ \text{s.a. } 2x_1 + 4x_2 &\leq 3, \\ 0 &\leq x_1 \leq 2, \\ 0 &\leq x_2 \leq 1. \end{aligned}$$

Vemos que a solução é $x^* = (\frac{3}{2}, 0)^t$, e $f(x^*) = -\frac{3}{2}$.

Consideremos a relaxação Lagrangiana

$$\begin{aligned} L(x, \mu) &= -x_1 - x_2 + \mu(2x_1 + 4x_2 - 3); \\ q(\mu) &= -3\mu + \min_{0 \leq x_1 \leq 2} \{(-1 + 2\mu)x_1\} + \min_{0 \leq x_2 \leq 1} \{(-1 + 4\mu)x_2\} \\ &= \begin{cases} -3 + 5\mu, & 0 \leq \mu \leq 1/4, \\ -2 + \mu, & 1/4 \leq \mu \leq 1/2, \\ -3\mu, & 1/2 \leq \mu. \end{cases} \end{aligned}$$

Vemos que $\mu^* = \frac{1}{2}$, e, portanto, que $q(\mu^*) = -\frac{3}{2}$. Para problemas lineares, temos dualidade forte, mas como podemos obter a solução ótima primal a partir de μ^* ? É claro que q não é diferenciável em μ^* . Vamos utilizar a caracterização dada no sistema (2-18).

Primeiro, em μ^* , é claro que o conjunto solução do subproblema Lagrangiano

$$\Omega(\mu^*) = \{x \in \Omega / f(x) + \mu^* g(x) = q(\mu^*)\}$$

é o conjunto $\{(2\alpha, 0)^t / 0 \leq \alpha \leq 1\}$.

Entre as soluções do subproblema Lagrangiano, temos que encontrar uma que seja primal viável, bem como satisfazer a condição de complementaridade.

Viabilidade primal significa que $2.2\alpha + 2.0 \leq 3 \iff \alpha \leq 3/4$.

Além disso, a complementaridade significa que

$$\mu^* \cdot (2x_1^* + 4x_2^* - 3) = 0 \iff \alpha = 3/4, \text{ desde que } \mu^* \neq 0.$$

Nós concluímos que o único vetor primal x que satisfaz o sistema (2-18) juntamente com a solução dual ótima $\mu^* = 1/2$ é $x^* = (3/2, 0)^t$ com $f^* = q^*$.

No primeiro exemplo, a função dual Lagrangiana é diferenciável desde que $x(\mu)$ seja único. O segundo mostra que, do contrário, pode haver pontos de não diferenciabilidade da função q , os quais implicam a existência de mais de uma solução. Uma

consequência desse fato é que obter uma solução primal ótima torna-se mais complexo.

2.5 Propriedades de diferenciabilidade da função dual

Nós estabelecemos que o problema dual Lagrangiano (2-15) é convexo, e ainda que em algumas circunstâncias, podemos gerar uma solução ótima dual que tem o mesmo valor ótimo do problema original. Passamos agora a estudar o problema Lagrangiano dual em detalhes, e em particular, como ele pode ser resolvido de forma eficiente. Primeiro, vamos estabelecer quando a função dual q é diferenciável. Veremos que a diferenciabilidade se satisfaz apenas em alguns casos especiais. Em muitos problemas práticos e teóricos, a função q é não diferenciável. Em tais situações, o método subgradiente pode ser uma alternativa bastante interessante. A seguir apresentamos a noção de subgradiente de uma função convexa e algumas propriedades básicas.

2.5.1 Subdiferenciabilidade de funções convexas

Ao longo desta seção supomos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa, e estudamos suas propriedades de subdiferenciabilidade.

Definição 2.49 (Subgradiente) *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Dizemos que um vetor $\gamma \in \mathbb{R}^n$ é um subgradiente de f em $x \in \mathbb{R}^n$ se*

$$f(y) \geq f(x) + \gamma'(y - x), \quad \forall y \in \mathbb{R}^n. \quad (2-24)$$

O conjunto de todos os subgradientes de f em x é denotado por $\partial f(x)$ e chamado de subdiferencial de f em x .

Observe a relação com a caracterização de uma função convexa de classe C^1 . É conhecido que $\partial f(x)$ contém mais de um elemento em um ponto não diferenciável. Assim como o gradiente gera hiperplanos suportando o gráfico de uma função convexa de classe C^1 , os subgradientes também gera hiperplanos suportando o gráfico de f , porém no caso não diferenciável existe mais de um vetor "suporte". Ilustramos este fato na Figura 2.15.

Observe também que um mínimo global x^* de f sobre \mathbb{R}^n fica caracterizado pelo fato que $0 \in \partial f(x^*)$.

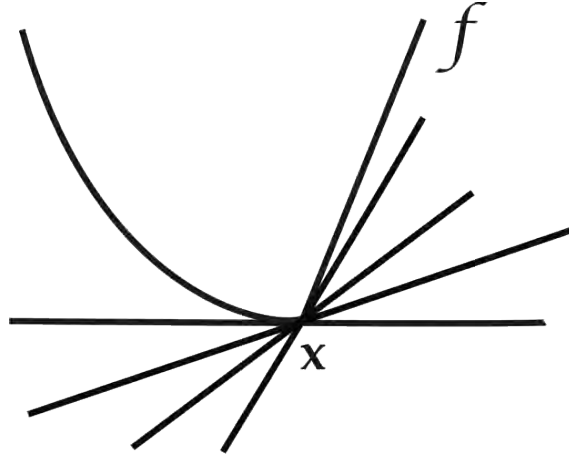


Figura 2.15: Três possíveis hiperplanos da função convexa f em x .

Definição 2.50 (Derivada direcional) A derivada direcional de f no ponto x na direção d é

$$f'(x; d) = \lim_{t \rightarrow 0^+} \frac{f(x + td) - f(x)}{t}.$$

Proposição 2.51 Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Um vetor x^* minimiza f sobre um conjunto convexo $\Omega \subset \mathbb{R}^n$ se, e somente se, existe um subgradiente $\gamma \in \partial f(x^*)$ tal que

$$\gamma^t(x - x^*) \geq 0, \quad \forall x \in \Omega.$$

Demonstração. Ver [4, Proposição 4.7.2]. □

Proposição 2.52 (Propriedades de uma função convexa) Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa.

- (a) Para cada $x \in \mathbb{R}^n$, $\partial f(x)$ é um conjunto não vazio, convexo e compacto. Se Ω é limitado então $\cup_{x \in \Omega} \partial f(x)$ é limitado.
- (b) A aplicação subdiferencial $x \mapsto \partial f(x)$ é uma aplicação fechada; i.e., se $\{x^k\}$ é uma sequência de vetores em \mathbb{R}^n convergindo a x , e vale que $\gamma^k \in \partial f(x^k)$ para cada k , então a sequência γ^k de subgradientes é limitada e cada ponto de acumulação pertence a $\partial f(x)$.
- (c) Para cada $x \in \mathbb{R}^n$ a derivada direcional de f em x na direção de $d \in \mathbb{R}^n$ satisfaz

$$f'(x; d) = \max_{\gamma \in \partial f(x)} \gamma^t d. \quad (2-25)$$

Em particular, f é diferenciável em x se, e somente se, o subdiferencial de f em x consiste de um único elemento, o $\nabla f(x)$. Neste caso,

$$f'(x; d) = \nabla f(x)^t d.$$

(d) [Teorema de Danskin] Sejam Z um subconjunto compacto de \mathbb{R}^m , e $\phi : \mathbb{R}^n \times Z \rightarrow \mathbb{R}$ contínua e tal que $\phi(\cdot, z) : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa para cada $z \in Z$. Seja a função $h : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

$$h(x) = \max_{z \in Z} \phi(x, z), \quad x \in \mathbb{R}^n. \quad (2-26)$$

A função h então é convexa sobre \mathbb{R}^n e tem derivada direcional em x na direção d igual a

$$h'(x; d) = \max_{z \in Z(x)} \phi'(x, z; d), \quad (2-27)$$

onde $\phi'(x, z; d)$ é a derivada direcional de $\phi(\cdot, z)$ em x na direção d , e $Z(x) = \{ z \in Z / \phi(x, z) = h(x) \}$.

Em particular, se $Z(x)$ contém um só ponto \bar{z} e $\phi(\cdot, \bar{z})$ é diferenciável em x , então h é diferenciável em x , e $\nabla h(x) = \nabla_x \phi(x, \bar{z})$, onde $\nabla_x \phi(x, \bar{z})$ é o vetor com componentes $\frac{\partial \phi(x, \bar{z})}{\partial x_i}$, $i = 1, \dots, n$.

Se ainda $\phi(x, z)$ é diferenciável para todo $z \in Z$ e $\nabla_x \phi(x, \cdot)$ é contínua sobre Z para cada x , então vale

$$\partial h(x) = \text{conv} \{ \nabla_x \phi(x, z) / z \in Z(x) \}, \quad x \in \mathbb{R}^n.$$

Demonstração. (a) Este é um caso especial de [22, Teorema 24.7].

(b) Ver [22, Teorema 24.5].

(c) Ver [22, Teorema 23.4 e 25.1].

(d) Ver [3, Proposição B.25]. □

Definição 2.53 Dizemos que $d \in \mathbb{R}^n$ é uma direção de descida de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto $x \in \mathbb{R}^n$, se existe $\varepsilon > 0$ tal que

$$f(x + td) < f(x), \quad \forall t \in (0, \varepsilon].$$

Denotamos por $D_f(x)$ o conjunto de todas as direções de descida da função f no ponto x .

Da parte (c) da Proposição 2.52, obtemos a seguinte condição que é necessária e suficiente para que $d \in \mathbb{R}^n$ seja uma direção de descida de f no ponto x :

$$d \in D_f(x) \iff \langle y, d \rangle < 0 \quad \forall y \in \partial f(x).$$

Uma consequência imediata disso é o seguinte fato, para obter uma direção de descida da função f num ponto dado x , necessitamos conhecer todo o conjunto subdiferencial $\partial f(x)$, o qual pela parte (a) da Proposição 2.52 é não vazio. Em particular, tomando um subgradiente qualquer, a direção oposta a um subgradiente pode não ser de descida, no entanto é bem conhecido que quando f é diferenciável a direção oposta ao gradiente de f é de descida.

Exemplo 2.54 Seja $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x) = |x_1| + 2|x_2|$. Consideremos um ponto $x = (x_1, 0)$, onde $x_1 > 0$ é arbitrário. Como é fácil de verificar,

$$y = (1, 2) \in \partial f(x).$$

No entanto, para todo $t > 0$ suficientemente pequeno,

$$f(x - ty) = |x_1 - t| + 2|0 - 2t| = x_1 - t + 4t = x_1 + 3t > x_1 = f(x),$$

o que mostra que $-y \notin D_f(x)$.

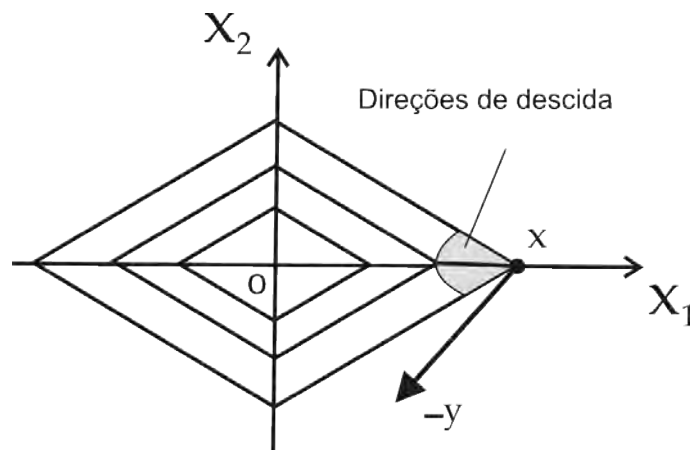


Figura 2.16: A direção $-y$ não é uma direção de descida.

2.5.2 Diferenciabilidade da função dual Lagrangiana

Nós consideraremos o problema de restrições de desigualdade (2-10), onde fazemos as seguintes hipóteses:

$$f, g_j, \quad j = 1, \dots, m, \quad \text{são funções contínuas; } \Omega \text{ é não vazio e compacto.} \quad (2-28)$$

Sob esta hipótese, o conjunto de soluções para o subproblema dual Lagrangiano,

$$\Omega(\mu) = \{ x \in \Omega / L(x, \mu) = q(\mu) \}, \quad \mu \in \mathbb{R}^m, \quad (2-29)$$

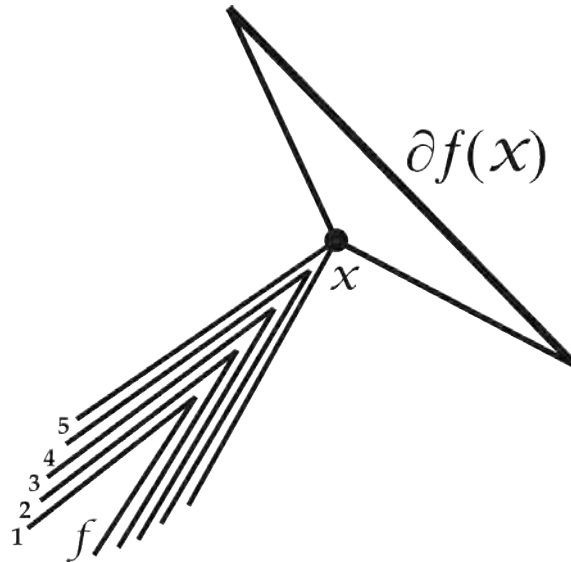


Figura 2.17: O subdiferencial de uma função convexa f em x .

é não vazio e compacto para qualquer escolha do vetor dual μ . Primeiro, desenvolvemos as propriedades de subdiferenciabilidade da função dual q , dada em (2-14). O primeiro resultado reforça o Teorema 2.41 sob algumas hipóteses adicionais. Com um abuso de notação, o conjunto dos elementos v , tal que $-v$ é um subgradiente da função convexa $-q$ em y , será chamado de subdiferencial da função côncava q em y . O subdiferencial de uma função côncava é também conhecido como superdiferencial.

Proposição 2.55 (Subdiferenciabilidade da função dual) *Suponha que no problema (2-10), valem as hipóteses em (2-28).*

- (a) *A função dual (2-14) é finita, contínua e côncava sobre \mathbb{R}^m . Se o seu supremo sobre \mathbb{R}_+^m é atingido, então o conjunto solução ótimo é fechado e convexo.*
- (b) *Seja $\mu \in \mathbb{R}^m$. Se $x \in \Omega(\mu)$, então $g(x)$ é um subgradiente para q em μ , i.e., $g(x) \in \partial q(\mu)$.*
- (c) *A derivada direcional de q em $\mu \in \mathbb{R}^m$ na direção de $d \in \mathbb{R}^m$ é*

$$q'(\mu; d) = \min_{\gamma \in \partial q(\mu)} d^t \gamma.$$

Demonstração. (a) Do Teorema 2.41 temos a côncavidade de q sobre seu domínio efetivo. Do Corolário [19, pág. 45] temos então que q é finito sobre \mathbb{R}^m o qual é também o

seu domínio efetivo.

(b) Seja $\mu, \bar{\mu} \in \mathbb{R}^m$ e $x \in \Omega(\mu)$. Temos que

$$\begin{aligned} q(\mu) &= \inf_{y \in \Omega} L(y, \mu) \leq f(x) + \mu^t g(x) \\ &= f(x) + (\mu - \bar{\mu})^t g(x) + \bar{\mu}^t g(x) \\ &= q(\bar{\mu}) + (\mu - \bar{\mu})^t g(x). \end{aligned}$$

o que implica que $g(x) \in \partial q(\mu)$.

(c) Ver Teorema de Danskin em 2.52 item (d). □

Em seguida, estabelecemos a diferenciabilidade da função dual sob hipóteses adicionais.

Proposição 2.56 (*Diferenciabilidade da função dual*) *Suponha que no problema (2-10), vale (2-28).*

(a) *Seja $\mu \in \mathbb{R}^m$. A função dual q é diferenciável em μ se e somente se o conjunto $\{g(x) / x \in \Omega(\mu)\}$ contém um único elemento, i.e., se o valor de $g(x)$ é invariável sobre o conjunto solução $\Omega(\mu)$ do subproblema Lagrangiano. Então temos que*

$$\nabla q(\mu) = g(x),$$

para cada $x \in \Omega(\mu)$.

(b) *O resultado em (a) é válido em particular se o subproblema Lagrangiano tem uma única solução, i.e., o conjunto $\Omega(\mu)$ contém um só elemento. Em particular, esta propriedade é satisfeita quando Ω é um conjunto convexo, f é estritamente convexa em Ω , e g_j , $j = 1, \dots, m$, são convexas, em cujo caso q sempre é de classe C1.*

Demonstração. (a) A função côncava q é diferenciável no ponto μ (onde é finita) se e somente se seu subdiferencial $\partial q(\mu)$ tem só um elemento, ver parte (b) da Proposição 2.55.

(b) Sob qualquer das hipóteses dadas, $\Omega(\mu)$ contém somente um elemento, onde o resultado segue de (a). A unicidade decorre da convexidade do conjunto viável e convexidade estrita da função primal, de acordo com [12, Proposição 3.1.5]. □

Método Subgradiente Condicional

Métodos subgradientes para minimizar funções convexas não diferenciáveis se originaram em um trabalho de N. Z. Shor no ano 1962; para o caso de otimização ir-restrita, Ermol'ev [9] provou a convergência do método usando comprimento de passos de séries divergentes, e Polyak [21] estendeu o método para um problema de otimização com restrições e apresentou várias propriedades adicionais. Estes métodos têm sido aplicados com sucesso na prática, principalmente em problemas duais via função Lagrangiana. Métodos subgradientes são também aplicados a problemas de relaxação para resolver sistemas de desigualdades lineares (Ver [10]). Neste capítulo apresentamos o método subgradiente condicional, o qual, como veremos, é uma generalização do método subgradiente (projetado) usual.

3.1 Subgradiente Condicional

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, uma função convexa não necessariamente diferenciável. Considere $\Omega \subseteq \mathbb{R}^n$ um conjunto não vazio, fechado e convexo. Assuma que existe $x^* \in \Omega$ uma solução do seguinte problema

$$f^* = \min_{x \in \Omega} f(x). \quad (3-1)$$

Denotaremos por

$$\Omega^* = \{x \in \Omega : f(x) = f^*\}, \quad (3-2)$$

o conjunto solução deste problema, o qual é convexo, não vazio e fechado (pois f é contínua). Note que a definição do subdiferencial (2.49) não leva em consideração o conjunto restrição, e sim todo o domínio da função. Dem'yanov e Shomesova [6], [7] generalizaram a definição do subdiferencial para que o conjunto restrição do problema (3-1) seja levado em consideração.

Definição 3.1 Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e $\Omega \subset \mathbb{R}^n$. O subdiferencial condicional de f em $x \in \Omega$ é o conjunto

$$\partial^\Omega f(x) = \{\gamma \in \mathbb{R}^n : f(y) \geq f(x) + \gamma^T (y-x), \text{ para todo } y \in \Omega\}. \quad (3-3)$$

Os elementos de $\partial^\Omega f(x)$ são ditos subgradientes condicionais de f em x .

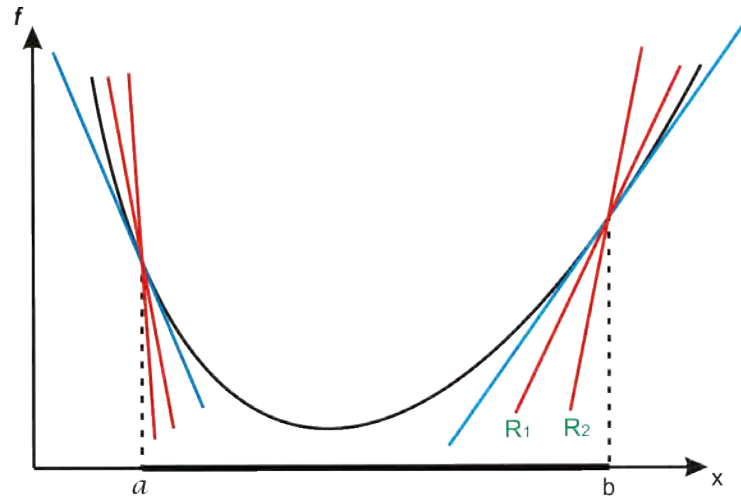


Figura 3.1: Os coeficientes angulares das retas R_1 e R_2 são subgradientes condicionais de f em b , onde $\Omega = [a, b]$.

Claramente $\partial f(x) \subseteq \partial^\Omega f(x)$ para todo $x \in \Omega$ e como $\partial f(x) \neq \emptyset$ então $\partial^\Omega f(x) \neq \emptyset$. O seguinte teorema contém propriedades básicas do subdiferencial condicional.

Teorema 3.2 O subdiferencial condicional $\partial^\Omega f(x)$ é não vazio, fechado e convexo para todo $x \in \Omega$.

Demonstração. O subdiferencial condicional é não vazio pois contém o subdiferencial usual, o qual é não vazio. As demais propriedades seguem da definição de subdiferencial condicional. □

As condições de otimalidade do problema (3-1) são dadas no seguinte teorema.

Teorema 3.3 (Condições de otimalidade)

- a) $x^* \in \Omega^*$ se, e somente se, $0 \in \partial^\Omega f(x^*)$.
- b) $x^* \in \Omega^*$ se e somente se, $-\partial f(x^*) \cap N_\Omega(x^*) \neq \emptyset$.

Demonstração. (a) Pela definição de subgradiente condicional, $0 \in \partial^\Omega f(x^*)$ para $x^* \in \Omega$ se e só se

$$f(x) - f(x^*) \geq 0(x - x^*) \quad \text{para todo } x \in \Omega,$$

o qual é equivalente a

$$f(x) \geq f(x^*) \quad \text{para todo } x \in \Omega.$$

(b) Ver [12, Teorema 3.4.13].

□

As figuras 3.2 e 3.3 ilustram as condições de otimalidade dadas pelo Teorema 3.3.

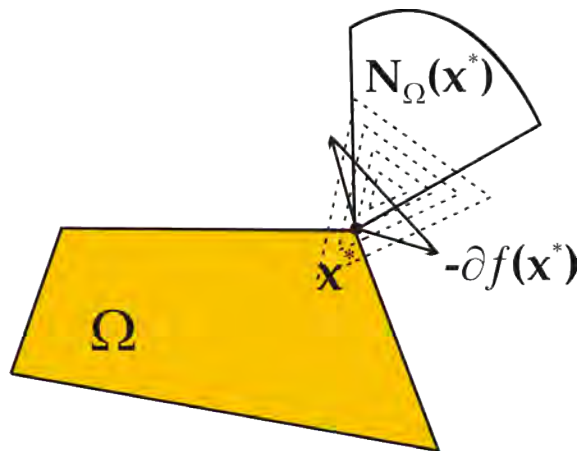


Figura 3.2: Uma solução ótima x^* .

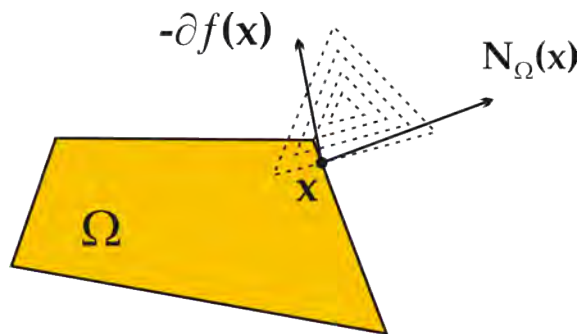


Figura 3.3: Uma solução não ótima x .

Teorema 3.4 (*Caracterização do subdiferencial condicional*)

$$\partial^\Omega f(x) = \partial f(x) + N_\Omega(x)$$

para cada $x \in \Omega$.

Demonstração. Suponha que $\gamma \in \partial^\Omega f(x)$, onde $x \in \Omega$ é fixo e arbitrário. Daí,

$$f(y) \geq f(x) + \gamma(y-x), \quad \text{para todo } y \in \Omega$$

Defina a função auxiliar $h : \Omega \rightarrow \mathbb{R}$ por

$$h(y) = f(y) - f(x) - \gamma(y-x).$$

Claramente,

- h é convexa em Ω ,
- $h(y) \geq 0$ para todo $y \in \Omega$, e
- $h(x) = 0$.

Então, x é um minimizador de h em Ω . Assim, pela condição de otimalidade, item (b) do Teorema 3.3, temos que

$$-\partial h(x) \cap N_\Omega(x) \neq \emptyset.$$

Ainda mais, da definição de h , $-\partial h(x) = -\partial f(x) + \{\gamma\}$. Daqui, existe $-\gamma_x \in -\partial f(x)$ tal que $\eta_x := -\gamma_x + \gamma \in N_\Omega(x)$. Portanto,

$$\gamma = \gamma_x + \eta_x \in \partial f(x) + N_\Omega(x).$$

Assim,

$$\partial^\Omega f(x) \subset \partial f(x) + N_\Omega(x).$$

Por outro lado, suponha que

$$\gamma = \gamma_1 + \gamma_2 \in \partial f(x) + N_\Omega(x),$$

onde $\gamma_1 \in \partial f(x)$ e $\gamma_2 \in N_\Omega(x)$. Assim, temos que

$$f(y) \geq f(x) + \gamma_1(y-x), \quad \text{para todo } y \in \mathbb{R}^n$$

e

$$\gamma_2(y-x) \leq 0 \quad \text{para todo } y \in \Omega.$$

Somando as duas desigualdades anteriores, obtemos

$$\begin{aligned} f(y) &\geq f(x) + (\gamma_1 + \gamma_2)(y-x) \\ &= f(x) + \gamma(y-x) \quad \text{para todo } y \in \Omega. \end{aligned}$$

o qual significa que $\gamma = \gamma_1 + \gamma_2 \in \partial^\Omega f(x)$. Portanto, $\partial f(x) + N_\Omega(x) \subset \partial^\Omega f(x)$, e o resultado segue. □

Em seguida, apresentamos uma outra forma de relacionar o subdiferencial usual e o subdiferencial condicional. Considere o seguinte problema "irrestrito" equivalente a (3-1):

(PA)

$$f^* = \min_{x \in \mathbb{R}^n} f^\Omega(x), \quad (3-4)$$

onde

$$f^\Omega(x) = f(x) + \delta_\Omega(x), \quad (3-5)$$

e $\delta_\Omega(x)$ é a função indicadora do conjunto Ω em x , definida em (2.27).

A função auxiliar $f^\Omega(x)$ é semi-continua inferior, própria e convexa.

Teorema 3.5 (*Caracterização do subdiferencial da função auxiliar*)

$$\partial f^\Omega(x) = \partial f(x) + \partial \delta_\Omega(x), \quad x \in \Omega$$

Demonstração. O resultado segue de [12, Proposição 3.4.8]. □

Uma vez que $\partial \delta_\Omega(x) = N_\Omega(x)$ para todo $x \in \Omega$, o Teorema 3.4 e o Teorema 3.5 implicam o seguinte resultado.

Teorema 3.6 (*Igualdade de subdiferenciais*)

$$\partial^\Omega f(x) = \partial f^\Omega(x), \quad \text{para todo } x \in \Omega.$$

Note que $\partial f^\Omega(x)$ denota o subdiferencial ordinário da função auxiliar f^Ω , enquanto $\partial^\Omega f$ denota o subdiferencial condicional de f com respeito a Ω . Como um subgradiente condicional de f é um subgradiente de f^Ω , o método do subgradiente condicional apresentado abaixo pode ser visto como o método do subgradiente usual aplicado ao problema auxiliar irrestrito (estendido) (PA).

3.2 Algoritmo Subgradiente Condicional

Considere o problema de otimização convexa

$$f^* = \min f(x) \quad (3-6a)$$

$$\text{s.a. } x \in \Omega \quad (3-6b)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa e o conjunto $\Omega \subset \mathbb{R}^n$ é não vazio, fechado e convexo.

O algoritmo subgradiente condicional é o seguinte:

Algoritmo 1 *Subgradiente Condicional*

Passo 1- Escolha um vetor inicial x^0 , $k := 0$.

Passo 2- Obtenha um subgradiente condicional de f em x^k :

$$\tilde{\gamma}^k = \gamma^k + v^k, \text{ onde } \gamma^k \in \partial f(x^k) \text{ e } v^k \in N_{\Omega}(x^k).$$

Passo 3-(Critério de parada) Se $\tilde{\gamma}^k = 0$ então pare o algoritmo, caso contrário vá ao passo seguinte.

Passo 4- Tome um tamanho de passo α^k e faça

$$x^{k+1} = P_{\Omega}(x^k - \alpha_k \tilde{\gamma}^k), \quad (3-7)$$

Passo 5- $k := k + 1$ e volte ao passo-2.

Observe que se $\tilde{\gamma}^k = 0$, então $0 \in \partial^{\Omega} f(x^k)$, e neste caso x^k é solução do problema (3-6), pelo Teorema 3.3. Portanto, nós assumiremos em nossa análise de convergência que o algoritmo gera uma sequência infinita. O parâmetro $\alpha_k > 0$ é um comprimento de passo, em seguida veremos algumas regras para este parâmetro.

Em geral, conhecer todo o subdiferencial $\partial f(x)$ é impossível do ponto de vista computacional. Uma maneira de contornar esse problema é simplesmente requerermos menos, ou seja, aceitarmos o conhecimento de pelo menos um subgradiente. A obtenção deste subgradiente está associada a uma caixa preta (também denominado oráculo), que é utilizada como base em diversos algoritmos para problemas não diferenciáveis. Dado $x^k \in \mathbb{R}^n$, a caixa preta é responsável por gerar um subgradiente $\gamma^k \in \partial f(x^k)$, juntamente com $f(x^k)$, o valor da função objetivo. No caso do método subgradiente condicional, aceitaremos que existe um oráculo capaz de retornar além do subgradiente da função, um elemento no cone dual do conjunto restrição, isto é um subgradiente condicional. Mencionamos que no Capítulo 4 aplicaremos este algoritmo a um problema dual, neste caso o cone dual possui uma fácil descrição.

Aqui aceitaremos somente esta informação do oráculo, não preocuparemos como ele está retornando o subgradiente condicional e o valor funcional.

Quando no Algoritmo 1 tomamos $v^k = 0$ obtemos o método subgradiente usual, o qual é uma generalização natural do método de Cauchy, o qual considera a direção oposta ao gradiente $\nabla f(x)$. No nosso caso, tomaremos então a direção oposta a um subgradiente γ fornecido pela caixa preta, e mais geral no caso do subgradiente condicional uma direção oposta a um subgradiente condicional e projetaremos em um conjunto restrição simples. Porém, a direção oposta ao subgradiente não é necessariamente de descida, conforme pode ser observado na Figura 3.4, que mostra curvas de níveis de uma função possuindo $0 \in \mathbb{R}^2$ como um minimizador. Direções de descida são aquelas que fazem um ângulo obtuso com todos os elementos do subdiferencial $\partial f(x)$, e não apenas com um único subgradiente γ .

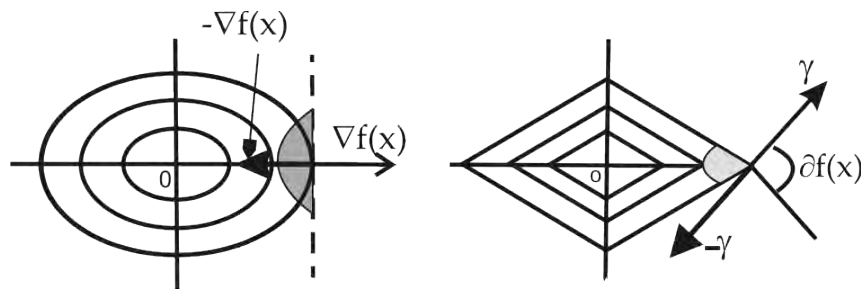


Figura 3.4: Direções de descida: caso diferenciável e não diferenciável.

A figura da direita (caso não diferenciável), mostra um exemplo em que a direção γ , fornecida pela caixa preta, é um vetor extremo do cone associado a $\partial f(x)$. Neste exemplo, a direção oposta $-\gamma$ claramente não é de descida. Apesar desse algoritmo não assegurar necessariamente o decréscimo da função objetivo em cada iteração, escolhas adequadas dos tamanhos de passo podem garantir a convergência do método.

A primeira regra é denominada regra de passos de série divergente ou passos exógenos, e exige que

$$\alpha_k \geq 0, \quad k = 0, 1, \dots, \quad \lim_{k \rightarrow \infty} \alpha_k = 0; \quad \sum_{k=0}^{\infty} \alpha_k = +\infty. \quad (3-8)$$

A segunda regra adiciona aos requisitos de (3-8) a restrição de quadrado somável

$$\sum_{k=0}^{\infty} \alpha_k^2 < +\infty. \quad (3-9)$$

A condição de soma infinita em (3-8) serve para assegurar que o tamanho do passo não é muito pequeno, enquanto que a condição adicional (3-9) assegura que o tamanho do passo não é muito grande. Ambas juntas, garantem a convergência do método independente do ponto inicial, porém em geral a convergência é lenta.

Exemplo 3.7 *Um exemplo das fórmulas do comprimento do passo que satisfaz ambas (3-8) e (3-9) é o seguinte:*

$$\alpha_k = \beta/(k+1), \quad k = 0, 1, \dots,$$

onde $\beta > 0$.

A terceira regra de comprimento de passo é

$$\alpha_k = \theta_k \frac{f(x^k) - f^*}{\|\tilde{\gamma}^k\|^2}, \quad 0 < \varepsilon_1 \leq \theta_k \leq (2 - \varepsilon_2) < 2, \quad \text{para todo } k, \quad (3-10)$$

onde f^* é o valor ótimo de (3-6). Nós nos referimos a esta regra como regra de Polyak, devido ao matemático russo Boris Polyak, um dos primeiros a estudar o método subgradiente na década de 1960, juntamente com Ermol'ev e Shor.

A convergência de algoritmos para problemas diferenciáveis é tipicamente baseada no fato que as direções são de descida, e as regras do comprimento do passo faz com que haja um decrescimento substancial nos valores funcionais dos iterados da sequência. Para o problema não diferenciável, gerar direções de descida é uma tarefa difícil, como vimos anteriormente, já que assumimos que os subgradientes são fornecidos por meio de um oráculo. Nós estabelecemos a seguir que, se o comprimento do passo é suficientemente pequeno, então uma iteração do método subgradiente leva a um vetor que está mais próximo do conjunto de soluções ótimas. Portanto, a convergência de métodos subgradientes são baseadas na redução da distância Euclidiana para as soluções ótimas e não no decréscimo da função objetivo.

Proposição 3.8 *(Redução da distância ao conjunto ótimo) Suponha que $x^k \in \Omega$ não é uma solução ótima de (3-6), e que x^{k+1} é gerado pelo método subgradiente condicional (Algoritmo) 1 com comprimento de passo α_k satisfazendo*

$$0 < \alpha_k < \frac{2(f(x^k) - f(x^*))}{\|\tilde{\gamma}^k\|^2}. \quad (3-11)$$

Então para cada solução ótima x^* de (3-6), vale

$$\|x^{k+1} - x^*\| < \|x^k - x^*\|. \quad (3-12)$$

Demonstração. Usando que o operador projeção é não expansivo temos que

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &= \|\mathbf{P}_\Omega(x^k - \alpha_k \tilde{\gamma}^k) - x^*\|^2 \\
&\leq \|x^k - \alpha_k \tilde{\gamma}^k - x^*\|^2 \\
&= \|x^k - x^*\|^2 - 2\alpha_k \tilde{\gamma}^k (x^k - x^*) + \alpha_k^2 \|\tilde{\gamma}^k\|^2 \\
&= \|x^k - x^*\|^2 + \alpha_k (\alpha_k \|\tilde{\gamma}^k\|^2 - 2\tilde{\gamma}^k (x^k - x^*)) \\
(\text{desigualdade de subgradiente}) &\leq \|x^k - x^*\|^2 + \alpha_k (\alpha_k \|\tilde{\gamma}^k\|^2 - 2(f(x^k) - f(x^*))) \\
(\text{hipótese}) &< \|x^k - x^*\|^2,
\end{aligned}$$

de onde concluímos que $\|x^{k+1} - x^*\| < \|x^k - x^*\|$. \square

O teorema seguinte estabelece a convergência da sequência $\{x^k\}$ gerada pelo método subgradiente condicional (Algoritmo) 1, impondo a condição (3-9) para a escolha do comprimento do passo α_k .

Teorema 3.9 *Seja $\{x^k\}$ a sequência gerada pelo método subgradiente condicional (Algoritmo) 1, com passo exógenos (3-8), (3-9). Se $\sup_k \{\|\tilde{\gamma}^k\|\} < \infty$, então $\{x^k\}$ converge para um elemento de Ω^* .*

Demonstração. Seja $x^* \in \Omega^*$ e $k \geq 1$. Em cada iteração k temos que

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &= \|\mathbf{P}_\Omega(x^k - \alpha_k \tilde{\gamma}^k) - x^*\|^2 \\
&\leq \|x^k - \alpha_k \tilde{\gamma}^k - x^*\|^2 \\
&= \|x^k - x^*\|^2 - 2\alpha_k \tilde{\gamma}^k (x^k - x^*) + \alpha_k^2 \|\tilde{\gamma}^k\|^2.
\end{aligned} \tag{3-13}$$

A aplicação repetida de (3-13) fornece

$$\|x^k - x^*\|^2 \leq \|x^0 - x^*\|^2 - 2 \sum_{j=0}^{k-1} \alpha_j \tilde{\gamma}^j (x^j - x^*) + \sum_{j=0}^{k-1} \alpha_j^2 \|\tilde{\gamma}^j\|^2. \tag{3-14}$$

Como $x^* \in \Omega^*$ e $\tilde{\gamma}^j \in \partial^\Omega f(x^j)$ para todo $j \geq 0$ obtemos que

$$f(x^j) \geq f^* \geq f(x^j) + (\tilde{\gamma}^j)^t (x^* - x^j), \quad j \geq 0, \tag{3-15}$$

e daí que $(\tilde{\gamma}^j)^t (x^* - x^j) \leq 0$ para todo $j \geq 0$.

Das relações (3-14) e (3-15), obtemos

$$\|x^k - x^*\|^2 \leq \|x^0 - x^*\|^2 + \sum_{j=0}^{k-1} \alpha_j^2 \|\tilde{\gamma}^j\|^2. \tag{3-16}$$

Definindo $c = \sup_j \{\|\tilde{\gamma}^j\|\}$ e $p = \sum_{j=0}^{\infty} \alpha_j^2$, obtemos

$$\|\tilde{\gamma}^j\| \leq c \text{ para qualquer } j \geq 0 \text{ e } \sum_{j=0}^{k-1} \alpha_j^2 \leq p.$$

Da relação (3-16) concluímos então que,

$$\|x^k - x^*\|^2 \leq \|x^0 - x^*\|^2 + pc^2 \text{ para qualquer } k \geq 1,$$

de onde obtemos que $\{x^k\}$ é limitada.

Suponha por contradição que não existe subsequência $\{x^{k_j}\}$ de $\{x^k\}$ satisfazendo $\{\tilde{\gamma}^{k_j}(x^{k_j} - x^*)\} \rightarrow 0$. Neste caso existe $\delta > 0$ tal que $(\tilde{\gamma}^k(x^k - x^*)) \geq \delta$ para todo $k \geq m_0$, onde $m_0 \in \mathbb{N}$ é suficientemente grande. Isto, juntamente com a condição $\sum_{j=0}^{\infty} \alpha_j = \infty$ implica que

$$\lim_{k \rightarrow \infty} \sum_{j=0}^{k-1} \alpha_{k_j} \tilde{\gamma}^{k_j}(x^{k_j} - x^*) = \infty.$$

Também temos que

$$\lim_{k \rightarrow \infty} \sum_{j=0}^{k-1} \alpha_{k_j}^2 \|\tilde{\gamma}^{k_j}\|^2 \leq c^2 \lim_{k \rightarrow \infty} \sum_{j=0}^{k-1} \alpha_{k_j}^2 < \infty.$$

As duas relações anteriores e (3-14) implicam que $\{\|x^k - x^*\|\} \rightarrow -\infty$, o que é impossível. Portanto a sequência $\{x^k\}$ possui uma subsequência $\{x^{k_j}\}$ tal que

$$\{\tilde{\gamma}^{k_j}(x^{k_j} - x^*)\} \rightarrow 0.$$

Da relação (3-15) segue-se que

$$\{f(x^{k_j})\} \rightarrow f^*.$$

Como $\{x^k\}$ é limitado, podemos supor sem perda de generalidade que $\{x^{k_j}\}$ converge a algum $\bar{x} \in \Omega$. Da continuidade de f segue-se que

$$f(\bar{x}) = f^*$$

e portanto, $\bar{x} \in \Omega^*$.

Agora, mostramos que toda a sequência $\{x^k\}$ converge para \bar{x} . Seja $\varepsilon > 0$,

podemos achar $n_0 = n_0(\varepsilon)$ tal que

$$\|x^{k_{n_0}} - \bar{x}\| < \frac{\varepsilon}{2} \quad e \quad \sum_{j=n_0}^{\infty} \alpha_{k_j}^2 < \frac{\varepsilon}{2c^2}.$$

Então, para qualquer $k > k_{n_0}$, por analogia com a relação (3-16), obtemos

$$\begin{aligned} \|x^k - \bar{x}\|^2 &\leq \|x^{k_{n_0}} - \bar{x}\|^2 + \sum_{j=k_{n_0}}^{k-1} \alpha_j^2 \|\tilde{\gamma}^j\|^2 \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2c^2} c^2 \\ &= \varepsilon. \end{aligned}$$

Como $\varepsilon > 0$ é arbitrário, obtemos que $\{x^k\}$ converge a \bar{x} , o qual é uma solução de (3-1), concluindo a demonstração do teorema. \square

Note que a condição $\sup_k \{\|\tilde{\gamma}^k\|\} < \infty$ é assegurada quando Ω é limitado.

Em geral, convergência do método subgradiente com a regra de passo exógenos é muito lenta. De fato, a regra do comprimento do passo (3-8) impede que a convergência seja rápida, como ilustraremos a seguir.

Dizemos que a sequência $\{x^k\}$ converge para x^* , com uma taxa geométrica se existe $M > 0$ e $q \in (0, 1)$ com

$$\|x^k - x^*\| \leq Mq^k, \quad k = 0, 1, \dots$$

Suponha que a sequência $\{x^k\}$ gerada pelo método subgradiente com regra de passo exógenos converge para x^* com uma taxa geométrica. Logo, para todo $k > 0$, temos

$$\|\tilde{\gamma}^k\| \alpha_k = \|x^{k+1} - x^k\| \leq \|x^{k+1} - x^*\| + \|x^k - x^*\| \leq M(q+1)q^k.$$

Somando essas desigualdades para todo k obtemos

$$\sum_{k=0}^{\infty} \|\tilde{\gamma}^k\| \alpha_k \leq M(q+1)/(1-q).$$

Por outro lado, é típico dos métodos subgradientes que a sequência de vetores direcionais $\{\tilde{\gamma}^k\}$, em geral, não converge a zero, devido a não diferenciabilidade da função objetivo. Neste caso, a desigualdade acima contradiz a condição das Séries divergentes (3-8).

O resultado do Teorema 3.9, pode ser usado para estabelecer a convergência de outras escolhas de comprimentos de passo, o seguinte corolário é imediato.

Corolário 3.10 *Considere sequências $\{\underline{\alpha}_k\}$, $\{\bar{\alpha}_k\}$ satisfazendo (3-8) e (3-9). Seja $\{\beta_k\}$ tal que $\underline{\alpha}_k \leq \beta_k \leq \bar{\alpha}_k$ para todo k . Então a sequência gerada pelo método subgradiente condicional com tamanho de passo dado pela sequência $\{\beta_k\}$ converge a uma solução de (3-6).*

Demonstração. Só temos que verificar que a sequência de comprimentos de passos $\{\beta_k\}$ satisfaz as condições de passos de series divergentes (3-8). Como $\underline{\alpha}_k \leq \beta_k \leq \bar{\alpha}_k$ para todo k , então

$$\sum_{k=0}^{\infty} \underline{\alpha}_k \leq \sum_{k=0}^{\infty} \beta_k \leq \sum_{k=0}^{\infty} \bar{\alpha}_k,$$

e

$$\sum_{k=0}^{\infty} \underline{\alpha}_k^2 \leq \sum_{k=0}^{\infty} \beta_k^2 \leq \sum_{k=0}^{\infty} \bar{\alpha}_k^2.$$

Além disso, $\{\underline{\alpha}_k\}$ e $\{\bar{\alpha}_k\}$ satisfazem (3-8) e (3-9), obtemos

$$\sum_{k=0}^{\infty} \beta_k = \infty, \quad e \quad \sum_{k=0}^{\infty} \beta_k^2 < \infty.$$

□

Como os elementos das sequências $\{\underline{\alpha}_k\}$ e $\{\bar{\alpha}_k\}$ podem ser arbitrariamente pequenos e grandes, respectivamente, enquanto satisfazem (3-8) e (3-9), por exemplo,

$$\underline{\alpha}_k = \frac{c}{\lambda + k} \quad e \quad \bar{\alpha}_k = \frac{\bar{c}}{\lambda + k}, \quad (3-17)$$

onde $c > 0$ é tão pequeno quanto se queira, $\bar{c} > 0$ é tão grande quanto se queira e $\lambda > 0$, a condição do Corolário (3.10) nos ajuda a ser flexíveis na seleção de comprimentos de passo.

Podemos assegurar que a sequência $\{\tilde{\gamma}^k\}$ de subgradientes condicionais seja limitada mediante uma escolha apropriada de v^k no método 1. Por exemplo,

$$v^k = P_{N_{\Omega}(x^k)}(-\gamma^k), \quad (3-18)$$

onde $\gamma^k \in \partial f(x^k)$. Como $\partial f(x^k)$ é um conjunto convexo e limitado e um v^k escolhido de acordo a (3-18) é também limitado, segue-se que a sequência dos $\tilde{\gamma}^k = \gamma^k + v^k$ é também limitada.

Apresentamos agora as propriedades de convergência do método subgradiente usando a regra de passo de Polyak.

Teorema 3.11 *Seja a sequência $\{x^k\}$ gerada pelo método subgradiente condicional (Algoritmo) 1 com passo de Polyak (3-10) aplicado ao problema (3-6). Se $\sup_k \{\|\tilde{\gamma}^k\|\} < \infty$, então $\{f(x^k)\} \rightarrow f^*$, e $\{x^k\}$ converge para um elemento de Ω^* .*

Demonstração. Seja $x^* \in \Omega^*$ e $k \geq 1$. Em cada iteração temos que

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &= \|\mathbf{P}_\Omega(x^k - \alpha_k \tilde{\gamma}^k) - x^*\|^2 \\
&\leq \|x^k - \alpha_k \tilde{\gamma}^k - x^*\|^2 \\
&= \|x^k - x^*\|^2 - 2\alpha_k \tilde{\gamma}^k (x^k - x^*) + \alpha_k^2 \|\tilde{\gamma}^k\|^2 \\
&= \|x^k - x^*\|^2 - 2\theta_k \frac{f(x^k) - f^*}{\|\tilde{\gamma}^k\|^2} \tilde{\gamma}^k (x^k - x^*) + \theta_k^2 \frac{(f(x^k) - f^*)^2}{\|\tilde{\gamma}^k\|^2} \\
(\text{Da desigualdade (3-15)}) &\leq \|x^k - x^*\|^2 - 2\theta_k \frac{(f(x^k) - f^*)^2}{\|\tilde{\gamma}^k\|^2} + \theta_k^2 \frac{(f(x^k) - f^*)^2}{\|\tilde{\gamma}^k\|^2} \\
&= \|x^k - x^*\|^2 - \theta_k(2 - \theta_k) \frac{(f(x^k) - f^*)^2}{\|\tilde{\gamma}^k\|^2} \tag{3-19}
\end{aligned}$$

Da definição do parâmetro θ_k temos que

$$-\theta_k(2 - \theta_k) \leq -\varepsilon_1 \varepsilon_2,$$

da relação anterior e da desigualdade (3-19), obtemos

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \frac{\varepsilon_1 \varepsilon_2}{c^2} (f(x^k) - f^*)^2, \tag{3-20}$$

onde $c = \sup_k \{\|\tilde{\gamma}^k\|\}$. Suponha que que exista uma subsequência $\{f(x^{k_j})\}$, $\delta > 0$ e $m_0 \geq 0$ tal que

$$(f(x^{k_j}) - f^*)^2 \geq \delta, \quad \text{para todo } k_j \geq m_0,$$

sem perda de generalidade, consideremos $m_0 = 0$. Usando a relação (3-20), obtemos

$$\|x^{k_j+1} - x^*\|^2 \leq \|x^{k_j} - x^*\|^2 - \alpha \delta$$

onde $\alpha = \frac{\varepsilon_1 \varepsilon_2}{c^2}$ é positivo. A desigualdade anterior escrita recursivamente da

$$\|x^{k_j+1} - x^*\|^2 \leq \|x^0 - x^*\|^2 - k_j \alpha \rightarrow -\infty \quad \text{quando } k_j \rightarrow \infty,$$

o qual é impossível. Portanto,

$$\{f(x^k)\} \rightarrow f^*. \tag{3-21}$$

Agora só resta mostrar que $\{x^k\} \rightarrow x^* \in \Omega^*$. Observe que pela propriedade (3-12), $\{\|x^k - x^*\|\}$ é monótona decrescente, o qual implica que a sequência $\{x^k\}$ é limitada. Daí, existe uma subsequência $\{x^{k_j}\}$ de $\{x^k\}$ convergente, *i.e.*, $\{x^{k_j}\} \rightarrow \bar{x}$. Segue-

se da continuidade de f que

$$\{f(x^{k_j})\} \rightarrow f(\bar{x})$$

Assim, da relação (3-21) temos que

$$f(\bar{x}) = f^*,$$

i.e., $\bar{x} \in \Omega^*$.

Nós mostramos agora que toda a sequência $\{x^k\}$ converge para \bar{x} . Como $\bar{x} \in \Omega^*$, a sequência $\|x^k - \bar{x}\|$ é limitada e monótona decrescente (Portanto convergente) com uma subsequência $\|x^{k_j} - \bar{x}\|$ convergindo a 0. Daí a sequência $\|x^k - \bar{x}\| \rightarrow 0$. assim,

$$\{x^k\} \rightarrow \bar{x}.$$

□

Observação 3.12 Se $\{v^k\} = \{0\}$, então o método subgradiente condicional 1 reduz-se ao método subgradiente tradicional.

Exemplo 3.13 (Convergência lenta do método subgradiente com passo de Polyak)
O seguinte exemplo ilustra a possível convergência lenta do método subgradiente tradicional usando o passo de Polyak (3-10).

Considere o problema

$$\begin{aligned} \min f(x) &= x_1 + 2x_2 - 1 \\ \text{s.a. } x &\in \Omega \end{aligned}$$

onde

$$\Omega = \{x \in \mathbb{R}^2 / x_1 + x_2 = 1, \quad x_1, x_2 \geq 0\}.$$

A solução ótima é dada por $x^* = (1, 0)^t$, com $f^* = 0$.

Aplicamos o método subgradiente usual com a regra de passo de Polyak tomando como vetor inicial $x^0 = (0, 1)^t$ e parâmetro de passo $\theta_k = 1$, o qual fornece o passo $\alpha_k = f(x^k)/5, \forall k$,

logo as iterações são dadas por

$$x^k = (1 - (9/10)^k, (9/10)^k)^t,$$

com valor funcional $f(x^k) = (9/10)^k, \forall k$. Ver Figura 3.5.

O método converge lentamente para a solução ótima.

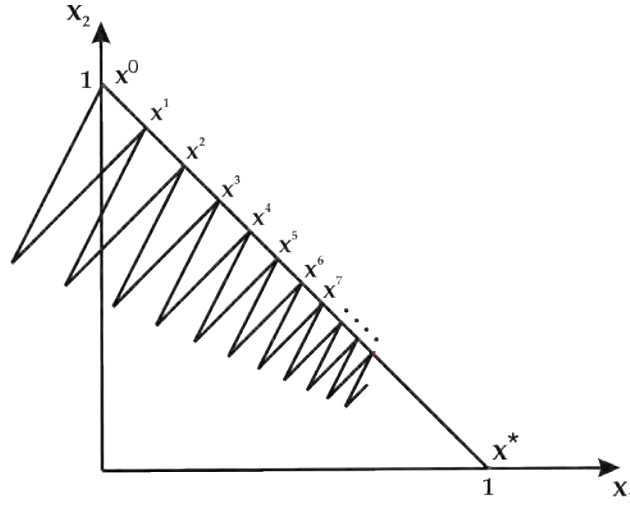


Figura 3.5: Ilustração das iterações usando passo de Polyak

O zigue zague neste exemplo é devido ao fato que os subgradientes são quase ortogonais ao conjunto Ω .

3.3 Método subgradiente com projeção

De agora em diante, consideraremos $\gamma^\Omega(x)$ definido como
(PS)

$$\gamma^\Omega(x) = \arg \min \{ \|\gamma\|^2 / \gamma \in \gamma(x) + N_\Omega(x) \}, \quad \gamma(x) \in \partial f(x).$$

Seja $v(x) \in N_\Omega(x)$, então o problema (PS) pode ser expresso equivalentemente por

$$\min_{v(x) \in N_\Omega(x)} \|\gamma(x) + v(x)\|^2 = \min_{v(x) \in N_\Omega(x)} \|v(x) - (-\gamma(x))\|^2$$

Portanto,

$$v(x) = P_{N_\Omega(x)}(-\gamma(x))$$

Lema 3.14 Seja $\gamma(x) \in \partial f(x)$, $x \in \Omega$. Então, $\|\gamma^\Omega(x)\| \leq \|\gamma(x)\|$.

Demonstração. Pela Proposição 2.37, com $v(x) = P_{N_\Omega(x)}(-\gamma(x))$ temos que

$$\langle -\gamma(x) - v(x), v(x) \rangle = 0$$

Assim, $\gamma^\Omega(x) \perp v(x)$, logo,

$$\begin{aligned} \|\gamma(x)\|^2 &= \|\gamma^\Omega - v(x)\|^2 \\ &= \|\gamma^\Omega\|^2 + \|v(x)\|^2 - 2\langle \gamma^\Omega, v(x) \rangle \\ &= \|\gamma^\Omega\|^2 + \|v(x)\|^2 \\ &\geq \|\gamma^\Omega\|^2 \end{aligned}$$

Portanto,

$$\|\gamma^\Omega\| \leq \|\gamma(x)\|, \quad x \in \Omega.$$

□

Em seguida apresentamos uma caracterização alternativa deste subgradiente condicional projetado.

Lema 3.15 (*Caracterização dual do subgradiente projetado*) *O subgradiente projetado pode ser obtido também por*

$$\gamma^\Omega = \arg \min \{ \|\gamma - \gamma(x)\|^2 / \gamma \in -T_\Omega(x) \}.$$

Demonstração. Segue da convexidade do conjunto Ω e da Proposição 2.14, que Ω é regular, i.e., $T_\Omega(x)^* = N_\Omega(x)$ e $N_\Omega(x)^* = T_\Omega(x)$.

Pela decomposição de Moreau, Teorema 2.38 com $K = -N_\Omega(x)$ e $K^* = -T_\Omega^*(x)$ temos que

$$\gamma(x) = P_K(\gamma(x)) + P_{K^*}(\gamma(x)). \quad (3-22)$$

Mas, como consequência da Proposição 2.37 temos que

$$P_K(-\gamma(x)) = -P_{-K}(\gamma(x))$$

logo, da relação (3-22) obtemos

$$-P_{-K}(-\gamma(x)) = P_K(\gamma(x)) = \gamma(x) - P_{K^*}(\gamma(x))$$

assim, $P_{K^*}(\gamma(x)) = \gamma(x) + v(x) = \gamma^\Omega(x)$.

Portanto,

$$\gamma^\Omega = \arg \min \{ \|\gamma - \gamma(x)\|^2 / \gamma \in -T_\Omega(x) \}.$$

□

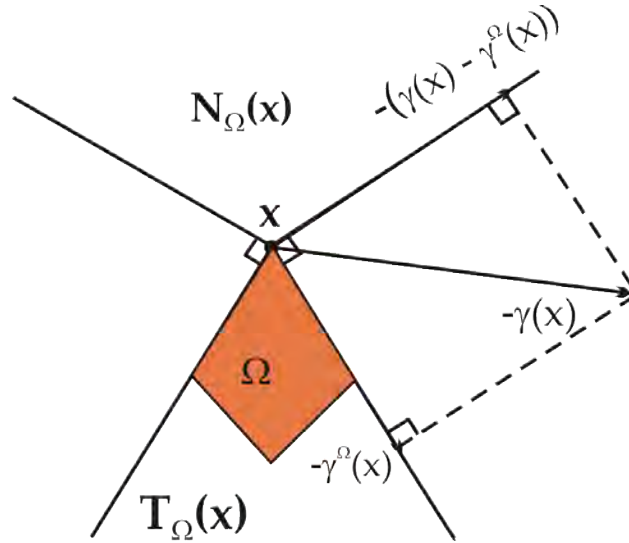


Figura 3.6: Ilustração da decomposição de Moreau.

3.3.1 Análise do método

Dado $x^k \in \Omega$, sejam $\gamma(x^k) \in \partial f(x^k)$ e $\gamma^\Omega(x^k)$ dados por (PS), o método de projeção subgradiente gera o novo iterado da seguinte forma

$$x^{k+1} = P_\Omega(x^k - \alpha_k \gamma^\Omega(x^k)), \quad k = 0, 1, \dots \quad (3-23)$$

O Corolário 3.16 decorre do Teorema 3.9 e dos Lemas 3.14 e 3.15.

Corolário 3.16 *Seja $\{x^k\}$ a sequência gerada pelo método (3-23), (3-8) e (3-9) aplicado ao problema (3-6). Se $\sup_k \{\|\gamma(x^k)\|\} < \infty$, então $\{x^k\}$ converge para um elemento de Ω^* .*

Corolário 3.17 *Seja $\{x^k\}$ a sequência gerada pelo método (3-23) e (3-10) (com $\gamma^\Omega(x^k)$ sendo um subgradiente projetado) aplicado ao problema (3-6). Então $\{f(x^k)\} \rightarrow f^*$, enquanto $\{x^k\}$ converge para um elemento de Ω^* .*

Demonstração. Pela Proposição 3.8 temos que a sequência $\{x^k\}$ é Fejér convergente, i.e.,

$$\|x^{k+1} - x^*\| < \|x^k - x^*\|,$$

daí, a sequência $\{x^k\}$ é limitada e, portanto, a sequência $\|\gamma(x^k)\|$ é limitada. Pelo Lema 3.14 obtemos a limitação de $\{\gamma^\Omega(x^k)\}$, e o corolário segue do Lema 3.15 e do Teorema 3.11. \square

Exemplo 3.18 (Convergência usando subgradiente projetado) Para mostrar o efeito da projeção do subgradiente, aplicamos o método (3-23), (3-10) ao exemplo 3.13, com $x^0 = (0, 1)$ e $\theta_k = 1 \ \forall k$.

Temos que $v^0 = P_{N_{\Omega}(x^0)}(-\gamma^0)$,

onde

$$N_{\Omega}(x) = \{y \in \mathbb{R}^n / y^T(z-x) \leq 0, \text{ para todo } z \in \Omega\}.$$

Como $x^0 = (0, 1)$, e $y \in \Omega = \{x \in \mathbb{R}^2 / x_1 + x_2 = 1, \ x_1, x_2 \geq 0\}$,

então temos que o cone normal de Ω no ponto $x^0 = (0, 1)$ é dado por

$$N_{\Omega}(x^0) = \{(v_1, v_2) \in \mathbb{R}^2 / v_1 - v_2 \leq 0, \forall z \in \Omega\}. \text{ Ver Figura 3.7.}$$

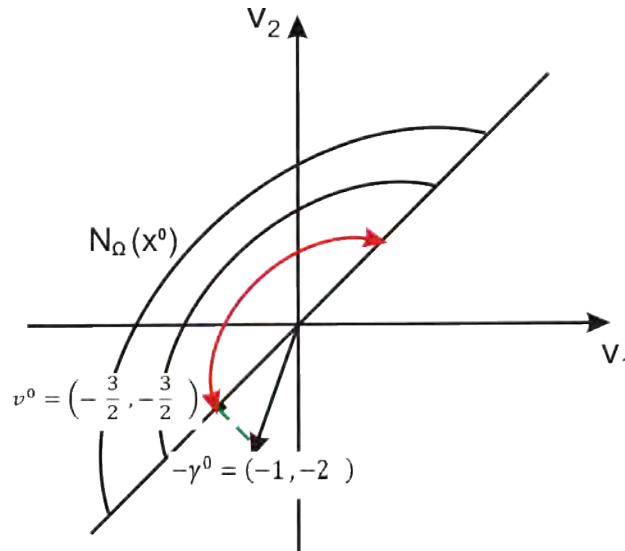


Figura 3.7: Ilustração do Cone normal de Ω no ponto x^0 .

Logo $v^0 = (\frac{-3}{2}, \frac{-3}{2})$,

e o subgradiente condicional é dado por

$$\begin{aligned} \tilde{\gamma}^0 &= \gamma^0 + v^0 \\ &= (1, 2) + (\frac{-3}{2}, \frac{-3}{2}) \\ &= (\frac{-1}{2}, \frac{1}{2}), \end{aligned}$$

onde o passo de Polyak é dado por $\alpha_0 = \frac{f(x^0)}{\|\tilde{\gamma}^0\|^2} = 2$.

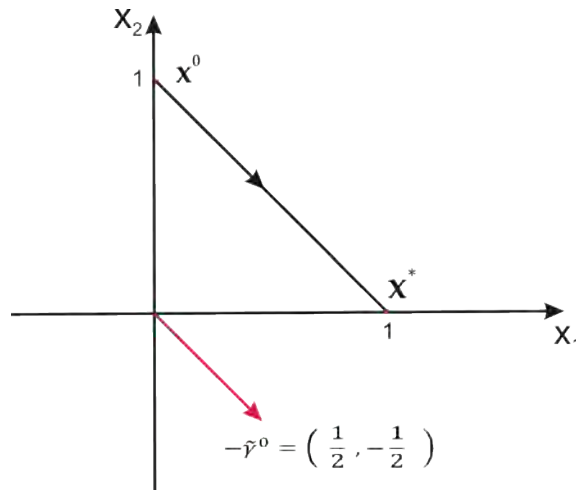


Figura 3.8: Ilustração da direção viável $-\tilde{\gamma}^0$.

Assim, obtemos

$$\begin{aligned}
 x^1 &= P_{\Omega}(x^0 - \alpha_0 \tilde{\gamma}^0) \\
 &= P_{\Omega}\left((0, 1) - 2\left(\frac{-1}{2}, \frac{1}{2}\right)\right) \\
 &= P_{\Omega}((1, 0)) \\
 &= (1, 0).
 \end{aligned}$$

Obtendo a solução em uma iteração.

Observe no exemplo que a direção $-\tilde{\gamma}^0 = (\frac{1}{2}, -\frac{1}{2})$ é uma direção viável para o conjunto Ω , no qual é um poliedro. Em seguida mostraremos que se o conjunto Ω é um poliedro, então a direção $-\tilde{\gamma}$ gerada pelo caso especial de projeção subgradiente (3-23) é sempre uma direção viável. Primeiro apresentaremos uma caracterização do cone normal de um poliedro.

Proposição 3.19 *Seja Ω um conjunto poliedral convexo e fechado definido por m restrições:*

$$\Omega = \{ x \in \mathbb{R}^n / a_i^t x \leq b_i \text{ para } i = 1, \dots, m \}$$

e defina

$$I(x) = \{ i = 1, \dots, m / a_i^t x = b_i \}$$

o conjunto de índices das restrições ativas em $x \in \Omega$.

Então,

$$N_{\Omega}(x) = \text{cone}\{ a_i / i \in I(x) \} = \left\{ \sum_{i \in I(x)} \alpha_i a_i / \alpha_i \geq 0 \right\}.$$

Demonstração. Seja $v \in \text{cone}\{ a_i / i \in I(x) \}$, i.e., $v = \{ \sum_{i \in I(x)} \alpha_i a_i / \alpha_i \geq 0 \}$. Então para cada $y \in \Omega$ temos que

$$\begin{aligned} \langle v, y - x \rangle &= \left\langle \sum_{i \in I(x)} \alpha_i a_i, y - x \right\rangle \\ &= \sum_{i \in I(x)} \alpha_i (\langle a_i, y \rangle - \langle a_i, x \rangle) \\ &\leq 0. \end{aligned}$$

Pois, para cada $i \in I(x)$ temos que $\langle a_i, x \rangle = b_i$ e $\langle a_i, y \rangle \leq b_i$ para todo $y \in \Omega$. Em consequência, $v \in N_{\Omega}(x)$.

Reciprocamente, seja $v \in N_{\Omega}(x)$. Então $\langle v, y - x \rangle \leq 0$ para todo $y \in \Omega$. Vamos denotar $\text{cone}(\Omega - x)$ o cone composto de vetores $t(y - x)$ com $y \in \Omega$ e $t \geq 0$, vemos que

$$\langle v, z \rangle \leq 0 \quad \text{para todo } z \in \text{cone}(\Omega - x). \quad (3-24)$$

Afirmamos que $\text{cone}(\Omega - x)$ coincide com o conjunto

$$X = \{ z \in \mathbb{R}^n / \langle a_i, z \rangle \leq 0 \quad \text{para todo } i \in I(x) \}.$$

De fato, seja $z \in \text{cone}(\Omega - x)$, $y = t(y - x)$ para algum $y \in \Omega$ e $t \geq 0$. Então para cada $i \in I(x)$ temos

$$\langle a_i, z \rangle = t \langle a_i, y - x \rangle \leq 0.$$

Isto implica que $\text{cone}(\Omega - x) \subset X$. Reciprocamente, seja $z \in X$. Seja $t \geq 0$ suficientemente pequeno tal que

$$\langle a_j, tz \rangle \leq -\min\{ \langle a_i, x \rangle - b_i : i \in \{1, \dots, m\} \setminus I(x) \}$$

para todo $j = 1, \dots, m$. Tal t existe pois $z \in X$ e $\langle a_i, x \rangle < b_i$ para $i \notin I(x)$. Logo temos

$$\langle a_i, tz + x \rangle = t \langle a_i, z \rangle + \langle a_i, x \rangle \leq b_i, \quad \text{para } i \in I(x),$$

e

$$\langle a_j, tz + x \rangle = \langle a_j, tz \rangle + \langle a_j, x \rangle \leq b_j, \quad \text{para } j \in \{1, \dots, m\} \setminus I(x).$$

Estas desigualdades mostram que $tz + x \in \Omega$, ou equivalentemente que $z \in \text{cone}(\Omega - x)$. Assim $\text{cone}(\Omega - x) = X$. Isto e (3-24) implicam que $v \in X^*$, i.e., v esta no polar de X . Pelo item (a) de [4, Proposição 3.2.1], temos que

$$v \in \text{cone}\{ a_i / i \in I(x) \}.$$

□

Observe que as condições de otimalidade dadas pelo Teorema 3.3 podem ser expressas como

$$\exists \gamma \in \partial f(x) \quad \text{tal que} \quad \gamma = - \sum_{i \in I} \alpha_i a_i.$$

Por outro lado, pela Decomposição de Moreau, Teorema 2.38 temos que

$$P_{N_{\Omega}(x)}(z) + P_{T_{\Omega}(x)}(z) = z, \quad z \in \mathbb{R}^n.$$

Tomando $z = -\gamma(x)$, i.e., um subgradiente de f no ponto x , temos que

$$P_{N_{\Omega}(x)}(-\gamma(x)) + P_{T_{\Omega}(x)}(-\gamma(x)) = -\gamma(x),$$

assim, obtemos

$$-\gamma(x) - P_{N_{\Omega}(x)}(-\gamma(x)) = P_{T_{\Omega}(x)}(-\gamma(x)),$$

portanto,

$$-\tilde{\gamma}(x) = P_{T_{\Omega}(x)}(-\gamma(x)). \quad (3-25)$$

Como Ω é um poliedro fechado, então os cones tangente e normal são poliedros fechados, o que significa que $-\tilde{\gamma}(x)$ é uma direção viável.

Observação 3.20 Como o conjunto \mathbb{R}_+^n é um poliedro da forma

$$\mathbb{R}_+^n = \{ x \in \mathbb{R}^n / \langle a_i, x \rangle \geq 0 \},$$

onde $a_i = ce_i$, $i = 1, \dots, n$ sendo e_i os vetores canônicos de \mathbb{R}^n e c uma constante em \mathbb{R}_+ , segue da Proposição 3.19 que

$$N_{\mathbb{R}_+^n}(x) = \text{cone}\{ -a_i / \langle a_i, x \rangle = 0 \},$$

ou, equivalentemente

$$N_{\mathbb{R}_+^n}(x) = \{ v \in \mathbb{R}_+^n / v_i x_i = 0, \quad i = 1, \dots, n \}.$$

Método subgradiente condicional com sequência ergódica

Métodos subgradientes são bem simples e geralmente são eficazes para obter boas aproximações tanto para soluções duais quanto para o valor ótimo do problema primal, principalmente quando o problema dual possui estruturas especiais. No entanto, é bem conhecido que, em geral, estes métodos não proporcionam soluções do problema primal. Como vimos anteriormente, para obter uma solução ótima primal a partir de uma solução ótima dual, convexidade e dualidade forte tem um papel fundamental, veja Teorema 2.44. A obtenção de uma solução ótima primal é direta quando a função dual q é diferenciável na solução dual. Infelizmente, em muitos problemas práticos este não é o caso (por exemplo, programação linear). Neste capítulo, apresentamos uma alternativa para obter uma solução primal usando sequência ergódica das soluções obtidas durante a resolução do subproblema dual Lagrangiano.

4.1 Método subgradiente dual

Sejam as funções $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$, $j = 1, 2, \dots, m$ convexas, não necessariamente diferenciáveis, e seja o conjunto $\Omega \subset \mathbb{R}^n$ não vazio, convexo e compacto. Considere o seguinte problema convexo

$$f^* = \min f(x) \tag{4-1a}$$

$$g_j(x) \leq 0, \quad j = 1, 2, \dots, m \tag{4-1b}$$

$$s.a \ x \in \Omega, \tag{4-1c}$$

com conjunto solução Ω^* . Vamos assumir que o conjunto viável

$$\{x \in \Omega / g_j(x) \leq 0, j = 1, 2, \dots, m\}$$

é não vazio.

4.1.1 Dualidade Lagrangiana

De acordo com os resultados básicos apresentados no capítulo 2, temos que a formulação dual Lagrangiana para o problema (4-1) é dada por

$$\begin{aligned} q^* &= \max_{\mu} q(\mu) \\ \text{s.a. } \mu &\geq 0, \end{aligned} \quad (4-2)$$

com conjunto solução U^* , onde

$$q(\mu) = \min_{x \in \Omega} f(x) + \mu^t g(x), \quad (4-3)$$

e o conjunto solução convexo e compacto para este subproblema em $\mu \in \mathbb{R}_+^m$ é dado por

$$\Omega(\mu) = \{x \in \Omega / f(x) + \mu^t g(x) = q(\mu)\}. \quad (4-4)$$

Os dois resultados seguintes são cruciais para o desenvolvimento de nossa análise. Nós utilizamos a seguinte noção de aplicação fechada. Uma aplicação ponto-conjunto $\Omega : \mathbb{R}^m \rightarrow 2^{\mathbb{R}^n}$ é fechada se $\{\mu^k\} \rightarrow \mu$, $x^k \in \Omega(\mu^k)$ para todo k , e $\{x^k\} \rightarrow x$ implicam que $x \in \Omega(\mu)$.

Lema 4.1 *Tome uma sequência $\{\mu^k\} \subset \mathbb{R}^m$ e considere a aplicação $\Omega : \mathbb{R}^m \rightarrow 2^{\mathbb{R}^n}$ dada pela relação (4-4), e a sequência dada pela inclusão $x^k \in \Omega(\mu^k)$. Se $\{\mu^k\} \rightarrow \mu$, então $\{dist(x^k, \Omega(\mu))\} \rightarrow 0$. Se, além disso, $\Omega(\mu) = \{x\}$, então $\{x^k\} \rightarrow x$.*

Demonstração. Para provar este resultado, mostraremos que todo ponto limite de $\{x^k\}$ pertence a $\Omega(\mu)$. Observe que $\{x^k\}$ é limitada, pois Ω é compacto. Seja $\{x^{k_j}\}$ uma subsequência convergente para algum $x \in \Omega$. Vamos mostrar que $f(x) + \mu^t g(x) = q(\mu)$. Como as funções f e g são contínuas, temos que

$$\{f(x^{k_j}) + (\mu^{k_j})^t g(x^{k_j})\} \rightarrow f(x) + \mu^t g(x) \quad (4-5)$$

Por outro lado, como a função q é contínua, segue-se que

$$q(\mu^{k_j}) \rightarrow q(\mu). \quad (4-6)$$

Assim, usando que $x^{k_j} \in \Omega(\mu^{k_j})$, (4-5) e (4-6) temos que

$$x \in \Omega \quad e \quad f(x) + \mu^t g(x) = q(\mu),$$

isto é, $x \in \Omega(\mu)$.

Portanto,

$$\{dist(x^k, \Omega(\mu))\} \rightarrow 0.$$

Segue-se que, quando $\Omega(\mu) = \{x\}$, então $\{x^k\} \rightarrow x$. □

Para cada $\mu \in \mathbb{R}^m$, definimos o conjunto $I(\mu)$ de índices, que correspondem aos multiplicadores μ_i estritamente positivos, isto é

$$I(\mu) = \{i \in I / \mu_i > 0\}, \quad (4-7)$$

onde $I = \{1, 2, \dots, m\}$.

Lema 4.2 *As funções f e g_i , $i \in I(\mu)$, são afins sobre $\Omega(\mu)$ para cada $\mu \geq 0$.*

Demonstração. Para qualquer $\mu \geq 0$, sejam $x, y \in \Omega(\mu)$ arbitrários e $\lambda \in [0, 1]$. Pela convexidade das funções f e g_i , $i \in I(\mu)$ e as relações (4-3) e (4-4),

$$\begin{aligned} q(\mu) &\leq f(\lambda x + (1 - \lambda)y) + \sum_{i \in I(\mu)} \mu_i g_i(\lambda x + (1 - \lambda)y) \\ &\leq \lambda f(x) + (1 - \lambda)f(y) + \lambda \sum_{i \in I(\mu)} \mu_i g_i(x) + (1 - \lambda) \sum_{i \in I(\mu)} \mu_i g_i(y) \\ &= \lambda \left(f(x) + \sum_{i \in I(\mu)} \mu_i g_i(x) \right) + (1 - \lambda) \left(f(y) + \sum_{i \in I(\mu)} \mu_i g_i(y) \right) \\ &= \lambda q(\mu) + (1 - \lambda)q(\mu) \\ &= q(\mu). \end{aligned}$$

Portanto o conjunto $\Omega(\mu)$ é convexo. Logo temos que,

$$\begin{aligned}
 f(\lambda x + (1 - \lambda)y) &= q(\mu) - \sum_{i \in I(\mu)} \mu_i g_i(\lambda x + (1 - \lambda)y) \\
 &\geq q(\mu) - \lambda \sum_{i \in I(\mu)} \mu_i g_i(x) - (1 - \lambda) \sum_{i \in I(\mu)} \mu_i g_i(y) \\
 &= \lambda q(\mu) - \lambda q(\mu) + q(\mu) - \lambda \sum_{i \in I(\mu)} \mu_i g_i(x) - (1 - \lambda) \sum_{i \in I(\mu)} \mu_i g_i(y) \\
 &= \lambda \left(q(\mu) - \sum_{i \in I(\mu)} \mu_i g_i(x) \right) + (1 - \lambda) \left(q(\mu) - \sum_{i \in I(\mu)} \mu_i g_i(y) \right) \\
 &= \lambda f(x) + (1 - \lambda)f(y).
 \end{aligned}$$

Assim, f é côncava e convexa sobre $\Omega(\mu)$. Portanto, f é afim sobre $\Omega(\mu)$.

De maneira análoga, temos que g_i é afim sobre Ω , $i \in I(\mu)$ □

Uma consequência do Lema 4.2 é que, para todo $\mu \geq 0$ e cada $i \in I(\mu)$, ∂g_i é constante sobre $\text{rint}\Omega(\mu)$; daí, por exemplo, para cada $\bar{x} \in \text{rint}\Omega(\mu)$, cada subgradiente $\gamma^i \in \partial g_i(\bar{x})$ define um hiperplano que suporta a função g_i em cada $x \in \Omega(\mu)$, i.e.,

$$\partial g_i(\bar{x}) = \cap_{x \in \Omega(\mu)} \partial g_i(x).$$

Proposição 4.3 (Subdiferencial da função dual) Para cada $\mu \in \mathbb{R}^m$, obtemos $\partial q(\mu) = \text{conv}\{g(x) / x \in \Omega(\mu)\}$. Além disso, q é diferenciável em μ se, e somente se, cada g_i é constante sobre $\Omega(\mu)$, neste caso $\nabla q(\mu) = g(x)$ para qualquer $x \in \Omega(\mu)$.

Demonstração. Note que a função $L(x, \cdot) : \mathbb{R}^m \rightarrow \mathbb{R}$ é uma função afim, portanto convexa e diferenciável para todo $x \in \mathbb{R}^n$ fixo. Da definição do conjunto solução do subproblema dual temos

$$\Omega(\mu) = \{ x \in \Omega / L(x, \mu) = q(\mu) \},$$

logo, pelo Teorema de Danskin 2.52 item (d), temos que

$$\partial q(\mu) = \text{conv} \{ g(x) / x \in \Omega(\mu) \}.$$

□

Da observação 3.20 temos que o cone normal para o conjunto \mathbb{R}_+^m no ponto $u \in \mathbb{R}_+^m$ é

$$N_{\mathbb{R}_+^m}(u) = \{ v \in \mathbb{R}_+^m / v_i u_i = 0, i = 1, 2, \dots, m \} \quad (4-8)$$

O seguinte resultado resume as condições de otimalidade para o problema dual Lagrangiano (4-2), e suas consequências para a viabilidade de uma solução ótima primal

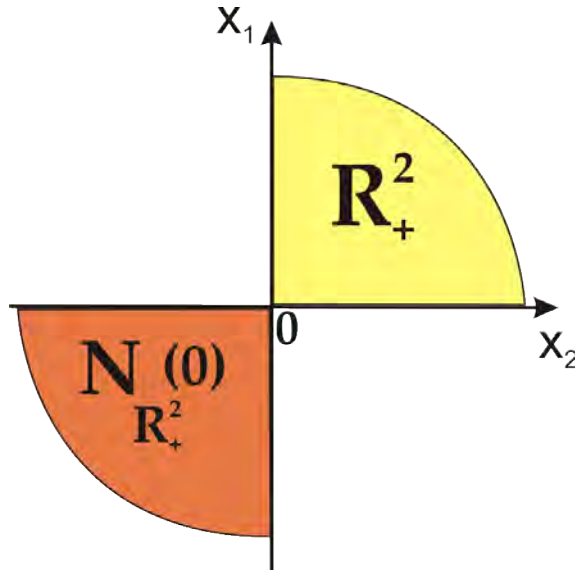


Figura 4.1: Ilustração do Cone Normal de \mathbb{R}_+^2 no ponto $\mu = 0$.

na ausência de brecha de dualidade.

Proposição 4.4 (Condições de otimalidade para o problema dual)

- (a) O ponto $\mu \in U^*$ se, e somente se, existe um $\gamma \in \partial q(\mu)$ tal que $\gamma \leq 0$ e $\mu^t \gamma = 0$, ou, equivalentemente, $\partial q(\mu) \cap N_{\mathbb{R}_+^m}(\mu) \neq \emptyset$.
- (b) Se q é diferenciável em μ^* , então não existe brecha de dualidade. Além disso, qualquer elemento em $\Omega(\mu^*)$ resolve o problema primal (4-1).

Demonstração. (a) É similar ao item (b) do Teorema 3.3.

(b) Seja \bar{x} qualquer vetor em $\Omega(\mu^*)$. Pelo item (a) do Teorema 2.56 segue que $\nabla q(\mu^*) = g(\bar{x})$. Pela parte (a) obtemos que

$$0 \leq \mu^* \perp g(\bar{x}) \leq 0.$$

Assim, o par (\bar{x}, μ^*) satisfaz todas as condições dadas em (2-18), logo, pelo Teorema 2.44, \bar{x} é uma solução ótima para (4-1). \square

Para obter relações de otimalidade primal-dual, o conjunto viável primal deve satisfazer a definição (2.45), que iremos usar como hipótese ao longo deste capítulo.

Hipótese 1 (Qualificação de Slater) Existe um vetor $x \in \Omega$ tal que

$$g_j(x) < 0 \text{ para todo } j = 1, 2, \dots, m. \quad (4-9)$$

Sob a hipótese 1, pelo Teorema 2.46 de Dualidade Forte, o conjunto convexo U^* é não vazio e compacto, e valem as condições de otimalidade (2-18).

Consequentemente, sob a hipótese 1, o conjunto solução do problema primal (4-1) pode ser expresso como

$$\Omega^* = \{x \in \Omega(\mu) / g(x) \leq 0, \mu^t g(x) = 0, \mu \geq 0\}. \quad (4-10)$$

As condições de otimalidade primal-dual podem ser expressas como

$$(x, \mu) \in \Omega^* \times U^* \iff g(x) \in \partial q(\mu) \cap N_{\mathbb{R}_+^m}(\mu). \quad (4-11)$$

Como vimos anteriormente, a função dual, em geral, não é diferenciável sobre U^* , e um subgradiente que possa ser usado para verificar a otimalidade de tal solução, de acordo com o item (a) da Proposição 4.4, não é fácil de obter.

4.2 Convergência do método subgradiente dual

Adaptando o método subgradiente condicional (estudado no Capítulo 3) aplicado ao problema dual Lagrangiano (4-2), obtemos o seguinte algoritmo:

Algoritmo 2 Subgradiente Condicional Dual

Passo 1- Escolha um vetor inicial $\mu^0 \geq 0$, $k:=0$.

Passo 2- Obtenha $x^k \in \Omega(\mu^k)$, solução de (4-3) em μ^k , e tome $v^k \in N_{\mathbb{R}_+^m}(\mu^k)$,

Passo 3-(Critério de parada) Se $g(x^k) - v^k = 0$ então pare o algoritmo, caso contrário vá ao passo seguinte.

Passo 4- Tome um tamanho de passo α^k e faça

$$\mu^{k+1} = [\mu^k + \alpha_k(g(x^k) - v^k)]^+, \quad (4-12)$$

($[\cdot]^+$ denota a projeção Euclidiana sobre o octante não negativo \mathbb{R}_+^m .)

Passo 5- $k:=k+1$ e volte ao passo-2.

Lembramos que se $\{v^k\} = \{0\}$, então o método (Algoritmo) 2 reduz-se ao método subgradiente usual aplicado ao problema dual. Escolhendo $v^k = P_{N_{\mathbb{R}_+^m}(\mu^k)}(g(x^k))$ temos que

$$g_j(x^k) - v_j^k = \begin{cases} 0, & \text{se } \mu_j^k = 0 \text{ e } g_j(x^k) < 0 \\ g_j(x^k), & \text{caso contrário,} \end{cases} \quad (4-13)$$

onde $j = 1, \dots, m$.

O método (Algoritmo) 2 usando esta direção é chamado *método de projeção subgradiente*. Lembre que como \mathbb{R}_+^m é um poliedro, similar à relação (3-25), a direção $g(x^k) - v^k$ de $\mu^k \in \mathbb{R}_+^m$, como é definida na relação (4-13), é viável para o problema dual (4-2).

No capítulo 3 provamos a convergência do algoritmo subgradiente condicional usando as regras de passos exógenos e passo de Polyak. Neste capítulo analisaremos o método de projeção subgradiente usando a regra de passo exógenos para a escolha do tamanho dos passos.

Proposição 4.5 *Suponha que vale a hipótese 1, e considere o método (Algoritmo) 2 aplicado ao problema dual (4-2), com a regra de passos exógenos (3-8) e (3-9). Se a sequência $\{\mu^k\}$ é limitada, então $\{\mu^k\} \rightarrow \mu^* \in U^*$ e $\{q(\mu^k)\} \rightarrow q^*$.*

Demonstração. Se vale a qualificação de Slater, então do Teorema 2.46 de Dualidade forte, temos que existe pelo menos um multiplicador de Lagrange, i.e., o conjunto U^* é não vazio e compacto.

Por outro lado, como o conjunto Ω é compacto, de [12, Proposição 3.4.6] temos que a sequência $\{g^k\}$ é limitada, e portanto, as sequências $\{v^k\}$ e $\{g^k - v^k\}$ são limitadas. O resultado decorre de maneira similar ao caso de minimização do Teorema 3.9. □

Das Proposições 4.4 e 4.5 segue-se que o conjunto $\partial q(\mu^*) \cap N_{\mathbb{R}_+^m}(\mu^*) \neq \emptyset$.

Definição 4.6 *Definimos a sequência $\{A_k\}$ de passos acumulativos por*

$$A_k = \sum_{s=0}^{k-1} \alpha_s, \quad k = 1, 2, \dots, \quad (4-14)$$

Em seguida, apresentamos um exemplo onde a sequência obtida do subproblema dual quando aplicado o método subgradiente não converge a uma solução ótima primal.

Exemplo 4.7 (*Sequência primal que não converge*)

Considere o seguinte problema de programação linear

$$\min f(x) = x$$

$$\text{s.a. } x \geq 1$$

$$0 \leq x \leq 2,$$

onde temos as restrições $g(x) = 1 - x \leq 0$ e $\Omega = [0, 2]$.

Dualizando a restrição $g(x) \leq 0$, obtemos o problema dual correspondente

$$\max q(\mu)$$

$$\text{s.a. } \mu \geq 0,$$

onde $q(\mu) = \min_{x \in \Omega} f(x) + \mu^t g(x)$.

Vamos considerar o caso em que $\{v^k\} = \{0\}$, i.e., consideramos o método subgradiente tradicional com passo de Polyak, com parâmetro de passo $\theta_k = 0.5 \forall k$, e vetor inicial dual $\mu^1 = 0$.

Vamos mostrar que o método gera uma sequência primal $\{x^k\}$ tal que $x^k = 0 \forall k$, a qual não é uma solução primal, e de fato não é uma solução viável.

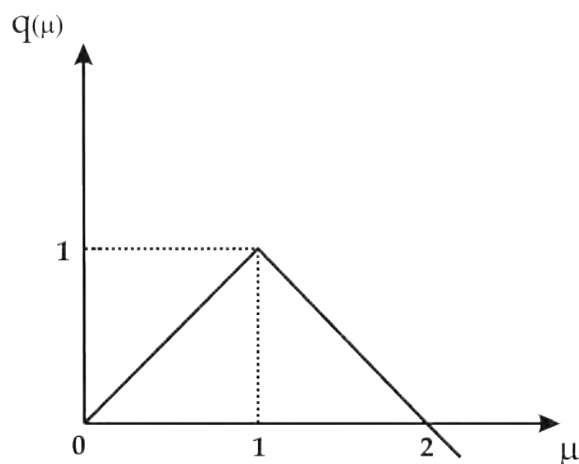


Figura 4.2: Função dual Lagrangiana.

Note que $x^k = 0$ se e somente se $0 \leq \mu^k < 1$. Daí, é suficiente mostrar que o método subgradiente gera uma sequência dual $\mu^k \in [0, 1)$, $\forall k$. Vamos mostrar isto por

indução.

Para $k = 1$, temos que $\mu^1 = 0 \in [0, 1)$.

Agora, suponha que $\mu^1, \mu^2, \dots, \mu^k$ estão no intervalo $[0, 1)$.

Logo,

$$\mu^{k+1} = \mu^k + 0.5 \frac{(q^* - q(\mu^k))}{\|g(x^k)\|^2}, \quad (4-15)$$

onde $q^* = 1$ e $q(\mu^k) = \min_{x \in \Omega} f(x^k) + \mu^k g(x^k)$.

Como da hipótese de indução $\mu^k \in [0, 1)$, então $x^k = 0$, assim temos $f(x^k) = 0$ e $g(x^k) = 1 - x^k = 1$.

Substituindo em (4-15), temos

$$\begin{aligned} \mu^{k+1} &= \mu^k + 0.5 \frac{(f^* - q(\mu^k))}{\|g(x^k)\|^2} \\ &= \mu^k + 0.5 \frac{1 - \mu^k}{1} \\ &= 0.5 + 0.5\mu^k. \end{aligned}$$

Como $\mu^k < 1$, então $0.5 + 0.5\mu^k < 0.5 + 0.5 = 1$.

Portanto, $\mu^{k+1} < 1$, concluindo por indução que $\mu^j < 1$ for all j . Portanto o método subgradiente com os parâmetros escolhidos com acima gera uma sequência $\{x^k\} = 0$, a qual não converge à solução ótima primal $x^* = 1$. Embora o método não gera uma sequência convergindo para uma solução primal, ele ainda gera uma sequência que converge para uma solução ótima dual, como era de se esperar. De fato, temos que $\mu^{k+1} = 0.5 + 0.5\mu^k$, logo com $\mu^1 = 0$ temos que

$$\begin{aligned} \lim_{k \rightarrow \infty} \mu^k &= 0.5 + 0.5^2 + 0.5^3 + \dots, \\ &= -1 + \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k, \\ &= -1 + \frac{1}{1 - 1/2}, \\ &= 1. \end{aligned}$$

Portanto, a sequência dual converge para a solução ótima $\mu^* = 1$, com $q(\mu^*) = 1$.

Na seguinte seção mostramos que podemos aproximar uma solução primal usando uma sequência ergódica relacionada a sequência obtida dos subproblemas Lagrangianos.

4.2.1 Convergência ergódica primal

O método subgradiente condicional (Algoritmo) 2 aplicado ao problema dual (4-2) produz uma sequência $\{x^k\}$ obtida dos subproblemas Lagrangianos (4-3). Como vimos no exemplo 4.7, não há garantia de convergência desta sequência a uma solução ótima. Em seguida, apresentamos um tipo de escolha, para gerar uma sequência ergódica de soluções do subproblema dual. Mostramos que esta sequência possui boas propriedades de convergência. Em particular, podemos obter uma solução primal por meio de pontos limites da sequência ergódica.

Definição 4.8 (*Sequência ergódica*) Uma sequência ergódica das soluções do subproblema dual Lagrangiano é definida pela média de pesos

$$\bar{x}^k = A_k^{-1} \sum_{s=0}^{k-1} \alpha_s x^s, \quad k = 1, 2, \dots, \quad (4-16)$$

onde a sequência $\{A_k^{-1}\}$ é definida na relação (4-14), $x^k \in \Omega(\mu^k)$, i.e., uma solução do subproblema dual Lagrangiano (4-3) no ponto μ^k , e α_k é um comprimento de passo usado no método 2.

Nós, iremos mostrar que cada um dos pontos de acumulação da sequência ergódica $\{\bar{x}^k\}$ definida por (4-16) é uma solução ótima do problema primal (4-1).

O seguinte Lema é importante para a nossa prova do teorema principal deste capítulo.

Lema 4.9 *Suponha que a sequência $\{\beta_{ks}\} \subset \mathbb{R}$ satisfaz as condições*

$$\beta_{ks} \geq 0, \quad s = 0, \dots, k-1, \quad \sum_{s=0}^{k-1} \beta_{ks} = 1, \quad k = 1, 2, \dots,$$

$$e \quad \lim_{k \rightarrow \infty} \beta_{ks} = 0, \quad s = 0, 1, \dots$$

Se a sequência $\{b^s\} \subset \mathbb{R}^m$ é tal que $\lim_{s \rightarrow \infty} b^s = b$, então $\lim_{k \rightarrow \infty} \left(\sum_{s=0}^{k-1} \beta_{ks} b^s \right) = b$.

Demonstração. Ver [14, Teorema 2]. □

A convergência da sequência $\{\bar{x}^k\}$ para o conjunto Ω^* é estabelecida em termos do cumprimento das condições de otimalidade dadas pela relação (4-10).

Teorema 4.10 *Suponha que vale a hipótese 1, e considere o método (Algoritmo) 2 com a escolha da direção como em (4-13), e regra de passos exógenos (3-8) e (3-9), aplicado ao problema dual (4-2). Tome $\{\bar{x}^k\}$ gerada por (4-16), e suponha que a sequência $\{v^k\}$ é limitada. Então,*

$$\{dist(\bar{x}^k, \Omega^*)\} \rightarrow 0.$$

Demonstração. Primeiro vamos mostrar que, para qualquer ponto de acumulação x^* da sequência ergódica $\{\bar{x}^k\}$, $x^* \in \Omega(\mu^*)$ para algum $\mu^* \in \Omega^*$, logo que esta sequência é viável no limite, i.e., $g_j(x^*) \leq 0 \quad \forall j = 1, 2, \dots, m$, e por ultimo, que a sequência satisfaz a condição de complementaridade, i.e., $\mu^{*t} g(x^*) = 0$.

Seja $\mu^* \in U^*$ o limite da sequência $\{\mu^k\}$ dada pela Proposição 4.5. Pela convexidade da função distância, temos

$$0 \leq dist(\bar{x}^k, \Omega(\mu^*)) \leq A_k^{-1} \sum_{s=0}^{k-1} \alpha_s dist(x^s, \Omega(\mu^*)), \quad \forall k.$$

Pelo Lema 4.1 e a convergência da sequência $\{\mu^k\} \rightarrow \mu^*$ obtemos

$$dist(x^s, \Omega(\mu^*)) \rightarrow 0 \quad \text{quando} \quad s \rightarrow \infty. \quad (4-17)$$

Observe que

$$\lim_{k \rightarrow \infty} \beta_{ks} = \lim_{k \rightarrow \infty} \frac{\alpha_k}{\sum_{s=0}^{k-1} \alpha_s} = 0, \quad \sum_{s=0}^{k-1} \beta_{ks} = \sum_{s=0}^{k-1} \frac{\alpha_s}{A_s} = 1, \quad \beta_{ks} \geq 0, \quad \forall k, s,$$

e, portanto, usando o Lema 4.9, com $b^s = dist(x^s, \Omega(\mu^*))$ e $b = 0$, seque-se que

$$0 \leq dist(\bar{x}^k, \Omega(\mu^*)) \leq A_k^{-1} \sum_{s=0}^{k-1} \alpha_s dist(x^s, \Omega(\mu^*)) \rightarrow 0, \quad \text{quando} \quad k \rightarrow \infty,$$

assim,

$$\{dist(\bar{x}^k, \Omega(\mu^*))\} \rightarrow 0 \quad \text{quando} \quad k \rightarrow \infty.$$

Segue-se que para qualquer ponto de acumulação x^* da sequência $\{\bar{x}^k\}$ temos que

$$x^* \in \Omega(\mu^*). \quad (4-18)$$

Em seguida, mostramos que para qualquer ponto de acumulação x^* da sequência ergódica $\{\bar{x}^k\}$, x^* é viável para o problema primal (4-1).

Pela convexidade das funções g_j , $j = 1, \dots, m$, temos que

$$g(\bar{x}^k) \leq A_k^{-1} \sum_{s=0}^{k-1} \alpha_s g(x^s) \quad \forall k,$$

e a partir da fórmula de iteração do método (Algoritmo) 2, segue-se que

$$\mu^{s+1} = [\mu^s + \alpha_s g(x^s)]^+ \geq \mu^s + \alpha_s g(x^s) \quad \forall s,$$

de onde obtemos

$$g(x^s) \leq \frac{\mu^{s+1} - \mu^s}{\alpha_s} \quad \forall s.$$

Por isso,

$$g(\bar{x}^k) \leq A_k^{-1} \sum_{s=0}^{k-1} (\mu^{s+1} - \mu^s) = A_k^{-1} (\mu^k - \mu^0) \quad \forall k.$$

Da Proposição 4.5 temos que a sequência $\{\mu^k - \mu^0\}$ é limitada e portanto,

$$\limsup_{k \rightarrow \infty} g_j(\bar{x}^k) \leq \limsup_{k \rightarrow \infty} \frac{\mu^k - \mu^0}{\sum_{s=0}^{k-1} \alpha_s} = 0, \quad \forall j = 1, \dots, m, \quad (4-19)$$

portanto,

$$g(x^*) \leq 0. \quad (4-20)$$

Agora queremos mostrar que $\mu^{*t} g(x^*) = 0$. Consideremos um $j \in I(\mu^*)$. Como μ^* é o limite da sequência $\{\mu^k\}$, segue que, para algum k_1 fixo suficientemente grande, $\mu_j^k > 0 \forall k \geq k_1$, e pela relação (4-13) temos que

$$v_j^k \mu_j^k = 0 \quad \text{para todo } k.$$

Portanto, da formula de iteração do método (Algoritmo) 2,

$$g_j(x^k) = \frac{\mu_j^{k+1} - \mu_j^k}{\alpha_k}, \quad \forall k \geq k_1. \quad (4-21)$$

Escolhendo $\bar{x} \in \text{rint}\Omega(\mu^*)$ e $\gamma_j \in \partial g_j(\bar{x})$, do Lema 4.2 temos que

$$g_j(x) = g_j(\bar{x}) + \gamma_j^t(x - \bar{x}), \quad \forall x \in \Omega(\mu^*).$$

Logo, pela continuidade da função g_j , para cada $\varepsilon > 0$ existe $\delta > 0$ tal que

$$g_j(x) \leq g_j(\bar{x}) + \gamma_j^t(x - \bar{x}) + \frac{\varepsilon}{3}, \quad \forall x : \text{dist}(x, \Omega(\mu^*)) \leq \delta. \quad (4-22)$$

Da relação (4-17) segue que

$$\text{dist}(x^s, \Omega(\mu^*)) \leq \delta \quad \text{para todo } s \geq k_2, \quad \text{para algum fixo } k_2 > k_1.$$

As relações (4-21) e (4-22) implicam que

$$\begin{aligned} g_j(\bar{x}) + \gamma'_j(x^s - \bar{x}) &\geq g_j(x^s) - \frac{\varepsilon}{3} \\ &= \frac{\mu_j^{s+1} - \mu_j^s}{\alpha_s} - \frac{\varepsilon}{3}, \quad \forall s \geq k_2. \end{aligned} \quad (4-23)$$

Usando a definição (4-16), temos que para todo $k \geq k_2$,

$$\begin{aligned} g_j(\bar{x}^k) &\geq g_j(\bar{x}) + \gamma'_j(\bar{x}^k - \bar{x}) \quad (\text{Da definição de subgradiente}) \\ &= g_j(\bar{x}) + \gamma'_j\left(A_k^{-1} \sum_{s=0}^{k-1} \alpha_s x^s - \bar{x}\right) \quad (\text{Da definição de sequência ergódica}) \end{aligned}$$

Como $A_k^{-1} \sum_{s=0}^{k-1} \alpha_s g_j(\bar{x}) = g_j(\bar{x}) A_k^{-1} \sum_{s=0}^{k-1} \alpha_s = g_j(\bar{x})$, então

$$\begin{aligned} g_j(\bar{x}^k) &\geq A_k^{-1} \sum_{s=0}^{k-1} \alpha_s g_j(\bar{x}) + A_k^{-1} \sum_{s=0}^{k-1} \alpha_s \gamma'_j(x^s - \bar{x}) \\ &= A_k^{-1} \sum_{s=0}^{k-1} \alpha_s (g_j(\bar{x}) + \gamma'_j(x^s - \bar{x})) \\ &= A_k^{-1} \sum_{s=0}^{k_2-1} \alpha_s (g_j(\bar{x}) + \gamma'_j(x^s - \bar{x})) + A_k^{-1} \sum_{s=k_2}^{k-1} \alpha_s (g_j(\bar{x}) + \gamma'_j(x^s - \bar{x})) \end{aligned}$$

Pela desigualdade (4-23), obtemos

$$\begin{aligned} &A_k^{-1} \sum_{s=0}^{k_2-1} \alpha_s (g_j(\bar{x}) + \gamma'_j(x^s - \bar{x})) + A_k^{-1} \sum_{s=k_2}^{k-1} \alpha_s (g_j(\bar{x}) + \gamma'_j(x^s - \bar{x})) \\ &\geq A_k^{-1} \sum_{s=0}^{k_2-1} \alpha_s (g_j(\bar{x}) + \gamma'_j(x^s - \bar{x})) + A_k^{-1} \sum_{s=k_2}^{k-1} \left(\mu_j^{s+1} - \mu_j^s - \frac{\alpha_s \varepsilon}{3} \right) \end{aligned}$$

Logo, da igualdade

$$\begin{aligned}
& A_k^{-1} \sum_{s=0}^{k_2-1} \alpha_s \gamma_j^t(x^s - \bar{x}) \\
&= A_k^{-1} \gamma_j^t \left(\frac{A_{k_2}}{A_k} \sum_{s=0}^{k_2-1} \alpha_s x^s - A_{k_2} \bar{x} \right) \\
&= \frac{A_{k_2}}{A_k} \gamma_j^t(\bar{x}^{k_2} - \bar{x})
\end{aligned}$$

temos que,

$$\begin{aligned}
g_j(\bar{x}^k) &\geq \frac{A_{k_2}}{A_k} \left(g_j(\bar{x}) + \gamma_j^t(\bar{x}^{k_2} - \bar{x}) \right) + \frac{\mu_j^k - \mu_j^{k_2}}{A_k} - \frac{\varepsilon}{3} A_k^{-1} \sum_{s=k_2}^{k-1} \alpha_s \\
&= \frac{A_{k_2}}{A_k} \left(g_j(\bar{x}) + \gamma_j^t(\bar{x}^{k_2} - \bar{x}) \right) + \frac{\mu_j^k - \mu_j^{k_2}}{A_k} - \frac{\varepsilon}{3} A_k^{-1} \left(\sum_{s=0}^{k-1} \alpha_s - \sum_{s=0}^{k_2-1} \alpha_s \right) \\
&= \frac{A_{k_2}}{A_k} \left(g_j(\bar{x}) + \gamma_j^t(\bar{x}^{k_2} - \bar{x}) \right) + \frac{\mu_j^k - \mu_j^{k_2}}{A_k} - \frac{\varepsilon}{3} A_k^{-1} (A_k - A_{k_2}) \\
&= \frac{A_{k_2}}{A_k} \left(g_j(\bar{x}) + \gamma_j^t(\bar{x}^{k_2} - \bar{x}) \right) + \frac{\mu_j^k - \mu_j^{k_2}}{A_k} - \frac{\varepsilon}{3} \left(1 - \frac{A_{k_2}}{A_k} \right).
\end{aligned}$$

Como $\lim_{k \rightarrow \infty} A_k = \lim_{k \rightarrow \infty} \sum_{s=0}^{k-1} \alpha_s = \infty$ e $\{\mu_j^k\} \rightarrow \mu_j^*$, temos que para k_2 suficientemente grande

- $A_k^{-1} A_{k_2} (g_j(\bar{x}) + \gamma_j^t(\bar{x}^{k_2} - \bar{x})) \geq -\frac{\varepsilon}{3}$
- $A_k^{-1} (\mu_j^k - \mu_j^{k_2}) \geq -\frac{\varepsilon}{3}$
- $-\frac{\varepsilon}{3} \left(1 - \frac{A_{k_2}}{A_k} \right) \geq -\frac{\varepsilon}{3}$.

Segue-se que $g_j(\bar{x}^k) \geq -\varepsilon$ para todo $k > k_2$, que são suficientemente grandes. Portanto, $\liminf_{k \rightarrow \infty} g_j(\bar{x}^k) \geq 0$ junto com a desigualdade (4-19) implicam que

$$\lim_{k \rightarrow \infty} g_j(\bar{x}^k) = 0.$$

Uma vez que este resultado é válido para todo $j \in I(\mu^*)$, segue-se que

$$\left\{ \mu^{*t} g(\bar{x}^k) \right\} \rightarrow 0 \quad \text{quando} \quad k \rightarrow \infty. \quad (4-24)$$

Assim, das relações (4-18), (4-20), e (4-24) segue que qualquer ponto de acumulação da sequência ergódica é uma solução primal para qualquer $\mu^* \in U^*$. \square

O seguinte resultado segue da proposição 4.5, Teorema 4.10, e da relação (4-11).

Corolário 4.11 *Sob as hipóteses do Teorema 4.10,*

$$\left\{ \text{dist} \left(g(\bar{x}^k), \partial q(\mu^*) \cap N_{\mathbb{R}_+^m}(\mu^*) \right) \right\} \rightarrow 0.$$

Exemplo 4.12 *(Contra-exemplo para passo de Polyak) Do exemplo 4.7 temos que $x^k = 0$ $\forall k$, assim, fazendo a sequência ergódica temos*

$$\bar{x}^k = 0,$$

logo $\{\bar{x}^k\} \rightarrow 0$, mas $x^ = 1$.*

Portanto, não podemos garantir os resultados de convergência provados nesta seção com sequência ergódica usando a regra de passo de Polyak para a escolha do tamanho do passo.

Soluções primais duais aproximadas

Neste capítulo, nosso interesse está em resolver aproximadamente o problema primal usando sequência ergódica dos vetores primais obtidos ao resolver o subproblema dual. Aproximamos soluções primais sem comportamento assintótico. Estamos focados no comprimento de passo *constante* para algoritmos subgradientes duais, principalmente por sua importância prática e simplicidade. Mostramos que a sequência dual gerada pelo método subgradiente é limitada sob a *qualificação do Slater*. Usamos este resultado para estimar a aproximação de erros em termos da viabilidade e otimalidade primal e proporcionamos estimativas por-iteração para a violação das restrições, e cotas superior e inferior sobre a função objetivo da sequência ergódica primal.

Consideramos um método subgradiente alternativo sob a *qualificação do Slater*. Este método usa o fato que o conjunto ótimo dual é limitado projetando os iterados duais em um conjunto relativamente simples contendo o conjunto ótimo dual. Neste método, também usamos sequência ergódica para gerar soluções primais aproximadas e proporcionamos estimativas dos erros sobre a quantidade da violação das restrições.

5.1 Problemas primal e dual

Consideremos o seguinte problema de otimização com estrutura:

$$f^* = \min f(x) \tag{5-1a}$$

$$s.a \ x \in \Omega \tag{5-1b}$$

$$g_j(x) \leq 0, \quad j = 1, 2, \dots, m, \tag{5-1c}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ ($j = 1, 2, \dots, m$) são funções convexas, e $\Omega \subset \mathbb{R}^n$ é um conjunto convexo fechado não vazio. Neste capítulo vamos supor que o valor ótimo primal f^* é finito.

Usaremos dualidade Lagrangiana para relaxar o problema primal (5-1). Como vimos no capítulo 2, a função Lagrangiana é

$$L(x, \mu) = f(x) + \mu^t g(x).$$

e a função dual é definida por

$$q(\mu) = \min_{x \in \Omega} L(x, \mu),$$

onde $\mu \in \mathbb{R}_+^m$ é um multiplicador de Lagrange. O problema dual é definido por

$$\begin{aligned} q^* &= \max q(\mu) \\ \text{s.a } &\mu \geq 0 \end{aligned} \quad (5-2)$$

onde denotamos o conjunto dual viável como $U = \{ \mu / \mu \geq 0 \mid -\infty < q(\mu) < \infty \}$, o valor ótimo dual por q^* e o conjunto ótimo dual por U^* .

Vamos assumir que o subproblema dual Lagrangiano tem solução para cada $\mu \geq 0$. Por exemplo, se Ω é compacto então essa hipótese é satisfeita.

5.1.1 Qualificação de Slater e conjunto de multiplicadores limitado

Nesta seção vamos considerar conjuntos da forma $\{ \mu \geq 0 / q(\mu) \geq q(\bar{\mu}) \}$ para um $\bar{\mu} \geq 0$ fixo, que são obtidos pela interseção do octante não negativo no \mathbb{R}^m e conjuntos de supernível da função dual q . Mostramos que esses conjuntos são limitados quando o problema primal satisfaz a definição 2.45, i.e., a qualificação de Slater a qual vai ser uma hipótese importante ao longo deste capítulo.

Hipótese 1 (Qualificação de Slater) Existe um vetor $\bar{x} \in \mathbb{R}^n$ tal que

$$g_j(\bar{x}) < 0 \text{ para todo } j = 1, 2, \dots, m. \quad (5-3)$$

Nós nos referimos a um vetor \bar{x} que satisfaz a qualificação de Slater como um vetor Slater.

Sob as hipóteses de que f é finita, e a qualificação de Slater pelo Teorema 2.46 temos que não há brecha de dualidade, e o conjunto de solução ótima dual é não vazio. O seguinte lema, em particular, mostra que a qualificação de Slater garante também que os conjuntos de nível $\{ \mu \geq 0 / q(\mu) \geq q(\bar{\mu}) \}$ sejam limitados.

Lema 5.1 *Suponha que vale a qualificação de Slater (Hipótese 1). Seja $\bar{\mu} \geq 0$ um vetor tal que o conjunto $Q_{\bar{\mu}} = \{ \mu \geq 0 / q(\mu) \geq q(\bar{\mu}) \}$ é não vazio. Então, o conjunto $Q_{\bar{\mu}}$ é limitado*

e em particular, temos que

$$\max_{\mu \in Q_{\bar{\mu}}} \|\mu\| \leq \frac{1}{\gamma} (f(\bar{x}) - q(\bar{\mu})), \quad (5-4)$$

onde $\gamma = \min_{1 \leq j \leq m} \{-g_j(\bar{x})\}$ e \bar{x} é um vetor Slater.

Demonstração. Seja $\mu \in Q_{\bar{\mu}}$, temos que

$$q(\bar{\mu}) \leq q(\mu) = \inf_{x \in \Omega} f(x) + \mu^T g(x) \leq f(\bar{x}) + \mu^T g(\bar{x}) = f(\bar{x}) + \sum_{j=1}^m \mu_j g_j(\bar{x}),$$

o que implica

$$-\sum_{j=1}^m \mu_j g_j(\bar{x}) \leq f(\bar{x}) - q(\bar{\mu}).$$

Como $g_j(\bar{x}) < 0$ e $\mu_j \geq 0$ para todo j , segue que

$$\min_{1 \leq j \leq m} \{-g_j(\bar{x})\} \sum_{j=1}^m \mu_j \leq -\sum_{j=1}^m \mu_j g_j(\bar{x}) \leq f(\bar{x}) - q(\bar{\mu}).$$

Portanto,

$$\sum_{j=1}^m \mu_j \leq \frac{f(\bar{x}) - q(\bar{\mu})}{\min_{1 \leq j \leq m} \{-g_j(\bar{x})\}}.$$

Como $\mu \geq 0$, temos que $\|\mu\| \leq \sum_{j=1}^m \mu_j$. (i.e., norma Euclidiana menor ou igual que norma da soma).

Logo,

$$\|\mu\| \leq \frac{1}{\gamma} (f(\bar{x}) - q(\bar{\mu})),$$

onde $\gamma = \min_{1 \leq j \leq m} \{-g_j(\bar{x})\}$ e daqui, o conjunto $Q_{\bar{\mu}}$ é limitado, além disso é fechado e, assim compacto, logo

$$\max_{\mu \in Q_{\bar{\mu}}} \|\mu\| \leq \frac{1}{\gamma} (f(\bar{x}) - q(\bar{\mu})).$$

□

Segue-se do Lema que sob a qualificação de Slater, o conjunto ótimo dual

$$U^* = \{\mu \geq 0 / q(\mu) \geq q^*\}$$

é não vazio, e

$$\max_{\mu^* \in U^*} \|\mu^*\| \leq \frac{1}{\gamma} (f(\bar{x}) - q^*), \quad (5-5)$$

com $\gamma = \min_{1 \leq j \leq m} \{-g_j(\bar{x})\}$.

Na prática, o valor ótimo dual q^* em geral não é conhecido. No entanto, com um valor funcional dual $q(\tilde{\mu})$ para algum $\tilde{\mu} \geq 0$, podemos proporcionar uma cota para a norma das soluções ótimas duais. Em particular, como $q^* \geq q(\tilde{\mu})$, da relação (5-5) obtemos a seguinte cota

$$\max_{\mu^* \in U^*} \|\mu^*\| \leq \frac{1}{\gamma} (f(\bar{x}) - q(\tilde{\mu})).$$

Além disso, com uma sequência de multiplicadores $\{\mu^k\}$, podemos usar os valores da função dual $q(\mu^k)$ para gerar uma sequência de cotas superiores da norma das soluções ótimas duais $\|\mu^*\|$. Formalmente, uma vez que $q^* \geq \max_{0 \leq i \leq k} q(\mu^i)$, da relação (5-5) temos que

$$\max_{\mu^* \in U^*} \|\mu^*\| \leq \frac{1}{\gamma} \left(f(\bar{x}) - \max_{0 \leq i \leq k} q(\mu^i) \right) \quad \text{para todo } k \geq 0.$$

Observação 5.2 *Do Lema 5.1 temos que se o vetor Slater \bar{x} é uma solução do problema primal, então da relação (5-5) obtemos*

$$\begin{aligned} \|\mu^*\| &\leq \frac{1}{\gamma} (f(\bar{x}) - q^*), \\ &= \frac{1}{\gamma} (f^* - q^*), \\ &= 0. \quad (\text{Pela dualidade forte, i.e., } f^* = q^*) \end{aligned}$$

Portanto, o conjunto ótimo dual é igual ao conjunto $\{0\}$.

5.2 Método Subgradiente

Para resolver o problema dual, consideramos o algoritmo subgradiente clássico com comprimento de passo constante, isto é consideramos o algoritmo 2 com $v^k = 0$ e regra de passo constante. Lembremos que dado μ^k obtemos o próximo iterado como

$$\mu^{k+1} = [\mu^k + \alpha g^k]^+ \quad \text{para } k = 0, 1, \dots, \quad (5-6)$$

onde o vetor $\mu^0 \geq 0$ é uma iteração inicial e o escalar $\alpha > 0$ é o comprimento de passo. Pela Proposição 4.3 temos que o conjunto $\partial q(\mu)$ de subgradientes de q em μ é dado por

$$\partial q(\mu) = \text{conv}\{ g(x) / x \in \Omega(\mu) \}, \quad (5-7)$$

onde $\Omega(\mu) = \{x \in \Omega / f(x) + \mu^t g(x) = q(\mu)\}$ é o conjunto solução do subproblema dual.

O comprimento de passo constante α é um parâmetro simples de controlar, e escolhendo um valor apropriado de α , o método subgradiente gera uma sequência que se aproxima do valor ótimo em um número finito de iterações, como veremos.

5.2.1 Multiplicadores limitados

Hipótese 2 (Subgradientes limitados) A sequência de subgradientes $\{g^k\}$ é limitada, *i.e.*, existe um escalar $L > 0$ tal que

$$\|g^k\| \leq L \text{ para todo } k \geq 0.$$

Esta hipótese é satisfeita, por exemplo, quando o conjunto Ω é compacto. Neste caso, devido à convexidade das funções g_j sobre \mathbb{R}^n , cada g_j é contínua sobre \mathbb{R}^n . Assim, $\max_{x \in \Omega} \|g(x)\|$ é finito e proporciona uma cota superior sobre as normas dos subgradientes g^k , e, portanto

$$L = \max_{x \in \Omega} \|g(x)\| \quad \text{ou} \quad L = \max_{1 \leq j \leq m} \max_{x \in \Omega} |g_j(x)|.$$

No seguinte lema, estabelecemos que a sequência de multiplicadores é limitada.

Lema 5.3 (Multiplicadores limitados) *Seja a sequência de multiplicadores $\{\mu^k\}$ gerada pelo método subgradiente usual (usando a fórmula de iteração (5-6)). Sob as hipóteses 1 e 2, temos que a sequência $\{\mu^k\}$ é limitada e, além disso,*

$$\|\mu^k\| \leq \frac{2}{\gamma}(f(\bar{x}) - q^*) + \max \left\{ \|\mu^0\|, (f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \alpha L \right\},$$

onde $\gamma = \min_{1 \leq j \leq m} \{-g_j(\bar{x})\}$, onde L é a cota da norma da sequência de subgradientes da hipótese 2, \bar{x} é um vetor que satisfaz a qualificação de Slater, e $\alpha > 0$ é o comprimento de passo.

Demonstração. Sob a qualificação de Slater o conjunto ótimo dual U^* é não vazio. Consideremos o conjunto Q_α definido por

$$Q_\alpha = \left\{ \mu \geq 0 / q(\mu) \geq q^* - \frac{\alpha L^2}{2} \right\},$$

que não é vazio, levando em conta que $U^* \subset Q_\alpha$. Fixamos um $\mu^* \in U^*$ arbitrário e vamos provar primeiro que para todo $k \geq 0$,

$$\|\mu^k - \mu^*\| \leq \max \left\{ \|\mu^0 - \mu^*\|, \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \|\mu^*\| + \alpha L \right\}, \quad (5-8)$$

Usamos indução em k . Note que a relação é válida para $k = 0$. Suponha que vale para algum $k > 0$, ou seja,

$$\|\mu^k - \mu^*\| \leq \max \left\{ \|\mu^0 - \mu^*\|, \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \|\mu^*\| + \alpha L \right\} \text{ para algum } k > 0 \quad (5-9)$$

Consideramos 2 casos: $q(\mu^k) \geq q^* - \alpha L^2/2$ e $q(\mu^k) < q^* - \alpha L^2/2$.

Caso 1: $q(\mu^k) \geq q^* - \alpha L^2/2$. Usando a fórmula de iteração (5-6) para μ^{k+1} e a hipótese 2, obtemos

$$\|\mu^{k+1} - \mu^*\| \leq \|\mu^k + \alpha g^k - \mu^*\| \leq \|\mu^k\| + \|\mu^*\| + \alpha L.$$

Como $q(\mu^k) \geq q^* - \alpha L^2/2$, segue-se que $\mu^k \in Q_\alpha$. Pelo Lema 5.1, o conjunto Q_α é limitado e, em particular, $\|\mu\| \leq \frac{1}{\gamma}(f(\bar{x}) - q^* + \alpha L^2/2)$ para todo $\mu \in Q_\alpha$. Portanto

$$\|\mu^k\| \leq \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma}.$$

Combinando as duas relações anteriores, obtemos

$$\|\mu^{k+1} - \mu^*\| \leq \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \|\mu^*\| + \alpha L,$$

mostrando assim que a estimativa em equação (5-8) é válido para $k + 1$.

Caso 2: $q(\mu^k) \leq q^* - \alpha L^2/2$. Usando a propriedade de não expansividade da projeção temos para todo k ,

$$\begin{aligned} \|\mu^{k+1} - \mu^*\|^2 &\leq \|\mu^k + \alpha g^k - \mu^*\|^2 \\ &\leq \|\mu^k - \mu^*\|^2 + 2\alpha g^k(\mu^k - \mu^*) + \alpha^2 \|g^k\|^2 \end{aligned}$$

Como g^k é um subgradiente (supergradiente) de q em μ^k , temos que

$$g^{k^t}(\mu^* - \mu^k) \geq q(\mu^*) - q(\mu^k),$$

o que implica que

$$g^{k^t}(\mu^k - \mu^*) \leq -(q(\mu^*) - q(\mu^k)).$$

Assim, obtemos

$$\|\mu^{k+1} - \mu^*\|^2 \leq \|\mu^k - \mu^*\|^2 - 2\alpha(q^* - q(\mu^k)) + \alpha^2 \|g^k\|^2. \quad (5-10)$$

Usando a hipótese 2, nós ainda obtemos

$$\|\mu^{k+1} - \mu^*\|^2 \leq \|\mu^k - \mu^*\|^2 - 2\alpha \left(q^* - q(\mu^k) - \frac{\alpha L^2}{2} \right),$$

como $q(\mu^k) < q^* - \alpha L^2/2$, segue-se que $q^* - q(\mu^k) - \alpha L^2/2 > 0$, que combinado com a relação anterior nos dá

$$\|\mu^{k+1} - \mu^*\| < \|\mu^k - \mu^*\|.$$

Pela hipótese de indução (5-9), segue-se que a estimativa da equação (5-8) é válida para $k + 1$. Assim, a estimativa da equação (5-8) vale para todo $k \geq 0$. Da equação (5-8) obtemos para todo $k \geq 0$,

$$\|\mu^k\| \leq \|\mu^k - \mu^*\| + \|\mu^*\| \leq \max \left\{ \|\mu^0 - \mu^*\|, \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \|\mu^*\| + \alpha L \right\} + \|\mu^*\|.$$

Usando $\|\mu^0 - \mu^*\| \leq \|\mu^0\| + \|\mu^*\|$, temos, ainda, para todo $k \geq 0$,

$$\begin{aligned} \|\mu^k\| &\leq \max \left\{ \|\mu^0\| + \|\mu^*\|, \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \|\mu^*\| + \alpha L \right\} + \|\mu^*\| \\ &= 2\|\mu^*\| + \max \left\{ \|\mu^0\|, \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \alpha L \right\}. \end{aligned}$$

Como $U^* = \{\mu \geq 0 / q(\mu) \geq q^*\}$, de acordo com o Lema 5.1, temos a seguinte cota sobre as soluções ótimas duais

$$\max_{\mu^* \in U^*} \|\mu^*\| \leq \frac{1}{\gamma}(f(\bar{x}) - q^*),$$

o que implica que para todo $k \geq 0$,

$$\|\mu^k\| \leq \frac{2}{\gamma}(f(\bar{x}) - q^*) + \max \left\{ \|\mu^0\|, \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \alpha L \right\}.$$

□

Como comentamos ao final do Lema 5.1, na prática, o valor ótimo dual q^* em geral não é conhecido, tomando $\tilde{\mu} = \mu^0$ temos $q^* \geq q(\mu^0)$, substituindo q^* com $q(\mu^0)$, obtemos a seguinte cota para a sequência de multiplicadores

$$\|\mu^k\| \leq \frac{2}{\gamma}(f(\bar{x}) - q(\mu^0)) + \max \left\{ \|\mu^0\|, \frac{1}{\gamma}(f(\bar{x}) - q(\mu^0)) + \frac{\alpha L^2}{2\gamma} + \alpha L \right\},$$

onde $\gamma = \min_{1 \leq j \leq m} \{-g_j(\bar{x})\}$. Note que esta cota só depende dos parâmetros do algoritmo e dados do problema. Especificamente, envolve a primeira iteração μ^0 do método, o passo constante α , o vetor Slater \bar{x} , e a cota da norma subgradiente L .

5.3 Soluções primais aproximadas

Nesta seção, nós proporcionamos soluções primais aproximadas considerando sequência ergódica.

Consideremos a sequência de multiplicadores $\{\mu^k\}$ gerada pelo algoritmo subgradiente usual (fórmula de iteração dada por (5-6)), e a sequência de vetores primais $\{x^k\}$ obtida durante a resolução do subproblema dual Lagrangiano. Lembremos que a sequência ergódica \bar{x}^k dos vetores x^0, \dots, x^{k-1} , é dada por

$$\bar{x}^k = \frac{1}{k} \sum_{s=0}^{k-1} x^s \quad \text{para todo } k \geq 1. \quad (5-11)$$

Note que a sequência ergódica \bar{x}^k fica no conjunto Ω , pois, Ω é convexo e cada $x^s \in \Omega$ para todo s . No entanto, estes vetores não precisam satisfazer as restrições primais $g_j(x) \leq 0$, $j=0, \dots, m$, e portanto, eles podem não ser viáveis.

No que segue desta seção, estudamos algumas propriedades básicas da sequência ergódica \bar{x}^k . Usando estas propriedades e a qualificação de Slater, nós proporcionamos estimativas para o valor ótimo primal e estimativas para a violação da viabilidade em cada iteração do método subgradiente.

5.3.1 Propriedades básicas da sequência ergódica primal

A seguinte proposição proporciona cotas superiores e inferiores sobre os valores funcionais primais da sequência ergódica \bar{x}^k e cotas superiores e inferiores para a quantidade da violação das restrições. Note que estas cotas são dadas por iteração, podendo assim obter um critério de parada prático para o método. Assim, essas cotas podem nos proporcionar informações práticas para estimar soluções ótimas.

Proposição 5.4 *Seja a sequência de multiplicadores $\{\mu^k\}$ gerada pelo método subgradiente usual com regra de passo constante. Seja a sequência ergódica $\{\bar{x}^k\}$. Então, para todo $k \geq 1$,*

- a) *uma cota superior sobre a quantidade da violação das restrições do vetor \bar{x}^k é dado por*

$$\|g(\bar{x}^k)^+\| \leq \frac{\|\mu^k\|}{k\alpha},$$

b) uma cota superior sobre os valores funcionais primais do vetor \bar{x}^k é dado por

$$f(\bar{x}^k) \leq q^* + \frac{\|\mu^0\|^2}{2k\alpha} + \frac{\alpha}{2k} \sum_{s=0}^{k-1} \|g(x^s)\|^2.$$

c) uma cota inferior para os valores funcionais primais do vetor \bar{x}^k é dado por

$$f(\bar{x}^k) \geq q^* - \|\mu^*\| \|g(\bar{x}^k)^+\|,$$

onde μ^* é uma solução dual ótima.

Demonstração.

(a) Usando a fórmula de iteração para μ^{k+1} , obtemos

$$\mu^k + \alpha g^k \leq [\mu^k + \alpha g^k]^+ = \mu^{k+1} \quad \text{para todo } k \geq 0.$$

Como $g^k = g(x^k)$ com $x^k \in \Omega$, segue-se que

$$\alpha g(x^k) \leq \mu^{k+1} - \mu^k \quad \text{para todo } k \geq 0. \quad (5-12)$$

Portanto,

$$\sum_{s=0}^{k-1} \alpha g(x^s) \leq \mu^k - \mu^0 \leq \mu^k \quad \text{para todo } k \geq 1,$$

logo,

$$\sum_{s=0}^{k-1} g(x^s) \leq \frac{\mu^k}{\alpha}, \quad \text{para todo } k \geq 1,$$

Como $x^k \in \Omega$ para todo k , pela convexidade de Ω , temos que $\bar{x}^k \in \Omega$ para todo k .

Assim, pela convexidade de cada uma das funções g_j , segue-se que

$$g(\bar{x}^k) \leq \frac{1}{k} \sum_{s=0}^{k-1} g(x^s) \leq \frac{\mu^k}{k\alpha} \quad \text{para todo } k \geq 1,$$

e como $\mu^k \geq 0$ para todo k , temos que $g(\bar{x}^k)^+ \leq \mu^k / (k\alpha)$ para todo $k \geq 1$ e, portanto,

$$\|g(\bar{x}^k)^+\| \leq \frac{\|\mu^k\|}{k\alpha} \quad \text{para todo } k \geq 1.$$

(b) Pela convexidade do funcional primal $f(x)$ e pela definição de \bar{x}^k como um minimizador da função Lagrangiana $f(x) + (\mu^k)^t g(x)$ sobre $x \in \Omega$, temos que

$$f(\bar{x}^k) \leq \frac{1}{k} \sum_{s=0}^{k-1} f(x^s) = \frac{1}{k} \sum_{s=0}^{k-1} \{f(x^s) + (\mu^s)^t g(x^s)\} - \frac{1}{k} \sum_{s=0}^{k-1} (\mu^s)^t g(x^s).$$

Como $q(\mu^s) = f(x^s) + (\mu^s)^t g(x^s)$ e $q(\mu^s) \leq q^*$ para todo s , segue-se que para todo $k \geq 1$,

$$f(\bar{x}^k) \leq \frac{1}{k} \sum_{s=0}^{k-1} q(x^s) - \frac{1}{k} \sum_{s=0}^{k-1} (\mu^s)^t g(x^s) \leq q^* - \frac{1}{k} \sum_{s=0}^{k-1} (\mu^s)^t g(x^s). \quad (5-13)$$

Da fórmula de iteração (5-6), usando a propriedade de não expansividade da projeção, e o fato que $0 \in \{\mu \in \mathbb{R}^m / \mu \geq 0\}$ e $g^s = g(x^s)$, obtemos

$$\|\mu^{s+1}\|^2 \leq \|\mu^s\|^2 + 2\alpha(\mu^s)^t g(x^s) + \alpha^2 \|g(x^s)\|^2 \quad \text{para todo } s \geq 0,$$

no qual implica que

$$-(\mu^s)^t g(x^s) \leq \frac{\|\mu^s\|^2 - \|\mu^{s+1}\|^2 + \alpha^2 \|g(x^s)\|^2}{2\alpha} \quad \text{para todo } s \geq 0.$$

Somando sobre $s=0, \dots, k-1$ para $k \geq 1$, temos que

$$-\frac{1}{k} \sum_{s=0}^{k-1} (\mu^s)^t g(x^s) \leq \frac{\|\mu^0\|^2 - \|\mu^k\|^2}{2k\alpha} + \frac{\alpha}{2k} \sum_{s=0}^{k-1} \|g(x^s)\|^2 \quad \text{para todo } k \geq 1.$$

Combinando a equação anterior e a equação (5-13), nós ainda obtemos

$$\begin{aligned} f(\bar{x}^k) &\leq q^* + \frac{\|\mu^0\|^2 - \|\mu^k\|^2}{2k\alpha} + \frac{\alpha}{2k} \sum_{s=0}^{k-1} \|g(x^s)\|^2 \quad \text{para todo } k \geq 1, \\ &\leq q^* + \frac{\|\mu^0\|^2}{2k\alpha} + \frac{\alpha}{2k} \sum_{s=0}^{k-1} \|g(x^s)\|^2 \quad \text{para todo } k \geq 1, \end{aligned}$$

implicando a estimativa desejada.

(c) Dada uma solução ótima dual μ^* , temos que

$$f(\bar{x}^k) = f(\bar{x}^k) + (\mu^*)^t g(\bar{x}^k) - (\mu^*)^t g(\bar{x}^k) \geq q(\mu^*) - (\mu^*)^t g(\bar{x}^k).$$

Já que $\mu^* \geq 0$ e $g(\bar{x}^k)^+ \geq g(\bar{x}^k)$, temos ainda que

$$-(\mu^*)^t g(\bar{x}^k) \geq -(\mu^*)^t g(\bar{x}^k)^+ \geq -\|\mu^*\| \|g(\bar{x}^k)^+\|.$$

Das duas relações anteriores e do fato que $q(\mu^*) = q^*$ segue-se que

$$f(\bar{x}^k) \geq q^* - \|\mu^*\| \|g(\bar{x}^k)^+\|.$$

□

Note que os resultados da proposição 5.4 indicam que as cotas sobre a violação de viabilidade e o valor primal $f(\bar{x}^k)$ são facilmente calculados se temos as cotas sobre as normas dos multiplicadores $\|\mu^k\|$, da solução ótima $\|\mu^*\|$, e subgradientes $\|g(\bar{x}^k)\|$. Isto é precisamente o que usamos na próxima seção para estabelecer novas estimativas.

5.3.2 Propriedades da sequência ergódica primal sob a qualificação de Slater

Aqui, iremos fortalecer as relações da Proposição 5.4, sob a qualificação de Slater e subgradientes limitados.

Proposição 5.5 *Seja a sequência $\{\mu^k\}$ gerada pelo algoritmo subgradiente (fórmula de iteração dada por (5-6)). Suponha que valem a qualificação de Slater e a hipótese de subgradientes limitados (hipóteses 1 e 2). Além disso, seja*

$$B^* = \frac{2}{\gamma}(f(\bar{x}) - q^*) + \max \left\{ \|\mu^0\|, \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \alpha L \right\}, \quad (5-14)$$

onde $\gamma = \min_{1 \leq j \leq m} \{-g_j(\bar{x})\}$, \bar{x} é um vetor Slater da hipótese 1, e L é a cota sobre as normas dos subgradientes da hipótese 2. Considere também a sequência ergódica $\{\bar{x}^k\}$. Então para todo $k \geq 1$,

a) *uma cota superior sobre a quantidade da violação das restrições do vetor \bar{x}^k é dado por*

$$\|g(\bar{x}^k)^+\| \leq \frac{B^*}{k\alpha}.$$

b) *uma cota superior sobre os valores funcionais primais do vetor \bar{x}^k é dado por*

$$f(\bar{x}^k) \leq f^* + \frac{\|\mu^0\|^2}{2k\alpha} + \frac{\alpha L^2}{2}.$$

c) *uma cota inferior para os valores funcionais primais do vetor \bar{x}^k é dado por*

$$f(\bar{x}^k) \geq f^* - \frac{1}{\gamma}(f(\bar{x}) - q^*) \|g(\bar{x}^k)^+\|.$$

Demonstração.

(a) Sob as hipóteses da qualificação de Slater e subgradiente limitado, pelo Lema (5.3) temos que

$$\|\mu^k\| \leq \frac{2}{\gamma}(f(\bar{x}) - q^*) + \max \left\{ \|\mu^0\|, \frac{1}{\gamma}(f(\bar{x}) - q^*) + \frac{\alpha L^2}{2\gamma} + \alpha L \right\} \quad \text{para todo } k \geq 0.$$

Pela definição de B^* na equação (5-14), a relação anterior é equivalente a

$$\|\mu^k\| \leq B^* \quad \text{para todo } k \geq 0. \quad (5-15)$$

Usando a parte (a) da Proposição 5.4, obtemos

$$\|g(\bar{x}^k)^+\| \leq \frac{\|\mu^k\|}{k\alpha} \leq \frac{B^*}{k\alpha} \quad \text{para todo } k \geq 1.$$

(b) Da parte (b) da Proposição 5.4, obtemos

$$f(\bar{x}^k) \leq q^* + \frac{\|\mu^0\|^2}{2k\alpha} + \frac{\alpha}{2k} \sum_{s=0}^{k-1} \|g(x^s)\|^2 \quad \text{para todo } k \geq 1.$$

Sob a qualificação de Slater, existe dualidade forte, *i.e.*, $q^* = f^*$. Além disso, os subgradientes são limitados por um escalar L , de modo que

$$f(\bar{x}^k) \leq f^* + \frac{\|\mu^0\|^2}{2k\alpha} + \frac{\alpha L^2}{2} \quad \text{para todo } k \geq 1.$$

(c) Sob a qualificação de Slater, existe uma solução ótima dual e temos dualidade forte, *i.e.*, $q^* = f^*$. Assim, pela parte (c) da Proposição 5.4, para qualquer solução dual μ^* temos que

$$f(\bar{x}^k) \geq f^* - \|\mu^*\| \|g(\bar{x}^k)^+\| \quad \text{para todo } k \geq 1.$$

Usando o Lema 5.1 com $\bar{\mu} = \mu^*$, vemos que o conjunto ótimo dual é limitado e, $\|\mu^*\| \leq \frac{1}{\gamma}(f(\bar{x}) - q^*)$ para todo $\mu^* \in U^*$. Daqui,

$$f(\bar{x}^k) \geq f^* - \frac{1}{\gamma}(f(\bar{x}) - q^*) \|g(\bar{x}^k)^+\| \quad \text{para todo } k \geq 1.$$

□

Como ocorre também na parte (a) da Proposição 5.5, a quantidade da violação de viabilidade $\|g(\bar{x}^k)^+\|$ do vetor \bar{x}^k diminui a zero quando o número de iterações subgradientes k aumenta. Combinando os resultados em (a) – (c), vemos que os valores funcionais $f(\bar{x}^k)$ convergem para f^* no nível de erro $\alpha L^2/2$ quando $k \rightarrow \infty$.

5.4 Método subgradiente modificado sob a qualificação de Slater

Nesta seção, nós iremos considerar uma versão modificada do método subgradiente sob a qualificação de Slater. A motivação vem do fato de que, sob a qualificação de Slater, podemos "encontrar" o conjunto de soluções ótimas duais U^* . Portanto, é interessante considerar um método subgradiente no qual as iterações duais são projetadas sobre um superconjunto limitado do conjunto de soluções ótimas duais. A seguinte proposição proporciona estimativas para a quantidade da violação das restrições e os valores funcionais da sequência ergódica primal e vamos comparar essas estimativas com os erros das estimativas obtidas para o método subgradiente ordinário da seção anterior.

Formalmente, podemos considerar métodos subgradientes da seguinte forma

$$\mu^{k+1} = P_D(\mu^k + \alpha g^k), \quad (5-16)$$

onde o conjunto D é um conjunto convexo e compacto contendo o conjunto de soluções ótimas duais $U^* \subset D$ e P_D denota a projeção sobre o conjunto D .

Sob a qualificação de Slater, o conjunto ótimo dual U^* é não vazio e limitado, e uma cota sobre as normas das soluções ótimas duais é dado por

$$\sum_{s=0}^m (\mu^s)^* \leq \frac{1}{\gamma} (f(\bar{x}) - q^*) \quad \text{para todo } \mu^* \in U^*,$$

com $\gamma = \min_{1 \leq j \leq m} \{-g_j(\bar{x})\}$ e \bar{x} um vetor Slater. Assim, tendo o valor dual $\tilde{q} = q(\tilde{\mu})$ para algum $\tilde{\mu} \geq 0$, como $q^* \geq \tilde{q}$, obtemos

$$\sum_{s=0}^m (\mu_j^s)^* \leq \frac{1}{\gamma} (f(\bar{x}) - \tilde{q}) \quad \text{para todo } \mu^* \in U^*. \quad (5-17)$$

Isto motiva a seguinte escolha para o conjunto D

$$D = \left\{ \mu \geq 0 / \|\mu\| \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma} + r \right\}, \quad (5-18)$$

com um escalar $r > 0$. Claramente, o conjunto D é convexo e compacto, e contém o conjunto de soluções ótimas duais tendo em conta a relação (5-17).

A seguinte proposição, nos dá cotas por-iteração para a violação das restrições e valores funcionais primais da sequência ergódica \bar{x}^k .

Proposição 5.6 *Sejam as hipóteses da condição de Slater e a de subgradientes limitados válidas. Seja a sequência dual $\{\mu^k\}$ gerada pelo método subgradiente modificado da equação (5-16). Considere a sequência ergódica $\{\bar{x}^k\}$. Então, para todo $k \geq 1$,*

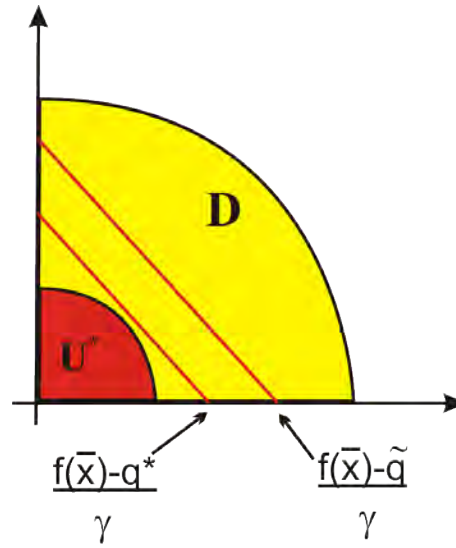


Figura 5.1: O conjunto ótimo dual U^* e o conjunto D .

a) uma cota superior sobre a quantidade da violação das restrições do vetor \bar{x}^k é dado por

$$\|g(\bar{x}^k)^+\| \leq \frac{2}{k\alpha r} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r \right)^2 + \frac{\alpha L^2}{2r}.$$

b) uma cota superior sobre os valores funcionais primais do vetor \bar{x}^k é dado por

$$f(\bar{x}^k) \leq f^* + \frac{\|\mu^0\|^2}{2k\alpha} + \frac{\alpha L^2}{2}.$$

c) uma cota inferior para os valores funcionais primais do vetor \bar{x}^k é dado por

$$f(\bar{x}^k) \geq f^* - \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right) \|g(\bar{x}^k)^+\|.$$

Aqui, os escalares $r > 0$ e \tilde{q} com $\tilde{q} \leq q^*$ são aqueles da definição do conjunto D na equação (5-18), $\gamma = \min_{1 \leq j \leq m} \{-g_j(\bar{x})\}$, \bar{x} é o vetor Slater da hipótese 1, e L é a cota sobre a norma subgradiente da hipótese 2.

Demonstração.

(a) Usando a definição da iteração μ^{k+1} na equação (5-16) e a propriedade de não expansividade da projeção sobre um conjunto convexo fechado, obtemos que para todo

$\mu \in D$ e todo $s \geq 0$,

$$\begin{aligned} \|\mu^{s+1} - \mu\|^2 &= \|P_D(\mu^s + \alpha g^s) - \mu\|^2 \\ &\leq \|\mu^s + \alpha g^s - \mu\|^2 \\ &\leq \|\mu^s - \mu\|^2 + 2\alpha(g^s)^t(\mu^s - \mu) + \alpha^2\|g^s\|^2 \\ &\leq \|\mu^s - \mu\|^2 + 2\alpha(g^s)^t(\mu^s - \mu) + \alpha^2L^2, \end{aligned}$$

onde a ultima desigualdade segue da hipótese 2 de subgradiente limitado. Portanto, para qualquer $\mu \in D$,

$$(g^s)^t(\mu - \mu^s) \leq \frac{\|\mu^s - \mu\|^2 - \|\mu^{s+1} - \mu\|^2}{2\alpha} + \frac{\alpha L^2}{2} \quad \text{para todo } s \geq 0. \quad (5-19)$$

Já que g^s é um subgradiente da função dual q em μ^s , usando a desigualdade subgradiente, obtemos para qualquer solução ótima dual μ^* ,

$$(g^s)^t(\mu^s - \mu^*) \leq q(\mu^s) - q(\mu^*) \leq 0 \quad \text{para todo } s \geq 0,$$

onde a ultima desigualdade segue da otimalidade de μ^* . Temos então, para todo $\mu \in D$ e todo $s \geq 0$,

$$(g^s)^t(\mu - \mu^*) = (g^s)^t(\mu - \mu^* - \mu^s + \mu^s) = (g^s)^t(\mu - \mu^s) + (g^s)^t(\mu^s - \mu^*) \leq (g^s)^t(\mu - \mu^s).$$

Da relação anterior e da equação (5-19), obtemos para qualquer $\mu \in D$,

$$(g^s)^t(\mu - \mu^*) \leq \frac{\|\mu^s - \mu\|^2 - \|\mu^{s+1} - \mu\|^2}{2\alpha} + \frac{\alpha L^2}{2} \quad \text{para todo } s \geq 0.$$

Somando sobre $s = 0, \dots, k-1$ para $k \geq 1$, obtemos para qualquer $\mu \in D$ e $k \geq 1$,

$$\sum_{s=0}^{k-1} (g^s)^t(\mu - \mu^*) \leq \frac{\|\mu^0 - \mu\|^2 - \|\mu^k - \mu\|^2}{2\alpha} + \frac{\alpha k L^2}{2} \leq \frac{\|\mu^0 - \mu\|^2}{2\alpha} + \frac{\alpha k L^2}{2}.$$

Portanto, para qualquer $k \geq 1$,

$$\max_{\mu \in D} \left\{ \sum_{s=0}^{k-1} (g^s)^t(\mu - \mu^*) \right\} \leq \frac{1}{2\alpha} \max_{\mu \in D} \|\mu^0 - \mu\|^2 + \frac{\alpha k L^2}{2}. \quad (5-20)$$

Vamos agora proporcionar uma estimativa inferior sobre o lado esquerdo da relação anterior. Seja $k \geq 1$ arbitrario e, por simplicidade,

$$\eta = \sum_{s=0}^{k-1} g^s. \quad (5-21)$$

A j -ésima entrada do vetor η^+ é dada por $(\eta_j)^+ = \max\{\eta_j, 0\}$. Se $\eta^+ = 0$, então a cota no item(a) desta proposição vale trivialmente. Assim, suponha que $\eta^+ \neq 0$ e defina um vetor $\bar{\mu}$ da seguinte forma

$$\bar{\mu} = \mu^* + r \frac{\eta^+}{\|\eta^+\|}.$$

Note que $\bar{\mu} \geq 0$ pois $\mu^* \geq 0$, $\eta^+ \geq 0$ e $r > 0$. Pelo Lema 5.1, o conjunto solução ótimo dual é limitado e, $\|\mu^*\| \leq \frac{f(\bar{x}) - q^*}{\gamma}$. Além disso, como $\tilde{q} \leq q^*$, segue-se que $\|\mu^*\| \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma}$ para qualquer solução dual μ^* . Portanto, pela definição do vetor $\bar{\mu}$, temos que

$$\|\bar{\mu}\| \leq \|\mu^*\| + r \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma} + r, \quad (5-22)$$

o qual implica que $\bar{\mu} \in D$. Usando a definição do vetor η na equação (5-21) e a relação (5-20), obtemos

$$\eta^t(\bar{\mu} - \mu^*) = \sum_{s=0}^{k-1} (g^s)^t(\bar{\mu} - \mu^*) \leq \max_{\mu \in D} \left\{ \sum_{s=0}^{k-1} (g^s)^t(\mu - \mu^*) \right\} \leq \frac{1}{2\alpha} \max_{\mu \in D} \|\mu^0 - \mu\|^2 + \frac{\alpha k L^2}{2}.$$

Como $\bar{\mu} - \mu^* = r \frac{\eta^+}{\|\eta^+\|}$, temos que $\eta^t(\bar{\mu} - \mu^*) = r \|\eta^+\|$. Assim, pela definição de η na equação (5-21) e o fato que $g^s = g(x^s)$, temos que

$$\eta^t(\bar{\mu} - \mu^*) = r \left\| \left[\sum_{s=0}^{k-1} g(x^s) \right]^+ \right\|.$$

Combinando as duas relações anteriores, segue-se que

$$\left\| \left[\sum_{s=0}^{k-1} g(x^s) \right]^+ \right\| \leq \frac{1}{2\alpha r} \max_{\mu \in D} \|\mu^0 - \mu\|^2 + \frac{\alpha k L^2}{2r}.$$

Dividindo ambos os lados desta relação por k , e usando a convexidade das funções g_j em $g = (g_1, \dots, g_m)$ e a definição de \bar{x}^k , obtemos

$$\|g(\bar{x}^k)^+\| \leq \frac{1}{k} \left\| \left[\sum_{s=0}^{k-1} g(x^s) \right]^+ \right\| \leq \frac{1}{2k\alpha r} \max_{\mu \in D} \|\mu^0 - \mu\|^2 + \frac{\alpha L^2}{2r}. \quad (5-23)$$

Como $\mu^0 \in D$, temos que

$$\max_{\mu \in D} \|\mu^0 - \mu\|^2 \leq \max_{\mu \in D} (\|\mu^0\| + \|\mu\|)^2 \leq 4 \max_{\mu \in D} \|\mu\|^2.$$

Usando a definição do conjunto D da equação (5-18), temos que

$$\max_{\mu \in D} \|\mu\| \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma} + r.$$

Substituindo as duas estimativas anteriores na relação (5-23), obtemos

$$\|g(\bar{x}^k)^+\| \leq \frac{2}{k\alpha r} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r \right)^2 + \frac{\alpha L^2}{2r}.$$

(b) A prova decorre a partir de um argumento idêntico ao utilizado na prova da parte (b) da Proposição 5.4.

(c) A prova decorre a partir de um argumento idêntico ao utilizado na prova da parte (c) da Proposição 5.4 e a cota sobre o conjunto solução ótimo dual que se segue, tendo em conta a qualificação de Slater (Hipótese 1 e Lema 5.1 com $\bar{\mu} = \mu^*$). \square

Notamos aqui que o método subgradiente da equação (5-16) com o conjunto D dado na equação (5-18) acopla o cálculo dos multiplicadores.

A seguir, consideramos selecionar o parâmetro r , que é utilizado na definição do conjunto D , tal que o lado direito da cota no item (a) da Proposição 5.6 é minimizada em cada iteração k . Dado algum $k \geq 1$, nós escolhemos r como a solução ótima do problema

$$\min_{r>0} \left\{ \frac{2}{k\alpha r} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r \right)^2 + \frac{\alpha L^2}{2r} \right\}.$$

Considere,

$$h(r) = \frac{2}{k\alpha r} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r \right)^2 + \frac{\alpha L^2}{2r} \quad (5-24)$$

derivando com respeito a r , obtemos

$$h'(r) = \frac{-2}{k\alpha r^2} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r \right)^2 + \frac{4}{k\alpha r} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r \right) - \frac{\alpha L^2}{2r^2}. \quad (5-25)$$

Considerando $y = \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r \right)$ e igualando $h'(r) = 0$ obtemos

$$\frac{-2}{k\alpha r^2} y^2 + \frac{4}{k\alpha r} y - \frac{\alpha L^2}{2r^2} = 0.$$

de onde temos

$$y^2 - 2ry + \frac{k\alpha^2 L^2}{4} = 0$$

assim, as raízes da equação anterior são

$$y = r \pm \frac{\sqrt{4r^2 - k\alpha^2 L^2}}{2}$$

substituindo y por $\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r\right)$ temos

$$\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r = r \pm \frac{\sqrt{4r^2 - k\alpha^2 L^2}}{2}$$

elevando ao quadrado ambos os membros, obtemos

$$4\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma}\right)^2 = 4r^2 - k\alpha^2 L^2$$

portanto,

$$r^*(k) = r(k) = \sqrt{\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma}\right)^2 + \frac{\alpha^2 L^2 k}{4}}. \quad (5-26)$$

vejamos que r^* é um minimizador de h

$$h''(r^*) = \frac{4}{k\alpha r^3} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r\right)^2 - \frac{8}{k\alpha r^2} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r\right) + \frac{4}{k\alpha r} + \frac{\alpha L^2}{r^3}.$$

Suponha que $h''(r^*) \leq 0$, logo temos

$$\frac{4}{k\alpha r^{*3}} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r^*\right)^2 + \frac{4}{k\alpha r^*} + \frac{\alpha L^2}{r^{*3}} \leq \frac{8}{k\alpha r^{*2}} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r^*\right).$$

de onde,

$$\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r^*\right)^2 + r^{*2} + \frac{k\alpha^2 L^2}{4} \leq 2r^* \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r^*\right).$$

logo,

$$\begin{aligned} 2r^* \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma}\right) + 2r^{*2} &\geq \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r^*\right)^2 + r^{*2} + \frac{k\alpha^2 L^2}{4} \\ &\geq \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma}\right)^2 + 2r^* \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma}\right) + 2r^{*2} + \frac{k\alpha^2 L^2}{4} \\ (\text{pela igualdade (5-26)}) &= 3r^{*3} + 2r^* \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma}\right). \end{aligned}$$

de onde obtemos

$$r^{*2} \leq 0$$

o qual é um absurdo. Portanto a solução ótima do problema anterior, denotado por $r^*(k)$,

é dada por

$$r^*(k) = \sqrt{\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma}\right)^2 + \frac{\alpha^2 L^2 k}{4}} \quad \text{para } k \geq 1. \quad (5-27)$$

Considere agora um algoritmo onde as iterações duais são obtidas por

$$\mu^{k+1} = P_{D_k}(\mu^k + \alpha g^k) \quad \text{para cada } k \geq 0,$$

com $\mu^0 \in D_0$ e o conjunto D_k dado por

$$D_k = \left\{ \mu \geq 0 / \|\mu\| \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma} + r^*(k) \right\},$$

onde $r^*(k)$ é dado pela equação (5-27). Daqui, em cada iteração k , o algoritmo projeta sobre o conjunto D_k , que contém o conjunto de soluções ótimas duais U^* . Substituindo $r^*(k)$ na cota da parte (a) da proposição 5.6, temos que

$$\begin{aligned} \|g(\bar{x}^k)^+\| &\leq \frac{2}{k\alpha r^*} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r^* \right)^2 + \frac{\alpha L^2}{2r^*} \\ &= \frac{2}{k\alpha r^*} \left(\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right)^2 + 2r^* \frac{f(\bar{x}) - \tilde{q}}{\gamma} + (r^*)^2 \right) + \frac{\alpha L^2}{2r^*} \\ &= \frac{2}{k\alpha r^*} \left(\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right)^2 + 2r^* \frac{f(\bar{x}) - \tilde{q}}{\gamma} + (r^*)^2 + \frac{\alpha^2 L^2 k}{4} \right). \end{aligned}$$

Da relação 5-27 temos que

$$(r^*(k))^2 = \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right)^2 + \frac{\alpha^2 L^2 k}{4} \quad \text{para } k \geq 0.$$

Logo,

$$\begin{aligned}
\|g(\bar{x}^k)^+\| &\leq \frac{2}{k\alpha r^*} \left(2(r^*)^2 + 2r^* \frac{f(\bar{x}) - \tilde{q}}{\gamma} \right) \\
&= \frac{4}{k\alpha} \left(r^* + \frac{f(\bar{x}) - \tilde{q}}{\gamma} \right) \\
&= \frac{4}{k\alpha} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + \sqrt{\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right)^2 + \frac{\alpha^2 L^2 k}{4}} \right) \\
&\leq \frac{4}{k\alpha} \left(\frac{2(f(\bar{x}) - \tilde{q})}{\gamma} + \frac{\alpha L \sqrt{k}}{2} \right) \\
&= \frac{8}{k\alpha} \left(\frac{f(\bar{x}) - \tilde{q}}{k\alpha} \right) + \frac{2L}{\sqrt{k}}.
\end{aligned}$$

Portanto,

$$\|g(\bar{x}^k)^+\| \leq \frac{8}{k\alpha} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right) + \frac{2L}{\sqrt{k}}. \quad (5-28)$$

Este resultado mostra que em um k dado, o erro da estimativa proporcionada na equação (5-28) pode ser alcançada se nós usamos um método de subgradiente modificado, onde cada iteração dual é projetado sobre o conjunto

$$D_k = \left\{ \mu \geq 0 / \|\mu\|_2 \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma} + \sqrt{\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right)^2 + \frac{\alpha^2 L^2 k}{4}} \right\}. \quad (5-29)$$

Vamos agora comparar a cota da violação da viabilidade da parte (a) da Proposição 5.5 para o método do subgradiente ordinario com o resultado da desigualdade (5-28) para o método do subgradiente modificado. Em particular, pela parte (a) da Proposição 5.5 e usando $q^* \geq q(\tilde{\mu})$ para $\tilde{\mu} \geq 0$ na definição da cota B^* na equação (5-14), obtemos que para todo $k \geq 1$

$$\|g(\bar{x}^k)^+\| \leq \frac{2}{k\alpha\gamma} (f(\bar{x}) - \tilde{q}) + \frac{1}{k\alpha} \max \left\{ \|\mu^0\|, \frac{1}{\gamma} (f(\bar{x}) - \tilde{q}) + \frac{\alpha L^2}{2\gamma} + \alpha L \right\}. \quad (5-30)$$

Esta cota, quando é comparada com a da desigualdade (5-28), é melhor quando k é muito grande. Porém, inicialmente, a violação da viabilidade na equação (5-30) para o método subgradiente ordinario pode ser pior, porque este depende da iteração inicial μ^0 . Suponha que a iteração inicial é $\mu^0 = 0$. Então, a cota na equação (5-30) se reduz a

$$\|g(\bar{x}^k)^+\| \leq \frac{3}{k\alpha\gamma} (f(\bar{x}) - \tilde{q}) + \frac{1}{k} \left(\frac{L^2}{2\gamma} + L \right).$$

Mesmo neste caso, inicialmente, esta cota pode ser pior do que da desigualdade (5-28), porque esta cota depende inversamente de γ que pode ser muito pequena.

Conclusões

No capítulo 2 apresentamos uma breve revisão da teoria de convexidade e dualidade Lagrangiana para programação convexa. No capítulo 3 estudamos o método subgradiente condicional para um problema convexo restrito sem estrutura, e vimos que este é uma generalização do método subgradiente tradicional, juntamente com resultados de convergência usando passos de série divergente e passo de Polyak.

Apresentamos depois resultados de convergência para o método subgradiente condicional aplicado ao problema dual de um problema convexo com estrutura, vimos que, a sequência gerada pelas soluções do subproblema dual, em geral não converge a uma solução ótima primal. Estudamos um esquema de sequência ergódica de soluções do subproblema que induzem a convergência para o conjunto solução do problema primal.

Finalmente, estudamos o método subgradiente tradicional aplicado ao problema dual, usando o comprimento de passo constante em virtude da sua simplicidade e significado prático. Apresentamos estimativas da sequência ergódica com o conjunto de soluções primais, proporcionando cotas para os valores funcionais primais e para a violação de viabilidade. Os principais resultados desta dissertação foram obtidos dos artigos [18] e [20].

Para trabalhos futuros, pretendemos investigar os principais resultados desta dissertação em dimensão infinita, fazendo modificações necessárias para obter convergência forte. Também pretendemos analisar esses resultados por meio de experimentos numéricos, fazendo comparações com outros métodos encontrados na literatura.

Referências Bibliográficas

- [1] ANSTREICHER, K; WOLSEY, L. *Two well-known properties of subgradient optimization*. European Journal of Operational Research, 120:213–220, 2009.
- [2] BAZARAA, M; SHERALI, H; SHETTY, C. *Nonlinear Programming, Theory and Algorithms*. John Wiley and Sons, New Jersey., 2006.
- [3] BERTSEKAS, D. P. *Nonlinear Programming*. Athena Scientific, Belmont, Mass., 1999.
- [4] BERTSEKAS, D; NEDIC, A; OZDAGLAR, A. *Convex Analysis and Optimization*. Athena Scientific, Belmont, Mass., 2003.
- [5] BONNANS, J; GILBERT, J; LEMARÉCHAL, C; SAGASTIZÁBAL, C. *Numerical Optimization*. Springer, Berlin, 2003.
- [6] DEM'YANOV, V; SHOMESOVA, V. *Conditional subdifferentials of convex functions*. Soviet Mathematics Doklady, 19:1181–1185, 1978.
- [7] DEM'YANOV, V. F; SHOMESOVA, V. K. *Subdifferentials of functions on sets*. Cybernetics, 16/1:24–31, 1980.
- [8] DEM'YANOV, V; VASIL'EV, L. *Nondifferentiable Optimization*. Optimization Software, New York, NY, 1985.
- [9] ERMOL'EV, Y. M. *Methods for solving nonlinear extremal problems*. Kibernetika, 2/4:1–17, 1966.
- [10] GOFFIN, J. *Nondifferentiable optimization and the relaxation method*. Nonsmooth Optimization, Oxford:31–49, 1978.
- [11] HIRIART-URRUTY, J; LEMARÉCHAL, C. *Fundamentals of Convex Analysis*. Springer, 2001.
- [12] IZMAILOV, A; SOLODOV, M. *Otimização - Volume 1, Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. Impa, Rio de Janeiro, 2005.

- [13] IZMAILOV, A; SOLODOV, M. *Otimização - Volume 2, Métodos Computacionais*. Impa, Rio de Janeiro, 2005.
- [14] KNOPP, K. *Infinite Sequences and Series*. Dover Publications, New York, NY, 1956.
- [15] LARSSON, T; LIU, Z. *Primal convergence result for dual subgradient optimization with application in multicommodity network flows*. Technical report, Dept. of Math., Linköping Institute of Technology, Linköping, Sweden, 1989.
- [16] LARSSON, T; PATRIKSSON, M; STRÖMBERG, A. *Conditional subgradient optimization - Theory and applications*. European Journal of Operational Research, 88:382–403, 1996.
- [17] LARSSON, T; PATRIKSSON, M; STRÖMBERG, A. *Ergodic convergence in subgradient optimization*. Optim. Methods Software, 9:93–120, 1998.
- [18] LARSSON, T; PATRIKSSON, M; STRÖMBERG, A. *Ergodic, primal convergence in dual subgradient schemes for convex programming*. Mathematical Programming, 86:283–312, 1999.
- [19] LIMA, E. L. *Curso de Análise Vol. 2*. Impa, Rio de Janeiro, RJ, 2009.
- [20] NEDIC, A; OZDAGLAR, A. *Approximate Primal Solutions and Rate Analysis for Dual Subgradient Methods*. SIAM Journal on Optimization, 19:1757–1780, 2008.
- [21] POLYAK, B. *Minimization of unsmooth functionals*. Computational Mathematics and Mathematical Physics, 9:14–29, 1969.
- [22] ROCKAFELLAR, T. *Convex Analysis*. Princeton University Press, New Jersey, NY, 1970.
- [23] SHOR, N. Z. *Minimization methods for nondifferentiable functions*. Springer, Berlin, 1985.