Lista 3 - Otimização Não Linear Inteira Mista
COS886 - 2021/3

Amanda Ferreira de Azevedo - afazevedo@cos.ufrj.br

PESC/UFRJ — 28 de janeiro de 2022

Exercício 4.1

Considere o problema de otimização:

$$\begin{array}{ll} \min & f_0(x_1, x_2) \\ \text{s.a.} & 2x_1 + x_2 \geq 1 \\ & x_1 + 3x_2 \geq 1 \\ & x \geq 0 \end{array}$$

Faça um desenho e apresente a região viável e duas curvas de nível para a função objetivo do problema. Para cada função objetivo, determine o conjunto ótimo e o valor ótimo da função objetivo, a partir do gráfico.

1. $f_0(x_1, x_2) = x_1 + x_2$.
2. $f_0(x_1, x_2) = -x_1 - x_2$.
3. $f_0(x_1, x_2) = x_1$.
4. $f_0(x_1, x_2) = \max\{x_1, x_2\}$.
5. $f_0(x_1, x_2) = x_1^2 + 9x_2^2$.

O conjunto de restrições nos dá a região viável delimitada por $(0, \infty)$, $(0, 1)$, $(2/5, 1/5)$, $(1, 0)$, $(\infty, 0)$, como mostra a Figura 1.

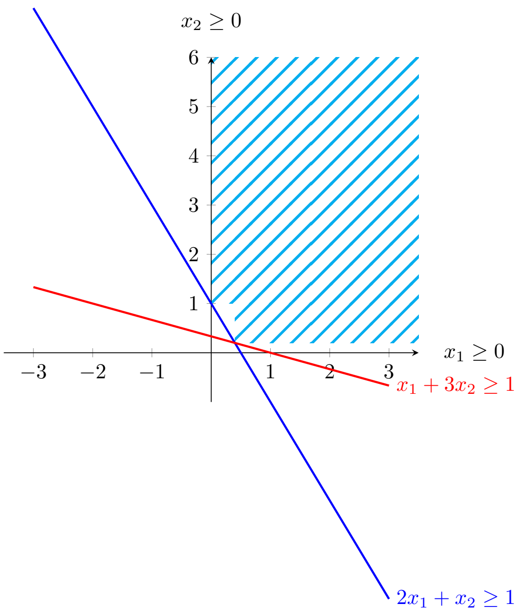
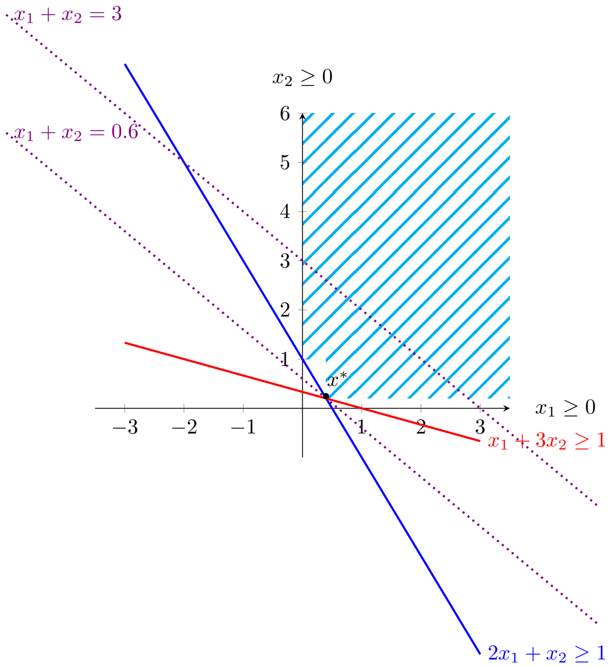


Figura 1: Região viável

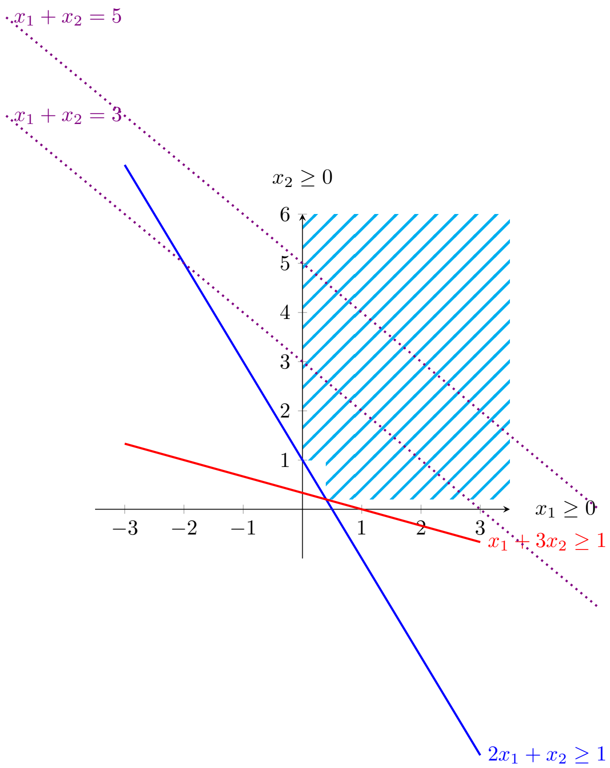
$$\boxed{f_0\left(x_1, x_2\right) = x_1 + x_2}$$

As curvas de nível são definidas pelo conjunto$\{x_1, x_2 \mid x_1 + x_2 = C\}$. Como o problema é de minimização, então a direção da solução ótima será oposta ao gradiente. Neste caso, $\nabla f(x) = (1, 1)$ e a curva de nível $x + y = 0.6$ bate exatamente no ponto $x^* = (2/5, 1/5)$, que é perpendicular a esta curva.



$$\boxed{f_0\left(x_1, x_2\right) = -x_1 - x_2.}$$

As curvas de nível são definidas pelo conjunto$\{x_1, x_2 \mid x_1 + x_2 = C\}$. Como o problema é de minimização, então a direção da solução ótima será oposta ao gradiente. Neste caso, $\nabla f(x) = (-1, -1)$ e as curvas de nível não encontrarão nenhum ponto, pois a região é ilimitada para este lado.



$$\boxed{f_0\left(x_1, x_2\right) = x_1.}$$

Observando a Figura **??**, basta encontrarmos o menor valor que x_1 pode assumir na região viável. Esta solução é atingida exatamente quando $x_1 = 0$ e $x_2 \geq 1$, uma vez que a função objetivo independe do valor de x_2.

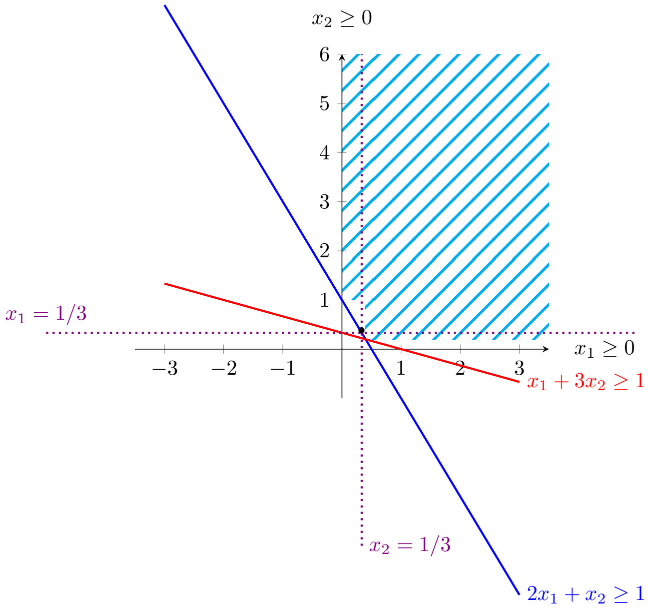
$$\boxed{f_0\left(x_1, x_2\right) = \max\left\{x_1, x_2\right\}.}$$

Suponha que $x_1 > x_2$. Teremos a mesma função objetivo do item anterior. No entanto, não podemos tomar $x_1 = 0$, uma vez que $x_2 \geq 1/5$ e $x_1 > x_2$. Dessa forma, $x_1 > 1/5$.

De forma análoga, se tomarmos $x_2 > x_1$, vamos minimizar x_2 e $x_2 > 2/5$, pois $x_1 \geq 2/5$.

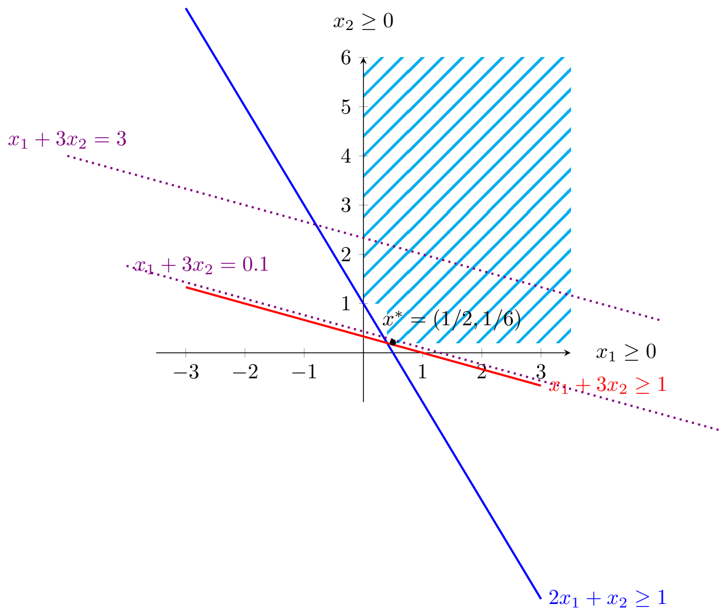
Se $x_1 = x_2$, o menor valor é atingido no ponto $x^* = (1/3, 1/3)$.

Dessa forma, o menor valor que a função objetivo pode atingir é exatamente no caso em que $x_1 = x_2$.



$$\boxed{f_0(x_1, x_2) = x_1^2 + 9x_2^2.}$$

Observe que podemos trocar a função objetivo por $x_1 + 3x_2$, uma vez que f_0 é exatamente isto elevado ao quadrado. Podemos fazer isso para calcular as curvas de nível e o gradiente. Derivando, temos que $\nabla f(x) = (2, 6)$.



Indo na direção contrária ao gradiente, perpendicular as curvas de nível, chegamos ao ponto $x^* = (1/2, 1/7)$.

Exercício 4.8

Apresente uma solução explicita para cada um dos PPLs

1. Minimizando a função linear em um conjunto afim.

$$\begin{aligned} &\min c^\top x \\ &\text{s.a.} \quad Ax = b \end{aligned}$$

2. Minimizando a função linear em um semi-espaço.

$$\begin{array}{ll} \min & c^\top x \\ \text{s.a.} & a^\top x \leq b \end{array}$$

onde $a \neq 0$.

3. Minimizando a função linear em um retângulo.

$$\begin{array}{ll} \min & c^\top x, \\ \text{s.a.} & l \preceq x \preceq u, \end{array}$$

onde l e u satisfazem $l \succeq u$.

4. Minimizando a função linear em um simplex probabilístico, definido com o conjunto de vetores $\left\{x \in \mathbb{R}^n : \mathbf{1}^\top x = 1, x \succeq 0\right\}$.

$$\begin{array}{ll} \min & c^\top x \\ \text{s.a.} & \sum_{i=1}^n x_i = 1 \\ & x \geq 0 \end{array}$$

Depois de resolver o problema, diga o que ocorre se a restrição de igualdade for substituída por uma inequação $\sum_{i=1}^n x_i \leq 1$.

5. Minimizando a função linear em uma caixa unitária com restrição de despesa total.

$$\begin{array}{ll} \min & c^\top x, \\ \text{s.a.} & \sum_{i=1}^n x_i = \alpha, \\ & 0 \preceq x \preceq 1, \end{array}$$

onde α é um inteiro entre 0 e n. Depois de resolver o problema, diga o que ocorre se a restrição de igualdade for substituída por uma inequação $\sum_{i=1}^n x_i \leq \alpha$.

1. Minimizando a função linear em um conjunto afim.

Podemos dividir em alguns casos:
1. Se $Ax = b$ não tiver solução, então o problema é **inviável**. Ou seja, a solução ótima é ∞.

2. Se c é ortogonal ao espaço nulo de A, então c é ortogonal ao hiperplano $Ax = b$ (basta transladar). Note que, como o núcleo de A é um subespaço vetorial, existe seu complemento ortogonal que será exatamente a imagem de $A^\top$. Em outras palavras, c pertence a $Im(A^\top)$. Dessa forma, existe um vetor λ tal que

$$c = A^\top \lambda$$

Nessa situação, qualquer solução satisfazendo $Ax = b$ resolve o problema uma vez que você não pode se mover por $Ax = b$ de forma a decrescer $c^T x$. Dessa forma, a solução ótima dada pela função objetivo será:

$$c^T x = \lambda^\top A x = \lambda^\top b$$

3. Se c não é ortogonal ao espaço nulo de A então a projeção de c no espaço nulo de A é uma projeção não-trivial e portanto, seguindo a direção contrária da projeção a função objetivo pode decrescer o tanto quanto for possível.

Dessa forma, a solução pode ser dividida em

$$p^* = \begin{cases} \infty & \text{if } Ax = b \text{ quando não há solução} \\ \lambda^\top b & c = A^\top \lambda \text{ para algum } \lambda \\ -\infty & \text{if } Ax = b \text{ caso contrário} \end{cases}$$

$\boxed{\text{2. Minimizando a função linear em um semi-espaço.}}$

Neste caso, esse problema é sempre viável. Novamente, suponha que c seja ortogonal ao espaço nulo. Dessa forma, podemos decompor c com uma componente do espaço nulo somada a uma componente da imagem de A^T.

$$c = a\lambda + \hat{c}$$

Onde $a^T\hat{c} = 0$. Vamos dividir em alguns casos

Se $\lambda > 0$, então podemos escolher um $x = -ta$, com $t \geq 0$ fazendo $t \to \infty$, temos

$$c^T x = -tc^T a = -t\lambda a^T a \to -\infty$$

Além disso,

$$a^T x - b = -ta^T a - b \leq 0$$

Logo, o problema é ilimitado inferiormente.

Se $\hat{c} \neq 0$, o problema também será ilimitado inferiormente. Basta escolher $x =$ e fazer $t \to \infty$.

Se $c = \lambda a + \hat{c}$ e $x = ba - t\hat{c}$, então

$$\begin{aligned} c^T x &= (\lambda a + \hat{c})^T (ba - t\hat{c}) \\ &= \lambda ba^T a + bc^T a - \lambda ta^T \hat{c} - t\hat{c}^T \hat{c} \\ &= \lambda ba^T a + bc^T a - t\hat{c}^T \hat{c} \end{aligned}$$

Como $\lambda ba^T a + bc^T a$ é constante, temos que $c^T x \to -\infty$ quando $t \to \infty$.

Se $\hat{c} = 0$, então $c = a\lambda$ para algum $\lambda \leq 0$. Logo,

$$c^T x = \lambda a^T x \underbrace{\geq \lambda b}_{a^T x \leq b}$$

Logo, $c^T x = \lambda b$. A solução final fica

$$p^\star = \begin{cases} \lambda b & c = a\lambda \text{ para algum } \lambda \leq 0 \\ -\infty & \text{caso contrário} \end{cases}$$

$\boxed{\text{3. Minimizando a função linear em um retângulo.}}$

Pela estrutura do problema, podemos **separá-los**, uma vez que a função objetivo é a soma de termos $c_i x_i$ e as restrições dependem apenas de uma única variável. Podemos resolver o problema minimizando cada componente de x independentemente. Ou seja, vamos resolver o seguinte problema

$$\begin{aligned} \min \quad & c_i x_i \\ \text{s.a} \quad & l_i \leq x_i \leq u_i \end{aligned}$$

Observe que:

- Se $c_i > 0$, então a solução ótima será $x_i^* = l_i$
- Se $c_i < 0$, então $x_i^* = u_i$
- se $c_i = 0$, então qualquer x_i no intervalo fechado $[l_i, u_i]$ é ótimo

Dessa forma, a solução ótima do problema original pode ser escrita como

$$p^* = l^T \max\{c, 0\} + u^T \max\{-c, 0\}$$

4. Minimizando a função linear em um simplex probabilístico, definido com o conjunto de vetores

Vamos ordenar as componentes do vetor c de forma crescente.

$$c_1 = c_2 = c_k < c_{k+1} \leq \cdots \leq c_n$$

Como c_1 é a componente de menor valor, podemos escrever

$$\sum_{i=1}^{n} c_i x_i \geq \underbrace{c_1 \sum_{i=1}^{n} x_i}_{z}$$

Que será o menor valor que $c^T x$ poderá assumir, com

$$x_1 + \cdots + x_k = 1$$

Uma vez que k componentes possuem o mesmo valor e o resto das componentes

$$x_{k+1} = \cdots = x_n = 0$$

Donde concluímos que $p^* = z$.

Para a versão de *budget*, como não temos necessariamente uma igualdade para garantir o investimento e queremos minimizar $c^T x$, bastará priorizar as variáveis associadas aos coeficientes c negativos. Em outras palavras, $p^* = \min\{0, z\}$.

5. Minimizando a função linear em uma caixa unitária com restrição de despesa total.

Novamente, vamos ordenar de forma crescente os coeficientes c.

$$c_1 \leq \cdots \leq c_{i-1} < c_i = \cdots = c_\alpha = \cdots = c_k < c_{k+1} \leq \cdots \leq c_n$$

A solução ótima será exatamente a soma dos α menores valores de c, ou seja

$$c_1 + c_2 + \cdots + c_\alpha$$

Com

$$x_1 = \cdots = x_{i-1} = 1$$

E, considerando que há alguns elementos iguais ao elemento na posição α, vamos fazer

$$x_i + \cdots + x_k = \alpha - i + 1$$

Para considerar **exatamente** os α variáveis associadas ao coeficientes de menor custo. E com o resto, zeramos

$$x_{k+1} = \cdots = x_n = 0$$

Para o segundo caso, basta tomarmos o somatório dos valores não positivos de c. Em outras palavras, faremos $x = 1$ toda vez que $c_i < 0$, não ultrapassando α e 0 caso contrário.

Exercício 4.17

Considere o problema de otimização

$$\begin{array}{ll} \max & \sum_{j=1}^{n} r_j\left(x_j\right), \\ \text{s.a.} & Ax \preceq c^{\max}, \\ & x \succeq 0, \end{array}$$

2 onde $x \in \mathbb{R}^n$ é a variável, $A \in \mathbb{R}^{m\times n}$ e $c^{\max} \in \mathbb{R}^m$ são dados, e $r_j\left(x_j\right)$ é uma função côncava linear por partes, dada por

$$r_j\left(x_j\right) = \begin{cases} p_j x_j & 0 \leq x_j \leq q_j \\ p_j q_j + p_j^{disc}\left(x_j - q_j\right) & x_j \geq q_j \end{cases}$$

onde $p, q \succeq 0$ e $0 < p_j^{\text{disc}} < p_j$ para todo $j = 1, \ldots, n$. Transforme o modelo em um problema de programação linear, apresente os passos detalhadamente.

Podemos escrever a função de receita como

$$r_j(x_j) = min_j\{p_j x_j, p_j q_j + p_j^{disc}(x_j - q_j)\}$$

Para vermos essa equivalência, considere

$$p_j x_j < p_j q_j + p_j^{disc}(x_j - q_j) \Leftrightarrow p_j(x_j - q_j) < p_j^{disc}(x_j - q_j)$$

Como $p_j > p_j^{disc}$, isso só será verdade se $x_j < q_j$. Por outro lado,

$$p_j x_j > p_j q_j + p_j^{disc}(x_j - q_j) \Leftrightarrow p_j(x_j - q_j) > p_j^{disc}(x_j - q_j)$$

Apenas se $x_j > q_j$.

Dessa maneira, podemos definir uma nova variável u_j que será um limite inferior para $r_j(x_j)$, de forma que podemos adicionar as seguintes restrições ao problema:

$$p_j x_j \geq u_j$$
$$p_j q_j + p_j^{disc}(x_j - q_j) \geq u_j$$

Como no problema original estamos querendo maximizar a receita total, podemos reescrever o problema da seguinte maneira

$$\max \quad \sum_{j=1}^{n} u_j \tag{1}$$
$$\text{s.a.} \quad x \succeq 0 \tag{2}$$
$$Ax \preceq c^{max} \quad \textit{resource limits} \tag{3}$$
$$p_j x_j \geq u_j l \tag{4}$$
$$p_j q_j + p_j^{disc}(x_j - q_j) \geq u_j \tag{5}$$

Em outras palavras, podemos crescer u_j até no máximo $p_j x_j$ ou $p_j q_j + p_j^{disc}(x_j - q_j) \geq u_j$, dependendo do valor de x_j.

Agora, temos um problema de programação linear com variáveis x e u. Para mostrar que esse novo problema é equivalente ao original, vamos fixar uma solução x.

As últimas duas restrições (4-5) garantem que $u_j \leq r_j(x_j)$, então pra toda solução viável nesse LP, a função objetivo nos dá um valor menor ou igual a receita total. Por outro lado, sempre podemos pegar, sem perda de generalidade, $u_j = r_j(x_j)$ que é exatamente o caso em que as duas funções se igualam.

Exercício 4.11

Formule os problemas a seguir como problemas de programação linear. Explique detalhadamente a relação entre a solução ótima de cada problema e a solução do seu LP equivalente. Considere $A \in \mathbb{R}^{m\times n}$ e $b \in \mathbb{R}^m$ dados do problema.

1. Minimizar $\|Ax - b\|_\infty$
2. Minimizar $\|Ax - b\|_1$
3. Minimizar $\|Ax - b\|_1$ sujeito a $\|x\|_\infty \leq 1$
4. Minimizar $\|x\|_1$ sujeito a $\|Ax - b\|_\infty \leq 1$
5. Minimizar $\|Ax - b\|_1 + \|x\|_\infty$

$$\boxed{\text{Minimizar } \|Ax - b\|_\infty}$$

Por definição, temos que $\|Ax - b\|_\infty = \max_i |Ax_i - b|$. Com isso, podemos criar uma nova variável u que será um limitante superior para $Ax - b$, ao passo que minimizamos u, descobrindo assim o maior valor para $Ax - b$. No entanto, como estamos lidando com módulo, também será necessário que $Ax - b$ esteja limitando inferiormente pelo simétrico de u.

Formalmente, seja $u \in \mathbb{R}$, o novo LP pode ser reescrito como

$$\begin{aligned} \min \quad & u && (6)\\ \text{s.a} \quad & Ax - b \preceq u && (7)\\ & Ax - b \succeq -u && (8) \end{aligned}$$

Agora, basta provarmos a equivalência. Para um x fixado as restrições 7 e 8 podem ser escritas, para algum k,

$$\begin{aligned} -u \leq a_k^\top x - b_k \leq u \Leftrightarrow u &\geq \left|a_k^\top x - b_k\right| \\ \Leftrightarrow u &\geq \max_k \left|a_k^\top x - b_k\right| \\ \Leftrightarrow u &\geq \|Ax - b\|_\infty \end{aligned}$$

Fazendo para todo x, chegamos no resultado minimizando u.

$$\boxed{\text{Minimizar } \|Ax - b\|_1}$$

Usaremos a mesma ideia, no entanto, como não estamos olhando para o valor máximo, a dimensão de u deverá ser modificada. Considere $u \in \mathbb{R}^m$, podemos reescrever como seguinte LP:

$$\begin{aligned} \min \quad & \sum_{i=1}^{m} u_i \\ \text{s.a} \quad & Ax - b \preceq u \\ & Ax - b \succeq -u \end{aligned}$$

Agora, provemos a equivalência. Para um x fixado arbitrário, as restrições adicionais podem ser escritas como

$$-u_k \leq a_k^T x - b_k \leq u_k \Leftrightarrow u_k \geq \left|a_k^\top x - b_k\right|$$

Observe que atingimos o ótimo com relação a u escolhendo exatamente $u_k = \left|a_k^\top x - b_k\right| = \|Ax - b\|_1$. Fazendo para todo x, chegamos no resultado.

$$\boxed{\text{Minimizar } \|Ax-b\|_1 \text{ sujeito a } \|x\|_\infty \leq 1}$$

Equivalente ao que fizemos anteriormente,

$$\begin{aligned}
\min \quad & \sum_{i=1}^{m} u_i \\
\text{s.a} \quad & Ax - b \preceq u \\
& Ax - b \succeq -u \\
& x \leq 1 \\
& x \geq -1
\end{aligned}$$

$$\boxed{\text{Minimizar } \|x\|_1 \text{ sujeito a } \|Ax-b\|_\infty \leq 1}$$

Equivalente ao que fizemos anteriormente,

$$\begin{aligned}
\min \quad & \sum_{i=1}^{m} u_i \\
\text{s.a} \quad & Ax - b \preceq 1 \\
& Ax - b \succeq -1 \\
& x \leq u \\
& x \geq -u
\end{aligned}$$

$$\boxed{\text{Minimizar } \|Ax-b\|_1 + \|x\|_\infty}$$

Para esta, vamos definir uma variável para cada componente da soma.

$$\begin{aligned}
\min \quad & \sum_{i=1}^{m} u_i + t \\
\text{s.a} \quad & Ax - b \preceq u \\
& Ax - b \succeq -u \\
& x \leq \sum_{i=1}^{m} t \\
& x \geq \sum_{i=1}^{m} -t
\end{aligned}$$

Exercicio adicional 4.1

Encontre condições necessárias e suficientes para $x \in \mathbb{R}^n$ minimizar a função convexa e diferenciável f no conjunto $\{x \in \mathbb{R}^n \mid \sum_{i=1}^n x = 1, x \succeq 0\}$.

Considere o conjunto

$$R = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x = 1, x \succeq 0\}$$

Pelas condições de otimalidade sabemos que se x é viável então $\nabla f(x)^\top (y - x) \geq 0$ para todo y viável.

Afirmação:

$$\nabla f(x)^\top (y - x) \geq 0 \Rightarrow \min_{i=1,\ldots,n} \nabla f(x)_i \geq \nabla f(x)^\top x$$

Demonstração: Por hipótese, temos que vale $\nabla f(x)^\top (y - x) \geq 0$ para todo y viável, em particular, tome $y = e_i$, onde e_i é o vetor canônico. Dessa forma, temos

$$\begin{aligned} \nabla f(x)^\top (y - x) \geq 0 &\Leftrightarrow \nabla f(x)_i^\top \underbrace{y}_{e_i} - \nabla f(x)^\top x \geq 0 \\ &\Leftrightarrow \nabla f(x)_i^\top \geq \nabla f(x)^\top x \end{aligned}$$

■

Afirmação:

$$\min_{i=1,\ldots,n} \nabla f(x)_i \geq \nabla f(x)^\top \Rightarrow \nabla f(x)^\top (y - x)$$

Demonstração: Seja y um ponto viável, temos por hipótese que para todo $i = 1, \ldots, n$ vale

$$\nabla f(x)_i \geq \nabla f(x)^\top x$$

Multiplicando ambos os lados por y_i e somando, temos

$$\underbrace{\sum_{i=1}^n y_i \nabla f(x)_i}_{y^T \nabla f(x)} \geq \underbrace{\sum_{i=1}^n y_i}_{=1} \nabla f(x)^\top x = \nabla f(x)^T x$$

Logo

$$y^T \nabla f(x) \geq \nabla f(x)^T x \Leftrightarrow \nabla f(x)^T (y - x) \geq 0$$

Reescrevendo,

$$\min_{i=1,\ldots,n} \nabla f(x)_i \geq \nabla f(x)^T x \Leftrightarrow \min_{i=1,\ldots,n} \frac{\partial f}{\partial x_i} \geq \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}$$

Como $\sum_{i=1}^n x = 1$ e $x \geq 0$, temos

$$\min_{i=1,\ldots,n} \frac{\partial f}{\partial x_i} \leq \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}$$

O que nos dá que

$$\min_{i=1,\ldots,n} \frac{\partial f}{\partial x_i} = \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}$$

Isso só é possível se quando $\frac{\partial f}{\partial x_k} > \min_i \frac{\partial f}{\partial x_i}$, para algum k, $x_k = 0$. Além disso, se $x_k > 0$ então $\frac{\partial f}{\partial x_k} = \min_{i=1,\ldots,n} \frac{\partial f}{\partial x_i}$.

■

Exercício adicional 5.1

Considere a matriz $X = X^\top \in \mathbb{R}^{m+n\times m+n}$ particionada como

$$X = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$$

onde $A \in \mathbb{R}^{n\times n}$. Se $\det(A) \neq 0$, a matriz $S = C - B^\top A^{-1}B$ é chamada de complemento de Schur de A em X.

1. O complemento de Schur aparece quando se minimiza uma função na forma quadrática em alguma das variáveis. Seja

$$f(u,v) = \left[u^\top v^\top\right] X \left[u^\top v^\top\right]^\top$$

onde $u \in \mathbb{R}^n$. Seja $g(v)$ o menor valor de f em u, ou seja, $g(v) = \inf_u f(u,v)$. Veja que $g(v)$ pode ser $-\infty$. Mostre que se $A \succ 0$, temos $g(v) = v^\top Sv$.

2. O complemento de Schur aparece em diversas caracterizações de matrizes semi-definidas positivas e definidas positivas de uma matriz em blocos. Mostre que:

(a) $X \succ 0$ se e somente se $A \succ 0$ e $S \succ 0$.

(b) Se $A \succ 0$, então $X \succeq 0$ se e somente se $S \succeq 0$.

(c) $X \succeq 0$ se e somente se $A \succeq 0, B^\top \left(I - AA^\dagger\right) = 0$ e $C - B^\top A^\dagger B \succeq 0$, onde $A^\dagger$ é a pseudo-inversa de Moore-Penrose de $A \cdot \left(C - B^\top A^\dagger B\right.$ serve como generalização do complemento de Schur no caso em que A é singular e semi-definida positiva).

Afirmação: Se $A \succ 0$, temos $g(v) = v^\top Sv$.

Temos que

$$f(u,v) = u^TAu + 2v^TBu + v^TCv$$

Por hipótese, $A \succ 0$, então podemos minimizar f com relação a u calculando o gradiente em relação a u e igualando a zero.

Pela definição, temos que

$$\alpha = u^TAu = \sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}u_iu_j$$

Derivando α com respeito ao k-ésimo elemento de u, temos

$$\frac{\partial\alpha}{\partial u_k} = \sum_{j=1}^{n} a_{kj}u_j + \sum_{i=1}^{n} a_{ik}u_i$$

Fazendo para todo $k = 1, 2, \ldots, n$, temos

$$\frac{\partial\alpha}{\partial u} = u^TA^T + u^TA = u^T(A^T + A)$$

Como as matrizes são simétricas então

$$\frac{\partial\alpha}{\partial u} = 2u^TA$$

Agora, derivando $\beta = 2v^TBu$ com respeito a u, temos que

$$\frac{\partial\beta}{\partial u} = 2v^TB$$

Logo, a derivada de f com respeito a u será

$$\frac{\partial f}{\partial u} = 2u^T A + 2v^T B$$

Igualando a zero, temos que

$$\begin{aligned} 2u^T A + 2v^T B &= 0 \\ 2u^T A &= -2v^T B \\ u^T A &= -v^T B \\ u^T &= -v^T B A^{-1} \\ u &= -A^{-1} B v \end{aligned}$$

Agora, substituindo em f, temos

$$f(u^*, v) = g(v) = -(A^{-1}Bv)^T A(-A^{-1}Bv) + 2v^T B(-A^{-1}Bv) + v^T Cv$$

Arrumando, temos

$$v^T B^T A^{-1} \underbrace{AA^{-1}}_{=1} Bv - 2v^T BA^{-1}Bv + v^T Cv$$

Botando v^T e v em evidência, temos

$$v^T(B^T A^{-1} B - 2 \underbrace{B}_{B=B^T} A^{-1}Bv + C)v = v^T(C - B^T A^{-1} B)v = v^T S v$$

Afirmação: $(\Rightarrow)$ Se $X \succ 0$ então $A \succ 0$ e $S \succ 0$.

Demonstração: Como $X \succ 0$ então $f(u, v) > 0$ para todo vetor não nulo (u, v). Em particular, se pegarmos $f(u, 0) = u^T Au > 0$ para todo vetor não nulo u. Dessa forma, $A \succ 0$.

Pela questão anterior, podemos usar o ponto $u^* = -A^{-1}Bv$ para analisar o que acontece com S. Com isso, $f(u^*, v) = v^T(C - B^T A^{-1}B)v > 0$, dessa forma, $S = C - B^T A^{-1} B \succ 0$ ■

Afirmação: $(\Leftarrow)$ Se $A \succ 0$ e $S \succ 0$ então $X \succ 0$

Demonstração: Segue da questão anterior que

$$f(u, v) \geq f(u^*, v) = v^T S v > 0 \quad \because S \succ 0$$

Se $v \neq 0$. Por outro lado, se $v = 0$ e $u \neq 0$

$$f(u, 0) = u^T Au > 0 \quad \because A \succ 0$$

Logo, $f(u, v) > 0$ e consequentemente, $X \succ 0$. ■

Afirmação: $(\Rightarrow)$ Se $A \succ 0$ (logo $X \succeq 0$) então $S \succeq 0$.

Demonstração: Por hipótese, $X \succ 0$, isto é, $f(u, v) \geq 0$ para todo u, v. Em particular, valerá para $f(u^*, v)$ que é exatamente $v^T Sv$. Dessa forma, $S \succ 0$. ■

Afirmação: $(\Leftarrow)$ $S \succeq 0$ então $A \succ 0$ (logo $X \succeq 0$).

Demonstração: Pelo item anterior, sabemos que se $A \succ 0$ então $f(u^*, v) = v^T Sv$. Como por hipótese $S \succ 0$ então

$$f(u, v) \geq f(u^*, v) = v^T Sv \geq 0$$

Para todo u, v. Dessa forma, $X \succ 0$ e por consequência, $A \succ 0$. ■

Afirmação: $X \succeq 0$ se e somente se $A \succeq 0, B^\top \left(I - AA^\dagger\right) = 0$ e $C - B^\top A^\dagger B \succeq 0$

Demonstração:

Vamos supor que $A \in \mathbb{R}^{k\times k}$ tem $rank(A) = r$. Dessa forma, existem matrizes $Q_1 \in \mathbb{R}^{k\times r}$ e $Q_2 \in \mathbb{R}^{k\times(k-r)}$ e uma matriz diagonal invetível $\Lambda \in \mathbb{R}^{r\times r}$ de tal forma que podemos decompor A da seguinte maneira

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^T$$

Onde $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^T \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} = I \qquad \because$ ortogonalidade.
Essa é a decomposição espectral de A.

Considere a seguinte matriz

$$M = \begin{bmatrix} \begin{matrix} Q_1 \\ \underbrace{\mathbf{0}}_{(n-k)\times r} \end{matrix} & \begin{matrix} Q_2 \\ \underbrace{\mathbf{0}}_{(n-k)\times(k-r)} \end{matrix} & \begin{matrix} \overbrace{\mathbf{0}}^{k\times(n-k)} \\ \underbrace{I}_{(n-k)\times(n-k)} \end{matrix} \end{bmatrix} \in \mathbb{R}^{n\times n}$$

Que é **não-singular** uma vez que

$$\begin{bmatrix} Q_1 & Q_2 & 0 \\ 0 & 0 & I \end{bmatrix} \cdot \begin{bmatrix} Q_1 & 0 \\ Q_2 & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = I$$

Agora, como X é semi-definida positiva, vale

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \Leftrightarrow \begin{bmatrix} Q_1 & Q_2 & 0 \\ 0 & 0 & I \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 & 0 \\ 0 & 0 & I \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} Q_1 & Q_2 & 0 \\ 0 & 0 & I \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} = \begin{bmatrix} [Q_1 Q_2]^T A & [Q_1 Q_2]^T B \\ B^T & C \end{bmatrix} =$$

Daí

$$\begin{bmatrix} [Q_1 Q_2]^T A & [Q_1 Q_2]^T B \\ B^T & C \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 & 0 \\ 0 & 0 & I \end{bmatrix} = \begin{bmatrix} [Q_1 Q_2]^T A [Q_1 Q_2] & [Q_1 Q_2]^T B \\ B^T [Q_1 Q_2] & C \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} \Lambda & 0 & Q_1^T B \\ 0 & 0 & Q_2^T B \\ B^T Q_1 & B^T Q_2 & C \end{bmatrix} \succeq 0$$

$$\Leftrightarrow Q_2^T B = 0, \begin{bmatrix} \Lambda & Q_1^T B \\ B^T Q_1 & C \end{bmatrix} \succeq 0.$$

Temos que $\Lambda \succ 0$ se e somente se $A \succeq 0$.

É verdade que

$$A^\dagger = Q_1 \Lambda^{-1} Q_1^T, \quad I - AA^\dagger = Q_2 Q_2^T.$$

Dessa forma,

$$Q_2^T B = 0 \Leftrightarrow Q_2 Q_2^T B = \left(I - A^\dagger A\right) B = 0$$

Além disso, como Λ é invertível,

$$\begin{bmatrix} \Lambda & Q_1^T B \\ B^T Q_1 & C \end{bmatrix} \succeq 0 \Leftrightarrow \Lambda \succ 0, \quad C - B^T Q_1 \Lambda^{-1} Q_1^T B = C - B^T A^\dagger B \succeq 0.$$

■

Exercício 1.1

Considere a matriz $X = X^\top \in \mathbb{R}^{m+n \times m+n}$ particionada como

$$X = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$$

onde $A \in \mathbb{R}^{n\times n}$. Se $\det(A) \neq 0$, a matriz $S = C - B^\top A^{-1} B$ é chamada de complemento de Schur de A em X. 1. Reconhecer complementos de Schur normalmente ajuda a representar restrições convexas não-lineares como inequações lineares de matrizes (LMI). Considere a função

$$f(x) = (Ax+b)^\top \left(P_0 + x_1 P_1 + \ldots + x_n P_n\right)^{-1} (Ax+b),$$

onde $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ e $P_i = P_i^\top \in \mathbb{R}^{m\times m}$, com o domínio

$$\text{dom } f = \{x \in \mathbb{R}^n \mid P_0 + x_1 P_1 + \cdots + x_n P_n \succ 0\}.$$

Essa é a composição da função fracionária da matriz e um mapeamento afim, e então é uma função convexa. Dada uma representação LMI de epi f, encontre a matriz simétrica $F(x,t)$ afim em (x,t), tal que

$$x \in \operatorname{dom} f, \quad f(x) \leq t \Longleftrightarrow F(x,t) \succeq 0.$$

O *epigrafo* de f é definida como

$$\mathbf{epi} f = \{(x, P, t) \mid P \succ 0, x^T P^{-1} x \leq t\}$$

Onde $P = P_0 + x_1 P_1 + \cdots + x_n P_n$. Da outra questão, provamos que

Se $A \succ 0$, então $X \succeq 0$ se e somente se $S \succeq 0$.

Como $P \succ 0$, podemos definir X como

$$X = \begin{bmatrix} P & Ax+b \\ (Ax+b)^T & t \end{bmatrix} \succ 0$$

Que é uma matriz simétrica afim em (x,t).

Exercício 1.1

Formule o problema de otimização como problema de programação semidefinida. A variável é $x \in \mathbb{R}^n$ e $F(x)$ é definida como

$$F(x) := F_0 + x_1F_1 + x_2F_2 + \cdots + x_nF_n,$$

$\text{com}\, F_i \in \mathbb{S}^m$, para $i = 1, \ldots, n$. O domínio de f em cada subproblema é dado por dom $f = \{x \in \mathbb{R}^n \mid F(x) \succ 0\}$.

1. Minimizar $f(x) = c^\top F(x)^{-1}c$, onde $c \in \mathbb{R}^m$.
2. Minimizar $f(x) = \max_{i=1,\ldots,K} c_i^\top F(x)^{-1}c_i$, onde $c_i \in \mathbb{R}^m, i = 1, \ldots, K$.
3. Minimizar $f(x) = \sup_{\|c\|_2 \leq 1} c^\top F(x)^{-1}c$.

Considere a matriz $X = X^\top \in \mathbb{R}^{m+n\times m+n}$ particionada como

$$X = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$$

onde $A \in \mathbb{R}^{n\times n}$. Se $\det(A) \neq 0$, a matriz $S = C - B^\top A^{-1}B$ é chamada de complemento de Schur de A em X.

Minimizar $f(x) = c^\top F(x)^{-1}c$, onde $c \in \mathbb{R}^m$.

Considere a seguinte matriz

$$X = \begin{bmatrix} F(x) & c \\ c^T & t \end{bmatrix} \succeq 0$$

Como $F(x) \succ 0$, pela demonstração da questão anterior, $X \succeq 0$ e então $S \succeq 0$, onde S é o complemento de Schur de A em X. Daí, segue que

$$S = t - c^tF^{-1}(x)c \succ 0 \Leftrightarrow t \succ c^tF^{-1}(x)c$$

Dessa forma, podemos reescrever o problema como

$$\begin{array}{ll} \text{min} & t \\ \text{s.a.} & \begin{bmatrix} F(x) & c \\ c^T & t \end{bmatrix} \succeq 0. \end{array}$$

$\boxed{\text{Minimizar } f(x) = \max_{i=1,\dots,K} c_i^\top F(x)^{-1} c_i, \text{ onde } c_i \in \mathbb{R}^m, i = 1, \dots, K.}$

Nesse caso, podemos usar a mesma ideia do item anterior, com a diferença que podemos definir K matrizes X, uma para cada i. Depois disso, podemos adicioná-las como K restrições no nosso problema, de forma que encontraremos o menor t que faz com que todas as matrizes definidas independentemente sejam menor do que t.

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \begin{bmatrix} F(x) & c_i \\ c_i^T & t \end{bmatrix} \succeq 0, \quad i = 1, \dots, K. \end{array}$$

$$t - c_i^t F^{-1}(x) c_i \succ 0 \Leftrightarrow t \succ c_i^t F^{-1}(x) c_i \quad \forall i = 1, \dots, K$$

$$\boxed{\text{Minimizar } f(x) = \sup_{\|c\|_2 \leq 1} c^\top F(x)^{-1} c.}$$

Segue da definição de autovalores e autovetores que

$$F^{-1}(x)c = \lambda c \Leftrightarrow \sup_{\|c\|=1} c^T F^{-1}(x)c = \lambda_{max}(F^{-1}(x))$$

Onde c são os autovetores associados aos autovalores λ. Dessa forma, considere a seguinte matriz

$$X = \begin{bmatrix} F(x) & I \\ I & tI \end{bmatrix} \succeq 0.$$

Como $F(x) \succ 0$, pela demonstração da questão anterior, $X \succeq 0$ e então $S \succeq 0$, onde S é o complemento de Schur de A em X. Daí, segue que

$$S = tI - F^{-1}(x) \succ 0 \Leftrightarrow tc^T Ic - c^T F^{-1}(x)v \succ 0 \Leftrightarrow t \succ c^T F^{-1}(x)c$$

Dessa forma, podemos reescrever o problema como

$$\begin{array}{ll} \text{min} & t \\ \text{s.a.} & \begin{bmatrix} F(x) & I \\ I & tI \end{bmatrix} \succeq 0. \end{array}$$