

**Algoritmo do volume e
otimização não diferenciável**

Ellen Hidemi Fukuda

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO DE MESTRE
EM
CIÊNCIAS

Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Paulo José da Silva e Silva

Durante a elaboração deste trabalho a autora recebeu auxílio financeiro da CAPES

São Paulo, março de 2007

Algoritmo do volume e otimização não diferenciável

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Ellen Hidemi Fukuda e aprovada pela comissão julgadora.

São Paulo, março de 2007.

Banca examinadora:

Prof. Dr. Paulo José da Silva e Silva

IME-USP

Prof. Dr. Carlos Humes Júnior

IME-USP

Profa. Dra. Claudia Alejandra Sagastizábal

IMPA

Agradecimentos

Ao Paulo Silva, pela sua excelente orientação e dedicação a este trabalho.

Ao Carlos Humes, Marcelo Queiroz, Claudia Sagastizábal e Roberto Andreani, pelas dicas e correções na dissertação e/ou na qualificação.

Ao grupo de Otimização Contínua do IME, pelo apoio que me deram.

Ao Marcelo Hashimoto, pelo companheirismo e afeto ao longo desses anos.

Ao Rafael, Ju, Leo, Otávio e Salem, pelas pizzadas, cinemas e amizade.

Aos amigos da “salinha vip” (Cris, Marcel, Domingos Jr., Cardonha, Chicão, Ricardo, Antonio Carlos, Goca, Marina, Fabrício, Domingos S., Felipe, Lobato, Márcio, Facci, Wanderley, Fernando T., Daniel e Maya), pelas risadas, incentivos e muito café ao longo de cada dia de trabalho.

Aos imescosTM (Wendel, Nelson, Andrei, David, Ulisses, Wagner, Sílvia, Luciana, Roberta e Sandro), pelas piadas e conversas que só nós entendemos.

Ao Mauricio, pelas conversas e grande incentivo que me deu na carreira.

À Priscilla, por eliminar o desgaste no trânsito durante o trajeto casa-USP.

E aos meus pais, pelo apoio constante que me deram em relação à carreira.

*Une œuvre d'homme n'est rien d'autre
que ce long cheminement pour retrouver,
par les détours de l'art, les deux ou trois
images simples et grandes sur lesquelles
le cœur, une première fois, s'est ouvert.*

Albert Camus ("L'envers et l'endroit")

Um trabalho de um homem não é nada além de seu longo caminho para redescobrir, através dos desvios da arte, as duas ou três imagens simples e grandes em cuja presença seu coração, pela primeira vez, se abre.

Albert Camus ("O avesso e o direito")

Resumo

Uma maneira de resolver problemas de programação linear de grande escala é explorar a relaxação lagrangeana das restrições “difíceis” e utilizar métodos de subgradientes. Populares por fornecerem rapidamente boas aproximações de soluções duais, eles não produzem diretamente as soluções primais. Para obtê-las com custo computacional adequado, pode-se construir seqüências ergódicas ou utilizar uma técnica proposta recentemente, denominada algoritmo do volume. As propriedades teóricas de convergência não foram bem estabelecidas nesse algoritmo, mas pequenas modificações permitem a demonstração da convergência dual. Destacam-se como adaptações o algoritmo do volume revisado, um método de feixes específico, e o algoritmo do volume incorporado ao método de variação do alvo.

Este trabalho foi baseado no estudo desses algoritmos e de todos os conceitos envolvidos, em especial, análise convexa e otimização não diferenciável. Estudamos as principais diferenças teóricas desses métodos e realizamos comparações numéricas com problemas lineares e lineares inteiros, em particular, o corte máximo em grafos.

Palavras-chave: algoritmo do volume, método de subgradientes, método de feixe, convergência ergódica, relaxação lagrangeana, problema do corte máximo em grafos.

Abstract

One way to solve large-scale linear programming problems is to exploit the Lagrangian relaxation of the difficult constraints and use subgradient methods. Such methods are popular as they give good approximations of dual solutions. Unfortunately, they do not directly yield primal solutions. Two alternatives to obtain primal solutions under reasonable computational cost are the construction of ergodic sequences and the use of the recently developed volume algorithm. While the convergence of ergodic sequences is well understood, the convergence properties of the volume algorithm is not well established in the original paper. This lead to some modifications of the original method to ease the proof of dual convergence. Three alternatives are the revised volume algorithm, a special case of the bundle method, and the volume algorithm incorporated by the variable target value method. The aim of this work is to study such algorithms and all related concepts, especially convex analysis and nondifferentiable optimization. We analysed the main theoretical differences among the methods and performed numerical experiments with linear and integer problems, in particular, the maximum cut problem on graphs.

Keywords: volume algorithm, subgradient method, bundle method, ergodic convergence, Lagrangian relaxation, max-cut problems on graphs.

Sumário

Notações	1
1 Introdução	2
2 Análise Convexa	5
2.1 Subdiferenciais	11
2.2 Continuidade do Subdiferencial Aproximado	16
3 Otimização Não Diferenciável	22
3.1 Condições de Otimalidade	23
3.2 Método de Descida	26
3.3 Método de Subgradientes	28
3.4 Método de Planos de Corte	34
4 Método de Feixe	38
4.1 Agregação	45
4.2 Atualização do Parâmetro de Penalização	47
4.3 Análise de Convergência	49
5 Método de Subgradientes e Convergência Ergódica	56
6 Algoritmo do Volume	66
6.1 Algoritmo do Volume Original	66
6.2 Algoritmo do Volume Revisado	72
6.3 Algoritmo do Volume com Variação do Alvo	86
7 Experimentos Numéricos	95
7.1 Corte Máximo em Grafos	96
7.2 Comparações Numéricas	99
8 Conclusão	103

Notações

Utilizaremos letras maiúsculas (A, B, \dots) para representar *matrizes*. A dimensão de uma matriz é $m \times n$ se ela possui m linhas e n colunas. Assim, $A \in \mathbb{R}^{m \times n}$ se A possui dimensão $m \times n$ e seus elementos são reais. Um *vetor* é uma matriz com $n = 1$, cuja representação é dada por letras minúsculas (a, b, \dots). Com isso, um vetor $v \in \mathbb{R}^n$ se $v = (v_1, \dots, v_n)^T$ possui dimensão n e seus elementos v_i são reais.

Considere vetores $x, y \in \mathbb{R}^n$. O *produto interno* (euclidiano) de x e y é denotado por $\langle x, y \rangle \doteq \sum_{i=1}^n x_i y_i$. A *norma* (euclidiana) de x , por sua vez, é dada por $\|x\| \doteq \langle x, x \rangle^{1/2}$. Além disso, dizemos que x e y são *ortogonais* entre si se $\langle x, y \rangle = 0$. Nesse caso, temos ainda que $\cos \gamma = \frac{\langle x, y \rangle}{\|x\| \|y\|}$, onde γ é o ângulo formado por x e y .

Dado um *conjunto* $C \subset \mathbb{R}^n$, denotamos \bar{C} e $\text{int}(C)$ como sendo, respectivamente, o seu *fecho* e o seu *interior*. Um conjunto que utilizaremos com frequência é o *simplex*, definido como $\Delta_n \doteq \{z \in [0, 1]^n : \sum_{i=1}^n z_i = 1\}$, onde n é sua dimensão. Um outro conjunto é a *bola* aberta (resp. fechada) de centro x e raio r , denotado por $B(x, r)$ (resp. $\bar{B}(x, r)$). Considere agora um ponto $x \in \mathbb{R}^n$. A *distância* de x a um conjunto $C \subset \mathbb{R}^n$ fechado, convexo, não vazio, é dada por $\text{dist}(x, C) \doteq \min_{y \in C} \|x - y\|$, e o elemento que realiza esse mínimo é chamado de *projeção* de x em C , o qual é denotado por $P_C(x) = \text{argmin}_{y \in C} \|x - y\|$.

Por fim, uma *seqüência de vetores* $x^k \in \mathbb{R}^n$ é denotada por $\{x^k\}$. Para especificar os índices k tomados, podemos escrever $\{x^k\}_{k=k_i}^{k_f}$, com $0 \leq k_i \leq k_f$, quando os índices são seqüenciais ou $\{x^k\}_{k \in K}$, com K sendo um conjunto de índices não necessariamente consecutivos. Observe a diferença entre x^k e x_k . O primeiro denota o k -ésimo iterado (que pertence a \mathbb{R}^n) de uma seqüência, enquanto que o segundo é um escalar, que indica o k -ésimo elemento de um vetor $x \in \mathbb{R}^n$. Por outro lado, um k -ésimo iterado de uma *seqüência de escalares* pode ser escrito simplesmente como α_k para não conflitar com a k -ésima potência.

Capítulo 1

Introdução

Considere o problema de programação linear de grande escala:

$$\text{minimizar } \langle c, x \rangle \quad \text{sujeito a } \begin{cases} Ax = b \\ x \in \Psi \end{cases} \quad (1.1)$$

onde

$$\Psi \doteq \{x \in \mathbb{R}^n : Zx = w, x \geq 0\}$$

corresponde às restrições “fáceis”, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $Z \in \mathbb{R}^{z \times n}$, $w \in \mathbb{R}^z$, $\text{posto}(Z) = z$, e as restrições de $Ax = b$ são “difíceis”.

Uma maneira de resolver tal problema é explorar a *relaxação lagrangeana* das restrições “difíceis”. Considere, então, a seguinte função lagrangeana:

$$\Psi \times \mathbb{R}^m \ni (x, \pi) \mapsto L(x, \pi) \doteq \langle c, x \rangle + \langle Ax - b, \pi \rangle, \quad (1.2)$$

onde $\pi \in \mathbb{R}^m$ é o vetor associado aos multiplicadores de Lagrange, ou seja, a variável dual. A idéia dessa técnica é substituir algumas restrições — no caso, as “difíceis” — por um “preço” linear. Dessa forma, o problema (1.1) pode ser escrito da seguinte forma:

$$\inf_{x \in \Psi} \sup_{\pi \in \mathbb{R}^m} L(x, \pi).$$

Segue ainda, da relação de dualidade fraca, que:

$$\inf_{x \in \Psi} \sup_{\pi \in \mathbb{R}^m} L(x, \pi) \geq \sup_{\pi \in \mathbb{R}^m} \inf_{x \in \Psi} L(x, \pi). \quad (1.3)$$

A função objetivo dual, denotada por

$$f(\pi) \doteq \inf_{x \in \Psi} L(x, \pi), \quad (1.4)$$

é o ínfimo de funções afins, e portanto é côncava e possivelmente não diferenciável. Logo, o problema dual

$$\text{maximizar } f(\pi) \quad \text{sujeito a } \pi \in \mathbb{R}^m \quad (1.5)$$

é um problema de otimização não diferenciável irrestrita. Ademais, é possível provar, nesse caso, que a folga de dualidade é nula, ou seja, que vale a igualdade em (1.3) sempre que um dos dois problemas for viável. Tendo em vista que o problema primal (1.1) é difícil de resolver, concentraremos na resolução do problema dual (1.5).

Para resolvê-lo, podemos utilizar, por exemplo, o *método de subgradientes*. Apesar de fornecerem rapidamente aproximações de soluções duais, eles não produzem diretamente as soluções primais. Para obtê-las com custo computacional adequado, podemos construir *seqüências ergódicas* que convergem a elas. Essa idéia foi dada por Shor [Sho85] e consiste essencialmente em realizar combinações convexas específicas. Posteriormente, Larsson e Liu [LL89] especificaram os tamanhos de passo e os parâmetros de convexidade, de modo a aumentar a eficiência da técnica.

Em seguida, Serali e Choi [SC96] estenderam esses resultados permitindo escolhas gerais de tamanhos de passo e de parâmetros de convexidade e utilizando métodos de subgradientes puros, projetados e com desvios. Larsson, Patriksson e Strömberg [LPS99] apresentaram depois resultados mais gerais, para programação convexa, utilizando o método de subgradientes condicionais [LPS96].

Mais recentemente, Barahona e Anbil [BA00] criaram uma extensão do método de subgradientes que produz uma aproximação de solução primal ao mesmo tempo em que resolve o dual. Denominado de *algoritmo do volume* (VA), ele obtém esse ponto primal estimando-se o volume associado às faces ativas das restrições “difíceis”. Sua convergência não foi demonstrada nesse artigo, mas experimentos numéricos realizados com relaxações de problemas combinatórios mostraram resultados bastante encorajadores.

Esse algoritmo motivou Bahiense, Maculan e Sagastizábal [BMS02] a modificá-lo, de forma a obter resultados de convergência dual. Basicamente, foram feitas pequenas alterações no VA de modo a ser possível interpretá-lo como um método de feixe. Eles mostraram também que o VA pode ser entendido como um método extra-gradiente baseado em ε -subgradientes. Dessa forma, criaram uma nova formulação do método, denominada *algoritmo do volume revisado* (RVA). Observaram ainda que uma simples modificação do RVA origina um método de feixes específico, denominado BVA.

Finalmente, Serali e Lim [SL04], também com o propósito de estabelecer resultados de convergência do VA, modificaram-no, utilizando uma técnica

denominada *método de variação do alvo* (VTVM) [SCT00]. Em seu artigo, os autores demonstraram um resultado de convergência dual do algoritmo modificado e também mostraram, através de um exemplo simples, que o VA original pode não convergir no espaço primal. No entanto, assim como Bahiense, Maculan e Sagastizábal, eles deixaram como lacuna as demonstrações para convergência primal dos algoritmos modificados.

Esse trabalho foca no estudo de técnicas de recuperação de soluções primais, em particular, os métodos de subgradientes com convergência ergódica e as variantes do algoritmo do volume. Além disso, realizamos suas implementações e fizemos algumas comparações numéricas, na qual verificamos seu bom desempenho em relação ao método simplex. Em particular, realizamos experimentos com o problema de corte máximo em grafos utilizando a técnica *branch-and-cut-and-price* [BL01].

O texto está dividido da seguinte forma: No capítulo 2, abordaremos os principais aspectos de análise convexa, em especial as propriedades relacionadas a subdiferenciais. No capítulo 3, trataremos os principais algoritmos de otimização não diferenciável, entre eles, os métodos de descida, de subgradientes e de planos de corte. O capítulo 4 será totalmente voltado a mais um método de otimização não diferenciável: o método de feixe.

Apenas no capítulo 5 começaremos a abordar maneiras de resolver o problema (1.1), com o método de subgradientes com construção de seqüência ergódica primal. No capítulo 6 veremos finalmente o algoritmo do volume e suas variações. Terminaremos no capítulo 7, onde apresentaremos experimentos numéricos com todos os métodos estudados.

Capítulo 2

Análise Convexa

Neste capítulo abordaremos alguns elementos necessários para o estudo de métodos de otimização não diferenciável. Focaremos nas propriedades do subdiferencial, baseando-se essencialmente em [Roc70] e [HUL01], com apoio da versão estendida deste último, [HUL93a, HUL93b] e de [Ber03].

Para que o estudo de métodos de otimização seja possível, consideramos aqui o conjunto dos *reais estendidos*, ou seja, adicionamos os elementos $+\infty$ e $-\infty$ aos reais. No entanto, em problemas de minimização (resp. maximização), devemos proibir que o valor da função objetivo seja igual a $-\infty$ (resp. $+\infty$). Isso significa que, em geral, as funções de nosso interesse são *próprias*, ou seja, para pelo menos um ponto seu valor é finito e nas demais são sempre maiores (resp. menores) que $-\infty$ (resp. $+\infty$) quando trabalhamos com problemas de minimização (resp. maximização).

Por padrão, pensaremos por enquanto em problemas de minimização e consideraremos o conjunto $\mathbb{R} \cup \{+\infty\}$ como contra-domínio das funções, a menos que seja indicado o contrário. Observe, por exemplo, na definição abaixo, que com isso evitamos indeterminações do tipo $\infty + (-\infty)$.

Definição 2.1. *Uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é convexa quando, para todo $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ e todo $\alpha \in]0, 1[$, vale:*

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

Se a desigualdade acima for estrita, então f é estritamente convexa.

Por outro lado, uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{-\infty\}$ é *côncava* (resp. *estritamente côncava*) quando $-f$ é convexa (resp. estritamente convexa). Considere agora uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ qualquer. O *domínio efetivo* de f é denotado por $\text{dom } f \doteq \{x \in \mathbb{R}^n: f(x) < +\infty\}$. Sua *epígrafe*, por

sua vez, é dada por $\text{epi } f \doteq \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq r\}$, e o conjunto de nível de f em $r \in \mathbb{R}$ é dado por $S_r(f) \doteq \{x \in \mathbb{R}^n : f(x) \leq r\}$. Observa-se, através dessas definições, que $(x, r) \in \text{epi } f$ se, e somente se, $x \in S_r(f)$. A desigualdade invertida da definição de epígrafe nos dá um conjunto denominado *hipógrafe*. Note que a epígrafe (resp. hipógrafe) ser convexa equivale a dizer que a função correspondente é convexa (resp. côncava).

Definição 2.2. *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*

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

Observa-se que a definição de semi-continuidade inferior acima equivale a dizer que a epígrafe da função é fechada ou que os conjuntos de nível são fechados [Roc70, teorema 7.1]. Outras definições essenciais são as seguintes:

Definição 2.3. *Uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é denominada fechada se $\text{dom } f$ é fechado e f é semi-continua inferior em $\text{dom } f$.*

Definição 2.4. *Uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é denominada Lipschitz contínua em $S \subset \mathbb{R}^n$ se existe um escalar $L \geq 0$ tal que*

$$|f(x) - f(y)| \leq L\|x - y\| \quad \text{para todo } x, y \in (S \cap \text{dom } f).$$

Definição 2.5. *Considere um conjunto $S \subset \mathbb{R}^n$ não vazio. Uma função $\sigma_S: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ dada por*

$$\sigma_S(x) \doteq \sup \{\langle s, x \rangle : s \in S\}$$

é denominada função suporte de S .

Assim como uma função suporte está, por definição, associada a um conjunto convexo, pode-se mostrar que todo conjunto convexo fechado é completamente determinado por uma função suporte. Em outras palavras, há uma correspondência bijetora entre a classe dessas funções e a classe desses conjuntos [HUL93a, teorema V.3.1.1]. Veremos um exemplo disso posteriormente. Por enquanto, considere outros dois exemplos de funções:

Definição 2.6. *Considere um conjunto $S \subset \mathbb{R}^n$ não vazio. Uma função $i_S: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ definida por*

$$i_S(x) \doteq \begin{cases} 0, & \text{se } x \in S \\ +\infty, & \text{caso contrário} \end{cases}$$

é denominada função indicadora de S .

Definição 2.7. Sejam $(a_1, b_1), \dots, (a_m, b_m) \in \mathbb{R}^n \times \mathbb{R}$. Então uma função $\tilde{f}: \mathbb{R}^n \rightarrow \mathbb{R}$ definida por

$$\tilde{f}(x) \doteq \max_{i=1, \dots, m} \{ \langle a_i, x \rangle - b_i \}$$

é denominada função afim por partes.

Considere agora um mapeamento F que associa cada $x \in X \subset \mathbb{R}^n$ a um conjunto do \mathbb{R}^n , ou seja, $X \ni x \mapsto F(x) \subset \mathbb{R}^n$. Denominamos tal mapeamento de *função multivalorada*. Seu *domínio* é definido por $\text{dom } F \doteq \{x \in X : F(x) \neq \emptyset\}$. Além disso, dizemos que F é *fechada* se seu gráfico (isto é, união de $\{x\} \times F(x) \subset X \times \mathbb{R}^n$) é um conjunto fechado. Dizemos ainda que ela é *localmente limitada* perto de x^* se para alguma vizinhança V de x^* e algum conjunto limitado $B \subset \mathbb{R}^n$, tem-se $V \subset \text{dom } F$ e $F(V) \subset B$.

Definição 2.8. Seja F uma função multivalorada fechada e localmente limitada em x^* . Então, F é *semi-contínua exterior* (resp. *semi-contínua interior*) se, para todo $\varepsilon > 0$, existe uma vizinhança $V(x^*)$ de x^* tal que $x \in V(x^*)$ implica

$$F(x) \subset F(x^*) + B(0, \varepsilon) \quad (\text{resp. } F(x^*) \subset F(x) + B(0, \varepsilon)).$$

Ademais, se F é *semi-contínua exterior* e *interior*, então ela é *contínua*.

Abordamos a partir de agora algumas definições e resultados referentes a conjuntos. Considere $S \subset \mathbb{R}^n$ um conjunto não vazio. Denotamos por $\text{conv}(S)$ o *fecho convexo* de S , ou seja, à intersecção de todos os conjuntos convexos que contêm S . De acordo com o teorema de Carathéodory [Roc70, teorema 17.1], podemos descrevê-lo da seguinte forma:

$$\text{conv}(S) \doteq \left\{ \sum_{k=1}^{n+1} \alpha_k x^k : \alpha \in \Delta_{n+1} \text{ e } x^k \in S \text{ para todo } k = 1, \dots, n+1 \right\}.$$

O conjunto S é *afim* se pode ser escrito na forma $x + Z$, onde $x \in \mathbb{R}^n$ e Z é um subespaço determinado unicamente por S , denominado *subespaço linear paralelo* a S . Além disso, o *fecho afim* de S , denotado por $\text{aff}(S)$, é a intersecção de todos os conjuntos afins que contêm S . Note que $\text{aff}(S)$ é também um conjunto afim e que ele contém $\text{conv}(S)$.

Por sua vez, um ponto $x \in \mathbb{R}^n$ pertence ao *interior relativo* de S , denotado por $\text{ri}(S)$, se $x \in S$ e existe $r > 0$ tal que $B(x, r) \cap \text{aff}(S) \subset S$, ou seja, se x é um ponto interior de S relativo a $\text{aff}(S)$. Utilizando essa definição, temos, por exemplo, o seguinte resultado: Se $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é convexa

e própria, e $S \subset \text{ri}(\text{dom } f)$ é compacto, então f é Lipschitz contínua em S . [Roc70, teorema 10.4]. Um outro resultado é enunciado a seguir, sendo que sua demonstração pode ser encontrada em [Roc70, teorema 6.1].

Proposição 2.9. *Seja $S \subset \mathbb{R}^n$ um conjunto convexo, com $\bar{x} \in \bar{S}$ e $x \in \text{ri}(S)$. Então, $\alpha\bar{x} + (1 - \alpha)x \in \text{ri}(S)$ para todo $\alpha \in [0, 1[$.*

Um conjunto $K \subset \mathbb{R}^n$ é um *cone* se $\alpha x \in K$, para qualquer $x \in K$ e $\alpha > 0$. Note que, por essa definição, nem todos os cones contêm o ponto 0. Além disso, denotamos $\text{cone}(S)$ como o *cone gerado* por S , ou seja, o conjunto de todas as combinações não negativas de elementos de S . Seu fecho será denotado por $\overline{\text{cone}(S)}$. Referente a cones, necessitamos ainda da seguinte definição:

Definição 2.10. *Seja $K \subset \mathbb{R}^n$ um cone não vazio. O (cone) polar de K é dado por*

$$K^\circ \doteq \{y \in \mathbb{R}^n : \langle y, x \rangle \leq 0 \text{ para todo } x \in K\}.$$

Pode-se mostrar que se um cone K é não vazio, fechado e convexo, então $(K^\circ)^\circ = K$ [Roc70, teorema 14.1]. Outros dois tipos de cones de nosso interesse são o *cone tangente* e o *cone normal*. Conforme indica a definição abaixo, o primeiro pode ser visto como uma aproximação local de um conjunto por cones.

Definição 2.11. *Seja $S \subset \mathbb{R}^n$ um conjunto não vazio. Dizemos que $d \in S$ é uma direção tangente a S em $x \in S$ se, e somente, se, existem seqüências $\{x^k\} \subset S$ e $\{t_k\} \subset \mathbb{R}$ tais que*

$$x^k \rightarrow x, \quad t_k \downarrow 0, \quad \frac{x^k - x}{t_k} \rightarrow d.$$

O conjunto de tais direções é chamado cone tangente e denotado por $T_S(x)$.

Uma formulação equivalente para o cone tangente é dada abaixo.

Proposição 2.12. *Seja $S \subset \mathbb{R}^n$ um conjunto não vazio. Dizemos que $d \in S$ é uma direção tangente a S em $x \in S$ se, e somente, existem seqüências $\{d^k\} \subset \mathbb{R}^n$ e $\{t_k\} \subset \mathbb{R}$ tais que*

$$d^k \rightarrow d, \quad t_k \downarrow 0, \quad x + t_k d^k \in S \text{ para todo } k.$$

Demonstração. Basta considerar $d^k \doteq (x^k - x)/t_k$ (ou, equivalentemente, $x^k = x + t_k d^k \in S$) na definição 2.11. \square

Por tais definições, pode-se mostrar que o cone tangente é fechado [HUL93a, proposição III.5.1.3]. A proposição abaixo mostra, ainda, que quando S é convexo, o cone tangente $T_S(x)$ em $x \in S$ pode ser escrito de outra forma.

Proposição 2.13. *Seja $S \subset \mathbb{R}^n$ um conjunto convexo não vazio. Então,*

$$T_S(x) = \overline{\text{cone}}(S - \{x\}).$$

Demonstração. Pela convexidade de S , se $x \in S$ e $x + t_k d^k \in S$, então $x + \lambda d^k \in S$ para qualquer $\lambda \in]0, t_k]$. Em particular, $S - \{x\} \subset T_S(x)$. Como $T_S(x)$ é um cone fechado, podemos escrever $\overline{\text{cone}}(S - \{x\}) \subset T_S(x)$. Considere agora $d \in T_S(x)$. Tome seqüências $\{x^k\}$ e $\{t_k\}$ conforme a definição 2.11. Então, o ponto $(x^k - x)/t_k$ pertence a $\text{cone}(S - \{x\})$ e seu limite d pertence a $\overline{\text{cone}}(S - \{x\})$. \square

Definição 2.14. *Seja $S \subset \mathbb{R}^n$ um conjunto convexo fechado não vazio. Dizemos que $d \in \mathbb{R}^n$ é uma direção normal a S em $x \in S$ quando*

$$\langle d, y - x \rangle \leq 0 \text{ para todo } y \in S.$$

O conjunto de tais direções é chamado cone normal e denotado por $N_S(x)$.

Um resultado interessante é que o cone normal é o polar do cone tangente, ou seja, que dado um conjunto convexo fechado $S \subset \mathbb{R}^n$, tem-se $N_S(x) = (T_S(x))^\circ$. Como os cones tangentes e normais são fechados e convexos, podemos também afirmar que $T_S(x) = (N_S(x))^\circ$ e, desse modo, temos que um é o polar do outro [HUL93a, proposição III.5.2.4]. Essa relação pode ser observada na figura 2, onde as partes hachuradas clara e escura correspondem aos cones tangente e normal, respectivamente.

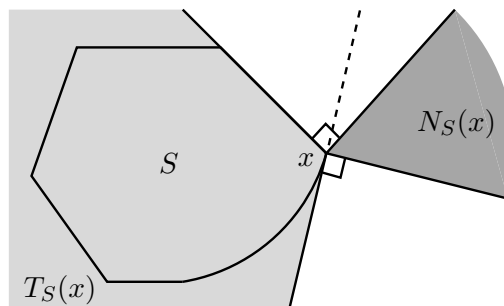


Figura 2.1: Relação entre cone tangente e cone normal.

O último elemento de análise convexa que veremos é uma generalização da definição 2.14.

Definição 2.15. Sejam $S \subset \mathbb{R}^n$ um conjunto convexo fechado não vazio, e $\varepsilon \geq 0$. Dizemos que $d \in \mathbb{R}^n$ é uma direção ε -normal a S em $x \in S$ quando

$$\langle d, y - x \rangle \leq \varepsilon \text{ para todo } y \in S.$$

Denominamos o conjunto de tais direções de cone ε -normal. Além disso, o denotamos por $N_{S,\varepsilon}(x)$.

Finalizamos a seção com o teorema de separação estrita de convexos.

Teorema 2.16. (Teorema de separação estrita de convexos) Sejam $S_1 \subset \mathbb{R}^n$ e $S_2 \subset \mathbb{R}^n$ conjuntos convexos e não vazios. Se S_1 é fechado, S_2 é compacto e $S_1 \cap S_2 = \emptyset$, então existe um hiperplano $\{x \in \mathbb{R}^n : \langle a, x \rangle = b\}$ que os separa estritamente. Em outras palavras,

$$\langle a, x^1 \rangle < b < \langle a, x^2 \rangle \text{ para todo } x^1 \in S_1 \text{ e } x^2 \in S_2. \quad (2.1)$$

Demonstração. Considere o conjunto

$$S \doteq S_2 - S_1 = \{x \in \mathbb{R}^n : x = x_2 - x_1, x_1 \in S_1, x_2 \in S_2\}.$$

Como S_2 é compacto e S_1 é fechado, temos que S é fechado. Considerando agora o vetor de norma mínima em $S_2 - S_1$, sabemos que ele é escrito como $\bar{x}^2 - \bar{x}^1$, para algum $\bar{x}^1 \in S_1$ e algum $\bar{x}^2 \in S_2$ (veja figura 2.2). Defina

$$a \doteq \frac{\bar{x}^2 - \bar{x}^1}{2}, \quad \bar{x} \doteq \frac{\bar{x}^1 + \bar{x}^2}{2},$$

e seja $b \doteq \langle a, \bar{x} \rangle$. Como $S_1 \cap S_2 = \emptyset$, temos que $a \neq 0$. Pela figura 2.2, observa-se também que $\bar{x}^1 = P_{S_1}(\bar{x})$ e que $\bar{x}^2 = P_{S_2}(\bar{x})$.

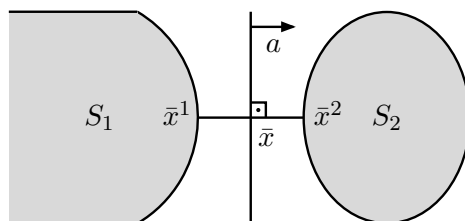


Figura 2.2: Construção de um hiperplano separador estrito.

Utilizando a definição de projeção $\bar{x}^1 = P_{S_1}(\bar{x})$, temos:

$$\langle \bar{x} - \bar{x}^1, x^1 - \bar{x}^1 \rangle \leq 0 \text{ para todo } x^1 \in S_1.$$

Equivalentemente, como $\bar{x} - \bar{x}^1 = a$, para todo $x^1 \in S_1$, temos:

$$\langle a, x^1 \rangle \leq \langle a, \bar{x}^1 \rangle = \langle a, \bar{x} \rangle + \langle a, \bar{x}^1 - \bar{x} \rangle = b - \|a\|^2 < b.$$

De maneira análoga, podemos provar que $b < \langle a, x^2 \rangle$ para todo $x^2 \in S_2$. \square

2.1 Subdiferenciais

O conjunto dos pontos não diferenciáveis de uma função convexa (ou côncava) a valores reais tem medida de Lebesgue nula [HUL93a, teorema IV.4.2.3]. Sabe-se, no entanto, que estes pontos tendem a ser, muitas vezes, as soluções de problemas de minimização (resp. maximização) de funções convexas (resp. côncavas). Essa observação pode ser ilustrada, por exemplo, com a função $\mathbb{R} \ni x \mapsto f(x) \doteq |x| + \alpha x$, com $\alpha \in [-1, 1]$. Note que para qualquer α desse intervalo, o ponto $x = 0$ é o minimizador do problema e é o (único) ponto não diferenciável da função.

Considerando que um dos objetivos deste trabalho é resolver o problema (1.5), concentraremos agora em descrever o comportamento de uma função em torno dos pontos de não diferenciabilidade. Para isso, devemos generalizar o conceito de gradiente.

Definição 2.17. *O subdiferencial $\partial f(x)$ de uma função convexa $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ no ponto $x \in \text{dom } f$ é dado por*

$$\partial f(x) \doteq \{v \in \mathbb{R}^n: f(y) \geq f(x) + \langle v, y - x \rangle \text{ para todo } y \in \mathbb{R}^n\}. \quad (2.2)$$

Um vetor $v \in \partial f(x)$ é denominado subgradiente de f em x .

Proposição 2.18. *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa e própria. Para todo $x \in \text{dom } f$, o subdiferencial $\partial f(x)$ pode ser escrito como*

$$\partial f(x) = S^\perp + \{v \in \partial f(x): v \in S\},$$

onde S é o subespaço paralelo a $\text{aff}(\text{dom } f)$.

Demonstração. Segue diretamente do fato de que dados $x \in \mathbb{R}^n$ e $S \subset \mathbb{R}^n$, podemos escrever $x = x^1 + x^2$, onde $x^1 \in S$ e $x^2 \in S^\perp$. \square

Utilizamos a notação $\partial f(\cdot)_S$ para indicar a parte de $\partial f(\cdot)$ paralela a $\text{aff}(\text{dom } f)$. É interessante notar também que um subdiferencial $\partial f(x)$ é um conjunto que reúne propriedades desejáveis, como ser convexo e fechado. Se $x \in \text{ri}(\text{dom } f)$, então $\partial f(x)$ é também não vazio e $\partial f(x)_S$ é limitado [Roc70, teorema 23.4]. Verifiquemos agora uma outra definição do subdiferencial, dada através do conceito de *derivada direcional*.

Definição 2.19. *A derivada direcional de uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ em um ponto $x \in \text{dom } f$ na direção $d \in \mathbb{R}^n$ é dada por*

$$f'(x; d) \doteq \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}.$$

Além disso, se f for convexa, podemos escrever

$$f'(x; d) = \inf_{t>0} \frac{f(x + td) - f(x)}{t}.$$

uma vez que esses coeficientes são decrescentes em t .

Proposição 2.20. *Sejam uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ convexa e um ponto $x \in \text{dom } f$. Então,*

$$\partial f(x) \doteq \{v \in \mathbb{R}^n : \langle v, d \rangle \leq f'(x; d) \text{ para todo } d \in \mathbb{R}^n\}. \quad (2.3)$$

Demonstração. Queremos mostrar que (2.2) e (2.3) são equivalentes.

$$\begin{aligned} & \langle v, d \rangle \leq f'(x; d) && \forall d \in \mathbb{R}^n \\ \Leftrightarrow & \langle v, d \rangle \leq \inf_{t>0} \frac{f(x + td) - f(x)}{t} && \forall d \in \mathbb{R}^n \\ \Leftrightarrow & \langle v, d \rangle \leq \frac{f(x + td) - f(x)}{t} && \forall d \in \mathbb{R}^n \text{ e } \forall t > 0 \end{aligned}$$

Tomando $y \doteq x + td$, temos $f(y) \geq f(x) + \langle v, y - x \rangle$ para todo $y \in \mathbb{R}^n$. \square

Corolário 2.21. *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa. Então, o subdiferencial $\partial f(x)$ de f em um ponto $x \in \text{ri}(\text{dom } f)$ é o conjunto convexo fechado não vazio de \mathbb{R}^n cuja função suporte é $f'(x; \cdot)$. Em outras palavras,*

$$f'(x; d) = \sup \{ \langle v, d \rangle : v \in \partial f(x) \}.$$

Referimo-nos a [HUL93a, teorema V.3.1.1] para verificar a validade do corolário acima. Observe que, conforme mencionado anteriormente, há uma correspondência bijetora entre a função suporte $f'(x; \cdot)$ e o conjunto $\partial f(x)$. O próximo corolário mostra que, de fato, o subgradiente generaliza o conceito de gradiente. Lembremos antes que uma função f é diferenciável em x se, e somente se, $\nabla f(x)$ existe e satisfaz $\langle \nabla f(x), y \rangle = f'(x; y)$ para todo y .

Corolário 2.22. *Uma função convexa $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é diferenciável em $x \in \mathbb{R}^n$ com gradiente $\nabla f(x)$ se, e somente se, $\partial f(x) = \{\nabla f(x)\}$.*

Demonstração. Sabemos que f é diferenciável em x se, e somente se, a derivada direcional $f'(x; y)$ é linear da forma $f'(x; y) = \langle \nabla f(x), y \rangle$. Logo, pelo corolário 2.21, temos que $\nabla f(x)$ é o único subgradiente de x . \square

Note agora que a definição 2.17 oferece-nos uma interpretação geométrica coerente com o cálculo diferencial. De fato, dada $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ convexa, a aproximação linear dada por um subgradiente $v \in \mathbb{R}^n$ em um ponto $x \in \mathbb{R}^n$ minora f no ponto $(x, f(x))$. Utilizando os conceitos de cone normal e cone tangente, podemos expressar esse resultado da seguinte maneira:

Proposição 2.23. *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa. Um vetor $v \in \mathbb{R}^n$ é um subgradiente de f em um ponto $x \in \text{dom } f$ se, e somente se, $(v, -1) \in \mathbb{R}^n \times \mathbb{R}$ é normal à epígrafe de f no ponto $(x, f(x))$, isto é, se*

$$N_{\text{epi } f}(x, f(x)) = \{(\lambda v, -\lambda) : v \in \partial f(x), \lambda \geq 0\}. \quad (2.4)$$

Ademais, o cone tangente à epígrafe de f em $(x, f(x))$ é a epígrafe da função derivada direcional $f'(x; \cdot)$. Em outras palavras,

$$T_{\text{epi } f}(x, f(x)) = \{(d, r) : r \geq f'(x; d)\}. \quad (2.5)$$

Demonstração. Provemos primeiro a igualdade (2.4). Por definição de subgradiente, temos que $f(y) \geq f(x) + \langle v, y - x \rangle$ para todo $y \in \mathbb{R}^n$. Sabendo que $r \geq f(y)$ para todo $(y, r) \in \text{epi } f$, temos:

$$\begin{aligned} f(x) + \langle v, y - x \rangle &\leq r \quad \forall y \in \mathbb{R}^n \text{ e } \forall r \geq f(y) \\ \Leftrightarrow \langle v, y - x \rangle + \langle -1, r - f(x) \rangle &\leq 0 \quad \forall (y, r) \in \text{epi } f \\ \Leftrightarrow \langle \lambda v, y - x \rangle + \langle -\lambda, r - f(x) \rangle &\leq 0 \quad \forall (y, r) \in \text{epi } f \text{ e } \forall \lambda \geq 0 \\ \Leftrightarrow \langle (\lambda v, -\lambda), (y, r) - (x, f(x)) \rangle &\leq 0 \quad \forall (y, r) \in \text{epi } f \text{ e } \forall \lambda \geq 0 \\ \Leftrightarrow (\lambda v, -\lambda) \in N_{\text{epi } f}(x, f(x)). \end{aligned}$$

Demonstremos agora a igualdade (2.5). Sabemos que o cone tangente é o polar do cone normal, isto é, que $T_{\text{epi } f}(x, f(x)) = (N_{\text{epi } f}(x, f(x)))^\circ$. Dessa forma, $(d, r) \in T_{\text{epi } f}(x, f(x))$ se, e somente se,

$$\begin{aligned} \langle (\lambda v, -\lambda), (d, r) \rangle &\leq 0 \quad \forall v \in \partial f(x) \text{ e } \forall \lambda \geq 0 \\ \Leftrightarrow \lambda(\langle v, d \rangle - r) &\leq 0 \quad \forall v \in \partial f(x) \text{ e } \forall \lambda \geq 0. \end{aligned}$$

Excluindo o caso trivial $\lambda = 0$, podemos dividir a desigualdade acima por λ e assim obtemos $r \geq \langle v, d \rangle$ para todo $v \in \partial f(x)$. Em particular,

$$r \geq \sup\{\langle v, d \rangle : v \in \partial f(x)\} = f'(x; d),$$

onde a última igualdade é verdadeira conforme o corolário 2.21. \square

Para ilustrar o resultado acima, observe o exemplo da figura 2.1.1. O subdiferencial $\partial f(x)$ é representado pelo conjunto de vetores $(v, -1)$ hachurados,

ou seja, por $N_{\text{epi } f}(x, f(x))$, enquanto que as linhas tracejadas mostram duas aproximações lineares, correspondentes aos hiperplanos cujas normais são os vetores extremos de $N_{\text{epi } f}(x, f(x))$. Note também que a epígrafe de f está contida em $T_{\text{epi } f}(x, f(x))$ e que este é determinado pelas linhas tracejadas.

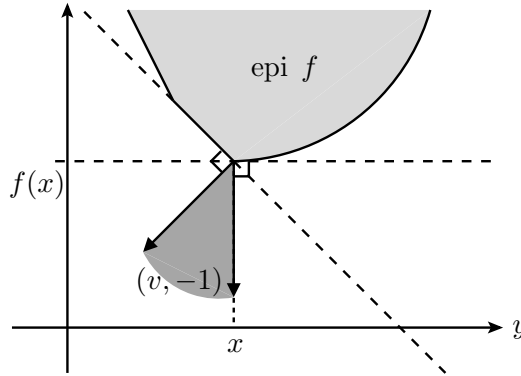


Figura 2.1.1: Interpretação geométrica do subgradiente.

Assim como no cálculo diferencial, existem regras de cálculo com subdiferenciais. A proposição abaixo mostra um exemplo de um subdiferencial de uma função f derivada de subdiferenciais de funções f_k que definem f . Outras regras de cálculo podem ser vistas, por exemplo, em [HUL93a, seção VI.4] e em [Roc70, seção V.23]. Referimo-nos também a [HUL93a, teorema VI.4.4.2] para a demonstração da proposição abaixo.

Proposição 2.24. *Considere $\{f_k\}_{k \in K}$ uma coleção de funções convexas e semi-contínuas superiores, com $f_k: \mathbb{R}^n \rightarrow \mathbb{R}$ para cada k e K compacto. Sejam $f(x) \doteq \sup\{f_k(x) : k \in K\} < +\infty$ e $K(x) \doteq \{k \in K : f_k(x) = f(x)\}$ para todo $x \in \mathbb{R}^n$. Então, $\partial f(x) = \text{conv}\{\cup \partial f_k(x) : k \in K(x)\}$. Se cada f_k é diferenciável, podemos escrever $\partial f(x) = \text{conv}\{\nabla f_k(x) : k \in K(x)\}$.*

Uma utilização dessa proposição pode ser vista abaixo.

Proposição 2.25. *Considere a função \tilde{f} da definição 2.7 e seja $x^* \in \mathbb{R}^n$ qualquer. Então, existe $r > 0$ tal que $\partial f(x) \subseteq \partial f(x^*)$ para todo $x \in B(x^*, r)$.*

Demonstração. Seja $K \doteq \{1, \dots, m\}$. Pela notação utilizada na proposição 2.24, podemos dizer que $K(x^*)$ é o conjunto das funções ativas em x^* . Assim, dado x^* , tome um índice inativo $k \in K \setminus K(x^*)$. Então, por continuidade, existem $\varepsilon_k > 0$ e $\delta_k > 0$ tais que $f_k(x) < \tilde{f}(x^*) - \delta_k$ para todo $x \in B(x^*, \varepsilon_k)$. Definindo $\varepsilon \doteq \min_{k \in K \setminus K(x^*)} \varepsilon_k$ e $\delta \doteq \min_{k \in K \setminus K(x^*)} \delta_k$, temos

$$f_k(x) < \tilde{f}(x^*) - \delta \text{ para todo } x \in B(x^*, \varepsilon) \text{ e todo } k \in K \setminus K(x^*). \quad (2.6)$$

Além disso, por continuidade de \tilde{f} , se $x \rightarrow x^*$ então $\tilde{f}(x) \rightarrow \tilde{f}(x^*)$. Logo, por (2.6), se $k \in K \setminus K(x^*)$, então $k \notin K(x^*)$ para x próximo de x^* . Com isso, existe $r > 0$ suficientemente pequeno tal que $K(x) \subseteq K(x^*)$. Finalmente, pela proposição 2.24 e sabendo que $\nabla f_k(x) = a_k$ para todo k ,

$$\partial f(x) = \text{conv}\{a_k : k \in K(x)\} \subseteq \text{conv}\{a_k : k \in K(x^*)\} = \partial f(x^*),$$

conforme queríamos. \square

Mostremos agora que o conceito de subdiferencial pode, por sua vez, ser generalizado através da introdução de um parâmetro (ou tolerância) $\varepsilon \geq 0$.

Definição 2.26. O ε -subdiferencial $\partial_\varepsilon f(x)$ de uma função convexa $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ em um ponto $x \in \text{dom } f$ é dado por

$$\partial_\varepsilon f(x) \doteq \{v \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x \rangle - \varepsilon \text{ para todo } y \in \mathbb{R}^n\}. \quad (2.7)$$

Um vetor $v \in \partial_\varepsilon f(x)$ é denominado ε -subgradiente de f em x .

Da mesma maneira, podemos afirmar que um vetor $v \in \mathbb{R}^n$ é um ε -subgradiente de f em x se, e somente se, $(v, -1) \in \mathbb{R}^n \times \mathbb{R}$ é ε -normal à epígrafe de f no ponto $(x, f(x))$. A figura 2.1.2 ilustra essa interpretação com o mesmo exemplo anterior. Visualmente podemos notar que $N_{\text{epi } f}(x, f(x)) \subset N_{\text{epi } f, \varepsilon}(x, f(x))$, ou ainda que $\partial f(x) \subset \partial_\varepsilon f(x)$. De fato, é fácil mostrar que $\partial_{\varepsilon'} f(x) \subset \partial_\varepsilon f(x)$ sempre que $0 \leq \varepsilon' \leq \varepsilon$.

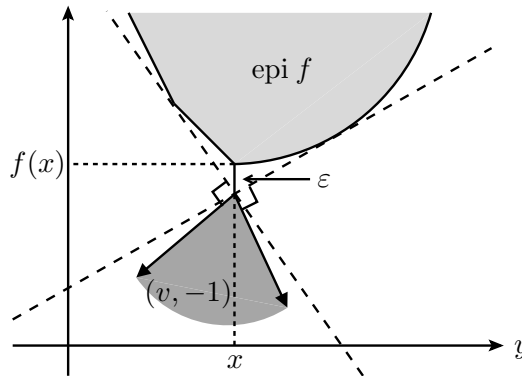


Figura 2.1.2: Interpretação geométrica do ε -subgradiente.

Da mesma forma que o subdiferencial, temos uma definição alternativa para o ε -subdiferencial, dada através da *derivada ε -direcional*.

Definição 2.27. A derivada ε -direcional de uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ em um ponto $x \in \mathbb{R}^n$ na direção $d \in \mathbb{R}^n$ é dada por

$$f'_\varepsilon(x; d) \doteq \inf_{t>0} \frac{f(x + td) - f(x) + \varepsilon}{t}.$$

Proposição 2.28. Sejam uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ convexa e um ponto $x \in \text{dom } f$. Então,

$$\partial_\varepsilon f(x) \doteq \{v \in \mathbb{R}^n : \langle v, d \rangle \leq f'_\varepsilon(x; d) \text{ para todo } d \in \mathbb{R}^n\}.$$

Corolário 2.29. O ε -subdiferencial $\partial_\varepsilon f(x)$ de uma função convexa $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ em um ponto $x \in \text{ri}(\text{dom } f)$ é o conjunto convexo fechado não vazio de \mathbb{R}^n cuja função suporte é $f'_\varepsilon(x; \cdot)$. Em outras palavras,

$$f'_\varepsilon(x; d) = \sup \{ \langle v, d \rangle : v \in \partial_\varepsilon f(x) \}.$$

A demonstração da proposição 2.28 está omitida por ser análoga a da proposição 2.20. Além disso, da mesma forma que o corolário 2.21, referimo-nos a [HUL93a, teorema V.3.1.1] para verificar a igualdade do corolário 2.29 acima. Utilizamos também a notação $\partial_\varepsilon f(\cdot)_S$ para indicar a parte de $\partial_\varepsilon f(\cdot)$ paralela a $\text{aff}(\text{dom } f)$. Podemos ainda mostrar que o subdiferencial aproximado $\partial_\varepsilon f(x)$ é um conjunto não vazio, convexo, fechado e que $\partial_\varepsilon f(x)_S$ é limitado, quando x pertence ao interior relativo do domínio efetivo de f [HUL93b, teorema XI.1.1.4].

Finalmente, ressaltamos que as definições 2.17 e 2.26, dadas nesta seção, são apropriadas apenas para funções convexas. Se, ao invés disso, trabalharmos com uma função côncava $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{-\infty\}$, devemos mudar a condição $x \in \text{dom } f$ para x tal que $f(x) > -\infty$, além de inverter as desigualdades de (2.2) e (2.7). Em (2.7), devemos também realizar a mudança de “ $-\varepsilon$ ” para “ $+\varepsilon$ ”. Nesse caso, a coerência exige que chamemos $\partial f(\cdot)$ de *superdiferencial* de f e $\partial_\varepsilon f(\cdot)$ de ε -*superdiferencial* de f .

2.2 Continuidade do Subdiferencial Aproximado

Veremos nesta seção que o subdiferencial exato $\partial f(x)$ e o subdiferencial generalizado $\partial_\varepsilon f(x)$ (com $\varepsilon > 0$), vistos como funções multivaloradas, são semi-contínuos exteriores. Além disso, mostraremos que o ε -subdiferencial (com $\varepsilon > 0$) possui uma propriedade de continuidade ainda mais forte, pois é também semi-contínuo interior. Tal estudo nos permite que justifiquemos a utilização de ε -subdiferenciais no contexto de otimização. De fato, observaremos na seção 3.2 que a semi-continuidade interior é importante em algoritmos de otimização baseados em descida.

Lema 2.30. *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa fechada. Então o ε -subdiferencial $(\varepsilon, x) \rightarrow \partial_\varepsilon f(x)$ possui um gráfico fechado para todo $\varepsilon \geq 0$, ou seja,*

$$\begin{cases} x^k \rightarrow x^* \\ \varepsilon_k \rightarrow \varepsilon_*, \varepsilon_k \geq 0 \\ v^k \in \partial_{\varepsilon_k} f(x^k) \rightarrow v^* \end{cases} \implies v^* \in \partial_{\varepsilon_*} f(x^*). \quad (2.8)$$

Demonstração. Utilizando a definição de ε_k -subgradiente, temos que

$$f(y) \geq f(x^k) + \langle v^k, y - x^k \rangle - \varepsilon_k \quad \text{para todo } y \in \mathbb{R}^n.$$

Tomando o limite em k , e considerando que f é semi-contínua inferior, temos

$$\begin{aligned} f(y) &\geq \liminf_{k \rightarrow \infty} (f(x^k) + \langle v^k, y - x^k \rangle - \varepsilon_k) \quad \text{para todo } y \in \mathbb{R}^n \\ \Leftrightarrow f(y) &\geq f(x^*) + \langle v^*, y - x^* \rangle - \varepsilon_* \quad \text{para todo } y \in \mathbb{R}^n \end{aligned}$$

ou seja, $v^* \in \partial_{\varepsilon_*} f(x^*)$. □

Lema 2.31. *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa fechada. Sejam $\delta > 0$ e L a constante Lipschitz de f em $B(x, \delta) \cap \text{aff}(\text{dom } f)$ com $x \in \text{ri}(\text{dom } f)$. Então, para todo $\delta' < \delta$,*

$$\|v\| \leq L + \frac{\varepsilon}{\delta - \delta'},$$

para qualquer $v \in \partial_\varepsilon f(y)_S$, com $y \in B(x, \delta')$.

Demonstração. Observe que a desigualdade a ser provada é trivial quando $y \notin \text{aff}(\text{dom } f)$, pois neste caso $\partial_\varepsilon f(y)_S = 0$ para todo $\varepsilon \geq 0$. Logo, consideramos aqui $y \in B(x, \delta') \cap \text{aff}(\text{dom } f)$. Dessa forma, fixe $v \in \partial_\varepsilon f(y)_S$ com $y \in B(x, \delta') \cap \text{aff}(\text{dom } f)$. Então,

$$f(y) - f(z) \leq \langle v, y - z \rangle + \varepsilon \quad \text{para todo } z \in \text{aff}(\text{dom } f). \quad (2.9)$$

Como f é Lipschitz contínua, para todo $\bar{y}, z \in B(x, \delta) \cap \text{aff}(\text{dom } f)$,

$$|f(z) - f(\bar{y})| \leq L\|z - \bar{y}\| \Rightarrow -L\|\bar{y} - z\| \leq f(\bar{y}) - f(z). \quad (2.10)$$

Dessa forma, de (2.9) e (2.10), com $\bar{y} \doteq y$ (note que $\delta' < \delta$), temos que

$$-L\|y - z\| \leq \langle v, y - z \rangle + \varepsilon \quad \text{para todo } z \in B(x, \delta) \cap \text{aff}(\text{dom } f). \quad (2.11)$$

Defina $z \doteq y + ((\delta - \delta')/\|v\|)v$ e observe que nesse caso $\|y - z\| = |\delta - \delta'|$. Lembrando que $\delta' < \delta$ e considerando $y \in B(x, \delta') \cap \text{aff}(\text{dom } f)$, temos que

essa definição de z nos garante que $z \in B(x, \delta) \cap \text{aff}(\text{dom } f)$. Assim, podemos substituir tal z em (2.11):

$$-L \left\| \frac{\delta - \delta'}{\|v\|} v \right\| \leq \left\langle v, -\frac{\delta - \delta'}{\|v\|} v \right\rangle + \varepsilon \Leftrightarrow \|v\| \leq L + \frac{\varepsilon}{\delta - \delta'},$$

obtendo portanto o resultado desejado. \square

Teorema 2.32. *Sejam $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa e $\varepsilon \geq 0$. Então o ε -subdiferencial $\partial_\varepsilon f(x)$ é semi-contínuo exterior em todo $x \in \mathbb{R}^n$, ou seja, para todo $r > 0$, existe $\delta > 0$ tal que $y \in B(x, \delta)$ implica*

$$\partial_\varepsilon f(y) \subset \partial_\varepsilon f(x) + B(0, r). \quad (2.12)$$

Demonstração. Pela proposição 2.18, podemos escrever (2.12) como:

$$\partial_\varepsilon f(y)_S + S^\perp \subset \partial_\varepsilon f(x)_S + S^\perp + B(0, r).$$

Assim, basta provarmos aqui que para todo $r > 0$, existe $\delta > 0$ tal que $y \in B(x, \delta)$ com $x \in \text{ri}(\text{dom } f)$ implica

$$\partial_\varepsilon f(y)_S \subset \partial_\varepsilon f(x)_S + B(0, r).$$

Suponha, por contradição, que para algum $x \in \text{ri}(\text{dom } f)$ existem $r > 0$ e seqüências $\{x^i\}$ e $\{v^i\}$ com $x^i \rightarrow x$ para $i \rightarrow +\infty$, e $v^i \in \partial_\varepsilon f(x^i)_S$ com $v^i \notin \partial_\varepsilon f(x)_S + B(0, r)$ para todo $i = 1, 2, \dots$. Mas pelos lemas 2.30 e 2.31, a seqüência $\{v^i\}$ é limitada e converge para $v \in \partial_\varepsilon f(x)_S$, o que é uma contradição. Logo, $\partial_\varepsilon f(x)$ é, de fato, semi-contínuo exterior. \square

Observe que os lemas 2.30 e 2.31 e o teorema 2.32 são apropriados para qualquer $\varepsilon \geq 0$, ou seja, mostramos que a semi-continuidade exterior é válida tanto para subdiferenciais aproximados quanto para os exatos (i.e., com $\varepsilon = 0$). Verifiquemos agora a semi-continuidade interior de $\partial_\varepsilon f(\cdot)$. Para isso, considere inicialmente a seguinte definição:

Definição 2.33. *Sejam $A, B \subset \mathbb{R}^n$ conjuntos compactos não vazios. Então*

$$\Delta_H(A, B) \doteq \max \left\{ \sup_{x \in B} \left\{ \text{dist}(x, A) \right\}, \sup_{x \in A} \left\{ \text{dist}(x, B) \right\} \right\}.$$

é denominada distância de Hausdorff entre A e B .

Lema 2.34. *Sejam $A, B \subset \mathbb{R}^n$ conjuntos compactos não vazios. Então*

$$\Delta_H(A, B) = \max_{\|d\| \leq 1} \left\{ |\sigma_A(d) - \sigma_B(d)| \right\}.$$

Demonstração. Geometricamente é fácil ver que $\sup_{d \in B} \{\text{dist}(d, A)\} \leq r$ significa $B \subset A + \bar{B}(0, r)$, ou seja, $x \in A + \bar{B}(0, r)$ para todo $x \in B$. Além disso, utilizando a definição de função suporte, temos:

$$\begin{aligned} x \in A + \bar{B}(0, r) & \qquad \qquad \qquad \forall x \in B \\ \Leftrightarrow \langle x, d \rangle \leq \sigma_{A+\bar{B}(0,r)}(d) & \qquad \qquad \qquad \forall x \in B \text{ e } \forall d \in \mathbb{R}^n \\ \Leftrightarrow \sigma_B(d) \leq \sup_{x \in A+\bar{B}(0,r)} \{\langle d, x \rangle\} & \qquad \qquad \qquad \forall d \in \mathbb{R}^n \\ \Leftrightarrow \sigma_B(d) \leq \sup_{x_1 \in A, x_2 \in \bar{B}(0,r)} \{\langle d, x_1 + x_2 \rangle\} & \qquad \qquad \qquad \forall d \in \mathbb{R}^n \\ \Leftrightarrow \sigma_B(d) \leq \sup_{x_1 \in A} \{\langle d, x_1 \rangle\} + \sup_{x_2 \in \bar{B}(0,r)} \{\langle d, x_2 \rangle\} & \qquad \qquad \qquad \forall d \in \mathbb{R}^n \\ \Leftrightarrow \sigma_B(d) \leq \sigma_A(d) + \sigma_{\bar{B}(0,r)}(d) & \qquad \qquad \qquad \forall d \in \mathbb{R}^n \end{aligned}$$

Mostremos agora que $\sigma_{\bar{B}(0,r)}(d) = r\|d\|$. Por definição de função suporte, temos, para qualquer $d \neq 0$:

$$\sigma_{\bar{B}(0,r)}(d) \geq \left\langle \frac{rd}{\|d\|}, d \right\rangle = r\|d\|.$$

Por outro lado, pela desigualdade de Cauchy-Schwarz,

$$\sigma_{\bar{B}(0,r)}(d) = \sup_{x \in \bar{B}(0,r)} \langle d, x \rangle \leq \sup_{x \in \bar{B}(0,r)} \|d\|\|x\| = r\|d\|.$$

Ou seja, $\sigma_{\bar{B}(0,r)}(d) = r\|d\|$. Dessa forma, temos:

$$\sigma_B(d) \leq \sigma_A(d) + r\|d\| \text{ para todo } d \in \mathbb{R}^n \Leftrightarrow \max_{\|d\| \leq 1} \{\sigma_B(d) - \sigma_A(d)\} \leq r.$$

Em outras palavras, mostramos que

$$\sup_{d \in B} \{\text{dist}(d, A)\} = \max_{\|d\| \leq 1} \{\sigma_B(d) - \sigma_A(d)\}.$$

De maneira análoga, podemos concluir que

$$\sup_{d \in A} \{\text{dist}(d, B)\} = \max_{\|d\| \leq 1} \{\sigma_A(d) - \sigma_B(d)\}.$$

Logo, $\Delta_H(A, B) = \max_{\|d\| \leq 1} \{|\sigma_A(d) - \sigma_B(d)|\}$. □

Lema 2.35. *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e Lipschitz em \mathbb{R}^n . Então, existe $M > 0$ tal que, para todo $x, y \in \mathbb{R}^n$ e $\varepsilon, \varepsilon' > 0$, tem-se:*

$$\Delta_H(\partial_\varepsilon f(x), \partial_{\varepsilon'} f(y)) \leq \frac{M}{\min\{\varepsilon, \varepsilon'\}} (\|x - y\| + |\varepsilon - \varepsilon'|).$$

Demonstração. Seja $d \in \mathbb{R}^n$ unitária, isto é tal que $\|d\| = 1$. Defina $q_\varepsilon(x, t) \doteq (f(x + td) - f(x) + \varepsilon)/t$. Então, para todo $\eta > 0$, existe um $t_\eta > 0$ tal que

$$q_\varepsilon(x, t_\eta) \leq f'_\varepsilon(x; d) + \eta. \quad (2.13)$$

Por hipótese, podemos tomar $\delta \rightarrow +\infty$ no lema 2.31 pois existe uma constante de Lipschitz global L tal que $\|v\| \leq L$ para todo $v \in \partial_\varepsilon f(x)$. Sabemos ainda, pela desigualdade de Cauchy-Schwarz, que $|\langle v, d \rangle| \leq \|v\| \|d\| = \|v\|$. Assim, pelo corolário 2.29,

$$f'_\varepsilon(x; d) = \sup\{\langle v, d \rangle : v \in \partial_\varepsilon f(x)\} \leq L.$$

Considerando a desigualdade abaixo, válida simplesmente pela definição,

$$-L + \frac{\varepsilon}{t_\eta} \leq q_\varepsilon(x, t_\eta) \leq L + \eta,$$

obtemos a relação

$$\frac{1}{t_\eta} \leq \frac{2L + \eta}{\varepsilon}. \quad (2.14)$$

Assim, utilizando (2.13), (2.14) e a definição de derivada direcional, temos:

$$\begin{aligned} & f'_{\varepsilon'}(y; d) - f'_\varepsilon(x; d) - \eta \\ & \leq q_{\varepsilon'}(y, t_\eta) - q_\varepsilon(x, t_\eta) \\ & = (f(y + t_\eta d) - f(x + t_\eta d) + f(x) - f(y) + \varepsilon' - \varepsilon)/t_\eta \\ & \leq (2L\|y - x\| + |\varepsilon' - \varepsilon|)/t_\eta \\ & \leq [(2L + \eta)/\varepsilon](2L\|y - x\| + |\varepsilon' - \varepsilon|). \end{aligned}$$

Analogamente, obtemos o seguinte resultado:

$$f'_\varepsilon(x; d) - f'_{\varepsilon'}(y; d) - \eta \leq [(2L + \eta)/\varepsilon'](2L\|y - x\| + |\varepsilon' - \varepsilon|).$$

Como $\eta > 0$ é arbitrário, temos que

$$|f'_{\varepsilon'}(y; d) - f'_\varepsilon(x; d)| \leq [2L/\min\{\varepsilon, \varepsilon'\}](2L\|y - x\| + |\varepsilon' - \varepsilon|).$$

Portanto, concluímos a demonstração tomando-se $M \doteq \max\{2L, 4L^2\}$. \square

Observe que o lema acima considera f definida apenas nos reais. O resultado abaixo mostra uma generalização desse lema para os reais estendidos. Omitiremos sua demonstração, mas referimo-nos a [HU80] para verificá-la.

Lema 2.36. *Sejam $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa e $K \subset \text{dom } f$ compacto. Suponha que para algum $r_0 > 0$, $\partial f(x) \cap B(0, r_0)$ é não vazio para todo $x \in K$. Então, para todo $r \geq r_0$, existe M_r tal que*

$$\Delta_H(\partial_\varepsilon f(x) \cap B(0, r), \partial_{\varepsilon'} f(y) \cap B(0, r)) \leq \frac{M_r}{\min\{\varepsilon, \varepsilon'\}} (\|x - y\| + |\varepsilon - \varepsilon'|),$$

para todo $x, y \in K$ e $\varepsilon, \varepsilon' > 0$. Em particular, podemos tomar $K \subset \text{ri}(\text{dom } f)$ e $r \doteq \max\{\sup_{v \in \partial_\varepsilon f(K)_S} \|v\|, \sup_{v \in \partial_{\varepsilon'} f(K)_S} \|v\|\}$.

Teorema 2.37. *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa e $\varepsilon > 0$. Então $\partial_\varepsilon f(x)$ é semi-contínua interior em todo compacto $K \subset \text{ri}(\text{dom } f)$.*

Demonstração. Conforme o lema 2.36, seja $r \doteq \sup_{v \in \partial_\varepsilon f(K)_S} \|v\|$ e note que para todo $x, y \in K$,

$$\partial_\varepsilon f(x) \cap B(0, r) \subset \partial_\varepsilon f(y) \cap B(0, r) + \|x - y\|B(0, M_r/\varepsilon). \quad (2.15)$$

Pela proposição 2.18 e a definição de r , temos:

$$\partial_\varepsilon f(x) \cap B(0, r) = (\partial_\varepsilon f(x)_S + S^\perp) \cap B(0, r) \supset \partial_\varepsilon f(x)_S \cap B(0, r) = \partial_\varepsilon f(x)_S.$$

Além disso, sabemos que $\partial_\varepsilon f(y) \cap B(0, r) \subset \partial_\varepsilon f(y)$. Logo, por (2.15), $\partial_\varepsilon f(x)_S \subset \partial_\varepsilon f(y) + \|x - y\|B(0, M_r/\varepsilon)$ e somando-se S^\perp nos dois lados,

$$\begin{aligned} \partial_\varepsilon f(x)_S + S^\perp &\subset \partial_\varepsilon f(y) + S^\perp + \|x - y\|B(0, M_r/\varepsilon) \\ \Rightarrow \partial_\varepsilon f(x) &\subset \partial_\varepsilon f(y) + \|x - y\|B(0, M_r/\varepsilon), \end{aligned}$$

o que mostra que $\partial_\varepsilon f(x)$ é semi-contínua interior. \square

Concluimos esse capítulo mostrando, através de um exemplo, que a propriedade de semi-continuidade interior de $\partial_\varepsilon f(x)$ é válida apenas para $\varepsilon > 0$. Considere a função módulo $\mathbb{R} \ni x \mapsto f(x) \doteq |x|$. Na figura 2.2.1, o gráfico da esquerda ilustra a função multivalorada $\partial f(x)$, enquanto que o da direita apresenta $\partial_\varepsilon f(x)$ para um $\varepsilon > 0$ fixo. Note, por exemplo, que o conjunto $\partial f(0) = [-1, 1]$ é muito “maior” do que $\partial f(x) = \{1\}$, quando $x > 0$. Por outro lado, $\partial_\varepsilon f(x)$ não “explode” quando x se aproxima do ponto 0.

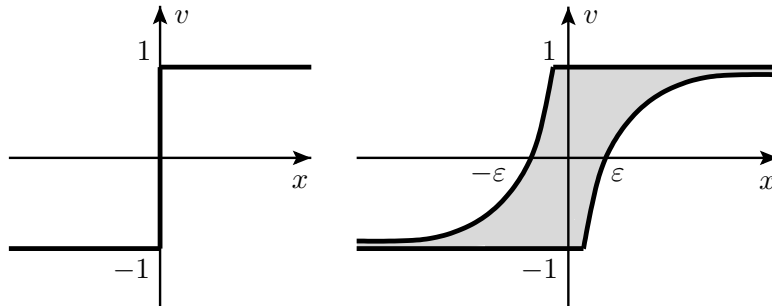


Figura 2.2.1: Semi-continuidade interior do ε -subdiferencial.

Capítulo 3

Otimização Não Diferenciável

Considerando os conceitos de Análise Convexa do capítulo anterior, estamos prontos para abordar os métodos nos quais se baseia o algoritmo do volume. Como referências para este capítulo, recomendamos [BGLS03], [HUL93a], [HUL93b] e [Sho85], bases deste próprio texto. Iniciaremos nosso estudo com o seguinte problema:

$$\text{minimizar } f(x) \quad \text{sujeito a } x \in X, \quad (3.1)$$

onde o conjunto $X \subset \mathbb{R}^n$, correspondente às restrições, é convexo e a função objetivo $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é convexa, semi-contínua inferior, própria, e não necessariamente diferenciável em todos os pontos. Quando $X = \mathbb{R}^n$, o problema correspondente é denominado irrestrito:

$$\text{minimizar } f(x) \quad \text{sujeito a } x \in \mathbb{R}^n. \quad (3.2)$$

De modo geral, os algoritmos que veremos aqui fundamentam-se na geração de iterados x^k através da busca de possíveis direções de descida d^k e tamanhos de passo t_k , e conseqüente atualização na forma $x^{k+1} \doteq x^k + t_k d^k$. Verificaremos na seção 3.1 as condições de otimalidade de (3.1) e (3.2) e em seguida estudaremos dois tipos básicos de métodos que propõem a resolução de (3.2): de descida (seção 3.2) e de caixa preta. Este último tipo, por sua vez, inclui os métodos de subgradientes (seção 3.3) e de planos de corte (seção 3.4). Finalizamos o assunto no capítulo 4, com o método de feixe.

3.1 Condições de Otimalidade

Teorema 3.1. (*Condições de otimalidade*) Considere o problema restrito (3.1). Então as seguintes afirmações são equivalentes:

- (a) x^* é solução ótima global, ou seja, $f(x^*) \leq f(x)$ para todo $x \in X$;
- (b) $f'(x^*; x - x^*) \geq 0$ para todo $x \in X$;
- (c) $f'(x^*; d) \geq 0$ para todo $d \in T_X(x^*)$;
- (d) $0 \in \partial f(x^*) + N_X(x^*)$.

Demonstração. Mostraremos as equivalências acima através das seguintes implicações: (a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (a).

[(a) \Rightarrow (b)] Seja $x \in X$. Como X é convexo, temos que $x^* + \lambda(x - x^*) \in X$ para todo $\lambda \in]0, 1[$. Assim, $f(x^* + \lambda(x - x^*)) \geq f(x^*)$ para todo $\lambda \in]0, 1[$. Isso implica que

$$[f(x^* + \lambda(x - x^*)) - f(x^*)]/\lambda \geq 0, \quad \text{para todo } \lambda \in]0, 1[.$$

Finalmente, tomando $\lambda \downarrow 0$, temos $f'(x^*; x - x^*) \geq 0$, para todo $x \in X$.

[(b) \Rightarrow (c)] Considere $d \doteq x - x^*$. Então, (b) implica $f'(x^*; d) \geq 0$ para todo $d \in X - \{x^*\}$. Observando a definição de derivada direcional, é fácil ver também que $f'(x^*; \alpha d) \geq 0$, para todo $d \in X - \{x^*\}$ e todo $\alpha > 0$. Em outras palavras, $f'(x^*; d) \geq 0$ para todo $d \in \overline{\text{cone}}(X - \{x^*\})$, que pela proposição 2.13, equivale a $T_X(x^*)$.

[(c) \Rightarrow (d)] Como $f'(x^*; \cdot) > -\infty$, temos,

$$f'(x^*; d) \geq 0 \quad \forall d \in T_X(x^*) \Leftrightarrow f'(x^*; d) + i_{T_X(x^*)}(d) \geq 0 \quad \forall d \in \mathbb{R}^n. \quad (3.3)$$

Pela definição de função suporte, temos:

$$\sigma_{N_X(x^*)}(d) = \sup_{s \in N_X(x^*)} \langle s, d \rangle = \begin{cases} 0, & \text{se } \langle s, d \rangle \leq 0, \quad \forall s \in N_X(x^*) \\ +\infty, & \text{caso contrário} \end{cases}$$

que é igual à indicadora $i_{(N_X(x^*))^\circ}(d) = i_{T_X(x^*)}(d)$. Com isso e pelo corolário 2.21, temos que (3.3) é equivalente a

$$\begin{aligned} & \sigma_{\partial f(x^*)}(d) + \sigma_{N_X(x^*)}(d) && \geq 0 \quad \forall d \in \mathbb{R}^n \\ \Leftrightarrow & \sup_{v_1 \in \partial f(x^*)} \langle v_1, d \rangle + \sup_{v_2 \in N_X(x^*)} \langle v_2, d \rangle && \geq 0 \quad \forall d \in \mathbb{R}^n \\ \Leftrightarrow & \sup_{v_1 \in \partial f(x^*), v_2 \in N_X(x^*)} \langle v_1 + v_2, d \rangle && \geq 0 \quad \forall d \in \mathbb{R}^n \\ \Leftrightarrow & \sup_{v \in \partial f(x^*) + N_X(x^*)} \langle v, d \rangle && \geq 0 \quad \forall d \in \mathbb{R}^n \end{aligned} \quad (3.4)$$

Suponha agora, por contradição, que $0 \notin \partial f(x^*) + N_X(x^*)$. Como $\{0\}$ é compacto e $\partial f(x^*) + N_X(x^*)$ é fechado, então, pelo teorema 2.16, $\{0\}$ e $\partial f(x^*) + N_X(x^*)$ podem ser separados estritamente, isto é,

$$\text{existe } \bar{d} \text{ tal que } \langle \bar{d}, 0 \rangle > \langle v, \bar{d} \rangle \text{ para todo } v \in \partial f(x^*) + N_X(x^*).$$

Em particular, $0 = \langle \bar{d}, 0 \rangle > \sup_{v \in \partial f(x^*) + N_X(x^*)} \langle v, \bar{d} \rangle$, o que contradiz a última desigualdade de (3.4). Logo, $0 \in \partial f(x^*) + N_X(x^*)$.

[(d) \Rightarrow (a)] Tome $v \in \partial f(x^*)$. Como $0 \in \partial f(x^*) + N_X(x^*)$, temos que $-v \in N_X(x^*)$. Portanto, por definição de $\partial f(x^*)$ e $N_X(x^*)$,

$$\begin{aligned} f(x) &\geq f(x^*) + \langle v, x - x^* \rangle \text{ e } \langle -v, x - x^* \rangle \leq 0 \quad \forall x \in X \\ \Rightarrow f(x) &\geq f(x^*) + \langle v, x - x^* \rangle + \langle -v, x - x^* \rangle \quad \forall x \in X. \end{aligned}$$

Portanto, $f(x) \geq f(x^*)$ para todo $x \in X$. \square

No caso de problemas irrestritos, as condições de otimalidade podem ser simplificadas, conforme se observa a seguir.

Corolário 3.2. *(Condições de otimalidade) Considere o problema irrestrito (3.2). Então as seguintes afirmações são equivalentes:*

- (a) x^* é solução ótima global, ou seja, $f(x^*) \leq f(x)$ para todo $x \in \mathbb{R}^n$;
- (b) $f'(x^*; d) \geq 0$ para todo $d \in \mathbb{R}^n$;
- (c) $0 \in \partial f(x^*)$.

Demonstração. Segue diretamente do teorema 3.1, substituindo X por \mathbb{R}^n . Observe apenas que $N_{\mathbb{R}^n}(x^*) = \emptyset$ e que os itens (b) e (c) do teorema 3.1 resumem-se ao item (b) deste corolário. \square

Pensemos por enquanto no problema irrestrito (3.2). Se $x \in \mathbb{R}^n$ não é ótimo global, então, pelo corolário acima, existe pelo menos uma direção d tal que $f'(x; d) < 0$. Tal direção é denominada *direção de descida* de f no ponto x . Da mesma forma que na otimização diferenciável, as direções de descida são, portanto, aquelas que garantem o decréscimo da função objetivo quando as utilizamos para gerar uma seqüência de iterados.

Lembrando que $f'(x; d) = \sup_{v \in \partial f(x)} \langle v, d \rangle$, podemos também dizer que uma direção d é de descida se $\langle v, d \rangle < 0$ para todo $v \in \partial f(x)$. Observe ainda, com auxílio da figura 3.1.1, uma outra caracterização de tal direção: dizemos que d é de descida se o hiperplano $H_\alpha \doteq \{z \in \mathbb{R}^n : \langle z, d \rangle = \alpha\}$, com $\alpha \in [f'(x; d), 0[$, separa estritamente os conjuntos $\partial f(x)$ e $\{0\}$ (isto é, $\langle v, d \rangle \leq \alpha < 0$ para todo $v \in \partial f(x)$).

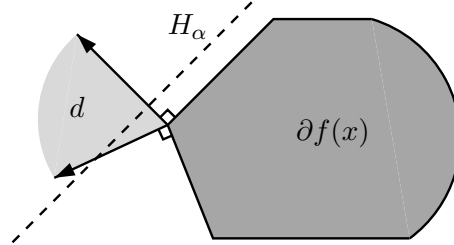


Figura 3.1.1: Direções de descida.

Do ponto de vista computacional, em algoritmos iterativos de otimização, é possível que não sejam obtidas soluções exatas, mas apenas suas aproximações. Desse modo, devemos requerer maneiras de tratar soluções ε -ótimas (i.e., próximas da solução ótima). Por esse motivo, enunciamos no teorema a seguir as condições de ε -otimalidade. Sua demonstração será omitida por ser análoga a do teorema 3.1.

Teorema 3.3. (Condições de ε -otimalidade) *Considere o problema restrito (3.1). Então as seguintes afirmações são equivalentes:*

- (a) x^* é solução ε -ótima global, ou seja, $f(x^*) - \varepsilon \leq f(x)$ para todo $x \in X$;
- (b) $f'_\varepsilon(x^*; x - x^*) \geq 0$ para todo $x \in X$;
- (c) $f'_\varepsilon(x^*; d) \geq 0$ para todo $d \in T_X(x^*)$;
- (d) $0 \in \partial_\varepsilon f(x^*) + N_X(x^*)$.

Analogamente, para o problema irrestrito, temos:

Corolário 3.4. (Condições de ε -otimalidade) *Considere o problema irrestrito (3.2). Então as seguintes afirmações são equivalentes:*

- (a) x^* é solução ε -ótima global, ou seja, $f(x^*) - \varepsilon \leq f(x)$ para todo $x \in \mathbb{R}^n$;
- (c) $f'_\varepsilon(x; d) \geq 0$ para todo $d \in \mathbb{R}^n$;
- (b) $0 \in \partial_\varepsilon f(x)$.

Novamente considerando o problema irrestrito (3.2), dizemos que uma direção $d \in \mathbb{R}^n$ é de ε -descida a partir de um ponto $x \in \mathbb{R}^n$ se existe $t > 0$ tal que $f(x + td) < f(x) - \varepsilon$ ou, equivalentemente, se $f'_\varepsilon(x; d) < 0$; mais ainda, se $\langle v, d \rangle < 0$ para todo $v \in \partial_\varepsilon f(x)$. Do ponto de vista geométrico, isso equivale a dizer que o hiperplano $H_\alpha \doteq \{z \in \mathbb{R}^n : \langle z, d \rangle = \alpha\}$, com $\alpha \in [f'_\varepsilon(x; d), 0[$, separa estritamente os conjuntos $\partial_\varepsilon f(x)$ e $\{0\}$.

Com as condições de otimalidade e as caracterizações de direções de descida e de ε -descida, estamos prontos para abordar os métodos que tentam resolver os problemas (3.1) e (3.2). No restante deste capítulo, teremos como objetivo a resolução do problema irrestrito (3.2), necessária para a compreensão dos capítulos posteriores.

3.2 Método de Descida

Os *métodos de descida* baseiam-se na geração de uma seqüência $\{x^k\}$ com a garantia de decréscimo da função objetivo f em cada iteração. As direções a serem tomadas são, portanto, as de descida, caracterizadas anteriormente. O algoritmo pode ser visto a seguir.

Algoritmo 3.5. *Métodos de Descida*

1. Tome $x^1 \in \mathbb{R}^n$ e seja $k = 1$.
2. (Critério de parada formal) Se $0 \in \partial f(x^k)$, pare.
3. (Descida) Encontre uma direção de descida d^k de f em x^k .
4. (Busca linear) Encontre um tamanho de passo $t_k > 0$ tal que

$$f(x^k + t_k d^k) < f(x^k).$$

5. (Próximo iterado) Defina $x^{k+1} \doteq x^k + t_k d^k$.
6. (Loop) Tome $k = k + 1$ e vá para o passo 2.

Vale ressaltar que o critério de parada oferecido pelo passo 2 é puramente formal, pois a obtenção de todo o subdiferencial $\partial f(x^k)$ é algo excessivo ou mesmo impossível computacionalmente. Critérios de parada implementáveis serão vistos posteriormente com detalhes nos algoritmos de planos de corte e de feixe (seção 3.4 e capítulo 4, respectivamente). Por ora, note apenas que a melhor descida d^k possível (i.e., a direção de máxima descida) é a solução do problema $\min_{\|d\|=1} f'(x^k; d)$, ou, equivalentemente, pelo corolário 2.21, de

$$\min_{\|d\|=1} \max_{v \in \partial f(x)} \langle v, d \rangle.$$

Geometricamente, isso significa que a direção de máxima descida é justamente a que está relacionada ao hiperplano H ortogonal à projeção de $\{0\}$ em $\partial f(x^k)$. Mais especificamente, $d^k = -\gamma^k / \|\gamma^k\|$, onde $\gamma^k \doteq P_{\partial f(x^k)}(0)$. Como ilustração, veja figura 3.2.1.

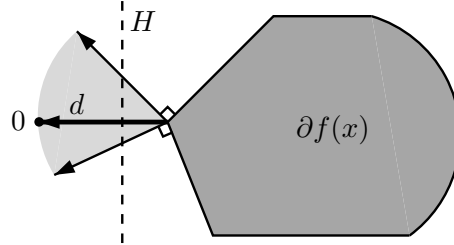


Figura 3.2.1: Direção de máxima descida.

Um problema do método de máxima descida é que a seqüência de iterados $\{x^k\}$ pode oscilar e convergir para um ponto não ótimo. Referimo-nos a [HUL93a, seção VIII.2.2] para verificar, através de um exemplo numérico, que o método pode, de fato, não convergir. Para compreendermos esse fato, lembremos primeiro que o algoritmo de descida converge se $\{f(x^k)\}$ é decrescente e se $\{x^k\}$ possui um ponto de acumulação x^* , que é minimizador de f . Considere, então, a seguinte seqüência:

$$\{\delta_k\} \doteq \{\text{dist}(0, \partial f(x^k))\} = \{\|\gamma^k\|\}.$$

Vimos na seção 2.2, que o subdiferencial $\partial f(x)$, visto como uma função multivalorada, é fechado. Isso significa que

$$\begin{cases} x^k \rightarrow x^* \\ v^k \in \partial f(x^k) \rightarrow v^* \end{cases} \implies v^* \in \partial f(x^*).$$

Dessa forma, se $\delta_k \rightarrow 0$, então $0 \in \partial f(x^*)$ e x^* é mínimo. Para assegurar que $\delta_k \rightarrow 0$, a função multivalorada $\partial f(\cdot)$ deveria ser contínua (i.e., semi-contínua interior e exterior). Vimos, no entanto, que ela não possui a propriedade de semi-continuidade interior, ou seja, não é verdade que

$$\begin{cases} x^k \rightarrow x^* \\ v^* \in \partial f(x^*) \end{cases} \implies \text{existe } \{v^k\} \rightarrow v^* \text{ tal que } v^k \in \partial f(x^k).$$

Ainda na seção 2.2, observamos que $\partial_\varepsilon f(\cdot)$ é semi-contínuo interior e exterior para $\varepsilon > 0$. Portanto, uma possível maneira de contornar a não convergência do método de máxima descida é utilizar $\partial_\varepsilon f(\cdot)$, com $\varepsilon > 0$, no lugar de $\partial f(\cdot)$. Tais algoritmos são denominados de ε -descida [HUL93b, capítulo XIII]. Eles geram seqüências $\{x^k\}$ tais que $\{f(x^k)\}$ é decrescente e $\{\text{dist}(0, \partial_\varepsilon f(x^k))\} \rightarrow 0$. O algoritmo é descrito a seguir.

Algoritmo 3.6. *Métodos de ε -Descida*

1. Tome $x^1 \in \mathbb{R}^n$, $\varepsilon > 0$ e seja $k = 1$.

2. (Critério de parada formal) Se $0 \in \partial_\varepsilon f(x^k)$, pare.
3. (Descida) Encontre uma direção de ε -descida d^k de f em x^k .
4. (Busca linear) Encontre um tamanho de passo $t_k > 0$ tal que

$$f(x^k + t_k d^k) < f(x^k) - \varepsilon.$$

5. (Próximo iterado) Defina $x^{k+1} \doteq x^k + t_k d^k$.
6. (Loop) Tome $k = k + 1$ e vá para o passo 2.

Nesse algoritmo, $f(x^k) \rightarrow -\infty$ ou o algoritmo termina numa iteração k_* tal que x^{k_*} é ε -ótimo. Esse é o algoritmo de ε -descida mais simples. Existem variantes do método que permitem escolhas de $\varepsilon = \varepsilon_k$ a cada iteração. Por fim, note que o ε -subdiferencial também não é usualmente conhecido por inteiro, ou seja, o método de ε -descida que descrevemos ainda não é implementável. Nas próximas seções, veremos outros algoritmos que tentam contornar essa questão.

3.3 Método de Subgradientes

Usualmente, a obtenção de um subdiferencial $\partial f(x)$ por inteiro é excessiva ou até impossível do ponto de vista computacional. Uma maneira de contornar esse problema é simplesmente requerermos menos, ou seja, pedirmos o cômputo de apenas um único subgradiente. Tal cômputo está associado 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 $v^k \in \partial f(x^k)$, juntamente com $f(x^k)$, o valor da função objetivo¹. Veremos, no entanto, que nem sempre será possível obter direções reais de descida. A performance de um algoritmo depende, portanto, da sua capacidade de gerar e reconhecer candidatos a direções de descida “suficientemente bons”. Nesta seção, estudaremos a primeira dessas metodologias, denominada *método de subgradientes*.

A idéia do método de subgradientes procede do método de Cauchy para otimização diferenciável, no qual se toma a direção oposta ao gradiente $\nabla f(x)$. No nosso caso, tomaremos então o vetor oposto ao subgradiente v fornecido pela caixa preta. Porém, tal direção não será necessariamente de

¹Observe que gerar uma seqüência de iterados $\{x^k\}$ independe da maneira na qual se calcula os valores $f(x^k)$ e v^k . Em outras palavras, não é necessário saber como se computa $f(x^k)$ e v^k . Daí o nome *caixa preta*.

descida. Conforme pode ser observado na figura 3.3.1, que mostra curvas de nível para funções minimizadas em $0 \in \mathbb{R}^2$, direções de descida (representadas pela parte hachurada clara) devem fazer produto escalar negativo com o subdiferencial $\partial f(x)$ inteiro, e não apenas com um único subgradiente v .

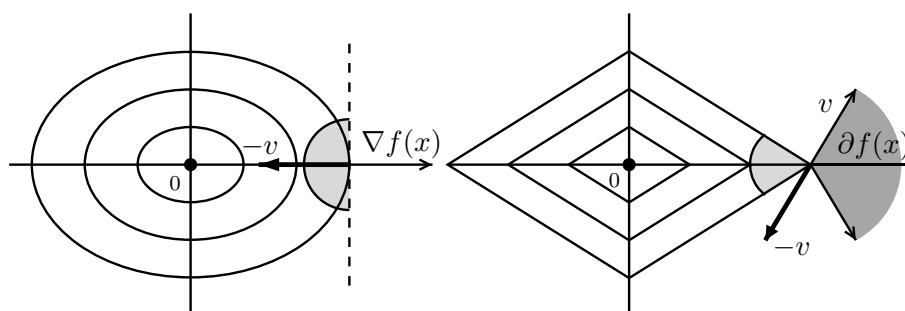


Figura 3.3.1: Direções de descida: caso diferenciável e não diferenciável.

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

Algoritmo 3.7. *Método de Subgradientes*

1. Tome $x^1 \in \mathbb{R}^n$ e seja $k = 1$.
2. (Caixa preta) Compute v^k e $f(x^k)$.
3. (Critério de parada) Se $0 \in \partial f(x^k)$ [i.e., $v^k = 0$], pare.
4. (Busca linear) Tome um passo $t_k > 0$ adequado.
5. (Próximo iterado) Defina $x^{k+1} \doteq x^k - t_k v^k$.
6. (Loop) Tome $k = k + 1$ e vá para o passo 2.

Para compreender o passo 4 do algoritmo acima, defina primeiro X^* como o conjunto de soluções ótimas e f_* como o valor ótimo da função objetivo, ou seja, $f_* = f(x^*)$ para todo $x^* \in X^*$. Em nossa análise, necessitaremos também das seguintes hipóteses:

Hipótese 3.8. A seqüência $\{v^k\}$ é limitada por uma constante M .

Hipótese 3.9. O problema (3.2) possui solução ótima, i.e., $X^* \neq \emptyset$.

Com isso, veremos no teorema 3.11 algumas condições para o tamanho do passo que garantem a convergência do método. Para isso, o lema abaixo é essencial.

Lema 3.10. Considere o algoritmo 3.7. Então para qualquer $y \in \mathbb{R}^n$ e iteração k , vale a seguinte desigualdade:

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 + t_k^2 \|v^k\|^2 - 2t_k(f(x^k) - f(y)).$$

Demonstração. Pela definição de x^{k+1} dada no algoritmo, temos:

$$\begin{aligned} \|x^{k+1} - y\|^2 &= \|x^k - t_k v^k - y\|^2 \\ &= \|x^k - y\|^2 + t_k^2 \|v^k\|^2 - 2t_k \langle v^k, x^k - y \rangle \\ &\leq \|x^k - y\|^2 + t_k^2 \|v^k\|^2 - 2t_k(f(x^k) - f(y)), \end{aligned}$$

sendo que a última desigualdade vem da definição de $v^k \in \partial f(x^k)$. \square

Teorema 3.11. Considere o algoritmo 3.7 sob as hipóteses 3.8 e 3.9, com os tamanhos de passo escolhidos por alguma das seguintes formas:

$$(a) \sum_{k=1}^{\infty} t_k = +\infty \text{ e } \sum_{k=1}^{\infty} t_k^2 < +\infty,$$

$$(b) t_k \doteq \beta_k \frac{f(x^k) - f_*}{\|v^k\|^2} \text{ com } 0 < \underline{\beta} \leq \beta_k \leq \bar{\beta} < 2 \text{ para todo } k.$$

Então, o único ponto de acumulação de $\{x^k\}$ pertence a X^* .

Demonstração.

(a) Utilizando o lema 3.10 para $y \doteq x^* \in X^*$ arbitrário,

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + t_k^2 \|v^k\|^2 - 2t_k(f(x^k) - f_*). \quad (3.5)$$

Aplicando essa desigualdade de maneira recursiva, temos:

$$\|x^{k+1} - x^*\|^2 \leq \|x^1 - x^*\|^2 + \sum_{i=1}^k t_i^2 \|v^i\|^2 - 2 \sum_{i=1}^k t_i(f(x^i) - f_*). \quad (3.6)$$

Como $f_* \leq f(x^i)$ e $t_i > 0$ para todo i , temos:

$$\|x^{k+1} - x^*\|^2 \leq \|x^1 - x^*\|^2 + \sum_{i=1}^k t_i^2 \|v^i\|^2.$$

Assim, da limitação de $\|v^i\|$ para todo i e de $\sum_{i=1}^{\infty} t_i^2 < +\infty$, temos que $\{x^k\}$ é limitada e que $\|x^k - x^*\|^2$ converge [Pol87, seção 2.2].

Defina agora $\hat{f}_k^j \doteq \min_{j \leq i \leq k} f(x^i)$ e mostremos que para todo j , $\hat{f}_k^j \rightarrow f_*$ quando $k \rightarrow \infty$. Tome um j arbitrário. De maneira análoga a (3.6), obtemos, aplicando recursivamente a desigualdade (3.5):

$$\|x^{k+1} - x^*\|^2 \leq \|x^j - x^*\|^2 + \sum_{i=j}^k t_i^2 \|v^i\|^2 - 2 \sum_{i=j}^k t_i (f(x^i) - f_*).$$

Como $\|x^{k+1} - x^*\|^2 \geq 0$, podemos escrever:

$$\begin{aligned} \|x^j - x^*\|^2 + \sum_{i=j}^k t_i^2 \|v^i\|^2 &\geq 2 \sum_{i=j}^k t_i (f(x^i) - f_*) \\ &\geq 2 \left(\sum_{i=j}^k t_i \right) \min_{i=j, \dots, k} (f(x^i) - f_*). \end{aligned}$$

Além disso, como $\|v^i\| \leq M$ para todo i ,

$$\hat{f}_k^j - f_* \leq \frac{\|x^j - x^*\|^2 + \sum_{i=j}^k t_i^2 \|v^i\|^2}{2 \sum_{i=j}^k t_i} \leq \frac{\|x^j - x^*\|^2 + M^2 \sum_{i=j}^k t_i^2}{2 \sum_{i=j}^k t_i}.$$

Logo, como x^* é arbitrário, podemos escrever

$$\hat{f}_k^j - f_* \leq \frac{\text{dist}(x^j, X^*) + M^2 \sum_{i=j}^k t_i^2}{2 \sum_{i=j}^k t_i}$$

e com isso, temos que $\hat{f}_k^j - f_* \rightarrow 0$ quando $k \rightarrow \infty$, pois $\text{dist}(x^j, X^*) + M^2 \sum_{i=j}^{\infty} t_i^2$ é limitado e $\sum_{i=j}^{\infty} t_i = +\infty$.

Finalmente, mostremos que um ponto de acumulação \bar{x} de $\{x^k\}$ pertence a X^* . Sabemos que para todo j ,

$$\hat{f}_k^j \doteq \min_{j \leq i \leq k} f(x^i) \geq \inf_{i \geq j} f(x^i) \geq f_*. \quad (3.7)$$

Como f é semi-contínua inferior,

$$f(\bar{x}) \leq \liminf_{i \rightarrow \infty} f(x^i) = \lim_{j \rightarrow \infty} \inf_{i \geq j} f(x^i) \leq \lim_{j \rightarrow \infty} \hat{f}_k^j,$$

onde a última desigualdade é dada por (3.7). Ademais, $\hat{f}_k^j \rightarrow f_*$, então $f(\bar{x}) \leq f_*$, ou mais precisamente, $f(\bar{x}) = f_*$. Logo, qualquer ponto de acumulação dessa seqüência (limitada) $\{x^k\}$ está em X^* , conforme queríamos. Para provar a unicidade de tal ponto de acumulação, suponha que existam \bar{x} e \tilde{x} distintos tais que $x^k \rightarrow \bar{x}$ e $x^k \rightarrow \tilde{x}$. Então, $\|\bar{x} - x^*\| = \|\tilde{x} - x^*\|$, o que é possível somente se $\bar{x} = \tilde{x}$.

(b) Utilizando o lema 3.10 para $y \doteq x^* \in X^*$ e a definição de t_k , temos:

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + t_k^2 \|v^k\|^2 - 2t_k(f(x^k) - f_*) \\ &= \|x^k - x^*\|^2 - \beta_k(2 - \beta_k) \frac{(f(x^k) - f_*)^2}{\|v^k\|^2} \end{aligned} \quad (3.8)$$

Assim, como $\beta_k \in]0, 2[$,

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2$$

e isso implica que $\{x^k\}$ é limitada. Mostremos agora que $f(x^k) \rightarrow f_*$. Suponha, por contradição, que $\{x^k\} \rightarrow \bar{x}$ e que $f(\bar{x}) > f_*$. Então, existem $\delta > 0$ e uma seqüência infinita de índices $k_1 < k_2 < \dots$ tais que $f(x^{k_i}) - f_* > \delta$ para todo $i = 1, 2, \dots$. Pela hipótese 3.8, $\|v^k\|$ é limitado. Além disso, como $0 < \underline{\beta} \leq \beta_k \leq \bar{\beta} < 2$ para todo k , temos, por (3.8), que

$$\|x^{k_i+1} - x^*\|^2 \leq \|x^{k_i} - x^*\|^2 - \varepsilon, \quad i = 1, 2, \dots$$

onde $\varepsilon > 0$. Logo, $\|x^{k_i+1} - x^*\|^2 \leq \|x^{k_1} - x^*\|^2 - i\varepsilon$. Como $\|x^{k_i+1} - x^*\|^2 \geq 0$, temos $\|x^{k_1} - x^*\|^2 - i\varepsilon \geq 0$. Assim, tomando $i \rightarrow \infty$, chegamos a uma contradição. Com isso, para qualquer ponto de acumulação \bar{x} , temos $\bar{x} \in X^*$. Ademais, como $\{\|x^k - x^*\|\}$ é decrescente, ela converge para $\|\bar{x} - x^*\|$ para todo $x^* \in X^*$. Para verificar a unicidade do ponto de acumulação de $\{x^k\}$, suponha que existam \bar{x} e \tilde{x} distintos tais que $\{x^k\} \rightarrow \bar{x}$ e $\{x^k\} \rightarrow \tilde{x}$. Então, $\bar{x}, \tilde{x} \in X^*$ e $\|\bar{x} - x^*\| = \|\tilde{x} - x^*\|$ para todo $x^* \in X^*$. Mas isso é possível somente se $\bar{x} = \tilde{x}$, o que nos leva novamente a uma contradição. \square

Naturalmente, outras condições de tamanho de passo são possíveis de modo a estabelecer convergência do método. De fato, o caso (b) do teorema 3.11 pode ser generalizado da seguinte forma:

$$t_k \doteq \beta_k \frac{f(x^k) - LB}{\|v^k\|^2}, \quad (3.9)$$

onde LB é um limitante inferior do valor ótimo f_* . A utilização de (3.9) é mais conveniente pois geralmente não possuímos, a priori, o valor de f_* .

Referimo-nos a [Pol69] para a demonstração de convergência nesse caso. Por sua vez, esse limitante inferior LB pode ser atualizado a cada iteração k . Veremos detalhes disso na seção 6.3.

Vale ressaltar que embora o método de subgradientes seja considerado base para diversos algoritmos de otimização, sua teoria de convergência não está ainda completa. Isso significa que algumas versões do algoritmo mostradas na teoria são lentas do ponto de vista computacional, enquanto que para muitas versões implementáveis do algoritmo não foram obtidas provas reais de convergência. Finalizamos o assunto listando duas variações do método de subgradientes.

Suponha primeiro que queremos resolver o problema restrito (3.1) e que o conjunto viável X é simples, no sentido de que é fácil projetar um ponto em X . Então, o *método de subgradientes projetados* é baseado na seguinte atualização de iterados:

$$x^{k+1} \doteq P_X(x^k - t_k v^k). \quad (3.10)$$

Observe que a projeção $P_X(\cdot)$ garante que, a cada passo, os iterados x^k sejam viáveis.

Com ou sem projeções, é possível que ocorra zigzags no método de subgradientes. Para ilustrar isso, considere o seguinte exemplo:

$$\begin{aligned} \text{minimizar} \quad & f(x) \doteq x_1 + 2x_2 - 1 \\ \text{sujeito a} \quad & x \in X \doteq \{(x_1, x_2)^T \in \mathbb{R}^2 : x_1 + x_2 = 1; x_1, x_2 \geq 0\}. \end{aligned}$$

Nesse caso, a solução ótima é $x^* = (1, 0)^T$ com $f(x^*) = f_* = 0$. Suponha que a cada passo $\beta_k = 1$. Considerando $x^0 \doteq (0, 1)^T$, teríamos $t_k = f(x^k)/5$,

$$x^k = \left(1 - \left(\frac{9}{10}\right)^k, \left(\frac{9}{10}\right)^k \right)^T \quad \text{e} \quad f(x^k) = \left(\frac{9}{10}\right)^k$$

para todo k . Assim, com essa instância, o algoritmo claramente converge para a solução ótima muito lentamente. Conforme [CFM75], o *método de subgradientes com desvios*, caracterizado por:

$$x^{k+1} \doteq P_X(x^k - t_k d^k) \quad \text{onde} \quad \begin{cases} d^1 \doteq v^1, \\ d^k \doteq v^k + \alpha_k d^{k-1}, \quad (\alpha_k \geq 0), \quad \forall k > 1, \end{cases}$$

tenta diminuir a ocorrência de tal fenômeno considerando uma combinação entre o subgradiente v^k e a direção anterior d^{k-1} . No capítulo 5 utilizaremos todas essas variações, bem como adaptações das mesmas.

3.4 Método de Planos de Corte

O método de planos de corte, ao contrário do método de subgradientes, aproveita as informações obtidas nas iterações anteriores para definir um modelo da função objetivo. Esse modelo é útil para se obter candidatos a direções de descida. Assim, a cada iteração k , temos um modelo linear por partes \tilde{f}_k de f construído da seguinte forma:

$$\tilde{f}_k(\cdot) \doteq \max_{i=1, \dots, k} \{f(x^i) + \langle v^i, \cdot - x^i \rangle\}. \quad (3.11)$$

Note que a cada iteração k , adicionamos uma função afim $f(x^k) + \langle v^k, \cdot - x^k \rangle$ ao modelo. O máximo de todas as funções afins que definem \tilde{f}_k é claramente uma função convexa e linear por partes. Além disso,

$$\tilde{f}_k \leq \tilde{f}_{k+1} \quad \text{e} \quad \tilde{f}_k \leq f \quad (3.12)$$

para todo k , ou seja, \tilde{f}_k se aproxima de f por baixo a cada iteração.

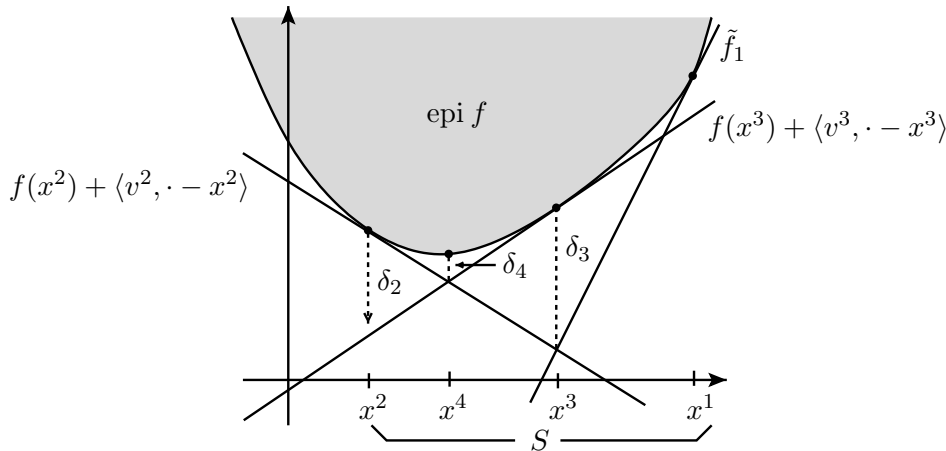


Figura 3.4.1: Iterações do método de planos de corte.

A idéia do método de planos de corte é utilizar esse modelo \tilde{f}_k para encontrar o próximo iterado x^{k+1} . Mais precisamente,

$$x^{k+1} \in \operatorname{argmin}_{x \in S} \tilde{f}_k(x), \quad (3.13)$$

onde S é um conjunto compacto convexo que contém um ponto mínimo de f . Ademais, o modelo ainda nos permite ter um critério de parada

implementável através do cômputo do *decréscimo nominal*, o qual é definido como

$$\delta_k \doteq f(x^k) - \tilde{f}_{k-1}(x^k). \tag{3.14}$$

Observe que o algoritmo termina quando δ_k é pequeno, ou seja, ele estima a otimalidade, conforme veremos no teorema 3.13. Para facilitar a compreensão do método, observe também as iterações do algoritmo na figura 3.4.1.

Algoritmo 3.12. *Método de Planos de Corte*

1. Sejam $\text{tol} \geq 0$ uma tolerância dada e $S \subset \mathbb{R}^n$.
2. Tome $x^1 \in S$ e seja $k = 1$. Defina $\tilde{f}_0 \equiv -\infty$.
3. (Caixa preta) Obtenha v^k e $f(x^k)$.
4. (Decréscimo nominal) Compute $\delta_k \doteq f(x^k) - \tilde{f}_{k-1}(x^k)$.
5. (Critério de parada) Se $\delta_k \leq \text{tol}$, pare.
6. (Próximo iterado) Defina $x^{k+1} \in \text{argmin}_{x \in S} \tilde{f}_k(x)$.
7. (Loop) Tome $k = k + 1$ e vá para o passo 3.

No algoritmo acima, o conjunto S é necessário para se ter uma solução para o problema de minimização do passo 6, além de ser importante para se evitar instabilidades. Referimo-nos a [BGLS03, exemplo 8.7] para se observar um exemplo numérico onde ocorrem tais instabilidades. Note ainda que nesse algoritmo o preço que se paga para se ter um critério de parada simples e implementável é resolver problemas de minimização (3.13). Em particular, se S é poliedral, então o problema do passo 6 é de programação linear:

$$\text{minimizar}_{(d,r)} r \quad \text{sujeito a} \quad \begin{cases} r \geq f(x^i) + \langle v^i, x^k - x^i \rangle + \langle v^i, d \rangle, & i = 1, \dots, k, \\ x^k + d \in S, \\ r \in \mathbb{R}. \end{cases}$$

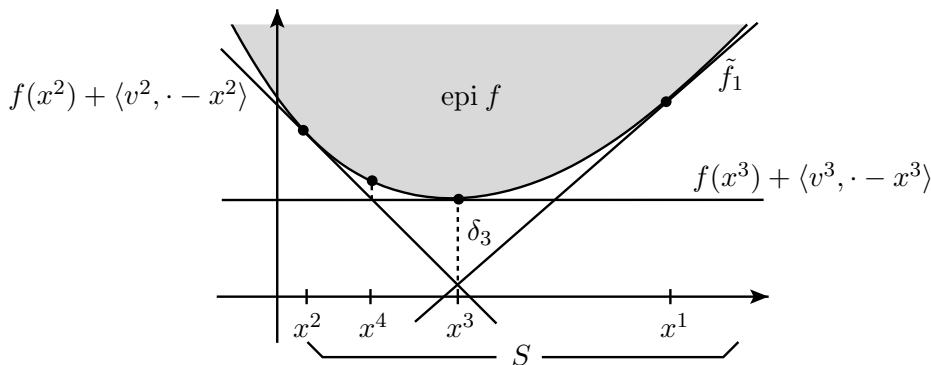


Figura 3.4.2: Introdução de uma função afim quase “horizontal”.

Assim como o método de subgradientes, o método de planos de corte não garante o decréscimo da função objetivo a cada iteração. Tal fato pode ser observado quando introduzimos uma função afim quase “horizontal” ao modelo \tilde{f}_k . Por essa razão, dizemos que o algoritmo não está livre de instabilidades. Para maiores esclarecimentos, veja a figura 3.4.2 acima. Observe que a introdução da função afim $f(x^3) + \langle v^3, \cdot - x^3 \rangle$ gera um ponto x^4 tal que $f(x^4) > f(x^3)$.

Convém verificarmos ainda que o algoritmo acumula um número crescente de funções afins que definem o modelo, o que dificulta a resolução dos problemas do passo 6, mesmo se essas forem lineares. A dificuldade não está apenas na existência de um grande número de funções afins, mas no fato de que várias dessas restrições são quase idênticas entre si. Veremos na seção posterior como resolver esse problema, “limpando-se” o modelo.

Concluiremos essa seção com as demonstrações de convergência do método. A análise está dividida em duas partes: quando $\{x^k\}$ é finita e quando $\{x^k\}$ é infinita. Seja f_* o valor ótimo da função objetivo e defina \hat{f}_k como o melhor valor da função objetivo encontrado até a iteração k . Em outras palavras, seja $\hat{f}_k \doteq \min_{i=1, \dots, k} f(x^i)$.

Teorema 3.13. *Considere uma seqüência $\{x^k\}$ gerada pelo algoritmo 3.12 sob as hipóteses 3.8 e 3.9, e suponha que ele termina em uma iteração k_f . Então,*

$$f(x^{k_f}) \leq f_* + \text{tol}.$$

Demonstração. Como o algoritmo termina na iteração k_f , temos:

$$\delta_{k_f} \leq \text{tol} \Leftrightarrow f(x^{k_f}) \leq \tilde{f}_{k_f-1}(x^{k_f}) + \text{tol}.$$

Por (3.13), temos $\tilde{f}_{k_f-1}(x^{k_f}) \leq \tilde{f}_{k_f-1}(x)$ para todo $x \in S$. Utilizando esse fato e pela desigualdade da direita de (3.12),

$$f(x^{k_f}) \leq \tilde{f}_{k_f-1}(x) + \text{tol} \leq f(x) + \text{tol} \text{ para todo } x \in S.$$

Em particular, tome $x \in \text{argmin}_{y \in S} f(y)$. Logo, $f(x^{k_f}) \leq f_* + \text{tol}$. \square

Teorema 3.14. *Considere uma seqüência $\{x^k\}$ gerada pelo algoritmo 3.12 sob as hipóteses 3.8 e 3.9, e suponha que $k \rightarrow \infty$. Então,*

$$\lim_{k \rightarrow \infty} \tilde{f}_{k-1}(x^k) = f_* = \lim_{k \rightarrow \infty} \hat{f}_k = \liminf_{k \rightarrow \infty} f(x^k).$$

Demonstração. Como $k \rightarrow \infty$, por (3.12), temos que $\hat{f}_{k-1}(x^k) \uparrow f_* - \gamma$ com $\gamma \geq 0$. Em outras palavras, para todo k ,

$$f_* - \gamma \geq \tilde{f}_{k-1}(x^k). \quad (3.15)$$

Suponha, por contradição, que $\gamma > 0$. Como $\{x^k\} \subset S$ é limitada (pois S é compacto), existe uma subsequência $\{x^{k_i}\}$ convergente. Lembrando a hipótese 3.8, para um índice k_i suficientemente grande, considere que $\|v^{k(i-1)}\| \leq M$ e que $\|x^{k(i-1)} - x^{k_i}\| \leq \gamma/(2M)$. Assim,

$$f_* - \gamma \geq \tilde{f}_{k_i-1}(x^{k_i}) \geq \tilde{f}_{k(i-1)}(x^{k_i}) \geq f(x^{k(i-1)}) + \langle v^{k(i-1)}, x^{k_i} - x^{k(i-1)} \rangle,$$

onde as desigualdades acima são obtidas respectivamente por (3.15), pelo fato de que $k_i - 1 \geq k(i-1)$ juntamente por (3.12), e por (3.11). Utilizando ainda a desigualdade de Cauchy-Schwarz, temos:

$$f_* - \gamma \geq f(x^{k(i-1)}) - M\|x^{k_i} - x^{k(i-1)}\| \geq f_* - M(\gamma/(2M)) = f_* - \gamma/2,$$

o que é uma contradição. Logo $\gamma = 0$ conforme queríamos. Considere agora a seqüência decrescente $\{\hat{f}_k\}$. Suponha, novamente por contradição, que existe uma constante $\xi > 0$ tal que $\hat{f}_k \downarrow f_* + \xi$. Já vimos que existe uma subsequência $\{x^{k_i}\}$ convergente de $\{x^k\}$. Tomando M como uma constante Lipschitz de f e \hat{f}_k , seja k_i suficientemente grande tal que $\|x^{k(i+1)} - x^{k_i}\| \leq \xi/(2M)$. Assim, pela definição de \hat{f}_{k_i} e por (3.11),

$$f_* + \xi \leq \hat{f}_{k_i} \leq f(x^{k_i}) = \tilde{f}_{k_i}(x^{k_i}) = \tilde{f}_{k_i}(x^{k_i}) - \tilde{f}_{k_i}(x^{k(i+1)}) + \tilde{f}_{k_i}(x^{k(i+1)}).$$

Sabendo que $k_i \leq k_{i+1} - 1$ e por (3.12),

$$\begin{aligned} f_* + \xi &\leq \tilde{f}_{k_i}(x^{k_i}) - \tilde{f}_{k_i}(x^{k(i+1)}) + \tilde{f}_{k(i+1)-1}(x^{k(i+1)}) \\ &\leq \tilde{f}_{k_i}(x^{k_i}) - \tilde{f}_{k_i}(x^{k(i+1)}) + f_* \\ &\leq M\|x^{k(i+1)} - x^{k_i}\| + f_*, \end{aligned}$$

onde a segunda desigualdade é dada por (3.15) (note que já provamos que $\gamma = 0$) e a última pelo fato de que f é Lipschitz contínua. Com k_i suficientemente grande, temos, então:

$$f_* + \xi \leq M(\xi/(2L)) + f_* = f_* + \xi/2,$$

o que é uma contradição. Logo, $\xi = 0$ conforme queríamos. \square

Capítulo 4

Método de Feixe

O *método de feixe* reconcilia as características de métodos de descida e de planos de corte, garantindo, ao mesmo tempo, o decréscimo da função objetivo e a estabilização. Baseia-se não apenas na construção de um modelo \tilde{f}_k da função objetivo, mas também na manutenção de uma aproximação poliedral de seu ε_k -subdiferencial $\partial_{\varepsilon_k} \tilde{f}_k(\cdot)$.

O modelo \tilde{f}_k pode ser construído de modo análogo ao modelo utilizado no método de planos de corte. No entanto, tenta-se agora evitar um acúmulo grande de funções afins que causam mau condicionamento. Isso é feito considerando-se duas seqüências distintas de pontos: os *candidatos* $\{x^k\}$, que são os iterados do algoritmo, e os *centros de estabilização* $\{\hat{x}^k\}$, que decrescem de fato a função objetivo. Uma definição mais precisa é dada abaixo.

Definição 4.1. *Um iterado (ponto candidato) x^{k+1} torna-se um centro de estabilização (i.e., $\hat{x}^{k+1} \doteq x^{k+1}$) somente se uma condição do tipo Armijo é satisfeita, ou seja, se*

$$f(\hat{x}^k) - f(x^{k+1}) \geq m\delta_{k+1}, \quad (4.1)$$

onde δ_{k+1} é o decréscimo nominal calculado na iteração k e $m \in]0, 1[$. Nesse caso, o passo realizado é denominado sério (ou de descida). Do contrário, $\hat{x}^{k+1} = \hat{x}^k$ e o passo é denominado nulo.

Note que com isso o conjunto $\{\hat{x}^1, \dots, \hat{x}^k\}$ está contido em $\{x^1, \dots, x^k\}$ para todo k . Assim como nos planos de corte, \tilde{f}_k é uma aproximação da função objetivo, utilizada para computar o decréscimo nominal δ_{k+1} , definido em (3.14), e o ponto candidato x^{k+1} . Como instabilidades ocorrem quando o movimento a partir de \hat{x}^k é muito grande, consideramos o decréscimo do modelo \tilde{f}_k não em um conjunto S fixo, mas ao redor de uma

bola de centro \hat{x}^k . Assim, ao invés de usarmos problemas do tipo (3.13), utilizamos o seguinte problema estabilizado:

$$x^{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \tilde{f}_k(x) + \frac{\mu_k}{2} \|x - \hat{x}^k\|^2, \quad (4.2)$$

onde $\mu_k > 0$ é um parâmetro, cuja atualização será vista na seção 4.2. Conforme (3.14) e (4.2), o respectivo decréscimo nominal será, portanto,

$$\delta_{k+1} = f(\hat{x}^k) - \left(\tilde{f}_k(x^{k+1}) + \frac{\mu_k}{2} \|x^{k+1} - \hat{x}^k\|^2 \right). \quad (4.3)$$

Com as definições acima, já é possível verificarmos a descrição do método.

Algoritmo 4.2. *Método de Feixe*

1. Seja $\text{tol} \geq 0$ uma tolerância dada e $m \in]0, 1[$.
2. Tome um ponto inicial x^1 e sejam $k = 1$ e $\delta_1 \doteq \infty$.
3. (Caixa preta) Obtenha v^1 e $f(x^1)$.
4. (Construção do modelo) Construa o modelo \tilde{f}_1 .
5. (Critério de parada) Se $\delta_k \leq \text{tol}$, pare.
6. (Ponto candidato) Resolva

$$x^{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \tilde{f}_k(x) + \frac{\mu_k}{2} \|x - \hat{x}^k\|^2.$$

7. (Decréscimo nominal) Defina

$$\delta_{k+1} \doteq f(\hat{x}^k) - \left(\tilde{f}_k(x^{k+1}) + \frac{\mu_k}{2} \|x^{k+1} - \hat{x}^k\|^2 \right).$$

8. (Teste de descida) Se $f(\hat{x}^k) - f(x^{k+1}) \geq m\delta_{k+1}$:
Então faça $\hat{x}^{k+1} \doteq x^{k+1}$ e atualize μ_{k+1} (*passo de descida*).
Senão, faça $\hat{x}^{k+1} \doteq \hat{x}^k$ (*passo nulo*).
9. (Atualização do modelo) Construa \tilde{f}_{k+1} adicionando x^{k+1} ao modelo.
10. (Loop) $k = k + 1$; e vá para o passo 5.

Analisemos agora o problema estabilizado (4.2). Tal problema é de programação quadrática convexa, usualmente resolvido através do seu dual, que possui um conjunto viável com uma estrutura mais simples. Para mostrarmos esse problema dual, convém antes considerar o modelo \tilde{f}_k referindo-o ao centro de estabilização. Para isso, considere as seguintes definições:

Definição 4.3. Dado uma iteração k do método de feixes, o conjunto das informações obtidas, isto é,

$$\left\{ (x^i, f(x^i), v^i) : v^i \in \partial f(x^i), i = 1, \dots, k \right\}$$

é denominado feixe.

Definição 4.4. Considere uma iteração k do método. Para cada índice i do feixe, o erro de linearização é dado por

$$e_i^k \doteq f(\hat{x}^k) - f(x^i) - \langle v^i, \hat{x}^k - x^i \rangle.$$

Usaremos a notação e_i ao invés de e_i^k quando não houver ambigüidades.

As definições acima nos permitem substituir a caracterização do modelo \tilde{f}_k dada em (3.11), pela seguinte forma:

$$\tilde{f}_k(x) = f(\hat{x}^k) + \max_{i=1, \dots, k} \{ -e_i + \langle v^i, x - \hat{x}^k \rangle \}. \quad (4.4)$$

Note também que, pela definição de ε -subdiferencial, $v^i \in \partial_{e_i} f(\hat{x}^k)$. Com essa notação, podemos caracterizar o feixe da seguinte forma:

Proposição 4.5. Dado uma iteração k do método, o feixe é o conjunto formado por

$$\left\{ (v^i, e_i) : v^i \in \partial_{e_i} f(\hat{x}^k), i = 1, \dots, k \right\} \cup \left\{ (\hat{x}^k, f(\hat{x}^k), v^k) \right\}.$$

Demonstração. Segue diretamente das definições 4.3 e 4.4. \square

Conforme mencionamos na seção 3.4, um número crescente de funções afins no modelo dificulta a resolução do problema. Ao contrário do método de planos de corte, o método de feixe possibilita o controle da quantidade de aproximações afins usadas por ele. Isso é feito com a diminuição do tamanho do feixe, de maneira que ele guarde apenas $p_k \leq k$ elementos. O modelo \tilde{f}_k não será dado, portanto por (4.4), mas por

$$\tilde{f}_k(x) = f(\hat{x}^k) + \max_{i=1, \dots, p_k} \{ -e_i + \langle v^i, x - \hat{x}^k \rangle \}, \quad (4.5)$$

onde os índices são reordenados de forma apropriada, se necessário. Veremos na seção 4.1 como selecionar os índices que permanecem no feixe. Podemos agora descrever o dual do problema estabilizado (4.2).

Proposição 4.6. *Seja x^{k+1} a solução única do problema (4.2) e suponha que $\mu_k > 0$. Então,*

$$x^{k+1} = \hat{x}^k - \frac{1}{\mu_k} \hat{v}^k \quad \text{onde} \quad \hat{v}^k \doteq \sum_{i=1}^{p_k} \bar{\alpha}_i v^i \quad (4.6)$$

e $\bar{\alpha} \in \mathbb{R}^{p_k}$ é uma solução do problema

$$\min_{\alpha \in \Delta_{p_k}} \frac{1}{2} \left\| \sum_{i=1}^{p_k} \alpha_i v^i \right\|^2 + \mu_k \sum_{i=1}^{p_k} \alpha_i e_i. \quad (4.7)$$

Demonstração. De (4.5), temos que $\tilde{f}_k(x) \geq f(\hat{x}^k) - e_i + \langle v^i, x - \hat{x}^k \rangle$ para todo $i = 1, \dots, p_k$. Podemos então escrever o problema (4.2) com uma variável extra y da seguinte forma:

$$\begin{aligned} \min_{(x,y) \in \mathbb{R}^n \times \mathbb{R}} \quad & y + \frac{\mu_k}{2} \|x - \hat{x}^k\|^2 \\ \text{sujeito a} \quad & y \geq f(\hat{x}^k) - e_i + \langle v^i, x - \hat{x}^k \rangle, \quad i = 1, \dots, p_k. \end{aligned} \quad (4.8)$$

A função lagrangeana correspondente é dada por

$$\begin{aligned} \mathcal{L}(x, y, \alpha) & \doteq y + \frac{\mu_k}{2} \|x - \hat{x}^k\|^2 + \sum_{i=1}^{p_k} \alpha_i (f(\hat{x}^k) - e_i + \langle v^i, x - \hat{x}^k \rangle - y) \\ & = \left(1 - \sum_{i=1}^{p_k} \alpha_i\right) y + \frac{\mu_k}{2} \|x - \hat{x}^k\|^2 + \sum_{i=1}^{p_k} \alpha_i (f(\hat{x}^k) - e_i + \langle v^i, x - \hat{x}^k \rangle) \end{aligned}$$

Com isso, podemos escrever o problema (4.2) da seguinte forma:

$$\min_{(x,y) \in \mathbb{R}^n \times \mathbb{R}} \max_{\alpha \geq 0} \mathcal{L}(x, y, \alpha). \quad (4.9)$$

Pela convexidade forte, o problema acima tem uma única solução x^{k+1} , com $\bar{\alpha}$ como seu multiplicador correspondente. Como o problema em questão é convexo e a condição de Slater é válida, a folga de dualidade é nula. Logo,

$$\min_{(x,y) \in \mathbb{R}^n \times \mathbb{R}} \max_{\alpha \geq 0} \mathcal{L}(x, y, \alpha) = \max_{\alpha \geq 0} \min_{(x,y) \in \mathbb{R}^n \times \mathbb{R}} \mathcal{L}(x, y, \alpha),$$

sendo que o problema da direita é o dual de (4.9) (ou, equivalentemente, de (4.2)). Observe que o problema dual pode assumir valor infinito devido ao termo $(1 - \sum_{i=1}^{p_k} \alpha_i)y$ do lagrangeano: basta que $(1 - \sum_{i=1}^{p_k} \alpha_i) < 0$ e

que $y \rightarrow \infty$. Para evitar isso, é necessário que $\sum_{i=1}^{p_k} \alpha_i = 1$. Além disso, como $\alpha_i \geq 0$ para todo $i = 1, \dots, p_k$, temos que $\alpha \in \Delta_{p_k}$. Com isso, podemos escrever a função lagrangeana da seguinte forma:

$$L(x, \alpha) \doteq f(\hat{x}^k) + \frac{\mu_k}{2} \|x - \hat{x}^k\|^2 + \sum_{i=1}^{p_k} \alpha_i (-e_i + \langle v^i, x - \hat{x}^k \rangle) \quad (4.10)$$

e podemos dizer que x^{k+1} e $\bar{\alpha}$ resolvem os problemas primal e dual abaixo:

$$\min_{x \in \mathbb{R}^n} \max_{\alpha \in \Delta_{p_k}} L(x, \alpha) = \max_{\alpha \in \Delta_{p_k}} \min_{x \in \mathbb{R}^n} L(x, \alpha).$$

Considere agora a seguinte definição:

$$\theta(\alpha) \doteq \operatorname{argmin}_{x \in \mathbb{R}^n} L(x, \alpha).$$

Fixando $\alpha \in \Delta_{p_k}$, observamos que a condição de otimalidade que define θ é dada por

$$\begin{aligned} \nabla_x L(\theta(\alpha), \alpha) &= 0 \\ \Leftrightarrow \mu_k(\theta(\alpha) - \hat{x}^k) + \sum_{i=1}^{p_k} \alpha_i v^i &= 0 \end{aligned} \quad (4.11)$$

Em particular, para $\alpha \doteq \bar{\alpha}$ (note que $\theta(\bar{\alpha}) = x^{k+1}$),

$$\mu_k(x^{k+1} - \hat{x}^k) + \sum_{i=1}^{p_k} \bar{\alpha}_i v^i = 0 \Leftrightarrow x^{k+1} = \hat{x}^k - \frac{1}{\mu_k} \hat{v}^k,$$

conforme queríamos. Mostremos agora que $\bar{\alpha}$ resolve (4.7). Multiplicando a igualdade (4.11) por $\theta(\alpha) - \hat{x}^k$, temos:

$$\mu_k \|\theta(\alpha) - \hat{x}^k\|^2 + \sum_{i=1}^{p_k} \alpha_i \langle v^i, \theta(\alpha) - \hat{x}^k \rangle = 0 \quad (4.12)$$

E multiplicando essa mesma igualdade (4.11) por $1/\mu_k \sum_{i=1}^{p_k} \alpha_i v^i$, temos:

$$\sum_{i=1}^{p_k} \alpha_i \langle v^i, \theta(\alpha) - \hat{x}^k \rangle + \frac{1}{\mu_k} \left\| \sum_{i=1}^{p_k} \alpha_i v^i \right\|^2 = 0 \quad (4.13)$$

Pela definição (4.10), com $x \doteq \theta(\alpha)$,

$$\begin{aligned}
L(\theta(\alpha), \alpha) &= f(\hat{x}^k) + \frac{\mu_k}{2} \|\theta(\alpha) - \hat{x}^k\|^2 + \sum_{i=1}^{p_k} \alpha_i (-e_i + \langle v^i, \theta(\alpha) - \hat{x}^k \rangle) \\
&= f(\hat{x}^k) + \frac{\mu_k}{2} \|\theta(\alpha) - \hat{x}^k\|^2 + \sum_{i=1}^{p_k} \alpha_i \langle v^i, \theta(\alpha) - \hat{x}^k \rangle - \sum_{i=1}^{p_k} \alpha_i e_i \\
&= f(\hat{x}^k) + \frac{1}{2} \sum_{i=1}^{p_k} \alpha_i \langle v^i, \theta(\alpha) - \hat{x}^k \rangle - \sum_{i=1}^{p_k} \alpha_i e_i \\
&= f(\hat{x}^k) - \frac{1}{2\mu_k} \left\| \sum_{i=1}^{p_k} \alpha_i v^i \right\|^2 - \sum_{i=1}^{p_k} \alpha_i e_i
\end{aligned}$$

sendo que a terceira igualdade é dada por (4.12) e a última por (4.13). Como

$$\bar{\alpha} \in \operatorname{argmax}_{\alpha \in \Delta_{p_k}} \min_{x \in \mathbb{R}^n} L(x, \alpha) = \operatorname{argmax}_{\alpha \in \Delta_{p_k}} L(\theta(\alpha), \alpha),$$

concluimos que $\bar{\alpha}$ é solução do problema

$$\max_{\alpha \in \Delta_{p_k}} f(\hat{x}^k) - \frac{1}{2\mu_k} \left\| \sum_{i=1}^{p_k} \alpha_i v^i \right\|^2 + \sum_{i=1}^{p_k} \alpha_i e_i \quad (4.14)$$

o que é equivalente a (4.7). \square

Seguem agora alguns resultados associados ao problema estabilizado.

Proposição 4.7. *Considere o problema (4.2) e o dual (4.7), e seja $\mu_k > 0$. Sejam x^{k+1} e $\bar{\alpha}$ as soluções ótimas de (4.2) e (4.7), respectivamente. Então, as seguintes afirmações são verdadeiras:*

- (a) $\delta_{k+1} = \varepsilon_k + \frac{1}{2\mu_k} \|\hat{v}^k\|^2$, onde $\varepsilon_k \doteq \sum_{i=1}^{p_k} \bar{\alpha}_i e_i$,
- (b) $\hat{v}^k \in \partial \tilde{f}_k(x^{k+1})$,
- (c) $\varepsilon_k = f(\hat{x}^k) - \tilde{f}_k(x^{k+1}) - \frac{1}{\mu_k} \|\hat{v}^k\|^2$,
- (d) $\hat{v}^k \in \partial_{\varepsilon_k} f(\hat{x}^k)$.

Demonstração.

(a) Como a folga de dualidade é nula, os valores ótimos de (4.2) e (4.14) são iguais, ou seja,

$$\tilde{f}_k(x^{k+1}) + \frac{\mu_k}{2} \|x^{k+1} - \hat{x}^k\|^2 = f(\hat{x}^k) - \frac{1}{2\mu_k} \left\| \sum_{i=1}^{p_k} \bar{\alpha}_i v^i \right\|^2 - \sum_{i=1}^{p_k} \bar{\alpha}_i e_i.$$

O resultado segue diretamente de (4.3), da definição de ε_k acima, e de \hat{v}^k , dado em (4.6).

(b) A condição de otimalidade de (4.2) é dada por

$$0 \in \partial \tilde{f}_k(x^{k+1}) + \mu_k(x^{k+1} - \hat{x}^k) \Leftrightarrow 0 \in \partial \tilde{f}_k(x^{k+1}) - \hat{v}^k$$

sendo que a última sentença é válida por (4.6).

(c) Por (4.6), temos:

$$x^{k+1} - \hat{x}^k = -\frac{1}{\mu_k} \hat{v}^k \Rightarrow \|x^{k+1} - \hat{x}^k\| = \frac{\|\hat{v}^k\|}{\mu_k}. \quad (4.15)$$

Pelo item (a) desta proposição, por (4.3) e depois por (4.15),

$$\begin{aligned} \varepsilon_k &= \delta_{k+1} - \frac{1}{2\mu_k} \|\hat{v}^k\|^2 \\ &= f(\hat{x}^k) - \tilde{f}_k(x^{k+1}) - \frac{\mu_k}{2} \|x^{k+1} - \hat{x}^k\|^2 - \frac{1}{2\mu_k} \|\hat{v}^k\|^2 \\ &= f(\hat{x}^k) - \tilde{f}_k(x^{k+1}) - \frac{1}{\mu_k} \|\hat{v}^k\|^2, \end{aligned}$$

conforme queríamos.

(d) Por (3.12), sabemos que $f \geq \tilde{f}_k$. Pelo item (b) desta proposição, temos,

$$\begin{aligned} f(x) &\geq \tilde{f}_k(x) \\ &\geq \tilde{f}_k(x^{k+1}) + \langle \hat{v}^k, x - x^{k+1} \rangle \\ &= \tilde{f}_k(x^{k+1}) + \langle \hat{v}^k, x - \hat{x}^k \rangle - \langle \hat{v}^k, x^{k+1} - \hat{x}^k \rangle. \end{aligned}$$

Além disso, de (4.6),

$$\begin{aligned} f(x) &\geq \tilde{f}_k(x^{k+1}) + \langle \hat{v}^k, x - \hat{x}^k \rangle + \frac{1}{\mu_k} \|\hat{v}^k\|^2 \\ &= f(\hat{x}^k) + \langle \hat{v}^k, x - \hat{x}^k \rangle - \left(f(\hat{x}^k) - \tilde{f}_k(x^{k+1}) - \frac{1}{\mu_k} \|\hat{v}^k\|^2 \right) \end{aligned}$$

Assim, pelo item (c) desta mesma proposição,

$$f(x) \geq f(\hat{x}^k) + \langle \hat{v}^k, x - \hat{x}^k \rangle - \varepsilon_k,$$

e o resultado segue da definição de ε_k -subgradiente. \square

Uma atualização no algoritmo 4.2 pode ser realizada quando analisamos a nova caracterização do decréscimo nominal, dada no item (a) da proposição acima. Pelo item (d) da mesma proposição, e lembrando que o ε -subdiferencial é fechado (conforme o lema 2.30), se $\{\varepsilon_k\}$ e $\{\hat{v}^k\}$ tendem ambos a zero, então $\{\hat{x}^k\}$ deve tender ao mínimo. Assim, o critério de parada, dado no passo 5 do algoritmo 4.2, pode ser melhorado para:

$$\varepsilon_k \leq \text{tol}_\varepsilon \text{ e } \|\hat{v}^k\| \leq \text{tol}_v,$$

com tol_ε e tol_v duas tolerâncias dadas.

4.1 Agregação

Veremos agora uma interpretação do vetor \hat{v}^k , definido em (4.6), relacionando-o com uma técnica de limpeza do modelo denominada *agregação*. Este mecanismo, responsável por diminuir o número de elementos do feixe (ou, equivalentemente, o número de funções afins que definem o modelo \tilde{f}_k), divide-se em duas fases: (i) a seleção de pares (v^i, e_i) do feixe que devem ser descartados, e (ii) o agrupamento de tais pares descartados em um único par, precisamente $(\hat{v}^k, \varepsilon_k)$.

Esse par $(\hat{v}^k, \varepsilon_k)$ conterà, então, um resumo de todas as informações essenciais dos descartados que foram acumuladas até a iteração k . A função afim correspondente a ele, denominada *linearização agregada* é dada por:

$$f_a(x) \doteq f(\hat{x}^k) - \varepsilon_k + \langle \hat{v}^k, x - \hat{x}^k \rangle. \quad (4.16)$$

Note que, embora cada par (v^i, e_i) do feixe original satisfaça $v^i \in \partial f(x^i)$, não há um ponto conhecido x^i para o qual $\hat{v}^k \in \partial f(x^i)$.

Proposição 4.8. *Considere a função f_a definida em (4.16). Então,*

$$(a) \quad f_a(x) = \tilde{f}_k(x^{k+1}) + \langle \hat{v}^k, x - x^{k+1} \rangle \text{ para todo } x \in \mathbb{R}^n.$$

$$(b) \quad f_a(x) \leq \tilde{f}_{k+1}(x) \text{ para todo } x \in \mathbb{R}^n.$$

Demonstração. (a) Pelo item (c) da proposição 4.7,

$$\begin{aligned}
 f_a(x) &= f(\hat{x}^k) - \varepsilon_k + \langle \hat{v}^k, x - \hat{x}^k \rangle \\
 &= \tilde{f}_k(x^{k+1}) + \langle \hat{v}^k, x - \hat{x}^k \rangle + \frac{1}{\mu_k} \|\hat{v}^k\|^2 \\
 &= \tilde{f}_k(x^{k+1}) + \left\langle \hat{v}^k, x - \hat{x}^k + \frac{1}{\mu_k} \hat{v}^k \right\rangle \\
 &= \tilde{f}_k(x^{k+1}) + \langle \hat{v}^k, x - x^{k+1} \rangle,
 \end{aligned}$$

sendo que a última igualdade é dada por (4.6).

(b) Pelo item (b) da proposição 4.7,

$$\tilde{f}_k(x) \geq \tilde{f}_k(x^{k+1}) + \langle \hat{v}^k, x - x^{k+1} \rangle \text{ para todo } x \in \mathbb{R}^n.$$

Com isto, e pelo item (a) desta proposição, temos que $f_a(x) \leq \tilde{f}_k(x)$ para todo $x \in \mathbb{R}^n$. O resultado segue do fato de que $\tilde{f}_k(x) \leq \tilde{f}_{k+1}(x)$ para todo $x \in \mathbb{R}^n$. \square

Existem diversas possibilidades para a escolha dos pares (v^i, e_i) a serem descartados do feixe. Podemos escolher, por exemplo, os mais antigos, pois estes tendem a ser menos importantes para o modelo. Também é possível escolhermos todos os elementos do feixe, conforme veremos na seção 6.1. É claro, no entanto, que escolhas distintas de pares resultam em velocidades de convergência diferentes. Considere agora o conjunto de índices

$$I \doteq \{i : (v^i, e_i) \text{ não foi descartado do feixe}\}.$$

Então, o modelo pode ser reescrito naturalmente como

$$\tilde{f}_{k+1}(x) = \max \left\{ f(\hat{x}^{k+1}) + \max_{i \in I} \{ -e_i + \langle v^i, x - \hat{x}^{k+1} \rangle \}, f_a(x) \right\}. \quad (4.17)$$

Observe que tal modelo continua sendo uma aproximação da função objetivo que não permite cortes em sua epígrafe. Finalizamos essa seção incorporando no passo 9 do algoritmo 4.2 a realização da agregação. No algoritmo abaixo, deve-se considerar que na iteração k há p_k elementos (v^i, e_i) no feixe. Ademais, definimos p_{\max} como o tamanho máximo do feixe permitido.

Algoritmo 4.9. *Agregação no Método de Feixe*

a) Sejam $p_{\max} > 0$ e $v^{p(k+1)} \doteq v(x^{k+1})$.

b) (Teste de compressão) Se $p_k \geq p_{\max}$, então:

1. Selecione dois ou mais pares (v^i, e_i) para descartar.
2. Agrupe tais pares em $(v^{p_k}, e_{p_k}) \doteq (\hat{v}^k, \varepsilon_k)$.

c) (Atualização do feixe) Agrupe $(v^{p(k+1)}, e_{p_{k+1}})$ no feixe com

$$e_{p_{k+1}} \doteq \begin{cases} 0, & \text{se o passo é sério,} \\ f(\hat{x}^k) - (f(x^{k+1}) + \langle v^{k+1}, \hat{x}^k - x^{k+1} \rangle), & \text{se o passo é nulo.} \end{cases}$$

d) (Atualização dos erros de linearização) Se o passo é sério, faça:

$$e_i \doteq e_i + f(x^{k+1}) - f(\hat{x}^k) - \langle v^i, x^{k+1} - \hat{x}^k \rangle, \quad i = 1, \dots, p_k.$$

e) Mude de p_k para $p_{(k+1)}$ e construa \tilde{f}_{k+1} conforme (4.17).

4.2 Atualização do Parâmetro de Penalização

Nesta seção veremos como atualizar o parâmetro μ_k do problema estabelecido (4.2). Para isso, considere a *regularização de Moreau-Yosida* de f em um ponto $x \in \mathbb{R}^n$, isto é,

$$F(x) \doteq \min_{z \in \mathbb{R}^n} f(z) + \frac{\mu}{2} \|z - x\|^2, \quad (4.18)$$

sendo que $\mu > 0$. Essa regularização possui propriedades interessantes [HUL93b, teorema XV.4.1.4]: é convexa, finita em todo lugar e diferenciável, com gradiente

$$\nabla F(x) = \mu(x - p(x)) \quad (4.19)$$

sendo que $p(x)$ é o *ponto proximal* de x , ou seja,

$$p(x) \doteq \operatorname{argmin}_{z \in \mathbb{R}^n} f(z) + \frac{\mu}{2} \|z - x\|^2.$$

Além disso, $\nabla F(\cdot)$ é globalmente Lipschitz contínua. Observe que $p(x)$ é o único minimizador, e que pela condição de otimalidade do problema (4.18),

$$0 \in \partial f(p(x)) + \mu(p(x) - x) \Rightarrow p(x) = x - \frac{1}{\mu}v, \quad (4.20)$$

onde $v \in \partial f(p(x))$. Logo, $\nabla F(x) = \mu(x - p(x)) \in \partial f(p(x))$. Sabe-se, ainda, que (3.2) é equivalente ao problema

$$\text{minimizar } F(x) \quad \text{sujeito a } x \in \mathbb{R}^n, \quad (4.21)$$

no sentido de que um minimizador de (3.2) também é de (4.21) [HUL93b, teorema XV.4.1.7].

Como a regularização de Moreau-Yosida é diferenciável e possui gradiente Lipschitz contínuo, é razoável o uso de informação de segunda ordem para a sua minimização. Considere então o *método quase-Newton* aplicado em (4.21), isto é,

$$\hat{x}^{k+1} = \hat{x}^k - M_k^{-1} \nabla F(\hat{x}^k), \quad (4.22)$$

onde M_k é uma matriz simétrica definida positiva atualizada a cada iteração. Restringindo M_k a um múltiplo da identidade, isto é, tomando $M_k \doteq \mu_k I$, com $\mu_k > 0$, e considerando a equação secante dada por

$$M_{k+1}(\hat{x}^{k+1} - \hat{x}^k) = \nabla F(\hat{x}^{k+1}) - \nabla F(\hat{x}^k),$$

temos a atualização de μ_k feita pela forma de quadrados mínimos, pois com tal M_k , a equação acima pode não ser obedecida com exatidão. Assim,

$$\mu_{k+1} \doteq \operatorname{argmin}_{\mu > 0} \frac{1}{2} \left\| \frac{1}{\mu} (\nabla F(\hat{x}^{k+1}) - \nabla F(\hat{x}^k)) - (\hat{x}^{k+1} - \hat{x}^k) \right\|^2. \quad (4.23)$$

Pela condição de otimalidade de (4.23),

$$\left\langle \frac{1}{\mu_{k+1}} (\nabla F(\hat{x}^{k+1}) - \nabla F(\hat{x}^k)) - (\hat{x}^{k+1} - \hat{x}^k), (\nabla F(\hat{x}^{k+1}) - \nabla F(\hat{x}^k)) \right\rangle = 0.$$

Portanto,

$$\mu_{k+1} = \frac{\|\nabla F(\hat{x}^{k+1}) - \nabla F(\hat{x}^k)\|^2}{\langle \nabla F(\hat{x}^{k+1}) - \nabla F(\hat{x}^k), \hat{x}^{k+1} - \hat{x}^k \rangle}. \quad (4.24)$$

Considere o método de feixe utilizando o problema estabilizado (4.2) em uma iteração cujo passo é sério. Conforme vimos em (4.6), isso pode ser interpretado como uma maneira de computar o ponto proximal, isto é,

$$p(\hat{x}^k) \approx \hat{x}^{k+1} = x^{k+1} = \hat{x}^k - \frac{1}{\mu_k} \hat{v}^k,$$

sendo que

$$\hat{v}^k \approx \nabla \tilde{F}_k(\hat{x}^k) = \mu_k(\hat{x}^k - \hat{x}^{k+1}) \in \partial \tilde{f}_k(\hat{x}^{k+1}),$$

onde \tilde{F}_k é a regularização de Moreau-Yosida do modelo \tilde{f}_k . O passo acima, por sua vez, pode ser visto como um passo do tipo (4.22), com $M_k \doteq \mu_k I$. Assim, a atualização do parâmetro de penalização μ_k do método de feixe pode ser dada exatamente por (4.24).

No entanto, por (4.19), para computar $\nabla F(\hat{x}^{k+1})$ necessitamos calcular primeiro $p(\hat{x}^{k+1})$. Ao invés disso, considere uma situação oposta, isto é, encontrar z^{k+1} tal que $\hat{x}^{k+1} = p(z^{k+1})$ [LS97]. Para isso, tome em (4.20), $x \doteq z^{k+1}$, $p(x) = \hat{x}^{k+1}$, $v \doteq \hat{v}^{k+1}$ e $\mu \doteq \mu_k$. Então,

$$\hat{x}^{k+1} = z^{k+1} - \frac{1}{\mu_k} \hat{v}^{k+1} \Leftrightarrow z^{k+1} = \hat{x}^{k+1} + \frac{1}{\mu_k} \hat{v}^{k+1},$$

com $\hat{v}^{k+1} = \mu_k(z^{k+1} - \hat{x}^{k+1}) = \nabla F(z^{k+1})$. Logo, substituindo tais termos em (4.24), temos:

$$\begin{aligned} \mu_{k+1} &= \frac{\|\hat{v}^{k+1} - \hat{v}^k\|^2}{\langle \hat{v}^{k+1} - \hat{v}^k, \hat{x}^{k+1} - \hat{x}^k + 1/\mu_k(\hat{v}^{k+1} - \hat{v}^k) \rangle} \\ \Leftrightarrow \frac{1}{\mu_{k+1}} &= \frac{1}{\mu_k} + \frac{\langle \hat{v}^{k+1} - \hat{v}^k, \hat{x}^{k+1} - \hat{x}^k \rangle}{\|\hat{v}^{k+1} - \hat{v}^k\|^2}. \end{aligned}$$

Essa última igualdade será, portanto, utilizada no passo 8 do algoritmo 4.2.

4.3 Análise de Convergência

Terminamos o assunto enunciando a convergência do método de feixe que utiliza o dual do problema estabilizado (4.2) e uma tolerância $\text{tol} = 0$ (ou, equivalentemente, $\text{tol}_\varepsilon = \text{tol}_v = 0$). Considere

$$K_S \doteq \{k: \text{ o passo dado na iteração } k \text{ é sério} \}.$$

Há duas possibilidades para a seqüência de passos sérios $\{\hat{x}^k\}_{k \in K_S}$, conforme veremos nos teoremas 4.11 e 4.12. Antes, definimos $f_* \doteq \lim_{k \in K_S} f(\hat{x}^k)$ e consideramos o seguinte resultado:

Lema 4.10. *Considere o algoritmo 4.2 e suponha que $f_* > -\infty$. Então,*

$$\sum_{k \in K_S} \delta_k \leq \frac{f(\hat{x}^1) - f_*}{m}.$$

Demonstração. Tome um $k \in K_S$ qualquer. Como $\hat{x}^{k+1} = x^{k+1}$, temos

$$f(\hat{x}^k) - f(\hat{x}^{k+1}) = f(\hat{x}^k) - f(x^{k+1}) \geq m\delta_{k+1}. \quad (4.25)$$

Seja ℓ o índice sucessor de k em K_S . Como entre as iterações k e ℓ ocorrem apenas passos nulos, temos que $\hat{x}^{k+1} = \hat{x}^{k+i}$, para todo $i = 2, \dots, \ell - k$. Em particular, $\hat{x}^{k+1} = \hat{x}^\ell$. Com tal fato e por (4.25) (note que $\ell \in K_S$), temos

$$f(\hat{x}^{k+1}) - f(\hat{x}^{\ell+1}) = f(\hat{x}^\ell) - f(\hat{x}^{\ell+1}) \geq m\delta_{\ell+1}.$$

Portanto,

$$\sum_{k \in K_S} (f(\hat{x}^k) - f(\hat{x}^{k+1})) \geq m \sum_{k \in K_S} \delta_k \Leftrightarrow f(\hat{x}^1) - f_* \geq m \sum_{k \in K_S} \delta_k,$$

conforme queríamos. \square

Teorema 4.11. *Considere o algoritmo 4.2 com $\text{tol} = 0$ e a realização da agregação dada no algoritmo 4.9. Suponha que $f_* > -\infty$ e que o método gera infinitos passos de descida \hat{x}^k .*

- (a) Se $\sum_{k \in K_S} \frac{1}{\mu_k} = +\infty$, então $\{\hat{x}^k\}$ é uma seqüência minimizante.
- (b) Se $\{1/\mu_k\}$ é limitada superiormente em K_S e se a hipótese 3.9 é válida, então a seqüência $\{\hat{x}^k\}$ é limitada.
- (c) Se as condições de (a) e (b) são simultaneamente satisfeitas, então a seqüência $\{\hat{x}^k\}$ tem um ponto de acumulação que é ótimo.

Demonstração.

(a) Seja $k \in K_S$. Então, para qualquer $x \in \mathbb{R}^n$,

$$\begin{aligned} \|x - \hat{x}^{k+1}\|^2 &= \|x - x^{k+1}\|^2 \\ &= \|x - \hat{x}^k + \hat{x}^k - x^{k+1}\|^2 \\ &= \|x - \hat{x}^k\|^2 + 2\langle x - \hat{x}^k, \hat{x}^k - x^{k+1} \rangle + \|\hat{x}^k - x^{k+1}\|^2 \\ &= \|x - \hat{x}^k\|^2 + \frac{2}{\mu_k} \langle \hat{v}^k, x - \hat{x}^k \rangle + \frac{\|\hat{v}^k\|^2}{\mu_k^2}, \end{aligned} \quad (4.26)$$

onde a última igualdade é dada por (4.6). Além disso, pelos itens (a) e (d) da proposição 4.7, temos:

$$\begin{aligned} f(x) &\geq f(\hat{x}^k) + \langle \hat{v}^k, x - \hat{x}^k \rangle - \varepsilon_k \\ &\geq f(\hat{x}^k) + \langle \hat{v}^k, x - \hat{x}^k \rangle - \delta_{k+1} - \frac{1}{2\mu_k} \|\hat{v}^k\|^2, \end{aligned}$$

em outras palavras,

$$\frac{2}{\mu_k} \langle \hat{v}^k, x - \hat{x}^k \rangle + \frac{\|\hat{v}^k\|^2}{\mu_k^2} \leq \frac{2}{\mu_k} (f(x) - f(\hat{x}^k) + \delta_{k+1}).$$

Substituindo esta última desigualdade em (4.26), temos:

$$\|x - \hat{x}^{k+1}\|^2 \leq \|x - \hat{x}^k\|^2 + \frac{2}{\mu_k} \left(f(x) - f(\hat{x}^k) + \delta_{k+1} \right).$$

Suponha agora, por contradição, que existem $y \in \mathbb{R}^n$ e $\eta > 0$ tais que

$$f(y) \leq f(\hat{x}^k) - \eta \quad \text{para todo } k \in K_S.$$

Então, por (4.26) com $x \doteq y$, obtemos:

$$\|y - \hat{x}^{k+1}\|^2 \leq \|y - \hat{x}^k\|^2 + \frac{2}{\mu_k} (\delta_{k+1} - \eta).$$

Considere aqui que $K_S \doteq \{k_0, k_1, \dots\}$, com $k_0 < k_1 < \dots$. Pelo lema 4.10, $\{\delta_k\}_{k \in K_S}$ tende a 0. Portanto, existe ℓ tal que

$$\|y - \hat{x}^{k_i+1}\|^2 \leq \|y - \hat{x}_i^k\|^2 - \frac{\eta}{\mu_{k_i}} \quad \text{para todo } i \geq \ell. \quad (4.27)$$

Observe que, para todo $i \geq \ell$, temos

$$\begin{aligned} 0 &\leq \sum_{\ell \leq j \leq i} \|y - \hat{x}^{k_j+1}\|^2 \leq \sum_{\ell \leq j \leq i} \|y - \hat{x}_j^k\|^2 - \eta \sum_{\ell \leq j \leq i} \frac{1}{\mu_{k_j}} \\ \Leftrightarrow 0 &\leq \|y - \hat{x}^{k_i+1}\|^2 \leq \|y - \hat{x}^{k_\ell}\|^2 - \eta \sum_{\ell \leq j \leq i} \frac{1}{\mu_{k_j}}, \end{aligned}$$

pois $\hat{x}^{k_i+1} = \hat{x}^{k_{(i+1)}}$ para qualquer $i \geq 0$. Assim, somando-se as desigualdades de (4.27) em $K_S \cap \{k : k \geq k_\ell\}$ e utilizando a hipótese deste teorema,

$$0 \leq \|y - \hat{x}^{k_\ell}\|^2 - \eta \sum_{k_\ell \leq k \in K_S} \frac{1}{\mu_k} = -\infty,$$

o que é um absurdo. Logo, $\{\hat{x}^k\}$ é, de fato, uma seqüência minimizante.

(b) Seja $k \in K_S$. Então, para qualquer $x \in \mathbb{R}^n$,

$$\begin{aligned} \|x - \hat{x}^{k+1}\|^2 &= \|x - x^{k+1}\|^2 \\ &= \|x - \hat{x}^k + \hat{x}^k - x^{k+1}\|^2 \\ &= \|x - \hat{x}^k\|^2 + 2\langle x - \hat{x}^k, \hat{x}^k - x^{k+1} \rangle + \|\hat{x}^k - x^{k+1}\|^2 \\ &= \|x - \hat{x}^k\|^2 + \frac{2}{\mu_k} \langle \hat{v}^k, x - \hat{x}^k \rangle + \frac{\|\hat{v}^k\|^2}{\mu_k^2}, \end{aligned} \quad (4.28)$$

onde a última igualdade é dada por (4.6). Além disso, pelos itens (a) e (d) da proposição 4.7, temos:

$$\begin{aligned} f(x) &\geq f(\hat{x}^k) + \langle \hat{v}^k, x - \hat{x}^k \rangle - \varepsilon_k \\ &\geq f(\hat{x}^k) + \langle \hat{v}^k, x - \hat{x}^k \rangle - \delta_{k+1} - \frac{1}{2\mu_k} \|\hat{v}^k\|^2, \end{aligned}$$

em outras palavras,

$$\frac{2}{\mu_k} \langle \hat{v}^k, x - \hat{x}^k \rangle + \frac{\|\hat{v}^k\|^2}{\mu_k^2} \leq \frac{2}{\mu_k} \left(f(x) - f(\hat{x}^k) + \delta_{k+1} \right).$$

Substituindo esta última desigualdade em (4.28), temos:

$$\|x - \hat{x}^{k+1}\|^2 \leq \|x - \hat{x}^k\|^2 + \frac{2}{\mu_k} \left(f(x) - f(\hat{x}^k) + \delta_{k+1} \right). \quad (4.29)$$

Pela hipótese 3.9, existe x^* , um minimizador de f . Considere $x \doteq x^*$ em (4.29). Então,

$$\|x^* - \hat{x}^{k+1}\|^2 \leq \|x^* - \hat{x}^k\|^2 + \frac{2\delta_{k+1}}{\mu_k} \quad \text{para todo } k \in K_S.$$

Considere agora $K_S \doteq \{k_0, k_1, \dots\}$, com $k_0 < k_1 < \dots$. Como $\hat{x}^{k_i+1} = \hat{x}^{k_{(i+1)}}$ para todo $i \geq 0$, temos, para qualquer $i \geq 0$,

$$\begin{aligned} \sum_{0 \leq j \leq i} \|x^* - \hat{x}^{k_j+1}\|^2 &\leq \sum_{0 \leq j \leq i} \|x^* - \hat{x}^{k_j}\|^2 + 2 \sum_{0 \leq j \leq i} \frac{\delta_{k_j+1}}{\mu_{k_j}} \\ \Leftrightarrow \|x^* - \hat{x}^{k_{i+1}}\|^2 &\leq \|x^* - \hat{x}^{k_0}\|^2 + 2 \sum_{0 \leq j \leq i} \frac{\delta_{k_j+1}}{\mu_{k_j}} \\ \Rightarrow \|x^* - \hat{x}^{k_{i+1}}\|^2 &\leq \|x^* - \hat{x}^{k_0}\|^2 + 2 \sum_{k \in K_S} \frac{\delta_{k+1}}{\mu_k}. \end{aligned}$$

A hipótese de que $\{1/\mu_k\}$ é limitada em K_S , juntamente com o lema 4.10, implica que a série $\sum_{k \in K_S} \delta_{k+1}/\mu_k$ é convergente. Logo, $\|x^* - \hat{x}^{k_{i+1}}\|$ é limitado para qualquer $i \geq 0$ (ou, equivalentemente, para qualquer $k_i \in K_S$). Logo, $\{\hat{x}^k\}_{k \in K_S}$ é limitada.

(c) Pela hipótese (b), $\{\hat{x}^k\}$ é limitada. Logo, existe uma subsequência $\{\hat{x}^{k_j}\}_{j \in \mathbb{N}}$ convergente a um ponto x^* . Pela hipótese (a), temos $\{f(\hat{x}^{k_j})\} \rightarrow f_*$. Em particular, $\{f(\hat{x}^{k_j})\}_{j \in \mathbb{N}} \rightarrow f_*$. Assim, pela continuidade inferior de f ,

$$f_* = \lim_{K \ni k \rightarrow \infty} f(\hat{x}^k) \geq f(x^*) \geq f_*,$$

e portanto $f(x^*) = f_*$, ou seja, x^* é ótimo. \square

Teorema 4.12. *Considere o algoritmo 4.2 com $\text{tol} = 0$ e a realização da agregação dada no algoritmo 4.9. Suponha que o método gera um último passo de descida \hat{x}^{k_f} , seguido de infinitos passos nulos $\{x^{k+1}\}_{k>k_f}$. Suponha ainda que as seguintes hipóteses são válidas:*

(a) $\{\mu_k\}_{k>k_f}$ é não decrescente.

(b) $\sum_{k>k_f} \frac{\mu_{k-1}}{\mu_k^2} = +\infty$.

Então, $\{\delta_k\}_{k>k_f}$ é não crescente e tende a zero, e \hat{x}^{k_f} minimiza f .

Demonstração. Para todo $k > k_f$, $\hat{x}^k = \hat{x}^{k_f}$. Assim, por (4.6),

$$\begin{aligned} \|x - \hat{x}^{k_f}\|^2 &= \|x - \hat{x}^k\|^2 \\ &= \left\| x - x^{k+1} - \frac{1}{\mu_k} v^k \right\|^2 \\ &= \|x - x^{k+1}\|^2 + \frac{\|v^k\|^2}{\mu_k^2} - \frac{2}{\mu_k} \langle \hat{v}^k, x - x^{k+1} \rangle \end{aligned}$$

Pela igualdade (4.15), temos:

$$\|x - \hat{x}^{k_f}\|^2 = \|x - x^{k+1}\|^2 + \|x^{k+1} - \hat{x}^k\|^2 - \frac{2}{\mu_k} \langle \hat{v}^k, x - x^{k+1} \rangle, \quad (4.30)$$

Defina agora

$$\varphi(x) \doteq f(\hat{x}^{k_f}) - \delta_{k+1} + \frac{\mu_k}{2} \|x - x^{k+1}\|^2. \quad (4.31)$$

Então, por (4.3),

$$\begin{aligned} \varphi(x) &= \tilde{f}_k(x^{k+1}) + \frac{\mu_k}{2} \|x^{k+1} - \hat{x}^k\|^2 + \frac{\mu_k}{2} \|x - x^{k+1}\|^2 \\ &= \tilde{f}_k(x^{k+1}) + \frac{\mu_k}{2} \left(\|x - \hat{x}^{k_f}\|^2 + \frac{2}{\mu_k} \langle \hat{v}^k, x - x^{k+1} \rangle \right) \\ &= \tilde{f}_k(x^{k+1}) + \frac{\mu_k}{2} \|x - \hat{x}^{k_f}\|^2 + \langle \hat{v}^k, x - x^{k+1} \rangle, \end{aligned}$$

sendo que a segunda igualdade é dada por (4.30). Tome $x \doteq x^{k+2}$ em (4.31).

Então, pelos itens (a) e (b) da proposição 4.8, temos:

$$\begin{aligned}
\varphi(x^{k+2}) &\doteq f(\hat{x}^{k_f}) - \delta_{k+1} + \frac{\mu_k}{2} \|x^{k+2} - x^{k+1}\|^2 \\
&= \tilde{f}_k(x^{k+1}) + \frac{\mu_k}{2} \|x^{k+2} - \hat{x}^{k_f}\|^2 + \langle \hat{v}^k, x^{k+2} - x^{k+1} \rangle \\
&= f_a(x^{k+2}) + \frac{\mu_k}{2} \|x^{k+2} - \hat{x}^{k_f}\|^2 \\
&\leq \tilde{f}_{k+1}(x^{k+2}) + \frac{\mu_k}{2} \|x^{k+2} - \hat{x}^{k_f}\|^2.
\end{aligned}$$

Pela primeira igualdade e última desigualdade acima,

$$f(\hat{x}^{k_f}) - \tilde{f}_{k+1}(x^{k+2}) - \frac{\mu_k}{2} \|x^{k+2} - \hat{x}^{k_f}\|^2 \leq \delta_{k+1} - \frac{\mu_k}{2} \|x^{k+2} - x^{k+1}\|^2 \quad (4.32)$$

Pela definição δ_{k+2} dada por (4.3) e como $\hat{x}^{k+1} = \hat{x}^{k_f}$ (pois $k > k_f$),

$$\begin{aligned}
\delta_{k+2} &= f(\hat{x}^{k+1}) - \tilde{f}_{k+1}(x^{k+2}) - \frac{\mu_{k+1}}{2} \|x^{k+2} - \hat{x}^{k+1}\|^2 \\
&\leq f(\hat{x}^{k_f}) - \tilde{f}_{k+1}(x^{k+2}) - \frac{\mu_k}{2} \|x^{k+2} - \hat{x}^{k_f}\|^2,
\end{aligned}$$

onde a última desigualdade é válida pois $\{\mu_k\}_{k>k_f}$ é não decrescente. Utilizando (4.32), temos

$$\delta_{k+2} \leq \delta_{k+1} - \frac{\mu_k}{2} \|x^{k+2} - x^{k+1}\|^2 \leq \delta_{k+1} \quad (4.33)$$

Como tomamos $k > k_f$ qualquer, concluímos que $\{\delta_k\}_{k>k_f}$ é não crescente.

Provemos agora que $\{x^k\}_{k>k_f}$ é limitado. Mais precisamente, mostremos que existe $r > 0$ tal que $x^k \in B(\hat{x}^{k_f}, r)$ para todo $k > k_f$. Considere $x \doteq \hat{x}^{k_f}$ em (4.31). Então, pelo item (a) da proposição 4.8,

$$\begin{aligned}
\varphi(\hat{x}^{k_f}) &= \tilde{f}_k(x^{k+1}) + \frac{\mu_k}{2} \|\hat{x}^{k_f} - \hat{x}^{k_f}\|^2 + \langle \hat{v}^k, \hat{x}^{k_f} - x^{k+1} \rangle \\
&= f_a(\hat{x}^{k_f}) - \langle \hat{v}^k, \hat{x}^{k_f} - x^{k+1} \rangle + \langle \hat{v}^k, \hat{x}^{k_f} - x^{k+1} \rangle \\
&= f_a(\hat{x}^{k_f})
\end{aligned}$$

Sabendo que $f_a(x) \leq f(x)$ para todo $x \in \mathbb{R}^n$, temos:

$$\begin{aligned}
f(\hat{x}^{k_f}) - \delta_{k+1} + \frac{\mu_k}{2} \|\hat{x}^{k_f} - x^{k+1}\|^2 &\leq f(\hat{x}^{k_f}) \\
\Leftrightarrow \|\hat{x}^{k_f} - x^{k+1}\|^2 &\leq \frac{2\delta_{k+1}}{\mu_k} \leq \frac{2\delta_{k_f}}{\mu_{k_f}} \doteq r^2,
\end{aligned}$$

sendo que a última desigualdade vem do fato de que $\{1/\mu_k\}_{k>k_f}$ e $\{\delta_k\}_{k>k_f}$ são ambos não crescentes. Logo, $\{x^k\}_{k>k_f}$ é limitado.

Mostremos agora que $\{\delta_k\}_{k>k_f}$ tende a 0. Para todo $k > k_f$, como temos apenas passos nulos,

$$\begin{aligned} f(\hat{x}^k) - f(x^{k+1}) &< m\delta_{k+1} \\ \Leftrightarrow (1-m)\delta_{k+1} &< -f(\hat{x}^k) + f(x^{k+1}) + \delta_{k+1} \end{aligned}$$

Pela definição de decréscimo nominal (4.3),

$$\begin{aligned} (1-m)\delta_{k+1} &< f(x^{k+1}) - \tilde{f}_k(x^{k+1}) - \frac{\mu_k}{2}\|x^{k+1} - \hat{x}^k\|^2 \\ &\leq f(x^{k+1}) - \tilde{f}_k(x^{k+1}) \end{aligned} \quad (4.34)$$

Sabemos que $f(x^k) = \tilde{f}_k(x^k)$ para todo $k > k_f$. Então,

$$f(x^{k+1}) - \tilde{f}_k(x^{k+1}) = f(x^{k+1}) + \tilde{f}_k(x^k) - f(x^k) - \tilde{f}_k(x^{k+1}) \quad (4.35)$$

Assim, de (4.34) e (4.35), temos:

$$0 \leq (1-m)\delta_{k+1} \leq f(x^{k+1}) + \tilde{f}_k(x^k) - f(x^k) - \tilde{f}_k(x^{k+1}),$$

para todo $k > k_f$. Pela Lipschitz continuidade de f e de \tilde{f}_k , temos:

$$0 \leq (1-m)\delta_{k+1} \leq L\|x^{k+1} - x^k\| \Rightarrow -\|x^{k+1} - x^k\|^2 \leq -\frac{(1-m)^2\delta_{k+1}^2}{L^2}$$

para todo $k > k_f$ e para algum $L > 0$. Substituindo essa desigualdade em (4.33), temos:

$$\frac{(1-m)^2\mu_k\delta_{k+1}^2}{2L^2} \leq \delta_k - \delta_{k+1}.$$

Somando-se tal desigualdade em $k \in K$, temos:

$$\frac{(1-m)^2}{2L^2} \sum_{k>k_f} \mu_k\delta_{k+1}^2 \leq \sum_{k>k_f} (\delta_k - \delta_{k+1}) < \delta_{k_f} < +\infty.$$

Pelo item (a) da proposição 4.7, $\varepsilon_k \leq \delta_{k+1}$ para todo $k > k_f$. Utilizando-se esse fato e pela hipótese (a) deste teorema, $\varepsilon_k \rightarrow 0$ quando $k \rightarrow \infty$ e

$$\sum_{k>k_f} \frac{\mu_{k-1}}{\mu_k^2} \|\hat{v}^k\|^4 < +\infty.$$

Portanto, pela hipótese (b), $\{\hat{v}^k\}_{k>k_f} \rightarrow 0$. Pelo item (d) da proposição 4.7, sabemos também que $\hat{v}^k \in \partial_{\varepsilon_k} f(\hat{x}^{k_f})$. Logo, como $\partial_{\varepsilon_k} f(\cdot)$ é fechado, temos que $0 \in \partial f(\hat{x}^{k_f})$, ou seja, \hat{x}^{k_f} é minimizador de f . \square

Capítulo 5

Método de Subgradientes e Convergência Ergódica

Neste capítulo, veremos uma forma de resolver problemas do tipo (1.1). Conforme [SC96], no qual baseamos este texto, optamos por abordar aqui problemas um pouco mais genéricos que (1.1). Mais especificamente, estudaremos uma maneira de resolver o seguinte problema:

$$\text{minimizar } \langle c, x \rangle \quad \text{sujeito a} \quad \begin{cases} Ax \leq b \\ x \in \Psi \end{cases} \quad (5.1)$$

onde Ψ é um poliedro não vazio que corresponde às restrições “fáceis”, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, e as restrições de $Ax \leq b$ são “difíceis”. Refirimo-nos também a [Sho85, LL89, LPS99] para aprofundar a idéia que veremos. Em particular, [LPS99] apresenta resultados para programação convexa em geral. Considere inicialmente a seguinte função lagrangeana:

$$\Psi \times \mathbb{R}^m \ni (x, \pi) \mapsto L(x, \pi) \doteq \langle c, x \rangle + \langle Ax - b, \pi \rangle,$$

De maneira análoga ao que vimos no capítulo 1, o problema (5.1) pode ser escrito da seguinte forma:

$$\inf_{x \in \Psi} \sup_{\pi \geq 0} L(x, \pi).$$

Nesse caso, a relação de dualidade forte é verdadeira, ou seja,

$$\inf_{x \in \Psi} \sup_{\pi \geq 0} L(x, \pi) = \sup_{\pi \geq 0} \inf_{x \in \Psi} L(x, \pi). \quad (5.2)$$

A função objetivo dual, denotada por $f(\pi) \doteq \inf_{x \in \Psi} L(x, \pi)$, é o ínfimo de funções afins, e portanto é côncava e possivelmente não diferenciável. Logo, o problema dual

$$\text{maximizar } f(\pi) \text{ sujeito a } \pi \geq 0 \quad (5.3)$$

é um problema de otimização não diferenciável do tipo (3.1), sendo que o conjunto X dado em (3.1) é, nesse caso, \mathbb{R}_+^m .

Como esse conjunto de restrições duais é simples, podemos usar o método de subgradientes projetados, abordado na seção 3.3, para resolvê-lo. Conforme (3.10), o passo é dado por

$$\pi^{k+1} = P_{\mathbb{R}_+^m}(\pi^k + t_k v^k),$$

onde $v^k \in \partial f(\pi^k)$ e $t_k > 0$ é o tamanho do passo. Mais especificamente, o iterado π^{k+1} é determinado por:

$$\pi_i^{k+1} = \max \{(\pi^k + t_k v^k)_i, 0\}, \quad (5.4)$$

para cada $i = 1, \dots, m$. Defina agora X^* e Π^* como os conjuntos de soluções ótimas primais e duais, respectivamente. Defina também o conjunto

$$\Psi(\pi) \doteq \operatorname{argmin}_{x \in \Psi} L(x, \pi) = \{x \in \Psi : L(x, \pi) = f(\pi)\},$$

ou seja, o conjunto dos minimizadores do problema que define a função objetivo dual. Observe que como Ψ corresponde às restrições “fáceis”, também é fácil obter um elemento de $\Psi(\pi^k)$ para algum π^k . A proposição abaixo mostra, ainda, que o cômputo de um supergradiente de f em um ponto π^k pode ser feito gratuitamente dado um ponto de $\Psi(\pi^k)$.

Proposição 5.1. *Seja $x^k \in \Psi(\pi^k)$. Então,*

$$v^k \doteq Ax^k - b \in \partial f(\pi^k), \quad (5.5)$$

Demonstração. Para todo $\pi \in \mathbb{R}^m$,

$$\begin{aligned} f(\pi) &= \min_{x \in \Psi} L(x, \pi) \leq L(x^k, \pi) \\ &= \langle c, x^k \rangle + \langle Ax^k - b, \pi \rangle + \langle Ax^k - b, \pi^k \rangle - \langle Ax^k - b, \pi^k \rangle \\ &= f(\pi^k) + \langle Ax^k - b, \pi - \pi^k \rangle, \end{aligned}$$

onda a última igualdade é válida pois $x^k \in \Psi(\pi^k)$ ou, equivalentemente, $f(\pi^k) = L(x^k, \pi^k)$. Logo, $v^k \doteq Ax^k - b \in \partial f(\pi^k)$, conforme queríamos. \square

Com isso, o problema dual pode ser resolvido pelo método de subgradientes (projetados). No entanto, em muitos problemas, não estamos interessados apenas no valor ótimo da função objetivo. Queremos também obter uma solução ótima primal $x^* \in X^*$. Considere então $x^i \in \Psi(\pi^i)$. Infelizmente, não é verdade que $x^i \rightarrow x^* \in X^*$, mas podemos construir $\{\hat{x}^k\}$ tal que $\hat{x}^k \rightarrow x^* \in X^*$. Tal $\{\hat{x}^k\}$ é denominada *seqüência ergódica*, e será construída da seguinte forma:

$$\hat{x}^k \doteq \sum_{i=1}^k \lambda_{i,k} x^i, \quad \text{onde } (\lambda_{1,k}, \dots, \lambda_{k,k})^T \in \Delta_k \text{ e } x^i \in \Psi(\pi^i). \quad (5.6)$$

para toda iteração k . Observe que $\lambda_{i,k}$, com $i = 1, \dots, k$, são coeficientes de combinação convexa.

Podemos resumir agora as idéias do método com o algoritmo abaixo.

Algoritmo 5.2. (*Método de Subgradientes com Convergência Ergódica*)

1. Seja k_{\max} o número máximo de iterações permitidas no algoritmo.
2. Tome $\pi^1 \in \mathbb{R}^n$ e seja $k = 1$.
3. (Caixa preta) Compute $x^k \in \Psi(\pi^k)$ e $v^k \doteq Ax^k - b$.
4. (Combinação convexa) Tome $(\lambda_{1,k}, \dots, \lambda_{k,k})^T \in \Delta_k$ adequado.
5. (Próximo iterado primal) Defina $\hat{x}^k \doteq \sum_{i=1}^k \lambda_{i,k} x^i$.
6. (Critério de parada) Se $k \geq k_{\max}$, pare.
7. (Busca linear) Tome um passo $t_k > 0$ adequado.
8. (Próximo iterado dual) Seja $\pi_i^{k+1} \doteq \max\{(\pi^k + t_k v^k)_i, 0\}$, $\forall i = 1, \dots, m$.
9. (Loop) Tome $k = k + 1$ e vá para o passo 3.

Os passos 4 e 7 do algoritmo acima são explicados no teorema 5.3 abaixo. Antes disso, defina, para todo k ,

$$\gamma_{i,k} \doteq \frac{\lambda_{i,k}}{t_i}, \quad \text{para todo } i = 1, \dots, k \quad (5.7)$$

e considere

$$\Delta\gamma_k^{\max} \doteq \max_{i=2, \dots, k} \left\{ \gamma_{i,k} - \gamma_{(i-1),k} \right\}. \quad (5.8)$$

Teorema 5.3. *Considere o algoritmo 5.2 com iterados $\pi^k \rightarrow \pi^*$ para algum $\pi^* \geq 0$. Se os tamanhos de passo t_k e os coeficientes de combinação convexa $\lambda_{i,k}$ são escolhidos de modo a satisfazer:*

- (a) $\gamma_{i,k} \geq \gamma_{(i-1),k}$ para todo $i = 2, \dots, k$ e todo k ;
- (b) $\Delta \gamma_k^{\max} \rightarrow 0$ quando $k \rightarrow \infty$;
- (c) $\gamma_{1,k} \rightarrow 0$ quando $k \rightarrow \infty$; e
- (d) $\gamma_{k,k} \leq \xi$ para todo k e para algum $\xi > 0$;

então qualquer ponto de acumulação x^* de $\{\hat{x}^k\}$ é um ponto viável. Ademais, x^* e π^* são, respectivamente, soluções ótimas primal e dual. Note que uma condição que garante a existência de tal x^* é que Ψ seja limitado.

Demonstração. Note que a viabilidade do ponto dual π^* é dada por hipótese. Começemos provando, então, a viabilidade primal de um ponto de acumulação x^* de $\{\hat{x}^k\}$. Pelas igualdades (5.5) e (5.6), temos:

$$A\hat{x}^k - b = \sum_{i=1}^k \lambda_{i,k}(Ax^i - b) = \sum_{i=1}^k \lambda_{i,k}v^i \quad (5.9)$$

Além disso, por (5.4), $v_j^i \leq (\pi^{i+1} - \pi^i)_j / t_i$ para todo $j = 1, \dots, n$, e assim,

$$\begin{aligned} (A\hat{x}^k - b)_j &\leq \sum_{i=1}^k \frac{\lambda_{i,k}}{t_i} (\pi^{i+1} - \pi^i)_j \\ &= -\frac{\lambda_{1,k}}{t_1} \pi_j^1 + \frac{\lambda_{k,k}}{t_k} \pi_j^{k+1} - \sum_{i=2}^k \frac{\lambda_{i,k}}{t_i} \pi_j^i + \sum_{i=1}^{k-1} \frac{\lambda_{i,k}}{t_i} \pi_j^{i+1} \\ &= -\frac{\lambda_{1,k}}{t_1} \pi_j^1 + \frac{\lambda_{k,k}}{t_k} \pi_j^{k+1} - \sum_{i=2}^k \left(\frac{\lambda_{i,k}}{t_i} - \frac{\lambda_{(i-1),k}}{t_{i-1}} \right) \pi_j^i \\ &= -\gamma_{1,k} \pi_j^1 + \gamma_{k,k} \pi_j^{k+1} - \sum_{i=2}^k (\gamma_{i,k} - \gamma_{(i-1),k}) \pi_j^i \\ &= -\gamma_{1,k} \pi_j^1 + \gamma_{k,k} \pi_j^{k+1} - \sum_{i=2}^k (\gamma_{i,k} - \gamma_{(i-1),k}) \pi_j^* + \\ &\quad + \sum_{i=2}^k (\gamma_{i,k} - \gamma_{(i-1),k}) (\pi^* - \pi^i)_j \\ &= \gamma_{1,k} (\pi^* - \pi^1)_j + \gamma_{k,k} (\pi^{k+1} - \pi^*)_j \\ &\quad + \sum_{i=2}^k (\gamma_{i,k} - \gamma_{(i-1),k}) (\pi^* - \pi^i)_j. \end{aligned} \quad (5.10)$$

Sabemos que para todo $\varepsilon > 0$, existe \bar{k} suficientemente grande tal que

$$\|\pi^* - \pi^i\| \leq \frac{\varepsilon}{2\xi} \quad \text{para todo } i > \bar{k}. \quad (5.11)$$

Além disso, pela hipótese (b) deste teorema, dado $\ell \leq \bar{k}$,

$$\Delta\gamma_k^{\max} \sum_{i=\ell}^{\bar{k}} \|\pi^* - \pi^i\| \leq \frac{\varepsilon}{2} \quad \text{para } k \geq \bar{k} \text{ suficientemente grande.} \quad (5.12)$$

Defina agora

$$z^k \doteq \sum_{i=\ell}^k (\gamma_{i,k} - \gamma_{(i-1),k}) (\pi^* - \pi^i). \quad (5.13)$$

Então, para todo $k \geq \bar{k}$ suficientemente grande, temos, pela hipótese (a),

$$\begin{aligned} \|z^k\| &\leq \sum_{i=\ell}^k (\gamma_{i,k} - \gamma_{(i-1),k}) \|\pi^* - \pi^i\| \\ &\leq \sum_{i=\ell}^{\bar{k}} (\gamma_{i,k} - \gamma_{(i-1),k}) \|\pi^* - \pi^i\| + \sum_{i=\bar{k}+1}^k (\gamma_{i,k} - \gamma_{(i-1),k}) \|\pi^* - \pi^i\| \\ &\leq \Delta\gamma_k^{\max} \sum_{i=\ell}^{\bar{k}} \|\pi^* - \pi^i\| + \sum_{i=\bar{k}+1}^k (\gamma_{i,k} - \gamma_{(i-1),k}) \|\pi^* - \pi^i\| \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2\xi} \sum_{i=\bar{k}+1}^k (\gamma_{i,k} - \gamma_{(i-1),k}) \\ &= \frac{\varepsilon}{2} + \frac{\varepsilon}{2\xi} (\gamma_{k,k} - \gamma_{\bar{k},k}), \end{aligned}$$

onde a última desigualdade é dada por (5.11) e (5.12). Utilizando agora a hipótese (d) deste teorema,

$$\|z^k\| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2\xi} (\gamma_{k,k} - \gamma_{\bar{k},k}) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2\xi} \gamma_{k,k} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Como $\varepsilon > 0$ é arbitrário, concluímos que $z^k \rightarrow 0$ quando $k \rightarrow \infty$. Assim, tomando-se $\ell = 2$ na definição de z^k , temos que o último termo da direita de (5.10) tende a 0. Seja x^* um ponto de acumulação de $\{\hat{x}^k\}$. Note que se Ψ fosse limitado, então $\{\hat{x}^k\}$ seria limitada e então existiria uma subsequência convergente a $x^* \in \Psi$. Tomando limite em (5.10) para $k \rightarrow \infty$ e para cada j ,

$$Ax^* - b \leq \lim_{k \rightarrow \infty} (\gamma_{1,k}(\pi^* - \pi^1) + \gamma_{k,k}(\pi^{k+1} - \pi^*)).$$

Assim, pela hipótese (c) e pelo fato de que $\pi^k \rightarrow \pi^*$, temos também que $Ax^* \leq b$, ou seja, o ponto de acumulação x^* é viável.

Demonstremos agora a otimalidade dos pontos x^* e π^* . Como f é côncava e linear por partes, pela proposição 2.25, existe $r > 0$ tal que $\partial f(\pi) \subset \partial f(\pi^*)$ para todo $\pi \in B(\pi^*, r)$. Logo, como $\pi^k \rightarrow \pi^*$, por (5.5), existe $k_1 < \bar{k}$ suficientemente grande tal que

$$(Ax^k - b) \in \partial f(\pi^k) \subset \partial f(\pi^*) \quad \text{para todo } k \geq k_1. \quad (5.14)$$

Além disso, de (5.5) e (5.4), existe $k_2 < \bar{k}$ também suficientemente grande tal que

$$\pi_j^k > 0 \quad \text{e} \quad v_j^k = (Ax^k - b)_j = \frac{(\pi^{k+1} - \pi^k)_j}{t_k} \quad \text{se } \pi_j^* > 0. \quad (5.15)$$

Defina $k' \doteq \max\{k_1, k_2\}$. Então, para $k \geq k'$, temos, por (5.8), que

$$\begin{aligned} 0 &\leq \sum_{i=1}^{k'} \gamma_{i,k} = \gamma_{1,k} + \sum_{i=2}^{k'} \gamma_{(i-1),k} + \sum_{i=2}^{k'} (\gamma_{i,k} - \gamma_{(i-1),k}) \\ &\leq \gamma_{1,k} + \sum_{i=1}^{k'-1} \gamma_{i,k} + \Delta \gamma_k^{\max} (k' - 1). \end{aligned}$$

Aplicando recursivamente a desigualdade em $\sum_{i=1}^{k'-1} \gamma_{i,k}$, temos:

$$\begin{aligned} 0 &\leq \sum_{i=1}^{k'} \gamma_{i,k} \leq (k' - 1)\gamma_{1,k} + \gamma_{1,k} + \Delta \gamma_k^{\max} \sum_{i=1}^{k'-1} (k' - i) \\ &= k' \gamma_{1,k} + \Delta \gamma_k^{\max} \frac{(k' - 1)k'}{2}. \end{aligned}$$

Pelas hipóteses (b) e (c), a desigualdade acima resulta que $\sum_{i=1}^{k'} \gamma_{i,k} \rightarrow 0$ para $k \rightarrow \infty$. Isso implica, por (5.7), que

$$\lambda_{i,k} \rightarrow 0 \quad \text{para todo } i = 1, \dots, k' \quad \text{com } k \rightarrow \infty. \quad (5.16)$$

Mostremos agora que $x^* \in \Psi(\pi^*)$. Seja $k \geq k'$. Então, por (5.14), temos:

$$f(\pi) \leq f(\pi^*) + \langle Ax^k - b, \pi - \pi^* \rangle, \quad \text{para todo } \pi \in \mathbb{R}^m.$$

Em particular, a desigualdade acima vale para $\pi \doteq \pi^k$. Sabemos também que $x^k \in \Psi(\pi^k)$. Então:

$$\begin{aligned} f(\pi^k) &\leq f(\pi^*) + \langle Ax^k - b, \pi^k - \pi^* \rangle \\ \Leftrightarrow L(x^k, \pi^k) &\leq f(\pi^*) + \langle Ax^k - b, \pi^k - \pi^* \rangle \\ \Leftrightarrow \langle c, x^k \rangle &\leq f(\pi^*) - \langle Ax^k - b, \pi^* \rangle \\ \Leftrightarrow L(x^k, \pi^*) &\leq f(\pi^*), \end{aligned}$$

onde as duas últimas desigualdades são válidas por definição da função lagrangeana L . Ademais, por definição de f , temos:

$$f(\pi^*) = \inf_{x \in \Psi} L(x, \pi^*) \leq L(x^k, \pi^*),$$

e assim concluímos que $f(\pi^*) = L(x^k, \pi^*)$ ou, equivalentemente, $x^k \in \Psi(\pi^*)$, quando $k \geq k'$. Por (5.6) e pelo fato de que $\Psi(\pi^*)$ é convexo, conclui-se também que $\hat{x}^k \in \Psi(\pi^*)$ para todo k , e portanto $x^* \in \Psi(\pi^*)$.

Mostremos agora que as folgas complementares são satisfeitas. Tome $\pi_j^* > 0$ para algum $j = 1, \dots, n$. Por (5.9) e (5.15), temos:

$$\begin{aligned} (A\hat{x}^k - b)_j &= \sum_{i=1}^k \lambda_{i,k} v_j^i = \sum_{i=1}^{k'} \lambda_{i,k} v_j^i + \sum_{i=k'+1}^k \lambda_{i,k} v_j^i \\ &= \sum_{i=1}^{k'} \lambda_{i,k} v_j^i + \sum_{i=k'+1}^k \frac{\lambda_{i,k}}{t_i} (\pi^{i+1} - \pi^i)_j \\ &= \sum_{i=1}^{k'} \lambda_{i,k} v_j^i + \gamma_{(k'+1),k} (\pi^* - \pi^{k'+1})_j + \gamma_{k,k} (\pi^{k+1} - \pi^*)_j \\ &\quad + \sum_{i=k'+2}^k (\gamma_{i,k} - \gamma_{(i-1),k}) (\pi^* - \pi^i)_j, \end{aligned} \tag{5.17}$$

sendo que a última igualdade é válida por analogia a (5.10). Veremos agora que o termo da direita de (5.17) converge para 0 quando $k \rightarrow \infty$. Observe que para $k \rightarrow \infty$, temos $\pi^k \rightarrow \pi^*$, o primeiro termo de (5.17) tende a 0 por (5.16) e o último termo tende a zero pois equivale à definição (5.13), com $\ell = k' + 2$.

Falta-nos mostrar que $\gamma_{(k'+1),k} \rightarrow 0$ quando $k \rightarrow \infty$. De fato,

$$0 \leq \gamma_{i,k} - \gamma_{(i-1),k} \leq \Delta \gamma_k^{\max} \text{ para todo } i = 2, \dots, k.$$

Como $\gamma_{1,k} \rightarrow 0$ e $\Delta\gamma_k^{\max} \rightarrow 0$ quando $k \rightarrow \infty$ pelos itens (c) e (d) deste teorema, temos, em particular, que $0 \leq \gamma_{2,k} - \gamma_{1,k} \leq \Delta\gamma_k^{\max}$ implica $\gamma_{2,k} \rightarrow 0$. Um processo simples de indução resulta que $\gamma_{i,k} \rightarrow 0$ para todo i , inclusive para $i = k' + 2$, o qual queríamos. Logo, quando $\pi_j^* > 0$, concluímos que $(Ax^* - b)_j = 0$ ou seja, as folgas complementares são satisfeitas. Isso completa a demonstração de otimalidade de π^* e x^* . \square

Mostraremos agora duas regras para escolha de tamanhos de passo t_k e coeficientes de combinação convexa $\lambda_{i,k}$. A primeira, abordada por Shor [Sho85], faz uma média ponderada de pontos $\{x^i\}_{i=1}^k$ em cada iteração k , enquanto que a segunda, apresentada por Larsson e Liu [LL89], realiza uma média aritmética dos mesmos.

Corolário 5.4. (Regra de Shor) *Considere o algoritmo 5.2 com iterados $\pi^k \rightarrow \pi^*$ para algum $\pi^* \geq 0$. Se os tamanhos de passo t_k são tais que*

$$t_k > 0, \quad \forall k, \quad \lim_{k \rightarrow \infty} t_k = 0, \quad \sum_{k=1}^{\infty} t_k = +\infty, \quad (5.18)$$

e os coeficientes de combinação convexa $\lambda_{i,k}$ são tais que

$$\lambda_{i,k} \doteq \frac{t_i}{\sum_{j=1}^k t_j}, \quad \text{para todo } i = 1, \dots, k \text{ e todo } k. \quad (5.19)$$

Então π^* e qualquer ponto de acumulação x^* de $\{\hat{x}^k\}$ são, respectivamente, soluções ótimas primal e dual.

Demonstração. O resultado segue do teorema 5.3. Para isso, verifiquemos que as condições (a) – (d) são verdadeiras. Pela definição (5.7) e por (5.19),

$$\gamma_{i,k} \doteq \frac{\lambda_{i,k}}{t_i} = \frac{1}{\sum_{j=1}^k t_j} = \frac{\lambda_{(i-1),k}}{t_{i-1}} \doteq \gamma_{(i-1),k} \quad \text{para todo } i = 2, \dots, k, \quad (5.20)$$

ou seja, a hipótese (a) é satisfeita. A igualdade (5.20), juntamente com a definição (5.8) implica também que $\Delta\gamma_k^{\max} = 0$ para todo k , e a hipótese (b) também é verdadeira. Ademais, $\sum_{i=1}^{\infty} t_k = \infty$ implica que $\gamma_{k,k} \rightarrow 0$ quando $k \rightarrow \infty$, e em particular, por (5.20), $\gamma_{1,k} = \gamma_{k,k}$ para todo k , e assim $\gamma_{1,k} \rightarrow 0$ quando $k \rightarrow \infty$, e as hipóteses (c) e (d) também são satisfeitas. \square

Corolário 5.5. (Regra de Larsson e Liu) *Considere o algoritmo 5.2 com iterados $\pi^k \rightarrow \pi^*$ para algum $\pi^* \geq 0$. Suponha que os tamanhos de passo t_k e os coeficientes de combinação convexa $\lambda_{i,k}$ são tais que*

$$t_k \in \left[\frac{a_1}{b + ck}, \frac{a_2}{b + ck} \right], \quad \lambda_{k,i} \doteq \frac{1}{k}, \quad \text{para todo } i = 1, \dots, k \text{ e todo } k, \quad (5.21)$$

com $0 < a_1 \leq a_2$, $b \geq 0$ e $c > 0$. Para cada k , podemos também escrever

$$t_k = \frac{\omega_k}{b + ck}, \quad \text{para algum } \omega_k \in [a_1, a_2]. \quad (5.22)$$

Agora suponha também que a seqüência $\{\omega_k\}$ é não crescente. Então π^* e qualquer ponto de acumulação x^* de $\{\hat{x}^k\}$ são, respectivamente, soluções ótimas primal e dual.

Demonstração. O resultado segue novamente pelo teorema 5.3, verificando-se que as condições (a) – (d) são verdadeiras. Por (5.7) e pelas definições de $\lambda_{k,i}$ em (5.21) de t_i em (5.22), temos:

$$\gamma_{i,k} \doteq \frac{\lambda_{i,k}}{t_i} = \frac{b + ci}{\omega_i k}, \quad \text{para todo } i = 1, \dots, k \text{ e todo } k.$$

Dessa forma,

$$\gamma_{i,k} - \gamma_{(i-1),k} = \frac{b + ci}{\omega_i k} - \frac{b + c(i-1)}{\omega_{i-1} k} = \frac{b + ci}{k} \left(\frac{1}{\omega_i} - \frac{1}{\omega_{i-1}} \right) + \frac{c}{\omega_{i-1} k} \quad (5.23)$$

para todo $i = 2, \dots, k$ e todo k . Como $\{\omega_k\}$ é não crescente por hipótese, temos $1/\omega_i - 1/\omega_{i-1} \geq 0$. Além disso, como $\omega_{i-1} \geq a_1 > 0$, concluímos que $\gamma_{i,k} - \gamma_{(i-1),k} > 0$ para todo $i = 2, \dots, k$ e todo k , ou seja a hipótese (a) do teorema é válida. Ademais, quando $k \rightarrow \infty$,

$$\gamma_{1,k} = \frac{b + c}{\omega_1 k} \rightarrow 0 \quad \text{e} \quad \gamma_{k,k} = \frac{b + ck}{\omega_k k} = \frac{b}{\omega_k k} + \frac{c}{\omega_k} \leq \frac{b}{a_1 k} + \frac{c}{a_1} \rightarrow \frac{c}{a_1}$$

isto é, as hipóteses (c) e (d) também são verdadeiras.

Mostremos agora que a hipótese (b) é satisfeita. Defina inicialmente

$$i_k \doteq \operatorname{argmax}_{i=2, \dots, k} \{\gamma_{i,k} - \gamma_{(i-1),k}\}.$$

Pela definição de $\Delta\gamma_k^{\max}$ dado em (5.8) e por (5.23),

$$\begin{aligned} \Delta\gamma_k^{\max} &= \max_{i=2, \dots, k} \left(\frac{b}{k} \left(\frac{1}{\omega_i} - \frac{1}{\omega_{i-1}} \right) + \frac{c}{\omega_{i-1} k} + \frac{ci}{k} \left(\frac{1}{\omega_i} - \frac{1}{\omega_{i-1}} \right) \right) \\ &= \left(\frac{b}{k} \left(\frac{1}{\omega_{i_k}} - \frac{1}{\omega_{i_k-1}} \right) + \frac{c}{\omega_{i_k-1} k} + \frac{ci_k}{k} \left(\frac{1}{\omega_{i_k}} - \frac{1}{\omega_{i_k-1}} \right) \right). \end{aligned}$$

Por conseqüência da hipótese (a), $\Delta\gamma_k^{\max} \geq 0$ para todo k . Então, basta provarmos aqui que $\limsup_{k \rightarrow \infty} \Delta\gamma_k^{\max} = 0$. Observe que pela limitação de ω_k , temos:

$$\frac{b}{k} \left(\frac{1}{\omega_{i_k}} - \frac{1}{\omega_{i_k-1}} \right) \rightarrow 0 \quad \text{e} \quad \frac{c}{\omega_{i_k-1} k} \rightarrow 0.$$

quando $k \rightarrow \infty$. Defina

$$y_k \doteq \frac{ci_k}{k} \left(\frac{1}{\omega_{i_k}} - \frac{1}{\omega_{i_k-1}} \right).$$

Considere uma subsequência K de índices tal que $\{y_k\}_{k \in K} \rightarrow \limsup_{k \rightarrow \infty} y_k$. Se $\{i_k\}_{k \in K}$ é limitada, então, $\{y_k\}_{k \in K} \rightarrow 0$ quando $k \rightarrow \infty$. Do contrário, tome $K' \subset K$ tal que $\{i_k\}_{k \in K'} \rightarrow +\infty$. Nesse caso, observe que como $\{\omega_k\}$ é não crescente e limitada por baixo, $\{1/\omega_{i_k} - 1/\omega_{i_k-1}\}_{k \in K'} \rightarrow 0$. Logo, $\{y_k\}_{k \in K'} \rightarrow 0$ e pela definição de K , $\limsup_{k \rightarrow \infty} y_k = 0$. Concluímos dessa forma que $\limsup_{k \rightarrow \infty} \Delta\gamma_k^{\max} = 0$, o que completa a demonstração. \square

A proposição 5.6 abaixo mostra que a regra para a escolha do tamanho de passo t_k de Larsson e Liu (corolário 5.5) é um caso particular da regra de Shor (corolário 5.4). Quanto aos coeficientes de combinação convexa $\lambda_{i,k}$, a regra de Larsson e Liu é considerada mais razoável que a de Shor, pois como $\lim_{k \rightarrow \infty} t_k = 0$, essa última regra atribui pesos grandes para $x^i \in \Psi(\pi^i)$ gerados nas primeiras iterações.

Proposição 5.6. *Se os tamanhos de passo t_k são definidos por (5.22), então, as condições (5.18) são verdadeiras.*

Demonstração. Como $\omega_k, c > 0$ e $b \geq 0$, claramente temos que $t_k > 0$ para todo k . Além disso,

$$\lim_{k \rightarrow \infty} t_k = \lim_{k \rightarrow \infty} \frac{\omega_k}{b + ck} = 0.$$

Ademais, sabendo que $t_k \doteq \omega_k/(b + ck) \geq \omega_k/[(b + c)k]$ para todo k , temos:

$$\sum_{k=1}^{\infty} t_k \geq \frac{1}{b + c} \sum_{k=1}^{\infty} \frac{\omega_k}{k} \geq \frac{a_1}{b + c} \sum_{k=1}^{\infty} \frac{1}{k} = +\infty \Rightarrow \sum_{k=1}^{\infty} t_k = +\infty.$$

Logo, os tamanhos de passos t_k definidos no corolário 5.5 satisfazem as condições (5.18). \square

Os tamanhos de passo t_k definidos no corolário 5.5 seguem a *estratégia de relaxação quase completa*, dado por Dem'yanov e Vasil'ev [DV85, seção 3.4]. Tal estratégia induz a convergência do método de subgradientes definindo previamente duas seqüências $\{\underline{t}_k\}$ e $\{\bar{t}_k\}$ com $\underline{t}_k \leq \bar{t}_k$ para todo k , com ambas as seqüências (salva-guardas) satisfazendo as condições (5.18). Supondo que \tilde{t}_k é gerado na iteração k , o tamanho de passo t_k usado será dado por

$$t_k \doteq P_{[\underline{t}_k, \bar{t}_k]}(\tilde{t}_k).$$

Pode-se mostrar que a seqüência $\{t_k\}$, por sua vez, também satisfaz as condições (5.18). Assim, obtemos dinamicamente (isto é, durante o processo do algoritmo) tamanhos de passo que satisfazem (5.18).

Capítulo 6

Algoritmo do Volume

O capítulo anterior aborda uma maneira de resolver problemas do tipo (1.1), utilizando o método de subgradientes em conjunto com seqüências ergódicas. Veremos agora uma outra forma, denominada *algoritmo do volume* (VA). Esse método é uma extensão do método de subgradientes, que produz uma aproximação de solução primal enquanto se resolve o dual. Consideramos ainda duas de suas modificações: *VA revisado* (RVA) e *VA com variação do alvo* (VA+VTVM). Neste capítulo, discutiremos tais técnicas fundamentando-se em [BA00, BMS02, SL04].

Conforme vimos no capítulo 1, o problema dual (1.5) é de maximização irrestrita de uma função f côncava não necessariamente diferenciável em todos os pontos. Assim, uma solução ótima dual será obtida através da geração de iterados com busca de possíveis direções de crescimento e de tamanhos de passo. Por outro lado, uma solução ótima primal será obtida estimando-se o volume associado às faces ativas das restrições “difíceis”, o que justifica o nome do algoritmo. Veremos a seguir detalhes disso.

Este capítulo está dividido em três partes, na qual abordaremos o VA original (seção 6.1), o RVA (seção 6.2) e o VA+VTVM (seção 6.3).

6.1 Algoritmo do Volume Original

Analogamente ao capítulo 5, definimos o conjunto

$$\Psi(\pi^k) \doteq \operatorname{argmin}_{x \in \Psi} L(x, \pi^k) = \left\{ x \in \Psi : L(x, \pi^k) = f(\pi^k) \right\},$$

com $L(x, \pi^k)$ e $f(\pi^k)$ especificados respectivamente em (1.2) e (1.4). Conforme a proposição 5.1, se $x^k \in \Psi(\pi^k)$, então $Ax^k - b \in \partial f(\pi^k)$. Em outras

palavras, um supergradiente de f em π^k é diretamente obtido depois que o problema $\min_{x \in \Psi} L(x, \pi^k)$ é resolvido. Note que quanto mais simples é o poliedro Ψ , mais simples será resolver tal problema.

Os supergradientes fornecidos pela caixa preta são representados aqui por $v^k \doteq Ax^k - b$, com $x^k \in \Psi(\pi^k)$. Assim como o método de feixe, o VA gera pontos candidatos π^k e centros de estabilização $\hat{\pi}^k$. Além disso, um candidato a direção de crescimento não será dado simplesmente por v^k , mas por combinações convexas de elementos de $\{v^i\}_{i=0}^k$, indicados por w^k . Denotamos também z^k como uma combinação convexa de pontos de $\{x^i\}_{i=0}^k$, que será usada para aproximar soluções primais. A estrutura do algoritmo original é dada a seguir.

Algoritmo 6.1. *Algoritmo do Volume*

1. Sejam tol_w e tol_f tolerâncias dadas e $\pi^0 \in \mathbb{R}^m$ o ponto inicial.
2. (Caixa preta) Compute $x^0 \in \Psi(\pi^0)$, $f(\pi^0)$ e $v^0 \doteq Ax^0 - b$.
3. (Inicialização) Sejam $z^1 \doteq x^0$, $w^1 \doteq v^0$, $\hat{\pi}^1 \doteq \pi^0$ e $k = 1$.
4. (Critério de parada) Se $\|w^k\| \leq \text{tol}_w$ e $\frac{\langle c, z^k \rangle - f(\hat{\pi}^k)}{f(\hat{\pi}^k)} \leq \text{tol}_f$, então pare.
5. (Busca linear) Tome um tamanho de passo t_k adequado.
6. (Atualização do candidato) Faça $\pi^k \doteq \hat{\pi}^k + t_k w^k$.
7. (Caixa preta) Compute $x^k \in \Psi(\pi^k)$, $f(\pi^k)$ e $v^k \doteq Ax^k - b$.
8. (Teste de crescimento) Se $f(\pi^k) > f(\hat{\pi}^k)$:
Então faça $\hat{\pi}^{k+1} \doteq \pi^k$ (*passo sério*).
Senão, $\hat{\pi}^{k+1} \doteq \hat{\pi}^k$ (*passo nulo*).
9. (Atualização de vetores) Dado um parâmetro $0 \leq \alpha_k \leq 1$, defina

$$z^{k+1} \doteq \alpha_k x^k + (1 - \alpha_k) z^k, \quad (6.1)$$

$$w^{k+1} \doteq \alpha_k v^k + (1 - \alpha_k) w^k. \quad (6.2)$$

10. (Loop) Faça $k = k + 1$ e vá para o passo 4.

Duas questões sobre o algoritmo 6.1 devem ser levantadas. A primeira está relacionada aos tamanhos de passo, requeridos pelo passo 5. O VA original especifica a escolha de t_k usando a versão de (3.9) para funções côncavas, ou seja,

$$t_k \doteq \beta \frac{UB - f(\pi^{k-1})}{\|w^k\|^2}, \quad (6.3)$$

onde UB é um limitante superior do valor ótimo de (1.1) e para algum $\underline{\beta}$ e $\bar{\beta}$, tem-se $0 < \underline{\beta} \leq \beta \leq \bar{\beta} < 2$. Porém, tal escolha pode ser substituída por algumas condições semelhantes às enunciadas no teorema 3.11. De fato, algumas condições serão usadas para que o estudo de convergência do algoritmo seja possível, conforme veremos no teorema 6.13 da seção 6.2.

A segunda questão se refere ao parâmetro α_k utilizado no passo 9 como um coeficiente de combinações convexas. Antes de verificarmos como se escolhe α_k , considere o seguinte resultado, que relaciona os pontos z^k e w^k :

Proposição 6.2. *Considere as seqüências $\{z^k\}$ e $\{w^k\}$ geradas pelo algoritmo 6.1. Então $w^k = Az^k - b$ para todo k .*

Demonstração. Provemos por indução em k . Para $k = 1$, basta considerarmos as definições de w^1 , v^0 e z^1 , ou seja,

$$w^1 = v^0 = Ax^0 - b = Az^1 - b.$$

Suponha agora que a igualdade é válida para uma iteração $k \geq 1$. Provemos que ela é verdadeira para $k + 1$. Por (6.2), pelo fato de que $v^k \doteq Ax^k - b$ por definição, e pela hipótese de indução, temos:

$$\begin{aligned} w^{k+1} &= \alpha_k v^k + (1 - \alpha_k)w^k \\ &= \alpha_k(Ax^k - b) + (1 - \alpha_k)(Az^k - b) \\ &= A(\alpha_k x^k + (1 - \alpha_k)z^k) - \alpha_k b - (1 - \alpha_k)b \\ &= Az^{k+1} - b, \end{aligned}$$

sendo que a terceira igualdade é válida devido à linearidade de A e a última pela definição (6.1). \square

Dessa forma, a idéia básica é escolher α_k de modo a tentar atingir a viabilidade primal, ou seja, de modo que $w^k = Az^k - b \approx 0$. Considerando a proposição acima, temos, por (6.2), que

$$\operatorname{argmin}_{\alpha} \|Az^{k+1} - b\| = \operatorname{argmin}_{\alpha} \|w^{k+1}\| = \operatorname{argmin}_{\alpha} \|\alpha v^k + (1 - \alpha)w^k\|.$$

Com isso, podemos descrever o algoritmo para obtenção do parâmetro α_k da seguinte forma:

Algoritmo 6.3. *Escolha de α_k no Algoritmo do Volume.*

- Seja $\alpha_k^{\max} \in]0, 1[$ um limitante superior e $\varsigma \in]0, 1[$ uma constante.
- Compute $\bar{\alpha} \doteq \operatorname{argmin}_{\alpha} \|\alpha v^k + (1 - \alpha)w^k\|$.

- (c) Se $\bar{\alpha} \geq 0$, então faça $\alpha_k \doteq \min\{\bar{\alpha}, \alpha_k^{\max}\}$.
Senão, faça $\alpha_k \doteq \alpha_k^{\max}/10$.
- (d) Se depois de um certo número de iterações, a função objetivo dual não cresceu suficientemente, faça $\alpha_{k+1}^{\max} = \varsigma \alpha_k^{\max}$.
Senão, faça $\alpha_{k+1}^{\max} = \alpha_k^{\max}$.

O passo (d) do algoritmo acima é realizado com o intuito de aumentar a precisão da solução primal. Ademais, dizer que k é um dos últimos iterados equivale a dizer que estamos próximos de atingir o critério de parada dado no passo 4 do algoritmo 6.1. Observe que tal critério visa não apenas a viabilidade primal, mas também a otimalidade dual.

Verifiquemos agora que, de fato, w^k e z^k são combinações convexas dos elementos de $\{v^i\}_{i=0}^k$ e $\{x^i\}_{i=0}^k$, respectivamente.

Lema 6.4. *Para qualquer $k \geq 1$, considere os coeficientes:*

$$\lambda_{k,i} \doteq \alpha_{k-i} \prod_{j=k-i+1}^k (1 - \alpha_j), \quad i = 0, \dots, k, \quad (6.4)$$

onde $\alpha_0 \doteq 1$ e $\prod_{j=j_i}^{j_f} (1 - \alpha_j) \doteq 1$ se $j_f < j_i$. Então:

$$(\lambda_{k,0}, \dots, \lambda_{k,k})^T \in \Delta_{k+1}, \quad \text{para todo } k.$$

Demonstração. Como $\alpha_j \in [0, 1]$ para todo j , claramente temos que $\lambda_{k,i} \geq 0$ para todo $i = 0, \dots, k$ e todo k . Além disso,

$$\begin{aligned} \sum_{i=0}^k \lambda_{k,i} &= \sum_{i=0}^k \alpha_{k-i} \prod_{j=k-i+1}^k (1 - \alpha_j) \\ &= \alpha_k + \alpha_{k-1}(1 - \alpha_k) + \alpha_{k-2}(1 - \alpha_{k-1})(1 - \alpha_k) + \\ &\quad \dots + \alpha_0 \prod_{j=1}^k (1 - \alpha_j) \\ &= \alpha_k + (1 - \alpha_k) \left(\alpha_{k-1} + (1 - \alpha_{k-1})(\alpha_{k-2} + (1 - \alpha_{k-2})(\dots \alpha_0)) \right) \\ &= \alpha_k + (1 - \alpha_k) (\alpha_{k-1} + (1 - \alpha_{k-1})) \\ &= 1. \end{aligned}$$

Logo, $(\lambda_{k,0}, \dots, \lambda_{k,k})^T \in \Delta_{k+1}$ para todo k , conforme queríamos. \square

Lema 6.5. *Considere as seqüências $\{z^k\}$ e $\{w^k\}$ geradas pelo algoritmo 6.1, respectivamente em (6.1) e (6.2). Então,*

$$z^{k+1} = \sum_{i=0}^k \lambda_{k,i} x^{k-i} \quad e \quad w^{k+1} = \sum_{i=0}^k \lambda_{k,i} v^{k-i}$$

para todo k , onde $\lambda_{k,i}$ é definido em (6.4).

Demonstração. Provemos por indução em k o resultado para z^k . Para $k = 0$, temos $z^1 = \lambda_{0,0} x^0 = x^0$. Suponha então que a igualdade vale para um $k \geq 0$ e demonstremos para $k + 1$. Observe inicialmente que

$$\lambda_{k,(i-1)} = \alpha_{k-i+1} \prod_{j=k-i+2}^k (1 - \alpha_j) \quad e \quad \lambda_{(k+1),i} = \alpha_{k-i+1} \prod_{j=k-i+2}^{k+1} (1 - \alpha_j),$$

o que implica que $\lambda_{(k+1),i} = \lambda_{k,(i-1)}(1 - \alpha_{k+1})$. Além disso, pela definição (6.1) e por hipótese de indução,

$$\begin{aligned} z^{k+2} &= \alpha_{k+1} x^{k+1} + (1 - \alpha_{k+1}) \sum_{i=0}^k \lambda_{k,i} x^{k-i} \\ &= \alpha_{k+1} x^{k+1} + (1 - \alpha_{k+1}) \sum_{i=1}^{k+1} \lambda_{k,(i-1)} x^{k-i+1} \\ &= \alpha_{k+1} x^{k+1} + (1 - \alpha_{k+1}) \sum_{i=1}^{k+1} \frac{\lambda_{(k+1),i}}{(1 - \alpha_{k+1})} x^{k-i+1} \\ &= \sum_{i=0}^{k+1} \lambda_{(k+1),i} x^{k-i+1}, \end{aligned}$$

sendo que a última igualdade é válida pois $\lambda_{(k+1),0} = \alpha_{k+1}$, e isso completa a prova. Note que a demonstração do resultado para w^k é análoga. \square

Proposição 6.6. *Os pontos w^k e z^k , gerados pelo algoritmo 6.1, são combinações convexas dos elementos de $\{v^i\}_{i=0}^k$ e $\{x^i\}_{i=0}^k$, respectivamente.*

Demonstração. Segue diretamente dos lemas 6.4 e 6.5 anteriores. \square

Conforme veremos no capítulo 7, o VA original apresenta bons resultados numéricos. No entanto, sua análise de convergência não foi realizada em [BA00]. De fato, em tal texto, a escolha de α_k no passo 9 do algoritmo 6.1 não é necessariamente dada pelo algoritmo 6.3. Assim, considerando $\alpha_k \doteq \alpha$ fixo para todo k , podemos obter exemplos de não-convergência

primal do VA. Finalizamos essa seção com o seguinte exemplo, o qual foi adaptado de [SL04, seção 2]:

$$\text{minimizar } x \quad \text{sujeito a} \quad \begin{cases} x \geq 1 \\ x \in \Psi \doteq \{x \in \mathbb{R}: 0 \leq x \leq 2\} \end{cases} \quad (6.5)$$

Observe que esse problema é do tipo (5.1), sendo que $A = -I$ e $b = -1$. Assim, o problema dual pode ser assim escrito:

$$\text{maximizar } f(\pi) \quad \text{sujeito a} \quad \pi \geq 0,$$

onde $f(\pi) \doteq \min\{x + (1-x)\pi : 0 \leq x \leq 2\}$ e cujo gráfico é dado abaixo.

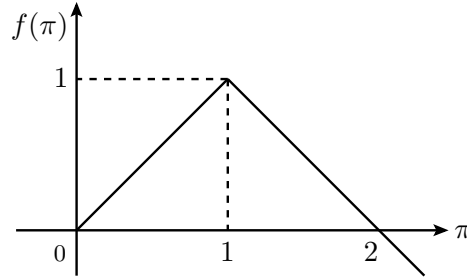


Figura 6.1.1: Não-convergência primal do VA: gráfico da função dual.

Observe que o valor ótimo e o ponto ótimo do problema são iguais a 1. Considere $\alpha_k \doteq \alpha$ para todo k , com $\alpha \in [0, 1]$, $UB \doteq 1$, $\pi^0 \doteq 0$ e $0 < \beta < 1$. Mostraremos que $z^k = 0$ para todo k e que $\lim_{k \rightarrow \infty} \pi^k = 1$, o que indica convergência apenas no espaço dual. Note também que se $\pi^k \in [0, 1[$, então $x^k \doteq \Psi(\pi^k) = 0$ e $v^k \doteq Ax^k - b = 1$. Assim, para concluirmos que $z^k = 0$ para todo k , basta mostrarmos que $\pi^k \in [0, 1[$ para todo k .

A demonstração será feita por indução em k . Considerando $k = 0$, sabemos que $\pi^0 = 0 \in [0, 1[$. Suponhamos então que $\pi^i \in [0, 1[$ para todo $i = 0, \dots, k$ e mostremos que $\pi^{k+1} \in [0, 1[$. Observe inicialmente que, pelo lema 6.5,

$$w^{k+1} = \sum_{i=0}^k \lambda_{k,i} v^{k-i},$$

onde $(\lambda_{k,0}, \dots, \lambda_{k,k})^T \in \Delta_{k+1}$. Como $v^i = 1$ para todo $i = 0, \dots, k$ (pois $x^i = 0$ para todo $i = 0, \dots, k$), temos que $w^{k+1} = 1$. Note também que $\pi^k \in [0, 1[$ quando $f(\pi^k) = \pi^k$. Com esses resultados e considerando o tamanho de passo dado por (6.3), temos no nosso exemplo que $t_{k+1} = \beta(1 - \pi^k)$ e

$$\pi^{k+1} = \hat{\pi}^k + t_{k+1} w^{k+1} = \pi^k + \beta(1 - \pi^k) = \beta + (1 - \beta)\pi^k.$$

Como $\pi^k \in [0, 1[$ e $0 < \beta < 1$, concluímos que $\pi^{k+1} \in [0, 1[$, conforme queríamos. Logo, a convergência primal não é estabelecida.

Mostremos agora que $\lim_{k \rightarrow \infty} \pi^{k+1} = 1$.

$$\begin{aligned}
 \pi^{k+1} &= \beta + (1 - \beta)\pi^k \\
 &= \beta + (1 - \beta)(\beta + (1 - \beta)\pi^{k-1}) \\
 &= \beta + \beta(1 - \beta) + (1 - \beta)^2\pi^{k-1} \\
 &= \beta + \beta(1 - \beta) + \beta(1 - \beta)^2 + \dots + \beta(1 - \beta)^k \\
 &= \beta \sum_{i=0}^k (1 - \beta)^i \\
 &= 1 - (1 - \beta)^{k+1}
 \end{aligned}$$

O resultado segue tomando-se $k \rightarrow \infty$ e lembrando que $0 < \beta < 1$.

É interessante notar que nesse exemplo, o primal não converge ao ótimo porque os pontos duais atingem apenas os valores correspondentes à face da esquerda do gráfico da função dual f (veja novamente a figura 6.1.1).

6.2 Algoritmo do Volume Revisado

Algumas modificações no VA podem ser feitas de modo a ser possível interpretá-lo como um método de feixe. Essa técnica, denominada *algoritmo do volume revisado* (RVA), obtém o coeficiente de combinação convexa α_k , basendo-se num modelo da função objetivo dual. Mostra-se também que ele pode ser visto como um método extra-gradiente. Nesta seção, mostraremos essas alterações, realizadas no algoritmo 6.1, e analisaremos sua convergência, fundamentando-se em [BMS02].

A primeira modificação que veremos é a adição de um novo ponto, definido recursivamente como em (6.1) e (6.2): $p^1 \doteq \pi^0$ e

$$p^{k+1} \doteq \alpha_k \pi^k + (1 - \alpha_k) p^k, \quad \text{para todo } k \geq 1. \quad (6.6)$$

Analogamente à proposição 6.6 e ao lema 6.5, concluímos que p^k é uma combinação convexa dos pontos de $\{\pi^i\}_{i=0}^k$, e, para todo $k \geq 1$,

$$p^{k+1} = \sum_{i=0}^k \lambda_{k,i} \pi^{k-i}.$$

Como resultado, mostramos na proposição abaixo que w^k , definido em (6.2), pode ser visto como um supergradiente aproximado de f em p^k . O

RVA usa, portanto, um ε_k -supergradiente que não é computado no iterado $\hat{\pi}^k$, mas em um ponto extra p^k . Por tal aspecto, consideramos que o algoritmo pode ser visto como um *método extra-gradiente* [SS99]. Além disso, dizemos que o passo realizado no item 6 do algoritmo 6.1 é “inexato”.

Proposição 6.7. *Considere a seqüência $\{\varepsilon_k\}$ tal que $\varepsilon_1 \doteq 0$ e para todo $k \geq 1$,*

$$\varepsilon_{k+1} \doteq \alpha_k(1 - \alpha_k)\langle v^k - w^k, p^k - \pi^k \rangle + (1 - \alpha_k)\varepsilon_k. \quad (6.7)$$

Então, para todo k , $\varepsilon_k \geq 0$ e

$$w^k \in \partial_{\varepsilon_k} f(p^k). \quad (6.8)$$

Demonstração. Provemos por indução em k . Para $k = 1$, temos, por definição, $\varepsilon_1 \doteq 0$. Além disso, como $w^1 = v^0 \in \partial f(\pi^0)$ e $p^1 \doteq \pi^0$, temos que $w^1 \in \partial_{\varepsilon_1} f(p^1)$. Suponha agora, por hipótese de indução, que os resultados são válidos para $k \geq 1$. Provemos inicialmente que $\varepsilon_{k+1} \geq 0$. Como $v^k \in \partial f(\pi^k)$,

$$f(\pi) \leq f(\pi^k) + \langle v^k, \pi - \pi^k \rangle, \quad \text{para todo } \pi \in \mathbb{R}^m. \quad (6.9)$$

Pela hipótese de indução, temos também que $w^k \in \partial_{\varepsilon_k} f(p^k)$, isto é,

$$f(\pi) \leq f(p^k) + \langle w^k, \pi - p^k \rangle + \varepsilon_k, \quad \text{para todo } \pi \in \mathbb{R}^m. \quad (6.10)$$

Assim (6.9) e (6.10), com $\pi \doteq p^k$ e $\pi \doteq \pi^k$, respectivamente, temos:

$$f(p^k) \leq f(\pi^k) + \langle v^k, p^k - \pi^k \rangle \quad \text{e} \quad f(\pi^k) \leq f(p^k) + \langle w^k, \pi^k - p^k \rangle + \varepsilon_k.$$

Somando-se as desigualdades acima, temos:

$$\langle v^k - w^k, \pi^k - p^k \rangle \leq \varepsilon_k.$$

Logo, por (6.7) e pela hipótese de indução,

$$\varepsilon_{k+1} \geq -\alpha_k(1 - \alpha_k)\varepsilon_k + (1 - \alpha_k)\varepsilon_k = (1 - \alpha_k)^2\varepsilon_k \geq 0.$$

Provemos agora que $w^{k+1} \in \partial_{\varepsilon_{k+1}} f(p^{k+1})$. Tomando-se uma combinação convexa das desigualdades (6.9) e (6.10), temos:

$$\begin{aligned} f(\pi) &\leq \alpha_k \left(f(\pi^k) + \langle v^k, \pi - \pi^k \rangle \right) + (1 - \alpha_k) \left(f(p^k) + \langle w^k, \pi - p^k \rangle + \varepsilon_k \right) \\ &= \alpha_k f(\pi^k) + (1 - \alpha_k) f(p^k) + \\ &+ \alpha_k \langle v^k, \pi - \pi^k \rangle + (1 - \alpha_k) \langle w^k, \pi - p^k \rangle + (1 - \alpha_k) \varepsilon_k. \end{aligned}$$

Pela concavidade de f , temos, ainda:

$$\begin{aligned}
f(\pi) &\leq f(\alpha_k \pi^k + (1 - \alpha_k) p^k) + \langle \alpha_k v^k + (1 - \alpha_k) w^k, \pi \rangle \\
&\quad - \alpha_k \langle v^k, \pi^k \rangle - (1 - \alpha_k) \langle w^k, p^k \rangle + (1 - \alpha_k) \varepsilon_k \\
&= f(p^{k+1}) + \langle w^{k+1}, \pi \rangle - \langle \alpha_k v^k, \pi^k \rangle - \langle (1 - \alpha_k) w^k, p^k \rangle + (1 - \alpha_k) \varepsilon_k \\
&= f(p^{k+1}) + \langle w^{k+1}, \pi - p^{k+1} \rangle + \langle w^{k+1}, p^{k+1} \rangle \\
&\quad - \langle \alpha_k v^k, \pi^k \rangle - \langle (1 - \alpha_k) w^k, p^k \rangle + (1 - \alpha_k) \varepsilon_k, \tag{6.11}
\end{aligned}$$

onde a primeira igualdade é dada por (6.2) e (6.6). Por (6.7), temos:

$$\begin{aligned}
\varepsilon_{k+1} &= \alpha_k (1 - \alpha_k) \langle v^k - w^k, p^k - \pi^k \rangle + (1 - \alpha_k) \varepsilon_k \\
&= \langle \alpha_k v^k, (1 - \alpha_k) (p^k - \pi^k) \rangle + \langle (1 - \alpha_k) w^k, \alpha_k (\pi^k - p^k) \rangle \\
&\quad + (1 - \alpha_k) \varepsilon_k. \tag{6.12}
\end{aligned}$$

Além disso, da definição de p^{k+1} ,

$$p^{k+1} - \pi^k = (1 - \alpha_k) (p^k - \pi^k) \quad \text{e} \quad p^{k+1} - p^k = \alpha_k (\pi^k - p^k). \tag{6.13}$$

Assim, substituindo (6.13) em (6.12), temos:

$$\begin{aligned}
\varepsilon_{k+1} &= \langle \alpha_k v^k, p^{k+1} - \pi^k \rangle + \langle (1 - \alpha_k) w^k, p^{k+1} - p^k \rangle + (1 - \alpha_k) \varepsilon_k \\
&= -\langle \alpha_k v^k, \pi^k \rangle + \langle \alpha_k v^k, p^{k+1} \rangle + \langle (1 - \alpha_k) w^k, p^{k+1} \rangle \\
&\quad - \langle (1 - \alpha_k) w^k, p^k \rangle + (1 - \alpha_k) \varepsilon_k \\
&= \langle w^{k+1}, p^{k+1} \rangle - \langle \alpha_k v^k, \pi^k \rangle - \langle (1 - \alpha_k) w^k, p^k \rangle + (1 - \alpha_k) \varepsilon_k.
\end{aligned}$$

Logo, substituindo essa última igualdade em (6.11), conclui-se que

$$f(\pi) \leq f(p^{k+1}) + \langle w^{k+1}, \pi - p^{k+1} \rangle + \varepsilon_{k+1},$$

ou seja, $w^{k+1} \in \partial_{\varepsilon_{k+1}} f(p^{k+1})$, o que completa a demonstração. \square

De maneira análoga a (6.1), (6.2) e (6.6), o escalar ε_k , definido em (6.7), também pode ser visto como uma combinação convexa, ou seja,

$$\varepsilon_{k+1} = \sum_{i=0}^k \lambda_{k,i} \sigma_{k-i}, \quad \text{com} \quad \sigma_k \doteq (1 - \alpha_k) \langle v^k - w^k, p^k - \pi^k \rangle, \forall k. \tag{6.14}$$

Uma outra mudança realizada pelo RVA está relacionada à condição utilizada no passo 8 do algoritmo 6.1. Para avaliarmos os ganhos produzidos pelos passos sérios, usamos uma condição do tipo Armijo

$$f(\pi^k) \geq f(\hat{\pi}^k) + m \delta_k, \tag{6.15}$$

onde $m \in]0, 1[$ e δ_k é o *ganho esperado* (que é uma versão do decréscimo nominal, visto na seção 3.4 e no capítulo 4, para funções côncavas).

Assim como no método de feixe, descrevemos δ_k utilizando um modelo linear por partes aproximado da função objetivo f . Nesse caso, ele é econômico, pois mantém apenas duas funções afins, a saber:

(i) Função afim gerada pelo último passo, dada por

$$f(\pi^k) + \langle v^k, \pi - \pi^k \rangle, \quad (6.16)$$

ou, equivalentemente, referindo-se ao último passo sério $\hat{\pi}^{k+1}$, por

$$f(\hat{\pi}^{k+1}) + \langle v^k, \pi - \hat{\pi}^{k+1} \rangle + e_k, \quad (6.17)$$

onde $e_k \doteq f(\pi^k) + \langle v^k, \hat{\pi}^{k+1} - \pi^k \rangle - f(\hat{\pi}^{k+1})$.

(ii) Função afim gerada pela agregação, definida por

$$f(p^k) + \langle w^k, \pi - p^k \rangle + \varepsilon_k, \quad (6.18)$$

ou, novamente referindo-se ao último passo sério $\hat{\pi}^{k+1}$, por

$$f(\hat{\pi}^{k+1}) + \langle w^k, \pi - \hat{\pi}^{k+1} \rangle + \hat{e}_k, \quad (6.19)$$

onde $\hat{e}_k \doteq f(p^k) + \langle w^k, \hat{\pi}^{k+1} - p^k \rangle - f(\hat{\pi}^{k+1}) + \varepsilon_k$.

Portanto, o modelo \tilde{f}_k utilizado na iteração k é escrito da seguinte forma:

$$\tilde{f}_k(\pi) \doteq \min \{ f(\pi^k) + \langle v^k, \pi - \pi^k \rangle, f(p^k) + \langle w^k, \pi - p^k \rangle + \varepsilon_k \}, \quad (6.20)$$

ou, de maneira equivalente,

$$\tilde{f}_k(\pi) \doteq f(\hat{\pi}^{k+1}) + \min \{ \langle v^k, \pi - \hat{\pi}^{k+1} \rangle + e_k, \langle w^k, \pi - \hat{\pi}^{k+1} \rangle + \hat{e}_k \}. \quad (6.21)$$

Conforme pode ser observado na figura 6.2.1, o hiperplano H correspondente a (6.16) suporta por cima a hipógrafe de f no ponto π^k , enquanto que o hiperplano H_ε associado a (6.18) suporta, com tolerância ε_k , a hipógrafe de f no ponto p^k . Isso implica que $\tilde{f}_k(\pi) \geq f(\pi)$ para todo $\pi \in \mathbb{R}^m$, ou seja, o modelo \tilde{f}_k não corta o gráfico da função objetivo f .

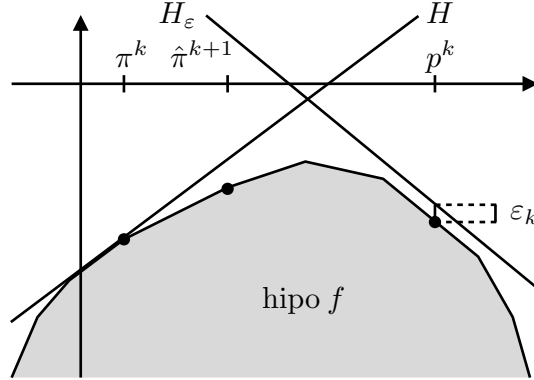


Figura 6.2.1: Modelo da função objetivo no algoritmo do volume.

Ademais, pela definição (3.14) e utilizando-se de (6.21), o ganho esperado pode ser escrito da seguinte maneira:

$$\delta_{k+1} \doteq \min \{ \langle v^k, \pi^{k+1} - \hat{\pi}^{k+1} \rangle + e_k, \langle w^k, \pi^{k+1} - \hat{\pi}^{k+1} \rangle + \hat{e}_k \}. \quad (6.22)$$

Agora que temos o modelo \tilde{f}_k em mente, estamos prontos para verificar a escolha do parâmetro α_k , necessário no passo 9 do algoritmo 6.1. Para isso, considere inicialmente o seguinte problema auxiliar:

$$\max_{\pi \in \mathbb{R}^m} \tilde{f}_k(\pi) - \frac{1}{2t_{k+1}} \|\pi - \hat{\pi}^{k+1}\|^2$$

e seja π' sua solução única. A condição de otimalidade nos diz que

$$\text{existe } w' \in \partial \tilde{f}_k(\pi') \text{ tal que } w' - \frac{1}{t_{k+1}}(\pi' - \hat{\pi}^{k+1}) = 0 \quad (6.23)$$

(ou, equivalentemente, que $\pi' = \hat{\pi}^{k+1} + t_{k+1}w'$). Como o modelo \tilde{f}_k é o mínimo de duas funções afins (6.16) e (6.18) (com gradiente v^k e w^k , respectivamente), temos, pela proposição 2.24, que

$$\partial \tilde{f}_k(\pi) = \{ \alpha v^k + (1 - \alpha)w^k : \alpha \in [0, 1] \}$$

para todo π . Portanto, (6.23) pode ser escrita como:

$$\text{existe } \alpha' \in [0, 1] \text{ tal que } \pi' = \hat{\pi}^{k+1} + t_{k+1}w' \text{ com } w' = \alpha'v^k + (1 - \alpha')w^k.$$

Note que se α_k for escolhido como sendo igual a α' , então w' será justamente igual a w^{k+1} . Logo, o próximo iterado π^{k+1} será exatamente o ponto π' que satisfaz a condição de otimalidade abaixo:

$$\text{existe } \alpha_k \in [0, 1] \text{ tal que } w^{k+1} \in \partial \tilde{f}_k(\pi^{k+1}) \text{ e } \pi^{k+1} = \hat{\pi}^{k+1} + t_{k+1}w^{k+1}.$$

O parâmetro α_k , por sua vez, pode ser encontrado resolvendo-se o dual do problema auxiliar, que é mais simples, por ser quadrático, diferenciável e com restrição de caixa:

$$\min_{\alpha \in [0,1]} \frac{t_{k+1}}{2} \|\alpha v^k + (1 - \alpha)w^k\|^2 + \alpha e_k + (1 - \alpha)\hat{e}_k. \quad (6.24)$$

Utilizando esse parâmetro α_k , obtemos também uma nova formulação do ganho esperado, mais concisa que (6.22):

$$\delta_{k+1} = t_{k+1} \|w^{k+1}\|^2 + \alpha_k e_k + (1 - \alpha_k)\hat{e}_k. \quad (6.25)$$

Voltemos agora à formulação do modelo \tilde{f}_k . Vimos, que a definição de \hat{e}_k exige o cômputo de $f(p^k)$, o qual não possuímos. Uma sugestão para resolver esse problema é usarmos uma aproximação de \hat{e}_k . Considerando que $f(\hat{\pi}^{k+1})$ é o melhor valor disponível da função objetivo, temos $\hat{e}_k \approx \langle w^k, \hat{\pi}^{k+1} - p^k \rangle + \varepsilon_k$. Apenas para sermos consistentes, aproximamos também o erro e_k e assim definimos os *erros de linearização aproximados* abaixo:

$$E_k \doteq \langle v^k, \hat{\pi}^{k+1} - \pi^k \rangle \quad \text{e} \quad \hat{E}_k \doteq \langle w^k, \hat{\pi}^{k+1} - p^k \rangle + \varepsilon_k. \quad (6.26)$$

Utilizando tais erros, podemos construir, conforme figura 6.2.2, um novo modelo da função objetivo:

$$\tilde{f}_k(\pi) \doteq f(\hat{\pi}^{k+1}) + \min \{ \langle v^k, \pi - \hat{\pi}^{k+1} \rangle + E_k, \langle w^k, \pi - \hat{\pi}^{k+1} \rangle + \hat{E}_k \}.$$

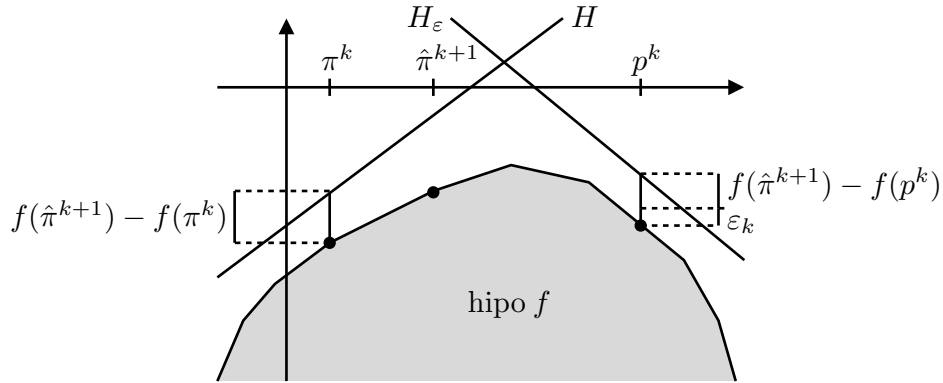


Figura 6.2.2: Modelo da função objetivo com aproximações.

Conseqüentemente, o problema dual (6.24) deve ser substituído por:

$$\min_{\alpha \in [0,1]} \frac{t_{k+1}}{2} \|\alpha v^k + (1 - \alpha)w^k\|^2 + \alpha E_k + (1 - \alpha)\hat{E}_k,$$

e podemos dar também uma nova expressão para o ganho esperado:

$$\delta_{k+1} \doteq t_{k+1}\|w^{k+1}\|^2 + \alpha_k E_k + (1 - \alpha_k)\hat{E}_k. \quad (6.27)$$

Proposição 6.8. *O ganho esperado, definido em (6.27), pode ser assim escrito:*

$$\delta_{k+1} \doteq t_{k+1}\|w^{k+1}\|^2 + \langle w^{k+1}, \hat{\pi}^{k+1} - p^{k+1} \rangle + \varepsilon_{k+1}, \quad (6.28)$$

onde E_k e \hat{E}_k são definidos em (6.26).

Demonstração. Mostraremos que

$$\alpha_k E_k + (1 - \alpha_k)\hat{E}_k = \langle w^{k+1}, \hat{\pi}^{k+1} - p^{k+1} \rangle + \varepsilon_{k+1}.$$

Pelas definições (6.26), temos:

$$\begin{aligned} & \alpha_k E_k + (1 - \alpha_k)\hat{E}_k \\ &= \alpha_k \langle v^k, \hat{\pi}^{k+1} - \pi^k \rangle + (1 - \alpha_k)(\langle w^k, \hat{\pi}^{k+1} - p^k \rangle + \varepsilon_k) \\ &= \langle \alpha_k v^k + (1 - \alpha_k)w^k, \hat{\pi}^{k+1} \rangle - \alpha_k \langle v^k, \pi^k \rangle - (1 - \alpha_k)\langle w^k, p^k \rangle + (1 - \alpha_k)\varepsilon_k \\ &= \langle w^{k+1}, \hat{\pi}^{k+1} \rangle - \alpha_k \langle v^k, \pi^k \rangle - (1 - \alpha_k)\langle w^k, p^k \rangle \\ & \quad + \langle w^{k+1}, p^{k+1} \rangle - \langle w^{k+1}, p^{k+1} \rangle + (1 - \alpha_k)\varepsilon_k \\ &= \langle w^{k+1}, \hat{\pi}^{k+1} - p^{k+1} \rangle - \alpha_k \langle v^k, \pi^k \rangle - (1 - \alpha_k)\langle w^k, p^k \rangle \\ & \quad + \langle \alpha_k v^k + (1 - \alpha_k)w^k, p^{k+1} \rangle + (1 - \alpha_k)\varepsilon_k \\ &= \langle w^{k+1}, \hat{\pi}^{k+1} - p^{k+1} \rangle + \alpha_k \langle v^k, p^{k+1} - \pi^k \rangle \\ & \quad + (1 - \alpha_k)\langle w^k, p^{k+1} - p^k \rangle + (1 - \alpha_k)\varepsilon_k, \end{aligned} \quad (6.29)$$

sendo que a terceira e a quarta igualdades são dadas válidas por (6.2). Pela definição (6.6), temos, novamente:

$$p^{k+1} - \pi^k = (1 - \alpha_k)(p^k - \pi^k) \quad \text{e} \quad p^{k+1} - p^k = \alpha_k(\pi^k - p^k),$$

e substituindo-os em (6.29), temos:

$$\begin{aligned} & \alpha_k E_k + (1 - \alpha_k)\hat{E}_k \\ &= \langle w^{k+1}, \hat{\pi}^{k+1} - p^{k+1} \rangle + \alpha_k(1 - \alpha_k)\langle v^k, p^k - \pi^k \rangle \\ & \quad + \alpha_k(1 - \alpha_k)\langle w^k, \pi^k - p^k \rangle + (1 - \alpha_k)\varepsilon_k \\ &= \langle w^{k+1}, \hat{\pi}^{k+1} - p^{k+1} \rangle + \alpha_k(1 - \alpha_k)\langle v^k - w^k, p^k - \pi^k \rangle + (1 - \alpha_k)\varepsilon_k \\ &= \langle w^{k+1}, \hat{\pi}^{k+1} - p^{k+1} \rangle + \varepsilon_{k+1}, \end{aligned}$$

onde a última igualdade é válida pela definição (6.7). \square

Conforme (6.15), para que a seqüência $\{f(\hat{\pi}^k)\}$ seja monótona, redefinimos o ganho esperado de modo que ele seja sempre positivo, ou seja,

$$\delta_k \doteq t_k \|w^k\|^2 + |\langle w^k, \hat{\pi}^k - p^k \rangle| + \varepsilon_k. \quad (6.30)$$

Essa será a definição utilizada no nosso algoritmo.

Devemos observar, agora, que com as aproximações, o hiperplano H_ε , correspondente a função afim agregada, poderá cortar o gráfico de f , o que não acontecia antes. Isso ocorre quando $f(p^k) - f(\hat{\pi}^{k+1}) > \varepsilon_k$. Observe a figura 6.2 para visualizar tal fato. Em [BMS02, seção 6.2], mostrou-se uma forma de organizar os cálculos para que o modelo não corte o gráfico de f . Como resultado, obtiveram um algoritmo equivalente ao método de feixes, denominado BVA, com número máximo de elementos do feixe igual a dois.

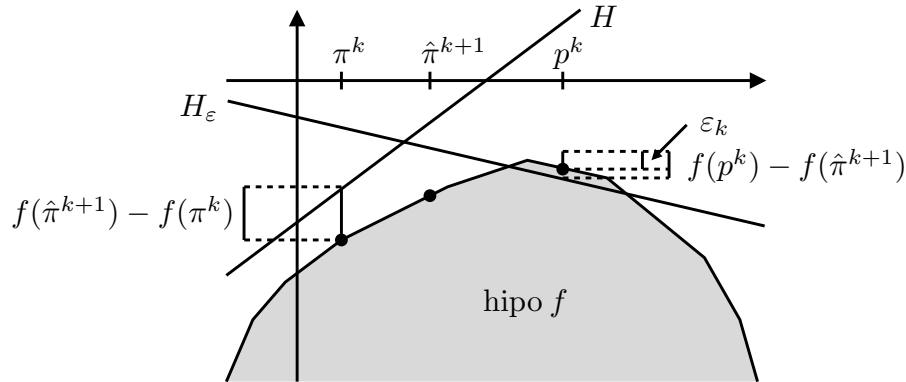


Figura 6.2.3: O modelo pode cortar o gráfico da função objetivo.

Finalmente, podemos estabelecer um critério de parada para o RVA herdando a idéia do método de feixes, isto é, utilizando uma tolerância $\text{tol} > 0$ dada. Em outras palavras, o algoritmo pára quando $\delta_k \leq \text{tol}$. Conforme (6.28), uma outra possibilidade é, dadas duas tolerâncias tol_w e tol_ε , usarmos as condições

$$\|w^k\|^2 \leq \text{tol}_w \quad \text{e} \quad |\langle w^k, \hat{\pi}^k - p^k \rangle| + \varepsilon_k \leq \text{tol}_\varepsilon, \quad (6.31)$$

que possuem a vantagem de não dependerem de t_k , cujo valor pode ser muito pequeno ao longo das iterações. O RVA é resumido a seguir.

Algoritmo 6.9. *Algoritmo do Volume Revisado*

1. Sejam $m \in]0, 1[$, tolerâncias tol_w , tol_ε e ponto inicial $\pi^0 \in \mathbb{R}^m$ dados.
2. (Caixa preta) Compute $x^0 \in \Psi(\pi^0)$, $f(\pi^0)$ e $v^0 \doteq Ax^0 - b$.
3. (Inicialização) Sejam $z^1 \doteq x^0$, $w^1 \doteq v^0$, $\hat{\pi}^1 \doteq \pi^0$, $p^1 \doteq \pi^0$, $\varepsilon_1 \doteq 0$, $E_0 \doteq 0$ e $\hat{E}_0 \doteq 0$ e $k = 1$.
4. (Tamanho de passo inicial) Tome t_1 adequado.
5. (Atualização do candidato) Faça $\pi^k \doteq \hat{\pi}^k + t_k w^k$.
6. (Ganho esperado) Compute $\delta_k \doteq t_k \|w^k\|^2 + |\langle w^k, \hat{\pi}^k - p^k \rangle| + \varepsilon_k$.
7. (Critério de parada) Se $\|w^k\|^2 \leq \text{tol}_w$ e $|\langle w^k, \hat{\pi}^k - p^k \rangle| + \varepsilon_k \leq \text{tol}_\varepsilon$, pare.
8. (Caixa preta) Compute $x^k \in \Psi(\pi^k)$, $f(\pi^k)$ e $v^k \doteq Ax^k - b$.
9. (Teste de crescimento) Se $f(\pi^k) \geq f(\hat{\pi}^k) + m\delta_k$:
Então faça $\hat{\pi}^{k+1} \doteq \pi^k$ (*passo sério*).
Senão, $\hat{\pi}^{k+1} \doteq \hat{\pi}^k$ (*passo nulo*).
10. (Busca linear) Tome um tamanho de passo t_{k+1} adequado.
11. (Erros) Compute $E_k \doteq \langle v^k, \hat{\pi}^k - \pi^k \rangle$ e $\hat{E}_k \doteq \langle w^k, \hat{\pi}^k - p^k \rangle + \varepsilon_k$.
12. (Subproblema) Resolva

$$\alpha_k \in \underset{\alpha \in [0,1]}{\text{argmin}} \frac{t_{k+1}}{2} \|\alpha v^k + (1 - \alpha)w^k\|^2 + \alpha E_k + (1 - \alpha)\hat{E}_k.$$

13. (Atualização de vetores) Defina

$$\begin{aligned} z^{k+1} &\doteq \alpha_k x^k + (1 - \alpha_k)z^k, & w^{k+1} &\doteq \alpha_k v^k + (1 - \alpha_k)w^k, \\ p^{k+1} &\doteq \alpha_k \pi^k + (1 - \alpha_k)p^k, & \varepsilon_{k+1} &\doteq \alpha_k \varepsilon_k + (1 - \alpha_k)\varepsilon_k. \end{aligned}$$

14. (Loop) Faça $k = k + 1$ e vá para o passo 5.

Mencionamos ainda que o subproblema do passo 12 do algoritmo acima não deve ser implementado apenas resolvendo-se projeção do ponto mínimo no intervalo $[0, 1]$. Note que devemos evitar o valor 0 para α_k , pois, do contrário, não atualizamos nenhum vetor no passo 13. Da mesma forma, o valor 1 para α_k indica que ignoramos completamente as iterações anteriores. Evitamos tais situações com a idéia dada no algoritmo 6.3. Mais especificamente, incorporamos um limitante superior $\alpha_k^{\max} \in]0, 1[$ no cálculo de α_k , conforme podemos observar no algoritmo abaixo.

Algoritmo 6.10. *Escolha de α_k no Algoritmo do Volume Revisado.*

- (a) Seja $\alpha_k^{\max} \in]0, 1[$ um limitante superior e $\varsigma \in]0, 1[$ uma constante.
- (b) Compute $\bar{\alpha} \doteq \operatorname{argmin}_{\alpha} \frac{t_{k+1}}{2} \|\alpha v^k + (1 - \alpha)w^k\|^2 + \alpha E_k + (1 - \alpha)\hat{E}_k$.
- (c) Se $\bar{\alpha} > 0$, então faça $\alpha_k \doteq \min\{\bar{\alpha}, \alpha_k^{\max}\}$.
Senão, faça $\alpha_k \doteq \alpha_k^{\max}/10$.
- (d) Se depois de um certo número de iterações, a função objetivo dual não cresceu suficientemente, faça $\alpha_{k+1}^{\max} = \varsigma \alpha_k^{\max}$.
Senão, faça $\alpha_{k+1}^{\max} = \alpha_k^{\max}$.

Enunciaremos agora alguns resultados de convergência que foram obtidos em [BMS02]. Considere o conjunto de índices

$$K_S \doteq \{k: \text{o passo dado na iteração } k \text{ é sério}\}$$

e seja Π^* o conjunto de soluções ótimas duais.

Lema 6.11. *Considere o problema (1.1) com Ψ limitado. Então, as seqüências $\{x^k\}$, $\{v^k\}$, $\{z^k\}$ e $\{w^k\}$ geradas pelo algoritmo 6.9 são limitadas. Ademais, se $\{\pi^k\}$ é limitada, então $\{p^k\}$ também é limitada.*

Demonstração. Como $x^k \in \Psi$ para todo k e Ψ é limitado, temos que $\{x^k\}$ é limitada. Além disso, por definição, $v^k \doteq Ax^k - b$, logo $\{v^k\}$ também é limitada. Pela proposição 6.6, $\{z^k\}$ e $\{w^k\}$ são também limitadas. Finalmente, por (6.6), se $\{\pi^k\}$ é limitada, então, $\{p^k\}$ também é limitada. \square

Lema 6.12. *Suponha que o algoritmo 6.9 gera infinitos passos sérios, isto é, $|K_S| = +\infty$, e que $\lim_{k \in K_S} f(\hat{\pi}^k) < +\infty$. Então, as seguintes afirmações são verdadeiras:*

- (a) $\lim_{k \in K_S} \delta_k = 0$ e $\lim_{k \in K_S} \varepsilon_k = 0$.
- (b) *Suponha que*
$$\sum_{k \in K_S} t_k = +\infty. \tag{6.32}$$

$$\text{Então, } \lim_{k \in K_S} w^k = 0.$$

- (c) *Suponha, juntamente com (6.32), que*

$$\text{a seqüência } \{t_k\} \text{ é limitada por } M_t \geq 0, \text{ e} \tag{6.33}$$

$$\Pi^* \neq \emptyset. \tag{6.34}$$

Então, a seqüência $\{\hat{\pi}^k\}$ é limitada.

(d) Suponha, juntamente com (6.32), (6.33) e (6.34), que

$$a \text{ seqüência } \{w^k\} \text{ é limitada por } M_w \geq 0. \quad (6.35)$$

Então, a seqüência $\{\pi^k\}$ é limitada.

Demonstração.

(a) Para todo $k \in K_S$, $\hat{\pi}^{k+1} = \pi^k$. Logo, somando-se (6.15) em K_S , temos:

$$m \sum_{k \in K_S} \delta_k \leq \sum_{k \in K_S} (f(\hat{\pi}^{k+1}) - f(\hat{\pi}^k)) = \lim_{k \in K_S} f(\hat{\pi}^k) - f(\hat{\pi}^1) < +\infty,$$

onde a última desigualdade é dada por hipótese. Logo, $\sum_{k \in K_S} \delta_k < +\infty$, o que implica que $\lim_{k \in K_S} \delta_k = 0$. Pela proposição 6.8, temos que $0 \leq \varepsilon_k \leq \delta_k$. Assim, $0 \leq \lim_{k \in K_S} \varepsilon_k \leq \lim_{k \in K_S} \delta_k = 0$, o que completa a demonstração.

(b) Pela proposição 6.8, $t_k \|w^k\|^2 \leq \delta_k$ para todo k . Portanto, pelo item (a), temos:

$$\sum_{k \in K_S} t_k \|w^k\|^2 \leq \sum_{k \in K_S} \delta_k < +\infty,$$

o que, juntamente com a hipótese (6.32), mostra que $\lim_{k \in K_S} w^k = 0$.

(c) Seja $p^* \in \Pi^* \neq \emptyset$. Por (6.8), podemos escrever:

$$\begin{aligned} f(p^*) &\leq f(p^k) + \langle w^k, p^* - p^k \rangle + \varepsilon_k, \quad \forall k \in K_S \\ \Leftrightarrow \langle w^k, p^k - p^* \rangle &\leq f(p^k) - f(p^*) + \varepsilon_k, \quad \forall k \in K_S \\ \Rightarrow \langle w^k, p^k - p^* \rangle &\leq \varepsilon_k, \quad \forall k \in K_S. \end{aligned} \quad (6.36)$$

Além disso, como $\hat{\pi}^{k+1} = \pi^k$ quando o passo é sério e $\hat{\pi}^{k+1} = \hat{\pi}^k$ quando é nulo, temos que $\hat{\pi}^{k+1} - \hat{\pi}^k$ admite valores $t_k w^k$ ou 0. Conseqüentemente,

$$\begin{aligned} \|\hat{\pi}^{k+1} - p^*\|^2 &= \|\hat{\pi}^{k+1} - \hat{\pi}^k + \hat{\pi}^k - p^*\|^2 \\ &= \|\hat{\pi}^k - p^*\|^2 + \|\hat{\pi}^{k+1} - \hat{\pi}^k\|^2 + 2\langle \hat{\pi}^{k+1} - \hat{\pi}^k, \hat{\pi}^k - p^* \rangle \\ &\leq \|\hat{\pi}^k - p^*\|^2 + t_k^2 \|w^k\|^2 + 2t_k \langle w^k, \hat{\pi}^k - p^* \rangle \\ &= \|\hat{\pi}^k - p^*\|^2 + t_k (t_k \|w^k\|^2 \\ &\quad + 2\langle w^k, \hat{\pi}^k - p^* \rangle + 2\langle w^k, p^k - p^* \rangle) \\ &\leq \|\hat{\pi}^k - p^*\|^2 + t_k (t_k \|w^k\|^2 + 2\langle w^k, \hat{\pi}^k - p^* \rangle + 2\varepsilon_k) \\ &\leq \|\hat{\pi}^k - p^*\|^2 + 2t_k (t_k \|w^k\|^2 + \langle w^k, \hat{\pi}^k - p^* \rangle + \varepsilon_k), \end{aligned}$$

sendo que a penúltima desigualdade é válida por (6.36). Ademais, por (6.28) e pela hipótese (6.33), para todo $k \in K_S$, temos:

$$\|\hat{\pi}^{k+1} - p^*\|^2 \leq \|\hat{\pi}^k - p^*\|^2 + 2t_k \delta_k \leq \|\hat{\pi}^k - p^*\|^2 + 2M_t \delta_k.$$

Considere agora $K_S \doteq \{k_0, k_1, \dots\}$, com $k_0 < k_1 < \dots$. Para todo $i \geq 0$, sabemos que $\hat{\pi}^{k_{i+1}} = \hat{\pi}^{k(i+1)}$. Então, para qualquer $i \geq 0$, temos:

$$\begin{aligned} \sum_{0 \leq j \leq i} \|\hat{\pi}^{k_{j+1}} - p^*\|^2 &\leq \sum_{0 \leq j \leq i} \|\hat{\pi}^{k_j} - p^*\|^2 + 2M_t \sum_{0 \leq j \leq i} \delta_{k_j} \\ \Leftrightarrow \|\hat{\pi}^{k_{i+1}} - p^*\|^2 &\leq \|\hat{\pi}^{k_0} - p^*\|^2 + 2M_t \sum_{0 \leq j \leq i} \delta_{k_j} \\ \Rightarrow \|\hat{\pi}^{k_{i+1}} - p^*\|^2 &\leq \|\hat{\pi}^{k_0} - p^*\|^2 + 2M_t \sum_{k \in K_S} \delta_k \end{aligned}$$

Pelo item (a), temos $\sum_{k \in K_S} \delta_k < +\infty$. Logo, essa última desigualdade mostra que para todo $i \geq 0$ (ou, equivalentemente, para todo $k_i \in K_S$), $\|\hat{\pi}^{k_{i+1}} - p^*\|^2 < +\infty$. Portanto, a seqüência $\{\hat{\pi}^k\}_{k \in K_S}$ é limitada. Como todos os elementos de $\{\hat{\pi}^k\}$ estão contidos no conjunto de elementos de $\{\hat{\pi}^k\}_{k \in K_S}$, concluímos que $\{\hat{\pi}^k\}$ é limitada.

(d) Pelo item (c), basta mostrarmos que a seqüência $\{\pi^k\}_{k \notin K_S}$ é limitada. Considere novamente que $K_S \doteq \{k_0, k_1, \dots\}$, com $k_0 < k_1 < \dots$. Sabemos que se k é tal que $k_i < k < k_{i+1}$ para algum $i \geq 0$, então na k -ésima iteração ocorreu um passo nulo. Por (6.33) e (6.35), temos:

$$\|\pi^k - \hat{\pi}^k\| = t_k \|w^k\| \leq M_t M_w.$$

para todo k . Logo, para qualquer $i \geq 0$,

$$\{\pi^k\}_{k_i < k < k_{i+1}} \subset \bar{B}(\hat{\pi}^k, M_t M_w),$$

o que implica que a seqüência $\{\pi^k\}_{k \notin K_S}$ é limitada. \square

Teorema 6.13. *Considere o problema (1.1) com Ψ limitado. Suponha que $\Pi^* \neq \emptyset$ e que $\text{dom } f = \mathbb{R}^m$. Suponha ainda que o algoritmo 6.9 produz infinitos passos sérios, isto é, $|K_S| = +\infty$. Se os passos t_k são escolhidos de forma que as hipóteses (6.32) e (6.33) são verdadeiras, então a seqüência $\{p^k\}_{k \in K_S}$ é maximizante.*

Demonstração. Pelo lema 6.11, a hipótese (6.35) é verdadeira pois Ψ é limitado. Como também temos (6.32), (6.33) e (6.34) válidas por serem

hipóteses deste teorema, concluímos, pelo item (d) do lema 6.12, que $\{\pi^k\}$ é limitada. Utilizando novamente o lema 6.11, temos também que $\{p^k\}$ é limitada. Logo, a subsequência (infinita) $\{p^k\}_{k \in K_S}$ também é limitada, o que implica que ela possui um ponto de acumulação p^* .

Ademais, pelos itens (a) e (b) do lema 6.12, temos $\lim_{k \in K_S} \varepsilon_k = 0$ e $\lim_{k \in K_S} w^k = 0$. Sabemos também, por (6.8), que $w^k \in \partial_{\varepsilon_k} f(p^k)$ e pelo lema 2.30, o ε_k -subdiferencial é um gráfico fechado. Logo, temos:

$$\begin{cases} \{p^k\}_{k \in K_S} \rightarrow p^* \\ \{\varepsilon_k\}_{k \in K_S} \rightarrow 0 \\ \{w^k\}_{k \in K_S} \rightarrow 0, \text{ com } w^k \in \partial_{\varepsilon_k} f(p^k) \end{cases} \implies 0 \in \partial f(p^*),$$

o que, juntamente com a continuidade de f , resulta que $\{p^k\}_{k \in K_S}$ é, de fato, maximizante. \square

Teorema 6.14. *Suponha que o algoritmo 6.9 com $\text{tol} = 0$ gera um último passo sério $\hat{\pi}^{k_f}$, seguido de infinitos passos nulos $\{\pi^{k+1}\}_{k > k_f}$. Então, as seguintes afirmações são válidas:*

- (a) *Se as hipóteses (6.33) e (6.35) são válidas, então $\{\pi^k\}$ é limitada.*
 (b) *Suponha, além das condições dadas em (a), que*

$$\{\pi^k\} \rightarrow \pi^* \quad e \quad \{v^k\} \rightarrow v^*. \quad (6.37)$$

Ademais, tome $m \in]0, 1/2[$ e suponha que existe

$$t_* > 0 \text{ tal que } \{t_k\}_{k \notin K_S} \rightarrow t_*. \quad (6.38)$$

Então, π^ e $\hat{\pi}^{k_f}$ resolvem (1.5).*

Demonstração.

- (a) Para todo $k > k_f$, como ocorrem apenas passos nulos, $\hat{\pi}^k = \hat{\pi}^{k_f}$. Assim,

$$\pi^k = \hat{\pi}^{k_f} + t_k w^k \implies \|\pi^k\| \leq \|\hat{\pi}^{k_f}\| + t_k \|w^k\| \leq \|\hat{\pi}^{k_f}\| + t_* M_w$$

ou seja, $\{\pi^k\}$ é uma seqüência limitada.

- (b) Pelas definições (6.2) e (6.6) e pela hipótese (6.37),

$$\lim_{k \rightarrow \infty} p^k = \lim_{k \rightarrow \infty} \pi^k = \pi^* \quad e \quad \lim_{k \rightarrow \infty} w^k = \lim_{k \rightarrow \infty} v^k = v^*. \quad (6.39)$$

Novamente, como para todo $k > k_f$, como ocorrem apenas passos nulos,

$$\pi^* = \lim_{k \rightarrow \infty} p^k = \lim_{k \rightarrow \infty} \pi^k = \lim_{k \rightarrow \infty} (\hat{\pi}^{k_f} + t_k w^k) = \hat{\pi}^{k_f} + t_* v^*. \quad (6.40)$$

Além disso, por (6.14), como $\lim_{k \rightarrow \infty} \sigma_k = 0$, temos:

$$\lim_{k \rightarrow \infty} \varepsilon_k = 0. \quad (6.41)$$

Com isso, pela definição (6.28) e pelos resultados (6.39), (6.40) e (6.41),

$$\lim_{k \rightarrow \infty} \delta_k = \lim_{k \rightarrow \infty} t_k \|w^k\|^2 + |\langle w^k, \hat{\pi}^{k_f} - p^k \rangle| + \varepsilon_k = 2t_* \|v^*\|^2.$$

Sabemos que $f(\pi^k) < f(\hat{\pi}^{k_f}) + m\delta_k$ para todo $k > k_f$. Tomando o limite em k , temos:

$$f(\pi^*) \leq f(\hat{\pi}^{k_f}) + 2mt_* \|v^*\|^2. \quad (6.42)$$

Ademais, pelo lema 2.30, o ε_k -subdiferencial é fechado. Logo, por (6.8), (6.39) e (6.41), temos:

$$\begin{cases} p^k \rightarrow p^* \\ \varepsilon_k \rightarrow 0 \\ w^k \rightarrow v^*, \text{ com } w^k \in \partial_{\varepsilon_k} f(p^k) \end{cases} \implies v^* \in \partial f(\pi^*),$$

o que implica, juntamente com (6.40), que

$$f(\hat{\pi}^{k_f}) \leq f(\pi^*) + \langle v^*, \hat{\pi}^{k_f} - \pi^* \rangle = f(\pi^*) - t_* \|v^*\|^2. \quad (6.43)$$

Somando-se as desigualdades (6.42) e (6.43), temos:

$$(1 - 2m)t_* \|v^*\|^2 \leq 0.$$

Por hipótese, $m < 1/2$. Logo, essa última desigualdade é válida apenas quando $v^* = 0$, em outras palavras, quando π^* é solução de (1.5). Por (6.43), conclui-se também que $f(\hat{\pi}^{k_f}) \leq f(\pi^*)$, o que implica que $f(\hat{\pi}^{k_f})$ também resolve (1.5). \square

Note que as hipóteses (6.37) são fortes, pois elas requerem que as seqüências $\{\pi^k\}$ e $\{v^k\}$ tenham apenas um único ponto de acumulação. Conforme menciona-se em [BMS02, seção 5.2], isso mostra que a análise de convergência do RVA não está completa. Introduzindo agora um critério de parada (isto é, tomando $\text{tol} > 0$) e considerando a análise das seqüências primais de (6.1), chegamos a um outro resultado.

Teorema 6.15. *Considere o algoritmo 6.9 com $\text{tol} > 0$ (em particular, com $\text{tol}_w > 0$). Suponha que z^{k_f} é o último ponto z^k gerado. Se o posto da matriz A for igual a m , então existe $L \in]0, +\infty[$, dependente da estrutura do problema (1.1), tal que, para todo x^* minimizador de (1.1), vale $\|x^* - z^{k_f}\| \leq L \text{tol}_w$. Em outras palavras, z^{k_f} está a uma distância menor que $L \text{tol}_w$ de soluções ótimas.*

Demonstração. Como x^* é viável, $Ax^* = b$. Logo, pela proposição 6.2,

$$A(z^{k_f} - x^*) = Az^{k_f} - b = w^{k_f}.$$

Portanto,

$$\|x^* - z^{k_f}\| = \|(A^T A)^{-1} A^T A(x^* - z^{k_f})\| = \|(A^T A)^{-1} A^T w^{k_f}\| \leq L \|w^{k_f}\|,$$

onde $L \doteq \|(A^T A)^{-1} A^T\|$, o que implica, pelo critério (6.31), que

$$\|x^* - z^{k_f}\| \leq L \text{tol}_w,$$

conforme queríamos. \square

6.3 Algoritmo do Volume com Variação do Alvo

Nesta seção realizaremos uma outra modificação no VA (algoritmo 6.1) com o intuito de possibilitar a análise de convergência do algoritmo. Denominamos essa técnica de *algoritmo do volume com variação do alvo* (VA+VTVM) e sua idéia principal consiste em utilizar uma variação do tamanho de passo definido em (6.3). Mais especificamente, consideramos, para uma iteração k ,

$$t_k \doteq \frac{\beta(\xi_\ell - f(\pi^k))}{\|w^k\|^2}, \quad (6.44)$$

onde w^k é definido em (6.2) e ξ_ℓ é um *alvo*, isto é, um limitante superior do valor ótimo f_* da função, que, ao contrário de (6.3), varia a cada iteração ℓ . Além disso, nesse caso, requerimos $\beta \in]0, 1]$. Baseado-se em [SCT00, SL04], referências principais desta seção, definimos abaixo algumas notações, de modo a facilitar a compreensão do algoritmo 6.20.

Notações 6.16. *Considere os seguintes contadores:*

- k : contador geral de iterações do algoritmo;
- ℓ : contador de iterações do laço externo;
- τ : contador de iterações do laço interno corrente;
- γ : contador do número de passos nulos consecutivos.

Notações 6.17. *Para toda iteração geral k , temos:*

- \bar{f}_k : melhor valor da função objetivo obtido até a iteração k ;
- χ_k : melhora (definido por $\max\{f(\pi^{k+1}) - \bar{f}_k, 0\}$) acumulada no laço interno corrente.

Notações 6.18. Para toda iteração externa ℓ , temos:

- ξ_ℓ : valor do alvo (isto é, limitante superior de f_*) na iteração ℓ ;
- ε_ℓ : tolerância de aceitação para declarar que o valor \bar{f}_k é suficientemente perto do alvo ξ_ℓ .

Notações 6.19. Considere os seguintes parâmetros:

- k_{\max} : número máximo de iterações gerais do algoritmo;
- tol : tolerância para declarar que um subgradiente v^k ou w^k é suficientemente perto de zero;
- ε : valor mínimo para a tolerância ε_ℓ ;
- σ : intervalo de tolerância de aceitação usada na atualização de ε_ℓ ;
- η : fração da melhora acumulada χ_k usada na atualização do alvo ξ_ℓ ;
- $\bar{\tau}$: número máximo de iterações no laço interno sem atualizar o alvo ξ_ℓ ;
- $\bar{\gamma}$: número máximo de passos nulos permitidos no laço interno corrente.

A atualização do alvo ξ_ℓ é dada seguindo-se um princípio semelhante aos métodos de região de confiança (ver, por exemplo, [NW99, capítulo 4] ou [BGLS03, seção 6.1]). Mais especificamente, no nosso caso, devemos atualizar ξ_ℓ em dois casos distintos: (i) quando \bar{f}_k é maior ou suficientemente próximo de ξ_ℓ , o que indica que ξ_ℓ não se comporta como um limitante superior adequado de f_* por ser pequeno demais; (ii) se $\tau \geq \bar{\tau}$ ou $\gamma \geq \bar{\gamma}$, ou seja, quando está demorando muito para atualizar o alvo, o que indica que ξ_ℓ é muito grande comparado aos valores $f(\pi^k)$ gerados.

Com essa idéia, as notações anteriores e o algoritmo 6.1 do VA em mente, podemos verificar agora a estrutura do método modificado.

Algoritmo 6.20. Algoritmo do Volume com Variação do Alvo

1. Inicialização:

- (a) Defina $\varepsilon, \text{tol} \geq 0, \beta \in]0, 1], \sigma \in]0, 1/3], \bar{\tau}, \bar{\gamma}, k_{\max}$ e $\eta \in]0, 1]$.
- (b) Escolha um ponto dual inicial π^1 .
- (c) Determine $f(\pi^1), x^1 \in \Psi(\pi^1)$ e $v^1 \doteq Ax^1 - b$.
- (d) Inicialize $w^0 \doteq v^1, z^0 \doteq x^1$ e $\bar{f}_1 \doteq f(\pi^1)$.
- (e) Se $\|v^1\| \leq \text{tol}$, pare.
- (f) Tome $\ell = k = 1$ e $\tau = \gamma = \chi_1 \doteq 0$.
- (g) Seja $\xi_1 \doteq \min\{UB, f(\pi^1) + \|v^1\|^2/2\}$, onde UB é um limitante superior para f_* , possivelmente $+\infty$.

- (h) Seja $\varepsilon_1 \doteq \sigma(\xi_1 - f(\pi^1))$.
2. Atualização dos iterados dual e primal:
- (a) Dado α tal que $0 < \beta \leq \alpha \leq 1$, defina
- $$w^k \doteq \alpha v^k + (1 - \alpha)w^{k-1} \quad \text{e} \quad z^k \doteq \alpha x^k + (1 - \alpha)z^{k-1}. \quad (6.45)$$
- (b) Se $\|w^k\| \leq \text{tol}$, então faça $w^k \doteq v^k$.
- (c) Compute o tamanho de passo $t_k \doteq \frac{\beta(\xi_\ell - f(\pi^k))}{\|w^k\|^2}$.
- (d) Atualize o próximo iterado dual: $\pi^{k+1} \doteq \pi^k + t_k w^k$.
- (e) Tome $\tau = \tau + 1$ e $k = k + 1$.
- (f) Compute $f(\pi^k)$, $x^k \in \Psi(\pi^k)$ e $v^k \doteq Ax^k - b$.
- (g) Faça $\chi_k \doteq \chi_{k-1} + \max\{(f(\pi^k) - \bar{f}_{k-1}), 0\}$.
3. Verificação do tipo de passo – sério ou nulo:
- (a) Se $f(\pi^k) > \bar{f}_{k-1}$, então vá para o passo 4.
Senão, vá para o passo 5.
4. Passo sério:
- (a) Se $k > k_{\max}$ ou $\|v^k\| \leq \text{tol}$, pare.
- (b) Tome $\bar{f}_k \doteq f(\pi^k)$ e $\gamma = 0$.
- (c) Se $\bar{f}_k \geq \xi_\ell - \varepsilon_\ell$, vá para o passo 6.
Senão, se $\tau \geq \bar{\tau}$ vá para o passo 7.
Senão, vá para o passo 2.
5. Passo nulo:
- (a) Se $k > k_{\max}$, pare.
- (b) Tome $\bar{f}_k \doteq \bar{f}_{k-1}$ e $\gamma = \gamma + 1$.
- (c) Se $\gamma \geq \bar{\gamma}$ ou $\tau \geq \bar{\tau}$, vá para o passo 7.
Senão, vá para o passo 2.
6. Aumento do valor do alvo (caso em que $\bar{f}_k \geq \xi_\ell - \varepsilon_\ell$):
- (a) Compute
- $$\xi_{\ell+1} \doteq \bar{f}_k + \varepsilon_\ell + \eta \chi_k \quad \text{e} \quad \varepsilon_{\ell+1} \doteq \max\{\sigma(\xi_{\ell+1} - \bar{f}_k), \varepsilon\}.$$

(b) Faça $\tau = 0$, $\chi_k \doteq 0$ e $\ell = \ell + 1$, e vá para o passo 2.

7. Diminuição do valor do alvo (nesse caso, $\bar{f}_k < \xi_\ell - \varepsilon_\ell$):

(a) Compute

$$\xi_{\ell+1} \doteq \frac{\bar{f}_k + \varepsilon_\ell + \xi_\ell}{2} \quad \text{e} \quad \varepsilon_{\ell+1} \doteq \max\{\sigma(\xi_{\ell+1} - \bar{f}_k), \varepsilon\}.$$

(b) Faça $\gamma = \tau = \chi_k \doteq 0$, $\ell = \ell + 1$ e vá para o passo 2.

Observe que, ao contrário do VA original, esse método não utiliza centros de estabilização. Além disso, sob o ponto de vista computacional, algumas modificações no algoritmo acima podem ser realizadas. Referimo-nos a [SCT00, observações 1 e 2] e [SL04, seção 3] para as considerações práticas. Tais modificações não prejudicam a análise de convergência dual abaixo. Para realizar essa análise, consideramos a seguinte hipótese:

Hipótese 6.21. *Além de $k_{\max} = +\infty$, o algoritmo 6.20 gera infinitos passos $\{\pi^k\}$ com $\text{tol} = 0$, ou seja, $\|v^k\| = 0$ não é válido para todo k .*

Lema 6.22. *Considere o algoritmo 6.20 com a hipótese 6.21. Defina $\hat{\xi}_k$ como o valor do alvo na iteração k e suponha que existem $\bar{\xi}$ e $\underline{\xi}$ tais que*

$$f(\pi^k) \leq \underline{\xi} < \hat{\xi}_k < \bar{\xi} \leq f_*, \quad \text{para todo } k. \quad (6.46)$$

Então, $\langle w^{k-1}, \pi^k - \pi^* \rangle \leq 0$ para todo k , onde π^* é uma solução ótima dual.

Demonstração. Observe inicialmente que pela definição de $v^k \in \partial f(\pi^k)$, temos, em particular,

$$f_* = f(\pi^*) \leq f(\pi^k) + \langle v^k, \pi^* - \pi^k \rangle \Rightarrow \langle v^k, \pi^k - \pi^* \rangle \leq f(\pi^k) - f_*, \quad (6.47)$$

e provemos o lema por indução em k . Por (6.47), temos que $\langle v^1, \pi^1 - \pi^* \rangle \leq f(\pi^1) - f_*$. Como $w^0 = v^1$, o resultado segue para $k = 1$. Suponha, então, que a desigualdade é válida para $k \geq 1$ e provemos que $\langle w^k, \pi^{k+1} - \pi^* \rangle \leq 0$. Pelo passo 2(d) do algoritmo e por (6.45), temos:

$$\begin{aligned} \langle w^k, \pi^{k+1} - \pi^* \rangle &= \langle w^k, \pi^k + t_k w^k - \pi^* \rangle \\ &= \langle w^k, \pi^k - \pi^* \rangle + t_k \|w^k\|^2 \\ &= \alpha \langle v^k, \pi^k - \pi^* \rangle + (1 - \alpha) \langle w^{k-1}, \pi^k - \pi^* \rangle + t_k \|w^k\|^2. \end{aligned}$$

Por (6.47) e utilizando a hipótese de indução, temos:

$$\begin{aligned}
\langle w^k, \pi^{k+1} - \pi^* \rangle &\leq \alpha(f(\pi^k) - f_*) + t_k \|w^k\|^2 \\
&= \alpha(f(\pi^k) - f_*) + \beta(\hat{\xi}_k - f(\pi^k)) \\
&\leq \alpha(f(\pi^k) - \bar{\xi}) + \beta(\bar{\xi} - f(\pi^k)) \\
&= (\beta - \alpha)(\bar{\xi} - f(\pi^k)), \tag{6.48}
\end{aligned}$$

onde a primeira igualdade é dada por (6.44) e a segunda desigualdade pelas condições (6.46). Além disso, pelo passo 2(a) do algoritmo, $\beta - \alpha \leq 0$ e novamente por (6.46), temos $\bar{\xi} - f(\pi^k) > 0$. Assim, concluímos que o termo da direita de (6.48) é menor ou igual a zero, ou seja, $\langle w^k, \pi^{k+1} - \pi^* \rangle \leq 0$. \square

Lema 6.23. *Considere o algoritmo 6.20 com a hipótese 6.21. Defina $\hat{\xi}_k$ como o valor do alvo na iteração k e suponha que existem $\bar{\xi}$ e $\underline{\xi}$ tais que as desigualdades de (6.46) são válidas. Então, $\hat{\xi}_k \rightarrow \underline{\xi}$ e $f(\pi^k) \rightarrow \underline{\xi}$.*

Demonstração. Provemos inicialmente que $\{\|\pi^k - \pi^*\|^2\}$ é monótona decrescente, onde π^* é uma solução ótima dual qualquer. Pelo passo 2(d) do algoritmo e por (6.45), temos:

$$\begin{aligned}
\|\pi^{k+1} - \pi^*\|^2 &= \|\pi^k + t_k w^k - \pi^*\|^2 \\
&= \|\pi^k - \pi^*\|^2 + t_k^2 \|w^k\|^2 + 2t_k \langle w^k, \pi^k - \pi^* \rangle \\
&= \|\pi^k - \pi^*\|^2 + t_k^2 \|w^k\|^2 + 2t_k \alpha \langle w^k, \pi^k - \pi^* \rangle \\
&\quad + 2t_k (1 - \alpha) \langle w^{k-1}, \pi^k - \pi^* \rangle.
\end{aligned}$$

Pelo lema 6.22 e por (6.47), e em seguida pela definição de tamanho de passo dado em (6.44), temos:

$$\begin{aligned}
\|\pi^{k+1} - \pi^*\|^2 &\leq \|\pi^k - \pi^*\|^2 + t_k^2 \|w^k\|^2 + 2t_k \alpha (f(\pi^k) - f_*) \\
&= \|\pi^k - \pi^*\|^2 \\
&\quad + \frac{\beta^2 (\hat{\xi}_k - f(\pi^k))^2}{\|w^k\|^2} + \frac{2\alpha \beta (\hat{\xi}_k - f(\pi^k)) (f(\pi^k) - f_*)}{\|w^k\|^2} \\
&= \|\pi^k - \pi^*\|^2 \\
&\quad + \frac{\beta (\hat{\xi}_k - f(\pi^k))}{\|w^k\|^2} \left(\beta (\hat{\xi}_k - f(\pi^k)) + 2\alpha (f(\pi^k) - f_*) \right) \\
&< \|\pi^k - \pi^*\|^2 \\
&\quad + \frac{\beta (\hat{\xi}_k - f(\pi^k))}{\|w^k\|^2} \left(2\alpha (\hat{\xi}_k - f(\pi^k)) + 2\alpha (f(\pi^k) - f_*) \right) \\
&= \|\pi^k - \pi^*\|^2 + \frac{\beta (\hat{\xi}_k - f(\pi^k))}{\|w^k\|^2} \left(2\alpha (\hat{\xi}_k - f_*) \right),
\end{aligned}$$

sendo que a última desigualdade estrita é válida pois $0 < \beta \leq \alpha < 2\alpha$ e $\hat{\xi}_k - f(\pi^k) > 0$ por (6.46). Assim, temos $\hat{\xi}_k - f(\pi^k) > 0$, $\hat{\xi}_k - f_* < 0$ e $\alpha, \beta > 0$. Logo, para todo k ,

$$0 \leq \|\pi^{k+1} - \pi^*\|^2 < \|\pi^k - \pi^*\|^2,$$

ou seja, $\{\|\pi^k - \pi^*\|^2\}$ é monótona decrescente. Isso implica que

$$\lim_{k \rightarrow \infty} \frac{\beta^2 (\hat{\xi}_k - f(\pi^k))^2}{\|w^k\|^2} = \lim_{k \rightarrow \infty} t_k^2 \|w^k\|^2 = \lim_{k \rightarrow \infty} \|\pi^{k+1} - \pi^k\|^2 = 0$$

Além disso, pela hipótese 3.8, $\|w^k\|$ é limitada. Assim, por (6.46), o limite acima só poderá ser zero quando $\hat{\xi}_k \rightarrow \underline{\xi}$ e $f(\pi^k) \rightarrow \underline{\xi}$. \square

Antes de prosseguir com a análise de convergência, observe que a sequência $\{\bar{f}_k\}$ é monótona não decrescente e limitada superiormente por f_* , e portanto, convergente. Definimos \bar{f} como sendo seu limite.

Lema 6.24. *Considere o algoritmo 6.20 com a hipótese 6.21. Então, existe uma iteração externa L tal que $\varepsilon_\ell = \varepsilon$ para todo $\ell \geq L$.*

Demonstração. Considere o algoritmo 6.20 para iterações $k \geq K$ tal que $\bar{f}_k > \bar{f} - \nu$, com $0 < \nu < \varepsilon$. Veremos o que acontece numa iteração externa ℓ para o qual $k \geq K$. Observe que dados ξ_ℓ e ε_ℓ , os próximos iterados $\xi_{\ell+1}$ e $\varepsilon_{\ell+1}$ são dados pelos passos 6(a) ou 7(a). Note também que $\eta\chi_k \leq \chi_k < \nu$, já que $\bar{f}_k \in]\bar{f} - \nu, \bar{f}]$. Assim, no caso do passo 6(a), temos:

$$\sigma(\xi_{\ell+1} - \bar{f}_k) = \sigma(\varepsilon_\ell + \eta\chi_k) < \sigma(\varepsilon_\ell + \nu) < \sigma(\varepsilon_\ell + \varepsilon) \leq (2\varepsilon_\ell)/3. \quad (6.49)$$

Já no caso do passo 7(a), temos:

$$\sigma(\xi_{\ell+1} - \bar{f}_k) = \sigma\left(\frac{(\bar{f}_k + \varepsilon_\ell + \xi_\ell)}{2} - \bar{f}_k\right) = \frac{\sigma\varepsilon_\ell}{2} + \frac{\sigma(\xi_\ell - \bar{f}_k)}{2}. \quad (6.50)$$

Temos agora dois casos a considerar. Primeiramente, vamos supor que $\varepsilon_\ell = \varepsilon$. Se $\varepsilon_{\ell+1}$ é atualizado em 6(a), temos por (6.49) que $\sigma(\xi_{\ell+1} - \bar{f}_k) \leq (2\varepsilon)/3 < \varepsilon$ e assim $\varepsilon_{\ell+1} = \varepsilon$. Analisemos o caso em que $\varepsilon_{\ell+1}$ é atualizado em 7(a). Para alguma iteração anterior $k' < k$, onde obtemos ε_ℓ por 6(a) ou por 7(a), temos $\sigma(\xi_\ell - \bar{f}_{k'}) \leq \varepsilon$. Então, como $\{\bar{f}_k\}$ é não decrescente,

$$\sigma(\xi_\ell - \bar{f}_k) \leq \sigma(\xi_\ell - \bar{f}_{k'}) \leq \varepsilon. \quad (6.51)$$

Dessa forma, se $\varepsilon_{\ell+1}$ é atualizado em 7(a), então, por (6.50) e por (6.51), $\sigma(\xi_{\ell+1} - \bar{f}_k) \leq \varepsilon/6 + \varepsilon/2 = (2\varepsilon)/3 < \varepsilon$ e assim $\varepsilon_{\ell+1} = \varepsilon$. Assim, se $\varepsilon_\ell = \varepsilon$, então $\varepsilon_{\ell+1} = \varepsilon$ em qualquer um dos passos (6(a) ou 7(a)).

Considere agora o segundo caso: $\varepsilon_\ell > \varepsilon$. Então, para alguma iteração anterior $k' < k$, na qual obtemos ε_ℓ por 6(a) ou 7(a), temos, considerando que $\{\bar{f}_k\}$ é não decrescente, que

$$\varepsilon_\ell = \sigma(\xi_\ell - \bar{f}_{k'}) \geq \sigma(\xi_\ell - \bar{f}_k). \quad (6.52)$$

Assim, se $\varepsilon_{\ell+1}$ é dado em 6(a), temos $\varepsilon_{\ell+1} = \varepsilon$ ou, por (6.49), temos $\varepsilon_{\ell+1} < (2\varepsilon_\ell)/3$. Da mesma forma, se $\varepsilon_{\ell+1}$ é dado por 7(a), então temos $\varepsilon_{\ell+1} = \varepsilon$ ou, por (6.50) e por (6.52), temos $\varepsilon_{\ell+1} = \sigma(\xi_{\ell+1} - \bar{f}_k) \leq \sigma\varepsilon_\ell/2 + \varepsilon_\ell/2 \leq (2\varepsilon_\ell)/3$. Note que em ambos os casos, $\varepsilon_{\ell+1} \leq \varepsilon_\ell$. Portanto, $\{\varepsilon_\ell\}$ decresce com fator $2/3$, ou seja, atingimos ε em alguma iteração externa L . Com isso, recaímos no caso analisado anteriormente, ou seja $\varepsilon_\ell = \varepsilon$ para todo $\ell \geq L$. \square

Lema 6.25. *Considere o algoritmo 6.20 com a hipótese 6.21. Suponha que existe K suficientemente grande tal que para todo $k \geq K$, existem apenas passos nulos. Então, ξ_ℓ decresce para $\bar{f}_k + \varepsilon = \bar{f} + \varepsilon$ para todo $k \geq K$.*

Demonstração. Considere L suficientemente grande tal que $\varepsilon_\ell = \varepsilon$ para todo $\ell \geq L$ e tal que as iterações gerais k correspondentes sejam maiores ou iguais que K . Observe que a existência de tal L é garantida pelo lema 6.24. Note que $\bar{f}_k = \bar{f}_{k+1} = \bar{f}$ para todo $k \geq K$ pois ocorrem apenas passos nulos. Então, dado um $\tilde{\ell} \geq L$ e $k \geq K$ temos o cômputo de $\xi_{\tilde{\ell}+1}$ no passo 7(a) do algoritmo e

$$\xi_{\tilde{\ell}+1} \doteq \frac{\bar{f}_k + \varepsilon_{\tilde{\ell}} + \xi_{\tilde{\ell}}}{2} = \frac{\bar{f} + \varepsilon + \xi_{\tilde{\ell}}}{2}. \quad (6.53)$$

Como ocorrem apenas passos nulos, por 5(c), observamos que ocorrem apenas diminuições do alvo no passo 7(a), pois durante o processo γ ou τ crescem, até serem novamente zerados no passo 7(b). Logo, na iteração externa $\tilde{\ell} + 1$, computaremos:

$$\xi_{\tilde{\ell}+2} = \frac{\bar{f} + \varepsilon + \xi_{\tilde{\ell}+1}}{2} = \frac{\bar{f} + \varepsilon}{2} + \frac{\bar{f} + \varepsilon + \xi_{\tilde{\ell}}}{4},$$

onde a última igualdade é válida por (6.53). Assim, para qualquer $\ell \geq 1$ podemos escrever:

$$\xi_{\tilde{\ell}+\ell} \doteq \sum_{j=1}^{\ell} \frac{\bar{f} + \varepsilon}{2^j} + \frac{\xi_{\tilde{\ell}}}{2^\ell}.$$

Tomando-se o limite para $\ell \rightarrow \infty$, essa última igualdade resulta $\xi_\ell \rightarrow \bar{f} + \varepsilon$, conforme queríamos. \square

Lema 6.26. *Considere o algoritmo 6.20 com a hipótese 6.21. Suponha que temos $f_* - \bar{f} - \varepsilon \geq \nu > 0$ para algum ν tal que $0 < \nu < \varepsilon$. Então, existe uma iteração $\tilde{\ell}$ tal que*

$$\bar{f} + \varepsilon - \nu < \xi_\ell < \bar{f} + \varepsilon + \nu \text{ para todo } \ell \geq \tilde{\ell}. \quad (6.54)$$

Demonstração. Considere o algoritmo nas iterações $k \geq K$ e $\ell \geq L$ suficientemente grandes, ou seja, tais que $\bar{f}_k > \bar{f} - \nu$ para todo $k \geq K$ e $\varepsilon_\ell = \varepsilon$ para todo $\ell \geq L$. Observe que a existência de tal L foi garantida pelo lema 6.24.

Suponha que $\xi_\ell \geq \bar{f} + \varepsilon + \nu$ para alguma iteração $\ell \geq L$. Então,

$$\bar{f}_k < \bar{f}_k + \nu < \bar{f} + \nu \leq \xi_\ell - \varepsilon$$

para cada iteração k correspondente. Sabemos que o aumento do alvo ocorre apenas quando a condição do passo 4(c) é satisfeita, o que não ocorre pela desigualdade acima. Dessa forma, o processo do algoritmo permanece entre os passos 2 e 5, com τ crescente no passo 2(e) até que τ atinja o valor $\bar{\tau}$ no passo 4(c) ou 5(c), o que nos leva ao passo 7. Assim, com sucessivas diminuições do alvo, chegamos numa iteração externa $\tilde{\ell}$ tal que $\xi_{\tilde{\ell}} < \bar{f} + \varepsilon + \nu$. Até obter tal iteração, temos $\xi_\ell \geq \bar{f} + \varepsilon + \nu$ e $\bar{f}_k > \bar{f} - \nu$ (pois $k \in K$). Logo,

$$\xi_{\ell+1} = \frac{\bar{f}_k + \varepsilon_\ell + \xi_\ell}{2} > \frac{\bar{f} - \nu + \varepsilon + \bar{f} + \varepsilon + \nu}{2} = \bar{f} + \varepsilon > \bar{f} + \varepsilon - \nu.$$

Em particular, temos também $\xi_{\tilde{\ell}} > \bar{f} + \varepsilon - \nu$. Logo, $\tilde{\ell}$ satisfaz as desigualdades de (6.54).

Suponha agora, por outro lado, que $\xi_\ell \leq \bar{f} + \varepsilon - \nu$ para algum $\ell \geq L$. Note que quando $k \geq K$, temos:

$$\bar{f}_k > \bar{f} - \nu \geq \xi_\ell - \varepsilon = \xi_\ell - \varepsilon_\ell.$$

Primeiramente, observe que a condição $\bar{f}_k > \xi_\ell - \varepsilon_\ell$ ocorre apenas no passo 4(c) do algoritmo. Supondo que ocorrem apenas passos nulos, essa condição nunca é atingida, porque pelo lema 6.25, ξ_ℓ decresce para $\bar{f}_k + \varepsilon_\ell$. Logo, podemos supor que ocorrem passos sérios, e conseqüentemente \bar{f}_k é atualizado e o passo 4(c) do algoritmo é atingido. Assim, quando \bar{f}_k cresce suficientemente de modo que $\bar{f}_k > \bar{f} - \nu$, temos que $\bar{f}_k > \xi_\ell - \varepsilon_\ell$, o que nos leva ao passo 6 do algoritmo. Logo,

$$\xi_{\ell+1} = \bar{f}_k + \varepsilon_\ell + \eta\chi_k > \bar{f} - \nu + \varepsilon + \eta\chi_k > \bar{f} + \varepsilon - \nu.$$

Esse $\ell + 1$ pode satisfazer as condições (6.54) ou resultar $\xi_{\ell+1} \geq \bar{f} + \varepsilon + \nu$. Este último recai no caso discutido no começo desta demonstração. Logo, concluímos que para algum $\tilde{k} \geq K$ que ocorre durante a iteração externa $\tilde{\ell}$,

$$\bar{f} + \varepsilon - \nu < \xi_{\tilde{\ell}} < \bar{f} + \varepsilon + \nu. \quad (6.55)$$

Falta-nos mostrar que as desigualdades de (6.54) são válidas para todo $\ell > \tilde{\ell}$. Seja j a primeira iteração depois de \tilde{k} para o qual o alvo muda na iteração externa $\tilde{\ell} + 1$. Mais especificamente, se $\xi_{\tilde{\ell}+1}$ é dado nos passos 6(a) ou 7(a), temos, respectivamente,

$$\xi_{\tilde{\ell}+1} = (\bar{f}_j + \varepsilon) + \eta\chi_j \quad (6.56)$$

ou

$$\xi_{\tilde{\ell}+1} = \frac{(\bar{f}_j + \varepsilon) + \xi_{\tilde{\ell}}}{2}. \quad (6.57)$$

Sabemos que $\eta\chi_j \leq \chi_j < \nu$, pois $\bar{f}_k > \bar{f} - \nu$. Logo, se $\xi_{\tilde{\ell}+1}$ é dado no passo 6(a), temos, por (6.56), que

$$\bar{f} + \varepsilon - \nu < \xi_{\tilde{\ell}} < \xi_{\tilde{\ell}+1} = \bar{f}_j + \varepsilon + \eta\chi_j < \bar{f} + \varepsilon + \nu,$$

onde a primeira desigualdade é válida por (6.55). Por outro lado, no caso em que $\xi_{\tilde{\ell}+1}$ é dado no passo 7(a), temos, por (6.57), que

$$\bar{f} + \varepsilon - \nu < \frac{\bar{f}_j + \varepsilon + \bar{f} + \varepsilon - \nu}{2} < \frac{\bar{f}_j + \varepsilon + \xi_{\tilde{\ell}}}{2} = \xi_{\tilde{\ell}+1} < \xi_{\tilde{\ell}} < \bar{f} + \varepsilon + \nu,$$

sendo que a primeira desigualdade é válida pois $\bar{f}_j > \bar{f} - \nu$, enquanto que a segunda e a última são verdadeiras por (6.55). Logo, em ambos os casos, temos $\xi_{\tilde{\ell}+1}$ satisfazendo as condições (6.54). O resultado segue por um processo simples de indução. \square

Teorema 6.27. *Considerando o algoritmo 6.20 sob a hipótese 6.21, conclui-se que $\bar{f} + \varepsilon \geq f_*$.*

Demonstração. Suponha, por contradição, que $\bar{f} + \varepsilon < f_*$. Podemos, então escolher um ν tal que $0 < \nu < \varepsilon$ tal que $f_* - \bar{f} - \varepsilon \geq \nu > 0$. Assim, pelo lema 6.26, existe uma iteração $\tilde{\ell}$ tal que

$$\bar{f} + \varepsilon - \nu < \xi_\ell < \bar{f} + \varepsilon + \nu \text{ para todo } \ell > \tilde{\ell}.$$

Defina $\underline{\xi} \doteq \bar{f} + \varepsilon - \nu$ e $\bar{\xi} \doteq \bar{f} + \varepsilon + \nu$. Como $\varepsilon - \nu > 0$ e $f(\pi^k) \leq \bar{f}_k \leq \bar{f}$, temos, para todo k e ℓ suficientemente grande,

$$f(\pi^k) \leq \bar{f} < \underline{\xi} < \xi_\ell < \bar{\xi} \leq f_*.$$

Logo, pelo lema 6.23, concluímos que $f(\pi^k) \rightarrow \underline{\xi}$, o que é uma contradição, pois $\bar{f} < \underline{\xi}$. Logo, $\bar{f} + \varepsilon \geq f_*$, conforme queríamos. \square

Finalmente, cabe-nos mencionar que o resultado de convergência obtido dessa variação do algoritmo do volume mostra apenas ε -otimalidade em tempo infinito.

Capítulo 7

Experimentos Numéricos

Nos capítulos anteriores, abordamos diversas formas de resolver problemas lineares do tipo (1.1). Observa-se que esses métodos são particularmente interessantes na resolução de problemas lineares inteiros. A idéia é usar relaxações desses problemas e obter soluções viáveis inteiras através de heurísticas e métodos de divisão e conquista. Tradicionalmente, utiliza-se o método simplex dual para resolver as relaxações. Porém, desde que Barahona e Anbil propuseram o algoritmo do volume, observou-se que nesse caso seu desempenho é em geral superior, sobretudo porque seu custo computacional é bem inferior ao do simplex para problemas de grande porte.

Em tal contexto, algumas aplicações do algoritmo do volume foram abordados na literatura, entre eles, o problema da árvore de Steiner [BL01, BMS02, BBP03], da partição de conjuntos [BA00, BA02], da cobertura por conjuntos [BA00] e da localização de facilidades [BC05]. Com o intuito de observar resultados referentes a todos os métodos estudados, realizamos também experimentos numéricos, em particular, com o problema do *corte máximo em grafos* (também denominado *max-cut*) que veremos na seção 7.1.

Seguindo a abordagem dada em [BL01], utilizamos também o método *branch-and-cut-and-price* para a obtenção de soluções viáveis inteiras. Para realizar os experimentos, usamos bibliotecas disponíveis pelo projeto COIN [CO07], mais especificamente as rotinas em C++ do método simplex, do algoritmo do volume original e do método *branch-and-cut-and-price*, com aplicação no problema *max-cut*. Os demais — os algoritmos do volume revisado e com variação do alvo, e os métodos de subgradientes com convergência ergódica — foram implementados tomando-se o código do COIN como base. Os resultados de comparações numéricas serão observadas na seção 7.2.

7.1 Corte Máximo em Grafos

Seja $G \doteq (V, E)$ um grafo não dirigido, com V e $E \doteq \{e_1, e_2, \dots, e_n\}$, sendo os conjuntos dos vértices e das arestas, respectivamente. Denotamos também $c_i \in \mathbb{R}$ como o custo da aresta e_i . Além disso, a seguinte definição é necessária para compreendermos o problema que queremos abordar:

Definição 7.1. O corte associado a um conjunto $S \subseteq V$ de vértices é o conjunto de todas as arestas que têm uma ponta em S e outra em $V \setminus S$.

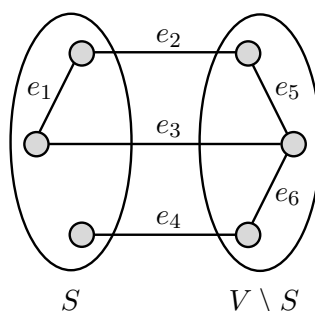


Figura 7.1.1: O conjunto $\{e_2, e_3, e_4\}$ é um corte associado a $S \subseteq V$.

O problema do *corte máximo em grafos* (também denominado *max-cut*) consiste então em encontrar um corte K de custo máximo, ou seja, tal que o valor de $\sum_{e \in K} c^e$ seja máximo. Considere agora em corte K fixado. Um vetor $x \doteq (x_1, \dots, x_n)^T$ é definido da seguinte forma: $x_e = 1$ se $e \in K$ e $x_e = 0$ caso contrário. Definindo também $c \doteq (c_1, \dots, c_n)^T$, podemos escrever o custo do corte K como sendo o produto interno $\langle c, x \rangle$. O problema do corte máximo pode ser, então, caracterizado da seguinte forma:

$$\text{maximizar } \langle c, x \rangle \quad \text{sujeito a} \quad \begin{cases} \sum_{e \in F} x_e - \sum_{e \in C \setminus F} x_e \leq |F| - 1 \\ x_e \in \{0, 1\}, \quad \forall e \in E \end{cases} \quad (7.1)$$

para todo ciclo C e todo subconjunto $F \subseteq C$ com cardinalidade ímpar. Conforme [BGJR88, seção 3], a primeira restrição acima descreve um corte.

O max-cut é um problema NP-completo [Kar72], o que faz com que tenhamos encontrar apenas uma solução aproximada de (7.1). Além disso, como o nosso intuito é utilizar os algoritmos estudados, resolveremos o problema a partir de sua relaxação, dada por:

$$\text{maximizar } \langle c, x \rangle \quad \text{sujeito a} \quad \begin{cases} \sum_{e \in F} x_e - \sum_{e \in C \setminus F} x_e \leq |F| - 1 \\ 0 \leq x_e \leq 1, \quad \forall e \in E \end{cases} \quad (7.2)$$

Com o método *branch-and-cut-and-price* resolve-se o problema linear inteiro (7.1) através da resolução de uma série de subproblemas lineares cujas (“poucas”) restrições pertencem ao conjunto de restrições de (7.1). Mais especificamente, a técnica é de divisão e conquista, com variáveis e cortes (isto é, restrições) gerados dinamicamente. Constrói-se, dessa forma, uma árvore onde cada nó representa um subproblema.

Mostraremos a seguir a estrutura do método *branch-and-cut* (onde apenas cortes são gerados) para problemas gerais de minimização. Por outro lado, o algoritmo 7.5, que trata a separação, é específica para o problema do corte máximo em grafos. Além disso, a geração de variáveis (conhecida como *price*) pode ser vista de maneira análoga ao algoritmo 7.2, pois corresponde à geração de cortes para o problema dual. Referimo-nos a [LRT01] para uma melhor compreensão e detalhes do método.

É importante salientar que esse método requer um algoritmo de resolução de problemas lineares que: (i) ofereça limitantes inferiores dos valores ótimos, para o *bounding*, (ii) permita reiniciar o algoritmo facilmente, visando eficiência e (iii) gere soluções primais (não necessariamente ótimas), para a geração de cortes e o *branching*. Note que as diversas variações do algoritmo do volume satisfazem as condições (i)–(iii), e portanto podem ser utilizadas para resolver aproximadamente os problemas relaxados.

Algoritmo 7.2. *Método Branch-and-Cut*

Entrada: Instância do problema original.

Saída: Uma solução ótima s^* aproximada.

1. Considere \hat{s} uma solução viável qualquer.
Denote α como limitante superior do valor ótimo e inicialize $\alpha \doteq \langle c, \hat{s} \rangle$.
2. Tome um subproblema P_0 com algumas “poucas” desigualdades.
Denote \mathcal{A} como o conjunto de subproblemas (i.e., a árvore).
Faça $\mathcal{A} \doteq \{P_0\}$.
3. Se $\mathcal{A} = \emptyset$, então faça $s^* \doteq \hat{s}$ e pare.
Senão, tome $P \in \mathcal{A}$, faça $\mathcal{A} \doteq \mathcal{A} \setminus \{P\}$ e vá para o algoritmo 7.3 com o subproblema P como entrada.
4. Se a rotina do algoritmo 7.3 devolve um \bar{s} viável, então $\langle c, \bar{s} \rangle < \langle c, \hat{s} \rangle$.
Assim, faça $\hat{s} \doteq \bar{s}$, $\alpha \doteq \langle c, \bar{s} \rangle$ e vá para o passo 3.
Senão, se tal rotina devolve *cortado*, então, vá para o passo 3.
Senão, continue.
5. Vá para o algoritmo 7.4 com P como entrada, adicione os subproblemas gerados em \mathcal{A} e vá para o passo 3.

Algoritmo 7.3. *Bounding*

Entrada: Um problema P descrito por um conjunto de desigualdades \mathcal{L}' e um limitante superior α da solução ótima.

Saída: Uma solução ótima s^* de P , ou um limitante inferior para a solução ótima, ou uma mensagem *cortado*, que indica que o problema pode ser descartado da árvore.

1. Seja $\mathcal{C} \doteq \mathcal{L}'$.
2. Resolva o programa linear com o conjunto de restrições \mathcal{C} usando, por exemplo, uma variação do algoritmo do volume.
3. Se devolver uma solução viável \hat{x} , então vá para o passo 4.
Senão, pare e devolva a mensagem *cortado*.
4. Se $\langle c, \hat{x} \rangle < \alpha$, então vá para o passo 5.
Senão, pare e devolva a mensagem *cortado*.
5. Se \hat{x} é o vetor de incidências de algum $\hat{s} \in P$, então \hat{s} é uma solução ótima desse subproblema. Assim, pare e faça $x^* \doteq \hat{x}$.
Senão, chame a rotina de separação (algoritmo 7.5) em \hat{x} e obtenha as desigualdades violadas \mathcal{C}' . Se $\mathcal{C}' = \emptyset$, então $\langle c, \hat{x} \rangle$ é limitante inferior do valor ótimo de P . Pare e devolva \hat{x} e o limitante inferior $\langle c, \hat{x} \rangle$.
Senão, faça $\mathcal{C} \doteq \mathcal{C} \cup \mathcal{C}'$ e vá para o passo 2.

Algoritmo 7.4. *Branching*

Entrada: Um problema P descrito por um conjunto de desigualdades \mathcal{L}' e um limitante inferior \hat{x} .

Saída: Problemas P_1, \dots, P_k gerados através de P .

1. Determine conjuntos de restrições $\mathcal{L}_1, \dots, \mathcal{L}_k$ tais que $\cup_{i=1}^k \mathcal{L}_i = \mathcal{L}'$ e \hat{x} não satisfaz $\cup_{i=1}^k \mathcal{L}_i$.
2. Defina os problemas P_1, \dots, P_k através de $\mathcal{L}_1, \dots, \mathcal{L}_k$ respectivamente.

Algoritmo 7.5. *Algoritmo de Separação Para o Problema do Corte Máximo*

Entrada: Um vetor \hat{x} .

Saída: Uma desigualdade que separa \hat{x} do conjunto viável do subproblema.

1. Defina $w_e \doteq c_e \max\{\hat{x}_e, 1 - \hat{x}_e\}$ para todo $e = 1, \dots, n$.
2. Encontre uma árvore geradora máxima T do grafo com custos definidos por w_e para todo $e = 1, \dots, n$.
3. Seja $S \subseteq V$. Considere A e B conjuntos de arestas de T , onde as arestas de A têm um extremo em S e outro em $V \setminus S$ e de B têm os dois extremos em S ou os dois em $V \setminus S$. Inicialize $A = B = \emptyset$.

4. Para cada $e \in T$, se $\hat{x}_e \geq 1 - \hat{x}_e$, faça $A \doteq A \cup \{e\}$.
Senão, faça $B \doteq B \cup \{e\}$.
5. As arestas de A definem o corte K .
6. Para cada aresta $e \notin T$, coloque-a em T e considere C como sendo o ciclo formado.
7. Se $e \in K$, teste a violação da primeira restrição de (7.2), com $F \doteq A$.
Senão, teste com $F \doteq A \cup \{e\}$.

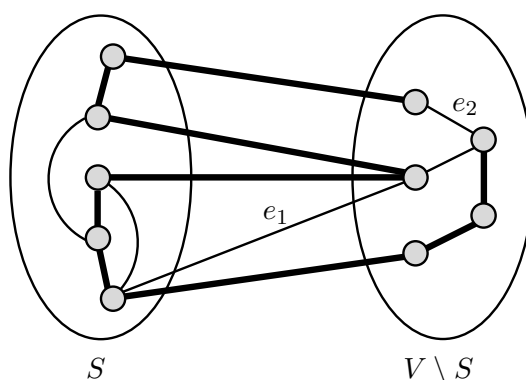


Figura 7.1.2: Processo do algoritmo de separação.

O algoritmo 7.5 de separação é uma heurística [BL01]. Além disso, para compreendermos melhor sua idéia, observemos a figura 7.1.2. Durante o processo, devemos encontrar uma árvore geradora de custo máximo T , representada pela linha em negrito. As arestas que têm uma ponta em S e outra em $V \setminus S$ pertencem a K . Note que podem existir arestas em K que não pertencem a A , mais especificamente, a T . Na figura, é o caso da aresta e_1 . Ademais, podem existir arestas como e_2 , que não pertencem a T , nem a K .

7.2 Comparações Numéricas

Nesta seção, mostraremos alguns resultados obtidos com o problema do corte máximo de grafos. A primeira comparação se refere aos métodos de sub-gradientes com convergência ergódica, vistos no capítulo 5. Em particular, vimos duas regras específicas para a escolha do tamanho de passo e dos coeficientes de combinação convexa: a regra de Shor (corolário 5.4), que realiza uma média ponderada, e a regra de Larsson e Liu (corolário 5.5), que faz uma média aritmética e se utiliza de salva-guardas.

Sabendo que esses métodos possuem convergência primal lenta, realizamos os testes com grafos gerados aleatoriamente, com dimensões pequenas,

ou seja, com no máximo 30 vértices e 400 arestas e densidades variadas. Além disso, como critério de parada, utilizamos um número máximo de iterações, como sendo de 5000. Em todos os casos, a regra de Larsson e Liu foi considerada superior ao método de Shor, no sentido de retornar soluções com menor folga de dualidade, ou seja, uma melhor solução primal. O resultado é condizente, pois, como vimos, a regra de Larsson e Liu é um caso particular da regra de Shor, que evita pesos grandes nas primeiras iterações do algoritmo. Apesar de pouco usual, fizemos um gráfico de perfil desempenho [DM02] com as folgas de dualidade e o resultado obtido pode ser visto na figura 7.2.1. Como a folga relativa é grande, os valores maiores ou iguais a 1024 foram colocados juntos nesse gráfico.

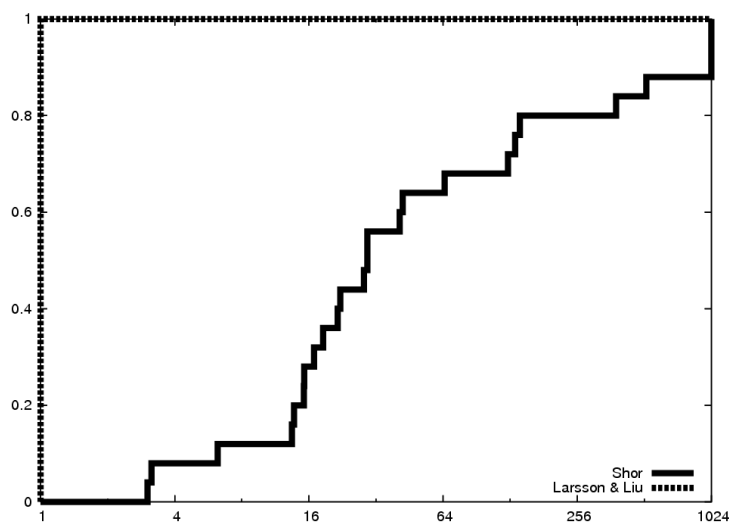


Figura 7.2.1: Perfil desempenho para as regras de Shor e Larsson e Liu.

O outro experimento que realizamos se refere ao algoritmo do volume, abordado no capítulo 6. Estudamos três algoritmos: o original (seção 6.1), o revisado (seção 6.2) e com variação do alvo (seção 6.3). Primeiramente consideramos grafos em formato de grade, com custos gerados de maneira aleatória. Note que ao contrário do método de subgradientes com convergência ergódica, esses métodos são mais eficientes e dessa forma, é interessante utilizarmos grafos com dimensões grandes nos experimentos.

No método *branch-and-cut-and-price* implementado, a rotina *branching* foi chamada sempre que a diferença entre os limitantes inferiores e superiores não decrescem de pelo menos 0.1% nas últimas três iterações. Além disso, os algoritmos do volume são usados para resolver os problemas lineares, enquanto a diferença entre esses limitantes inferiores e superiores diminuem

suficientemente. Do contrário, o método utiliza a rotina do simplex. Esse procedimento é necessário devido aos limitantes inferiores dos algoritmos do volume serem aproximados, e não exatos. Os resultados podem ser vistos na tabela 7.2.1, onde cada coluna referente a um método representa o tempo gasto (em segundos) por ele.

Vértices	Arestas	Simplex	VA	VA+VTVM	RVA
10×10	180	0.020	0.020	0.016	0.028
20×20	760	0.208	0.112	0.104	0.156
30×30	1740	1.344	0.360	0.296	0.396
40×40	3120	3.564	0.400	0.580	0.456
50×50	4900	13.929	1.016	0.612	1.052
60×60	7080	37.934	1.636	1.576	2.948
70×70	9660	89.746	2.488	4.324	3.944
80×80	12640	609.102	5.532	4.472	7.464
90×90	16020	3341.229	13.353	8.653	13.957

Tabela 7.2.1: Desempenho dos algoritmos do volume e do simplex.

Observa-se que quanto maior o problema, mais notória é a diferença entre o método simplex e as variações dos algoritmos do volume. Realizamos ainda medições com várias outras instâncias geradas aleatoriamente, e o resultado pode ser observado na figura 7.2.2, onde mostramos o perfil desempenho dos algoritmos do volume e do simplex.

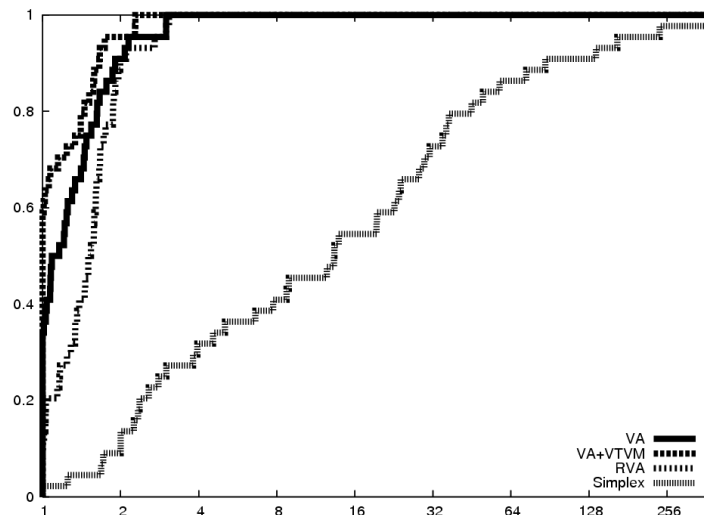


Figura 7.2.2: Perfil desempenho dos algoritmos do volume e do simplex.

Para uma melhor comparação dos três tipos de algoritmos do volume, referimo-nos à figura 7.2.3, na qual eliminamos o gráfico correspondente ao método simplex.

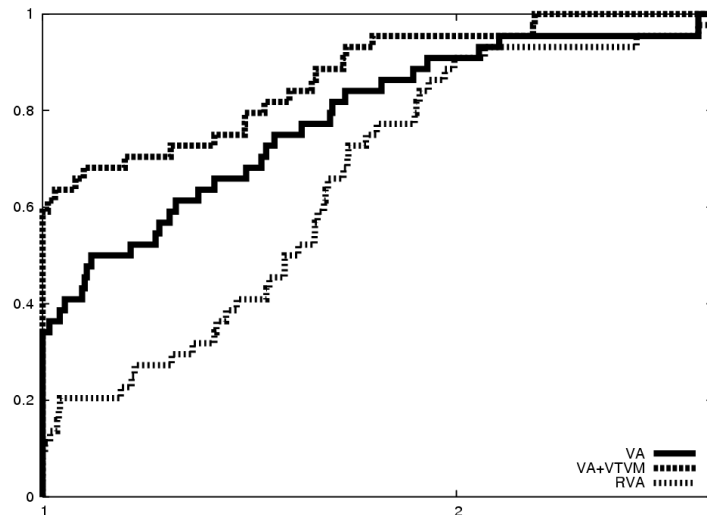


Figura 7.2.3: Perfil desempenho para VA, VA+VTVM e RVA.

Podemos observar no gráfico acima que o VA+VTVM é mais rápido que os demais algoritmos do volume e que o RVA é o mais lento. De fato, em nossos experimentos, o VA+VTVM foi, em média, 7% mais rápido que o VA, enquanto que o RVA foi aproximadamente 13% mais lento que o VA. Esse último resultado está de acordo com os resultados obtidos em [BMS02], onde se realizou experimentos com o problema da árvore de Steiner.

Em [SL04], utilizou-se o VTVM também nos algoritmos RVA e o método de feixes específico denominado BVA. Seus experimentos mostraram que a incorporação do VTVM no RVA e no BVA não fornece melhores resultados em problemas gerais, em particular o problema do corte máximo em grafos. Juntando-se seus resultados com a comparação que realizamos, podemos concluir que o algoritmo mais promissor na prática é o VA+VTVM, ao menos na classe de problemas que abordamos.

Capítulo 8

Conclusão

Procuramos apresentar aqui um texto auto-contido, tratando, nos primeiros capítulos, assuntos essenciais para o nosso estudo: análise convexa (no capítulo 2) e métodos de otimização não diferenciável (nos capítulos 3 e 4), como os métodos de descida, de subgradientes, de planos de corte e de feixe. Procuramos selecionar os tópicos e as demonstrações existentes de forma a facilitar a compreensão dos assuntos abordados.

Apresentamos depois algumas maneiras de resolver problemas de programação linear baseando-se na relaxação das restrições “difíceis”, com cômputo de solução dual através de métodos de otimização não diferenciável como base, e com a geração de possíveis soluções primais. Nos capítulos 5 e 6, vimos duas dessas técnicas: o método de subgradientes com convergência ergódica e o algoritmo do volume. No primeiro método, foi possível analisarmos tanto a convergência primal quanto a dual. No segundo, somente a convergência dual dos métodos modificados (isto é, RVA e VA+VTVM) foi estabelecida, sendo que no VA+VTVM garantiu-se ε -otimalidade apenas em tempo infinito.

Finalmente, realizamos experimentos numéricos com o problema do corte máximo em grafos. Conforme indicava nossas referências, constatamos que a regra de Larsson e Liu possui, de fato, desempenho superior à regra de Shor. Utilizando a técnica *branch-and-cut-and-price*, também vimos, conforme esperávamos, o bom desempenho do algoritmo do volume quando comparado ao simplex. Um resultado interessante que constatamos é a melhora do algoritmo do volume quando incorporamos nele o método de variação do alvo.

Referências Bibliográficas

- [BA00] F. Barahona e R. Anbil. The volume algorithm: producing primal solutions with a subgradient method. *Mathematical Programming*, 87(3):385–399, 2000.
- [BA02] F. Barahona e R. Anbil. On some difficult linear programs coming from set partitioning. *Discrete Applied Mathematics*, 118(1–2):3–11, 2002.
- [BBP03] L. Bahiense, F. Barahona e O. Porto. Solving Steiner tree problems in graphs with Lagrangian relaxation. *Journal of Combinatorial Optimization*, 7(3):259–282, 2003.
- [BC05] F. Barahona e F.A. Chudak. Near-optimal solutions to large-scale facility location problems. *Discrete Optimization*, 2(1):35–50, 2005.
- [Ber03] D. Bertsekas. *Convex Analysis And Optimization*. Athena Scientific, Belmont, MA, 2003.
- [BGJR88] F. Barahona, M. Grötschel, M. Jünger e G. Reinelt. An application of combinatorial optimization to statistical physics and circuit layout design. *Operations Research*, 36(3):493–513, 1988.
- [BGLS03] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal e C.A. Sagastizábal. *Numerical Optimization: Theoretical and Practical Aspects*. Springer Verlag, 2003.
- [BL01] F. Barahona e L. Ladányi. Branch and cut based on the volume algorithm: Steiner trees in graphs and max-cut. Technical report, IBM Thomas J. Watson Research Center, Yorktown Heights, 2001.

- [BMS02] L. Bahiense, N. Maculan e C.A. Sagastizábal. The volume algorithm revisited: relation with bundle methods. *Mathematical Programming*, 94(1):41–69, 2002.
- [CFM75] P. Camerin, L. Fratta e F. Maffioli. On improving relaxation methods by modified gradient techniques. 3:26–34, 1975.
- [CO07] COIN-OR. The computational infrastructure for operations research, 2007.
- [DM02] E.D. Dolan e J.J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.
- [DV85] V.F. Dem'yanov e L.V. Vasil'ev. *Nondifferentiable Optimization*. Optimization Software, New York, 1985.
- [HU80] J.B. Hiriart-Urruty. Lipschitz r -continuity of the approximate subdifferential of a convex function. *Math. Scand.*, 47:123–134, 1980.
- [HUL93a] J.B. Hiriart-Urruty e C. Lemaréchal. *Convex Analysis and Minimization Algorithms I: Fundamentals*. Springer Verlag, 1993.
- [HUL93b] J.B. Hiriart-Urruty e C. Lemaréchal. *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*. Springer Verlag, 1993.
- [HUL01] J.B. Hiriart-Urruty e C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer Verlag, Heidelberg, 2001.
- [Kar72] R.M. Karp. Reducibility among combinatorial problems. In J.W. Thatcher e R.E. Miller, editores, *Complexity of Computer Computations*, p. 85–103. Plenum Press, New York, 1972.
- [LL89] T. Larsson e Z. Liu. A 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.
- [LPS96] T. Larsson, M. Patriksson e A. Strömberg. Conditional subgradient optimization - theory and applications. *European Journal of Operational Research*, 88:382–403, 1996.

- [LPS99] T. Larsson, M. Patriksson e A. Strömberg. Ergodic, primal convergence in dual subgradient schemes for convex programming. *Mathematical Programming*, 86:283–312, 1999.
- [LRT01] L. Ladányi, T.K. Ralphs e L.E. Trotter Jr. Branch, cut, and price: Sequential and parallel. In D. Naddef e M. Juenger, editores, *Computational Combinatorial Optimization*, p. 223. Springer, Berlin, 2001.
- [LS97] C. Lemaréchal e C.A. Sagastizábal. Variable metric bundle methods: from conceptual to implementable forms. *Mathematical Programming*, 76(3):393–410, 1997.
- [NW99] J. Nocedal e S. Wright. *Numerical Optimization*. Springer Verlag, New York, 1a. edição, 1999.
- [Pol69] B.T. Polyak. Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*, 9:14–29, 1969.
- [Pol87] B.T. Polyak. *Introduction to Optimization*. Optimization Software, 1987.
- [Roc70] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.
- [SC96] H. D. Sherali e G. Choi. Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs. *Operations Research Letters*, 19(3):105–113, 1996.
- [SCT00] H. D. Sherali, G. Choi e C.H. Tuncbilek. A variable target value method for nondifferentiable optimization. *Operations Research Letters*, 26(1):1–8, 2000.
- [Sho85] N.Z. Shor. *Minimization Methods for Non-Differentiable Functions*. Springer, New York, NY, 1985.
- [SL04] H.D. Sherali e C. Lim. On embedding the volume algorithm in a variable target value method. *Operations Research Letters*, 32(5):455–462, 2004.
- [SS99] M.V. Solodov e B.F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of an operator. *Set Valued Analysis*, 24(7):323–345, 1999.

Índice Remissivo

- ε -subdiferencial, 15
- ε -superdiferencial, 16
- algoritmo de separação, 98
- algoritmo do volume, 66, 101
 - com variação do alvo, 86
 - original, 66
 - revisado, 72
- alvo, 86
- aresta, 96
- bounding, 98
- branching, 98
- BVA, 79, 102
- caixa preta, 28
- COIN, 95
- condição
 - de ε -otimalidade, 25
 - de Armijo, 74
 - de otimalidade, 24
- cone, 8
 - ε -normal, 10
 - gerado por um conjunto, 8
 - normal, 9, 13
 - polar, 8
 - tangente, 8, 13
- conjunto
 - afim, 7
 - bola aberta, 1
 - bola fechada, 1
 - de nível, 6
 - dos reais estendidos, 5
 - fecho de, 1
 - interior de, 1
 - simplex, 1
- corte em grafos, 96
- corte máximo, 96
- decrécimo nominal, 35, 38
- derivada ε -direcional, 16
- derivada direcional, 11
- direção
 - ε -normal, 10
 - de ε -descida, 25
 - de máxima descida, 26
 - normal, 9
 - tangente, 8
- direção de descida, 24
- distância de Hausdorff, 18
- distância euclidiana, 1
- domínio efetivo, 5
- dual, 2
- epígrafe, 5
- erro de linearização, 40, 77
- fecho afim, 7
- fecho convexo, 7
- feixe, 40
- folgas complementares, 62
- função
 - afim por partes, 7, 14
 - côncava, 5

- convexa, 5, 22
 - diferenciável, 12
 - estritamente côncava, 5
 - estritamente convexa, 5
 - fechada, 6
 - indicadora, 6, 23
 - lagrangeana, 2, 41, 56
 - Lipschitz contínua, 6, 8, 17, 19
 - multivalorada, 7, 16
 - própria, 5, 22
 - semi-contínua inferior, 6, 22, 31
 - semi-contínua superior, 6, 14
 - suporte, 6, 12, 16
- ganho esperado, 75, 77
- gráfico
- de função multivalorada, 7
 - fechado, 17
 - perfil desempenho, 100
- grafo, 96
- hipógrafo, 6, 75
- hiperplano separador, 10
- interior relativo, 7
- Larsson e Liu, *veja* regra
- linearização agregada, 45
- método
- branch-and-cut, 97
 - branch-and-cut-and-price, 97
 - de descida, 26
 - de feixe, 38
 - de planos de corte, 34
 - de região de confiança, 87
 - de subgradientes, 28
 - com desvios, 33
 - projetados, 33, 57
 - quase-Newton, 48
 - simplex, 101
- max-cut, *veja* corte máximo
- medida de Lebesgue, 11
- multifunção, *veja* função
- norma euclidiana, 1
- oráculo, 28
- ponto proximal, 47
- primal, 3
- produto interno, 1
- programação linear, 2
- projeção euclidiana, 1, 10, 65
- regra
- de Larsson e Liu, 63, 99
 - de Shor, 63, 99
- regularização de Moreau-Yosida, 47
- relaxação, 96
- relaxação lagrangeana, 2
- RVA, 72
- salva-guardas, 65
- semi-continuidade exterior, 7, 27
- semi-continuidade interior, 7, 27
- seqüência ergódica, 58
- Shor, *veja* regra
- subdiferencial, 11
- subespaço linear paralelo, 7, 11, 16
- subgradiente, 11
- superdiferencial, 16
- supergradiente, 67, 72
- tamanho de passo, 22, 30, 63, 67
- teorema
- de Carathéodory, 7
 - de separação de convexos, 10, 24
- vértice, 96
- VA, 66
- VA+VTVM, 86
- VTVM, 86