Otimização convexa

Aluno: Bruno Henrique Cervelin
Orientadora: Profa. Dra. Maria Aparecida Diniz Ehrhardt

DMA – IMECC – UNICAMP

Resumo

Recentemente otimização convexa tem se tornado uma ferramenta muito importante para a resolução de problemas de engenharia, graças à sua eficácia e confiabilidade. Neste relatório apresentamos resultados estudados sobre análise convexa e dualidade, formando uma base teórica sobre o assunto, e também apresentamos resultados sobre alguns métodos de pontos interiores, a fim de obtermos algumas formas de resolução dos problemas estudados.

1 Introdução

Funções convexas aparecem constantementes em problemas reais, tais funções possuem características muito importantes como por exemplo, todo o minimizador de funções convexas em um conjunto convexo é um minimizador global neste conjunto, no Capítulo 3 são apresentados resultados sobre funções convexas, conjuntos convexos e algumas condições necessárias (suficientes) para a otimalidade destas funções.

Frequentemente associamos um problema de otimização convexa com outro problema, o problema dual, que pode ser equivalente ao problema original, e também mais fácil de se resolver. No capítulo 4 apresentamos resultados sobre dualidade em problemas lineares (4.1) e também em problemas gerais (4.2).

A resolução de problemas convexos, muitas vezes, mesmo que irrestrito, se dá através de métodos iterativos. Geralmente estes são métodos de descida. No capítulo 5, tomando como referência principalmente [3] e [4], mostramos um método de descida muito eficiente, o Método das Barreiras.

2 Definições

Neste capítulo apresentaremos separadamente as definições que nos serão úteis ao longo deste trabalho.

Definição **1** *Dados $x^i \in I\!R^n$, $\alpha_i \in [0,1]$, $i = 1\cdots,p$, tais que $\sum_{i=1}^{p}\alpha_i = 1$, o ponto $\sum_{i=1}^{p}\alpha_i x^i$ chama-se combinação convexa dos pontos x^i com parâmetros α_i.*

Definição **2** *Um conjunto $D \subset I\!R^n$ é chamado conjunto convexo se para quaisquer $x \in D$, $y \in D$ e $\alpha \in [0,1]$, $\alpha x + (1-\alpha)y \in D$.*

Definição **3** *Um conjunto $K \subset I\!R^n$ chama-se cone quando $\forall d \in K$ e $\forall t \in I\!R_+$, $td \in K$.*

Definição **4** *O cone dual de um cone $K \subset I\!R^n$ é definido por*

$$K^* = \{y \in I\!R^n | y^T d \leq 0, \forall d \in K\}.$$

Definição **5** *Dado um conjunto $D \subset I\!R^n$ qualquer, o fecho cônico de D, denotado por cone D, é o menor cone convexo em $I\!R^n$ que contém D.*

Definição **6** *Dizemos que $d \in I\!R^n$ é uma direção tangente em relação ao conjunto D no ponto $x^* \in D$ quando $dist(x^* + td, D) = o(t)^1$, $t \in I\!R_+$.*

O conjunto de todas as direções tangentes em relação ao conjunto D que passam pelo ponto x, denotado por $\tau_D(x)$, é chamado cone tangente.

Definição **7** *Dizemos que $V_D(x)$, é um conjunto de direções viáveis[2], no ponto $x \in D$, em relação ao conjunto D se*

$$V_D(x) = \{d \in I\!R^n | \exists \alpha \in I\!R_+ \, t.q. \, x + \alpha d \in D\}.$$

Definição **8** *Se D é um conjunto convexo, f é função convexa em D se $\forall x \in D, y \in D$ e $\alpha \in [0,1]$,*

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

Se a desigualdade for estrita, f é uma função estritamente convexa. Se tivermos

[1] $f(t) = o(t)$ implica que a função $f : I\!R_+ \to I\!R$ possui a propriedade de que $lim_{t \to 0_+} f(t)/t = 0$

[2] a expressão factível será utilizada neste texto com o mesmo significado de viável

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

então a função é côncava.

Definição **9** *Seja $D \subset \mathbb{R}^n$ um conjunto qualquer. O fecho convexo de D, denotado $\operatorname{conv} D$, é o menor conjunto convexo em $\mathbb{R}^n$ que contém D.*

Definição **10** *Dizemos que $d \in \mathbb{R}^n$ é uma direção de recessão do conjunto convexo D quando $\forall t \in \mathbb{R}_+$ e $\forall x \in D$, $x + td \in D$.*

Definição **11** *Dizemos que o hiperplano $H(a,c) = \{x \in \mathbb{R}^n | a^T x = c\}$, $a \neq 0$, separa os conjuntos D_1 e D_2 se*

$$a^T x^1 \leq c \leq a^T x^2$$
$$\forall x^1 \in D_1,$$
$$\forall x^2 \in D_2.$$

Se tivermos desigualdade estrita dizemos que o hiperplano separa estritamente D_1 e D_2.

Definição **12** *O ponto x é ponto extremo do conjunto convexo D quando não pode ser representado por uma conbinação convexa de outros pontos de D. O conjunto de todos os pontos extremos é denotado por $E(D)$.*

Definição **13** *Seja f uma função convexa. Dizemos que $y \in \mathbb{R}^n$ é subgradiente de f no ponto x se*

$$f(z) \geq f(x) + y^T(z-x)\, \forall z.$$

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

Definição **14** *Dizemos que $(\bar{x}, \bar{y}) \in A \times B$ é um ponto de sela de uma função $f\ A \times B \to \mathbb{R}$ quando*

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq f(x, \bar{y})$$
$$\forall x \in A \text{ e } \forall y \in B.$$

3 Elementos da Análise Convexa

Neste capítulo apresentamos os resultados estudados sobre análise convexa, As demonstrações para os fatos apresentados podem ser encontrados em [1].

Começamos nosso estudo com algumas propriedades dos conjuntos convexos.

- A intersecção de conjuntos convexos é um conjunto convexo;
- O fecho e o interior de um conjunto convexo são convexos;
- Se D_1 e D_2 são conjuntos convexos fechados e um deles é limitado então $D_1 + D_2$ é convexo e fechado;
- Um conjunto é convexo se, e somente se, todo ponto desse conjunto puder ser escrito como combinação convexa de outros pontos do conjunto.

Com estas propriedades obtemos que um conjunto poliedral é convexo, ou seja, conjuntos delimitados por restrições lineares são convexos.

Se D for um conjunto convexo, fechado e não vazio e R_D for o conjunto de todas as direções de recessão do conjunto D então:

- R_D é um cone convexo, fechado e não vazio.
- $d \in R_D$ se, e somente se, $\exists x \in D$ t.q. $x + td \in D \; \forall t \in I\!R_+$.
- $R_D \neq 0$ se, e somente se, D é ilimitado.

Estudamos também propriedades sobre fechos convexos:

- O fecho convexo de um conjunto qualquer é o conjunto de todas as combinações convexas de pontos deste conjunto;
- O fecho convexo de um conjunto compacto é compacto;
- O dual de um cone é igual ao dual de seu fecho convexo;

Estudamos alguns fatos sobre cones duais e fechos cônicos:

- O cone dual de um cone não vazio sempre é convexo e fechado;
- O fecho cônico de um conjunto convexo pode ser escrito como

$$cone\, D = \{d \in I\!R^n | d = \alpha x, x \in D, \alpha \in I\!R_+\};$$

- O cone tangente de um conjunto convexo pode ser escrito como

$$\tau_D(x^*) = cl\, V_d(x^*) = cl\, cone(D - \{x^*\}),$$

  onde $cl\, D =$ fecho do conjunto D;
- O cone tangente, de um conjunto convexo, é convexo e fechado;
- O cone e seu fecho possuem o mesmo dual.

Com essas propriedades conseguimos condições necessárias de primeira ordem de otimalidade para um problema com conjunto viável convexo:

TEOREMA **1** *Sejam $D \subset I\!R^n$ um conjunto convexo e $f : I\!R^n \to I\!R$ uma função diferenciável no ponto $x^* \in D$, se x^* é um minimizador local de f no conjunto D, então*

$$(x - x^*)^T \nabla f(x^*) \geq 0 \; \forall x \in D.$$

Se f é convexa, então esta condição também é suficiente.

Quando D é um conjunto convexo e f é uma função convexa em D também convexo, dizemos que

$$\begin{array}{c} \min\, f(x) \\ \text{s.a } x \in D \end{array}$$

é um *problema de minimização convexa*, e vimos que todo minimizador deste problema é global, e ainda, se f for estritamente convexa o minimizador é único.

Como estudamos problema com funções convexas, vimos algumas propriedades básicas destas:

- A soma de múltiplos não negativos de um número finito de funções convexas é uma função convexa;
- O supremo de funções convexas é uma função convexa;

- Se g é uma função convexa e h é uma função convexa e não decrescente, então $f(x) = g(h(x))$ também é convexa;

- Os conjuntos de níveis de uma função convexa são convexos;

- Uma função convexa é contínua em qualquer subconjunto aberto de seu domínio.

Fazendo uma extensão da definição de funções convexas com imagem em $\mathbb{R}$ para funções com imagem em $\mathbb{R}^m$, temos que uma função é convexa se todas as suas componentes forem funções convexas. Daí chegamos à conclusão que um conjunto definido por funções convexas, na forma de inequações, e funções afins, na forma de equações, é um conjunto convexo.

Se a função $f : \mathbb{R}^n \to \mathbb{R}$ é fortemente convexa no $\mathbb{R}^n$, então seu conjunto de nível

$$L_{f,R^n}(c) = \{x \in \mathbb{R}^n | f(x) \leq c\}$$

é compacto para todo $c \in \mathbb{R}$.

Desta afirmação tiramos que uma função fortemente convexa tem um, e apenas um, minimizador em um conjunto fechado não vazio qualquer.

Vimos que se o conjunto de nível de uma função convexa é não vazio e limitado para alguma constante, então este conjunto de nível é limitado para qualquer constante.

Se estivermos maximizando uma função convexa sobre um conjunto convexo, então temos uma solução que é um ponto extremo do conjunto convexo. Logo se temos um problema de minimização em programação linear, como o $min(c^T x) = -max(-c^T x)$, uma solução ótima é um ponto extremo do conjunto viável.

Se f é uma função diferenciável em $D \subset \mathbb{R}^n$ convexo e aberto, as seguintes propriedades são equivalentes.

- f é convexa em D.

- $\forall x \in D$ e $\forall y \in D$, $f(y) \geq f(x) + (\nabla f(x))^T (y - x)$.

- $\forall x \in D$ e $\forall y \in D$, $(\nabla f(y) - \nabla f(x))^T (y - x) \geq 0$.

- A Hessiana de f é semidefinida positiva em todo ponto em D (Se f for duas vezes diferenciável em D).

Podemos apresentar agora as condições necessárias e suficientes para problemas de minimização convexa.

Teorema **2** *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : C \to \mathbb{R}$ uma função convexa e diferenciável no conjunto aberto C que contém D. Então x^* é um minimizador de f em D se, e somente se,*

$$(\nabla f(x^*))^T (x - x^*) \geq 0 \, \forall x \in D,$$

que é equivalente a

$$(\nabla f(x))^T (x - x^*) \geq 0 \, \forall x \in D.$$

Ou, equivalentemente, se D é fechado,

$$x^* = P_D(x^* - \alpha \nabla f(x^*)) \text{ para algum } \alpha > 0.$$

O gradiente da função objetivo em um problema de minimização convexa é constante no conjunto das soluções. Daí pode ser provado que o conjunto de soluções de um problema de programação quadrática é um conjunto poliedral.

Se f é convexa, então

$$f(x+\alpha d) \geq f(x) + \alpha \nabla_d f(x),\ \forall \alpha \in I\!R_+,$$

onde d é uma direção de $I\!R^n$ e $\nabla_d f(x)$ é a derivada direcional na direção d.

Vimos que a derivada direcional de uma função é semicontínua superiormente, e que se uma função convexa é diferenciável, então a sua derivada é contínua.

Vemos que o subdiferencial de uma função convexa definida em todo espaço é um conjunto convexo, compacto e não vazio em cada ponto.

Temos que f é uma função convexa e diferenciável no ponto x se, e somente se, o subdiferencial neste ponto possui apenas um elemento.

Com estas informações conseguimos obter uma condição de otimalidade para minimização de uma função convexa num conjunto convexo:

Teorema **3** *Sejam f uma função convexa e D um conjunto convexo. Então $x^* \in I\!R^n$ é um minimizador de f em D se, e somente se,*

$$\exists y \in \partial f(x^*) \text{ tal que } y^T(x-x^*) \geq 0\ \forall x \in D.$$

Podemos ver que o subdiferencial de uma função convexa é limitado em conjuntos limitados. E que se f é uma função convexa, a sequência $\{x^k\}$ converge para x e $y^k \in \partial f(x)\, \forall k$, então a sequência $\{y^k\}$ é limitada e todos os seus pontos de acumulação pertencem ao subdiferencial de f no ponto x.

E ainda vimos que o subdiferencial da soma de funções convexas é a soma dos subdiferenciais destas funções.

Estudamos também uma outra forma para o Teorema 1, mas para chegar nela foi preciso antes estudar o operador projeção.

A projeção de um ponto x sobre um conjunto D é o ponto $y \in D$ tal que a distância, dada pela norma Euclidiana, entre x e y é mínima, ou seja, é o ponto y que satisfaz:

$$\begin{array}{c} \min \|y-x\|_2 \\ \text{s.a } y \in D. \end{array}$$

Se $D \subset I\!R^n$ é fechado então todo $x \in I\!R^n$ possui uma projeção, denotada por $P_D(x)$, se D for convexo e fechado, então a projeção é única. Assim conseguimos a outra forma para o Teorema 1, que é mais útil, no sentido de aplicações a métodos computacionais.

Teorema **4** *Sejam $D \subset I\!R^n$ um conjunto convexo e fechado e $f : I\!R^n \to I\!R$ diferenciável no ponto $x^* \in D$.*

Se x^* é um minimizador local de f no conjunto D, então

$$P_D(x^* - \alpha \nabla f(x^*)) = x^*\ \forall \alpha \in I\!R_+.$$

Quando f é convexa esta condição é também suficiente.

Estudamos alguns fatos sobre separação de conjuntos, tais como:

- se um ponto pertence à fronteira de um conjunto convexo, então ele pode ser separado deste conjunto;

- se um ponto não pertence ao fecho de um conjunto, então este ponto pode ser separado estritamente deste conjunto;

- dois conjuntos convexos, não vazios, com intersecção vazia são separáveis;

- se no item acima temos que os conjuntos são fechados e pelo menos um deles é limitado, então a separação é estrita.

Com estes fatos podemos demonstrar que o dual de um cone dual não vazio é o fecho convexo do cone primal.

Vimos que um conjunto convexo possui pelo menos um ponto extremo se, e somente se, ele não contém nenhuma reta.

Um conjunto convexo compacto é igual ao fecho convexo de seus pontos extremos, logo podemos dizer que todo ponto de um conjunto convexo compacto pode ser representado por uma combinação convexa de, no máximo, $n+1$ de seus pontos extremos.

Vimos que em conjuntos delimitados por inequações lineares, ou seja, conjuntos com a forma

$$D = \{x \in I\!R^n | Ax \leq b\},$$

onde $A \in I\!R^{m \times n}$ e $b \in I\!R^m$, temos que x é ponto extremo de D se, e somente se, as linhas de A que correspondem às restrições ativas, ou seja, as restrições satisfeitas na igualdade, em x formam uma matriz de posto n.

Outra característica de um conjunto poliedral é que quando possui pontos extremos, ele pode ser gerado pela soma do fecho convexo de seus vértices com o cone de suas direções de recessão.

Estudamos também teoremas de alternativa, para isto tomamos dois sistemas de restrições lineares, equações e inequações, tais que um, e somente um, dos sistemas possui solução.

Teorema **5** *Teorema de Motzkin*

Para quaisquer matrizes $A_i \in I\!R^{m_i \times n}$, $i = 0, 1, 2$, um, e somente um, dos seguintes sistemas possui solução:

$$A_0 x > 0,\ A_1 x = 0,\ A_2 x \geq 0,\ x \in I\!R^n,$$

ou

$$A_0^T y^0 + A_1^T y^1 + A_2^T y^2 = 0,$$
$$y^0 \in I\!R_+^{m_0} \backslash \{0\},\ y^1 \in I\!R^{m_1}, y^2 \in I\!R_+^{m_2}.$$

Com o Teorema de Motzkin obtivemos o *Lema de Farkas*

Lema **1** *Lema de Farkas*

Para qualquer matriz $A \in I\!R^{m \times n}$, um, e somente um, dos dois sistemas possui solução:

$$Ax \leq 0,\ b^T x > 0\ x \in I\!R^n,$$

ou

$$A^T y = b,\ y \in I\!R_+^m.$$

Com isto podemos obter as condições de otimalidade em forma primal-dual no caso de restrições lineares.

Corolário **1** *Seja $x^* \in I\!R^n$ uma solução local do problema*

$$\begin{array}{c} min\ f(x) \\ s.a\ Ax \leq b, \end{array}$$

onde f é diferenciável em x^*, $A \in I\!R^{m \times n}$, $b \in I\!R^m$.

Então existe $y^* \in I\!R_+^m$ tal que

$$\nabla f(x^*) + A^T y^* = 0, y_i^*, i \notin I(x^*),$$

onde $I(x)$ é o conjunto de índices das restrições ativas.

Foram estudados também outros teoremas de Alternativa, como o *Teorema da Alternativa de Tucker*, o *Teorema da Alternativa de Gale I* e *Teorema da Alternativa de Gordan*, que serão omitidos neste texto, pois os consideramos de menor importância para a perfeita compreensão do estudo do nosso tema. Estes resultados podem ser encontrados em [1].

4 Dualidade

Em problemas de otimização convexa podemos associar o problema original, primal, com um outro problema, o dual, que é sob certa perspectiva equivalente ao primal, e pode ser mais fácil de resolver. Começamos estudando a dualidade em problemas lineares e depois estudamos casos mais gerais para problemas não-lineares convexos. As demosntrações para os resultados apresentados neste capítulo podem ser encontrados em [1] e [2] por exemplo.

4.1 Dualidade em Programação Linear

Tomomemos um problema de programação linear:

$$\begin{array}{c} \min c^T x \\ \text{s.a } Ax \geq b, \end{array} \tag{1}$$

onde $A \in I\!R^{m \times n}$, $c \in I\!R^n$, $b \in I\!R^m$ e $x \in R^n$, seu dual é dado por:

$$\begin{array}{c} \max b^T y \\ \text{s.a } A^T y = c \\ y \geq 0 \end{array} \tag{2}$$

com $y \in I\!R^m$. Daí temos que se $\bar{x}$ é viável para (1) e $\bar{y}$ viável para (2), então $b^T \bar{y} \leq c^T \bar{x}$.

Esta desigualdade é chamada de dualidade fraca. Quando temos a igualdade os pontos $\bar{x}$ e $\bar{y}$ são ótimos; chamamos este fato de dualidade forte.

Para qualquer par primal-dual de programação linear temos uma destas possibilidades satisfeitas:

- Os dois problemas são infactíveis;
- Um dos problemas é ilimitado, então o outro é infactível;
- Um dos problemas é limitado não vazio, então os dois possuem solução.

4.2 Dualidade para um problema geral

Veremos, a seguir, algumas propriedades do dual para problemas não lineares. Consideremos um problema de minimização (não linear) na forma geral:

$$\begin{array}{c} \min f(x) \\ \text{s.a } x \in D \\ D = \{x \,|\, g(x) \leq 0, h(x) = 0\} \end{array} \tag{3}$$

onde $x \in I\!R^n$, $f : I\!R^n \to I\!R$, $g : I\!R^n \to I\!R^m$ e $h : I\!R^n \to I\!R^l$. Chamamos o problema (3) de *problema primal.*

Usaremos a função *Lagrangeana* para produzir o dual deste problema, ou seja, passaremos as restrições do problema, com pesos, para a função objetivo. Asim podemos definir a Lagrangiana como

$$L(x, \lambda, \nu) = f(x) + \lambda^T h(x) + \nu^T g(x).$$

Primeiro vemos que

$$sup_{(\lambda,\nu)} L(x, \lambda, \nu) = \begin{cases} f(x) & \text{se} \quad x \in D, \\ \infty & \quad\quad \text{c.c} \end{cases}$$

Logo podemos escrever o primal como

$$\begin{aligned} &\min\{sup_{(\lambda,\nu)}L(x,\lambda,\nu)\} \\ &\text{s.a } x\in D=\{x\,|\,sup_{(\lambda,\nu)}L(x,\lambda,\nu)<\infty\}. \end{aligned} \tag{4}$$

Obtemos o *problema dual* trocando a ordem de minimização com a maximização, o que resulta em:

$$\begin{aligned} &\max\{inf_x L(x,\lambda,\nu)\} \\ &\text{s.a } (\lambda,\nu)\in\Delta=\{(\lambda,\nu)\,|\,inf_x L(x,\lambda,\nu)>-\infty\}. \end{aligned} \tag{5}$$

A *função dual* é definida como:

$$\phi(\lambda,\nu)=inf_x L(x,\lambda,\nu).$$

Daí obtemos uma outra forma de escrever o problema dual:

$$\begin{aligned} &\max\ \phi(\lambda,\nu) \\ &\text{s.a } (\lambda,\nu)\in\Delta. \end{aligned}$$

Um resultado importante obtido é que o conjunto viável Δ do problema dual é convexo e a função dual é côncava, independentemente da forma do problema primal, logo podemos escrever o problema dual como um problema de minimização convexa.

Outro resultado importante é a *dualidade fraca*, que nos diz que para todo par primal dual, temos

$$\begin{aligned} &\phi(\lambda,\nu)\le f(x) \\ &\forall x\in D,\ (\lambda,\nu)\in\Delta. \end{aligned}$$

Tomando $p^*=inf_x f(x)$, $d^*=sup_{(\lambda,\nu)}\phi(\lambda,\nu)$, $x^*\in D$ e $(\lambda^*,\nu^*)\in\Delta$, tiramos do resultado acima que:

- Se $f(x^*)=d^*$, então x^* é solução do problema primal;
- Se $\phi(\lambda^*,\nu^*)=p^*$, então (λ^*,ν^*) é solução do problema dual;
- Se $f(x^*)=\phi(\lambda^*,\nu^*)$, então x^* é solução do problema primal e (λ^*,ν^*) é solução do problema dual.

Em problemas não lineares nem sempre podemos satisfazer esta terceira afirmação, ou seja, temos $d^*<p^*$, mas ainda assim podemos tirar conclusões bastante úteis do problema dual, como por exemplo, se $d^*=-\infty$ então o conjunto $D=\emptyset$, e também qualquer ponto viável do problema dual nos fornece um limitante inferior para o valor ótimo do problema primal.

Em um problema de minimização convexa, onde é safisfeita a condição de qualificação de Slater,

$$\exists\bar{x}\in I\!\!R^n\ t.q.\ h(\bar{x})=0\ e\ g(\bar{x})<0, \tag{6}$$

temos que $d^*=p^*$, ou seja não há *gap de dualidade.*

Ainda considerando um problema de minimização convex satisfazendo (6), temos

Teorema **6** *Teorema de Kuhn-Tucker*
$x^*\in D$ é uma solução do primal se, e somente se, existe $(\lambda^*,\nu^*)\in I\!\!R^l\times I\!\!R^m_+$ tal que

$$\begin{aligned} &L(x^*,\lambda^*,\nu^*)=min_x L(x,\lambda^*,\nu^*), \\ &\nu_i^* g_i(x^*)=0,\ i=1,\cdots,m. \end{aligned}$$

Usando ainda a definição de ponto de sela (definição 14), podemos reescrever a afirmação anterior dizendo que $x^*\in I\!\!R^n$ é uma solução do problema primal se, e somente se, existe $(\lambda^*,\nu^*)\in I\!\!R^l\times I\!\!R^m_+$ tal que (x^*,λ^*,ν^*) é um ponto de sela em $I\!\!R^n\times(I\!\!R^l\times I\!\!R^m_+)$.

Concluimos este capítulo com o fato que, se $(\lambda,\nu)\in\Delta$ e $x(\lambda,\nu)\in I\!\!R^n$ tais que

$$L(x(\lambda,\nu),\lambda,\nu)=\phi(\lambda,\nu)=min_{x\in I\!\!R^n}L(x,\lambda,\nu),$$

então

$$-(h(x(\lambda,\nu)),g(x(\lambda,\nu)))\in\partial(-\phi)(\lambda,\nu).$$

5 Métodos de Pontos Interiores

Nesta seção apresentaremos alguns métodos estudados para resolver problemas de otimização convexa, começamos estudando um método para resolver problemas irrestritos.

Tomando o problema

$$min\, f(x), \tag{7}$$

onde $f : I\!R^n \to I\!R$ é convexa e duas vezes diferenciável, estamos supondo que existe solução para este problema, e que esta seja única.

Como f é convexa e diferenciável a condição, necessária e suficiente, para que x^* seja ótimo é que

$$\nabla f(x^*) = 0.$$

Definição **15** *$d \in I\!R^n$ é um direção de descida para f a partir de $x \in I\!R^n$ se existe $\bar{t}$ tal que $f(x + td) < f(x)$ para todo $t \in (0, \bar{t}\,]$.*

Teorema **7** *Se f é diferenciável e $\nabla f(x)^T d < 0$, então d é direção de descida para f a partir de x.*

Então usando uma sequência $x^{(k+1)} = x^{(k)} + t_k d^k$, com $d^{(k)} \in I\!R^n$ direções de descida e $0 < t_k \in I\!R$, temos um método de direção de descida:

Método de Descida Geral

Dados: $x^{(0)}$, $k = 0$, $\epsilon > 0$.

Repetir:

- Determinar uma direção de descida $d^{(k)}$.
- Escolher tamanho do passo $t_k > 0$, de modo que $f(x^{(k)} + t_k d^{(k)}) < f(x^{(k)})$.
- Fazer $x^{(k+1)} = x^{(k)} + t_k d^{(k)}$.
- Se $|f(x^{(k+1)} - f(x^{(k)})| < \epsilon$, parar: $x^{(k+1)}$ é solução ótima.
- Fazer $k = k + 1$.

Outro método estudado foi o *Método de Newton*. Neste caso estaremos resolvendo problemas com restrições lineares de igualdade,

$$\begin{array}{c} min\, f(x) \\ s.a\, Ax = b, \end{array} \tag{8}$$

onde $f : I\!R^n \to I\!R$ é convexa e duas vezes diferenciável, $A \in I\!R^{m \times n}$, com $posto\, A = p < n$. (novamente estamos suponto que exista uma solução para o sistema).

Logicamente todo problema com restrições de igualdade podem ser resolvidos através de um problema irrestrito equivalente; aqui iremos resolver o problema com as restrições.

Temos que x^* é ótimo se, e somente se, $\exists \nu^* \in I\!R^m$

$$\begin{array}{c} Ax^* = b, \\ \nabla f(x^*) + A^T \nu^* = 0. \end{array}$$

O método de Newton é amplamente utilizado em métodos de pontos interiores dada a sua rápida taxa de convergência, que vem do fato deste método utilizar informações da segunda derivada da função objetivo no cálculo das direções de descida. Esta direção de descida é escolhida de modo a minimizar uma aproximação quadrática da função objetivo na última iteração.

Quando, usamos uma aproximação quadrática no lugar de f, a direção de descida, d_{nt}, é dada por:

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_{nt} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}, \tag{9}$$

onde w é o valor ótimo da variável dual associada ao problema quadrático.

Definimos o decréscimo de Newton para um problema com restrições de igualdade como

$$\lambda(x) = (d_{nt}^T \nabla^2 f(x) d_{nt}).$$

Pode-se provar [2] que $\lambda(x)^2/2$ nos dá uma estimativa para $f(x)-p^*$ então $\lambda(x)$ pode ser utilizado como base de um critério de parada.

A direção d_{nt} é uma direção facttível, ou seja, $Ad_{nt} = 0$, vemos isto pela segunda linha do sistema (9), e também é uma direção de descida pois a derivada direcional de f na direção d_{nt} é negativa, na verdade vale $-\lambda(x)^2$.

Método de Newton para problemas com restrições de igualdade
Dados: $x^{(0)}$ t.q. $Ax^{(0)} = b$, $k = 0$, $\epsilon > 0$.
Repetir:

- Determinar uma direção de Newton, $d_{nt}^{(k)}$, e o decréscimo, $\lambda(x^{(k)})$.
- Se $\lambda^2/2 \leq \epsilon$, parar.
- Escolher tamanho do passo $t_k > 0$, de modo que $f(x^{(k)} + t_k d^{(k)}) < f(x^{(k)}$.
- Fazer $x^{(k+1)} = x^{(k)} + t_k d_{nt}^{(k)}$.
- Fazer $k = k + 1$.

Este método é chamado método de descida factível, já que toda iteração é factível, com $f(x^{(k+1)}) < f(x^{(k)})$.

Passamos então ao *Método das Barreiras* para resolução de problemas convexos com restrições de desigualdade,

$$\begin{gathered} min\, f(x) \\ s.a\, g(x) \leq 0 \\ Ax = b, \end{gathered} \tag{10}$$

onde $f : I\!R^n \to I\!R$, $g : I\!R^n \to I\!R^m$ e $A \in I\!R^{p\times n}$. Assumimos que existe alguma solução ótima, e que existem pontos viáveis que satisfazem as desigualdades estritamente. Isto implica que as condições de Slater (6) são satisfeitas; logo existe ponto ótimo do dual, $(\lambda^*, \nu^*) \in I\!R^m \times I\!R^p$, que junto com o ótimo do primal, x^*, satisfaz as condições

$$\begin{gathered} Ax^* = b \\ g(x^*) \leq 0 \\ \lambda_i g_i(x^*) = 0,\, i = 1, \cdots, m \\ \lambda^* \geq 0 \\ \nabla f(x^*) + \lambda^T \nabla g(x^*) + A^T \nu^* = 0. \end{gathered} \tag{11}$$

Os métodos de pontos interiores resolvem o problema (10) aplicando o método de Newton para uma sequência de problemas com restrição de igualdade. Nosso interesse é um tipo especial de método de pontos interiores, o método das barreiras.

Vamos então aproximar o problema (10) por um problema apenas com restrições de igualdade:

$$\begin{gathered} min\, f(x) + (-1/t)\mathbf{1}^T log(-f(x)) \\ s.a\, Ax = b. \end{gathered} \tag{12}$$

Definindo a função logaritmo para um vetor como o vetor cujas componentes são o logaritmo de cada componente do vetor original, o problema (12) tem função objetivo convexa e diferenciável.

A função

$$\phi(x) = \mathbf{1}^T log(-f(x)),$$

com domínio em $\{x | f(x) < 0\}$, é chamada de barreira logarítmica.

Vemos que quanto maior for o valor da função objetiva em (12), melhor é a aproximação, porém mais difícil é a resolução do problema através do método de Newton.

O que faremos neste método é ir aumentando o valor de t, e resolvendo a cada iteração o problema resultante usando o método de Newton, com ponto inicial dado pela solução dada pela iteração anterior.

Definimos o *caminho central* como o conjunto formado por $x^*(t)$, $t > 0$, onde $x^*(t)$ é solução de (12), e podemos ver que $x^*(t)$ é estritamente factível para o problema original e existe $\bar{\nu} \in I\!R^p$, tal que

$$0 = t\nabla f(x^*(t)) + \nabla \phi(x^*(t)) + A^T \bar{\nu}$$

Com encontrando o gradiente de ϕ podemos encontrar um valor para a variável dual associada a cada ponto do caminho central:

$$\begin{array}{c} \lambda_i(t) == -\frac{1}{tf_i(x^*(t))}\, i = 1, \cdots, m, \\ \nu^*(t) = \bar{\nu}/t, \end{array} \tag{13}$$

e este par, $(\lambda^*(t), \nu^*(t))$, é dual factível, e o gap de dualidade associado com $x^*(t)$ e $(\lambda^*(t), \nu^*(t))$ vale m/t.

Com isto podemos escrever o método em pseudo código:

Método das Barreiras
Dados: $x^{(0)}$ estritamente factível, $t_0 > 0$, $\mu_0 > 1$, $k = 0$, $\epsilon > 0$.
Repetir:

- Calcular $x^*(t_k)$ resolvendo o sistema dado por (12) começando com $x^{(k)}$.
- Fazer $x^{(k+1)} = x^*(t_k)$.
- Se $m/t < \epsilon$, parar.
- Fazer $t_{k+1} = \mu t_k$.
- Atualizar $k = k + 1$.

Este método exige que tenhamos um ponto estritamente factível como ponto inicial; aplicamos então um processo preeliminar, chamado *Fase 1.* Para isto criamos o seguinte problema

$$\begin{array}{c} min\, s \\ s.a\, g(x) \leq s\mathbf{1} \\ Ax = b. \end{array} \tag{14}$$

onde $\mathbf{1}$ é um vetor com todas as componentes iguais a 1 e $s \in I\!R$. Esta variável é conhecida como infactibilidade máxima, e a meta é deixá-la abaixo de 0, de modo a conseguir $x^{(0)}$ estritamente factível para usarmos no problema original.

6 Conclusão

Portanto podemos ver os benefícios de se trabalhar com problemas convexos como, por exemplo, não termos que nos preocupar se o método utilizado está convergindo para um minimizador local, pois todo minimizador em um problema convexo é global. Outra vantagem é que temos as condições necessárias e suficientes de otimalidade bem definidas, o que também nos dá possibilidades de desenvolvimento de métodos para resolver estes problemas.

Trabalhar com o problema dual também oferece vantagens, tanto para problemas lineares quanto para problemas não lineares, uma destas vantagens são os métodos primal-dual, por exemplo o método das barreiras, que usam informações tanto do primal quanto do dual para resolver o problema de forma muito eficiente.

Mesmo quando não é possível aplicar um método primal-dual o dual nos é muito útil dado que a função objetivo do dual de qualquer problema é sempre uma função côncava, logo podemos transformá-la em uma função convexa e daí pela desigualdade fraca obteremos pelo menos um limitante para o valor ótimo do primal, e ainda se o gap de dualidade for nulo este será o valor ótimo do problema primal.

Referências

[1] A. Izmailov, M. Solodov, *Otimização - volume 1.* IMPA, Rio de Janeiro, RJ, 2005.

[2] S.P. Boyd, L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[3] D.G. Luenberguer, Y.Ye, *Linear and Nonlinear Programming.* Springer Verlag, 2008.

[4] H. Hindi, A tutorial on convex optimization II: Duality and interior point methods, *American Control Conference*, 2006.