Fundamentos de otimização restrita

Marina Andretta

ICMC-USP

6 de novembro de 2018

Baseado no livro Numerical Optimization, de J. Nocedal e S. J. Wright.

Problema restrito

Em minimização restrita, queremos minimizar uma função de variáveis reais a um valor real com restrições aos valores das variáveis.

Ou seja, estamos interessados em resolver o seguinte problema:

$$\boxed{\begin{array}{lll} \text{Minimizar} & f(x) & \\ \text{Sujeita a} & c_i(x) = 0, & i \in \mathcal{E}, \\ & c_i(x) \geq 0, & i \in \mathcal{I}, \end{array}} \qquad (1)$$

onde

- $x \in \boldsymbol{R}^n$,
- $f, c_i : \boldsymbol{R}^n \to \boldsymbol{R}$ funções suaves,
- $\mathcal{E}$ e $\mathcal{I}$ conjuntos de índices das restrições de igualdade e desigualdade, respectivamente.

Problema restrito

Como no caso de problemas irrestritos, chamamos f de função objetivo.

Definimos o conjunto viável Ω como o conjunto dos pontos x que satisfazem as restrições. Ou seja,

$$\boxed{\Omega = \{x \mid c_i(x) = 0,\ i \in \mathcal{E};\ c_i(x) \geq 0,\ i \in \mathcal{I}\}.}$$

Soluções locais e globais

Já vimos que soluções globais são difíceis de encontrar quando estamos lidando com problemas irrestritos.

Esta situação poderia mudar ao acrescentarmos restrições, já que o conjunto viável pode excluir minimizadores locais e pode ser comparativamente mais fácil encontrar minimizadores globais.

No entanto, acrescentar restrições pode deixar o problema muito mais difícil de ser resolvido.

Soluções locais e globais

Por exemplo, considere o problema

$$\boxed{\begin{array}{ll} \text{Minimizar} & \|x\|_2^2 \\ \text{Sujeita a} & \|x\|_2^2 \geq 1. \end{array}}$$

Sem a restrição, a função objetivo é uma quadrática convexa, com apenas um minimizador $x^* = 0$.

Com a restrição, qualquer vetor x que satisfaça $\|x\| = 1$ é solução do problema. Há infinitos minimizadores locais para $n \geq 2$.

Soluções locais e globais

Considere agora o seguinte exemplo:

$$\begin{array}{ll} \text{Minimizar} & (x_2 + 100)^2 + 0.01x_1^2 \\ \text{Sujeita a} & x_2 - \cos x_1 \geq 0. \end{array}$$

Sem a restrição, o problema tem uma única solução $x^* = (0, -100)^T$.

Com a restrição, há soluções locais perto dos pontos

$$(x_1, x_2) = (k\pi, -1), \; k = \pm 1, \pm 3, \pm 5, \ldots$$

Definição de diferentes tipos de soluções locais são simplesmente extensões das definições correspondentes às de otimização irrestrita. No entanto, agora nos restringimos a pontos viáveis na vizinhança de x^*.

Temos a seguinte definição:

Um ponto x^* é uma solução local para o problema (1) se $x^* \in \Omega$ e existe uma vizinhança $\mathcal{N}$ de x^* tal que $f(x^*) \leq f(x)$ para todo $x \in \mathcal{N} \cap \Omega$.

Soluções locais e globais

De maneira similar, temos também as seguintes definições:

Um ponto x^* é uma solução local estrita (ou forte) para o problema (1) se $x^* \in \Omega$ e existe uma vizinhança $\mathcal{N}$ de x^* tal que $f(x^*) < f(x)$ para todo $x \in \mathcal{N} \cap \Omega$, $x^* \neq x$.

Um ponto x^* é uma solução local isolada para o problema (1) se $x^* \in \Omega$ e existe uma vizinhança $\mathcal{N}$ de x^* tal que x^* é a única solução local em $\mathcal{N} \cap \Omega$.

Usaremos os termos solução e minimizador como sinônimos.

Suavidade

Suavidade, tanto da função objetivo como das restrições, é importante para caracterizar soluções, como no caso de problemas sem restrições.

Isso garante que a função objetivo e as restrições se comportam de maneira razoavelmente previsível, o que permite que os algoritmos façam boas escolhas para direções de busca.

Quando uma função não é suave, ela possui “quinas” e “pulos” nos pontos não-diferenciáveis.

Se fizermos o desenho de uma região viável, geralmente observamos várias “quinas”. Isso significa que as funções de restrição que definem a região viável não são suaves?

A resposta, em geral, é não. Bordas não suaves podem, muitas vezes, ser descritas por uma coleção de funções de restrição suaves.

Suavidade

Considere, por exemplo, a região definida pela restrição não suave

$$\boxed{\|x\|_1 = |x_1| + |x_2| \leq 1.}$$

Ela pode ser descrita como o conjunto de restrições suaves

$$\boxed{x_1 + x_2 \leq 1, \ x_1 - x_2 \leq 1, \ -x_1 + x_2 \leq 1, \ -x_1 - x_2 \leq 1.}$$

Em geral, as funções de restrição são escolhidas de modo que cada uma represente uma parte suave da borda de Ω.

Suavidade

Problemas de otimização irrestrita não suaves podem, às vezes, ser reformulados como problemas restritos suaves.

Considere o problema

$$\boxed{\text{Minimizar} \quad f(x) = \max\{x^2, x\}.}$$

A função f possui "quinas" em $x = 0$ e $x = 1$. A solução é dada por $x^* = 0$.

Suavidade

Podemos obter uma formulação restrita e suave para o problema adicionando uma variável artificial t e escrevendo

$$\boxed{\begin{array}{ll} \text{Minimizar} & t \\ \text{Sujeita a} & t \geq x, \\ & t \geq x^2. \end{array}}$$

Estas técnicas de reformulação são comumente usadas quando f é o máximo de uma coleção de funções ou quando f é a norma 1 ou a norma infinito de uma função vetorial.

Soluções de problemas restritos

Para verificar como são caracterizadas as soluções de problemas de otimização restrita, veremos três exemplos.

Uma definição que será bastante usada é: dado um ponto viável x, a restrição de desigualdade $i \in \mathcal{I}$ é dita

- ativa, se $c_i(x) = 0$ e
- inativa, se $c_i(x) > 0$.

Exemplo de problema com uma restrição de igualdade

Considere o problema de duas variáveis e uma única restrição de igualdade

$$\boxed{\begin{array}{ll} \text{Minimizar} & f(x) = x_1 + x_2 \\ \text{Sujeita a} & c_1(x) = x_1^2 + x_2^2 - 2 = 0. \end{array}} \tag{2}$$

Temos, então, que $\mathcal{E} = \{1\}$ e $\mathcal{I} = \emptyset$.

Exemplo de problema com uma restrição de igualdade

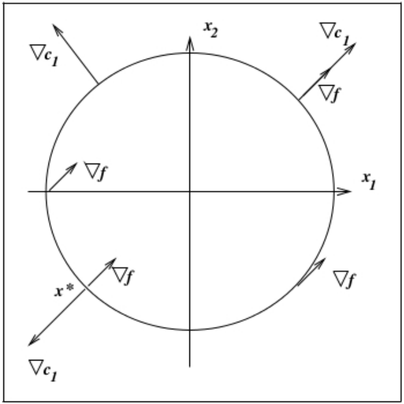

Figura: Exemplo de problema com uma restrição de igualdade (Figura 12.3 de Numerical Optimization, de J. Nocedal e S. J. Wright)

Exemplo de problema com uma restrição de igualdade

Podemos verificar que o conjunto viável é uma circunferência de raio $\sqrt{2}$ centrada na origem - apenas a borda da circunferência, não seu interior.

A solução é $x^* = (-1, -1)^T$. É fácil ver que, a partir de qualquer outro ponto na circunferência, podemos "andar" por pontos viáveis que fornecem valor de função objetivo menor.

Por exemplo, a partir do ponto $x = (\sqrt{2}, 0)^T$, qualquer movimento em sentido horário na circunferência produz decréscimo na função objetivo.

Exemplo de problema com uma restrição de igualdade

Podemos ver que, na solução x^*, a normal da restrição $\nabla c_1(x^*)$ é paralela a $\nabla f(x^*)$. Ou seja, existe um escalar λ_1^* tal que

$$\boxed{\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*).} \tag{3}$$

Neste caso, $\lambda_1^* = -\frac{1}{2}$.

Exemplo de problema com uma restrição de igualdade

Podemos obter (3) examinando a aproximação de primeira ordem da série de Taylor da função e da restrição.

Para manter viabilidade em relação à restrição $c_1(x) = 0$, pedimos que $c_1(x + d) = 0$. Ou seja,

$$\boxed{0 = c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d = \nabla c_1(x)^T d.}$$

Exemplo de problema com uma restrição de igualdade

Portanto, a direção d mantém viabilidade em relação a c_1, em primeira ordem, quando satisfaz

$$\boxed{\nabla c_1(x)^T d = 0.} \tag{4}$$

Exemplo de problema com uma restrição de igualdade

Além disso, queremos que a direção d produza decréscimo na função objetivo:

$$\boxed{0 > f(x+d) - f(x) \approx \nabla f(x)^T d.}$$

Ou seja,

$$\boxed{\nabla f(x)^T d < 0.} \tag{5}$$

Exemplo de problema com uma restrição de igualdade

Se, em um ponto x_k, existe uma direção d que satisfaz (4) e (5), podemos diminuir o valor da função objetivo mantendo a viabilidade. Ou seja, x_k não é um minimizador.

Daí segue uma condição de otimalidade necessária de primeira ordem para problemas restritos:

> se x^* é solução de (1) então não existe direção d que satisfaça tanto (4) como (5).

É possível mostrar que, se uma direção d não satisfaz nem (4) nem (5), temos que $\nabla f(x)$ e $\nabla c_1(x)$ são paralelos. Ou seja, $\nabla f(x) = \lambda_1 \nabla c_1(x)$, para algum λ_1.

Se esta condição não é satisfeita, a direção definida por

$$\boxed{d = -\left(I - \frac{\nabla c_1(x) \nabla c_1(x)^T}{\|\nabla c_1(x)\|^2} \right) \nabla f(x)}$$

satisfaz (4) e (5).

Exemplo de problema com uma restrição de igualdade

Defina a função Lagrangiana como

$$\boxed{\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x).}$$

Note que

$$\boxed{\nabla_x \mathcal{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x).}$$

Podemos, então, reformular a condição de otimalidade necessária de primeira ordem como

se x^* é solução de (1), então existe um escalar λ_1^* tal que

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0. \tag{6}$$

O escalar λ_1 é chamado de multiplicador de Lagrange para a restrição $c_1(x) = 0$.

Exemplo de problema com uma restrição de igualdade

Esta observação sugere que podemos encontrar soluções para o problema restrito (1) procurando pontos estacionários da função Lagrangiana.

Apesar de a condição (3) ser necessária para que um ponto x seja solução, ela não é suficiente.

No exemplo (2), o ponto $(1,1)^T$ satisfaz (3) com $\lambda_1 = \frac{1}{2}$. No entanto, este ponto não é uma solução do problema. De fato, $(1,1)^T$ é um maximizador da função f restrita a c_1.

Exemplo de problema com uma restrição de igualdade

Não podemos tornar a condição (3) suficiente apenas impondo uma condição sobre o sinal de λ_1.

Tome o problema (2), trocando a restrição $x_1^2 + x_2^2 - 2 = 0$ por $-x_1^2 - x_2^2 + 2 = 0$.

A solução para este novo problema continua sendo $(-1, -1)^T$, já que o conjunto viável é o mesmo do exemplo (2).

No entanto, o valor de λ_1^* que satisfaz (3) passa de $\lambda_1^* = -\frac{1}{2}$ para $\lambda_1^* = \frac{1}{2}$.

Exemplo de problema com uma restrição de desigualdade

Considere agora uma pequena variação do exemplo (2):

$$\boxed{\begin{array}{ll} \text{Minimizar} & f(x) = x_1 + x_2 \\ \text{Sujeita a} & c_1(x) = 2 - x_1^2 - x_2^2 \geq 0. \end{array}} \quad (7)$$

Temos, então, que $\mathcal{E} = \emptyset$ e $\mathcal{I} = \{1\}$.

Exemplo de problema com uma restrição de desigualdade

Neste problema, o conjunto viável é tanto a borda como o interior do círculo centrado na origem, de raio $\sqrt{2}$.

Note que, para todo ponto x na borda da região viável, a normal da restrição $\nabla c_1(x)$ aponta para o interior da região viável.

Podemos verificar que a solução para este problema ainda é $x^* = (-1, -1)^T$ e que a condição (3) vale para $\lambda_1^* = \frac{1}{2}$.

No entanto, este problema com uma restrição de desigualdade difere do problema (2), que tem uma restrição de igualdade, pelo fato de o sinal do multiplicador de Lagrange ter significado, como veremos a seguir.

Exemplo de problema com uma restrição de desigualdade

Mais uma vez, conjecturamos que um ponto x não é solução se podemos encontrar uma direção d que mantém viabilidade e diminui o valor da função objetivo em primeira ordem.

A principal diferença entre os problemas (2) e (7) está em como lidar com a viabilidade.

Como em (4), a direção d melhora a função objetivo, em primeira ordem, se $\nabla f(x)^T d < 0$.

A direção d mantém viabilidade se

$$\boxed{c_1(x) + \nabla c_1(x)^T d \geq 0.} \tag{8}$$

Para determinar se uma direção d satisfaz (5) e (8), vamos considerar dois casos:

1. x está estritamente no interior do círculo.
2. x está na borda do círculo.

Exemplo de problema com uma restrição de desigualdade

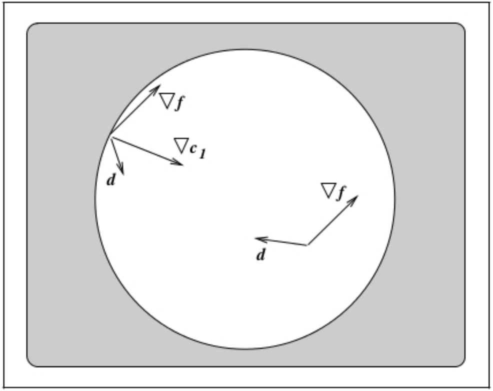

Figura: Exemplo de problema com uma restrição de desigualdade (Figura 12.4 de Numerical Optimization, de J. Nocedal e S. J. Wright)

Se x está estritamente no interior do círculo, temos que $c_1(x) > 0$.

Neste caso, qualquer direção d satisfaz (8), dado que seu tamanho seja suficientemente pequeno.

Em particular, sempre que $\nabla f(x) \neq 0$, podemos obter uma direção d que satisfaz tanto (5) como (8) fazendo

$$\boxed{d = -c_1(x)\frac{\nabla f(x)}{\|\nabla f(x)\|}.}$$

Esta direção não existe apenas quando $\nabla f(x) = 0$.

Exemplo de problema com uma restrição de desigualdade

Se x está na borda do círculo, temos que $c_1(x) = 0$.

Neste caso, as condições (5) e (8) se tornam

$$\boxed{\nabla f(x)^T d < 0, \quad \nabla c_1(x)^T d \geq 0.}$$

A primeira destas condições define um semi-espaço aberto e a segunda define um semi-espaço fechado.

Exemplo de problema com uma restrição de desigualdade

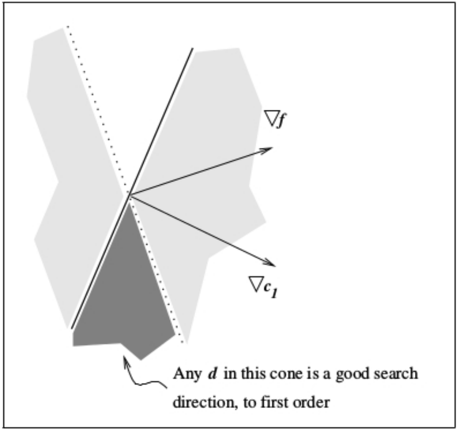

Figura: Exemplo de problema com uma restrição de desigualdade (Figura 12.5 de Numerical Optimization, de J. Nocedal e S. J. Wright)

Podemos verificar que estas duas regiões não se intersectam apenas quando $\nabla f(x)$ e $\nabla c_1(x)$ apontam para a mesma direção. Ou seja, quando

$$\boxed{\nabla f(x) = \lambda_1 \nabla c_1(x), \text{ para algum } \lambda_1 \geq 0.}$$

Note que o sinal do multiplicador é importante neste caso.

Se a condição (5) fosse satisfeita para um valor negativo de λ_1, $\nabla f(x)$ e $\nabla c_1(x)$ apontariam para direções opostas. Neste caso, as direções que satisfazem as condições (5) e (8) formam um semi-plano aberto.

Exemplo de problema com uma restrição de desigualdade

As condições de otimalidade tanto para o caso 1 (x no interior da região viável) como para o caso 2 (x na borda da região viável) podem ser resumidas usando a função Lagrangiana.

Quando não existe uma direção de descida de primeira ordem viável a partir de um ponto x^*, temos que

$$\boxed{\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0, \quad \text{para algum } \lambda_1^* \geq 0.} \tag{9}$$

Exemplo de problema com uma restrição de desigualdade

Também pedimos que

$$\boxed{\lambda_1^* c_1(x^*) = 0.} \tag{10}$$

Esta condição é conhecida como condição de complementaridade.

Ela implica que os multiplicadores de Lagrange λ_1 podem ser estritamente positivos apenas quando a restrição correspondente é ativa.

No primeiro caso, temos que $c_1(x^*) > 0$. Então (10) pede que $\lambda_1^* = 0$.

Portanto, por (9), temos que $\nabla f(x^*) = 0$, como queríamos.

No segundo caso, (10) permite que λ_1^* seja negativo, o que faz com que (9) seja equivalente a (6).

Exemplo de problema com duas restrições de desigualdade

Considere agora uma variação do exemplo (7):

$$\boxed{\begin{array}{ll} \text{Minimizar} & f(x) = x_1 + x_2 \\ \text{Sujeita a} & c_1(x) = 2 - x_1^2 - x_2^2 \geq 0, \\ & c_2(x) = x_2 \geq 0. \end{array}} \tag{11}$$

Neste caso, temos $\mathcal{E} = \emptyset$ e $\mathcal{I} = \{1, 2\}$.

Exemplo de problema com duas restrições de desigualdade

O conjunto viável é composto pelos pontos com $x_2 \geq 0$ que pertencem ao círculo centrado na origem e de raio $\sqrt{2}$.

A solução para este problema é $x^* = (-\sqrt{2}, 0)^T$. Neste ponto, ambas as restrições são ativas.

Exemplo de problema com duas restrições de desigualdade

Pelos argumentos usados nos exemplos anteriores, concluímos que uma direção d é viável e de descida, em primeira ordem, se d satisfaz as condições

$$\boxed{\nabla f(x)^T d < 0, \quad \nabla c_i(x)^T d \geq 0, \; i \in \mathcal{I}.} \tag{12}$$

É fácil ver que nenhuma direção que satisfaz estas condições pode ser encontrada em $x = (-\sqrt{2}, 0)^T$.

As condições $\nabla c_i(x)^T d \geq 0$, $i = 1, 2$, são ambas satisfeitas apenas quando d está no quadrante definido por $\nabla c_1(x)$ e $\nabla c_2(x)$. Mas todo vetor d neste quadrante satisfaz $\nabla f(x)^T d \geq 0$.

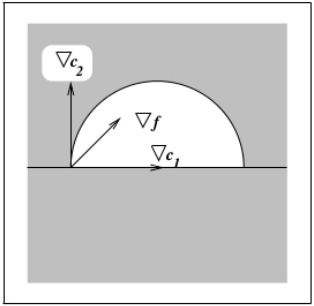

Figura: Exemplo de problema com duas restrições de desigualdade (Figura 12.6 de Numerical Optimization, de J. Nocedal e S. J. Wright)

Vejamos agora como a função Lagrangiana e sua derivada se comportam no problema (11) e na solução $(-\sqrt{2}, 0)^T$.

Primeiramente, incluímos um termo adicional $\lambda_i c_i(x)$ na função Lagrangiana para cada restrição adicional. Temos então

$$\boxed{\mathcal{L}(x, \lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x),}$$

onde $\lambda = (\lambda_1, \lambda_2)^T$ é o vetor de multiplicadores de Lagrange.

A extensão da condição (9), neste caso, é

$$\boxed{\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \quad \text{para algum } \lambda^* \geq 0.} \tag{13}$$

A inequação $\lambda^* \geq 0$ diz que todas as componentes de λ^* devem ser não-negativas.

A condição de complementaridade para ambas as restrições é

$$\boxed{\lambda_1^* c_1(x^*) = 0, \quad \lambda_2^* c_2(x^*) = 0.} \tag{14}$$

Exemplo de problema com duas restrições de desigualdade

Para $x^* = (-\sqrt{2}, 0)^T$, temos que

$$\boxed{\nabla f(x^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.}$$

É fácil verificar que, usando $\lambda^* = (1/(2\sqrt{2}), 1)^T$, temos que $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$ e $\lambda^* \geq 0$.

Além disso, como $c_1(x^*) = 0$ e $c_2(x^*) = 0$, valem as condições de complementaridade.

Exemplo de problema com duas restrições de desigualdade

Vamos considerar agora outros pontos viáveis do problema (11) que não são solução e analisar as propriedades da função Lagragiana e de seu gradiente nestes pontos.

Tomando o ponto $x = (\sqrt{2}, 0)^T$, temos novamente que ambas as restrições são ativas.

No entanto, o gradiente da função objetivo $\nabla f(x)$ não mais fica no quadrante definido pelas condições $\nabla c_i(x)^T d \geq 0, \; i = 1, 2$.

Exemplo de problema com duas restrições de desigualdade

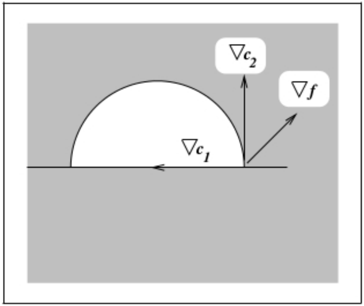

Figura: Exemplo de problema com duas restrições de desigualdade (Figura 12.7 de Numerical Optimization, de J. Nocedal e S. J. Wright)

Exemplo de problema com duas restrições de desigualdade

Uma das direções viáveis de descida de primeira ordem a partir deste ponto (ou seja, um vetor d que satisfaz (12)) é $d = (-1, 0)^T$.

Para este valor de x é fácil verificar a condição $\nabla_x \mathcal{L}(x, \lambda) = 0$ é satisfeita apenas quando $\lambda = (-1/(2\sqrt{2}), 1)^T$.

Note que a primeira componente de λ é negativa, então a condição (13) não é satisfeita para x.

Exemplo de problema com duas restrições de desigualdade

Analisemos agora o ponto $x = (1,0)^T$, para o qual apenas a segunda restrição do problema (11) é ativa.

Neste ponto, a linearização de f e c fornece as seguintes condições para que uma direção d seja viável e de descida em primeira ordem

$$\boxed{\begin{aligned} c_1(x) + \nabla c_1(x)^T d &\geq 0, \\ c_2(x) + \nabla c_2(x)^T d \geq 0 = \nabla c_2(x)^T d &\geq 0, \\ \nabla f(x)^T d &< 0. \end{aligned}} \tag{15}$$

Exemplo de problema com duas restrições de desigualdade

A primeira desigualdade se torna

$$\boxed{1 + \nabla c_1(x)^T d \geq 0,}$$

que pode ser satisfeita para qualquer d multiplicado por um escalar positivo suficientemente pequeno.

Exemplo de problema com duas restrições de desigualdade

Como

$$\boxed{\nabla f(x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \ \nabla c_2(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},}$$

é fácil verificar que $d = \left(-\frac{1}{2}, \frac{1}{4}\right)^T$ satisfaz (15) e é, portanto, uma direção viável de descida.

Exemplo de problema com duas restrições de desigualdade

Para verificar que x não satisfaz as condições de otimalidade (13) e (14), note que, como $c_1(x) > 0$, temos que $\lambda_1 = 0$.

Portanto, para satisfazer $\nabla_x \mathcal{L}(x, \lambda) = 0$, precisamos encontrar λ_2 tal que $\nabla f(x) - \lambda_2 \nabla c_2(x) = 0$.

Como não existe um valor de λ_2 que satisfaça esta equação, o ponto x não satisfaz as condições de otimalidade.

Condições de otimalidade de primeira ordem

Os 3 exemplos apresentados sugerem algumas condições para a caracterização de soluções para problemas de minimização restrita (1).

Entre elas, $\nabla_x \mathcal{L}(x, \lambda) = 0$, a não-negatividade de λ_i para todas as restrições de desigualdade $c_i(x)$ e a condição de complementaridade $\lambda_i c_i(x) = 0$.

Vamos agora generalizar estas condições e definir as condições de otimalidade de primeira ordem de maneira mais rigorosa.

Para o problema geral de minimização com restrições (1), a função Lagrangiana é dada por

$$\boxed{\mathcal{L}(x,\lambda) = f(x) - \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i c_i(x).}$$

O conjunto de (índices das) restrições ativas $\mathcal{A}(x)$, em qualquer ponto viável x, é a união do conjunto de (índices das) restrições de igualdade $\mathcal{E}$ e os índices das restrições de desigualdade que valem por igualdade em x:

$$\boxed{\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(x) = 0\}}$$

O vetor $\nabla c_i(x)$ é comumente chamado de vetor normal à restrição $c_i(x)$, pois ele geralmente é perpendicular às curvas de nível de c_i em x e, no caso de restrições de desigualdade, ele aponta para o lado viável desta restrição.

No entanto, é possível que $\nabla c_i(x)$ se anule devido à representação algébrica de c_i.

Com isso, $\lambda_i \nabla c_i(x) = 0$ para todo valor de λ_i e não aparece no gradiente da função Lagrangiana $\nabla_x \mathcal{L}$.

Condições de otimalidade de primeira ordem

Por exemplo, se trocamos a restrição $c_1(x) = x_1^2 + x_2^2 - 2 = 0$ pela restrição equivalente $c_1(x) = (x_1^2 + x_2^2 - 2)^2 = 0$ no exemplo (2), temos que $\nabla c_1(x) = 0$ para todo ponto viável x.

Em particular, a condição $\nabla f(x) = \lambda_1 \nabla c_1(x)$ não vale no ponto ótimo $(-1, -1)^T$.

Geralmente fazemos suposições chamadas condições de qualificação de restrição (*constraint qualifications*) para garantir que este tipo de comportamento degenerado não aconteça no ponto x em questão.

Uma condição de qualificação de restrição, provavelmente a mais usada no desenvolvimento de algoritmos, é a seguinte:

Definição 1 (LICQ): *Dados um ponto x^* e o conjunto de restrições ativas $\mathcal{A}(x^*)$, dizemos que a condição de qualificação de restrição de independência linear (LICQ) vale se o conjunto dos gradientes das restrições ativas $\{\nabla c_i(x^*),\ i \in \mathcal{A}(x^*)\}$ é linearmente independente.*

Note que, quando esta condição vale, nenhum dos gradientes das restrições ativas pode ser nulo.

Condições de otimalidade de primeira ordem

A condição LICQ nos permite definir condições de otimalidade para o problema de programação não-linear geral (1)

As condições a seguir são ditas de primeira ordem pois utilizam apenas informação do gradiente da função objetivo e das restrições.

Condições necessárias de primeira ordem

Teorema 1 (Condições necessárias de primeira ordem): *Suponha que x^* seja uma solução local de (1) e que LICQ vale em x^*. Então, existe um vetor de multiplicadores de Lagrange λ^*, com componentes λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, tal que as seguintes condições são satisfeitas:*

$$
\begin{array}{rcll}
\nabla_x \mathcal{L}(x^*, \lambda^*) & = & 0, & \\
c_i(x^*) & = & 0, & \forall i \in \mathcal{E}, \\
c_i(x^*) & \geq & 0, & \forall i \in \mathcal{I}, \\
\lambda_i^* & \geq & 0, & \forall i \in \mathcal{I}, \\
\lambda_i^* c_i(x^*) & = & 0, & \forall i \in \mathcal{E} \cup \mathcal{I}.
\end{array} \tag{16}
$$

Condições necessárias de primeira ordem

As condições (16) são conhecidas como condições Karush-Kuhn-Tucker, ou condições KKT.

Como a complementaridade implica que os multiplicadores de Lagrange correspondentes às restrições de desigualdade inativas são zero, podemos omitir os termos com índices $i \notin \mathcal{A}(x^*)$ na primeira equação das condições KKT.

Ou seja, ela pode ser escrita como

$$\boxed{0 = \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*).}$$

Um caso especial de complementaridade é importante e merece uma definição.

Definição 2 (Complementaridade estrita): *Dada uma solução local x^* de (1) e um vetor λ^* que satisfaz (16), dizemos que vale a complementaridade estrita se exatamente um dos λ_i^* e $c_i(x^*)$ é zero para cada índice $i \in \mathcal{I}$. Em outras palavras, temos que $\lambda^* > 0$ para todo $i \in \mathcal{I} \cap \mathcal{A}(x^*)$.*

Para um dado problema de programação não-linear (1) e uma solução x^*, vários vetores λ^* podem satisfazer as condições KKT (16).

No entanto, quando vale LICQ, o vetor λ^* ótimo é único.

Como exemplo, considere o problema

$$\boxed{\begin{array}{ll} \text{Minimizar} & f(x) = \left(x_1 - \frac{3}{2}\right)^2 + \left(x_2 - \frac{1}{2}\right)^4 \\ \text{Sujeita a} & c_1(x) = 1 - x_1 - x_2 \geq 0, \\ & c_2(x) = 1 - x_1 + x_2 \geq 0, \\ & c_3(x) = 1 + x_1 - x_2 \geq 0, \\ & c_4(x) = 1 + x_1 + x_2 \geq 0. \end{array}} \tag{17}$$

A solução para este problema é $x^* = (1, 0)^T$.

Condições necessárias de primeira ordem

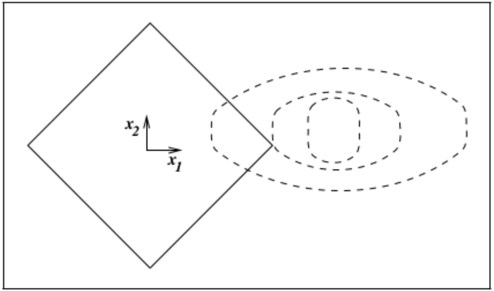

Figura: Exemplo de problema com restrições de desigualdade (Figura 12.8 de Numerical Optimization, de J. Nocedal e S. J. Wright)

Condições necessárias de primeira ordem

As duas primeiras restrições são ativas em x^*. Temos então que

$$\boxed{\nabla f(x^*) = \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.}$$

Portanto, as condições KKT são satisfeitas com

$$\boxed{\lambda^* = \left(\tfrac{3}{4}, \tfrac{1}{4}, 0, 0\right)^T.}$$

Condições de otimalidade de segunda ordem

Para definir as condições de otimalidade de segunda ordem, precisamos definir o conjunto $F_2(\lambda^*)$:

$$\boxed{w \in F_2(\lambda^*) \Leftrightarrow \begin{cases} \nabla c_i(x^*)^T w = 0, & \forall i \in \mathcal{E}, \\ \nabla c_i(x^*)^T w = 0, & \forall i \in \mathcal{A} \cap \mathcal{I} \text{ com } \lambda_i^* > 0, \\ \nabla c_i(x^*)^T w \geq 0, & \forall i \in \mathcal{A} \cap \mathcal{I} \text{ com } \lambda_i^* = 0, \end{cases}}$$

com x^* e λ^* que satisfazem as condições KKT (16).

Condições de otimalidade de segunda ordem

O conjunto $F_2(\lambda^*)$ contém as direções w que tendem a "aderir" às restrições de igualdade e às restrições de desigualdade ativas para as quais o multiplicador de Lagrange é positivo.

Com esta definição, podemos definir as condições necessárias e suficientes de segunda ordem para solução do problema (1).

Teorema 2 (Condições necessárias de segunda ordem): *Suponha que x^* seja uma solução local de (1) e que LICQ vale em x^*. Seja λ^* o vetor de multiplicadores de Lagrange tal que as condições KKT (16) são satisfeitas e seja $F_2(\lambda^*)$ definido como anteriormente. Então*

$$w^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0, \quad \forall w \in F_2(\lambda^*).$$

Teorema 3 (Condições suficientes de segunda ordem): *Suponha que, para um ponto viável $x^* \in \mathbb{R}^n$, exista um vetor de multiplicadores de Lagrange λ^* tal que as condições KKT (16) são satisfeitas. Suponha também que*

$$w^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) w > 0, \quad \forall w \in F_2(\lambda^*), \quad w \neq 0.$$

Então, x^* é uma solução local estrita.