Convex Optimization Basics

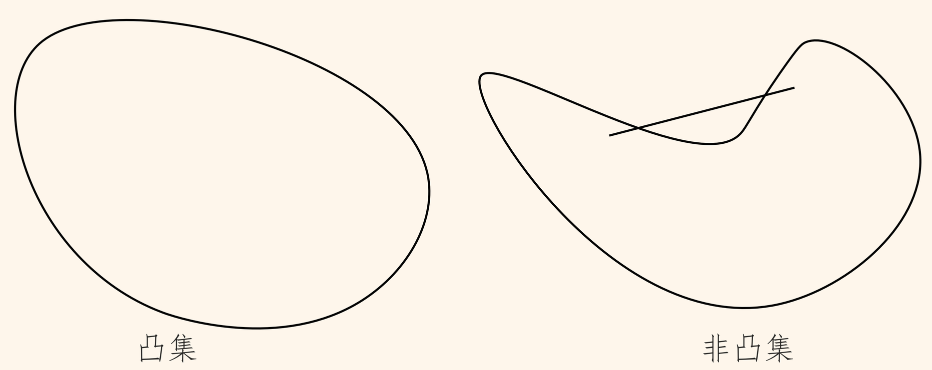
数据嗨客

凸集

如果一个集合C中的任意两点之间的线段也在该集合中，即集合C满足以下条件：

$$\forall x_1, x_2 \in C, \theta \in \mathbb{R}, 0 \le \theta \le 1$$
$$\theta x_1 + (1-\theta)x_2 \in C$$

那么集合C可以称为凸集.



推广至一般情况，对于参数θ的不同约束可得到

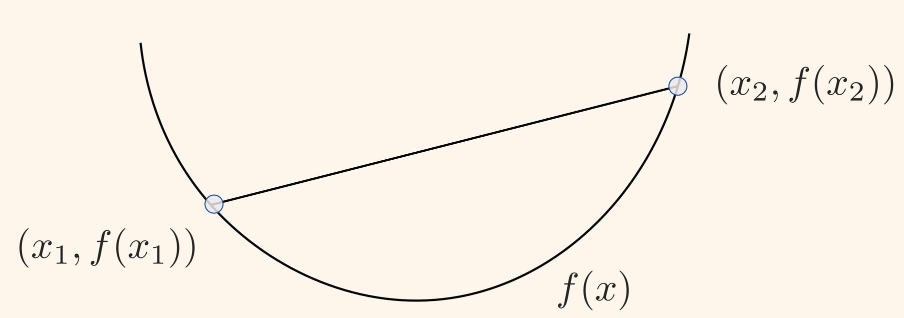
名称	约束
凸集	$\sum_i^n \theta_i = 1, \theta_i \ge 0$
仿射集	$\sum_i^n \theta_i = 1$
凸锥	$\forall \theta_i \ge 0$

凸函数

如果函数$f:\mathbb{R}^n \to \mathbb{R}$满足以下条件：

$$\forall x_1, x_2 \in dom(f), \theta \in \mathbb{R}, 0 \le \theta \le 1$$
$$f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta) f(x_2)$$

其中，$dom(f)$是凸集，那么函数f是一个凸函数.



严格凸 | 当$x_1 \ne x_2, 0 < \theta < 1$.

凹函数 | 当$-f$是一个凸函数.

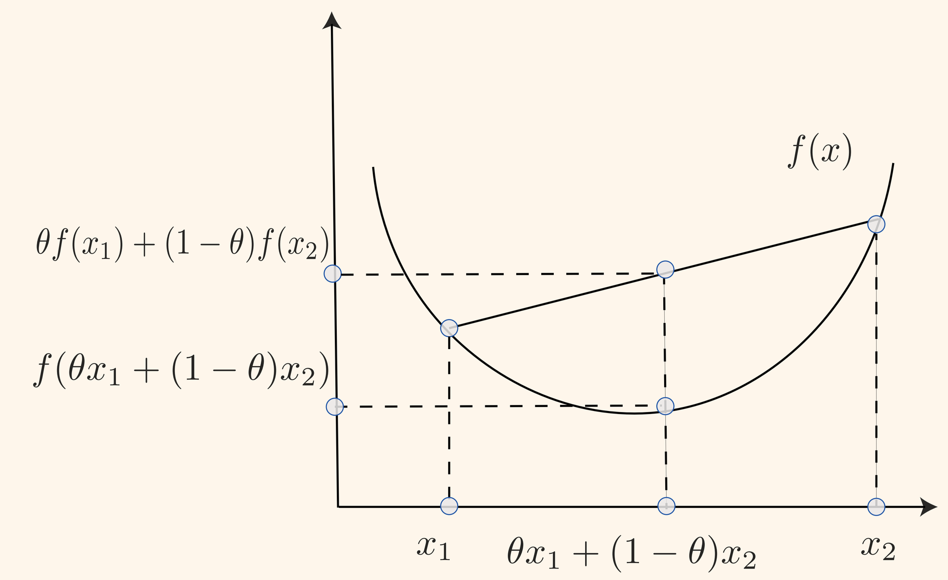
严格凹 | 当$-f$是一个严格凸函数.

推广至多个数据点，$\sum_i^n \theta_i = 1, \theta_i \ge 0$

$$f(\sum_{i=1}^n \theta_i x_i) \le \sum_{i=1}^n \theta_i f(x_i).$$

即$f(\mathbb{E}[x]) \le \mathbb{E}[f(x)]$，称为**Jensen不等式**.

凸函数的几何意义

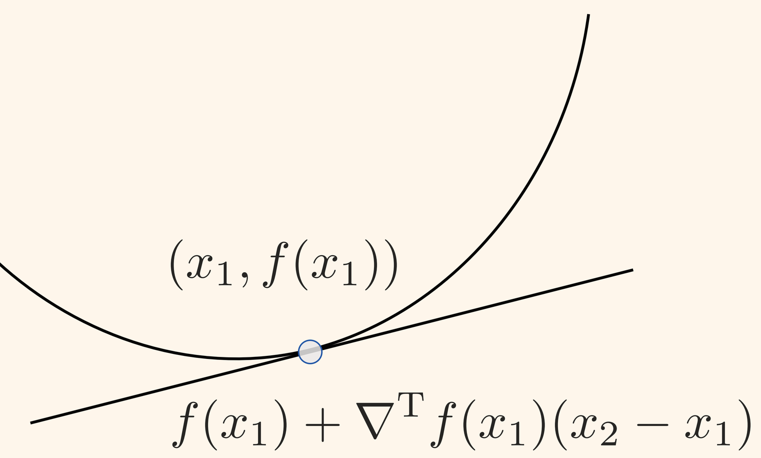


判断函数凸性的方式 ✓ 一阶条件 ✓ 二阶条件

一阶条件 | 如果一个函数$f:\mathbb{R}^n \to \mathbb{R}$在$dom(f)$上处处可导，满足以下条件：

$$\forall x_1, x_2 \in dom(f)$$
$$f(x_2) \ge f(x_1) + \nabla f(x_1)^{\mathrm{T}}(x_2 - x_1)$$

则f为凸函数



二阶条件 | 如果一个函数$f:\mathbb{R}^n \to \mathbb{R}$在$dom(f)$上的所有点二阶导数存在，并满足以下条件：

$$\forall x \in dom(f) \quad \nabla^2 f(x) \succeq 0$$

即如果f的Hessian矩阵为半正定矩阵，那么f为凸函数.

优化问题

标准形式 | 给定$x \in \mathbb{R}^n$，在满足条件$f_i(x) \le 0$以及$h_i(x) = 0$的所有x中寻找极小化$f_0(x)$的取值，即

$$\begin{array}{lll} \min\limits_x & f_0(x) & | \text{ 目标函数} \mathbb{R}^n \to \mathbb{R} \\ \text{s.t.} & f_i(x) \le 0, i = 1, \cdots m & | \text{ 不等式约束} \mathbb{R}^n \to \mathbb{R} \\ & h_i(x) = 0, i = 1, \cdots, p & | \text{ 等式约束} \mathbb{R}^n \to \mathbb{R} \end{array}$$

当$m = p = 0$，该问题称为**无约束问题**

如果问题定义域中的x满足约束条件，那么x是可行的；所有可行点的集合称为**可行集**

最优值 | 满足以下条件的$p^* \in [-\infty, \infty]$:

$$p^* = \inf\{f_0(x) | f_i(x) \le 0, h_i(x) = 0\}$$

✓ 如果问题不可行，则$p^* = \infty$

✓ 如果x_k为可行解，满足 $k \to 0, f_0(x+k) \to -\infty$

则$p^* = -\infty$，该优化问题**无下界**

最优点 | 最优点x^*是可行解，且$f_0(x^*) = p^*$

局部最优 | 可行解x为局部最优解，如果存在$R > 0$，使得以下条件成立

$$f_0(x) = \inf\{f_0(y) | f_i(y) \le 0, i = 1, \cdots, m,$$
$$h_i(y) = 0, i = 1, \cdots, p, \|y - x\|_2 \le R\}$$

松弛变量 | 引入$s \in \mathbb{R}^m$，将不等式约束替换为等式约束，原优化问题可以表示为

$$\begin{array}{ll} \min\limits_x & f_0(x) \\ \text{s.t.} & s_i \ge 0, i = 1, \cdots m \\ & f_i(x) + s_i = 0, i = 1, \cdots m \\ & h_i(x) = 0, i = 1, \cdots, p \end{array}$$

凸优化问题

凸优化问题 | 形如以下格式的优化问题

$$\begin{array}{ll} \min\limits_x & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \cdots, m \\ & a_i^T x = b_i, i = 1, \cdots, p \end{array}$$

其中，$f_0, \cdots, f_m$ 均为凸函数

凸优化问题的特点 | 满足以下条件：

✓ 目标函数必须是凸函数

✓ 不等式约束函数必须是凸函数

✓ 等式约束函数必须是放射函数

✓ **局部最优解**和**全局最优解**等价

如果目标函数**可微**，根据凸函数的定义，有

$$\forall x, y \in \mathrm{dom}(f)$$
$$f(y) \ge f(x) + \nabla f(x)^T (y - x)$$

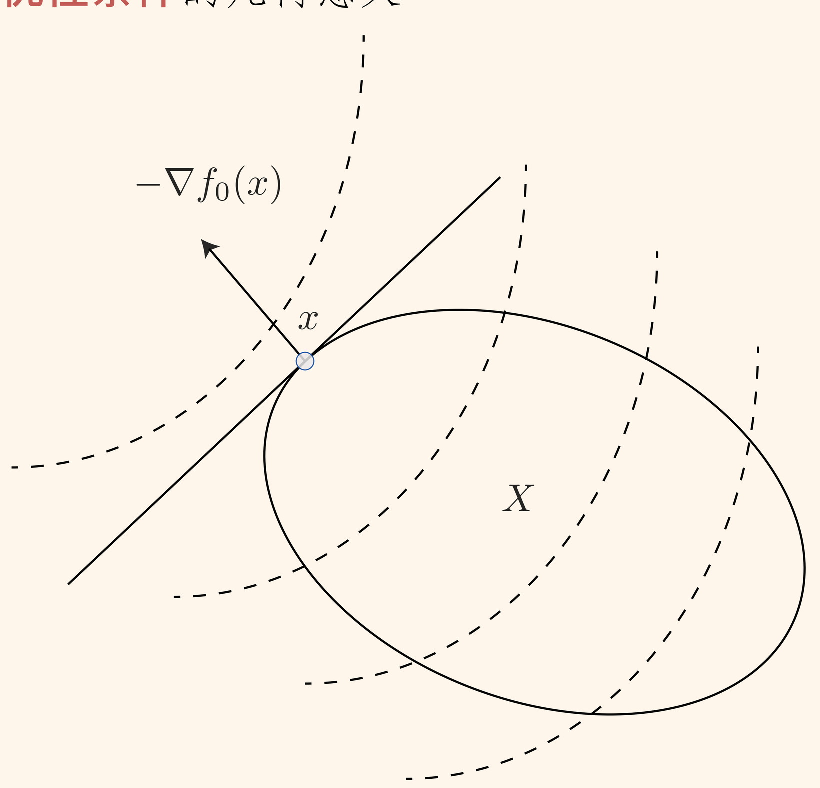
凸优化问题的**可行集**X可表示为

$$X = \{x | f_i(x) \le 0, h_i = 0\}$$

最优解 | x 为最优解，当且仅当满下条件

$$\begin{cases} x \in X \\ f(y) \ge f(x) + \nabla f(x)^T (y - x) \end{cases}$$

$f_0(x)$**最优性条件**的几何意义



常见的凸优化问题形式：

线性规划 | 仿射目标函数和约束函数.

二次规划 | 二次型目标函数和仿射约束函数.

二次约束二次规划 | 二次型目标、约束函数.

半定规划 | n维对称矩阵.

常见凸优化问题的数学描述

线性规划	二次规划	二次约束二次规划	半定规划
$G \in \mathbb{R}^{m \times n}, A \in \mathbb{R}^{p \times n}$	$G \in \mathbb{R}^{m \times n} A \in \mathbb{R}^{p \times n}, P \in \mathbb{R}^n_+$	$P_i \in \mathbb{S}^n_+, i = 1, \cdots, m$	$C, A_1, \cdots, A_p \in \mathbb{S}^n$
$\begin{array}{ll} \min\limits_x & c^T x + d \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{array}$	$\begin{array}{ll} \min\limits_x & \frac{1}{2} x^T P x + q^T x + r \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{array}$	$\begin{array}{ll} \min\limits_x & \frac{1}{2} x^T P_0 x + q_0^T x + r_0 \\ \text{s.t.} & Ax = b \\ & \frac{1}{2} x^T P_i x + q_i^T x + r_i \le 0 \end{array}$	$\begin{array}{ll} \min\limits_x & \mathrm{Tr}(CX) \\ \text{s.t.} & \mathrm{Tr}(A_i X) = b_i \\ & X \succeq 0 \end{array}$

Convex Optimization Basics

数 据 嗨 客

对偶

Lagrange函数 | 获取优化问题约束条件的加权和，放入原有目标函数，得到增广的目标函数

$\mathcal{L}:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^p\to\mathbb{R}$

$$\mathcal{L}(x,\lambda,\upsilon)=f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\upsilon_i h_i(x)$$

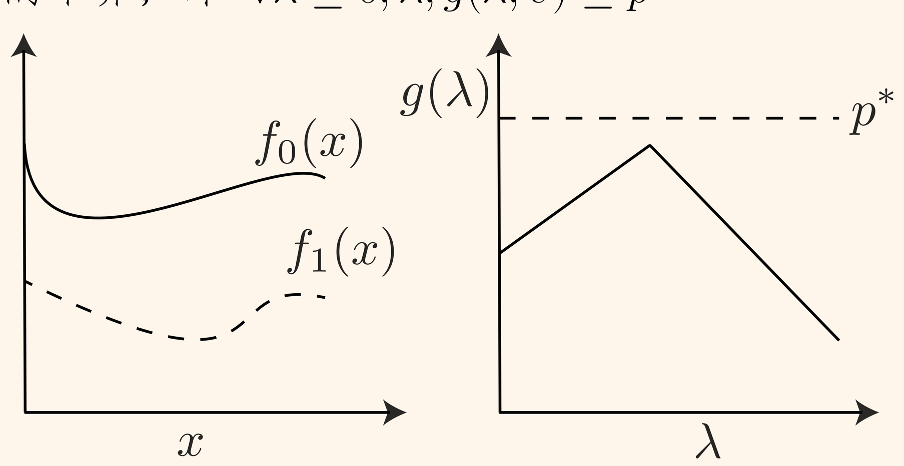
其中，λ_i,υ_i为对应的Lagrange乘子

Lagrange对偶函数 | 对于$\lambda\in\mathbb{R}^m,\upsilon\in\mathbb{R}^p$，Lagrange函数有关$x$可以取得的最小值

$g:\mathbb{R}^m\times\mathbb{R}^p\to\mathbb{R}$

$$g(\lambda,\upsilon)=\inf_x\Big(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\upsilon_i h_i(x)\Big)$$

最优值下界 | Lagrange对偶函数为原始优化问题最优值p^*的下界，即 $\forall\lambda\succeq 0,\lambda,g(\lambda,\upsilon)\le p^*$



对偶可行解给出的下界　　对偶函数的取值情况

共轭函数 | 给定函数$f:\mathbb{R}^n\to\mathbb{R}$，共轭函数$f^*$为

$$f^*(y)=\sup\Big(y^Tx-f(x)\Big)$$

Lagrange函数和共轭函数的联系 | 给定具有等式约束的优化问题 $\min\limits_x f(x)\quad \text{s.t.}\,x=0$

✓该问题的Lagrange函数为$\mathcal{L}=f(x)+\upsilon^Tx$

✓ Lagrange函数的对偶函数为

$$\begin{aligned}g(\upsilon)&=\inf_x\Big(f(x)+\upsilon^Tx\Big)\\&=-\sup_x\Big(-f(x)+(-\upsilon)^T\Big)=-f^*(-\upsilon)\end{aligned}$$

Lagrange对偶问题 | 原问题最优值p^*的下界，与λ,υ相关，可以描述为一个新的优化问题

$$\max_{\lambda,\upsilon}\ g(\lambda,\upsilon)\quad \text{s.t.}\ \lambda\succeq 0$$

由于目标函数为凹函数，约束集合为凸集，对偶问题仍是凸优化问题；对偶问题的凸性和原问题无关；如果(λ^*,υ^*)为对偶问题的最优解，那么(λ^*,υ^*)称为最优Lagrange乘子.

弱对偶性 | 如果d^*表示Lagrange对偶问题的最优取值，那么不等式 $d^*\le q^*$；即使原问题不是凸优化问题，该不等式依然成立.

最优对偶间隙 | p^*-d^*

如果原问题很难求解，那么基于弱对偶性，可以将求解原问题转化为求解其对偶问题

强对偶性 | 等式$d^*=p^*$成立，即最优对偶间隙为零；一般情况下，强对偶性不成立；如果原问题是一个凸问题，强对偶性通常成立

Lagrange对偶的鞍点解释

对偶性的极大极小描述 | 假设优化问题没有等式约束，可以有

$$\sup_{\lambda\succeq 0}\mathcal{L}(\lambda,x)=\sup_{\lambda\succeq 0}\Big(f_0(x)+\textstyle\sum_{i=1}^{m}\lambda_i f_i(x)\Big)=\begin{cases}f_0(x) & f_i(x)\le 0,i=1,\cdots,m\\ \infty & \text{其他}\end{cases}$$

基于上式，原问题的最优值可以表示为

$$p^*=\inf_x\sup_{\lambda\succeq 0}\mathcal{L}(x,\lambda)$$

同样地，对偶问题则可以表示为

$$d^*=\sup_{\lambda\succeq 0}\inf_x\mathcal{L}(x,\lambda)$$

弱对偶性可以使用下列不等式表示

$$\sup_{\lambda\succeq 0}\inf_x\mathcal{L}(x,\lambda)\le\inf_x\sup_{\lambda\succeq 0}\mathcal{L}(x,\lambda)$$

强对偶性可以使用下列等式表示

$$\sup_{\lambda\succeq 0}\inf_x\mathcal{L}(x,\lambda)=\inf_x\sup_{\lambda\succeq 0}\mathcal{L}(x,\lambda)$$

极大极小不等式 | $\forall f:\mathbb{R}^n\times\mathbb{R}^m\to R$，以下不等式成立，且 $W\subseteq\mathbb{R}^n,Z\subseteq\mathbb{R}^m$

$$\sup_{z\in Z}\inf_{w\in W}f(w,z)\le\inf_{w\in W}\sup_{z\in Z}f(w,z)$$

如果等式成立，f则满足鞍点性质

鞍点 | $\forall w\in W,z\in Z$，下列不等式成立

$$f(\tilde{w},z)\le f(\tilde{w},\tilde{z})\le f(w,\tilde{z})$$

$(\tilde{w},\tilde{z})$称为函数f的鞍点；如果x^*为原问题的最优点，λ^*为对偶问题的最优点，且强对偶性成立，那么(x^*,λ^*)为Lagrange函数的鞍点.

最优性条件

互补松弛性 | 若优化问题的强对偶性成立，给定x^*、(λ^*,υ^*)为原问题和对偶问题的最优解，有

$$\begin{aligned}f_0(x^*)&=g(\lambda^*,\upsilon^*)\\&=\inf_x\Big(f_0(x)+\sum_{i=1}^{m}\lambda_i^*f_i(x)+\sum_{i=1}^{p}\upsilon_i^*h_i(x)\Big)\\&=f_0(x^*)+\sum_{i=1}^{m}\lambda_i^*f_i(x)+\sum_{i=1}^{p}\upsilon_i^*h_i(x)\\&\le f_0(x^*)\end{aligned}$$

从上式可以得知$\sum_{i=1}^m\lambda_i^*f_i(x)=0$，即$\lambda_i^*f_i(x)=0$ 互补松弛条件可以表示为

$$\begin{cases}\lambda^*>0\Longrightarrow f_i(x^*)=0 & \text{或者}\\ \lambda^*<0\Longrightarrow f_i(x^*)=0\end{cases}$$

KKT条件 | 若优化问题的目标函数和约束函数可微，对偶间隙为零。那么，$\mathbb{L}(x,\lambda^*,\upsilon^*)$在$x=x^*$处取得最小值，则有

$$\nabla f_0(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla f_i(x^*)+\sum_{i=1}^{p}\upsilon_i^*\nabla h_i(x^*)=0$$

因此，列出优化问题的KKT条件

$$\begin{cases}f_i(x)\le 0,i=1,\cdots,m\\ h_i(x^*)=0,i=1,\cdots,p\\ \lambda_i^*\ge 0,i=1,\cdots,m\\ \lambda_i^*f_i(x^*)=0,i=1,\cdots,m\\ \nabla f_0(x^*)+\sum_{i=1}^m\lambda_i^*\nabla f_i(x^*)+\sum_{i=1}^p\upsilon_i^*\nabla h_i(x^*)=0\end{cases}$$

凸问题的KKT条件 | 如果原问题为凸问题，满足KKT条件的点也是原问题和对偶问题的最优解.

Slater条件 | $\exists\in\operatorname{relint}\mathcal{D}$，使得下式成立

$$f_i(x)<0,i=1,\cdots,m\quad Ax=b$$

其中$\mathcal{D}$为目标函数定义域，relint$\mathcal{D}$为定义域的相对内部. 当Slater条件成立且原问题为凸问题时，强对偶性成立

✓ 如果凸优化问题的目标函数和约束函数可微，且满足Slater条件，那么KKT条件为该问题最优性的充要条件

✓ 在实际的情形下，求解凸问题的方法可以转换为求解其KKT条件的方法.

凸优化算法概要

无约束优化 | 优化问题的目标函数$f:\mathbb{R}^n\to\mathbb{R}$二次可微，最优点满足充要条件

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

通过计算上述方程的解，得到优化问题的最优解；实际情形中，使用迭代算法获得最优解，计算数列$x^0,x^1,\cdots,x^k$，使得$k\to\infty$， $f(x^k)\to p^*$

下降方法 | 该方法产生一个数列$\{x^k\}k=1,\cdots,$，满足以下的等式

$$x^{k+1}=x^k+t^k\Delta x^k$$

其中，Δx表示搜索方向；t^k表示第k次迭代的步长；下降方法的搜索方向必须满足

$$\nabla f(x^k)^T\Delta x^k<0$$

下降方法算法框架 | 重复以下步骤，直至满足终止条件

✓ 确定搜索方向Δx

✓ 使用合适的搜索方法确定步长$t>0$

✓ 更新x，$x:=x+t\Delta x$

梯度下降法 | 负梯度为搜索方向 $\Delta x=-\nabla f(x)$

Newton法 | Newton法的搜索方向：

$$\Delta x_{Newton}=-\nabla^2 f(x)^{-1}\nabla f(x)$$

等式约束优化 | 优化问题的目标函数$f:\mathbb{R}^n\to\mathbb{R}$为二次连续可微凸函数，最优点满足充要条件

$$Ax^*=b,A\in\mathbb{R}^{p\times n},\nabla f(x^*)+A^T\upsilon^*=0$$

等式约束优化问题可以通过以下两种方式求解：

✓ 消除等式约束转化为等价的无约束优化问题

✓ 求解对偶问题，从对偶解中复原问题最优解

Newton法 | 假设KTT矩阵非奇异，确定Newton方向

$$\begin{bmatrix}\nabla^2 f(x) & A^T\\ A & 0\end{bmatrix}\begin{bmatrix}\Delta x_{newton}\\ w\end{bmatrix}=\begin{bmatrix}-\nabla f(x)\\ 0\end{bmatrix}$$

其中，w为优化问题的最优对偶变量

内点法 | 使用内点法求解不等式约束问题，实质是使用Newtorn法求解一系列等式约束问题

障碍法 | 顺序求解一系列等式约束或者无约束优化问题，每次使用获取的最新点作为求解下一个优化问题的起始点，该方法称为序列无语色输极小化技术，也称为障碍方法，或者路径跟踪法.