1. 定义

广义：凸目标，凸集约束的极小化问题.

一般优化问题： $\min f_0(x)$ 凸函数

s.t. $f_i(x) \le 0$, $i=1,\cdots,m$ 凸函数

$\boxed{h_i(x)=0}$, $i=1,\cdots,p$ 仿射函数

狭义凸问题

凸集 { ∂-sublevel set；子空间 $a_i^T x = b_i \Leftarrow h_i(x)=0$ }

一些定义：① $x \in R^n$, optimization variable

② $f_0: R^n \to R$，目标函数 objective function，

损失函数 lost function，效用函数 utility function

③ $f_i(x) \le 0$ 不等式约束 Inequality constraint

$h_i(x)=0$ 等式约束 equality constraint

④ $m=p=0$ 无约束 unconstrained

⑤ 优化问题的域 domain: $D = \bigcap_{i=1}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$

⑥ 可行解集 feasible set

若 $f_i(x) \le 0$, $i=1,\cdots,m$

$h_i(x)=0$, $i=1,\cdots,p$

☆ $x \in D$ 为可行解，

$X_f = \{x \text{ 为可行解}\}$

⑦ 问题的最优值 optimal value: $p^* = \inf\{f_0(x) \mid x \in X_f\}$

若 X_f 为空集，$p^* = +\infty$.

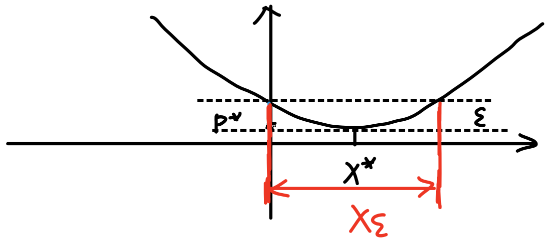
⑧ 最优解 optimal point/solution：若 x^* 可行，且 $f_0(x^*) = p^*$.

⑨ 最优解集 optimal set: $X_{opt} = \{x \mid x \in X_f, f_0(x) = p^*\}$

⑩ ε-次优解集. ε-suboptimal set

补充：satisficing (satisfy + suffice) solutions

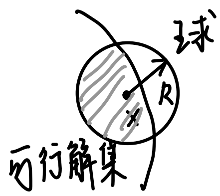
$X_\varepsilon = \{x \mid x \in X_f, f_0(x) \le p^* + \varepsilon\}$

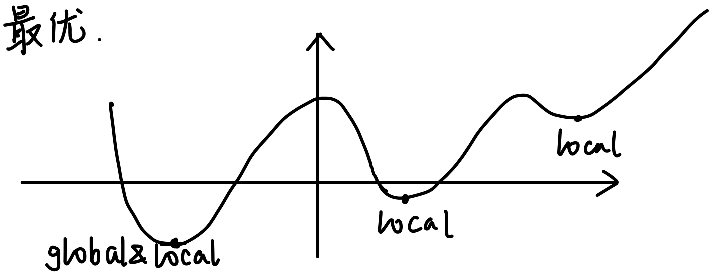


① 局部最优解 locally optimal

$$f_0(x) = \inf \left\{ f_0(z) \;\middle|\; \begin{array}{l} f_i(z) \le 0,\ i=1,\cdots,m \\ h_i(z)=0,\ i=1,\cdots,p \\ \|z-x\| \le R \end{array} \right.$$

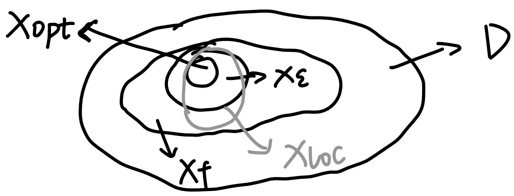
$\exists R>0$ 使 x 局部最优.





X_{loc}（局部最优解集）

☆ 各种解之间的关系



☆ why $f_i(x) \le 0$ rather than $f_i(x) < 0$?

若 $x \in X_f$，$f_i(x)=0$，则 $f_i(x) \le 0$ 为活动约束（active）

$f_i(x)<0$，则 $f_i(x) \le 0$ 为不活动约束（inactive）

eg： min $-$utility

s.t. money ≤ 100 （不能全都花完）

$-|\log(100-\text{money})| \le 100$ （使用域的方法）

2. 可行性优化问题 feasibility Problems

find x min 0

s.t. $f_i(x) \le 0, \ i=1,\cdots,m$

$h_i(x)=0, \ i=1,\cdots,p$

等价问题 例：Box Constraints

$\min f_0(x)$

s.t. $l_i \le x_i \le u_i, \ i=1,\cdots,n$

$$\boxed{\begin{aligned} & l_i - x_i \le 0, \ i=1,\cdots,n \\ & x_i - u_i \le 0, \ i=1,\cdots,n \end{aligned}}$$

例：$\min \alpha_0 f_0(x)$ （量纲标准化）

s.t. $\alpha_i f_i(x) \le 0, \ i=1,\cdots,m, \ \alpha_i > 0$

$\beta_i h_i(x)=0, \ i=1,\cdots,p, \ \beta_i \ne 0$

例：$\psi_0: \mathbb{R} \to \mathbb{R}$ 单调递增.

$\psi_1,\cdots,\psi_m: \mathbb{R} \to \mathbb{R} \quad \psi_i(u) \le 0 \Leftrightarrow u \le 0$

$\varphi_1,\cdots,\varphi_m: \mathbb{R} \to \mathbb{R} \quad \varphi_i(u)=0 \Leftrightarrow u=0$

$\min \ \psi_0(f_0(x))$

s.t. $\psi_i(f_i(x)) \le 0, \ i=1,\cdots,m$

$\varphi_i(h_i(x))=0, \ i=1,\cdots,p$

例：$\min \|Ax-b\|_2$

$\min \|Ax-b\|_2^2$

例：消除等式约束.

$\{h_i(x)=0,\ i=1,\cdots,p\}$ 一组方程，$z\in R^k$，$\varphi: R^k\to R^n$，$x=\varphi(z)$ 代入

$\min f_0(\varphi(z))$

$s.t.\ f_i(\varphi(z))\leqslant 0,\ i=1,\cdots,m$

例：消除线性等式约束. $Ax-b=0$，$A\in R^{p\times n}$ $\Leftrightarrow$ $x=\varphi(z)$ ①无解 → 原问题无解

化零空间

$z\in R^{n-r}$ (降维，降秩) $x=Fz+x_0$ $Ax_0=b$ 等(由z决定的子空间) ②A可逆，$x=A^{-1}b$

$\min\ f_0(Fz+x_0)$ 约束维数减少，不再有等式约束

$s.t.\ f_i(Fz+x_0)\leqslant 0,\ i=1,\cdots,m$

例：$\min f_0(x)$ 是否关于x和s都是凸的(联合凸)

$s.t.\ s_i\leqslant 0,\ i=1,\cdots,m$

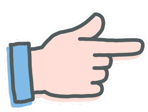
$f_i(x)-s_i=0,\ i=1,\cdots,m$ slack variable

$a_i^T x=b_i,\ i=1,\cdots,p$

例：$\min f_0(x)=x_1^2+x_2^2$

$s.t.\ f_1(x)=\dfrac{x_1}{1+x_2^2}\leqslant 0$

$h_1(x)=(x_1+x_2)^2=0$



$\min f_0(x)=x_1^2+x_2^2$

$s.t.\ f_1(x)=x_1\leqslant 0$

$\tilde{h}_1(x)=x_1+x_2=0$

☆ 区分 Quasi-covex optimization & Non-convex optimization

3. 重要性质：局部最优 = 全局最优

① 局部最优：$\exists R>0$，$f_0(x)=\inf\{f_0(z)\mid z\text{可行}, x\text{可行}, \|x-z\|\leq R\}$

证明：（反证法）设 x 不是全局最优，$\exists y$ 可行，$f_0(y)<f_0(x)$

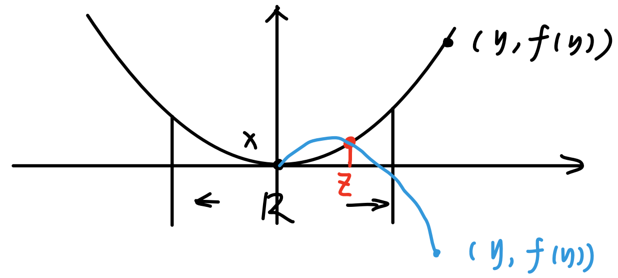
因 x 局部最优 $\|y-x\|_2>R$

$z=(1-\theta)x+\theta y$，取 $\theta=\dfrac{R}{2\|y-x\|_2}$（由于 $\theta\in(0,1)$，则 z 必为凸组合）

z 可行，且 $f_0(z)\leq(1-\theta)f_0(x)+\theta f_0(y)$ $<f_0(z)$

$\|z-x\|_2=\theta\|x-y\|_2=\frac{R}{2}<R$，then $f_0(x)\leq f_0(z)$

$f_0(y)<f_0(x)\leq f_0(z)$，与上述条件矛盾.



② 可微目标函数情况下的最优性

f_0 可微，则 f_0 凸 $\Leftrightarrow$ $\mathrm{dom}f_0$ 为凸，$f_0(y)\geq f_0(x)+\nabla f_0^T(x)(y-x)$，$\forall x,y\in\mathrm{dom}f_0$

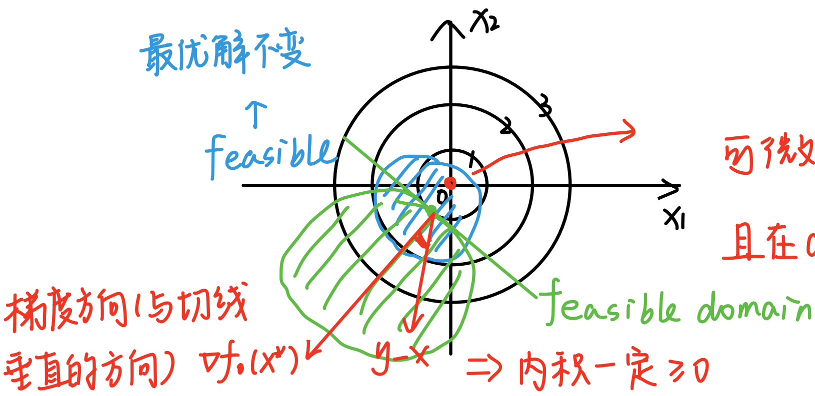
设凸问题可行域 $X_f=\left\{x\middle|\begin{array}{l}f_i(x)\leq 0, i=1,\cdots,m\\ h_i(x)=0, i=1,\cdots,p\end{array}\right\}\cap\mathrm{dom}f_0$

梯度方向

$x^*\in X_f$ 最优 $\Leftrightarrow$ $\nabla f_0^T(x^*)(y-x^*)\geq 0$，$\forall y\in X_f$

直观解释：令 $\mathrm{dom}f_0\triangleq X_f$，则 $\exists x^*$，$\forall y\in\mathrm{dom}f_0$

若 $f(x^*)$ 最小，则 $f_0(y)\geq f_0(x^*)$，则 $\nabla f_0^T(x^*)(y-x^*)\geq 0$



可微，一阶偏导等于0

且在 $\mathrm{dom}f$，则必为全局最优解

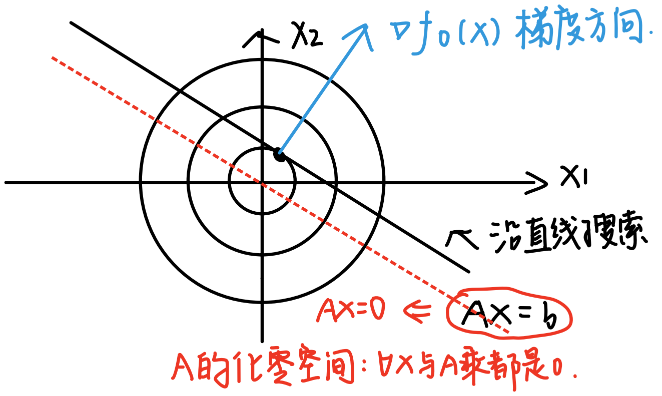
例：约束仅为等式约束

$\min f_0(x)$, $\mathrm{dom} f_0 = R^n$

s.t. $Ax=b$

若 $\exists x$, $Ax=b$, x 最优 $\Leftrightarrow$ $\exists y$, $Ay=b$

$\Rightarrow \nabla f_0^T(x)(y-x) \ge 0$



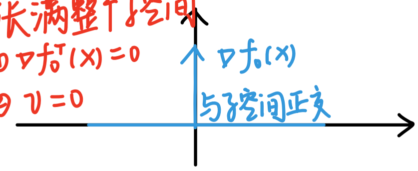
$Ax=b$, $Ay=b \Rightarrow y=x+v$, $v \in N(A)$

$\nabla f_0^T(x) v \ge 0$, $\forall v \in N(A)$ A的化零空间

张满整个空间

① $\nabla f_0^T(x)=0$

② $v=0$



① $v=0$，则A可逆，$x=A^{-1}b$

② $\nabla f_0(x)$ 与 $N(A)$ 正交

例：约束仅为非负约束

$\min f_0(x)$

s.t. $x \ge 0$

若 $\exists x \ge 0$, x 最优 $\Leftrightarrow$ $\forall y \ge 0$, $\nabla f_0^T(x)(y-x) \ge 0$

即 $\nabla f_0^T(x) y - \nabla f_0^T(x) x \ge 0$　$\sum_i a_i y_i \ge 0$

① 若 $\forall i$ $(\nabla f_0(x))_i < 0$，则 $\nabla f_0^T(x) y$ 必可取到无穷小. 则必有 $\nabla f_0(x) \ge 0$

② $\forall y$ 均有 $\nabla f^T(x)(y-x) \ge 0 \Rightarrow \nabla f_0^T(x) x \le 0$

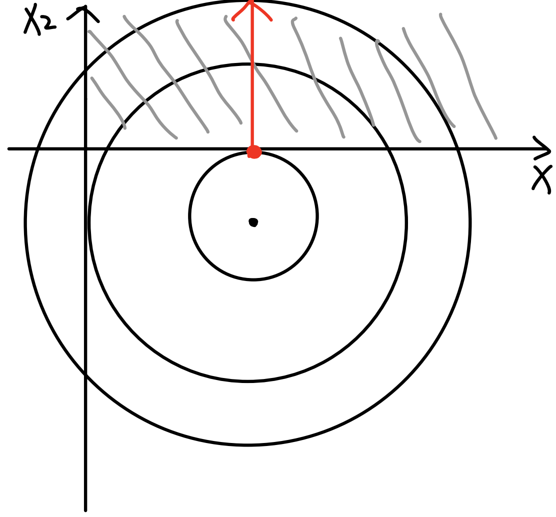
③ $\nabla f_0^T(x) \ge 0$, $x \ge 0 \Rightarrow \nabla f_0^T(x) x \ge 0$

$\Big\}$ $\nabla f_0^T(x) x = 0$

总结：如果x是最优解，则 $\begin{cases} x \ge 0 \\ \nabla f_0(x) \ge 0 \\ (\nabla f_0(x))_i x_i = 0 \end{cases}$

Complementarity

互补条件



最优解的直观理解：能使等高线到达最低。

$x_1>0,\ x_2=0$

$(\nabla f_0(x))_1=0,\ (\nabla f_0(x))_2>0$

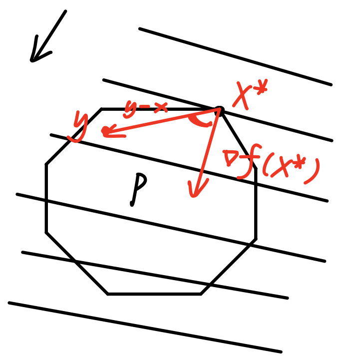
4. 典型凸问题

(1) 线性规划

$$
\begin{aligned}
\min\ & c^T x+d && c\in R^n,\ d\in R && \text{仿射} \\
\text{s.t.}\ & Gx\preceq h && G\in R^{m\times n},\ h\in R^m && \text{仿射} \\
& Ax=b && A\in R^{k\times n},\ b\in R^k && \text{仿射}
\end{aligned}
\quad \Big\}\ \text{凸问题}
$$

目标与约束均为线性 linear program (LP问题)

历史：1930-1940 调度问题 Dantzig、Von Neumann、Kantorovizh



$\nabla f(x^*)(y-x)$ 一定非负.

(等价变换)

$$
\begin{aligned}
\min\ & c^T x+d \\
\text{s.t.}\ & Gx+s=h \\
& Ax=b \\
& s\succeq 0
\end{aligned}
\quad \Longleftrightarrow \quad
\begin{aligned}
\min\ & c^T x^+ - c^T x^- + d \\
& G_1 x^+ - G_1 x^- + s = h \\
& Ax^+ - Ax^- = b \\
& s\succeq 0,\ x^+\succeq 0,\ x^-\succeq 0
\end{aligned}
$$

eg: $x=[1,-1,2,-2]^T$, $x^+=[1,0,2,0]^T$, $x^-=[0,1,0,2]$

$x=x^+-x^-$, $x^+, x^{-1}\succeq 0$

$$\begin{cases} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{cases}$$ (linprog 函数的命名规划)

例：食谱问题.

一份健康的饮食包含 m 种不同的营养，每种至少需要 $b_1, \cdots, b_m$。我们可以从n 种食物中选择非负的量 $x_1, \cdots, x_n$ 以构成一份食谱。单位第 j 种食品含有营养 i 的量为 a_{ij}，而价格为 c_j。我们希望设计出一份最便宜的满足营养需求的食谱。这一问题可以描述为线性规划

目标：单位食物价格 c_j，找到总价最低的食谱.

$$\begin{aligned} \min \quad & \sum_{j=1}^n c_j x_j \\ \text{s.t.} \quad & \sum_{j=1}^n a_{ij} x_j \geq b_i, \quad i=1,\cdots,m \\ & x_j \geq 0, \quad j=1,\cdots,n \end{aligned} \xrightarrow{\text{向量形式}} \begin{aligned} \min \quad & [c_1 \cdots c_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ \text{s.t.} \quad & \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \geq \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \\ & \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \geq 0 \end{aligned}$$

(2) 线性分数规划 linear fractional programming

(P_0)
$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & Gx \leq h \\ & Ax = b \end{aligned}$$

$f_0(x) = \dfrac{c^T x + d}{e^T x + f}$ 是拟凸函数，不是凸函数

$\mathrm{dom} f_0 = \{x \mid c^T x + f > 0\}$

(P_1)
$$\begin{aligned} \min \quad & c^T x + dz \\ \text{s.t.} \quad & Gy - hz \leq 0 \\ & Ay - bz = 0 \\ & e^T y + fz = 1 \\ & z \geq 0 \end{aligned}$$

验证：① x^* 是 P_0 的optimal $\Leftrightarrow$ y^*、z^* 是 P_1 的optimal

② $\forall x$，能找 y, z 对于 P_1 可行，且函数值相等

$\forall y, z$，能找到 x 对于 P_0 可行，且函数值相等

证明: 1) 若 x 在 P_0 可行. 取 $y = \dfrac{x}{e^T x+f}$, $z = \dfrac{1}{e^T x+f}$

$$\begin{cases} Gx \le h \\ Ax = b \\ e^T x + f > 0 \end{cases} \qquad \begin{aligned} Gy - hz &= \frac{Gx-h}{e^T x+f} \le 0 \\ Ay - bz &= \frac{Ax-b}{e^T x+f} = 0 \\ e^T x + f z &= 1 \\ z &\ge 0 \end{aligned}$$

目标函数值都 $= \dfrac{c^T x + d}{e^T x + f}$

2) 若 y, z 在 P_1 中可行

若 $z>0$, 则 $x = \dfrac{y}{z}$, then x 在 P_0 中可行且两问题在目标函数相同.

点　　　　　射线

若 $z=0$, 设 x_0 为 P_0 可行解, 则 $x = x_0 + ty$ 对 P_0 可行, $\forall t \ge 0$

$Gy \le 0$, $Ay = 0$, $e^T y = 1$,

$$Gx = Gx_0 + tGy \le h$$

$$Ax = Ax_0 + tAy = 1$$

$$e^T x + f = e^T x_0 + f + te^T y > 0$$

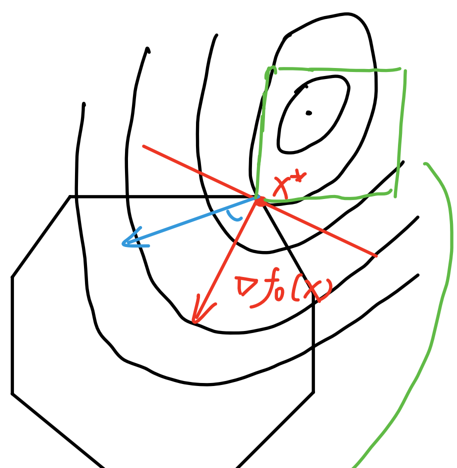
$$f_0(x) = f_0(x_0 + ty) = \frac{c^T x_0 + c^T t y + d}{e^T x + e^T t y + f} \xrightarrow{t \to \infty} c^T y$$

(3) 二次规划 Quadratic Programming

$$\min \ \frac{1}{2} x^T P x + q^T x + r, \quad P \in S_+^n$$

$$\text{s.t.} \quad Gx \le h$$

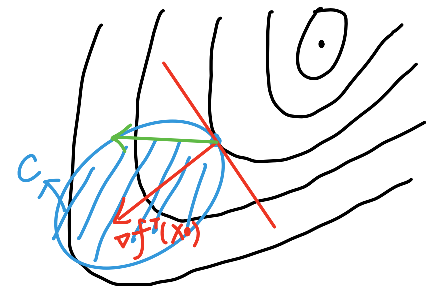
$$Ax = b$$



最优解不一定在边缘 or 顶点

还可能在集合内部.

(4) 二次约束二次规划 QCQP



$$\min \ \frac{1}{2}x^TPx + q^Tx + r$$

$$\text{s.t.} \ \frac{1}{2}x^TP_ix + q_i^Tx + r_i \le 0$$

$$Ax = b$$

例：带噪声的测量系统 $b = A\underline{x} + e$ 待测量向量

$$\hat{x} = \arg\min_x \|b - Ax\|_2 = \arg\min_x \|b-Ax\|_2^2$$

$$= \arg\min_x \ x^T\underbrace{A^TA}_{\text{半正定}}x - 2b^TAx + b^Tb$$ （无约束的QP问题）

$$= (A^TA)^{-1}A^{-1}b$$ （一阶偏导=0来计算）

（推广）若 x 稀疏 Lasso

$$\hat{x} = \arg\min_x \|b-Ax\|_2^2 + \lambda_0\|x\|_0$$ 非凸函数（用凸近似）

凸问题/QP ⇐ $$= \arg\min_x \|b-Ax\|_2^2 + \lambda_1\|x\|_1$$ （l_1-Regularized least squares）

（技巧：$x = x^+ - x^-$）☆ $$= \arg\min_{x^+,x^-} \|b - Ax^+ + Ax^-\|_2^2 + \lambda_1\|x^+ - x^-\|_1 + \lambda_1 1^Tx^+ - \lambda_1 1^Tx^-$$

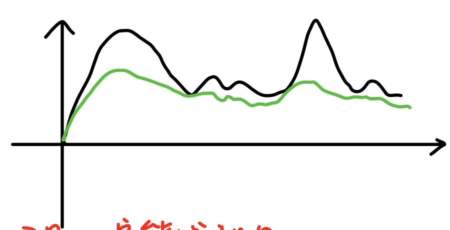
非光滑→光滑，可微

$$\text{s.t.} \ x^+, x^- \ge 0$$ （本质是二次规划问题）

（再推广）x 中元素幅度类似 ☆ l_2-regularized least square ☆ 岭回归

判断二次项是否一定是正定矩阵

$$\hat{x} = \arg\min_x \|b-Ax\|_2^2 + \lambda_2\|x\|_2^2$$



半正定 $\underline{A^TA} + \lambda_2 I$ $\lambda>0$，正定 ⇒ 确认二次规划

QCQP问题 ⇐ $$= \arg\min_x \|b-Ax\|_2^2$$ ⟺ 转化：$\forall \lambda_2 \ge 0$，一定能找到 θ

$$\text{s.t.} \ \|x\|_2^2 \le \underline{\theta}^{\ge 0}$$ 使两个问题最优解一样

例：投资组合问题

B	Bucket	
	beginning	ending
#1	x_1	$p_1 x_1$
⋮	⋮	⋮
#n	x_n	$p_n x_n$

$$\max \quad p_1 x_1 + \cdots + p_n x_n$$
$$\text{s.t.} \quad x_1 + \cdots + x_n \leq B$$
$$x_1, \cdots, x_n \geq 0$$

eg: $\bar{p}^T = [1.05,\ 1.05,\ 1]$, $\Sigma = \begin{bmatrix} 1.2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$$\min \quad x^T \Sigma^{-1} x$$
$$\text{s.t.} \quad \bar{p}^T x \geq r_{min}$$
$$1^T x = B$$
$$x \geq 0$$

确定收益下限 最小化风险

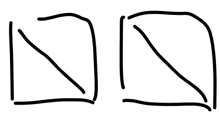
(1952, Markovitz) ⇒ 出名要趁早.

(5) 半正定规划 Semi-Definite Programming　　SDP

$$\min \quad tr(CX)$$ 矩阵空间的线性规划　　(优化矩阵)

$$\text{s.t.} \quad tr(A_i X) = b_i,\ i = 1, \cdots, p$$
$$X \succeq 0$$
$$X \in S_+^n,\ C \in R^{n \times n},\ A_i \in R^{n \times n},\ b_i \in R$$

eg: 特例. 对角矩阵 $diag\{x\}$



$$\min \quad (diag\{c\})^T diag\{x\}$$
$$\text{s.t.} \quad (diag\{A_i\})^T diag\{x\} = b_i,\ i = 1, \cdots, p$$
$$diag\{x\} \succeq 0$$

$\min \ c^T x$ （优化向量）

s.t. $x_1A_1+\cdots+x_nA_n \preceq B$ 半正定约束，左-右=非正定矩阵

$x\in R$, $B, A_1,\cdots,A_n \in S^k$, $c\in R^n$ 对称的k维的矩阵

例：$A(x)=A_0+x_1A_1+\cdots\cdots+x_nA_n$

$A_i\in R^{p\times q}$, $i=0,\cdots,n$, $x\in R^n$

谱范数 $\|A(x)\|_2$，即 $A(x)$最大的奇异值. → $A^{-1}(x)A(x)$

$\min \ \|A(x)\|_2$

标量

补充线代：$\|A(x)\|_2\le\sqrt{s}$, $s>0 \Leftrightarrow A^T(x)A(x)-sI \preceq 0$

 （用对角矩阵直观理解）

$\min \sqrt{s}$

凸函数 线性

s.t. $A^T(x)A(x)\preceq sI$ 二次项系数是半正定矩阵

$\Updownarrow$ 令 $\sqrt{s}=t$

$\min \ t$

s.t. $A^T(x)A(x)-t^2I\preceq 0$

$t\ge 0$ 非凸，很难

$\Longleftrightarrow$

$\min \ t$

s.t. $\begin{bmatrix} tI & A(x) \\ A^T(x) & tI \end{bmatrix} \succeq 0$

$t\ge 0$

$\Updownarrow$

$\min \ t$ 转化成标准的SDP问题

s.t. $Y=\begin{bmatrix} tI & A(x) \\ A(x) & tI \end{bmatrix} \succeq 0$

$Y\succeq 0$ 半正定

$t\ge 0$

(6)多目标优化问题

$$\min f_0(x)$$

$$\text{s.t. } f_i(x) \le 0, \quad i=1,\cdots,m$$

$$h_i(x) = 0, \quad i=1,\cdots,p$$

$f_0: R^n \to R^q$, $f_i: R^n \to R$, $h_i: R^n \to R$

例：投资组合问题 $\min$ Risk

$\min$ $-$income

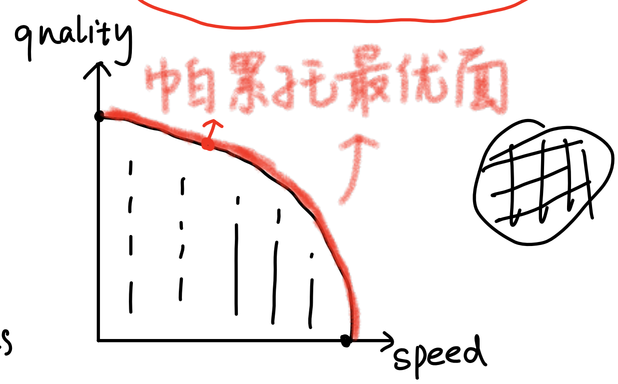
s.t. Resources

包含两个目标函数值和一个最优策略（帕累托最优值、点）

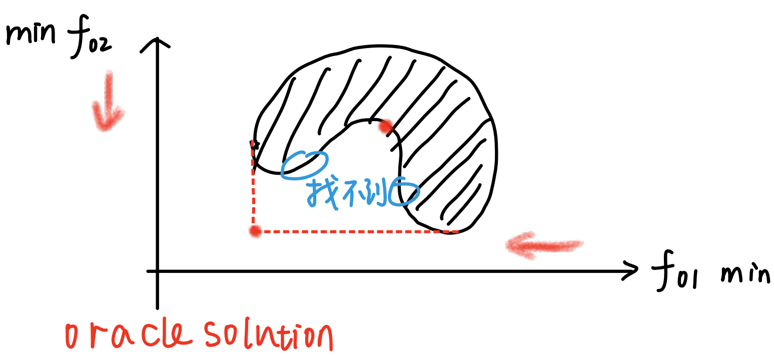
例：又好又快发展 $\min$ $-$speed

$\min$ $-$quality

s.t. resources



$\begin{cases} \text{Pareto optimal front：任意点，若能找到另一解，在某指标更优，则其他更差.} \\ \text{Pareto optimal value/point：对应的值和点.} \end{cases}$



不要试图解多目标优化

将其转化为单目标

若$\{f_0(x)\}$在R^k中为凸，$f_i(x)$为凸，$h_i(x)$为仿射，则必可以通过下述方法求得 Pareto front 中一点

$$\min \sum_{i=1}^{n} \lambda_i f_{0i}(x) \qquad \lambda_i \geq 0$$

$$\text{s.t. } f_i(x) \leq 0, \quad i=1,\cdots,m$$

$$h_i(x) = 0, \quad i=1,\cdots,p$$

通过遍历 $\{\lambda_i\}$ 可找出所有点.

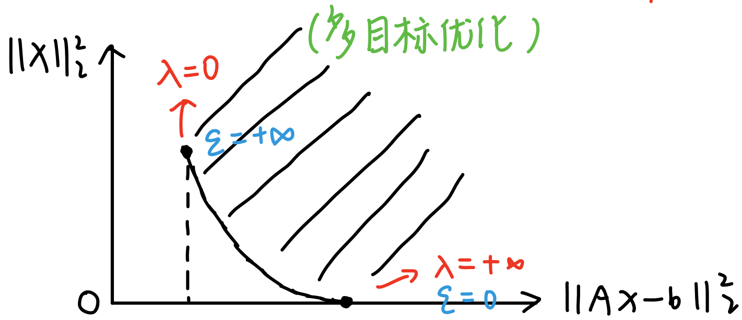
例: Ridge Regression

$b = Ax+e$, $A \in R^{m\times n}$, $b \in R^m$, $e \in R^m$, $x \in R^n$

$\min \|b-Ax\|_2^2$ ① 误差平方和最小 $\Leftrightarrow$ $\min \|b-Ax\|_2^2 + \lambda\|x\|_2^2$

$\min \|x\|_2^2$ ② x 的能量最小. (标准岭回归)

(多目标优化)



☆ 通过拉格朗日乘子联系起来.

有约束 → 无约束.

$\min \|b-Ax\|_2^2$ (岭回归的另一种形式)

$\text{s.t. } \|x\|_2^2 \leq \varepsilon$