Universidade Federal do Piauí
Centro de Ciências da Natureza
Pós-Graduação em Matemática
Mestrado em Matemática

Métodos de linearização parcial em programação não linear

Tiago da Costa Menezes

Teresina - 2017

Tiago da Costa Menezes

Dissertação de Mestrado:

Métodos de linearização parcial em programação não linear

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Matemática, da Universidade Federal do Piauí, como requisito parcial para obtenção do grau de Mestre em Matemática.

Orientador:
Prof. Dr. Paulo Sérgio Marques dos Santos

Teresina - 2017

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Métodos de linearização parcial em programação não-linear

TIAGO DA COSTA MENEZES

Esta Dissertação foi submetida como parte dos requisitos necessários à obtenção do grau de **Mestre em Matemática**, outorgado pela Universidade Federal do Piauí.

A citação de qualquer trecho deste trabalho é permitida, desde que seja feita em conformidade com as normas da ética científica.

Dissertação aprovada em 01 de Março de 2017.

Banca Examinadora:

Paulo Sérgio Marques dos Santos

Prof. Dr. Paulo Sérgio Marques dos Santos - Orientador

João Carlos de Oliveira Souza

Prof. Dr. João Carlos de Oliveira Souza - Membro interno

Afonso Norberto da Silva

Prof. Dr. Afonso Norberto da Silva - Membro externo

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca Setorial do CCN

M543m Menezes, Tiago da Costa.
Métodos de linearização parcial em programação não linear / Tiago da Costa Menezes. – Teresina, 2017.
71f. il.

Dissertação (Mestrado) – Universidade Federal do Piauí, Centro de Ciências da Natureza, Pós-Graduação em Matemática, 2017.
Orientador: Prof. Dr. Paulo Sérgio Marques dos Santos.

1. Análises. 2. Otimização Matemática. 3. Programação Não Linear. I. Título

CDD 515

Dedico este trabalho aos meus pais e irmão;
à toda minha família.

Agradecimentos

Ao fim dessa etapa, agradeço acima de tudo a Deus por estar comigo hoje e sempre compartilhando de minhas derrotas e vitórias, dando força de vontade para chegar até aqui. Agradeço aos meus pais José Filho e Maria Bernadete, ao meu irmão Felipe, por serem o alicerce em todos os momentos, nunca interferindo em minhas escolhas acadêmicas, aconselhando e apoiando sempre.

Agradeço a toda minha família a qual compartilho essa vitória, pelo apoio e acompanhamento durante o percurso de estudos e aos que um dia serviram de grande inspiração, em especial aos meus avós José Francisco (In memoriam) e Rosa Dália, às minhas tias Elizabete, Francisca, Lideane, aos meus tios Djacir e Francisco (In memoriam) e primos.

A todos os professores do mestrado, em especial os professores Gleison, Humberto, José Francisco, Kelton e Marcos Vinicio pelo acompanhamento durante as disciplinas. Meu agradecimento em especial para o professor Paulo Sérgio pela excelente orientação, apoio e amizade desde a graduação.

A todos meus amigos de Parnaíba/Ilha Grande, em especial a Maria das Graças, Fabiana Nascimento, Gleison Pinho e sua família pela velha e sincera amizade. Aos amigos de graduação, em particular José Filho, Merison Santos, André Felipe e Hércules Bezerra pela enorme amizade. Por fim, agradeço a todos os amigos que tive o prazer de conhecer em Teresina, em especial a Maria do Carmo e Natasha Falcão a quem também agradeço por toda a confiança depositada em mim, pela amizade e companheirismo incondicional, apesar dos contratempos.

Aos amigos do mestrado Atécio, Cabeça (Hércules), Jeferson, Juliana, Kelvin, Lucas, Lucas Cassiano, Pádua, Quaresma, Rafael, Ray, Raul, Veludo (Antônio Aguiar) e aos demais, pelo companheirismo diário e por proporcionarem uma ótima convivência.

Agradeço (a CAPES) pelo apoio financeiro.

...E nossa história não estará pelo avesso

Assim, sem final feliz.

Teremos coisas bonitas pra contar.

E até lá, vamos viver

Temos muito ainda por fazer

Não olhe pra trás

Apenas começamos.

O mundo começa agora

Apenas começamos.

Renato Russo.

Resumo

Neste trabalho, caracterizamos uma classe de métodos de direções viáveis na programação não-linear, através do conceito de linearização parcial da função objetivo. Baseado em um ponto viável, a função objetivo é substituída por uma função arbitrária convexa e continuamente diferenciável, e o erro é levado em conta por uma aproximação de primeira ordem. Um novo ponto viável é definido através de uma busca linear com respeito ao objetivo original, na direção da solução do problema aproximado. Os resultados de convergência global são obtidos para buscas lineares exatas e aproximadas. Apresentamos alguns casos particulares do algoritmo geral e discutimos extensões para programação não diferenciável.

Palavras chave: Linearização parcial. Métodos de direções viáveis. Programação não-linear. Regularização.

Abstract

In this work, we characterize a class of feasible direction methods in nonlinear programming through the concept of partial linearization of the objective function. Based on a feasible point, the objective function is replaced by an arbitrary convex and continuously differentiable function, and the error is taken into account by a first-order approximation. A new feasible point is defined through a line search with respect to the original objective, toward the solution of the approximate problem. Global convergence results are given for exact and approximate line searches. We present some instances of the general algorithm and discuss extensions to nondifferentiable programming.

Key Words: Feasible direction methods, partial linearization, regularization, nondifferentiable programming.

Sumário

Introdução

Problemas de programação não-linear tem motivado um grande aumento da pesquisa científica acadêmica em modelagem matemática e no projeto de algoritmos numéricos, nas últimas décadas. A procura pela solução de um problema de otimização e pela estratégia que obtém uma solução com o menor custo, são tarefas que podem ser realizadas através do projeto de algoritmos computacionais.

Uma dificuldade existente em métodos de otimização iterativos é a de encontrar uma direção tal que os valores da função objetivo vão sempre decrescendo quando nos movimentamos naquela direção, são exemplos os métodos de descida explorados em [2], [3] e [5].

Uma das estratégias usadas para a solução de problemas de programação não-linear são os métodos de linearização parcial, onde na iteração k, usamos uma função $g(x, x^{(k)})$ além do objetivo original de minimizar a função objetivo f. Exemplos de casos como estes são os métodos de descida explorados em [2], [3] e [5]. Uma característica desejável é que sendo o problema um problema de otimização com restrições, a direção de descida não viole as restrições. Assim, é muito importante o estudo dos métodos de direções viáveis.

Essa dissertação tem como objetivo o estudo da unificação de uma série de métodos de direções viáveis na programação não-linear. A função objetivo f, inicialmente pseudoconvexa e continuamente diferenciável, em um conjunto convexo e compacto, é substituída por uma função arbitrária convexa e continuamente diferenciável, e o erro é levado em conta por uma aproximação de primeira ordem do mesmo, assim como em [8] e [9].

Para entendermos tal generalização o presente trabalho apresenta, no Capítulo 1, alguns resultados preliminares de análise convexa, em especial a noção de subdiferencial de uma função convexa.

No Capítulo 2, baseado em [2], [3] e [5] recordamos alguns métodos de descida entre eles, o método do gradiente, Newton e quase-Newton, o método do ponto proximal clássico e o de Frank-Wolfe.

No Capítulo 3, apresentamos o resultado principal, o algoritmo de linearização parcial bem como sua boa definição e convergência. Sob hipóteses adicionais, garantimos sua convergência global. Além disso, discutimos alguns casos particulares apresentados no capítulo anterior.

No Capítulo 4, estendemos o algoritmo de linearização parcial para o caso não diferenciável. Então, provamos a convergência de duas versões do algoritmo, onde a busca linear é feita aplicando uma modificação da regra de Armijo para o comprimento do passo.

No Capítulo 5, apresentamos algumas conclusões e nossas considerações finais.

Capítulo 1

Preliminares

Neste capítulo estudaremos alguns fatos básicos que serão necessários ao longo deste trabalho. Iremos abordar resultados que garantem a existência de soluções do seguinte problema:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & x \in \mathbb{R}^n, \end{aligned} \tag{1.1}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$. Veremos sob quais hipóteses o problema (1.1) tem solução. Além disso, estudaremos condições necessárias e suficientes de otimalidade para o Problema (1.1). Por fim, abordaremos alguns tópicos de análise convexa, procurando esclarecer a importância da convexidade na otimização. Maiores detalhes dos resultados abordados neste capítulo podem ser encontrados em [1], [4], [6], [7] e [11].

1.1 Definições e alguns fatos básicos

Definição 1.1.1. *([4]) Dizemos que um ponto $\bar{x} \in \mathbb{R}^n$ é*

1. ***Minimizador global*** *de (1.1) se*

$$f(\bar{x}) \leqslant f(y), \ \forall \ y \in \ \mathbb{R}^n;$$

2. ***Minimizador local*** *de (1.1) se existe uma vizinhança U de $\bar{x}$ tal que*

$$f(\bar{x}) \leqslant f(y), \ \forall \ y \in U \cap \mathbb{R}^n.$$

Se $\bar{x}$ é minimizador global então $f(\bar{x})$ é chamado **valor ótimo global**.

Definição 1.1.2. *([4]) Dizemos que* $\bar{v} \in [-\infty, +\infty)$ *definido por*

$$\bar{v} = \inf_{x \in \mathbb{R}^n} f(x),$$

é o valor ótimo do Problema (1.1).

Uma função pode admitir vários minimizadores globais, mas o valor ótimo (global) do problema é único. Apresentamos a seguir uma condição necessária de primeira ordem para caracterizar um minimizador para o problema (1.1).

Teorema 1.1.1. *([4])* ***(Condição necessária de primeira ordem)*** *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *diferenciável no ponto* $\bar{x} \in \mathbb{R}^n$. *Se* $\bar{x}$ *é um minimizador local de* f*, então*

$$\nabla f(\bar{x}) = 0. \tag{1.2}$$

Demonstração. Seja $d \in \mathbb{R}^n$ arbitrário, pela definição de minimizador local, temos que existe $\epsilon > 0$ tal que

$$f(\bar{x}) \leqslant f(\bar{x} + td), \ \forall \ t \in (0, \epsilon].$$

Daí, pela diferenciabilidade de f em $\bar{x}$,

$$f(\bar{x} + td) = f(\bar{x}) + t(\nabla f(\bar{x}))^T d + \theta(t),$$

onde $\theta(t)$ segue da definição de função diferenciável e é tal que $\lim_{t \to 0^+} \dfrac{\theta(t)}{t} = 0.$

$$0 \leqslant t(\nabla f(\bar{x}))^T d + \theta(t).$$

Dividindo por $t > 0$, obtemos

$$0 \leqslant (\nabla f(\bar{x}))^T d + \frac{\theta(t)}{t},$$

passando o limite quando $t \longrightarrow 0^+$, temos que

$$0 \leqslant (\nabla f(\bar{x}))^T d.$$

Como $d \in \mathbb{R}^n$ é fixo, porém arbitrário, podemos escolher $d = -\nabla f(\bar{x})$, o que resulta na condição $0 \leqslant (\nabla f(\bar{x}))^T d = -\|\nabla f(\bar{x})\|^2$. Com isso, segue que $\nabla f(\bar{x}) = 0$. □

Definição 1.1.3. *Um ponto* $\bar{x} \in \mathbb{R}^n$ *que satisfaz a condição (1.2) é chamado de ponto crítico ou estacionário da função* f.

A partir do Teorema 1.1.1 e Definição 1.1.3, temos que se f é diferenciável então as soluções locais do problema de minimização (1.1) são pontos estacionários. O mesmo valendo para problemas de maximização.

Abordaremos a seguir um resultado que é conhecido na literatura como Fórmula de Taylor, onde a prova pode ser encontrada em ([6]).

Teorema 1.1.2. *Seja* $f : U \longrightarrow \mathbb{R}^n$ *de classe* C^2 *no aberto* $U \subset \mathbb{R}^n$*. Fixado* $a \in U$*, para todo* $v = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$ *tal que* $a + v \in U$*, escrevamos*

$$f(a+v) - f(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\alpha_i + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\alpha_i\alpha_j + \theta(v),$$

as derivadas sendo calculadas no ponto a*. Então* $\lim_{v\to 0} \frac{\theta(v)}{|v|^2} = 0$.

Teorema 1.1.3. *([4]) **(Condição necessária de segunda ordem)***
Suponhamos que $f : \mathbb{R}^n \to \mathbb{R}$ *seja duas vezes diferenciável em* $\bar{x} \in \mathbb{R}^n$*. Se* $\bar{x}$ *é um minimizador local irrestrito de* f*, então vale (1.2) e a matriz hessiana de* f *em* $\bar{x}$ *é semidefinida positiva, ou seja,*

$$(\nabla^2 f(\bar{x})d)^T d \geqslant 0, \quad \forall\ d \in \mathbb{R}^n. \tag{1.3}$$

Demonstração. A condição (1.2) é satisfeita pelo Teorema 1.1.1. Agora, seja $d \in \mathbb{R}^n$ qualquer, porém fixo. Se $\bar{x}$ é minimizador local do problema (1.1), então para todo t estritamente positivo e suficientemente pequeno temos que

$$\begin{aligned} 0 &\leqslant f(\bar{x}+td) - f(\bar{x}) \\ &= (\nabla f(\bar{x}))^T td + \frac{1}{2}(\nabla^2 f(\bar{x})td)^T td + \theta(t^2) \\ &= \frac{t^2}{2}(\nabla^2 f(\bar{x})d)^T d + \theta(t^2), \end{aligned}$$

onde acima usamos a expansão da função em série de Taylor (ver [4]) e novamente o Teorema 1.1.1. Daí, dividindo ambos os lados por $t^2 > 0$, temos que

$$0 \leqslant \frac{1}{2}(\nabla^2 f(\bar{x})d)^T d + \frac{\theta(t^2)}{t^2}.$$

Por fim, passando o limite quando $t \to 0^+$ e usando o Teorema 1.1.2 citado acima temos que $(\nabla^2 f(\bar{x})d)^T d \geqslant 0,\ \forall\ d \in \mathbb{R}^n$. □

A recíproca do teorema anterior é verdadeira sob uma condição mais forte sobre a matriz hessiana de f, como enunciaremos abaixo.

Teorema 1.1.4. *([4]) (**Condição suficiente de segunda ordem**)*
Suponhamos que $f : \mathbb{R}^n \to \mathbb{R}$ *seja duas vezes diferenciáveis em* $\bar{x} \in \mathbb{R}^n$*. Se* $\bar{x}$ *é um ponto estacionário e se a matriz hessiana de* f *em* $\bar{x}$ *é definida positiva, então existem uma vizinhança* U *de* $\bar{x}$ *e* $\beta > 0$ *tais que*

$$f(x) - f(\bar{x}) \geqslant \beta\|x - \bar{x}\|^2, \ \forall \ x \in U. \tag{1.4}$$

Em particular, $\bar{x}$ *é um minimizador local estrito do problema (1.1). Reciprocamente, se vale (1.4), então* $\bar{x}$ *é um ponto estacionário do problema (1.1) e a matriz hessiana de* f *em* $\bar{x}$ *é definida positiva.*

Demonstração. Seja $\bar{x}$ ponto estacionário do problema (1.1) e que a hessiana de f em $\bar{x}$ seja definida positiva, i.e.,

$$(\nabla^2 f(\bar{x})d)^T d > 0, \ \forall \ d \in \mathbb{R}^n \setminus \{0\}. \tag{1.5}$$

Suponha por absurdo que não ocorra (1.4). Então, existe $\{x^k\} \subset \mathbb{R}^n \setminus \{x^k\}$ tal que $x^k \to \bar{x}$ e

$$f(x^k) - f(\bar{x}) < \frac{1}{k}\|x^k - \bar{x}\|^2, \quad \forall \ k.$$

Daí, como $\left\{\dfrac{x^k - \bar{x}}{\|x^k - \bar{x}\|}\right\}$ é limitada, ela possui pontos de acumulação. Portanto, a menos de subsequência, podemos supor sem perda de generalidade que $\left\{\dfrac{x^k - \bar{x}}{\|x^k - \bar{x}\|}\right\} \to d \in \mathbb{R}^n \setminus \{0\}$. Logo, temos que

$$\begin{aligned}
\frac{1}{k}\|x^k - \bar{x}\|^2 &> f(x^k) - f(\bar{x}) \\
&= \nabla f(\bar{x})^T(x^k - \bar{x}) + \frac{1}{2}(\nabla^2 f(\bar{x})(x^k - \bar{x}))^T(x^k - \bar{x}) + \theta(\|x^k - \bar{x}\|^2) \\
&= \frac{1}{2}(\nabla^2 f(\bar{x})(x^k - \bar{x}))^T(x^k - \bar{x}) + \theta(\|x^k - \bar{x}\|^2),
\end{aligned}$$

onde acima usamos que $\nabla f(\bar{x}) = 0$. Dividindo ambos os membros desta desigualdade por $\|x^k - \bar{x}\|^2$ e tomando o limite quando $k \to +\infty$, obtemos

$$0 \geqslant (\nabla^2 f(\bar{x})d)^T d,$$

o que gera uma contradição, por (1.5). Portanto, temos que vale (1.4). Em particular,

$$f(x) \geqslant f(\bar{x}), \ \forall \ x \in U,$$

ou seja, $\bar{x}$ é um minimizador local.

Reciprocamente, suponha que vale (1.4), então como tal condição implica que $\bar{x}$ é minimizador local de f então temos que tal ponto é estacionário (vide Teorema 1.1.1). Seja $d \in \mathbb{R}^n$ arbitrário porém fixo, então para todo $t \in \mathbb{R}$ suficientemente pequeno $\bar{x}+td \in U$. Logo, usando (1.4) e o fato de $\bar{x}$ ser ponto estacionário obtemos que

$$\begin{aligned} \beta t^2\|d\|^2 &= \beta\|(\bar{x}+td)-\bar{x}\|^2 \\ &\leqslant f(\bar{x}+td)-f(\bar{x}) \\ &= t(\nabla f(\bar{x}))^T d + \frac{t^2}{2}(\nabla^2 f(\bar{x})d)^T d + \theta(t^2) \\ &= \frac{t^2}{2}(\nabla^2 f(\bar{x})d)^T d + \theta(t^2). \end{aligned}.$$

Dividindo ambos os lados por t^2 e fazendo $t \to 0$, temos que

$$\beta\|d\|^2 \leqslant \frac{1}{2}(\nabla^2 f(\bar{x})d)^T d.$$

Portanto, como d é arbitrário temos que a hessiana é definida positiva. □

Note que as condições (1.2) e (1.3) não são suficientes para garantir que um ponto $\bar{x}$ é minimizador. Por exemplo, a função $f : \mathbb{R} \longrightarrow \mathbb{R}$ dada por $f(x) = x^3$, satisfaz tais condições em $\bar{x} = 0$, mas tal ponto não é minimizador local do problema (1.1). Por outro lado, (1.2) e o fato da hessiana ser definida positiva não são condições necessárias para que um ponto $\bar{x}$ seja minimizador. Como exemplo, podemos tomar $f : \mathbb{R} \longrightarrow \mathbb{R}$ dada por $f(x) = x^6$. Note que $\bar{x} = 0$ é ponto minimizador de (1.1). Mas essa função não possui hessiana definida positiva nesse ponto.

1.2 Existência de soluções

No caso em que f é ilimitada inferiormente na Definição 1.1.2, temos que $\bar{v} = -\infty$, e o problema de minimização não possui solução global. Porém, mesmo quando $\bar{v}$ é finito o minimizador pode não existir, como é o caso de $f(x) = e^x$ em $\mathbb{R}$. Nesta seção teremos por objetivo estudar alguns resultados básicos, porém indispensáveis para garantir a existência de soluções para o problema (1.1). Por exemplo, a continuidade da função estudada e a compacidade do conjunto em questão garantem a existência de um minimizador em tal conjunto, como veremos a seguir. Alguns desses resultados são bem conhecidos na literatura e por isso algumas demonstrações serão omitidas.

O primeiro resultado desta seção traz a importância da compacidade do domínio e continuidade da função estudada para garantirmos a existência de solução do problema de maximização e minimização. Tal resultado é bem conhecido na literatura, em especial na Análise Matemática, como Teorema de Weierstrass, onde sua demonstração pode ser encontrada em ([4]).

Teorema 1.2.1. *([4]) Sejam $X \subset \mathbb{R}^n$ um conjunto compacto não-vazio e $f : X \to \mathbb{R}$ uma função contínua. Então, os problemas de minimização e maximização de f em X têm soluções globais.*

Como explica o autor em [4] o exemplo da função $f(x) = e^x$ mencionada acima reforça a importância do conjunto (viável) em questão ser compacto. Daí, tal hipótese só poderá ser retirada e ainda assim garantirmos a existência de soluções se reforçamos as hipóteses com respeito a função objetivo em questão. Daí, surge a importância da seguinte definição:

Definição 1.2.1. *([4]) O conjunto de nível da função $f : X \to \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por*

$$L_{f,X}(c) = \{x \in X \mid f(x) \leqslant c\}.$$

Corolário 1.2.1. *([4]) Sejam $X \subset \mathbb{R}^n$ e $f : X \to \mathbb{R}$ contínua no conjunto X. Suponhamos que existe $c \in \mathbb{R}$ tal que o conjunto de nível $L_{f,X}(c)$ seja não-vazio e compacto. Então o problema de minimização f em X possui uma solução global.*

Demonstração. Pelo Teorema 1.2.1 o problema (1.1) tem solução global em $L_{f,X}(c)$, digamos $\bar{x}$, isto é, $f(\bar{x}) \leqslant f(x), \forall\, x \in L_{f,X}(c)$. Agora seja $x \in X \setminus L_{f,X}(c)$ qualquer, temos que $f(x) > c \geqslant f(\bar{x})$, em outras palavras $\bar{x}$ é um minimizador global de f em todo conjunto X.

□

Definição 1.2.2. *Dizemos que $f : X \to \mathbb{R}$ é coerciva no conjunto X, quando para cada sequência $\{x^k\} \subset X$ tal que ou $\|x^k\| \to \infty$ ou $\{x^k\} \to x \in \bar{X} \setminus X (k \to \infty)$,tem-se $\lim_{k\to\infty} \sup\ f(x^k) = +\infty$.*

Corolário 1.2.2. *([4]) Sejam $X \subset \mathbb{R}^n$ e $f : X \to \mathbb{R}$ uma função contínua coerciva em $X \neq \emptyset$. Então, o problema de minimizar f em D possui uma solução global.*

Demonstração. Com $x \in X$ arbitrário, definimos $c = f(x)$. Daí, obtemos que $L_{f,X}(c) \neq \emptyset$. Iremos mostrar que tal conjunto é compacto. Suponhamos inicialmente que exista $\{x^k\} \subset L_{f,X}(c)$ tal que $\{x^k\} \longrightarrow x$, onde $x \notin L_{f,X}(c)$. Usando que

$$L_{f,X}(c) = X \cap \{x \in \mathbb{R}^n \mid f(x) \leqslant c\},$$

onde o segundo conjunto é fechado pelo fato de f ser contínua por hipótese. Isto nos diz que $\{x^k\} \longrightarrow x \in \bar{X} \setminus X$. Com isso, devido a coercividade de f em X temos que $\lim_{k\to\infty} \sup f(x^k) = +\infty$. Mas isso contradiz o fato de $\{x^k\} \subset L_{f,X}(c)$. Portanto, obtemos que $L_{f,X}(c)$ é fechado.
Agora suponha $L_{f,X}(c)$ ilimitado, isto é, podemos tomar uma sequência $\{x^k\} \subset L_{f,X}(c)$ tal que $\{x^k\} \longrightarrow +\infty$. Então pela fato de f ser coerciva em X e $\{x^k\} \subset X$, $\lim_{k\to\infty} \sup f(x^k) = +\infty$. Mas $f(x^k) \leqslant c, \ \forall \ k \in \mathbb{N}$. Portanto, $L_{f,X}(c)$ é limitado. □

1.3 Elementos de análise convexa

Nesta seção apresentamos tópicos de análise convexa, abordando alguns fatos básicos sobre conjuntos convexos e funções convexas que serão importantes para o desenvolvimento deste trabalho. É grande a importância da convexidade na otimização, pois com tal hipótese, as condições necessárias de otimalidade se tornam suficientes . Isto é, todo ponto estacionário é solução do problema de minimização. Além disso, o estudo de alguns resultados de análise convexa são de grande valia para termos um maior entendimento do teorema principal deste trabalho.

Definição 1.3.1. *([4]) Um conjunto* $X \subset \mathbb{R}^n$ *é convexo se, para quaisquer elementos* $x, \ y \in X$, *temos*

$$tx + (1-t)y \in X \ \textit{para todo} \ 0 \leqslant t \leqslant 1.$$

Em outras palavras, se x e y são pontos quaisquer do conjunto X então o segmento de reta com extremos x e y pertence à X. É chamado de combinação convexa de x e y o ponto $tx + (1-t)y$. Exemplos triviais de conjunto convexo é o conjunto vazio, o espaço n-dimensional $\mathbb{R}^n$, um conjunto que contém só um ponto.

Exemplo 1.3.1. *Todo semi-espaço em* $\mathbb{R}^n$*, isto é,*

$$C = \{x \in \mathbb{R}^n : a^T x \leqslant c\},$$

onde $a \in \mathbb{R}^n$ *e* $c \in \mathbb{R}$ *é convexo.*

De fato, sejam $x, y \in C$, logo $a^T x \leqslant c$ e $a^T y \leqslant c$. Então, para $tx + (1-t)y$, com $t \in [0,1]$, temos:

$$a^T(tx + (1-t)y) \leqslant tc + (1-t)c = c.$$

Portanto, $tx + (1-t)y \in C$ e, assim, C é um conjunto convexo.

Definição 1.3.2. *([4]) Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é dita **convexa** quando para todos $x, y \in \mathbb{R}^n$ tivermos*

$$f((1-t)x + ty) \leqslant (1-t)f(x) + tf(y), \tag{1.6}$$

para todo $t \in [0,1]$. A função f diz-se estritamente convexa quando a desigualdade acima é estrita para qualquer $x \neq y$ e $t \in (0,1)$.

Observação 1.3.1. *Note que a condição (1.6) é equivalente a $f(x + t(y-x)) \leqslant f(x) + t\,(f(y) - f(x))$.*

Definição 1.3.3. *([4]) Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é dita **fortemente convexa** com módulo $L > 0$ quando para todos $x, y \in \mathbb{R}^n$ tivermos*

$$f((1-t)x + ty) \leqslant (1-t)f(x) + tf(y) - Lt(1-t)\|x-y\|^2, \tag{1.7}$$

para todo $t \in [0,1]$.

Observação 1.3.2. *Note que se f é fortemente convexa então é estritamente convexa, em particular também é convexa.*

Exemplo 1.3.2. *A função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^2$ é convexa.*

De fato,

$$\begin{aligned} f(x + t(y-x)) &= x^2 + 2tx(y-x) + t^2(y-x)^2 \\ &\leqslant x^2 + 2tx(y-x) + t(y-x)^2 \\ &= x^2 + t(y^2 - x^2) \\ &= (1-t)x^2 + ty^2, \ \forall\ t \in [0,1]. \end{aligned}$$

Exemplo 1.3.3. *([4]) Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função convexa e sejam $x \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$ quaisquer. Então a função $\psi : \mathbb{R} \longrightarrow \mathbb{R}$, $\psi(\alpha) = f(x + \alpha d)$, é convexa.*

De fato, sejam $\alpha_1,\ \alpha_2 \in \mathbb{R}$ e $\lambda \in [0,1]$, temos que

$$\begin{aligned}
\psi(\lambda\alpha_1 + (1-\lambda)\alpha_2) &= f(x + (\lambda\alpha_1 + (1-\lambda)\alpha_2)d)\\
&= f(x + \lambda\alpha_1 d + (1-\lambda)\alpha_2 d)\\
&= f(\lambda(x + \alpha_1 d) + (1-\lambda)(x + \alpha_2 d))\\
&\leqslant \lambda f(x + \alpha_1 d) + (1-\lambda)f(x + \alpha_2 d)\\
&= \lambda\psi(\alpha_1) + (1-\lambda)\psi(\alpha_2).
\end{aligned}$$

Portanto, ψ é convexa.

Exemplo 1.3.4. *A função $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(x) = \|x\|^2$ é fortemente convexa com módulo $L = 1$.*

Dados $x, y \in \mathbb{R}^n$ e $t \in [0,1]$ temos que,

$$\begin{aligned}
f(tx + (1-t)y) &= t^2\|x\|^2 + 2t(1-t)x^T y + (1-t)^2\|y\|^2\\
&\leqslant t^2\|x\|^2 + 2t(1-t)\|x\|\|y\| + (1-t)^2\|y\|^2\\
&= t^2\|x\|^2 + 2t(1-t)\|x\|\|y\| + \|y\|^2 - 2t\|y\|^2 + t^2\|y\|^2\\
&+ (1-t)\|y\|^2 + (t-1)\|y\|^2 + t\|x\|^2 - t\|x\|^2\\
&= tf(x) + (1-t)f(y) + t\|y\|^2(t-1) + t\|x\|^2(t-1)\\
&+ 2t(1-t)\|x\|\|y\|\\
&= tf(x) + (1-t)f(y) - t\|y\|^2(1-t) - t\|x\|^2(1-t)\\
&+ 2t(1-t)\|x\|\|y\|\\
&= tf(x) + (1-t)f(y) - t(1-t)(\|x\| - \|y\|)^2\\
&\leqslant tf(x) + (1-t)f(y) - t(1-t)\|x - y\|^2
\end{aligned}$$

Definição 1.3.4. *([4]) O epígrafo da função $f : X \to \mathbb{R}$ é o conjunto*

$$E_f = \{(x, c) \in X \times \mathbb{R} \mid f(x) \leqslant c\}.$$

Teorema 1.3.1. *([4]) Seja $X \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : X \to \mathbb{R}$ é convexa em D se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Demonstração. Seja f uma função convexa e (x, c_1) e (y, c_2) pontos de E_f. Como pela Definição 1.3.4, $f(x) \leqslant c_1$ e $f(y) \leqslant c_2$ temos pela convexidade de f que para todo $\alpha \in [0,1]$ ocorre o seguinte:

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y) \leqslant \alpha c_1 + (1-\alpha)c_2$$

logo,

$$\alpha(x, c_1) + (1 - \alpha)(y, c_2) = (\alpha x + (1 - \alpha)y, \alpha c_1 + (1 - \alpha)c_2) \in E_f.$$

Em outras palavras, temos que E_f é convexo.

Reciprocamente, suponha E_f convexo. Daí, dados $x, y \in X$, como $(x, f(x)), (y, f(y)) \in E_f$ temos que para todo $\alpha \in [0, 1]$

$$(\alpha x + (1 - \alpha)y, \alpha f(x) + (1 - \alpha)f(y)) = \alpha(x, f(x)) + (1 - \alpha)(y, f(y)) \in E_f.$$

Logo, pela definição anterior temos que

$$f(\alpha x + (1 - \alpha)y) \leqslant \alpha f(x) + (1 - \alpha)f(y),$$

ou seja, f é convexa. $\square$

Levando em conta que estamos interessados em resolver o problema (1.1), as seguintes definições que envolve o comportamento da função objetivo se tornam essenciais.

Definição 1.3.5. *Dizemos que* $d \in \mathbb{R}^n$ *é uma direção viável em relação ao conjunto* X *no ponto* $\bar{x} \in X$*, quando existe* $\epsilon > 0$ *tal que*

$$\bar{x} + td \in X, \forall\ t \in [0, \epsilon].$$

Denotamos por $V_X(\bar{x})$ *o conjunto de todas as direções viáveis em relação ao conjunto* X *no ponto* $\bar{x} \in X$.

Definição 1.3.6. *Dizemos que* $d \in \mathbb{R}^n$ *é uma direção de descida de* $f : \mathbb{R}^n \to \mathbb{R}$ *no ponto* $\bar{x} \in \mathbb{R}^n$*, se existe* $\epsilon > 0$ *tal que*

$$f(\bar{x} + td) < f(\bar{x}), \forall\ t \in (0, \epsilon].$$

Denotamos por $X_f(\bar{x})$ *o conjunto de todas as direções de descida da função* f *no ponto* $\bar{x}$.

Note que o conjunto $X_f(\bar{x})$ pode ser vazio, por exemplo, este é o caso quando $\bar{x}$ é um minimizador irrestrito de f, mesmo que seja local.

Proposição 1.3.1. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável no ponto* $\bar{x} \in \mathbb{R}^n$*. Então:*

a) Para todo $d \in X_f(\bar{x})$*, tem-se que* $\nabla f(\bar{x})^T d \leqslant 0$.

b) Se $d \in \mathbb{R}^n$ *satisfaz* $\nabla f(\bar{x})^T d < 0$*, tem-se que* $d \in X_f(\bar{x})$.

Demonstração. Seja $d \in X_f(\bar{x})$. Para todo $t > 0$ suficientemente pequeno,

$$0 > f(\bar{x} + td) - f(\bar{x}) = t(\nabla f(\bar{x})^T d + o(t)/t).$$

Dividindo os dois lados da desigualdade acima por $t > 0$ e passando ao limite quando $t \to 0^+$, obtemos $0 \geqslant \nabla f(\bar{x})^T d$, o que prova o item (a).

Suponhamos agora que $\nabla f(\bar{x})^T d < 0$. Temos que

$$f(\bar{x} + td) - f(\bar{x}) = t(\nabla f(\bar{x})^T d + o(t)/t)$$

Em particular, para todo $t > 0$ suficientemente pequeno, temos

$$\nabla f(\bar{x})^T d + o(t)/t \leqslant \frac{1}{2} \nabla f(\bar{x})^T d < 0,$$

o que implica que $f(\bar{x} + td) - f(\bar{x}) < 0$, i.e., $d \in X_f(\bar{x})$. □

Se $\bar{x} \in X$ é um minimizador local de f sujeito a $x \in X$, é óbvio que não pode existir uma direção de descida de f no ponto $\bar{x}$ que ao mesmo tempo seja viável em relação ao conjunto X, i.e., necessariamente temos

$$X_f(\bar{x}) \cap V_X(\bar{x}) = \emptyset.$$

Pela propriedade 1.3.1, temos então que

$$\nabla f(\bar{x})^T d \geqslant 0, \forall\ d \in V_X(\bar{x}).$$

Definição 1.3.7. *([4]) Um conjunto $K \subset \mathbb{R}^n$ chama-se cone quando ele contém todos os múltiplos não-negativos de seus elementos, i.e.,*

$$d \in K \Rightarrow td \in K, \forall\ t \in \mathbb{R}_+.$$

Note que se K é um cone não-vazio, podemos concluir que $0 \in K$ e que de certa maneira, podemos interpretar um cone como um conjunto de direções. Daí, temos que quando $X_f(\bar{x})$ é não-vazio, ele não é um cone, pois $0 \notin X_f(\bar{x})$. No entanto, $X_f(\bar{x}) \cup \{0\}$ é um cone não-vazio, no caso em que $X_f(\bar{x}) \neq \emptyset$.

Definição 1.3.8. *([4]) Sejam $X \subset \mathbb{R}^n$ um conjunto convexo e $\bar{x} \in X$. O cone normal no ponto $\bar{x}$ em relação ao conjunto X é dado por:*

$$\aleph_X(\bar{x}) = \{d \in \mathbb{R}^n \mid d^T(x - \bar{x}) \leqslant 0, \forall\ x \in X\}.$$

Outros dois tipos específicos de cone bastante usados na literatura, os quais iremos nos utilizar ao decorrer deste capítulo são os chamados cone tangente e cone de Bouligand, abordados abaixo.

Definição 1.3.9. *([4]) Dado $X \subset \mathbb{R}^n$ e $\bar{x} \in X$, chama-se de cone tangente no ponto $\bar{x}$ o conjunto de todas as direções tangentes ao conjunto X no ponto $\bar{x}$. De forma equivalente, pode ser definido como*

$$\mathcal{T}_X(\bar{x}) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{c} \forall \{t_k\} \subset \mathbb{R}_+, \{t_k\} \to 0^+, \\ \exists \{d^k\} \subset \mathbb{R}^n, \{d^k\} \to d, \text{ tal que} \\ \bar{x} + t_k d^k \in X \text{ para todo } k \in \mathbb{N} \end{array} \right\}.$$

Outra noção útil é a de cone (tangente) de Bouligand ou contingent cone, dado por:

$$\mathcal{B}_X(\bar{x}) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{c} \exists \{t_k\} \subset \mathbb{R}_+, \{t_k\} \to 0^+, \\ \exists \{d^k\} \subset \mathbb{R}^n, \{d^k\} \to d, \text{ tal que} \\ \bar{x} + t_k d^k \in X \text{ para todo } k \in \mathbb{N} \end{array} \right\}.$$

Note que pela Definição 1.3.9 temos que $\mathcal{T}_X(\bar{x}) \subset \mathcal{B}_X(\bar{x})$. Ou seja, a noção de cone (tangente) de Bouligand é mais geral que a de cone tangente.

Teorema 1.3.2. *([4]) Se $f : \mathbb{R}^n \to \mathbb{R}$ é convexa então todo minimizador local é global. Além disso, o conjunto dos minimizadores é convexo.*

Demonstração. Suponhamos que $x^* \in \mathbb{R}^n$ seja mínimo local, e suponha por contradição que este não é global, então existe algum $y \in \mathbb{R}^n$ tal que $f(y) < f(x^*)$. Como f é convexa, temos:

$$\begin{aligned} f((1-t)x^* + ty) &\leqslant (1-t)f(x^*) + tf(y) \\ &< (1-t)f(x^*) + tf(x^*) \\ &= f(x^*), \ \forall t \in [0,1]. \end{aligned}$$

Tomando t suficientemente pequeno temos que $(1-t)x^* + ty$ é arbitrariamente próximo de x^* com $f((1-t)x^*+ty) < f(x^*)$. Isso contradiz o fato de x^* ser mínimo local. Portanto, se f é convexa todo mínimo local é global.

Agora, seja $S \subset \mathbb{R}^n$ o conjunto dos minimizadores e $\bar{v} \in \mathbb{R}$ o valor ótimo do nosso Problema (1.1), isto é, $f(x) = \bar{v}, \forall\, x \in S$. Tomando $x, y \in S$ e $t \in [0,1]$, pela convexidade de f obtemos

$$\begin{aligned} f(tx + (1-t)y) &\leqslant tf(x) + (1-t)f(y) \\ &= t\bar{v} + (1-t)\bar{v} \\ &= \bar{v}. \end{aligned}$$

Como $\bar{v}$ é valor ótimo temos que

$$f(tx + (1-t)y) = \bar{v}.$$

Em outras palavras $tx + (1-t)y \in S$. □

Teorema 1.3.3. *([4]) Se* $f : \mathbb{R}^n \to \mathbb{R}$ *é estritamente convexa, então não pode haver mais de um minimizador.*

Demonstração. Seja $S \subset \mathbb{R}^n$ o conjunto dos minimizadores de f. Suponha por contradição que existam $x \in S$ e $\bar{x} \in S$, com $x \neq \bar{x}$. Seja $t \in (0, 1)$, fato de f ser estritamente convexa temos que

$$\begin{aligned} f(tx + (1-t)\bar{x}) &< tf(x) + (1-t)f(\bar{x}) \\ &= t\bar{v} + (1-t)\bar{v} \\ &= \bar{v}. \end{aligned}$$

Como S é convexo (pelo teorema anterior) temos que $tx + (1 - t\bar{x})$ é uma solução para o problema com valor menor do que $\bar{v}$. Mas isso gera uma contradição. Portanto, o minimizador deve ser único. □

O próximo resultado nos diz que toda função convexa é contínua em qualquer subconjunto aberto do seu domínio. Além disso, ainda podemos garantir que tal função será também localmente Lipschitz-contínua no interior do seu domínio como pode ser visto em [4].

Teorema 1.3.4. *(**Continuidade de funções convexas**) Sejam* $\Omega \subset \mathbb{R}^n$ *um conjunto convexo e aberto e* $f : \Omega \longrightarrow \mathbb{R}$ *função convexa em* Ω*. Então* f *é localmente Lipschitz-contínua em* Ω*. Em particular, segue que* f *é contínua em* Ω*.*

No caso em que a função estudada é diferenciável, a convexidade admite algumas caracterizações que são utéis para saber se uma função é convexa ou não. Tais caracterizações também serão essenciais no prosseguimento deste trabalho.

Teorema 1.3.5. *([4]) Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável. Então as seguintes propriedades são equivalentes:*

(i) A função f *é convexa;*

(ii) Para todo $x, y \in \mathbb{R}^n$,

$$f(y) \geqslant f(x) + \nabla f(x)^T (y - x).$$

(iii) Para todo $x, y \in \mathbb{R}^n$,

$$(\nabla f(y) - \nabla f(x))^T (y - x) \geqslant 0.$$

Quando f *é duas vezes diferenciável, as propriedades acima também são equivalentes a*

(iv) A matriz hessiana de f *é semi-definida positiva em todo* $x \in \mathbb{R}^n$*,*

$$(\nabla^2 f(x) d)^T d \geqslant 0, \quad \forall \ d \in \mathbb{R}^n.$$

Demonstração. Suponha f convexa, então para $x, y \in \mathbb{R}^n, \ \alpha \in (0, 1]$ quaisquer, definindo $d = y - x$, temos que

$$\begin{aligned} f(x + \alpha d) &= f(\alpha y + (1 - \alpha) x) \\ &\leqslant \alpha f(y) + (1 - \alpha) f(x), \end{aligned}$$

donde

$$\alpha(f(y) - f(x)) \geqslant f(x + \alpha d) - f(x).$$

Dividindo ambos os membros por $\alpha > 0$,e passando o limite com $\alpha \to 0^+$, obtemos

$$\begin{aligned} f(y) - f(x) &\geqslant \lim_{\alpha \to 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha} \\ &= \nabla f(x)^T d = \nabla f(x)^T (y - x). \end{aligned}$$

Logo,

$$f(y) \geqslant f(x) + \nabla f(x)^T (y - x).$$

Trocando agora x por y no item (ii), temos

$$f(x) \geqslant f(y) + \nabla f(y)^T (x - y).$$

Somando essa desigualdade com a de (ii), obtemos (iii).

Agora, sejam $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^n$. Pelo Teorema do Valor Médio, existe $\alpha \in (0, 1)$ tal que

$$f(y) - f(x) = \nabla f(x + \alpha(y - x))^T (y - x).$$

Usando (iii) para os pontos $(x+\alpha(y-x))$ e x, obtemos

$$\begin{aligned}\nabla f(x+\alpha(y-x))^T(y-x) &= \alpha^{-1}(\nabla f(x+\alpha(y-x)))^T(\alpha(y-x))\\ &\geqslant \alpha^{-1}(\nabla f(x))^T(\alpha(y-x))\\ &= \nabla f(x)^T(y-x)\end{aligned}$$

Combinando esta desigualdade com a igualdade anterior, obtemos (ii).

Reciprocamente, fazendo novamente $d = y - x$ e usando (ii) para os pontos x e $x+\alpha d$; y e $x+\alpha d$ temos que

$$f(x) \geqslant f(x+\alpha d) - \alpha(\nabla f(x+\alpha d))^T d$$

e

$$f(y) \geqslant f(x+\alpha d) + (1-\alpha)(\nabla f(x+\alpha d))^T d.$$

Multiplicando a primeira desigualdade por $1-\alpha \geqslant 0$ e a segunda por $\alpha \geqslant 0$ e somando ambas, obtemos

$$\begin{aligned}(1-\alpha)f(x)+\alpha f(y) &\geqslant (1-\alpha)(f(x+\alpha d) - \alpha(\nabla f(x+\alpha d))^T d\\ &+ \alpha(f(x+\alpha d) + (1-\alpha)(\nabla f(x+\alpha d))^T d)\\ &= f(x+\alpha d)\\ &= f((1-\alpha)x+\alpha y),\end{aligned}$$

ou seja, f é convexa.

No caso em que f é duas vezes diferenciável é suficiente mostrar que $(ii) \Leftrightarrow (iv)$. De fato, fixemos $x, d \in \mathbb{R}^n$ quaisquer, daí por (ii),

$$f(x+\alpha d) - f(x) \geqslant \alpha(\nabla f(x))^T d.$$

Usando ainda o fato de que f é duas vezes diferenciável,

$$\begin{aligned}0 &\leqslant f(x+\alpha d) - f(x) - \alpha(\nabla f(x))^T d\\ &= \frac{\alpha^2}{2}(\nabla^2 f(x)d)^T d + \theta(\alpha^2).\end{aligned}$$

Dividindo por $\alpha^2 > 0$ e fazendo $\alpha \to 0^+$, obtemos (iv). Supondo agora que ocorre (iv), sejam $x, y \in \mathbb{R}^n$, pelo Teorema do Valor Médio existe $\alpha \in (0,1)$ tal que

$$f(y) - f(x) - \nabla f(x)^T(y-x) = \frac{1}{2}(\nabla^2 f(x+\alpha(y-x))(y-x))^T(y-x) \geqslant 0.$$

Logo, segue (ii). $\square$

Exemplo 1.3.5. *Considere a seguinte função,* $f(x_1, x_2) = 2x_1^2 + 3x_1x_2 + 5x_2^2 - 2x_1 + 6x_2$. *Note que*

$$y^T \nabla^2 f(x) y = (y_1 \quad y_2) \begin{pmatrix} 4 & 3 \\ 3 & 10 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 4y_1^2 + 6y_1y_2 + 10y_2^2 \geqslant 0.$$

Então, pelo resultado anterior, a função f *é convexa.*

Vejamos agora um outro conceito de análise convexa que será utilizado no decorrer deste trabalho.

Definição 1.3.10. *([1]) Seja* X *um conjunto aberto não vazio em* $\mathbb{R}^n$*, e seja* $f : \mathbb{R}^n \to \mathbb{R}$ *diferenciável em* X*. A função* f *é dita ser pseudoconvexa se para cada* $x_1, x_2 \in X$ *com* $\nabla f(x_1)^T (x_2 - x_1) \geqslant 0$*, temos* $f(x_2) \geqslant f(x_1)$*, ou, de forma equivalente, se* $f(x_2) < f(x_1)$*, então* $\nabla f(x_1)^T (x_2 - x_1) < 0$*. A função* f *é dita pseudoconcava se* $-f$ *é pseudoconvexa.*
A função f *é dita ser estritamente pseudoconvexa se para cada* $x_1, x_2 \in X$ *distintos satisfazendo* $\nabla f(x_1)^T (x_2 - x_1) \geqslant 0$*, temos* $f(x_2) \geqslant f(x_1)$*,ou equivalentemente, se* $f(x_2) \leqslant f(x_1)$*, então* $\nabla f(x_1)^T (x_2 - x_1) < 0$*. A função f é dita estritamente pseudoconcava se* $-f$ *é estritamente pseudoconvexa.*

Observação 1.3.3. *Note que toda função* $f : \mathbb{R}^n \to \mathbb{R}$ *convexa e diferenciável em* X *é pseudoconvexa.*
De fato, para todo $x_1, x_2 \in X$ *tal que* $f(x_2) < f(x_1)$*, usando o teorema 1.3.5,* $f(x_2) \geqslant f(x_1) + \nabla f(x_1)^T (x_2 - x_1) \Rightarrow 0 > f(x_2) - f(x_1) > \nabla f(x_1)^T (x_2 - x_1)$*. Portanto,* f *é pseudoconvexa.*

Exemplo 1.3.6. *A função real* $f : \mathbb{R} \to \mathbb{R}$ *dada por* $f(x) = -e^{-x^2}$ *é um exemplo de função pseudoconvexa.*

Pois se $f'(\bar{x}) > 0$, então f é crescente para $x > \bar{x}$. Se $f'(\bar{x}) < 0$, então f é decrescente para $x < \bar{x}$ e se $f'(\bar{x}) = 0$, então $f(\bar{x}) < f(x)$ para todo $x \in \mathbb{R}$. Mas f não é convexa pois existe um $\hat{x}$, digamos $\hat{x} = 1$, tal que $f''(\hat{x}) < 0$.

Teorema 1.3.6. *([4])* ***(Derivada direcional de uma função convexa)*** *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *convexa. Então para todo* $x \in \mathbb{R}^n$*,* f *é diferenciável em cada direção* $d \in \mathbb{R}^n$*. Além disso,*

$$f(x + \alpha d) \geqslant f(x) + \alpha f'(x; d), \forall \alpha \in \mathbb{R}_+.$$

Demonstração. Fixamos $x \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$ arbitrários. Defina,

$$\psi : \mathbb{R} \longrightarrow \mathbb{R}, \ \psi(\alpha) = f(x + \alpha d).$$

Daí, precisamos mostrar a existência do seguinte limite

$$\begin{aligned} f'(x; d) &= \lim_{\alpha \to 0_+} \frac{f(x + \alpha d) - f(x)}{\alpha} \\ &= \lim_{\alpha \to 0_+} \frac{\psi(\alpha) - \psi(0)}{\alpha}. \end{aligned} \tag{1.8}$$

Como f é localmente lipschitz, pelo Teorema (1.3.4), existem $L > 0$ e $\bar{\alpha} > 0$ tais que

$$|f(x + \alpha d) - f(x)| \leqslant L\alpha \|d\| \ \ \forall \ \alpha \in [0, \bar{\alpha}].$$

Portanto,

$$\frac{|\psi(\alpha) - \psi(0)|}{\alpha} \leqslant L\|d\|,$$

ou seja, a função $\alpha^{-1}(\psi(\alpha) - \psi(0))$ é limitada no conjunto $[0, \bar{\alpha}]$. Iremos mostrar agora que essa mesma função também é não-decrescente em $[0, \bar{\alpha}]$. De fato, sejam $\beta \geqslant \alpha > 0$, então existe $t \in (0, 1]$ tal que $\alpha = t\beta = (1 - t)0 + t\beta$. Daí, devido a convexidade de ψ, Exemplo (1.3.3), temos que

$$\psi(\alpha) \leqslant (1 - t)\psi(0) + t\psi(\beta).$$

Portanto,

$$\begin{aligned} \frac{|\psi(\alpha) - \psi(0)|}{\alpha} &\leqslant \frac{(1 - t)\psi(0) + t\psi(\beta) - \psi(0)}{\alpha} \\ &= \frac{t(\psi(\beta) - \psi(0))}{t\beta} \\ &= \frac{\psi(\beta) - \psi(0)}{\beta}. \end{aligned}$$

Isso nos diz que a função $\alpha^{-1}(\psi(\alpha) - \psi(0))$ é não-decrescente quando $\alpha \to 0_+$. Como já vimos que ela é limitada, em particular inferiormente, segue que o limite (1.8) existe. Além disso,

$$\alpha^{-1}(\psi(\alpha) - \psi(0)) \geqslant f'(x; d),$$

ou seja,

$$f(x + \alpha d) \geqslant f(x) + \alpha f'(x; d).$$

□

Definição 1.3.11. *([4]) Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Dizemos que* $y \in \mathbb{R}^n$ *é um subgradiente de* f *no ponto* $x \in \mathbb{R}^n$ *se*

$$f(z) \geqslant f(x) + y^T(z - x), \quad \forall\, z \in \mathbb{R}^n. \tag{1.9}$$

O conjunto de todos os subgradientes de f *em* x *se chama o* ***subdiferencial*** *de* f *em* x*, o denotamos por* $\partial f(x)$*.*

Considere agora $z = x + \alpha d$, onde $d \in \mathbb{R}^n$, $\alpha > 0$, a definição acima de subgradiente pode ser escrita da seguinte forma,

$$y \in \partial f(x) \Leftrightarrow \frac{f(x + \alpha d) - f(x)}{\alpha} \geqslant y^T d, \ \forall\, d \in \mathbb{R}^n, \forall\, \alpha > 0. \tag{1.10}$$

Daí, pelo que foi visto na demonstração do Teorema (1.3.6), $\alpha^{-1}(f(x + \alpha d) - f(x))$ é uma função não-decrescente; além disso, ela possui um limite quando fazemos $\alpha \to 0_+$. Logo, (1.10) é equivalente à:

$$f'(x; d) = \lim_{\alpha \to 0_+} \frac{f(x + \alpha d) - f(x)}{\alpha} \geqslant y^T d, \forall\, d \in \mathbb{R}^n.$$

Em resumo temos que a definição de subgradiente pode também ser vista de uma forma mais prática, que é a seguinte:

$$\partial f(x) = \{y \in \mathbb{R}^n \mid f'(x; d) \geqslant y^T d, \ \forall\, d \in \mathbb{R}^n\}. \tag{1.11}$$

Os próximos dois resultados nos dizem que dois conjuntos convexos sem pontos em comum são separáveis. Tais resultados serão utilizados em algumas demonstrações no decorrer deste capítulo, mas omitiremos suas demonstrações, que podem ser encontrada em [4].

Teorema 1.3.7. ***(Separação)*** *Sejam* $X_1 \subset \mathbb{R}^n$ *e* $X_2 \subset \mathbb{R}^n$ *conjuntos convexos não-vazios tais que* $X_1 \cap X_2 = \emptyset$*. Então existem* $a \in \mathbb{R}^n \setminus \{0\}$ *e* $c \in \mathbb{R}$ *tais que*

$$a^T x^1 \leqslant c \leqslant a^T x^2 \ \forall\, x^1 \in X_1, \ \forall\, x^2 \in X_2.$$

Teorema 1.3.8. ***(Separação Estrita)*** *Sejam* $X_1 \subset \mathbb{R}^n$ *e* $X_2 \subset \mathbb{R}^n$ *conjuntos convexos, fechados e não-vazios. Suponhamos que um deles também seja limitado (logo, compacto). Então* $X_1 \cap X_2 = \emptyset$ *se, e somente se,existem* $a \in \mathbb{R}^n \setminus \{0\}$ *e* $c \in \mathbb{R}$ *tais que*

$$a^T x^1 < c < a^T x^2 \ \forall\, x^1 \in X_1, \ \forall\, x^2 \in X_2.$$

Teorema 1.3.9. *([4]) Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Então para todo* $x \in \mathbb{R}^n$*, o conjunto* $\partial f(x)$ *é convexo, compacto e não-vazio. Além disso, para todo* $d \in \mathbb{R}^n$*, tem-se*

$$f'(x; d) = max_{y \in \partial f(x)} y^T d. \tag{1.12}$$

Demonstração. Por (1.11), temos que

$$\begin{aligned} \partial f(x) &= \{y \in \mathbb{R}^n \mid y^T d \leqslant f'(x; d),\ \forall\ d \in \mathbb{R}^n\} \\ &= \cap_{d \in \mathbb{R}^n} \{y \in \mathbb{R}^n \mid y^T d \leqslant f'(x; d)\}. \end{aligned}$$

Como a intersecção de semi-espaços fechados é convexa e fechada, obtemos que $\partial f(x)$ é convexo e fechado. Resta mostrarmos que $\partial f(x)$ é limitado. Até então, ainda não sabemos que $\partial f(x) \neq \emptyset$, daí caso $\partial f(x) = \emptyset$, o subdiferencial é limitado. Suponhamos que exista uma sequência $\{y^k\} \subset \partial f(x)$ tal que $||y^k|| \to \infty$. Daí, considere para todo k, $d^k = \dfrac{y^k}{||y^k||}$, onde podemos supor, a menos de subsequência que $\{d^k\} \to d$, pois d^k é limitada. Então por (1.11),

$$||y^k|| = (y^k)^T d^k \leqslant f'(x; d^k).$$

Portanto,

$$\lim_{k \to \infty} \sup ||y^k|| \leqslant \lim_{k \to \infty} \sup\ f'(x; d^k) \leqslant f'(x; d) < \infty,$$

onde a segunda desigualdade segue de ([4], pag 173). Isto contradiz o fato de $||y^k|| \to \infty$. Portanto, segue que $\partial f(x)$ é limitado.

Por fim provemos que $\partial f(x) \neq \emptyset$. Para isso, mostremos (1.12) e ao mesmo tempo verifiquemos que $\partial f(x) \neq \emptyset$. De fato, fixemos $d \in \mathbb{R}^n$ arbitrário, então pelos cálculos acima

$$f'(x; d) \geqslant \max_{y \in \partial f(x)} y^T d,$$

onde a existência do máximo acima segue pelo fato de $\partial f(x)$ ser compacto, junto com o Teorema (1.2.1). Defina os seguintes dois conjuntos:

$$X_1 = \{(z, c) \in \mathbb{R}^n \times \mathbb{R} \mid c > f(z)\}$$

e

$$X_2 = \Big\{(z, c) \in \mathbb{R}^n \times \mathbb{R} \ \Big|\ c = f(x) + \alpha f'(x; d), z = x + \alpha d,\ \alpha \in \mathbb{R}_+\Big\}.$$

Note que ambos os conjuntos são convexos, onde X_1 é o epígrafo de f sem a sua fronteira, logo é convexo, pois f é convexa.

Se $X_1 \cap X_2 \neq \emptyset$, teríamos que

$$f(x + \alpha d) < f(x) + \alpha f'(x; d),$$

para algum $\alpha \in \mathbb{R}_+$, mas isso gera uma contradição com o Teorema (1.3.6). Portanto, temos que $X_1 \cap X_2 = \emptyset$. Então, pelo Teorema (1.3.7), existe $(u, \beta) \in \mathbb{R}^n \times \mathbb{R} \setminus \{0\}$ tal que

$$u^T z + \beta c \leqslant u^T(x + \alpha d) + \beta(f(x) + \alpha f'(x; d)), \tag{1.13}$$

$\forall\ z \in \mathbb{R}^n$, $\forall\ \alpha \in \mathbb{R}_+$, $\forall\ c > f(x)$. Se $\beta = 0$, então

$$u^T z \leqslant u^T(x + \alpha d),\ \forall\ d \in \mathbb{R}^n,$$

mas isso só pode acontecer quando $u = 0$. Como $(u, \beta) \neq 0$, logo $\beta \neq 0$.
Suponha $\beta > 0$, então escolhendo $z = x$ e $\alpha = 0$ em (1.13), temos que $\beta c \leqslant \beta f(x)$ para todos os c tais que $c > f(x)$, o que gera uma contradição. Portanto, obtemos que $\beta < 0$. Dividindo ambos os lados em (1.13) por $\beta < 0$, temos que

$$c + \left(\frac{u}{\beta}\right)^T (z - x) \geqslant f(x) + \alpha f'(x; d) + \alpha \left(\frac{u}{\beta}\right)^T d, \tag{1.14}$$

$\forall\ z \in \mathbb{R}^n$, $\forall\ \alpha \in \mathbb{R}_+$, $\forall\ c > f(z)$. Tomando os limites quando $\alpha \to 0^+$ e $c \to f(z)^+$, obtemos que

$$f(z) \geqslant f(x) - \left(\frac{u}{\beta}\right)^T (z - x),\ \forall\ z \in \mathbb{R}^n,$$

ou seja, $\dfrac{-u}{\beta} \in \partial f(x)$. Em outras palavras, segue que $\partial f(x) \neq \emptyset$.
Tomando ainda $\alpha = 1$ e $z = x$ em (1.14) e passando o limite com $c \to f(z)^+$, obtemos

$$0 \geqslant f'(x; d) + \left(\frac{u}{\beta}\right)^T d.$$

Logo,

$$-\left(\frac{u}{\beta}\right)^T d \geqslant f'(x; d),\ \frac{-u}{\beta} \in \partial f(x).$$

Daí, como já sabemos que $f'(x; d) \geqslant \max_{y \in \partial f(x)} y^T d$, então

$$f'(x; d) = \max_{y \in \partial f(x)} y^T d.$$

Segue o resultado. $\square$

Proposição 1.3.2. *([4]) Uma função convexa* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é diferenciável no ponto* $x \in \mathbb{R}^n$ *se, e somente se, o conjunto* $\partial f(x)$ *contém um elemento só. Neste caso,* $\partial f(x) = \{\nabla f(x)\}$ *(onde* $\nabla f(x)$ *denota o gradiente de* f *em* x*).*

Demonstração. Considere f diferenciável no ponto $x \in \mathbb{R}^n$. Então, pelo Teorema (1.3.5)

$$f(z) \geqslant f(x) + \nabla f(x)^T(z-x), \forall\, z \in \mathbb{R}^n,$$

ou seja, $\nabla f(x) \in \partial f(x)$. Seja agora $y \in \partial f(x)$, então para todo $d \in \mathbb{R}^n$, fazendo $z = x + d$ na Definição 1.3.11 temos

$$\begin{aligned} f(x) + y^T d &\leqslant f(x+d) \\ &= f(x) + \nabla f(x)^T d + \theta(\|d\|). \end{aligned}$$

Portanto,

$$(y - \nabla f(x))^T d \leqslant \theta(\|d\|).$$

Tomando,

$$d^k = \frac{y - \nabla f(x)}{k\|y - \nabla f(x)\|},$$

temos que $d^k \to 0$ e

$$\frac{\|y - \nabla f(x)\|}{k} \leqslant \theta(\frac{1}{k}),$$

mas isso implica que $y - \nabla f(x) = 0$. Concluímos então que $\partial f(x) = \{\nabla f(x)\}$.
Seja $\partial f(x) = \{y\}$, pelo Teorema (1.3.9),

$$f'(x; d) = y^T d, \ \forall\, d \in \mathbb{R}^n.$$

Escolhendo como d os elementos da base canônica do $\mathbb{R}^n$, temos que $y_i = \dfrac{\partial f(x)}{\partial x_i}$, $i = 1, ..., n$. Temos que $f'(x; d)$ é uma função linear em d de forma $\nabla f(x)^T d$, o que implica a diferenciabilidade de f em x.

□

A importância do estudo do subdiferencial e suas propriedades na teoria de otimização pode ser vista no seguinte resultado, onde iremos nos utilizar da Definição 1.3.8. Como foi visto anteriormente temos que sempre vale a inclusão $\mathcal{T}_X(\bar{x}) \subset \mathcal{B}_X(\bar{x})$. Mas ainda, se o conjunto X em questão for convexo temos por [4] que vale a igualdade $\mathcal{T}_X(\bar{x}) = \mathcal{B}_X(\bar{x})$. Usando este fato, vejamos o próximo resultado.

Teorema 1.3.10. *([4]) **(Condição de otimalidade para minimização de uma função convexa num conjunto convexo)***
Sejam $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função convexa e $X \subset \mathbb{R}^n$ um conjunto convexo. Então $\bar{x} \in \mathbb{R}^n$ é um minimizador de f em X se, e somente se,

$$\exists y \in \partial f(\bar{x}) \text{ tal que } y^T(x - \bar{x}) \geqslant 0, \quad \forall\, x \in X, \tag{1.15}$$

ou equivalentemente,

$$0 \in \partial f(\bar{x}) + \mathcal{N}_X(\bar{x}). \tag{1.16}$$

Em particular, $\bar{x}$ é minimizador de f em $\mathbb{R}^n$ se, e somente se, $0 \in \partial f(\bar{x})$.

Demonstração. Suponha inicialmente que $y^T(x - \bar{x}) \geqslant 0$ para todo $x \in X$, então temos que $-y \in \mathcal{N}_X(\bar{x})$. Pela definição de subgradiente, segue que

$$f(x) \geqslant f(\bar{x}) + y^T(x - \bar{x}) \geqslant f(\bar{x}), \ \forall \ x \in X,$$

isto é, $\bar{x}$ é minimizador de f em X.

Reciprocamente, suponhamos que $\bar{x}$ é um minimizador de f em X. Escolhemos qualquer $d \in \mathcal{T}_X(\bar{x}) = \mathcal{B}_X(\bar{x}),\ d \neq 0$. Logo, existem sequências $\{t_k\} \subset \mathbb{R}_+ \setminus \{0\}$ e $\{d^k\} \subset \mathbb{R}^n$ tais que $\{d^k\} \to d,\ \{t_k\} \to 0^+, \bar{x} + t_k d^k \in X$ para todo k. Daí, para todo k temos que

$$0 \leqslant f(\bar{x} + t_k d^k) - f(\bar{x}) = t_k f'(\bar{x}; d^k) + \theta(t_k).$$

Dividindo ambos os lados por $t_k > 0$ e passando o limite com $k \to +\infty$ concluímos que

$$f'(x; d) \geqslant 0, \ \forall \ d \in \mathcal{T}_X(\bar{x}). \tag{1.17}$$

Suponhamos que não ocorra (1.16), logo não existe $y \in \mathbb{R}^n$ tal que $-y \in \partial f(\bar{x})$ e $y \in \mathcal{N}_X(\bar{x})$. Logo,

$$(-\partial f(\bar{x})) \cap \mathcal{N}_X(\bar{x}) = \emptyset.$$

Como $\partial f(\bar{x})$ é convexo, compacto e não-vazio e $\mathcal{N}_d(\bar{x})$ é convexo, fechado e não vazio, pelo Teorema (1.3.8), existem $a \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tais que

$$-a^T y > c > a^T d, \ \forall \ y \in \partial f(\bar{x}), \ \forall \ d \in \mathcal{N}_X(\bar{x}). \tag{1.18}$$

Como $0 \in \mathcal{N}_X(\bar{x})$, temos que $c > 0$. Portanto, $a^T y < -c < 0$ para todo $y \in \partial f(\bar{x})$. Daí, usando (1.12), obtemos que

$$f'(\bar{x}; a) = \max_{y \in \partial f(\bar{x})} a^T y < 0. \tag{1.19}$$

Suponhamos que $a^T d > 0$ para algum $d \in \mathcal{N}_X(\bar{x})$. Tomando $td \in \mathcal{N}_X(\bar{x}),\ t > 0$ e fazendo $t \to +\infty$, temos uma contradição em (1.18), pois $c \in \mathbb{R}$ é fixo. Então, $a^T d \leqslant 0$ para todo $d \in \mathcal{N}_X(\bar{x})$. Portanto,

$$a \in (\mathcal{N}_X(\bar{x}))^* = ((\mathcal{T}_X(\bar{x}))^*)^* = \mathcal{T}_X(\bar{x}), \tag{1.20}$$

onde as desigualdades seguem de ([4], pag 118).

Daí, temos que (1.19) e (1.20) contradizem (1.17). Por fim, se $\bar{x}$ é um minimizador de f em $\mathbb{R}^n$ então

$$f(\bar{x}) \leqslant f(x), \; \forall\, x \in \mathbb{R}^n.$$

Logo,

$$f(x) \geqslant f(\bar{x}) + 0^T(x - \bar{x}),$$

isto é, $0 \in \partial f(\bar{x})$. Reciprocamente, se $0 \in \partial f(\bar{x})$ segue que $\bar{x}$ é minimizador de f em $\mathbb{R}^n$. □

Corolário 1.3.1. *Nas mesmas hipóteses do teorema, suponha ainda que f seja diferenciável. Então $\bar{x}$ é uma solução ótima de f em X se, e somente se, $\nabla f(\bar{x})^T(x - \bar{x}) \geqslant 0$ para todo $x \in X$.*

Observação 1.3.4. *Como consequência do resultado acima, obtemos que a condição $\nabla f(\bar{x}) = 0$ é necessária e suficiente para otimalidade no caso de minimização irrestrita de uma função convexa diferenciável.*

A seguir mostraremos que o subdiferencial de uma função convexa é limitado em conjuntos limitados e que tem certas propriedades de continuidade.

Proposição 1.3.3. *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e $X \subset \mathbb{R}^n$ um conjunto limitado. Então o conjunto $\cup_{x \in X} \partial f(x)$ também é limitado.*

Demonstração. Suponhamos que $X \subset \mathbb{R}^n$ seja limitado, mas que existam $\{x^k\} \subset X$ e $\{y^k\} \subset \mathbb{R}^n$ tais que $y^k \in \partial f(x^k)$ para todo k e $\|y^k\| \to \infty (k \to \infty)$. Definindo $d^k = y^k/\|y^k\|$ e tomando subsequências (se necessário), podemos admitir que $\{x^k\} \to x$ e $\{d^k\} \to d$ $(k \to \infty)$, sendo $\|d\| = 1$. Pelo teorema (1.3.9),

$$f'(x^k; d^k) \geqslant (y^k)^T d^k = \|y^k\| \to +\infty.$$

No entanto, segue de ([4], pag 173) que $\limsup_{k \to \infty} f(x^k; d^k) \leqslant f'(x; d) < \infty$. Esta Contradição mostra que não podem existir sequências com as propriedades descritas acima. Portanto, o conjunto $\cup_{x \in X} \partial f(x)$ é limitado. □

Proposição 1.3.4. *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa, $\{x^k\} \to x$ $(k \to \infty)$ e $y^k \in \partial f(x^k)$ para todo k. Então a sequência $\{y^k\}$ é limitada e todos os seus pontos de acumulação pertencem ao $\partial f(x)$.*

Demonstração. Como $y^k \in \partial f(x^k)$, temos que

$$f'(x^k; d) \geqslant (y^k)^T d, \ \forall\ d \in \mathbb{R}^n.$$

Pela proposição (1.3.4), a sequência $\{y^k\}$ é limitada. Seja $y \in \mathbb{R}^n$ qualquer ponto de acumulação de $\{y^k\}$, i.e., $\{y^{k_j}\} \to y$ $(j \to \infty)$. Tomando limite superior quando $j \to \infty$ na relação acima, temos que

$$(y^k)^T d \leqslant \limsup_{k\to\infty} f'(x^k; d) \leqslant f'(x; d), \ \forall\ d \in \mathbb{R}^n,$$

Portanto, $y \in \partial f(x)$. □

Ainda no que diz respeito ao subdiferencial de uma função convexa, e usando os resultados abordados acima, nosso objetivo agora será provar o seguinte resultado:

Proposição 1.3.5. *([4]) Sejam $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}, i = 1, \cdots, p$, funções convexas. Então vale que*

$$\partial\Big(\sum_{i=1}^p f_i(x)\Big) = \sum_{i=1}^p \partial f_i(x). \tag{1.21}$$

Demonstração. Provaremos o resultado para $p = 2$ (a generalização para o caso de $p > 2$ é óbvia). Seja $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = f_1(x)+f_2(x)$. Supondo que $y^i \in \partial f_i(x),\ i = 1, 2$, temos que

$$f_i(z) \geqslant f_i(x) + (y^i)^T(z - x), \ \forall\ z \in \mathbb{R}^n.$$

Somando as duas desigualdades, obtemos

$$f(z) = f_1(z) + f_2(z) \geqslant f_1(x) + f_2(x) + (y^1 + y^2)^T(z - x) = f(x) + (y^1 + y^2)^T(z - x),$$

Assim, $y^1 + y^2 \in \partial f(x)$. Portanto, $\partial f_1(x) + \partial f_2(x) \subset \partial f(x)$. Suponha agora, por contradição, que não vale a inclusão $\partial f(x) \subset \partial f_1(x) + \partial f_2(x)$, i.e., existe $y \in \partial f(x)$, com $f(x) = f_1(x) + f_2(x)$ tal que $y \notin \partial f_1(x) + \partial f_2(x)$. Os conjuntos $\partial f_1(x)$ e $\partial f_2(x)$ são convexos, compactos e não-vazios (Teorema 1.3.9). Portanto, $\partial f_1(x) + \partial f_2(x)$ é convexo, compacto e não-vazio. Como $y \notin \partial f_1(x)+\partial f_2(x)$, pelo Teorema 1.3.8, existem $a \in \mathbb{R}^n\backslash\{0\}$ e $c \in \mathbb{R}$ tais que

$$a^T(y^1 + y^2) < c < a^T y, \ \ \forall\ y^i \in \partial f_i(x), i = 1, 2.$$

Daí,

$$\begin{aligned} y^T a &< \sup_{y^i \in \partial f_i(x), i=1,2} (y^1 + y^2)^T a \\ &= \sup_{y^1 \in \partial f_1(x)} (y^1)^T a + \sup_{y^2 \in \partial f_2(x)} (y^2)^T a \\ &= \max_{y^1 \in \partial f_1(x)} (y^1)^T a + \max_{y^2 \in \partial f_2(x)} (y^2)^T a \\ &= f_1'(x; a) + f_2'(x; a) \\ &= f'(x; a). \end{aligned}$$

Porém, esta desigualdade gera uma contradição pelo fato de $y \in \partial f(x)$ (vide 1.11). $\square$

Capítulo 2

Métodos de programação não linear

Nossa intenção neste capítulo é abordar as ideias principais e os fundamentos matemáticos do desenvolvimento e da análise dos algoritmos. Serão tratados métodos de descida e técnicas de busca linear, que representam ideias muito importantes em otimização computacional.

2.1 Métodos de descida

Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função dada. Uma das estratégias mais naturais para resolver o problema irrestrito

$$\min f(x), x \in \mathbb{R}^n \tag{2.1}$$

é a seguinte. Dada uma aproximação $x^k \in \mathbb{R}^n$ da solução do problema, encontramos um ponto $x^{k+1} \in \mathbb{R}^n$ tal que

$$f(x^{k+1}) < f(x^k).$$

Uma realização desta estratégia básica é tomar uma direção $d^k \in \mathbb{R}^n$ tal que f é decrescente a partir do ponto x^k nessa direção, e computar um comprimento de passo $\alpha_k > 0$ que fornece um valor de f menor do que no ponto x^k, ou seja,

$$f(x^k + \alpha_k d^k) < f(x^k).$$

Assim obtemos o seguinte método iterativo $x^{k+1} = x^k + \alpha_k d^k$. Métodos deste tipo se chamam métodos de descida, e os apresentaremos a seguir.

A seguir, apresentamos as principais regras de busca linear, supondo que para um ponto x^k dado, uma direção $d^k \in X_f(x^k)$ já foi escolhida.

Regra da minimização uni-dimensional. Uma estratégia natural é minimizar a função objetivo na semi-reta $x^k + \alpha d^k$, $\alpha \geqslant 0$. O comprimento de passo α_k vem dado pela condição

$$f(x^k + \alpha_k d^k) = \min_{\alpha \geqslant 0} f(x^k + \alpha d^k),$$

i.e., α_k é uma solução do problema

$$\min \varphi_k(\alpha), \alpha \in \mathbb{R}_+, \tag{2.2}$$

onde

$$\varphi_k : \mathbb{R}_+ \to \mathbb{R}, \varphi_k(\alpha) = f(x^k + \alpha d^k).$$

Observe que o fato de $d^k \in X_f(x^k)$ garante que soluções do problema (2.2) estão no interior do conjunto viável. Portanto, para uma função f diferenciável no ponto x^{k+1}, segue que

$$0 = \varphi_k'(\alpha_k) = \nabla f(x^k + \alpha_k d^k)^T d^k = \nabla f(x^{k+1})^T d^k. \tag{2.3}$$

Em particular, se $\nabla f(x^{k+1}) \neq 0$, do ponto de vista geométrico, x^{k+1} é o ponto de interseção da semi-reta a partir de x^k da direção d^k com a curva de nível da função f que passa pelo ponto x^{k+1}. Este valor de α_k é o melhor possível, no sentido de que no ponto correspondente obtemos o menor valor de f entre todos os pontos da forma $x^k + \alpha d^k$, $\alpha \geqslant 0$.

A regra de Armijo. Outra estratégia natural é computar um comprimento de passo que resulta em decréscimo suficiente da função f em relação ao valor $f(x^k)$.
Suponhamos que f seja diferenciável no ponto x^k. Fixamos os parâmetros $\hat{\alpha} > 0$ e $\sigma, \theta \in (0, 1)$. Tomamos $\alpha := \hat{\alpha}$.

1. Verificamos se a desigualdade

   $$f(x^k + \alpha d^k) \leqslant f(x^k) + \sigma\alpha\nabla f(x^k)^T d^k \tag{2.4}$$

   se satisfaz ou não.

2. Se (2.4) não se satisfaz, tomamos $\alpha := \theta\alpha$ e retornamos ao Passo 1. Caso contrário, aceitamos $\alpha_k = \alpha$ como valor do comprimento de passo.

O resultado seguinte mostra que quando d^k satisfaz a condição suficiente de descida dada na proposição 1.3.1, então a regra de Armijo está bem definida.

Lema 2.1.1. ***(A regra de Armijo está bem definida)*** *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável no ponto* $x^k \in \mathbb{R}^n$. *Suponhamos que* $d^k \in \mathbb{R}^n$ *satisfaça* $\nabla f(x^k)^T(d^k) < 0$. *Então a desigualdade (2.4) é satisfeita para todo* $\alpha > 0$ *suficientemente pequeno. Em particular, a regra de Armijo está bem definida e termina com um* $\alpha_k > 0$.

Demonstração. Veja em [5]. □

A desigualdade (2.4), com $\alpha = \alpha_k$, fornece uma estimativa de quanto o valor de $f(x^{k+1})$ é menor do que $f(x^k)$. A prova da convergência fica mais fácil quando os valores do comprimento de passo α_k são uniformemente limitados inferiormente por algum $\bar{\alpha} > 0$. Neste sentido, o seguinte resultado é importante.

Lema 2.1.2. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável no* $\mathbb{R}^n$, *com derivada Lipschitz-contínua no* $\mathbb{R}^n$ *com módulo* $L > 0$.
Se $x^k, d^k \in \mathbb{R}^n$ *satisfazem a condição* $\nabla f(x^k)^T(d^k) < 0$, *então a desigualdade (2.4) é valida para todo* $\alpha \in (0, \bar{\alpha}_k]$, *onde*

$$\bar{\alpha}_k = -\frac{2(1-\sigma)(\nabla f(x^k))^T(d^k)}{L\|d^k\|^2} > 0. \tag{2.5}$$

Demonstração. Veja em [5]. □

A regra de Wolfe. O comprimento de passo $\alpha_k > 0$ é calculado para satisfazer as seguintes duas desigualdades(em relação a α):

$$f(x^k + \alpha d^k) \leqslant f(x^k) + \sigma_1 \alpha (\nabla f(x^k))^T(d^k) \tag{2.6}$$

$$\nabla f(x^k + \alpha d^k)^T(d^k) \geqslant \sigma_2 (\nabla f(x^k))^T d^k, \tag{2.7}$$

onde $0 < \sigma_1 < \sigma_2 < 1$.
A seguir apresentamos um procedimento de implementação da regra de wolfe.
Fixamos os parâmetros $0 < \sigma_1 < \sigma_2 < 1$. Tomar $\hat{\alpha} := 0$ e $\breve{\alpha} := 0$. Tomar um valor inicial $\alpha > 0$.

1. Se (2.6) e (2.7) se satisfazem, passar ao Item 6.

2. Se (2.6) é violada, tomar $\hat{\alpha} := \alpha$ e passar ao Item 5.

3. Se (2.7) é violada, tomar $\breve{\alpha} := \alpha$.

4. Se $\hat{\alpha} = 0$, escolher um novo valor $\alpha > \breve{\alpha}$ e passar ao Item 1.

5. Escolher um novo valor $\alpha \in (\check{\alpha}, \hat{\alpha})$ e passar ao Item 1.

6. Tomar $\alpha_k := \alpha$.

Lema 2.1.3. ***(a regra de Wolfe está bem definida)*** *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função limitada inferiormente no* $\mathbb{R}^n$*, diferenciável com derivada contínua. Suponhamos que* $x^k \in \mathbb{R}^n$ *e* $d^k \in \mathbb{R}^n$ *satisfaçam* $\nabla f(x^k)^T(d^k) < 0$.
Então qualquer implementação da regra de Wolfe, para a qual tem-se que $\check{\alpha} \to +\infty$ *no caso de número infinito de extrapolações e* $(\hat{\alpha} - \check{\alpha}) \to 0$ *no caso de número infinito de interpolações, está bem definida, no sentido de que ela termina com um valor de comprimento de passo* $\alpha_k > 0$.

Demonstração. Veja em [5]. □

2.2 O método do gradiente

Suponhamos que a função f seja diferenciável no $\mathbb{R}^n$. No caso de métodos de descida $x^{k+1} = x^k + \alpha_k d^k$, falamos do método do gradiente quando a direção de descida d^k se escolhe como sendo a direção de menos gradiente de f em x^k. Em particular, o esquema se reduz ao seguinte:

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k), \ k = 0, 1, \ldots \tag{2.8}$$

O estudo deste método é um passo importante para se entender os fundamentos da otimização computacional, embora seja pouco eficiente para resolver problemas práticos.

Algoritmo 2.2.1. ***(Método do gradiente)***
Escolher $x^0 \in \mathbb{R}^n$ *e tomar* $k := 0$. *Escolher uma regra de busca linear e os parâmetros associados:* $\hat{\alpha} > 0$ *(ou* $\hat{\alpha} = +\infty$*) no caso da minimização uni-dimensional,* $\hat{\alpha} > 0$, σ, $\theta \in (0, 1)$ *no caso da regra de Armijo.*

1. *Calcular o comprimento de passo* α_k *na direção do antigradiente* $d^k = -\nabla f(x^k)$ *de acordo com a regra escolhida (se* $\nabla f(x^k) = 0$*, na prática o método para; para fins teóricos, podemos tomar* α_k *arbitrário).*

2. *Calcular* x^{k+1} *pela fórmula (2.8).*

3. *Tomar* $k := k + 1$ *e retornar ao Passo 1.*

Teorema 2.2.1. ***(Convergência global do método do gradiente)***
Seja $f : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável no* $\mathbb{R}^n$*, com derivada contínua. Suponhamos que o Algoritmo 2.2.1 utiliza a minimização uni-dimensional ou a regra de Armijo. Então cada ponto de acumulação de qualquer sequência* $\{x^k\}$ *gerada pelo Algoritmo 2.2.1 é um ponto estacionário do problema (2.1). Se a sequência* $\{x^k\}$ *é limitada, então vale*

$$\{\nabla f(x^k)\} \to 0 \ (k \to \infty). \tag{2.9}$$

Demonstração. Veja em [5]. □

No caso de uma função convexa, as propriedades de convergência global do método do gradiente são bem mais fortes. Ressaltamos que a afirmação é válida mesmo se o conjunto de minimizadores for ilimitado.

Teorema 2.2.2. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa, diferenciável no* $\mathbb{R}^n$*, com derivada contínua. Suponhamos que o Algoritmo 2.2.1 utiliza a regra de Armijo com* $\hat{\alpha} \leqslant 1$*. Se o conjunto de minimizadores irrestritos de* f *é não-vazio, então qualquer sequência* $\{x^k\}$ *gerada pelo Algoritmo 2.2.1 converge a uma solução de (2.1).*

Demonstração. Veja em [5]. □

2.3 O método de Newton

O método clássico de Newton foi introduzido para encontrar uma solução da equação

$$\Phi(x) = 0, \tag{2.10}$$

onde $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ é diferenciável. Seja $x^k \in \mathbb{R}^n$ uma aproximação de alguma solução de (2.10). Então é natural aproximar a equação (2.10) próxima ao ponto x^k pela sua linearização

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) = 0. \tag{2.11}$$

A equação linearizada (2.11) fornece o sistema de iteração do método clássico de Newton. A idéia é transparente - a equação não-linear (2.10) é substituída pela equação linear (computacionalmente muito mais simples) (2.11).

Algoritmo 2.3.1. ***(O método de Newton)***
Escolher $x^0 \in \mathbb{R}^n$ *e tomar* $k = 0$.

1. *Calcular* $x^{k+1} \in \mathbb{R}^n$ *como solução da equação linear (2.11).*

2. *Tomar* $k := k+1$ *e retornar ao item 1.*

Assumindo que o $\Phi'(x^k)$ é não-singular (neste caso a solução da equação de Newton (2.11) é única), o método de Newton é muitas vezes apresentado na forma do esquema iterativo explícito

$$x^{k+1} = x^k - (\Phi'(x^k))^{-1}\Phi(x^k),\ k = 0, 1, \dots \tag{2.12}$$

Com a compreensão de que uma implementação real do método não precisa exigir o cálculo do inverso completo da matriz $\Phi(x^k)$.

Sob hipóteses apropriadas, o método de Newton é muito eficiente, o que se reflete nas seguintes declarações de convergência. Ao mesmo tempo, é claro que, na sua forma pura, o método pode não convergir se iniciar em um ponto que não está suficientemente próximos de uma solução, mesmo que este satisfaça todas as suposições necessárias. O que se segue descreve as propriedades de convergência essenciais do método de Newton. Para maiores informações veja [5].

Teorema 2.3.1. ***(Convergência local do método de Newton)*** *Seja* $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ *diferenciável em uma vizinhança de um ponto* $\bar{x} \in \mathbb{R}^n$*, com derivada contínua neste ponto. Seja* $\bar{x}$ *uma solução da equação (2.10), e suponha que* $\Phi'(\bar{x})$ *é uma matriz não-singular. Então qualquer ponto de partida* $x^0 \in \mathbb{R}^n$ *suficientemente próximo de* $\bar{x}$*, o Algoritmo 2.3.1 gera uma sequência* $\{x^k\}$ *bem definida que converge para* $\bar{x}$*. A taxa de convergência é superlinear, e se a derivada de* Φ *é Lipschitz-contínua em uma vizinhança de* $\bar{x}$*, então a taxa de convergência é quadrática.*

Demonstração. Veja em [5] □

Agora consideremos o problema de otimização irrestrita

$$\min f(x), x \in \mathbb{R}^n, \tag{2.13}$$

onde $f : \mathbb{R}^n \to \mathbb{R}^n$ é uma função duas vezes diferenciável no $\mathbb{R}^n$. Os pontos estacionários deste problema são caracterizados pela equação (2.10), onde

$$\Phi : \mathbb{R}^n \to \mathbb{R}^n,\ \Phi(x) = \nabla f(x).$$

Portanto, podemos tentar computar pontos estacionários de (2.13) aplicando o método de Newton à equação $\nabla f(\bar{x}) = 0$.

Seja $x^k \in \mathbb{R}^n$ uma aproximação a um ponto estacionário $\bar{x}$ do problema (2.13). A aproximação seguinte x^{k+1} é computada como solução do sistema de equações lineares

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0, \tag{2.14}$$

em relação a $x \in \mathbb{R}^n$. Supondo que $\nabla^2 f(x^k)$ seja não-singular para todo k, obtemos o esquema iterativo seguinte:

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k), k = 0, 1, \ldots \tag{2.15}$$

Observamos que esta iteração admite uma interpretação em termos de problemas de otimização. Em particular, em torno do ponto x^k, o problema (2.13) pode ser aproximado pelo seguinte problema de otimização irrestrita de uma função quadrática:

$$\min f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2} (\nabla^2 f(x^k)(x - x^k))^T (x - x^k), \;\; x \in \mathbb{R}^n. \tag{2.16}$$

Como é fácil ver, a equação de Newton (2.14) caracteriza os ponto estacionários do problema quadrático (2.16). Portanto, o método de Newton para um problema de otimização irrestrito é o seguinte.

Algoritmo 2.3.2. ***(O método de Newton para otimização irrestrita)***
Escolher $x^0 \in \mathbb{R}^n$ *e tomar* $k = 0$.

1. *Calcular* $x^{k+1} \in \mathbb{R}^n$ *como ponto estacionário do problema (2.16).*

2. *Tomar* $k := k + 1$ *e retornar ao Passo 1.*

Segue do Teorema 2.3.1, obtemos as seguintes propriedades do Algoritmo 2.3.2.

Corolário 2.3.1. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função duas vezes diferenciável em uma vizinhança de um ponto* $\bar{x} \in \mathbb{R}^n$*, com segunda derivada contínua neste ponto. Seja* $\bar{x}$ *um ponto estacionário do problema (2.13), que satisfaz a condição suficiente de segunda ordem.*
Então para qualquer ponto de partida $x^0 \in \mathbb{R}^n$ *suficientemente próximo de* $\bar{x}$*, o Algoritmo 2.3.2 gera uma sequência* $\{x^k\}$ *bem definida que converge para* $\bar{x}$*. A taxa de convergência é superlinear, e se a derivada segunda de* f *é Lipschitz-contínua em uma vizinhança de* $\bar{x}$*, então a taxa de convergência é quadrática.*

2.4 Métodos quase-Newton

Uma classe importante de métodos de otimização irrestrita é dada pelo esquema iterativo

$$x^{k+1} = x^k + \alpha_k d^k, \; d^k = -Q_k \nabla f(x^k), \; k = 0, 1, \dots, \tag{2.17}$$

onde, para todo k, $Q_k \in \mathbb{R}(n, n)$ é uma matriz simétrica definida positiva, e $\alpha_k > 0$ é o comprimento de passo calculado utilizando uma das regras de busca linear apresentadas na seção 2.1. Notemos que o método (2.17) é um método de descida. Com efeito, se x^k não é um ponto estacionário do problema (2.13), como Q_k é definida positiva, tem-se que

$$\nabla f(x^k)^T (d^k) = -(Q_k \nabla f(x^k))^T \nabla f(x^k) < 0, \tag{2.18}$$

e portanto, pela Proposição 1.3.1, d^k é uma direção de descida de f em x^k, i.e., $d^k \in X_f(x^k)$. No caso de $Q_k = E^n$ (matriz identidade de dimensão $n \times n$), (2.17) se reduz a um dos métodos de gradiente, e no caso de $Q_k = (\nabla^2 f(x^k))^{-1}$ com $\alpha_k = 1$, obtemos o método de Newton puro. Existem, porém, outras possibilidades que são mais atrativas computacionalmente do que as duas mencionadas acima.

Algoritmo 2.4.1. ***(Método de descida com métrica variável)***
Escolher $x^0 \in \mathbb{R}^n$ *e tomar* $k := 0$. *Escolher uma das regras de busca linear apresentadas na seção 2.1 e os parâmetros associados:* $\hat{\alpha} > 0$ *(ou* $\hat{\alpha} = +\infty$*) no caso da minimização uni-dimensional;* $\hat{\alpha} > 0$, σ, $\theta \in (0, 1)$ *no caso da regra de Armijo; e* $\bar{\alpha} > 0$ *no caso de comprimento de passo fixo.*

1. *Escolher uma matriz* $n \times n$ *simétrica definida positiva* Q_k *e tomar* $d^k = -Q_k \nabla f(x^k)$. *Calcular o comprimento de passo* α_k *na direção* d^k*, pela regra de busca linear escolhida (se* $\nabla f(x^k) = 0$*, formalmente definimos* $\alpha_k = 1$*).*

2. *Tomar* $x^{k+1} = x^k + \alpha_k d^k$.

3. *Tomar* $k := k + 1$ *e retornar ao Passo 1.*

Os métodos quase-Newton são métodos de forma (2.17), onde as matrizes Q_k são escolhidas para aproximar, num certo sentido, a matriz $(\nabla^2 f(\bar{x}))^{-1}$ para alguma solução $\bar{x}$.

2.5 O método do ponto proximal clássico

Neste capítulo iremos abordar o chamado Método do Ponto Proximal clássico. Tal método busca resolver o problema de minimização de uma função $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Na literatura existem diversas variações do método do ponto proximal, de acordo com a função distância adotada. Por exemplo, em [3] além do caso clássico que iremos abordar, onde adota-se a distância euclidiana, é possível encontrar quando adota-se a função de Bregman.

Nosso objetivo aqui será além de definir o método clássico, mostrar que o mesmo está bem definido e analisar a sua convergência. Em outras palavras, estamos interessados em saber sob quais hipóteses o método do ponto proximal clássico converge para uma solução do problema (1.1). Para maiores informações veja [3].

Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Considere a sequência $\{x^k\} \subset \mathbb{R}^n$ dada por:

$$\begin{aligned} &x^0 \in \mathbb{R}^n \\ &x^{k+1} := \operatorname{argmin}_{x \in \mathbb{R}^n}\{f(x) + \lambda_k \|x - x^k\|^2\} \end{aligned} \tag{2.19}$$

onde $\lambda_k \in \mathbb{R}$; $0 < \lambda_k \leqslant \tilde{\lambda}$, para algum $\tilde{\lambda} > 0$. Acima $\|.\|$ denota a norma euclidiana.

Proposição 2.5.1. *O método do ponto proximal definido em (2.19) está bem definido.*

Demonstração. Por indução, considere $f_k(x) = f(x)+\lambda_k\|x-x^k\|^2$. Como estamos supondo que f atinge o mínimo, temos que f é limitada inferiormente, logo $\lim\limits_{j\to\infty} f_k(y^j) = +\infty$, onde $\{y^j\} \subset \mathbb{R}^n$ com $\|y^j\| \longrightarrow +\infty$. De fato, como

$$\|y^j\| \leqslant \|y^j - x^k\| + \|x^k\|.$$

Então, $\|y^j - x^k\| \longrightarrow +\infty$, quando $j \longrightarrow +\infty$. Como sabemos que $\{f(y^j)\}$ é limitada inferiormente, usando a definição de f_k temos que $\lim\limits_{j\to\infty} f_k(y^j) = +\infty$. Em outras palavras, f_k é coerciva. Portanto, como $f_k(x)$ é claramente contínua, atinge o mínimo e é único, pois pelo Teorema 1.3.3, f_k é estritamente convexa, logo x^{k+1} é único.

□

A seguir mostraremos que sob algumas hipóteses a sequência gerada por (2.19) converge para o minimizador de f. Para isso iremos precisar da seguinte definição e proposição.

Definição 2.5.1. *([3]) Uma sequência* $\{y^k\} \subset \mathbb{R}^n$ *é dita Fejér convergente para o conjunto* $U \subset \mathbb{R}^n$*, com respeito a norma euclidiana se,*

$$\|y^{k+1} - u\| \leqslant \|y^k - u\| \ \forall k \geqslant 0, \ \forall \ u \in U. \tag{2.20}$$

Proposição 2.5.2. *([3]) Se* $\{y^k\}$ *é Fejér convergente para* $U \neq \emptyset$*, então* $\{y^k\}$ *é limitada. Se um ponto de acumulação* y *de* $\{y^k\}$*, tal que* $y \in U$*, então* $\lim_{k\to\infty} y^k = y$.

Demonstração. De (2.20) temos que,

$$\|y^{k+1} - u\| \leqslant \|y^0 - u\|,$$

$\forall \ u \in U$, isto é, $\{y^k\} \in B[u, r]$, para cada $k \in \mathbb{N}$, com $r = \|y^0 - u\|$. Onde $B[u, r]$ denota a bola fechada de centro em u e raio r, isto é, $\{y^k\}$ é limitada.

Agora seja $y \in U$ um ponto de acumulação de $\{y^k\}$, logo $\exists \{y^{k_j}\}$ tal que $y^{k_j} \longrightarrow y$. Por (2.20) temos que a sequência $\{\|y^k - y\|\}$ é decrescente e não-negativa (aqui fazemos $u = y$ em (2.20), já que por hipótese $y \in U$), logo temos que tal sequência converge com

$$\lim_{k\to\infty} \|y^k - y\| = 0.$$

Portanto, $\lim_{k\to\infty} y^k = y$.

□

Teorema 2.5.1. *([3]) Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *convexa e continuamente diferenciável. Suponha que o conjunto* U *dos minimizadores de* f *em* $\mathbb{R}^n$ *é não-vazio. Então a sequência* $\{x^k\}$ *gerada por (2.19) converge para algum ponto* $\bar{x} \in U$.

Demonstração. Sendo $\{x^k\}$ a sequência gerada em (2.19), dividiremos a demonstração em três etapas. Primeiramente iremos mostrar que a sequência é Fejér convergente para U, na segunda etapa mostraremos que $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$ e na terceira etapa iremos mostrar que qualquer ponto de acumulação da sequência em (2.19) é solução do nosso problema e a sequência converge para tal ponto.

$1^\circ \mathrm{Etapa}:$

Afirmação: $\|x^{k+1} - \bar{x}\|^2 \leqslant \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2, \forall \ k \geqslant 0, \forall \ \bar{x} \in U$.

Note que,

$$\|x^k - \bar{x}\|^2 = \|x^k - x^{k+1}\|^2 + \|x^{k+1} - \bar{x}\|^2 + 2(x^k - x^{k+1})^T(x^{k+1} - \bar{x}). \tag{2.21}$$

Como já sabemos, x^{k+1} é solução de (2.19) logo,

$$0 = \nabla f_k(x^{k+1}) = \nabla f(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k) \tag{2.22}$$

De (2.21), (2.22) e pela convexidade de f

$$\begin{aligned} \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2 - \|x^{k+1} - \bar{x}\|^2 &= 2(x^k - x^{k+1})^T(x^{k+1} - \bar{x}) \\ &= \frac{1}{\lambda_k}(\nabla f(x^{k+1}))^T(x^{k+1} - \bar{x}) \\ &\geqslant \frac{1}{\lambda_k}[\, f(x^{k+1}) - f(\bar{x})] \\ &\geqslant 0, \end{aligned}$$

onde a última desigualdade é devido ao Teorema 1.3.5 e pelo fato de $\bar{x}$ ser minimizador. Com isso obtemos que,

$$\begin{aligned} \|x^{k+1} - \bar{x}\|^2 &\leqslant \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2 \\ &\leqslant \|x^k - \bar{x}\|^2. \end{aligned}$$

Então, $\|x^{k+1} - \bar{x}\| \leqslant \|x^k - \bar{x}\|, \forall \bar{x} \in U$. Portanto, $\{x^k\}$ é Fejér convergente para U.

2°$\mathtt{Etapa}$: Da etapa 1 temos que

$$0 \leqslant \|x^{k+1} - x^k\|^2 \leqslant \|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2. \tag{2.23}$$

Pelo fato de $\{x^k\}$ ser Fejér convergente temos que $\{\|x^k - \bar{x}\|\}$ é não-crescente e não-negativa, logo convergente, isto é,

$$\|x^k - \bar{x}\| \longrightarrow 0. \tag{2.24}$$

Daí, por (2.23)

$$\lim_{k\to\infty}(x^{k+1} - x^k) = 0. \tag{2.25}$$

3°$\mathtt{Etapa}$: A existência de pontos de acumulação de $\{x^k\}$ segue da etapa 1, onde mostramos que $\{x^k\}$ é Fejer convergente e da Proposição 2.5.2. Agora, seja $\bar{x}$ um ponto de acumulação de $\{x^k\}$, então $x^{k_j} \longrightarrow \bar{x}$ para alguma subsequência x^{k_j}. Daí, por (2.22)

$$\nabla f(x^{k_j+1}) = 2\lambda_{k_j}(x^{k_j} - x^{k_j+1}).$$

Fazendo $j \longrightarrow +\infty$ e usando (2.24) e (2.25) temos que

$$\nabla f(\bar{x}) = 0.$$

Como $f \in C^1$ e é convexa, pelo Teorema 1.3.5 temos que $\bar{x} \in U$. Portanto, pela Proposição 2.5.2

$$\lim_{k\to\infty} x^k = \bar{x}.$$

□

2.6 O método de Frank-Wolfe

Em 1956, Frank e Wolfe desenvolveram um algoritmo para resolver problemas de programação quadrática com restrições lineares. É aplicável a problemas de programação não linear com funções objetivas convexas e discutimos essa versão aqui, para mais detalhes veja em [2].

Considere o problema

$$\begin{aligned} \min \quad & f(x), \\ s.a \quad & x \in X \end{aligned} \tag{2.26}$$

onde f é convexa e X uma região poligonal limitada.

No Algoritmo de Frank-Wolfe, a direção de busca d^k em cada estágio é determinada resolvendo um problema de programação linear apropriado. Começando com um ponto inicial $x^0 \in X$, no passo k substituímos f pela sua expansão da série Taylor de primeira ordem em torno do ponto x^k, isto é,

$$f^{(k)}(x) := f(x^{(k)}) + \nabla f(x^{(k)})^T (x - x^{(k)}) \tag{2.27}$$

Como x^k é fixo nesta fase do algoritmo, a minimização é equivalente ao problema

$$(FW - P)^{(k)} \begin{aligned} \min \quad & f^{(k)}(x) := \nabla f(x^{(k)})^T (x - x^{(k)}), \\ s.a \quad & x \in X \end{aligned} \tag{2.28}$$

Suponhamos agora que $\bar{x}^k$ é uma solução deste último problema. Então $\bar{x}^k$ é um ponto extremo do polígono X e, portanto, como X é convexo, a linha que une $\bar{x}^k$ e x^k está contida em X, então o vetor $\bar{x}^k - x^k$ é uma direção viável.

O que queremos fazer agora é mostrar que essa direção é também uma direção de descida. Para isso, devemos lembrar que a convexidade de f implica que uma condição necessária para um mínimo local é a desigualdade $\nabla f(x^{(k)})^T (x - x^{(k)}) > 0, \ \forall x \in X$ (veja corolário 1.3.1).

Sendo $\bar{x}^{(k)}$ uma solução do problema $(FW - P)^{(k)}$ temos

$$f^{(k)}(\bar{x}^{(k)}) \leqslant f^{(k)}(x^{(k)}) = \nabla f(x^{(k)})^T (x^{(k)} - x^{(k)}) = 0,$$

segue que ou $f^{(k)}(\bar{x}^{(k)}) = 0$ ou $f^{(k)}(\bar{x}^{(k)}) < 0$. No primeiro caso

$$f^{(k)}(\bar{x}^{(k)}) = 0 \leqslant f^{(k)}(x) = \nabla f(x^{(k)})^T (x - x^{(k)}), \ \forall x \in X.$$

Assim, $x^{(k)}$ é um mínimo de f em X, pelo corolário 1.3.1.

No segundo caso, temos que $\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) < 0$ isso garante que $d^{(k)} := \bar{x}^{(k)} - x^{(k)}$ é uma direção de descida.

Observe que, em cada etapa x^k é fixo, minimizar $\nabla f(x^k)^T(x - x^k)$ em relação a x é o mesmo que o problema de minimizar $\nabla f(x^k)^T x$.

Capítulo 3

Algoritmo de linearização parcial

O objetivo deste capítulo é unificar uma série de métodos de direção viável na programação não-linear: exemplos disso são os métodos de Frank e Wolfe e método de Newton com restrições. O problema estudado é

$$(\mathsf{P}) \begin{array}{ll} \min & f(x), \\ \mathrm{s.a} & x \in X \end{array} \tag{3.1}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é contínua, e X é assumido não vazio, compacto e convexo. Essas condições garantem a existência de uma solução para o problema (P). Seja uma função $g(x, y)$ convexa e continuamente diferenciável em relação a x, e contínua em relação a y. A ideia do método de linearização parcial, baseado em [9], é usar a função $g(x, x^{(k)})$ na iteração k em vez do objetivo original f, supostamente com um problema mais fácil do que (P) como resultado. O erro feito, $f(x) - g(x, x^{(k)})$, é levado em conta aproximando-o com uma expansão de Taylor de primeira ordem em torno do ponto viável $x^{(k)}$. A solução para este problema aproximado define uma direção de descida para a função objetivo original f, uma busca linear nessa direção dá um novo ponto de iteração. Estamos agora prontos para indicar o algoritmo e estabelecer a sua convergência global.

3.1 O algoritmo de linearização parcial

Para estabelecer as propriedades do algoritmo de linearização parcial, primeiro será indicado para um problema sobre um conjunto viável, compacto e convexo, onde o objetivo original f é pseudoconvexa e continuamente diferenciável. O algoritmo geral funciona da seguinte maneira. Na iteração k, a função objetivo é expressa como

$$f(x) = g(x, x^{(k)}) + (f(x) - g(x, x^{(k)})). \tag{3.2}$$

O segundo termo pode ser visto como a expressão do erro ao substituir a função objetivo original f pela função $g(\cdot, x^{(k)})$. A ideia é então levar esse erro em consideração, substituindo-o por uma expansão de Taylor de primeira ordem en torno de $x^{(k)}$. O subproblema resolvido na iteração k é então

$$(SUB - P)^{(k)} \begin{array}{ll} \min & f^{(k)}(x), \\ s.a & x \in X \end{array} \tag{3.3}$$

$$f^{(k)}(x) = g(x, x^{(k)}) + f(x^{(k)}) - g(x^{(k)}, x^{(k)}) + [\nabla f(x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})]^T (x - x^{(k)}),$$

Onde $\nabla_x g(x, y)$ denota o gradiente de $g(x, y)$ com relação a x. Uma solução ótima $\bar{x}^{(k)}$ para este problema convexo define uma direção $d^{(k)} := \bar{x}^{(k)} - x^{(k)}$, que é uma direção de descida viável em relação a f; Se não, $x^{(k)}$ resolve $(SUB-P)^{(k)}$, em que $x^{(k)}$ é uma solução ótima para (P), como é mostrado no Teorema 3.1.1. Uma busca linear com respeito a f é então feita na direção $d^{(k)}$ e um novo ponto de itereção x^{k+1} é obtido. Sob as hipóteses feitas em (P), qualquer ponto de acumulação da sequência $\{x^{(k)}\}_{k \geqslant 0}$ é ótimo para (P), para qualquer ponto inicial viável x^0, como mostra o Teorema 3.1.2.

Observação 3.1.1. *Note que $f^{(k)}$ é convexa, pois é soma de uma função convexa com uma aproximação linear. Segue da observação 1.3.3 que $f^{(k)}$ é pseudoconvexa.*

Teorema 3.1.1. *([9]) Assuma que $x^{(k)}$ é um ponto viável de (P), e que $\bar{x}^{(k)}$ resolve $(SUB - P)^{(k)}$. Se $x^{(k)}$ resolve $(SUB - P)^{(k)}$, então $x^{(k)}$ é ótimo em P. De outra forma, a direção $d^{(k)} = \bar{x}^{(k)} - x^{(k)}$ é uma direção de descida viável com respeito a f.*

Demonstração. Como $\bar{x}^{(k)}$ resolve $(SUB - P)^{(k)}$ e pela pseudoconvexidade da $f^{(k)}$ (Definição 1.3.10), temos:

$$f^{(k)}(\bar{x}^{(k)}) < f^{(k)}(x), \forall x \in X \Rightarrow \nabla f^{(k)}(x)^T (\bar{x}^{(k)} - x) < 0. \tag{3.4}$$

Se $x^{(k)}$ não resolve $(SUB - P)^{(k)}$, como é viável, temos:

$$\nabla f^{(k)}(x^{(k)})^T (\bar{x}^{(k)} - x^{(k)}) < 0. \tag{3.5}$$

Agora note que:

$$\nabla f^{(k)}(x) = \nabla g(x, x^{(k)}) + [\nabla f(x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})] \Rightarrow \nabla f^{(k)}(x^{(k)}) = \nabla f(x^{(k)}) \tag{3.6}$$

Por (3.5), temos $\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) < 0$. Portanto, $d^{(k)} = \bar{x}^{(k)} - x^{(k)}$ é uma direção viável pela proposição 1.3.1.

Como $\bar{x}^{(k)}$ é solução de $(SUB - P)^{(k)}$, pelo corolário 1.3.1, temos:

$$[\nabla g(\bar{x}^{(k)}, x^{(k)}) + \nabla f(x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})]^T(x - \bar{x}^{(k)}) \geqslant 0, \forall x \in X. \tag{3.7}$$

Caso $x^{(k)}$ seja solução de $(SUB - P)^{(k)}$, substituindo $\bar{x}^{(k)}$ por $x^{(k)}$, temos:

$$\nabla f(x^{(k)})^T(x - x^{(k)}) \geqslant 0, \forall x \in X.$$

Pela pseudoconvexidade de f, temos que $x^{(k)}$ resolve (P). □

Teorema 3.1.2. *([9]) O algoritmo de linearização parcial ou termina em um número finito de iterações ou gera uma sequência infinita $\{x^{(k)}\}_{k\geqslant 0}$, tal que qualquer ponto de acumulação é solução de (P).*

Demonstração. Se o algoritmo terminar em um número finito de iterações, temos pelo teorema (3.1.1) que:

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

e portanto $x^{(k)}$ é solução de (P). Assuma agora que $f^{(k)}(\bar{x}^{(k)}) \neq f^{(k)}(x^{(k)})$, para toda interação k. Pelo teorema anterior, $\{d^{(k)}\}$ é uma sequência de direções de descida. Portanto,

$$f(x^{(k+1)}) = \min_{\alpha_k \in [0, \bar{\alpha}_k]} f(x^{(k)} + \alpha_k(\bar{x}^{(k)} - x^{(k)})) < f(x^{(k)}), \tag{3.8}$$

onde $\bar{\alpha}_k$ é o comprimento máximo do passo na direção $d^{(k)}$. Assim $\{f(x^{(k)})\}$ é monótona decrescente e limitada inferiormente, pois X é compacto e f é contínua. Além disso, pelo Teorema 1.2.1, $\{x^{(k)}\}$ tem pelo menos um ponto de acumulação $x^* \in X$ (da qual é limite de alguma subsequência $\{x^{(k)}\}_{k\in K}$).

Temos então que

$$\lim_{k\in K} f(x^{(k)}) = f(x^*). \tag{3.9}$$

Segue que

$$\lim_{k\in K} \left[f(x^{(k+1)}) - f(x^{(k)})\right] = 0. \tag{3.10}$$

Novamente, pelo teorema de Weierstrass, seja $\bar{x}$ um ponto de acumulação da sequência $\{\bar{x}^{(k)}\}$, para todo $k \in \hat{K} \subset K$.

Por (3.8), temos que

$$f(x^{(k)} + \alpha_k(\bar{x}^{(k)} - x^{(k)})) \geqslant f(x^{(k+1)}),\ \forall\ \alpha_k \in [0, \bar{\alpha}_k],$$

Daí, $f(x^{(k)} + \alpha_k(\bar{x}^{(k)} - x^{(k)})) - f(x^{(k)}) \geqslant f(x^{(k+1)}) - f(x^{(k)}),\ \forall\ \alpha_k \in [0, \bar{\alpha}_k]$.

Por (3.10),

$$\lim_{k \in \hat{K}} \left[f(x^{(k)} + \alpha_k(\bar{x}^{(k)} - x^{(k)})) - f(x^{(k)}) \right] \geqslant 0;$$

e escolhendo α_k tal que $\lim\limits_{k \in \hat{K}} \alpha_k = 0$, temos

$$\lim_{k \in \hat{K}} \left[(f(x^{(k)} + \alpha_k(\bar{x}^{(k)} - x^{(k)})) - f(x^{(k)}))/\alpha_k \right] \geqslant 0$$

e o limite existe devido a diferenciabilidade estrita de f (vide [11], Teorema 4F). Tal limite é igual a derivada direcional de $f(x^*)$ na direção $\bar{x} - x^*$, assim

$$\nabla f(x^*)^T (\bar{x} - x^*) \geqslant 0. \tag{3.11}$$

Agora, como $\bar{x}^k$ resolve $(SUB - P)^{(k)}$, então por (3.7) (corolário 1.3.1)

$$\left[\nabla g(\bar{x}^{(k)}, x^{(k)}) + \nabla f(x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)}) \right]^T (x - \bar{x}^{(k)}) \geqslant 0,\ \forall\ x \in X.$$

Tomando o limite sobre $\hat{K}$ temos

$$\left[\nabla g(\bar{x}, x^*) + \nabla f(x^*) - \nabla_x g(x^*, x^*) \right]^T (x - \bar{x}) \geqslant 0,\ \forall\ x \in X. \tag{3.12}$$

Note que pela convexidade de g, usando o Teorema 1.3.5, temos $\left[\nabla_x g(\bar{x}, x^*) - \nabla_x g(x^*, x^*) \right]^T (\bar{x} - x^*) \geqslant 0$. Portanto, de 3.12, escolhendo $x = x^*$ temos

$$0 \geqslant \left[\nabla_x g(\bar{x}, x^*) - \nabla_x g(x^*, x^*) \right]^T (x^* - \bar{x}) \geqslant \nabla f(x^*)^T (\bar{x} - x^*).$$

Ou seja,

$$\nabla f(x^*)^T (x^* - \bar{x}) \geqslant 0.$$

Consequentemente, por (3.11), temos

$$\nabla f(x^*)^T (x^* - \bar{x}) = 0. \tag{3.13}$$

Se $\nabla f(x^*) = 0$ ou $x^* = \bar{x}$, então pela Definição 1.3.10 ou por (3.12) x^* resolve (P). Sendo assim, vamos assumir que $x^* \neq \bar{x}$ e que $\nabla f(x^*) \neq 0$.

De acordo com (3.12), $\bar{x}$ resolve o subproblema definido por x^*, que denotaremos por $f^*(x)$ (vide corolário 1.3.1). Então temos que $f^*(\bar{x}) \leqslant f^*(x^*)$, i.e.,

$$g(\bar{x}, x^*) + f(x^*) - g(x^*, x^*) + \left[\nabla f(x^*) - \nabla_x g(x^*, x^*) \right]^T (\bar{x} - x^*) \leqslant g(x^*, x^*) + f(x^*) - g(x^*, x^*) \Rightarrow$$

$$\Rightarrow g(\bar{x}, x^*) - g(x^*, x^*) + \left[\nabla f(x^*) - \nabla_x g(x^*, x^*) \right]^T (\bar{x} - x^*) \leqslant 0.$$

Note que, por outro lado, da convexidade de g, temos

$$g(\bar{x}, x^*) - g(x^*, x^*) - \nabla_x g(x^*, x^*)^T(\bar{x} - x^*) \geqslant 0.$$

Usando (3.13), obtemos

$$g(\bar{x}, x^*) = g(x^*, x^*) + \nabla_x g(x^*, x^*)^T(\bar{x} - x^*).$$

Mas isso, por sua vez, implica que

$$\begin{aligned}
f^*(\bar{x}) &= \lim_{k \in \hat{K}} f^{(k)}(\bar{x}^{(k)}) \\
&= \lim_{k \in \hat{K}} g(\bar{x}^{(k)}, x^{(k)}) + f(x^{(k)}) - g(x^{(k)}, x^{(k)}) \\
&+ \left[\nabla f(x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T (\bar{x}^{(k)} - x^{(k)}) \\
&= g(\bar{x}, x^*) + f(x^*) - g(x^*, x^*) + \left[\nabla f(x^*) - \nabla_x g(x^*, x^*)\right]^T (\bar{x} - x^*) \\
&= g(x^*, x^*) + \nabla_x g(x^*, x^*)^T(\bar{x} - x^*) + f(x^*) - g(x^*, x^*) - \nabla_x g(x^*, x^*)^T(\bar{x} - x^*) \\
&= f(x^*).
\end{aligned}$$

Mas como $f^*(x) = g(x, x^*) + f(x^*) - g(x^*, x^*) + \left[\nabla f(x^*) - \nabla_x g(x^*, x^*)\right]^T (x - x^*) \Rightarrow f^*(x^*) = f(x^*)$, implicando que x^* também resolve o problema definido por x^*, pois $f^*(x^*) = f^*(\bar{x})$. Reescrevendo $\bar{x}$ por x^* em (3.12), obtemos

$$\nabla f(x^*)^T(x - x^*) \geqslant 0, \ \forall \ x \in X,$$

que novamente pelo corolário 1.3.1 segue o resultado desejado. □

Agora iremos supor de que $f(x)$ tem gradiente Lipschitziano, i.e., que existe uma constante positiva L tal que

$$\|\nabla f(x^1) - \nabla f(x^2)\| \leqslant L\|x^1 - x^2\|, \ \forall \ x^1, x^2 \in X. \tag{3.14}$$

Mostraremos que o algoritmo de linearização parcial é globalmente convergente mesmo que as buscas lineares sejam feitas apenas aproximadamente. Uma condição suficiente para ∇f Lipschitz é f ser continuamente diferenciável, sendo o conjunto viável compacto (vide [11], Proposição 4A). Como por exemplo, considere o caso de substituir a busca linear exata pela regra do comprimento de passo dado por Armijo (veja a seção 2.1)onde o ponto $x^{(k+1)}$ é calculado como

$$x^{(k+1)} = x^{(k)} + \beta^{-\bar{l}}(\bar{x}^{(k)} - x^{(k)}),$$

onde $\bar{i}$ é o menor inteiro i para os quais

$$f(x^{(k)} + \beta^{-i}(\bar{x}^{(k)} - x^{(k)})) - f(x^{(k)}) \leqslant \epsilon\beta^{i}\nabla f(x^{(k)})^{T}(\bar{x}^{(k)} - x^{(k)}), \tag{3.15}$$

onde $\epsilon \in (0,1)$ e $\beta > 1$. O próximo resultado, estabelecemos a convergência global do algoritmo modificado.

Teorema 3.1.3. *([9]) Sob a condição adicional (3.14), o algoritmo de linearização parcial é globalmente convergente com a busca linear exata dada pela regra do comprimento de passo de Armijo (3.15).*

Demonstração. Vamos provar que qualquer subsequência convergente $\hat{K}$ da sequência $\{x^{(k)}\}_{k=1}^{\infty}$ e $\{\bar{x}^{(k)}\}_{k=1}^{\infty}$ satisfaz

$$\lim_{k\in\hat{K}} \nabla f(x^{(k)})^{T}(\bar{x}^{(k)} - x^{(k)}) = \nabla f(x^{*})^{T}(\bar{x} - x^{*}) = 0,$$

i.e., é equivalente a (3.13), daí a prova segue exatamente usando os argumentos na demonstração do Teorema 3.1.2.

Pelo Teorema 3.1.1, a sequência $\{d^{(k)}\}$ consiste em direções de descida viáveis. Consequentemente, pela proposição 1.3.1

$$\nabla f(x^{(k)})^{T}(\bar{x}^{(k)} - x^{(k)}) < 0,$$

e por (3.15), $f(x^{(k+1)}) \leqslant f(x^{(k)}),\ \forall\ k$.

De fato, pela construção da regra de comprimento do passo,

$$f(x^{(k+1)}) \leqslant f(x^{(k)}) + \epsilon\alpha_k \nabla f(x^{(k)})^{T}(\bar{x}^{(k)} - x^{(k)}). \tag{3.16}$$

Essa desigualdade é válida, desde que (3.15) possa ser sempre satisfeita dentro de um número finito de etapas.

Usando a Fórmula de Taylor com resto integral (vide [7], página 262), a desigualdade de Cauchy-Schwarz e a condição (3.14), temos que

$$f(x^{(k)} + \alpha(\bar{x}^{(k)} - x^{(k)})) - f(x^{(k)}) = \int_0^{\alpha} \nabla f(x^{(k)} + s(\bar{x}^{(k)} - x^{(k)}))^{T}(\bar{x}^{(k)} - x^{(k)}))ds$$

ou seja,

$$
\begin{aligned}
&= \int_0^\alpha \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})ds \\
&+ \int_0^\alpha \left[\nabla f(x^{(k)} + s(\bar{x}^{(k)} - x^{(k)})) - \nabla f(x^{(k)})\right]^T (\bar{x}^{(k)} - x^{(k)})ds \\
&\leqslant \alpha \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) \\
&+ \int_0^\alpha \|\nabla f(x^{(k)} + s(\bar{x}^{(k)} - x^{(k)})) - \nabla f(x^{(k)})\| \|(\bar{x}^{(k)} - x^{(k)})\|ds \\
&\leqslant \alpha \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) \\
&+ \int_0^\alpha L\|s(\bar{x}^{(k)} - x^{(k)})\| \cdot \|(\bar{x}^{(k)} - x^{(k)})\|ds \\
&= \alpha \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) + \int_0^\alpha sL\|\bar{x}^{(k)} - x^{(k)}\|^2 ds \\
&= \alpha \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) + \tfrac{1}{2}\alpha^2 L\|\bar{x}^{(k)} - x^{(k)}\|^2.
\end{aligned}
\tag{3.17}
$$

escolhendo

$$\alpha \leqslant \frac{2(\epsilon - 1)\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})}{L\|\bar{x}^{(k)} - x^{(k)}\|^2},$$

obtemos de (3.17)

$$
\begin{aligned}
f(x^{(k+1)}) - f(x^{(k)}) &\leqslant \left[\frac{2(\epsilon - 1)\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})}{L\|\bar{x}^{(k)} - x^{(k)}\|^2}\right] \cdot \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) \\
&+ \frac{1}{2}\left[\frac{2(\epsilon - 1)\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})}{L\|\bar{x}^{(k)} - x^{(k)}\|^2}\right]^2 \cdot L\|\bar{x}^{(k)} - x^{(k)}\|^2 \\
&= 2(\epsilon - 1)\frac{\left[\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})\right]^2}{L\|\bar{x}^{(k)} - x^{(k)}\|^2} \\
&+ 2(\epsilon - 1)^2\frac{\left[\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})\right]^2}{L\|\bar{x}^{(k)} - x^{(k)}\|^2} \\
&= \left[2(\epsilon - 1) + (2\epsilon^2 - 4\epsilon + 2)\right]\frac{\left[\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})\right]^2}{L\|\bar{x}^{(k)} - x^{(k)}\|^2} \\
&= (2\epsilon^2 - 2\epsilon)\frac{\left[\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})\right]^2}{L\|\bar{x}^{(k)} - x^{(k)}\|^2} \\
&= \frac{2(\epsilon - 1)\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})}{L\|\bar{x}^{(k)} - x^{(k)}\|^2} \cdot \epsilon \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}).
\end{aligned}
$$

Ou seja, $f(x^{(k+1)}) \leqslant f(x^{(k)}) + \alpha\epsilon\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})$.

Substituindo α por β^{-i}, vemos que $\bar{i}$ é menor inteiro que satisfazendo

$$\beta^{-i} \leqslant \frac{2(\epsilon - 1)\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})}{L\|\bar{x}^{(k)} - x^{(k)}\|^2}.$$

Como $\beta^{-i} < \beta^{-(i-1)}$, consequentemente

$$\beta\alpha_k = \beta^{-(i-1)} > \frac{2(\epsilon - 1)\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})}{L\|\bar{x}^{(k)} - x^{(k)}\|^2},$$

i.e.,

$$\alpha_k > \left[\frac{2(\epsilon-1)}{\beta}\right] \frac{\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})}{L\|\bar{x}^{(k)} - x^{(k)}\|^2}. \tag{3.18}$$

Assim, (3.15) será satisfeita após um número finito de passos com o comprimento satisfazendo (3.18).

Note que $\|\bar{x}^{(k)} - x^{(k)}\|$ é limitado superiormente por uma constante, digamos $C > 0$, pois X é compacto. De (3.18) temos

$$\alpha_k > \left[\frac{2(\epsilon-1)}{\beta}\right] \frac{\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})}{LC^2}. \tag{3.19}$$

Por (3.16) obtemos

$$f(x^{(k+1)}) - f(x^{(k)}) \leqslant \left[\frac{2\epsilon(\epsilon-1)}{\beta}\right] \frac{\left[\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})\right]^2}{LC^2}. \tag{3.20}$$

Somando (3.20) para todo $k \leqslant m-1$, obtemos

$$f(x^{(m)}) - f(x^{(0)}) \leqslant \left[\frac{2\epsilon(\epsilon-1)}{\beta LC^2}\right] \sum_{k=1}^{m-1} \left[\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})\right]^2.$$

Usando o fato de que $f(x)$ é limitada inferiormente em X, por digamos $\bar{f}$, temos

$$\begin{aligned}\sum_{k=1}^{m-1} \left[\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})\right]^2 &\leqslant \left[\frac{\beta LC^2}{2\epsilon(1-\epsilon)}\right](f(x^{(0)}) - f(x^{(m)})) \\ &\leqslant \left[\frac{\beta LC^2}{2\epsilon(1-\epsilon)}\right](f(x^{(0)}) - \bar{f}).\end{aligned}$$

Segue que a série $\displaystyle\sum_{k=1}^{\infty} \left[\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)})\right]^2$ é convergente, e consequentemente

$$\lim_{l\to\infty} \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) = 0. \tag{3.21}$$

Usando os mesmos argumentos como no Teorema 3.1.2, podemos escolher subsequências de $\{x^{(k)}\}$ e $\{\bar{x}^{(k)}\}$ tal que

$$\lim_{k\in K} x^{(k)} = x^*, \quad \lim_{k\in \hat{K}} \bar{x}^{(k)} = \bar{x}.$$

Por (3.21), $\nabla f(x^*)^T(\bar{x} - x^*) = 0$, completando a prova. □

3.2 Interpretações e casos particulares

Ao escolher a função $g(x, y)$ de várias maneiras, conseguimos reconhecer métodos de direções viáveis bem conhecidos. A seguir listamos alguns casos particulares.

Caso particular 3.2.1. *O método de Frank-Wolfe*

O primeiro, trivialmente basta tomar $g(x,y) = 0$, assim o subproblema $(\mathsf{SUB}-\mathsf{P})^{(k)}$ é equivalente ao proposto no método de Frank-Wolfe.

Caso particular 3.2.2. *O método de Newton e quase-Newton*

Note que tomando a função $g(x,y) = \frac{1}{2}x^T\nabla^2 f(y)x$, onde f é também duas vezes diferenciável, conseguimos ter o método de Newton como caso particular. De fato,

$$\begin{aligned}
f^{(k)}(x) &= \tfrac{1}{2}x^T\nabla^2 f(x^{(k)})x + f(x^{(k)}) - \tfrac{1}{2}{x^{(k)}}^T\nabla^2 f(x^{(k)})x^{(k)} \\
&+ \left[\nabla f(x^{(k)}) - \nabla^2 f(x^{(k)})x^{(k)}\right]^T(x - x^{(k)}) \\
&= \tfrac{1}{2}x^T\nabla^2 f(x^{(k)})x + f(x^{(k)}) + \nabla f(x^{(k)})^T(x - x^{(k)}) + \tfrac{1}{2}{x^{(k)}}^T\nabla^2 f(x^{(k)})x^{(k)} \\
&- x^T\nabla^2 f(x^{(k)})x^{(k)} \\
&= \tfrac{1}{2}x^T\nabla^2 f(x^{(k)})(x - x^{(k)}) + f(x^{(k)}) + \nabla f(x^{(k)})^T(x - x^{(k)}) \\
&+ \tfrac{1}{2}{x^{(k)}}^T\nabla^2 f(x^{(k)})x^{(k)} - \tfrac{1}{2}x^T\nabla^2 f(x^{(k)})x^{(k)} \\
&= \tfrac{1}{2}x^T\nabla^2 f(x^{(k)})(x - x^{(k)}) + f(x^{(k)}) + \nabla f(x^{(k)})^T(x - x^{(k)}) \\
&- \tfrac{1}{2}(x - x^{(k)})^T\nabla^2 f(x^{(k)})x^{(k)} \\
&= f(x^{(k)}) + \nabla f(x^{(k)})^T(x - x^{(k)}) + \tfrac{1}{2}(x - x^{(k)})^T\nabla^2 f(x^{(k)})(x - x^{(k)}).
\end{aligned}$$

Ou seja,

$$f^{(k)}(x) = f(x^{(k)}) + \nabla f(x^{(k)})^T(x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T\nabla^2 f(x^{(k)})(x - x^{(k)}).$$

Note que usamos a simetria nas operações acima e que quando $f(x)$ é apenas pseudo-convexa, a hessiana $\nabla^2 f(x)$ não é necessariamente semidefinida positiva. Entretando, o método de Newton com restrições pode sempre ser modificado adicionando uma matriz definida positiva a matriz hessiana, i.e.,

$$g(x,y) = \frac{1}{2}x^T(\nabla^2 f(y) + \epsilon I)x,$$

onde $\epsilon > 0$ é suficientemente grande tal que a matriz $\nabla^2 f(y) + \epsilon I$ seja pelo menos semidefinida positiva.

No caso do método Quase-Newton, basta tomar $g(x,y) = \frac{1}{2}x^T B(y)x$ e verifica-se do mesmo modo feito no caso anterior.

Caso particular 3.2.3. *O método do ponto proximal*

Agora considere a função

$$g(x,y) := f(x) + \left(\frac{c}{2}\right)\|x-y\|^2,$$

onde $c > 0$. Note que reescrevendo o $(SUB-P)^{(k)}$ temos agora que

$$\begin{aligned} f^{(k)}(x) &= f(x) + \left(\frac{c}{2}\right)\|x-x^{(k)}\|^2 + f(x^{(k)}) - \left[f(x^{(k)}) + \left(\frac{c}{2}\right)\|x^{(k)}-x^{(k)}\|^2\right] \\ &+ \left[\nabla f(x^{(k)}) - (\nabla f(x^{(k)}) + c(x^{(k)}-x^{(k)}))\right]^T (x-x^{(k)}), \end{aligned}$$

ou seja,

$$f^{(k)}(x) = f(x) + \frac{c}{2}\|x-x^{(k)}\|^2.$$

Com a escolha de g dada acima garantimos que o método do ponto proximal dado na seção 2.5 também é um caso particular do algoritmo de linearização parcial.

Caso particular 3.2.4. *O método de Jacobi*

Assim como no método de Newton, apresentamos outro método para encontrar solução de sistemas de equações. O método de Jacobi é bem conhecido pela solução de sistemas de equações lineares e não-lineares e problemas de otimização sobre conjuntos de produtos cartesianos. Suponha que o conjunto viável $X \subset R^n$ tal que $X = \prod_{i\in C} X_i$, onde $X_i \in R^{n_i}$ e $\sum_{i\in C} n_i = n$. Dado um ponto $x^{(k)} \in X$, o subproblema de Jacobi consiste então no seguinte problema independente

$$(P_i^{(k)}) \qquad x_i^{(k+1)} \in \arg\min_{x_i\in X_i} f(x_1^{(k)}, x_2^{(k)}, \ldots, x_{i-1}^{(k)}, x_i, x_{i+1}^{(k)}, \ldots, x_C^{(k)}) \tag{3.22}$$

Note que tomando $g(x,y) = \sum_{i\in C} f(y_1^{(k)}, y_2^{(k)}, \ldots, y_{i-1}^{(k)}, x_i, y_{i+1}^{(k)}, \ldots, y_C^{(k)})$ no $(SUB-P)^{(k)}$, temos que o problema em questão é um caso particular do algoritmo de linearização parcial. De fato,

$$\begin{aligned} f^{(k)}(x) &= g(x,x^{(k)}) + f(x^{(k)}) - g(x^{(k)},x^{(k)}) + \left[\nabla f(x^{(k)}) - \nabla_x g(x^{(k)},x^{(k)})\right]^T (x-x^{(k)}) \\ &= \sum_{i\in C} f(x_1^{(k)}, x_2^{(k)}, \ldots, x_{i-1}^{(k)}, x_i, x_{i+1}^{(k)}, \ldots, x_C^{(k)}) + f(x^{(k)}) \\ &- \sum_{i\in C} f(x_1^{(k)}, x_2^{(k)}, \ldots, x_{i-1}^{(k)}, x_i^{(k)}, x_{i+1}^{(k)}, \ldots, x_C^{(k)}) \\ &+ \left[\nabla f(x^{(k)}) - (\frac{\partial f}{\partial x_1}(x^{(k)}), \frac{\partial f}{\partial x_2}(x^{(k)}), \ldots, \frac{\partial f}{\partial x_C}(x^{(k)}))\right]^T (x-x^{(k)}) \\ &= \sum_{i\in C} f(x_1^{(k)}, x_2^{(k)}, \ldots, x_{i-1}^{(k)}, x_i, x_{i+1}^{(k)}, \ldots, x_C^{(k)}) + f(x^{(k)}) \\ &- \mid C \mid f(x_1^{(k)}, x_2^{(k)}, \ldots, x_{i-1}^{(k)}, x_i^{(k)}, x_{i+1}^{(k)}, \ldots, x_C^{(k)}). \end{aligned}$$

Ou seja, a menos de uma constante o subproblema $(SUB-P)^{(k)}$ é equivalente a (3.22).

Capítulo 4

Extensões para o caso não diferenciável

4.1 O algoritmo de linearização parcial (caso não diferenciável)

Nessa seção, vamos estender o algoritmo de linearização parcial para o caso não diferenciável. Seja a função objetivo $f(x)$ dada por

$$f(x) = p(x) + q(x), \tag{4.1}$$

onde $p : \mathbb{R}^n \mapsto (-\infty, +\infty]$ é uma fechada e apropriada função convexa e $q : \mathbb{R}^n \mapsto (-\infty, +\infty]$ é continuamente diferenciável em X. O problema de minimizar f sobre o não vazio, compacto e convexo conjunto X será referido como (NP). Na iteração k, a função objetivo é reescrita como

$$f(x) = (p(x) + g(x, x^{(k)})) + (q(x) - g(x, x^{(k)})),$$

portanto, o subproblema definido pela linearização do segundo termo é o seguinte,

$$(SUB - NP)^{(k)} \begin{array}{ll} \min & f^{(k)}(x), \\ s.a & x \in X \end{array} \tag{4.2}$$

$$f^{(k)}(x) = p(x)+g(x, x^{(k)})+q(x^{(k)})-g(x^{(k)}, x^{(k)})+\left[\nabla q(x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T (x-x^{(k)}),$$

O resultado é dado em três teoremas. O primeiro, mostramos que as soluções do subproblema geram direções de descida. Então, mostramos a convergência de duas versões do algoritmo, onde a busca linear é feita exatamente e aplicando uma modificação da regra de Armijo para o comprimento do passo, respectivamente. O primeiro teorema é uma generalização do Teorema 3.1.1.

Teorema 4.1.1. *([9]) Assuma que $x^{(k)}$ é um ponto viável de (NP) e que $\bar{x}^{(k)}$ resolve $(SUB-NP)^{(k)}$. Se $x^{(k)}$ resolve $(SUB-NP)^{(k)}$, então $x^{(k)}$ é um ponto crítico do problema (NP). De outra forma, a direção $d^{(k)} := \bar{x}^{(k)} - x^{(k)}$ é uma direção de descida viável com respeito a f.*

Demonstração. Mostraremos que, para qualquer $x^{(k)}$, a derivada direcional de f na direção $d^{(k)}$ é não positiva, e é estritamente negativa se $x^{(k)}$ não é um ponto crítico, i.e., que

$$f^{(k)}(\bar{x}^{(k)}) < f^{(k)}(x^{(k)}) \Rightarrow \nabla f(x^{(k)}; d^{(k)}) < 0,$$

$$\begin{aligned}
\nabla f(x^{(k)}; d^{(k)}) &:= \lim_{\lambda \to o_+} \frac{1}{\lambda}\left[f(x^{(k)} + \lambda d^{(k)}) - f(x^{(k)})\right] \\
&= \sup_{u \in \partial p(x^{(k)})} \left[\nabla q(x^{(k)}) + u\right]^T d^{(k)} \\
&= \nabla q(x^{(k)})^T d^{(k)} + \sup_{u \in \partial p(x^{(k)})} u^T d^{(k)},
\end{aligned}$$

na qual a segunda igualdade segue do Teorema 1.3.9. Se $x^{(k)}$ não resolve o subproblema $(SUB - NP)^{(k)}$, então usando a convexidade da $g(x, y)$ com respeito a x, ou seja, que $0 \leqslant g(\bar{x}, x^{(k)}) - g(x^{(k)}, x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})^T(\bar{x} - x^{(k)})$ temos,

$$\begin{aligned}
& f^{(k)}(\bar{x}^{(k)}) && < \quad f^{(k)}(x^k) \\
& \Leftrightarrow p(\bar{x}^{(k)}) + g(\bar{x}^{(k)}, x^{(k)}) + \left[\nabla q(x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T d^{(k)} && < \quad p(x^{(k)}) + g(x^{(k)}, x^{(k)}).
\end{aligned}$$

Equivalentemente,

$$\begin{aligned}
& \Leftrightarrow p(\bar{x}^{(k)}) + g(\bar{x}^{(k)}, x^{(k)}) + \nabla q(x^{(k)})^T d^{(k)} - \nabla_x g(x^{(k)}, x^{(k)})^T d^{(k)} < p(x^{(k)}) + g(x^{(k)}, x^{(k)}) \\
& \Leftrightarrow p(\bar{x}^{(k)}) + g(\bar{x}^{(k)}, x^{(k)}) - g(x^{(k)}, x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})^T d^{(k)} + \nabla q(x^{(k)})^T d^{(k)} < p(x^{(k)}) \\
& \Leftrightarrow p(\bar{x}^{(k)}) + \nabla q(x^{(k)})^T d^{(k)} < p(x^{(k)})
\end{aligned}$$

i.e., $\nabla q(x^{(k)})^T d^{(k)} < p(x^{(k)}) - p(\bar{x}^{(k)})$. Usando esse fato e a definição de subgradiente em p temos,

$$f'(x^{(k)}; d^{(k)}) < p(x^{(k)}) + \sup_{u \in \partial p(x^{(k)})} u^T d^{(k)} - p(\bar{x}^{(k)}) \leqslant 0.$$

Portanto, a direção $d^{(k)}$ é uma direção de descida viável com respeito a f. Contudo, se $x^{(k)}$ resolve o subproblema, temos então a desigualdade variacional

$$\left[u^{(k)} + \nabla q(x^{(k)})\right]^T (x - x^{(k)}) \geqslant 0, \ \forall\, x \in X,$$

para algum $u^{(k)} \in \partial p(x^{(k)})$. Consequentemente, $x^{(k)}$ é um ponto crítico do problema (NP), e no caso de q ser uma função convexa, isso é uma condição suficiente para a otimalidade global do ponto $x^{(k)}$ no problema (NP). (veja Teorema 1.3.10) □

Teorema 4.1.2. *([9]) Seja* p *contínua em* x*. O algoritmo de linearização parcial aplicado ao problema* (NP) *ou termina em um número finito de interações ou gera uma sequência infinita* $\{x^{(k)}\}_{k\geqslant 0}$*, tal que qualquer ponto de acumulação é um ponto crítico no problema* (NP)*. Em particular, se a função* $q(x)$ *é convexa, então o algoritmo é globalmente convergente.*

Demonstração. A prova é uma modificação dada no Teorema 3.1.2. De acordo com o Teorema 4.1.1, o algoritmo termina em um número finito de interações se o subproblema na interação k é resolvido por $x^{(k)}$. Assumimos sendo assim que

$$f^{(k)}(\bar{x}^{(k)}) < f^{(k)}(x^{(k)}),$$

para toda iteração k.

Pelos mesmos argumentos na prova do Teorema 3.1.2, o algoritmo gera direções de descida, e podemos escolher uma subsequência $\hat{K}$ tal que $x^{(k)}$ e $\bar{x}^{(k)}$ tem pontos limites x^* e $\bar{x}$, respectivamente. Segue que

$$f(x^{(k)} + \alpha_k(\bar{x}^{(k)} - x^{(k)})) - f(x^{(k)}) \geqslant f(x^{(k+1)}) - f(x^{(k)}),$$

para todo $\alpha_k \in [0, \bar{\alpha}_k]$. Como $f(x^{(k+1)}) - f(x^{(k)})$ tende para zero em $\hat{K}$, temos

$$\lim_{k\in\hat{K}} f(x^{(k)} + \alpha_k(\bar{x}^{(k)} - x^{(k)})) - f(x^{(k)}) \geqslant 0. \tag{4.3}$$

Escolhendo a sequência de α_k tal que α_k tenda para zero em $\hat{K}$. A subderivada superior de f em x^* na direção $d^* = \bar{x} - x^*$ é definida como (veja [11], página 32)

$$f^{\uparrow}(x^*; d^*) = \lim_{\substack{x^{(k)}\to_f x^* \\ k\in\hat{K}}} \sup \inf_{d^{(k)}\to d^*} \frac{f(x^{(k)} + \alpha_k(\bar{x}^{(k)} - x^{(k)})) - f(x^{(k)})}{\alpha_k}.$$

A função convexa p é contínua no compacto X. Consequentemente, pela Proposição 3G (veja em [11], página 34), a subderivada superior de p em x^* com respeito a d^* satisfaz a relação

$$g^{\uparrow}(x^*; \bar{x} - x^*) = g'(x^*; \bar{x} - x^*).$$

Agora, a função q é continuamente diferenciável, e portanto o limite

$$\frac{q(x^{(k)} + \alpha_k(\bar{x}^{(k)} - x^{(k)})) - q(x^{(k)}}{\alpha_k}$$

existe e é igual a $\nabla q(x^*)^T(\bar{x} - x^*)$ (veja [11], Teorema 4F). Pelo Teorema 4.4a, [12],

$$\begin{aligned} f^{\uparrow}(x^*; d^*) &= p^{\uparrow}(x^*; (\bar{x} - x^*)) + q^{\uparrow}(x^*; (\bar{x} - x^*)) \\ &= p'(x^*; (\bar{x} - x^*)) + \nabla q(x^*)^T(\bar{x} - x^*) \end{aligned} \tag{4.4}$$

da qual é não negativo por (4.3).

Pelo Teorema 1.3.9 segue que

$$p'(x^*; (\bar{x} - x^*)) = \sup_{u \in \partial p(x^*)} u^T(\bar{x} - x^*),$$

e o supremo é atingido em, digamos, $u^* \in \partial p(x^*)$, desde que $\partial p(x^*)$ seja compacto. Portanto, por (4.4), existe um subgradiente u^* de p no ponto x^* tal que

$$[u^* + \nabla q(x^*)]^T (\bar{x} - x^*) \geqslant 0. \tag{4.5}$$

Consequentemente, no limite o subproblema não gera uma direção de descida.

Na iteração k, o subproblema é equivalente ao seguinte problema de desigualdade variacional (vide Teorema 1.3.10): existe um subgradiente $\bar{u}^{(k)}$ de p em $\bar{x}^{(k)}$ tal que, para todo $x \in X$,

$$\left[\bar{u}^{(k)} + \nabla_x g(\bar{x}^{(k)}, x^{(k)}) + \nabla q(x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T (x - \bar{x}^{(k)}) \geqslant 0. \tag{4.6}$$

Note que, se $\bar{u}^{(k)} \in \partial p(\bar{x}^{(k)})$, então pela compacidade de $\partial p(\bar{x}^{(k)})$ para k, uma subsequência de $\hat{K}$ dos subgradientes $\bar{u}^{(k)}$ converge para um ponto $\bar{u}$, da qual satisfaz $\bar{u} \in \partial p(\bar{x})$. Daí, tomando o limite sobre essa subsequência, (4.6) gera

$$[\bar{u} + \nabla_x g(\bar{x}, x^*) + \nabla q(x^*) - \nabla_x g(x^*, x^*)]^T (x - \bar{x}) \geqslant 0, \ \forall \ x \in X, \tag{4.7}$$

na qual implica que $\bar{x}$ resolve o subproblema do algoritmo de linearização parcial definido no ponto x^*.

Note que pela monotonicidade de $\nabla_x g(x, y)$ com respeito a x e substituindo $x = x^*$, i.e.,

$$0 \geqslant [\nabla_x g(\bar{x}, x^*) - \nabla_x g(x^*, x^*)]^T (x^* - \bar{x}) \geqslant [\bar{u} + \nabla q(x^*)]^T (\bar{x} - x^*)$$

temos que,

$$[\bar{u} + \nabla q(x^*)]^T (x^* - \bar{x}) \geqslant 0. \tag{4.8}$$

Somando (4.5) e (4.8), obtemos

$$[\bar{u} + \nabla q(x^*) - u^* - \nabla q(x^*)]^T (x^* - \bar{x}) \geqslant 0, \text{ ou seja, } [\bar{u} - u^*]^T (x^* - \bar{x}) \geqslant 0.$$

Por outro lado, pela definição de convexidade da p nos pontos $\bar{x}$ e x^*, respectivamente, obtemos a desigualdade invertida. Assim,

$$[\bar{u} - u^*]^T (x^* - \bar{x}) = 0,$$

e $\bar{u} = u^*$, substituindo em 4.8 temos que

$$[u^* + \nabla q(x^*)]^T (x^* - \bar{x}) \geqslant 0,$$

juntamente com 4.5 implica

$$[u^* + \nabla q(x^*)]^T (x^* - \bar{x}) = 0.$$

Assim, usando os mesmos argumentos na prova do Teorema 3.1.2, concluímos que

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

daí por (4.7) existe um subgradiente u^* de p em x^* tal que

$$[u^* + \nabla q(x^*)]^T (x - x^*) \geqslant 0, \ \forall \ x \in X.$$

Note que, se q é convexa, então a desigualdade variacional acima é, pela convexidade de f, tanto necessária quanto suficiente para a otimalidade global de x^* no problema (NP) (vide Teorema 1.3.10). Completando assim a prova. □

Se p é a função caracteristica de um conjunto convexo e fechado $S \subset \mathbb{R}^n$, segue da prova acima que o algoritmo é globalmente convergente também quando q é apenas pseudoconvexa.

O método de linearização parcial pode ser estendido em uma busca linear não exata também no caso da não diferenciabilidade, introduzindo uma condição de Lipschitz na parte diferenciável da função objetivo e por restringir $g(x, y)$ a ser estritamente convexa com respeito a x. A regra do comprimento do passo segundo Armijo é modificada como segue. O ponto $x^{(k+1)}$ é obtido encontrando o menor inteiro $\bar{i}$ tal que

$$f(x^{(k)} + \beta^{-i} d^{(k)}) - f(x^{(k)}) \leqslant -\epsilon\beta^{-i} \left[\nabla_x g(\bar{x}^{(k)}, x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T d^{(k)}, \quad (4.9)$$

onde $d^{(k)} = \bar{x}^{(k)} - x^{(k)}$.

Teorema 4.1.3. *([9]) Sob as hipóteses adicionais de que* $q(x)$ *tem um gradiente Lipschitz e* $g(x, y)$ *é estritamente convexa com respeito a* x*, o algoritmo de linearização parcial, aplicado ao problema* (NP) *usando a formula (4.9) do comprimento do passo modificado de Armijo, é convergente no mesmo sentido do Teorema 4.1.2.*

Demonstração. A condição de Lipschitz assumida em ∇q implica por (3.17) que, para todo $\alpha \geqslant 0$,

$$q(x^{(k)} + \alpha d^{(k)}) - q(x^{(k)}) \leqslant \alpha \nabla q(x^{(k)})^T d^{(k)} + \frac{1}{2}\alpha^2 L \|d^{(k)}\|^2, \tag{4.10}$$

onde L denota a constante de Lipschitz e $d^{(k)} = \bar{x}^{(k)} - x^{(k)}$. Além disso, para todo $\alpha \in [0, 1]$,

$$\begin{aligned} p(x^{(k)} + \alpha d^{(k)}) - p(x^{(k)}) &= \textstyle\int_0^\alpha p'(x^{(k)} + s d^{(k)}; d^{(k)}) ds \\ &\leqslant \textstyle\int_0^\alpha p'(x^{(k)} + d^{(k)}; d^{(k)}) ds \\ &\leqslant -\alpha p'(x^{(k)} + d^{(k)}; -d^{(k)}) \\ &= \alpha \inf_{u \in \partial p(\bar{x}^{(k)})} u^T d^{(k)} \\ &\leqslant \alpha \bar{u}^{(k)^T} d^{(k)} \end{aligned} \tag{4.11}$$

onde a primeira igualdade segue do corolário 24.2.1 de [10] (Página 231) e a primeira desigualdade do Teorema 24.1 de [10] (Página 227). Note que o subgradiente $\bar{u}^{(k)}$ é arbitrário em $\partial p(\bar{x}^{(k)})$, mas será escolhido um em particular.

Da relação

$$\bar{x}^{(k)} \in \min_{x \in X} f^{(k)}(x),$$

temos por (4.6) que existe um subgradiente $\bar{u}^{(k)}$ de p no ponto $\bar{x}^{(k)}$ tal que

$$\left[\bar{u}^{(k)} + \nabla_x g(\bar{x}^{(k)}, x^{(k)}) + \nabla q(x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T d^{(k)} \leqslant 0. \tag{4.12}$$

Note que daí, $-\alpha\left[\nabla_x g(\bar{x}^{(k)}, x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T d^{(k)} \geqslant \alpha\left[\bar{u}^{(k)} + \nabla q(x^{(k)})\right]^T d^{(k)}$ isso implica, usando (4.10) e (4.11), que

$$\begin{aligned} f(x^{(k)} + \alpha d^{(k)}) - f(x^{(k)}) &\leqslant \alpha\left[\bar{u}^{(k)} + \nabla q(x^{(k)})\right]^T d^{(k)} + \tfrac{1}{2}\alpha^2 L \|d^{(k)}\|^2 \\ &\leqslant -\alpha\left[\nabla_x g(\bar{x}^{(k)}, x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T d^{(k)} + \frac{1}{2}\alpha^2 L \|d^{(k)}\|^2. \end{aligned} \tag{4.13}$$

Seguindo a mesma linha de raciocínio do Teorema 3.1.3, apenas substituindo $\nabla f(x^{(k)})^T d^{(k)}$ por $-\left[\nabla_x g(\bar{x}^{(k)}, x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T d^{(k)}$, podemos concluir que a regra modificada de Armijo para o comprimento do passo (4.9) é finita e que a série

$$\sum_{k=1}^{\infty} (\left[\nabla_x g(\bar{x}^{(k)}, x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T d^{(k)})^2$$

é convergente. Consequentemente,

$$\lim_{k\to\infty} \left[\nabla_x g(\bar{x}^{(k)}, x^{(k)}) - \nabla_x g(x^{(k)}, x^{(k)})\right]^T d^{(k)} = 0,$$

e podemos escolher subsequências de $\{x^{(k)}\}$ e $\{\bar{x}^{(k)}\}$ tal que

$$\lim_{k\in k} x^{(k)} = x^* \quad \text{e} \quad \lim_{k\in\hat{K}} \bar{x}^{(k)} = \bar{x},$$

Assim obtemos que

$$[\nabla_x g(\bar{x}, x^*) - \nabla_x g(x^*, x^*)]^T (\bar{x} - x^*) = 0.$$

Pela convexidade estrita de $g(x, y)$ com respeito x, isso implica que

$$x^* = \bar{x}.$$

Usando que no limite em (4.12), existe um subgradiente u^* de $\partial p(x^*)$ tal que

$$[u^* + \nabla q(x^*)]^T (x - x^*) \geqslant 0, \ \forall \ x \in X,$$

na qual completa a prova.

$\square$

Capítulo 5

Considerações Finais

Nesta dissertação estudamos a boa definição do método de linearização parcial, assim como a convergência global do algoritmo de linearização parcial, apartir da hipótese adicional do gradiente da função objetivo ser lipschitz e com uma busca linear predeterminada. Dessa forma, unificamos alguns métodos de descida clássicos existentes na literatura tais como Newton, quase-Newton, ponto proximal, Frank-wolfe e o método de Jacobi. Além disso, estendemos o algoritmo para o caso não diferenciável onde, assim como no caso diferenciável, a convergência é garantida assumindo que o gradiente da parte diferenciável da função objetivo seja lipschitziano, acrescentando a convexidade estrita sobre a função $g(x, y)$ com respeito a x e usando de uma busca linear modificada.

Referências Bibliográficas

[1] Bazaraa, M. S., and Shetty, C. M., Nonlinear Programming: Theory and Algorithms. John Wiley and Sons, New York.(1979).

[2] Frank, M., and Philip W. An algorithm for quadratic programming. Naval research logistics quarterly 3.1-2 (1956): 95-110.

[3] Iusem, A, N., Métodos de Ponto Proximal em Otimização, 20º Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro, (1995).

[4] Izmailov, A., Solodov, M. Otimização, Volume 1, Rio de Janeiro: IMPA, (2014).

[5] Izmailov, A., Solodov, M. Otimização, volume 2, Rio de Janeiro, IMPA, (2012).

[6] Lima, E. L., Análise Real, Volume 2, Funções de n várias variáveis. 11ªed. Rio de Janeiro: IMPA, (2012).

[7] Lima, E. L., Curso de Análise, volume 2. Projeto Euclides. IMPA, Rio de Janeiro, Brasil, (2012).

[8] Patriksson, M. A unified description of iterative algorithms for traffic equilibria. European Journal of Operational Research 71.2 (1993): 154-176.

[9] Patriksson, M. Partial linearization methods in nonlinear programming. Journal of Optimization Theory and Applications 78.2 (1993): 227-246.

[10] Rockafellar, R. T., Convex Analysis. Princeton Univ. Press (1970).

[11] Rockafellar, R. T.,The Theory of Subgradients and Its Applications to Problems of Optimization: Convex and Nonconvex Functions, Heldermann-Verlag, Berlin, Germany, (1981).

[12] Rudin, W.,Principles of Mathematical Analysis, 3rd Edition, McGraw-Hill, Singapore, Republic of Singapore, (1985).