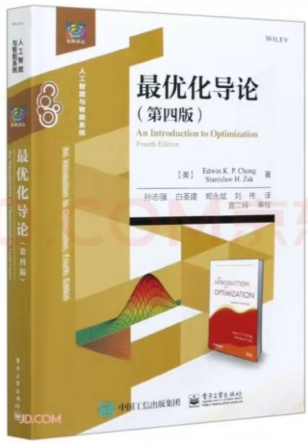
M05M11084 最优化理论、算法与应用

6-4 凸优化问题

讲义和程序下载
（随课程进度更新）

链接：
https://pan.baidu.com/s/1NynYva56GiPsj2gLI59k0Q?pwd=yuan
提取码：yuan



凸优化问题

参考：

1.最优化导论，第22章，，Edwin K.P.Chong, Stanislaw H. Żak著，孙志强等译

1. 引言
2. 凸函数
3. 凸优化问题
4. 半定规划

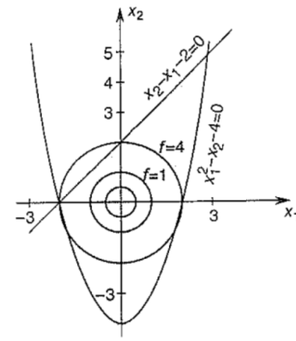
例

例22.1 考虑 $\min x_1^2 + x_2^2$

$$\text{s.t.}\ x_2 - x_1 - 2 \le 0$$
$$x_1^2 - x_2 - 4 \le 0$$

极小点处，所有约束都不起作用

无约束优化问题

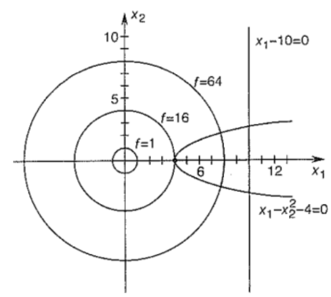


例22.2 考虑 $\min x_1^2 + x_2^2$

$$\text{s.t.}\ x_1 - 10 \le 0$$
$$x_1 - x_2^2 - 4 \ge 0$$

极小点处，只有1个约束起作用

等式约束优化问题



例22.3 考虑 $\min x_1^2 + x_2^2$

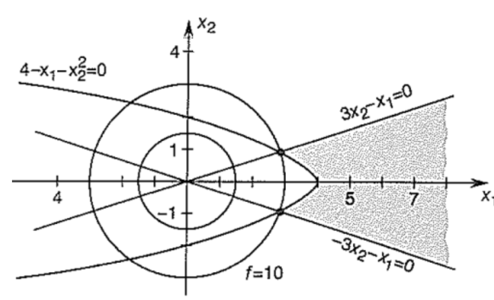
$$\text{s.t.}\ 4 - x_1 - x_2^2 \le 0$$
$$3x_2 - x_1 \le 0$$
$$-3x_2 - x_1 \le 0$$

无约束时，目标函数有1个全局极小点
约束条件下，有2个局部极小点

原因：可行域非凸

凸优化问题 $\begin{cases}\text{目标函数是凸函数}\\ \text{可行域是凸集}\end{cases}$

唯一极小点



约束条件的存在导致局部极小点的数量增加的情况

函数f的上图 epi f　　凸函数

实值函数$f:\Omega \to \mathcal{R}$， $\Omega \subset \mathcal{R}^n$ 的图像为集合$\Omega \times \mathcal{R} \subset \mathcal{R}^{n+1}$ 中的点集：

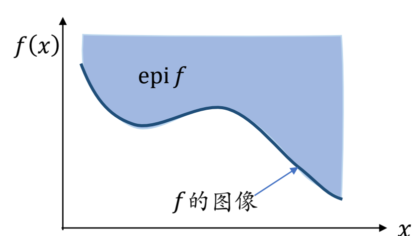
$$\left\{ \begin{bmatrix} x \\ f(x) \end{bmatrix} \middle| x \in \Omega \right\}$$

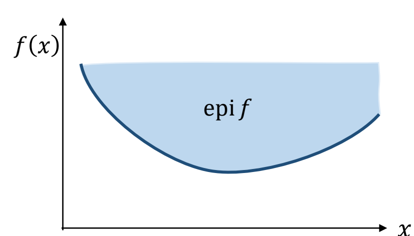
函数f的图像可以形象地描绘成$f(x)$关于x的“图形”上的点集

实值函数 $f:\Omega \to \mathcal{R}$， $\Omega \subset \mathcal{R}^n$ 的上图(epigraph)，记为 epi f，是集合$\Omega \times \mathcal{R} \subset \mathcal{R}^{n+1}$中的点集：

$$\operatorname{epi} f = \left\{ \begin{bmatrix} x \\ \beta \end{bmatrix} \middle| x \in \Omega, \beta \in \mathcal{R}, \beta \geq f(x) \right\}$$

函数f的上图epi f就是位于集合$\Omega \times \mathcal{R}$中、在函数$f$的图像上和图像上方的点集





如果函数$f:\Omega \to \mathcal{R}$， $\Omega \subset \mathcal{R}^n$， f的上图是凸集，那么函数f是集合 Ω上的凸函数

定理22.1

如果函数$f:\Omega \to \mathcal{R}$， $\Omega \subset \mathcal{R}^n$是集合$\Omega$上的凸函数，那么 Ω是凸集

证明： 反证法.

假设Ω不是凸集，那么集合Ω中存在两点y_1和y_2

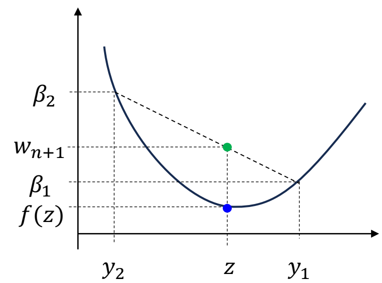
存在某个$\alpha \in (0,1)$，有 $z = \alpha y_1 + (1-\alpha)y_2$，使$z \notin \Omega$

令 $\beta_1 = f(y_1)$， $\beta_2 = f(y_2)$

可知 $\begin{bmatrix} y_1 \\ \beta_1 \end{bmatrix}$和$\begin{bmatrix} y_2 \\ \beta_2 \end{bmatrix}$位于$f$的图像上，因此也位于$f$的上图

令 $$w = \alpha \begin{bmatrix} y_1 \\ \beta_1 \end{bmatrix} + (1-\alpha) \begin{bmatrix} y_2 \\ \beta_2 \end{bmatrix}$$

有 $$w = \begin{bmatrix} z \\ \alpha\beta_1 + (1-\alpha)\beta_2 \end{bmatrix}$$



注意 $z \notin \Omega$，所以， $w \notin \mathrm{epi}\, f$，故， $\mathrm{epi}\, f$不是凸集

从而， f不是凸函数 矛盾

定理22.2

对于定义在凸集 $\Omega \subset \mathcal{R}^n$上的函数$f:\Omega \to \mathcal{R}$， f是凸函数当且仅当对于任意$x, y \in \Omega$和任意$\alpha \in (0,1)$ ，都有

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

证明 充分性

已知$\forall x, y \in \Omega$和$\alpha \in (0,1)$ ，有

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

令$\begin{bmatrix} x \\ u \end{bmatrix}$和$\begin{bmatrix} y \\ v \end{bmatrix}$是$\mathrm{epi}\, f$中的两点，$u, v \in \mathcal{R}$，由$\mathrm{epi}\, f$的定义知

$$f(x) \le u, f(y) \le v$$

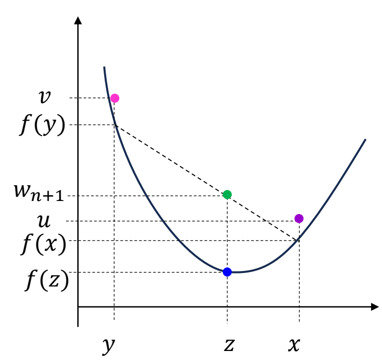
代入上面的不等式，得

$$f(\alpha x + (1-\alpha)y) \le \alpha u + (1-\alpha)v$$

因为Ω是凸集，故有 $\alpha x + (1-\alpha)\, y \in \Omega$

结合函数上图的定义，可知

$$\begin{bmatrix} \alpha x + (1-\alpha)\, y \\ \alpha u + (1-\alpha) v \end{bmatrix} \in \mathrm{epi}\, f$$



这表明$\mathrm{epi}\, f$是一个凸集，因此，f是凸函数

定理22.2

对于定义在凸集 $\Omega \subset \mathcal{R}^n$ 上的函数$f: \Omega \to \mathcal{R}$， f是凸函数当且仅当对于任意$x, y \in \Omega$ 和任意$\alpha \in (0,1)$，都有

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

必要性

已知$f: \Omega \to \mathcal{R}$是凸函数. 设$x, y \in \Omega$和$f(x) = u, f(y) = v$

因此，

$$\begin{bmatrix} x \\ u \end{bmatrix} \in \operatorname{epi} f\text{，} \begin{bmatrix} y \\ v \end{bmatrix} \in \operatorname{epi} f$$

由于f是凸函数，因此其上图为$\mathcal{R}^{n+1}$中的一个凸子集

由此可知，对于任意$\alpha \in (0,1)$都有

$$\alpha \begin{bmatrix} x \\ u \end{bmatrix} + (1-\alpha) \begin{bmatrix} y \\ v \end{bmatrix} = \begin{bmatrix} \alpha x + (1-\alpha)\, y \\ \alpha u + (1-\alpha) v \end{bmatrix} \in \operatorname{epi} f$$

这意味着，对于任意$\alpha \in (0,1)$都有

$$f(\alpha x + (1-\alpha)y) \le \alpha u + (1-\alpha) v = \alpha f(x) + (1-\alpha) f(y)$$

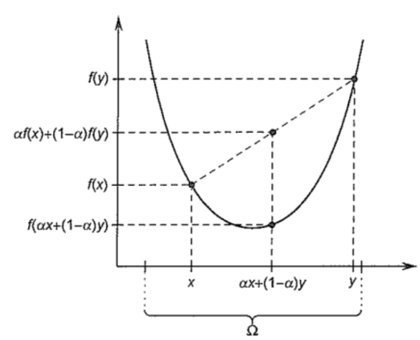
几何解释

如果函数$f: \Omega \to \mathcal{R}$是定义在凸集 Ω上的凸函数，

那么，对于任意$x, y \in \Omega$

$\mathcal{R}^{n+1}$中的连接两点$\begin{bmatrix} x \\ f(x) \end{bmatrix}$和$\begin{bmatrix} y \\ f(y) \end{bmatrix}$之间线段上的所有点，

都位于函数f的图像或上图



定理22.3 假设函数f, f_1和f_2都是凸函数，那么，对于$\forall\alpha \geq 0$，函数αf也是凸函数；
$f_1 + f_2$也是凸函数.
$f_i, i = 1, \dots, l$都是凸函数，$\max\{f_1, f_2, \dots, f_l\}$是凸函数

定义22.4 对于定义在凸集$\Omega \subset \mathcal{R}^n$上的函数$f: \Omega \to \mathcal{R}$，如果对于任意$x, y \in \Omega, x \neq y$和$\alpha \in (0,1)$，都有

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y)$$

则函数f是Ω上的严格凸函数

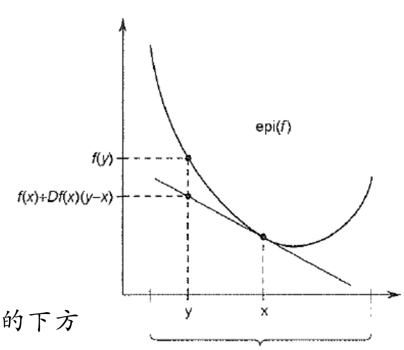
定义22.5 对于定义在凸集$\Omega \subset \mathcal{R}^n$上的函数$f: \Omega \to \mathcal{R}$，当$-f$是(严格)凸函数时，$f$是(严格)凹函数

凸函数的一阶充要条件

命题22.1 如果函数$f: \Omega \to \mathcal{R}$， $\Omega \subset \mathcal{R}^n$是二次型函数$f(x) = x^T Q x, Q \in \mathcal{R}^{n \times n}, Q = Q^T$，
那么，f是Ω上的凸函数，当且仅当对所有$x, y \in \Omega$，恒有$(x-y)^T Q (x-y) \geq 0$成立

定理22.4 设$f: \Omega \to \mathcal{R}$，$f \in C^1$是定义在开凸集 $\Omega \subset \mathcal{R}^n$上的可微函数，那么$f$是$\Omega$上的凸函数，当且仅当对于任意$x, y \in \Omega$，有

$$f(y) \geq f(x) + Df(x)(y-x)$$



几何解释

给定$x \in \Omega$，

$l(y) = f(x) + Df(x)(y-x)$是f在点x处的线性近似函数（切线）

函数f的图像总是位于线性近似函数的上方

即，在定义域内任意一点处，凸函数f的切线总是位于其上图epi f的下方

不可微函数f的广义梯度

函数$f: \Omega \to \mathcal{R}$定义在开凸集$\Omega \subset \mathcal{R}^n$上，如果对于所有$y \in \Omega$，都有

$$f(y) \geq f(x) + g^T(y-x)$$

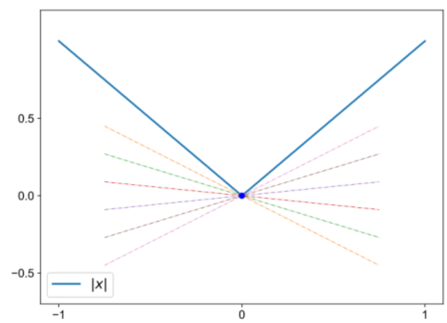
则称向量$g \in \mathcal{R}^n$为函数f定义在点$x \in \Omega$处的次梯度

与标准梯度一样，如果g是次梯度，那么对于给定的$x \in \Omega$，函数$l(y) = f(x) + g^T(y-x)$位于上图epi f的下方

函数f在点x 处的所有次梯度构成的集合称为f 在x 的次微分，记为$\partial f(x)$

$$\lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0} = a, \lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0} = b$$

$$\partial f(x_0) = [a \quad b], \text{if } a < b$$



f是凸集上的凸函数 $\Leftrightarrow$ $\nabla^2 f$半正定

定理22.5 函数$f: \Omega \to \mathcal{R}$，$f \in C^2$定义在开凸集$\Omega \subset \mathcal{R}^n$上. f是定义在Ω上的凸函数，当且仅当对于任意$x \in \Omega$，f在点x处的Hesse矩阵$F(x)$半正定

证明：充分性

令$x, y \in \Omega$，由于$f \in C^2$，由泰勒定理可知，存在$\xi \in (0,1)$，满足

$$f(y) = f(x) + Df(x)(y-x) + \frac{1}{2}(y-x)^T F\big(x + \xi(y-x)\big)(y-x)$$

由于$F\big(x + \xi(y-x)\big)$是半正定的($\leftarrow F(x) \succcurlyeq 0, f \in C^2$)，故有

$$(y-x)^T F\big(x + \xi(y-x)\big)(y-x) \geq 0$$

因此

$$f(y) \geq f(x) + Df(x)(y-x)$$

根据定理22.4，可知f是凸函数

f是凸集上的凸函数 $\Leftrightarrow$ $\nabla^2 f$半正定

定理22.5 函数$f: \Omega \to \mathcal{R}$，$f \in C^2$定义在开凸集$\Omega \subset \mathcal{R}^n$上。$f$是定义在$\Omega$上的凸函数，当且仅当对于任意$x \in \Omega$，$f$在点$x$处的Hesse矩阵$F(x)$半正定

证明：必要性. 利用反证法进行证明

假设$\exists\, x \in \Omega$，使$F(x) \ngeq 0$. 即 $\exists\, d \in \mathcal{R}^n$ 使$d^T F(x) d < 0$

已知Ω是开集，所以，x是内点

由$\nabla^2 f(x) = F(x)$的连续性知，$\exists$非零实数s，使 $y = x + sd \in \Omega$

那么，$\forall\, z = x + \alpha(y - x) = x + \alpha s d$，$\forall \alpha \in (0,1)$，有

$$d^T F(z) d < 0$$

由泰勒定理知，$f(y) = f(x) + Df(x)(y - x) + \frac{1}{2}(y - x)^T F\big(x + \alpha(y - x)\big)(y - x)$

$$= f(x) + Df(x)(y - x) + \frac{1}{2} s^2 d^T F(x + \alpha s d) d$$

$$< f(x) + Df(x)(y - x)$$

根据定理 22.4，可知f不是凸函数　　　矛盾

注：

定理22.5 可以扩展到定义域是非开集的情况

把条件修改为对于任意$x, y \in \Omega$，有

$$(x - y)^T Q (x - y) \geq 0$$

(假设$f \in C^2$定义在包含Ω的某个开集上，比如，$f \in C^2$定义在$\mathcal{R}^n$上)

根据凹函数的定义可知，函数$f: \Omega \to \mathcal{R}$，$f \in C^2$在凸集$\Omega \subset \mathcal{R}^n$上是凹的，当且仅当对于任意$x \in \Omega$，$f$的Hesse矩阵$F(x)$半负定

凸优化问题的特点

凸优化问题或凸规划：

目标函数是凸函数、约束集是凸集的优化问题. 如：

- ✓ 线性规划
- ✓ 二次规划(目标函数为二次型函数、约束方程为线性方程)

凸优化问题有很多独特之处

- ✓ 局部极小点就是全局极小点
- ✓ 极小点的一阶必要条件是凸优化问题的充分条件

凸优化问题的局部极小点等价于全局极小点

定理22.6 已知$f:\Omega\rightarrow\mathcal{R}$ 是定义在凸集$\Omega\subset\mathcal{R}^n$上的凸函数，集合$\Omega$中某一点是$f$的全局极小点，当且仅当它是$f$的局部极小点.

证明：必要性. 显然成立

充分性. 全局极小点 $\Longleftarrow$ 局部极小点

反证法

设$x^*\in\Omega$是局部极小点但不是的全局极小点，

那么，$\exists y\in\Omega$，使 $f(y)<f(x^*)$

因为f是可行域Ω上的凸函数，$\forall\alpha\in(0,1)$

$$f(\alpha y+(1-\alpha)x^*)\le\alpha f(y)+(1-\alpha)f(x^*)=\alpha[f(y)-f(x^*)]+f(x^*)<f(x^*)$$

由此可知，存在一个任意接近于x^*的点，其对应的目标函数值更小

比如，对于收敛于 x^*的序列$\{y_n\}$ $\quad y_n=\frac{1}{n}y+\left(1-\frac{1}{n}\right)x^*$

有 $\quad f(y_n)<f(x^*)$

因此，x^*不是局部极小点，与已知条件矛盾

全局极值点组成的集合是凸集

引理22.1 函数$g:\Omega\rightarrow\mathcal{R}$是定义在凸集 $\Omega\subset\mathcal{R}^n$上的凸函数，那么，对于任意$\alpha\in\mathcal{R}$，水平集$L_\alpha=\{x\in\Omega|g(x)\le\alpha\}$是凸集.

推论22.1 函数$f:\Omega\rightarrow\mathcal{R}$是定义在凸集$\Omega\subset\mathcal{R}^n$上的凸函数，那么$f$在$\Omega$上的全局极小点组成的集合是凸集.

证明：取$\alpha=\min\limits_{x\in\Omega}f(x)$，由引理22.1得出

任意一点与全局极小点的差与梯度的内积非负

引理22.2 函数$f: \Omega \to \mathcal{R}$为定义在凸集$\Omega \subset \mathcal{R}^n$上的凸函数，$f \in C^1$定义在包含$\Omega$的开集上.
选定点$x^* \in \Omega$，如果$\forall x \in \Omega, x \neq x^*$，都有

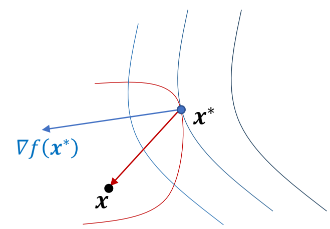
$$Df(x^*)(x - x^*) \geq 0$$

那么，x^*是在Ω上的全局极小点

证明：由于f是凸函数，依据定理22.4 “凸函数的一阶必要条件” 可知，
对于任意$x \in \Omega$，有

$$f(x) \geq f(x^*) + Df(x^*)(x - x^*)$$

由 $Df(x^*)(x - x^*) \geq 0$可知，$f(x) \geq f(x^*)$ ■



$\forall x \in \Omega$，将向量$x - x^*$看作点x^*处的一个可行方向
利用引理22.2，可推出下面的定理

全局极小点处的梯度与可行方向的内积非负

定理22.7 设函数$f: \Omega \to \mathcal{R}$是定义在凸集$\Omega \subset \mathcal{R}^n$上的凸函数，$f \in C^1$定义在包含$\Omega$的开集上.
$x^* \in \Omega$，对于点x^*处的任意可行方向d有

$$d^T \nabla f(x^*) \geq 0$$

那么，x^*是f在Ω上的一个全局极小点

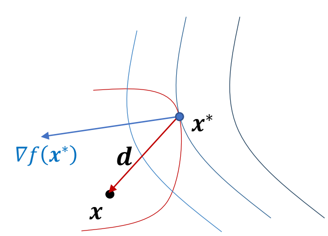
证明：令$x \in \Omega, x \neq x^*$，由于Ω是凸集，因此，$\forall \alpha \in (0,1)$，有

$$x^* + \alpha(x - x^*) \in \Omega$$

因此，向量$d = x - x^*$是在点x^*处的可行方向
根据已知条件，有

$$Df(x^*)(x - x^*) = d^T \nabla f(x^*) \geq 0$$

由引理22.2可知，x^*是f在Ω上的一个全局极小点 ■



推论22.2 函数$f: \Omega \to \mathcal{R}$，$f \in C^1$是定义在凸集$\Omega \subset \mathcal{R}^n$上的凸函数.
存在点$x^* \in \Omega$，使得

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

那么，x^*是f在Ω上的全局极小点

凸优化问题，Lagrange条件是充分条件

考虑带约束优化问题:

$$\min f(x)$$
$$\text{s.t.}\, h(x) = 0$$

假定可行域是凸集，如， $h(x) = Ax - b$

定理22.8 函数$f\colon \mathcal{R}^n \to \mathcal{R}, f \in C^1$是可行域 $\Omega = \{x \in \mathcal{R}^n | h(x) = 0\}$ 上的凸函数.
$h\colon \mathcal{R}^n \to \mathcal{R}^m, h \in C^1$ ，且Ω是凸集. 假设存在$x^* \in \Omega$和$\lambda^* \in \mathcal{R}^m$，使得

$$Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T$$

那么， x^*是f在Ω上的全局极小点

凸优化问题, x^*是f在Ω上的全局极小点$\Leftrightarrow Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T$

证明: $\Leftarrow$ 充分性　　$\Rightarrow$必要性 即，拉格朗日定理

由“凸函数的一阶充要条件”，$\forall x \in \Omega$，有 $f(x) \geq f(x^*) + Df(x^*)(x - x^*)$

已知$Df(x^*) = -\lambda^{*T} Dh(x^*)$，得　$f(x) \geq f(x^*) - \lambda^{*T} Dh(x^*)(x - x^*)$

由于Ω是凸集，对于任意$\alpha \in (0,1)$，有　$x^* + \alpha(x - x^*) = \alpha x + (1 - \alpha)x^* \in \Omega$

有　$h\big(x^* + \alpha(x - x^*)\big) = 0$

左乘λ^{*T}，减去$\lambda^{*T} h(x^*) = 0$，再除以α，得　$\dfrac{\lambda^{*T} h\big(x^* + \alpha(x - x^*)\big) - \lambda^{*T} h(x^*)}{\alpha} = 0$

$$\lambda^{*T} \frac{\big[h\big(x^* + \alpha(x - x^*)\big) - h(x^*)\big]}{\alpha \|x - x^*\|}(x - x^*) = 0$$

对上式取$\alpha \to 0$的极限，得　$\lambda^{*T} Dh(x^*)(x - x^*) = 0$

因此，　$f(x) \geq f(x^*)$　这说明x^*是f在Ω上的全局极小点

带约束优化问题可行集的凸性分析

考虑一般约束条件下的优化问题

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } h(x) = 0 \\ &\qquad g(x) \le 0 \end{aligned}$$

假设可行集是凸集$\Omega \triangleq \{x|h(x) = 0; g(x) \le 0\}$

$$\Omega = \{x|h(x) = 0; g(x) \le 0\} = \{x|h(x) = 0\} \cap \{x|g(x) \le 0\}$$

- ✓ 如果$\{x|h(x) = 0\}$和$\{x|g(x) \le 0\}$都是凸集，可行集Ω作为这两个凸集的交集也是凸集
- ✓ 等式约束，如：$\{x|h(x) = Ax - b = 0\}$是凸集
- ✓ 对于$g = [g_1 \quad \cdots \quad g_p]^T$，如果$g_i, i = 1, \dots, p$是凸函数，那么集合$\{x|g(x) \le 0\}$就是凸集
  这是因为

$$\{x|g(x) \le 0\} = \bigcap_{i \in \mathcal{J}} \{x|g_i(x) \le 0\}$$

而$L_0^{(i)} = \{x|g_i(x) \le 0\}$是凸集， 凸函数$g_i$的0-水平集$L_0^{(i)}$是凸集

凸优化问题，KKT条件是充分条件

定理22.9 $f: \mathcal{R}^n \to \mathcal{R}, f \in C^1$是可行域 Ω

$$\Omega = \{x|h(x) = 0; g(x) \le 0\}$$

上的凸函数. $h: \mathcal{R}^n \to \mathcal{R}^m, g: \mathcal{R}^n \to \mathcal{R}^p$， $h, g \in C^1$， Ω 是凸集.
假设存在点$x^* \in \Omega$, $\lambda^* \in \mathcal{R}^m$和$\mu^* \in \mathcal{R}^p$，使得

1. $\mu^* \ge 0$.
2. $Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$.
3. $\mu^{*T} g(x^*) = 0$.

那么， x^*是f在Ω上的全局极小点

证明： 令$x \in \Omega$， f是凸函数，根据“凸函数的一阶充要条件” 得

$$f(x) \ge f(x^*) + Df(x^*)(x - x^*)$$

将条件2代入上式，可得

$$f(x) \ge f(x^*) - \underbrace{\boxed{\lambda^{*T} Dh(x^*)(x - x^*)}}_{= 0} - \underbrace{\boxed{\mu^{*T} Dg(x^*)(x - x^*)}}_{\le 0\ ?}$$

采用定理22.8的证明过程，可得 $\lambda^{*T} Dh(x^*)(x - x^*) = 0$

证明: 令$x\in\Omega$，f是凸函数，根据“凸函数的一阶充要条件” 得

$$f(x)\geq f(x^*)+Df(x^*)(x-x^*)$$

将条件2代入上式，可得

$$f(x)\geq f(x^*)-\lambda^{*T}Dh(x^*)(x-x^*)-\mu^{*T}Dg(x^*)(x-x^*)$$

采用定理22.8的证明过程，可得 $\lambda^{*T}Dh(x^*)(x-x^*)=0$

接下来证明 $\mu^{*T}Dg(x^*)(x-x^*)\leq 0$

由于Ω是凸集，$\forall\alpha\in(0,1)$，有$\alpha x+(1-\alpha)x^*\in\Omega$

因此， $\forall x\in\Omega$，有 $\alpha x+(1-\alpha)x^*=x^*+\alpha(x-x^*)\in\Omega$

有 $g\big(x^*+\alpha(x-x^*)\big)\leq 0$

左乘μ^{*T} ($\geq 0^T$ 条件1)，再同时减去$\mu^{*T}g(x^*)=0$(条件3)，然后，同除以α，得

$$\frac{\mu^{*T}g\big(x^*+\alpha(x-x^*)\big)-\mu^{*T}g(x^*)}{\alpha}\leq 0 \quad\Rightarrow\quad \mu^{*T}\frac{\big[g\big(x^*+\alpha(x-x^*)\big)-g(x^*)\big]}{\alpha\|x-x^*\|}(x-x^*)\leq 0$$

两边取$\alpha\to 0$的极限，得 $\lambda^{*T}Dg(x^*)(x-x^*)\leq 0$

由此得， $\forall\, x\in\Omega$，有

$$f(x)\geq f(x^*)-\lambda^{*T}Dh(x^*)(x-x^*)-\mu^{*T}Dg(x^*)(x-x^*)\geq f(x^*)$$

1. 引言
2. 凸函数
3. 凸优化问题
4. 半定规划
   ① 线性矩阵不等式及其性质
   ② 线性矩阵不等式求解器
   ③ 线性矩阵不等式的可行性问题
   ④ 线性矩阵不等式约束下线性目标函数极小化
   ⑤ 线性矩阵不等式约束下的广义特征值极小化

半定规划被认为是自20 世纪50 年代著名的线性规划以后的另一个数学规划领域革命性的研究进展

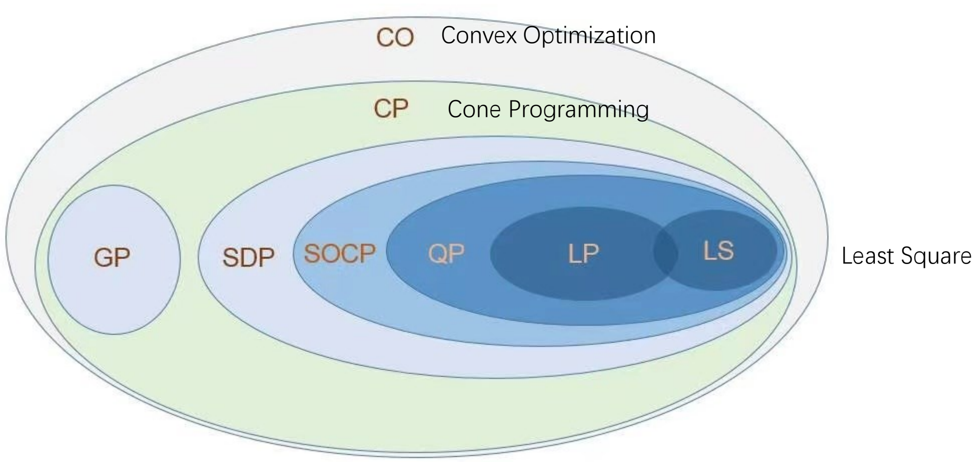
半定规划的发展得益于线性规划的充分研究

半定规划作为一类特殊的矩阵优化问题，不同于经典线性与非线性优化问题

在实际应用方面，半定规划和线性规划一样，作为重要的凸优化量化建模工具被应用于工程、经济学等领域

优化问题中的许多类型，如LP、SOCP和QP，可以看作是SDP问题
许多其它凸优化中的实用问题也可以转化表示为SDP问题

凸优化中LP问题的内点法可以有效地推广应用于SDP问题



二阶锥规划*second-order cone programming* (SOCP)

例

两个椭圆集合间最短距离

$\min t$

$\text{s.t. } d^2 = (x_1 - x_3)^2 + (x_2 - x_4)^2 \le t$

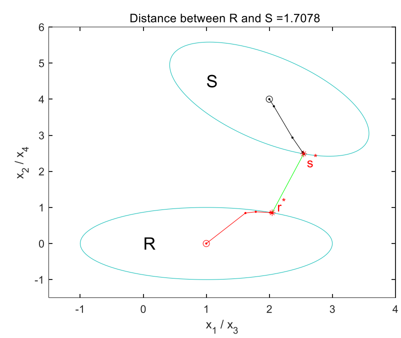
$c_1(x) = -[x_1 \quad x_2]\begin{bmatrix}0.25 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + [x_1 \quad x_2]\begin{bmatrix}0.5\\ 0\end{bmatrix} + 0.75 \ge 0$

$c_2(x) = -0.125[x_3 \quad x_4]\begin{bmatrix}5 & 3\\ 3 & 5\end{bmatrix}\begin{bmatrix}x_3\\ x_4\end{bmatrix} + [x_3 \quad x_4]\begin{bmatrix}5.5\\ 6.5\end{bmatrix} - 17.5 \ge 0$

$x = [t \;\; x_1 \;\; x_2 \;\; x_3 \;\; x_4]^T$

SOCP是SDP的特例

可以转化为SDP问题



线性矩阵不等式及其性质

考虑 $F_i \in S^m, i = 0,1,\dots,n$　　m阶实对称矩阵　　常数矩阵

$x \in \mathcal{R}^n$　　n维向量　　变量

$F(x) = F_0 + x_1F_1 + \cdots + x_nF_n = F_0 + \sum_{i=1}^{n} x_iF_i \in S^m$　是向量x的一个仿射函数

不等式约束　$F(x) = F_0 + x_1F_1 + \cdots + x_nF_n \succcurlyeq 0$

表示满足$z^TF(x)z \ge 0, \forall z \in \mathcal{R}^m$的点$x$的集合，
即，保证$F(x)$是半正定的点x的集合

容易验证，
集合$\{x \in \mathcal{R}^n | F(x) \succcurlyeq 0\}$是凸集

$F(x) = F_0 + x_1F_1 + \cdots + x_nF_n \succcurlyeq 0$ 称为线性矩阵不等式(Linear Matrix Inequality，LMI)
亦称为仿射矩阵不等式

$F(x) = F_0 + x_1F_1 + \cdots + x_nF_n \succ 0$ 的线性矩阵不等式　　集合$\{x \in \mathcal{R}^n | F(x) \succ 0\}$是凸集

一组由多个线性矩阵不等式构成的约束

$$F_1(x) \succcurlyeq 0, \dots, F_k(x) \succcurlyeq 0$$

等价于一个的线性矩阵不等式的约束

$$F(x) = \begin{bmatrix} F_1(x) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & F_k(x) \end{bmatrix} \succcurlyeq 0$$

例　$Ax \leq b$　$\Leftrightarrow$　$b_i - a_i^T x \geq 0, i = 1, \dots, m$　看作一个线性矩阵不等式

a_i^T 是A的第i行

把m个线性不等式转换为一个线性矩阵不等式

$$F(x) = \begin{bmatrix} b_1 - a_1^T x & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_m - a_m^T x \end{bmatrix} \succcurlyeq 0$$

半定规划　凸规划问题的一个分支

半定规划：求解的是线性矩阵不等式约束下的线性目标函数的极小值

$$\min c^T x$$
$$\text{s.t. } F(x) \succcurlyeq 0$$

线性矩阵不等式约束定义了一个凸可行集，要求在该可行集上使得目标函数达到极小值

半定规划可视为线性规划的推广，其中，向量不等式约束替换为矩阵不等式约束

优化、控制系统设计以及信号处理中的许多问题都可以转化为线性矩阵不等式的形式

- ✓ 确定是否存在一个点x使得$F(x) \succ 0$成立的问题称为可行性问题
- ✓ 如果不存在这样一个x，则称线形矩阵不等式问题是不可行的

正定矩阵的性质

$P \in \mathcal{R}^{n\times n}$非奇异，$M \in \mathcal{R}^{n\times n}$, $\det M \neq 0$，$x = Mz$，则有

$$x^T P x \geq 0 \iff z^T M^T P M z \geq 0 \qquad P \succcurlyeq 0 \iff M^T P M \succcurlyeq 0$$

$$P \succ 0 \iff M^T P M \succ 0$$

对于方阵 $\begin{bmatrix} A & B \\ B^T & D \end{bmatrix}$

$$\begin{bmatrix} A & B \\ B^T & D \end{bmatrix} \succcurlyeq 0 \iff \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}\begin{bmatrix} A & B \\ B^T & D \end{bmatrix}\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \succcurlyeq 0$$

$$\begin{bmatrix} A & B \\ B^T & D \end{bmatrix} \succcurlyeq 0 \iff \begin{bmatrix} D & B^T \\ B & A \end{bmatrix} \succcurlyeq 0$$

矩阵的Schur补 与 矩阵的正定性

$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$，$A_{11}$和$A_{22}$是子方阵

假设 A_{11}可逆

$$\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$$

$\triangle_{11} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ 为 A_{11}的Schur补

当 $A_{12} = A_{21}^T$ 时，可得

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \succ 0 \iff \begin{bmatrix} A_{11} & 0 \\ 0 & \triangle_{11} \end{bmatrix} \succ 0$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \succ 0 \iff A_{11} \succ 0, \qquad \triangle_{11} \succ 0$$

矩阵的Schur补 与 矩阵的正定性

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},\ A_{11}\text{和}A_{22}\text{是子方阵}$$

假设 A_{22} 可逆

$$\begin{bmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{bmatrix} = \begin{bmatrix} \triangle_{22} & 0 \\ 0 & A_{22} \end{bmatrix}$$

$\triangle_{22} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ 为 A_{22} 的Schur补

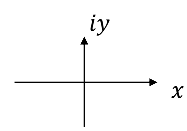
当 $A_{12} = A_{21}^T$ 时，可得

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \succ 0 \Leftrightarrow \begin{bmatrix} \triangle_{22} & 0 \\ 0 & A_{22} \end{bmatrix} \succ 0$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \succ 0 \Leftrightarrow A_{22} \succ 0, \quad \triangle_{22} \succ 0$$

例22.8 线性矩阵不等式的可行性问题

令$A \in \mathcal{R}^{m\times m}$是已知的常实数方阵，$A$的特征值是否都在复平面的左半平面上?



矩阵A的特征值全部位于复平面的左半平面 $\Leftrightarrow$ 存在对称的常实数正定矩阵 P，使 $A^TP + PA \prec 0$ 或 $-A^TP - PA \succ 0$

李雅普诺夫不等式

因此，

A的所有特征值都位于复平面的左半平面 $\Leftrightarrow$ 矩阵不等式 $\begin{bmatrix} P & 0 \\ 0 & -A^TP - PA \end{bmatrix} \succ 0$ 是可行的

即，$\exists\, P = P^T \succ 0$ 使 $A^TP + PA \prec 0$

下面证明，确定$P = P^T \succ 0$使 $A^TP + PA \prec 0$，实际上就是求解线性矩阵不等式问题

证明，确定$P=P^T\succ 0$使 $A^TP+PA\prec 0$，实际上就是求解线性矩阵不等式问题

为此，设
$$P=\begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_{m+1} & \cdots & x_{2m-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_{2m-1} & \cdots & x_n \end{bmatrix},\qquad n=\frac{m(m+1)}{2}$$

定义矩阵 $P_1=\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}\quad P_2=\begin{bmatrix} 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}\quad \cdots\ P_n=\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$

P_i中非零元素是与P中的x_i相对应的

$$F_i=-A^TP_i-P_iA,\qquad i=1,2,\dots,n$$

$$\begin{aligned} A^TP+PA &= x_1(A^TP_1+P_1A)+x_2(A^TP_2+P_2A)+\cdots+x_n(A^TP_n+P_nA) \\ &= -x_1F_1-x_2F_2-\cdots-x_nF_n \\ &\prec 0 \end{aligned}$$

令 $F(x)=x_1F_1+\cdots+x_nF_n$，

$$F(x)\succ 0 \iff P=P^T\succ 0 \quad \text{和} \quad A^TP+PA\prec 0$$

线性矩阵不等式求解器

不等式约束 $\qquad F(x)=F_0+x_1F_1+\cdots+x_nF_n\succcurlyeq 0$

MATLAB工具箱包含3种求解器

1. 求取线性矩阵不等式约束下的可行解

$$N^T L(X_1,X_2,\dots,X_k) N \preccurlyeq M^T R(X_1,X_2,\dots,X_k) M$$

左侧外部因子

矩阵变量

左侧内部因子

对称块矩阵

右侧内部因子

对称块矩阵

右侧外部因子

符号式“$L\preccurlyeq R$”的左侧是指L

符号式“$R\succcurlyeq L$”的左侧是指L

setlmis([]) 初始化线性矩阵不等式
lmivar 声明矩阵变量
lmiterm 指定当前问题的所有线性矩阵不等式约束 LMI
lmiterm ⋮

getlmis 来获取问题的内部表达式
feasp 求解线性矩阵不等式问题的可行解
dec2mat 提取矩阵变量的值

具体参考：线性矩阵不等式的_matlab求解.pdf

线性矩阵不等式的可行性问题的求解

一般情况下，线性矩阵不等式可行性问题

确定 x
使得 $L(x) \preccurlyeq R(x)$

构造一个辅助凸规划问题

$$\min t$$
$$\text{s.t. } L(x) \preccurlyeq R(x) + tI$$

当t的极小值是负数时，上述线性矩阵不等式系统有可行解
命令 feasp可求解这一问题，在求解过程中，每次迭代中的t值都可以显示出来

例22.9

已知 $A_1 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, A_2 = \begin{bmatrix} -2 & 0 \\ 1 & -1 \end{bmatrix}$

试利用前面提到的MATLAB的线性矩阵不等式工具箱中的命令，编写程序求出矩阵P使得$P \succcurlyeq 0.5I_2$，且

$$A_1^T P + PA_1 \preccurlyeq 0$$
$$A_2^T P + PA_2 \preccurlyeq 0$$

$$\min t$$
$$\text{s.t. } L(x) \preccurlyeq R(x) + tI$$

```
A_1 = [-1,0;0,-1];
A_2 = [-2,0;1,-1];
setlmis( [] )
P = lmivar(1,[2,1])
lmiterm([1 1 1 P],A_1',1,'s')  % A1
lmiterm([2 1 1 P],A_2',1,'s')  % A2
lmiterm([3 1 1 0], .5)         % I
lmiterm([-3 1 1 P],1,1)        % P
lmis=getlmis;
[tmin,xfeas] = feasp(lmis);
P = dec2mat(lmis,xfeas,P)
```

```
Solver for LMI feasibility problems L(x) < R(x)
   This solver minimizes  t  subject to  L(x) < R(x) + t*I
   The best value of t should be negative for feasibility

 Iteration   :    Best value of t so far
     1                    -1.406913

 Result:  best value of t:    -1.406913
          f-radius saturation:  0.000% of R =  1.00e+09

P =
   18.5489    6.3969
    6.3969   25.4731
```

线性矩阵不等式约束下线性目标函数极小化

凸规划问题 $\min c^T x$
$\text{s.t. } A(x) \preccurlyeq B(x)$

MATLAB的线性矩阵不等式工具箱中有此类凸规划问题的求解器，调用函数mincx实现

- ✓ 指定线性矩阵的不等式约束
- ✓ 声明线性目标函数
- ✓ 调用函数mincx

首先解决可行性问题，即利用求解器feasp，求出一个x满足$A(x) \preccurlyeq B(x)$

然后，利用求解器 mincx求得该问题的极小点

例22.10

考虑优化问题

$$\min c^T x \quad \text{s.t. } Ax \preccurlyeq b \qquad c^T = [4 \quad 5] \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 2 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 8 \\ 18 \\ 14 \end{bmatrix}$$

首先解决可行性问题，即利用求解器feasp，求出一个x满足 $Ax \preccurlyeq b$

然后，利用求解器 mincx求得该问题的极小点

```
% Enter problem data
A = [1,1;1,3;2,1];
b = [8,18,14]';
c = [-4,-5]';
setlmis([]);
X = lmivar(2,[2 1]);
lmiterm([1 1 1 X],A(1,:),1);
lmiterm([1 1 1 0],-b(1));
lmiterm([1 2 2 X],A(2,:),1);
lmiterm([1 2 2 0],-b(2));
lmiterm([1 3 3 X],A(3,:),1);
lmiterm([1 3 3 0],-b(3));
lmis = getlmis;
%---------------------------
disp('------feasp result-----------')
[tmin,xfeas] = feasp(lmis);
x_feasp = dec2mat(lmis,xfeas,X)
disp('-------- mincx result ------------')
[objective,x_mincx] = mincx(lmis,c,[0.0001 1000 0 0 1])
```

```
------feasp result------------
 Solver for LMI feasibility problems L(x) < R(x)
    This solver minimizes  t  subject to  L(x) < R(x) + t*I
    The best value of t should be negative for feasibility

 Iteration   :    Best value of t so far
     1                   -10.552963

 Result:  best value of t:   -10.552963
          f-radius saturation:  0.000% of R =  1.00e+09

x_feasp =    -64.3996
             -25.1712

-------- mincx result -------------
objective =    -37.0000

x_mincx =     3.0000
              5.0000
```

例22.11

利用函数mincx求解优化问题

$$\min \operatorname{trace}(P) \quad \text{s.t. } A^T P + PA \preccurlyeq 0, \quad P \succcurlyeq 0 \qquad \operatorname{trace}(P) = \sum p_{ii}$$

首先确定向量c，满足$c^T x = \operatorname{trace}(P)$

在指定了线性矩阵不等式且通过命令(如命令lmisys=getlmis)获得了内部表达式之后，就可以用下面的MATLAB代码获得期望的c

然后，利用函数mincx来求解此问题

```
q = decnbr(lmisys); %Total number of decision variables in system of LMIs
c = zeros(q,1);
for j = 1:q
    Pj = defcx(lmisys,j,P);% Help specify c^Tx objectives for mincx solver
    c(j) = trace(Pj);
end
```

此例说明函数defcx的用法，该函数可以用来构造向量c，供线性矩阵不等式求解器mincx使用

线性矩阵不等式约束下的广义特征值极小化问题

$$\begin{aligned} &\min \lambda \\ &\text{s.t. } C(x) \preccurlyeq D(x) \\ &\qquad 0 \preccurlyeq B(x) \\ &\qquad A(x) \preccurlyeq \lambda B(x) \end{aligned}$$

两种不同形式的线性矩阵不等式约束:

标准形式 $C(x) \preccurlyeq D(x)$

线性分式形式 $A(x) \preccurlyeq \lambda B(x)$

线性矩阵不等式约束下的广义特征值极小化问题，可以使用求解器 gevp 解决

函数 gevp的调用方式为

[lopt,xopt] =gevp{lmisys,nflc}

返回值lopt是广义特征值的全局极小值，xopt是最优决策向量变量

参数lmisys表示的线性矩阵不等式系统$C(x) \preccurlyeq D(x)$，$0 \preccurlyeq B(x)$或者当$\lambda = 1$时的$A(x) \preccurlyeq \lambda B(x)$

与前面的两个求解器一样，矩阵变量形式的最优值可由dec2mat得到

线性分式约束的数目由nflc指定

函数gevp还有其他的输入参数，都是可选的 （参见线性矩阵不等式的_matlab求解.pdf)

例22.12

求能够使得

$$\begin{aligned} &P \succ 0 \\ &A^T P + PA \preccurlyeq -\alpha P \end{aligned}$$

成立的最小的α

$$A = \begin{bmatrix} -1.1853 & 0.9134 & 0.2785 \\ 0.9058 & -1.3676 & 0.5469 \\ 0.1270 & 0.0975 & -3.0000 \end{bmatrix}$$

该问题是关于求解稳定线性微分方程$\dot{x} = Ax$衰减率的

可以使用如下代码构造相应的线性矩阵不等式系统，

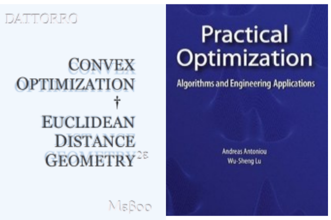
求取满足条件的最小α

注意，用$P \succ 0.01I$替换了$P \succ 0$

```
close;clear;clc;
A = [-1.1853,0.9134,0.2785;
0.9058,-1.3676,0.5469;
0.1270,0.0975,-3.0000];
setlmis([]);
P = lmivar(1,[3 1])
lmiterm([-1 1 1 P],1,1)    % P
lmiterm([1 1 1 0],.01)     % P >= 0.01*I
lmiterm([2 1 1 P],1,A,'s') % linear fractional constraint -- LSH
lmiterm([-2 1 1 P],1,1)    % linear fractional constraint   -- RSH
lmis = getlmis;
[gamma,P_opt] = gevp(lmis,1);
P = dec2mat(lmis,P_opt,P)
alpha = -gamma
```

```
P =    0.6996   -0.7466   -0.0296
      -0.7466    0.8537   -0.2488
      -0.0296   -0.2488    3.2307

alpha =      0.6561
```



1. Convex Optimization + Euclidean Distance Geometry, Dattorro, Chapter 4
2. Practical Optimization Algorithms and Engineering Applications, Chapter 14, A. Antoniou, W. LU

例 线性系统的稳定性 (Stability of linear systems)

问题描述

分析一个线性系统的稳定性，Lyapunov 方法：

对于线性系统$\frac{dx}{dt}=Ax$ 稳定当且仅当：存在Lyapunov函数$V(x)$ 满足

$$\begin{cases} V(x)=x^TPx>0 \\ \dot{V}(x)=x^TPx+x^TP\dot{x}=x^T(A^TP+PA)x<0 \end{cases}$$

根据给定的系统矩阵A，构造满足条件的正定阵 $P\in\mathcal{R}^{n\times n}$ ，即证明了闭环系统的稳定性

转化为 LMI

$A^TP+PA\prec 0$是一个关于 $P\in\mathcal{R}^{n\times n}$的LMI，最终将转化为含有 $m=\frac{n(n+1)}{2}$个自由度的LMI

隐含条件P是对称的

以$n = 3$为例

$$P = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{bmatrix}$$

$$= \sum_{i=1}^{3} \sum_{j=1}^{3} x_{ij} E_{ij}$$

$E_{ij} \in \mathcal{R}^{3\times 3}$在$(i,j)$为1、其余为0的基矩阵

Lyapunov 方法对时间的导数的约束转化为

$$A^T P + PA = A^T \left(\sum_{i=1}^{3} \sum_{j=1}^{3} x_{ij} E_{ij} \right) + \left(\sum_{i=1}^{3} \sum_{j=1}^{3} x_{ij} E_{ij} \right) A$$

$$= \sum_{i=1}^{3} \sum_{j=1}^{3} x_{ij} \left(A^T E_{ij} + E_{ij} A \right) < 0$$

对应于$F_0 = \emptyset, m = 6$的情况

特征值问题(EVP)：求矩阵G(x)的特征值的最小化问题

$$\begin{array}{c} \min \lambda \\ s.t.\ \ G(x) < \lambda I \end{array} \quad \overset{c^T x = \lambda}{\Longleftrightarrow} \quad \begin{array}{c} \min c^T x \\ s.t.\ \ G(x) - c^T x < 0 \end{array}$$

$$\Big\Updownarrow \ F(x) = G(x) - c^T x$$

$$\begin{array}{c} \min c^T x \\ s.t.\ \ F(x) < 0 \end{array}$$